ISSN 1913-8989 (Print)
ISSN 1913-8997 (Online)

COMPUTER AND INFORMATION SCIENCE

Vol. 3, No. 1
February 2010

Canadian Center of Science and Education

Editorial Board

Computer and Information Science

www.ccsenet.org/cis

Vol. 3, No. 1

February 2010

Contents

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Contents

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Improvement in Network Lifetime for On-Demand Routing in Mobile Ad hoc Networks Using either On-Demand Recharging or Transmission Power Control or Both

Natarajan Meghanathan (Corresponding Author)
Department of Computer Science, Jackson State University
P. O. Box 18839, 1400 John R. Lynch Street, Jackson, MS 39217, USA
Tel: 01-601-979-3661 E-mail: natarajan.meghanathan@jsums.edu

Levon Paul Judon, Jr.
Department of Computer Science, Jackson State University
E-mail: solidp2001@yahoo.com

Abstract

Given a fixed energy budget for the operation of a mobile ad hoc network (MANET), on-demand recharging is the technique of charging the nodes initially with identical, but reduced energy level called the recharge quantum, and then recharging the nodes with the recharge quantum of energy whenever the energy level at a node goes below a threshold level. Transmission power control is the technique of adjusting the transmission power at a sender node depending on the distance to the receiver node. The high-level contribution of this paper is a simulation-based analysis of the network lifetime obtained for each of the following four scenarios: [a] No power control, No on-demand recharging; [b] Power control, but no on-demand recharging; [c] On-demand recharging, but no power control and [d] Both power control and on-demand recharging. Network lifetime is defined as the time of first node failure due to the exhaustion of energy level at the node and the inability to further charge the node. The on-demand routing protocols studied are: Dynamic Source Routing (DSR), Flow-Oriented Routing Protocol (FORP) and the Min-Max Battery Cost Routing (MMBCR) algorithm run on the top of DSR. We illustrate the improvement obtained in network lifetime as we transition from scenarios [a] through [d]. Simulation results illustrate that scenarios involving on-demand recharging ([c] and [d]) yield a higher network lifetime than scenarios [a] and [b]. When we operate the network with both on-demand recharging and power control, we obtain the maximum improvement in network lifetime. The percentage of the supplied energy that has been consumed in the network at the time of first node failure for each of the four scenarios and the three routing protocols is also measured to illustrate the effectiveness of on-demand recharging in maximizing the usage of the available energy budget.

Keywords: On-demand Recharging, Mobile Ad hoc Networks, Transmission Power Control, Network Lifetime, Routing Protocols

1. Introduction

A mobile ad hoc network (MANET) is a dynamic distributed system of wireless nodes where in the nodes move independent of each other. MANETs have several operating constraints such as: limited battery charge per node, limited transmission range per node and limited bandwidth. Routes in MANETs are often multi-hop in nature. Packet transmission or reception consumes the battery charge at a node. Nodes forward packets for their peers in addition to their own. In other words, nodes are forced to expend their battery charge for receiving and transmitting packets that are not intended for them. Given the limited energy budget for MANETs, inadvertent over usage of the energy resources of a small set of nodes at the cost of others can have an adverse impact on the node lifetime.

There exist two classes of MANET routing protocols: proactive and reactive. The reactive (also called on-demand) routing protocols tend to be more economical in terms of energy and bandwidth usage (Broch, Maltz, Johnson, Hu & Jetcheva, 1998; Johansson, Larsson, Hedman, Mielczarek & Degermark, 1999) in dynamically changing scenarios, characteristic of MANETs. Hence, in this paper, we restrict ourselves to the class of on-demand routing protocols. We use three on-demand routing protocols: Dynamic Source Routing (DSR – Johnson, Maltz & Broch, 2001), Flow-Oriented Routing Protocol (FORP – Su & Gerla, 1999) and Min-Max Battery Cost Routing (MMBCR – Toh,

2001). DSR, FORP and MMBCR are respectively selected as representatives of the minimum-hop, stable path and power-aware routing strategies. In an earlier work (Meghanathan, 2008), it has been observed that FORP discovers the sequence of most stable (long-living) routes among the currently available MANET stable path routing protocols. DSR discovers minimum-hop routes with a lower routing overhead (Broch, Maltz, Johnson, Hu & Jetcheva, 1998). MMBCR is a power-aware routing algorithm that attempts to equally distribute energy consumption to all the nodes in an ad hoc network and takes into consideration the available energy level at the nodes before route selection. The residual energy of a path is the minimum of the energy levels of the intermediate nodes on the path and MMBCR chooses the path with the maximum residual energy.

On-demand routing protocols discover a route between two nodes, only when required, using a Route Request-Reply cycle involving flooding of route discovery packets. An optimum route, depending on the route selection principles of the routing protocol, is discovered as a result of flooding. To reduce the route discovery overhead, on-demand routing protocols adopt a Least Overhead Routing Approach (LORA) wherein the discovered route is used as long as it exists. With LORA, all nodes are unfortunately not equally used and there is a tendency to overuse a few "centrally located" nodes (Toh, 2002). Especially, protocols that are designed to use minimum hop routes or stable long-living routes are more likely to unequally exhaust the battery charge of the nodes. Also, at low mobility, a node is highly likely to remain in a source-destination route for a sufficiently long time. In mobile ad hoc networks, the nodes (for example: laptops, personal digital assistants) typically belong to different users and hence the loss of connectivity to even one node is significant. Assuming that the failure of a node occurs only due to the node running out of its battery charge, we define network lifetime to be the time of first node failure.

Traditionally, MANET routing protocols have often been studied without recharging where in the fixed energy budget for the network is divided equally and all nodes are charged upfront with identical energy levels. Meghanathan & Farago (2005) introduced the idea of on-demand recharging (referred here after simply as recharging) in MANETs and showed that given a fixed energy budget, supplying the energy in quanta to the nodes over time, rather than all of it initially, can yield significant improvement in network lifetime. Meghanathan & Farago (2005) assumed the use of a fixed transmission power per hop and did not employ transmission power control. Transmission power control (Gomez & Campbell, 2004) is a power-saving technique in which the transmission power at the sender is adjusted depending on the distance to the receiver node. If two nodes are closer to each other, then the transmission power needed to send a data packet between each other can be relatively smaller compared to the transmission power needed to send data between two nodes that are far away from each other. Transmission power control (referred here after simply as power control) is one of the commonly used energy-efficient techniques to optimize network lifetime in MANETs.

In this paper, we explore the use of both on-demand recharging and transmission power control to maximize network lifetime for on-demand MANET routing protocols. We conduct extensive simulations of DSR, FORP and MMBCR under each of the following four scenarios: [a] No power control, No recharging; [b] Power control, but no recharging; [c] Recharging, but no power control and [d] Both power control and recharging. Our objective is to study the improvement in network lifetime brought about by scenarios [b], [c] and [d] over scenario [a] and with respect to each other. We observe scenario [c] (recharging only) can often yield a network lifetime relatively larger than that has been obtained with scenario [b] (power control only). We also observe that operating the network with both recharging and power control leads to a significantly larger network lifetime compared to the other three scenarios.

The rest of the paper is organized as follows: Section 2 briefly reviews the idea of on-demand recharging for ad hoc networks. Section 3 reviews the DSR, FORP and MMBCR routing protocols. Section 4 explains the simulation environment. Section 5 illustrates the simulation results obtained for each of the four scenarios [a] through [d] and interprets them. Section 6 presents the conclusions.

2. On-Demand Recharging

Given a fixed energy budget for the operation of a MANET, the basic idea of on-demand recharging is to charge the nodes initially with identical, but reduced energy level called the recharge quantum, and then recharge a node only when it is about to run out of energy. On-demand recharging is basically a dynamic resource allocation strategy for networks with a common fixed supply of resources whose consumption across the network is unpredictable. A real-world analogical example would be the case of an investor with a fixed amount of money adjusting his/her investments in stock market according to the changing trends in the market value of the stocks.

The amount of energy added per recharge operation is called the recharge quantum and under a fixed energy budget, it is inversely related to the total number of recharges that could be done in the network before a node fails. The smaller the recharge quantum, the larger is the network lifetime and larger is the percentage of the fixed energy budget usefully consumed within the network before the time of first node failure. In this paper, in order to determine the maximum possible improvement in network lifetime, we use a recharge quantum of 1 Joule for every recharge operation. To avoid any change in performance and unnecessary route transitions, we do not let the energy level of a node to reach zero, before adding a recharge quantum. If the energy level at a node goes below the expiration threshold energy level, the

node is supplied a recharge quantum of energy. The expiration threshold energy level is chosen in such a way that the node is always available for packet forwarding. We use an expiration threshold level of 0.05 Joules in our simulations.

On-demand recharging need not mean just physical recharging alone (Meghanathan & Farago, 2005). For example, it could even emulate the on-demand deployment of sensors (from an initial fixed supply) to effectively track the movement of a mobile hotspot. A single phase uniform random distribution of the sensor nodes may unnecessarily increase the density of the network and also cannot guarantee that all hotspots are effectively covered. Sensors have to be deployed in a sequence of phases based on the information gathered from the sensors already deployed. On-demand sensor deployment can extend the lifetime of the network by effectively tracking a dynamically moving hotspot with immobile sensor nodes and at the same time maintain the density of the network within a reasonable magnitude.

3. Review of the MANET Routing Protocols

In this section, we provide a brief overview of the minimum-hop based Dynamic Source Routing (DSR) protocol, stability-based Flow-Oriented Routing Protocol (FORP) and the power-aware Min Max Battery Cost Routing (MMBCR) algorithm. In this paper, we implemented MMBCR on the top of DSR.

3.1 Dynamic Source Routing Protocol

The unique feature of DSR (Johnson, Maltz & Broch, 2001) is source routing: data packets carry information about the route from the source to the destination in the packet header. As a result, intermediate nodes do not need to store up-to-date routing information in their forwarding tables. This avoids the need for beacon control neighbor detection packets that are used in the stability-oriented routing protocols. Route discovery is by means of the broadcast query-reply cycle. A source node *s* wishing to send a data packet to a destination *d*, broadcasts a Route-Request (RREQ) packet throughout the network. The RREQ packet reaching a node contains the list of intermediate nodes through which it has propagated from the source node. After receiving the first RREQ packet, the destination node waits for a short time period for any more RREQ packets, then chooses a path with the minimum hop count and sends a Route-Reply Packet (RREP) along the selected path. If any RREQ is received along a path whose hop count is lower than the one on which the RREP was sent, another RREP would be sent on the latest minimum hop path discovered. To minimize the route acquisition delay, DSR lets intermediate nodes to promiscuously listen to the channel, store the learnt routes (from the RREQ and data packets) in a route cache and use these cached route information to send the RREP back to the source. We do not use this feature as promiscuous listening dominates the energy consumed at each node and DSR could still effectively function without promiscuous listening and route caching. Also, in networks of high node mobility, cached routes are more likely to become stale, by the time they are used.

3.2 Flow-Oriented Routing Protocol (FORP)

FORP (Su & Gerla, 1999) utilizes the mobility and location information of nodes to approximately predict the expiration time (LET) of a wireless link. The minimum of LET values of all wireless links on a path is termed as the route expiration time (RET). The route with the maximum RET value is selected. Each node is assumed to be able to predict the LET values of each of its links with neighboring nodes based on the information regarding the current position of the nodes, velocity, the direction of movement, and transmission range. FORP assumes the availability of location-update mechanisms like Global Positioning System (GPS – Hofmann-Wellenhof, Lichtenegger & Collins, 2004) to identify the location of the nodes and also assumes that the clocks across all nodes are synchronized. Route discovery is similar to the flooding-based query-reply cycle described in Section 3.1, with the information propagated in the RREQ packet being the predicted LET of each link in a path.

RREQ packets are propagated as described before, from the source node *s* to the destination node *d*. The information recorded in this case by a node *j* receiving a RREQ packet from a node *i* is the predicted lifetime of the link *i-j*. The destination *d* will receive several RREQ packets with the predicted link lifetimes in the paths traversed being listed. The residual expiration time of a path is the minimum of the predicted lifetimes of its constituent links. The *s-d* path that has the maximum predicted residual lifetime is then selected. If more than one path has the same maximum predicted residual lifetime, the tie is broken by selecting the shortest (minimum hop path) of such paths.

3.3 Min-Max Battery Cost Routing (MMBCR)

The residual battery charge of an *s-d* path is the minimum of the battery charges of the intermediate nodes of the path. The MMBCR algorithm (Toh, 2001) chooses the *s-d* path with the largest residual battery charge. The route selection metric recorded in an *s-d* path is the residual battery charge (available battery charge) of each of the intermediate nodes on the *s-d* path through which the RREQ packet got forwarded. The residual battery charge of a path is the minimum of the residual battery charge of all the constituent intermediate nodes, other than the source and destination, of the path. From all the *s-d* paths learnt, the destination chooses the path that has the largest residual battery charge value.

4. Simulation Environment

The simulations of DSR, FORP and MMBCR were conducted in ns-2: version 2.28 (Fall & Varadhan, 2001). We used

the implementation of DSR that comes with ns-2 and implemented FORP and MMBCR in ns-2. The network dimensions are 1000m x 1000m. The transmission range of each node is 250m. We vary the network density by conducting simulations with 25 nodes (low density network with an average of 5 neighbors per node) and 50 nodes (high density network with an average of 10 neighbors per node). The fixed energy budget for the low-density network and high-density network is 2500 Joules and 5000 Joules respectively. If we do not conduct any recharging (scenarios [a] and [b]), each node in these networks gets upfront 100 Joules of energy. When we conduct recharging (scenarios [c] and [d]), each node gets an initial energy of 1 Joule to start with and whenever the energy level at a node goes below the expiration threshold energy level of 0.05 Joules, the node is charged with the recharge quantum of 1 Joule, taken from the available energy budget for the network. For all the four scenarios, whenever a node needs energy and the available energy budget is zero, we declare the node has failed. Note in scenarios [a] and [b], the available energy budget is zero after the nodes are charged fully upfront (i.e., without any recharging). The time of first node failure is recorded as the network lifetime.

Besides network lifetime, we also measure the percentage of energy consumed in the network at the time of first node failure. This metric is basically the ratio, expressed in percentage, of the total energy actually consumed across all the nodes in the network to the initial fixed energy budget. This metric is especially of interest in scenarios [a] and [b] where there is no recharging. As illustrated in the performance figures 2.1 through 2.8, one can observe that for both DSR and FORP, only 35% - 60% of the energy supplied to the nodes have been consumed at the time of first node failure and even for MMBCR, the percentage of energy consumed at the time of first node failure is mostly within 75-80%. On-demand recharging actually exploits this unfairness in the routing protocols and attempts to efficiently use the fixed energy budget only at nodes that need energy.

Traffic sources are continuous bit rate (CBR). Number of source-destination (*s-d*) sessions used is 15 (low traffic load) and 30 (high traffic load). The starting times of the *s-d* sessions is uniformly distributed between 1 to 20 seconds. Data packets are 512 bytes in size; the packet sending rate is 4 data packets per second. While distributing the source-destination roles for each node, we saw to it that a node does not end up as source of more than two sessions and also not as destination for more than two sessions.

While DSR does not use beacons, FORP and MMBCR require the use of beacon control messages for neighborhood awareness. In these two protocols, beacons are periodically broadcast by a node within the transmission range to let each node advertise its presence to neighbors and learn about the neighborhood. In our simulations, beacons are exchanged for every one second. In the case of FORP, each node sends information about its location and current velocity in the beacons. Each node keeps track of the previously advertised location of its neighbor nodes. This will help to determine the direction in which the neighbor node is moving. For MMBCR, each node includes in the beacon packets information about the current battery charge available at the node.

4.1 MAC Layer Model

The MAC layer uses the distributed co-ordination function (DCF) of the IEEE Standard 802.11 (Bianchi, 2000) for wireless LANs. For scenarios with transmission power control, the channel negotiation is dealt as follows: the sender node transmits the Request-To-Send (RTS) packet with a transmission power corresponding to the fixed maximum transmission range of 250m. The receiver node on receiving the RTS packet, estimates the distance to the sender based on the strength of the signal received for the RTS packet. The receiver node includes this distance information in the Clear-To-Send (CTS) packet, which is transmitted with the transmission power that is just enough to reach the sender with signal strength above the receiving signal strength threshold of $3 * 10^{-10}$ W. The sender node on receiving the CTS packet uses the distance information in the CTS packet and estimates the transmission power that would be just sufficient to send the DATA packet to the receiver node. The transmission power employed is calculated using the formula (Park & Sivakumar, 2002; Meghanathan, 2008): $1.1182 + 7.2 * 10^{-11}(d)^4$, d – distance between the transmitter and the receiver. The receiver node upon successfully receiving the DATA packet will send an ACK packet using the transmission power that was spent to send the CTS packet. The neighbors of the receiver that had earlier received the CTS packet and the neighbors of the sender that had received the RTS packet are free to start their own channel negotiations after they receive the ACK packet within a certain time period.

The neighbor nodes of the sender that received the RTS packet and not the CTS packet within a certain time are free to start having their own transmissions while the DATA packet transmission is taking place. Similarly, the neighbors of the receiver that did not receive the RTS packet are also free to start having their own transmissions while the DATA packet transmission is taking place. Thus, transmission power control also helps us to increase the usage of bandwidth and minimize the delay in packet transmissions.

4.2 Energy Consumption Model

The energy consumption at a node in an ad hoc network can be divided into three categories: (i) Energy utilized for transmitting a message, (ii) Energy utilized for receiving a message and (iii) Energy utilized in idle state. It has been

shown that in the presence of overhearing, no real optimization in the energy consumption or node lifetime can be achieved (Kim, Garcia-Luna-Aceves, Obraczka, Cano & Manzoni, 2003). That is, the energy consumption at a node would be dominated by the energy lost when the node is in idle state (also referred to as being in the promiscuous mode). Thus, in this paper, we do not consider the energy lost in the idle state and focus only on the energy consumed during the transmission and reception of messages (the DATA packets, the MAC layer RTS-CTS-ACK packets and the periodic beacons), and the energy consumed due to route discoveries. We model the energy consumed due to broadcast traffic and point-to-point traffic as linear functions of the packet transmission time, network density, transmission and reception powers per hop. A similar linear modeling for energy consumption has also been earlier used (Feeney, 2001; Meghanathan, 2006). For simulations without transmission power control, the fixed transmission power per hop is 1.4W. For simulations with transmission power control, the transmission power per hop is dynamically adjusted using the formula $1.1182 + 7.2*10^{-11}*(d)^4$, which includes power required to drive the circuit (1.1182W) and transmission power from the antenna computed using the two-ray ground reflection model (Fall & Varadhan, 2001) and distance *d* between the sender and receiver. The reception power per hop is fixed for all situations and it is 0.967W. The above values/ equations for transmission and reception power per hop are commonly used (Feeney, 2001; Meghanathan, 2006; Park & Sivakumar, 2002) in simulation studies.

4.3 Node Mobility Model

The node mobility model used is the Random Waypoint model (Bettstetter, Hartenstein & Perez-Costa, 2004), a widely used mobility model in MANET simulation studies. According to this model, each node starts moving from an arbitrary location to a randomly selected destination location at a speed uniformly distributed in the range $[v_{min},\ldots,v_{max}]$. Once the destination is reached, the node may stop there for a certain time called the pause time and then continue to move by choosing a different target location and a different velocity. In this paper, we set $v_{min} = 0$, and each node chooses speed uniformly distributed between 0 and v_{max}. The v_{max} values used are 5 m/s (representing low node mobility condition) and 50 m/s (representing high node mobility condition). Pause time is 0 seconds. Note that as the node velocities for each direction change are randomly picked from [0 … 5 m/s] and [0 … 50 m/s], the average velocity per node in the low node mobility and high node mobility conditions is about 2.5 m/s and 25 m/s respectively. These two values translate to 6 miles per hour and 55 miles per hour representing vehicular speeds in a school environment and interstate highways respectively.

5. Simulation Results

Each data point in Figures 1 and 2 is an average of data collected using 5 mobility trace files and 5 sets of randomly selected 15 and 30 *s-d* sessions.

5.1 Network Lifetime

Considering scenarios [b] and [c], wherein only one of the two energy-efficient techniques is implemented, we observe that on-demand recharging yields a significant improvement in network lifetime compared to transmission power control. On average, when we implement only on-demand recharging (scenario [c]), DSR, FORP and MMBCR show an improvement of 86%, 92% and 36% respectively and when we implement only power-control (scenario [b]), DSR and FORP show an improvement of only 52% and 59% respectively. The network lifetime for MMBCR when operated with power control is even slightly lower than when operated without power control (by a factor of 12%).

When we operate the network with both on-demand recharging and transmission power control (scenario [d]), DSR, FORP and MMBCR showed an improvement of (a) about 150-200%, 150-300% and 50-70% respectively compared to the scenario of operating without both on-demand recharging and transmission power control, (b) about 75-85%, 75-140% and 78-100% respectively compared to the scenario of operating only with transmission power control and (c) about 35-64%, 54-90% and 18-20% respectively compared to the scenario of operating only with on-demand recharging. So, even though the absolute network lifetime is relatively lower, FORP incurred the maximum improvement in network lifetime and this can be attributed to the supply of energy to the nodes lying on the long-living stable paths as long as the available energy budget is positive.

When we operate the network without both power control and on-demand recharging (scenario [a]), MMBCR incurs the largest network lifetime for low-density networks and DSR incurs the largest network lifetime for high-density networks. FORP performs very poorly in this scenario for both low-density and high-density networks. FORP, being a stable path routing protocol, tends to overuse nodes lying on the stable path at the cost of others. To a certain extent, DSR also overuses nodes (especially in low-density networks) lying in the center of the network compared to nodes in the periphery.

MMBCR shows the least improvement in network lifetime and DSR and FORP show the a relatively larger improvement in network lifetime as we move from scenarios of no power control and no on-demand recharging towards scenarios of power control and on-demand recharging. This can be attributed to the power-aware nature of MMBCR. The gain in network lifetime with the introduction of the energy-efficient techniques is relatively low with MMBCR.

On the other hand, the performance of DSR and FORP with respect to network lifetime improved as we introduce scenarios involving power control and/ or on-demand recharging. For both these non power-aware routing protocols, with on-demand recharging, nodes that are being used more frequently receive more energy and nodes that are used less frequently receive less energy. DSR preferentially uses nodes that lie in the center of the network to be part of several minimum-hop routes and FORP preferentially uses nodes maintain stable links with its neighbors to be part of several stable routes. As MMBCR inherently attempts to ensure fairness of node usage at the cost of hop count and route stability, with on-demand recharging, energy is still likely to be equally distributed and used across all the nodes.

For a given level of node mobility and network density, as we increase the offered traffic load to 30 *s-d* pairs, the network lifetime is reduced by almost half of the value observed with 15 *s-d* pairs for each of the four scenarios [a] through [d]. Nevertheless, the rate of improvement in network lifetime as we transition from scenarios [a] through [d] remains the same, irrespective of the offered traffic load.

For a given level of network density and offered traffic load, in general, for scenarios [a] and [b], the routing protocols yield a relatively lower network lifetime when operated under low mobility conditions rather than the high mobility conditions. Even though, frequent route changes result in higher energy consumption, route changes are preferred to yield diversity in the routes and achieve better fairness of node usage. But, as we transition to scenario [d], the routing protocols yield a larger network lifetime under low mobility conditions rather than high mobility conditions. Under low-mobility conditions, the energy consumed due to route discoveries is relatively low and with the introduction of on-demand recharging, nodes that are being heavily used get more energy compared to nodes that are lightly used.

For a given level of node mobility and offered traffic load, under each of the scenarios [a] through [d], we observe the network lifetime with DSR as the routing protocol increases with increase in network density (from 25 nodes to 50 nodes) and the network lifetime with FORP decreases with increase in network density. With increase in network density, the number of nodes in the center of the network increases and DSR has several candidate nodes to be in the minimum hop paths. On the other hand, FORP makes use of the increase in the number of nodes to find paths with a relatively longer lifetime compared to those discovered in low-density networks. Hence, even though the FORP route lifetime increases with increase in network density, the nodes that have been used in the stable path for a longer time, will fail early. With MMBCR (as it attempts to balance the energy consumption load across the network), the network lifetime remains almost the same or is slightly high with increase in network density, for a given level of node mobility and offered traffic load.

5.2 Percentage of Energy Consumed in the Network

With on-demand recharging (either in the absence or in the presence of power control, representing scenarios [c] and [d] respectively), for each of the routing protocols, we observe that almost the entire fixed energy budget is completely consumed (at least 99.5%) by the nodes in the network. In the absence of on-demand recharging (scenarios [a] and [b]), the fixed energy budget is equally divided and all the nodes are charged upfront. Due to the stochastic nature of ad hoc networks and random node movements, many nodes are lightly used and have abundant energy left at them during the time of first node failure resulting from the exhaustion of the supplied energy.

Overall, we observe that in the absence of on-demand recharging, MMBCR has the highest percentage of energy consumed in the network and this is obtained in the absence of power control (scenario [a]). Even though the use of power control (scenario [b]) leads to increase in the network lifetime for MMBCR, the percentage of energy consumed in the network at the time for first node failure is less than that obtained in the absence of power control. On the other hand, the energy consumed in the network for DSR in the presence of power control is greater than that observed in the absence of power control. The percentage of energy used in the network remains almost the same for FORP for both the scenarios of operating with and without power control.

For a given level of node mobility and network density, for each of the scenarios [a] and [b], as we increase the offered traffic load to 30 *s-d* pairs, we observe that the percentage of energy consumed in the network remains the same for both FORP and DSR as observed for 15 *s-d* pairs. The percentage of energy consumed in the network for MMBCR at higher offered traffic load of 30 *s-d* pairs is 20% more than that incurred at a lower traffic load of 15 *s-d* pairs.

For a given level of network density and offered traffic load, the percentage of energy consumed in the network for both DSR and FORP at high node mobility conditions is about 30% and 15% more than that obtained at low node mobility conditions for scenarios [a] (no power control, no on-demand recharging) and [b] (power control, no on-demand recharging) respectively. The relative decrease in the magnitude of energy consumption with scenario [b] can be attributed to the reduction in the transmission power per hop. In the case of MMBCR, the percentage of energy consumed in the network at high node mobility conditions is about 35% more than that consumed at low node mobility conditions for both scenarios [a] and [b].

For a given level of node mobility and offered traffic load, under each of the scenarios [a] and [b], as we increase the network density from 25 to 50 nodes, we notice that the percentage of energy consumed in the network for FORP at

high network density is about 20-25% lower than that consumed at low network density. On the other hand, for both DSR and MMBCR, the percentage of energy consumed in the network at high network density is about 5% more than that consumed at low network density.

6. Conclusions and Future Work

The simulation results highlight the improvement in network lifetime obtained with on-demand recharging compared to transmission power control. Due to the stochastic nature of ad hoc networks and random node movements, many nodes are lightly used and have abundant energy left at them during the time of first node failure due to exhaustion of the supplied energy. On-demand recharging actually exploits this unfairness in the routing protocols and attempts to efficiently use the fixed energy budget by providing energy only at nodes that need energy. With on-demand recharging (either in the absence or in the presence of power control representing scenarios [c] and [d] respectively), for each of the routing protocols, we observe that almost the entire fixed energy budget is completely consumed (at least 99.5%) by the nodes in the network.

MMBCR shows the least improvement in network lifetime and DSR and FORP show a relatively larger improvement in network lifetime as we move from scenarios of no power control and no on-demand recharging towards scenarios of power control and on-demand recharging. This can be attributed to the power-aware nature of MMBCR. The gain in network lifetime with the introduction of the energy-efficient techniques is relatively low in the case of MMBCR. On the other hand, the performance of DSR and FORP with respect to network lifetime improved in scenarios involving power control and/ or on-demand recharging. For both these non power-aware routing protocols, with on-demand recharging, nodes that are being used more frequently receive more energy and nodes that are used less frequently receive less energy. DSR preferentially uses nodes that lie in the center of the network to be part of several minimum-hop routes and FORP preferentially uses nodes that maintain stable links with its neighbors to be part of several stable routes. As future work, we plan to study the improvement in the network lifetime obtained using on-demand recharging for multicast routing protocols.

References

Bettstetter, C., Hartenstein, H., and Perez-Costa, X. (2004). Stochastic Properties of the Random Waypoint Mobility Model. *Wireless Networks*, *10*, 5, 555 – 567.

Bianchi, G. (2000). Performance Analysis of the IEEE 802.11 Distributed Coordination Function. *IEEE Journal of Selected Areas in Communications*, *18*, 3, 535 – 547.

Broch, J., Maltz, D. A., Johnson, D. B., Hu, Y. C., and Jetcheva, J. (1998). *A Performance Comparison of Multi-hop Wireless Ad hoc Network Routing Protocols*. Paper presented at the 4th ACM/IEEE Conference on Mobile Computing and Networking, USA.

Fall, K. and Varadhan, K. (2001). *NS-2 Notes and Documentation*, The VINT Project at LBL, Xerox PARC, UCB, and USC.ISI, http://www.isis.edu/nsnam/ns.

Feeney, L. M. (2001). An Energy Consumption Model for Performance Analysis of Routing Protocols for Mobile Ad hoc Networks. *Journal of Mobile Networks and Applications*, *3*, 6, 239 – 249.

Gomez, J., and Campbell, A. T. (2004). *A Case for Variable-Range Transmission Power Control in Wireless Multi-hop Networks*. Paper presented at the IEEE International Conference on Communications, Hong Kong.

Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J. (2004). *Global Positioning System: Theory and Practice*. (5th rev. ed.) Springer.

Johansson, P., Larsson, T., Hedman, N., Mielczarek, B., and Degermark, M. (1999). *Scenario-based Performance Analysis of Routing Protocols for Mobile Ad hoc Networks*. Paper presented at the 5th International Conference on Mobile Computing and Networking, USA.

Johnson, D. B., Maltz, D. A., and Broch, J. (2001). *Ad hoc Networking*, Addison-Wesley.

Kim, D., Garcia-Luna-Aceves, J. J., Obraczka, K., Cano, J.-C., and Manzoni, P. (2003). Routing Mechanisms for Mobile Ad hoc Networks based on the Energy Drain Rate. *IEEE Transactions on Mobile Computing*, *2*, 2, 161 – 173.

Meghanathan, N. (2008). Exploring the Stability-Energy Consumption-Delay-Network Lifetime Tradeoff of Mobile Ad hoc Network Routing Protocols. *Academy Publisher Journal of Networks*, *3*, 2, 17 – 28.

Meghanathan, N., and Farago, A. (2005). *Maximizing Network Lifetime under Fixed Energy Budget in Ad hoc Networks*. Paper presented at the 6th IEEE-Southeast Conference, USA.

Meghanathan, N. and Farago, A. (2006). *Comparison of Routing Strategies for Minimizing Energy Consumption in Mobile Ad hoc Networks*. Paper presented at the 4th Asian International Mobile Computing Conference, India.

Park, S.-J., and Sivakumar, R. (2002). *Load Sensitive Transmission Power Control in Wireless Ad hoc Networks*. Paper

presented at the 4th IEEE Global Communications Conference, Taiwan.

Su, W., and Gerla, M. (1999). *IPv6 Flow Handoff in Ad hoc Wireless Networks using Mobility Prediction*. Paper presented at the IEEE Global Telecommunications Conference, Brazil.

Toh, C.-K. (2001). Maximum Battery Life Routing to Support Ubiquitous Mobile Computing in Wireless Ad hoc Networks. *IEEE Communications Magazine*, *39*, 6, 138 – 147.

Toh, C.-K. (2002). *Ad hoc Mobile Wireless Networks*, Prentice Hall.

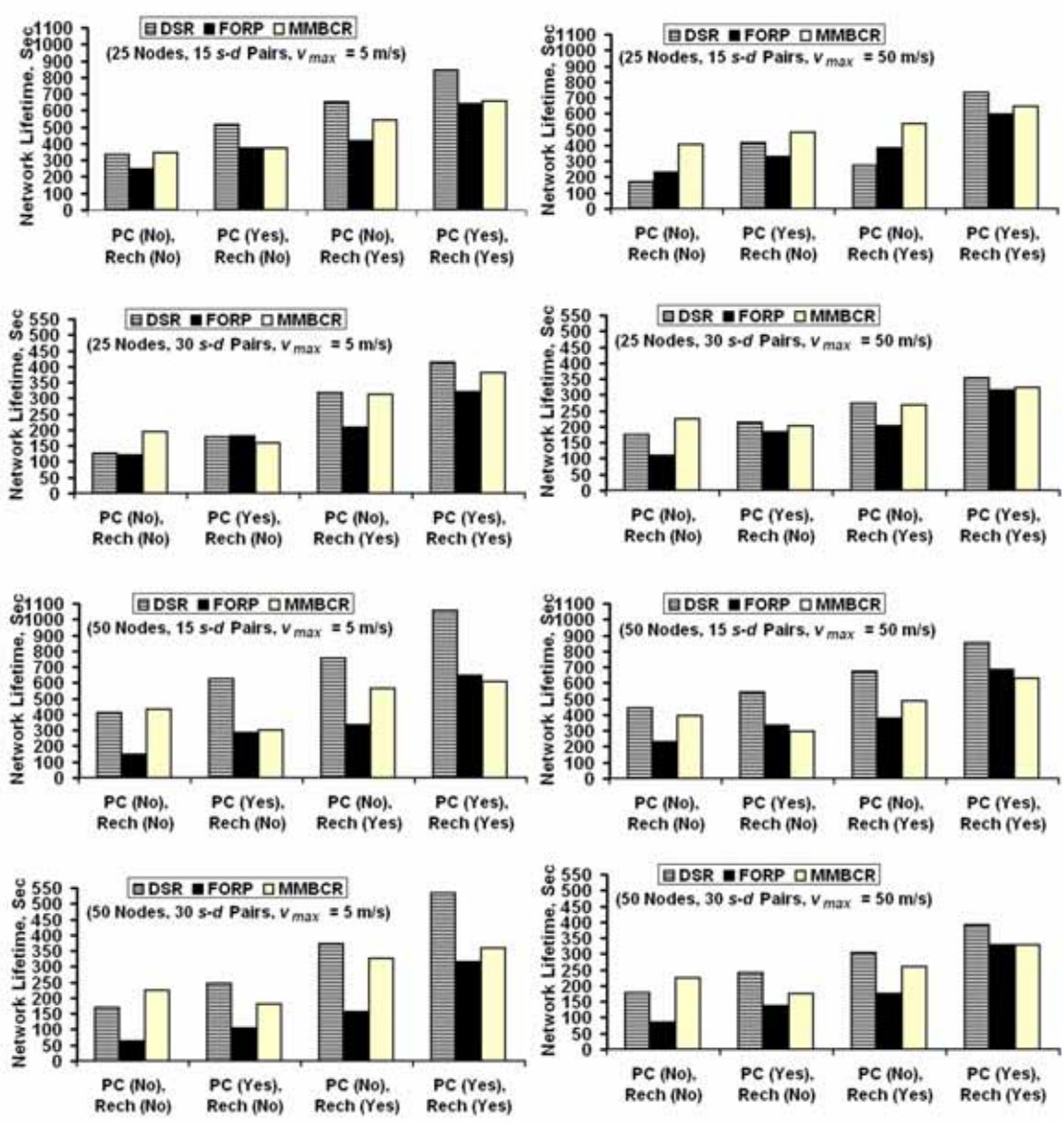


Figure 1. Network Lifetime for DSR, FORP and MMBCR

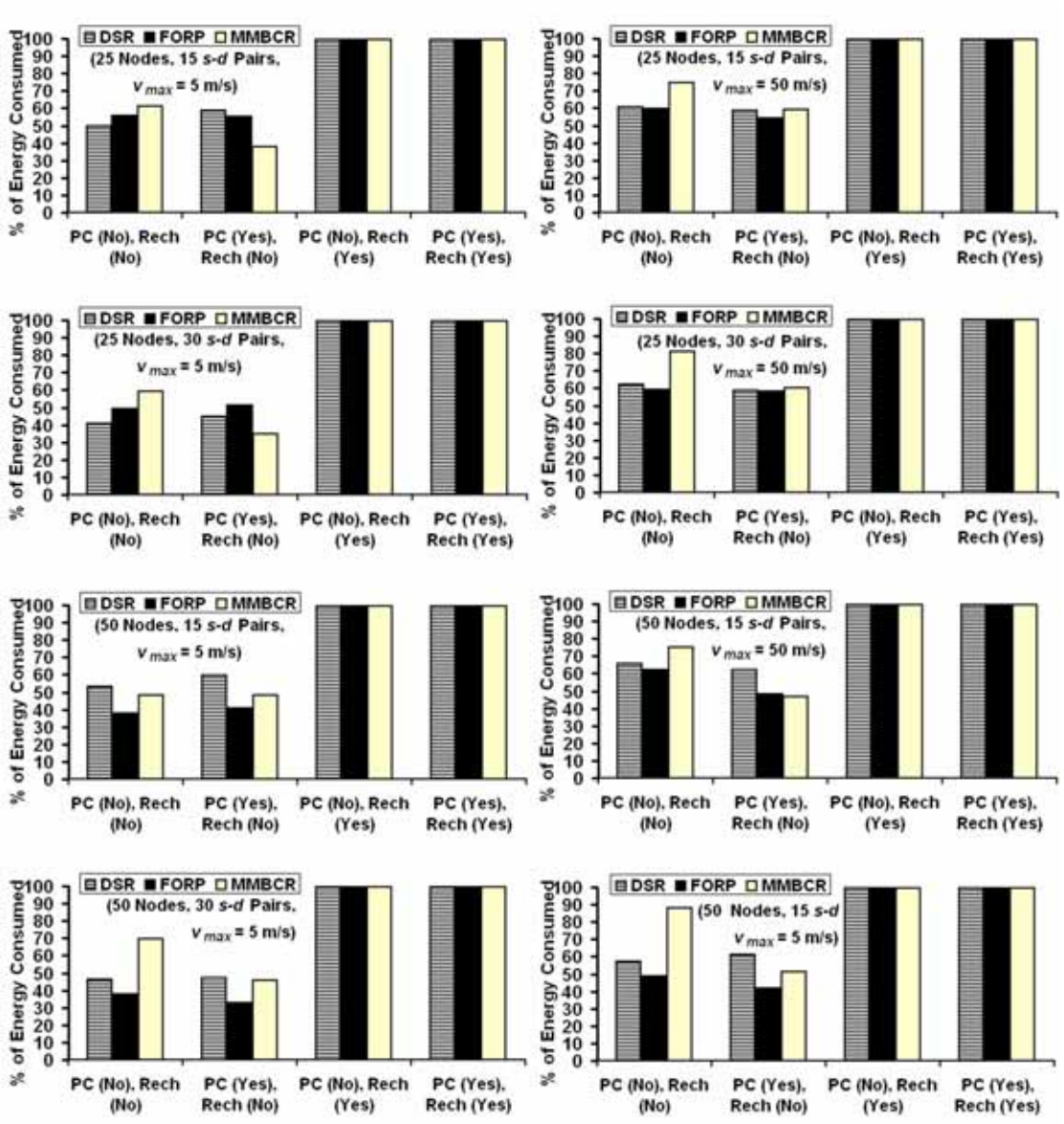


Figure 2. Percentage of the Fixed Energy Budget Actually Consumed for DSR, FORP and MMBCR

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Noise Removal of Spaceborne SAR Image Based on the FIR Digital Filter

Wei Zhang & Jinzhong Yang
China Aero Geophysical Survey & Remote Sensing Center for Land and Resources, Beijing 100083, China
Tel: 86-10-6206-0250 E-mail: zhangwei@agrs.cn

Conghui Fang
China University of Geosciences, Beijing 100083, China

Dongli Fan
China University of Mining & Technology, Beijing 100083, China

Jianyong Wu
Zhejiang Institute of Geology & Mineral Resources, Hangzhou 310007, China

Abstract

The speckle effect inevitably exists in the image of the Synthetic Aperture Radar (SAR). The removal of the speckle noise is the necessary approach before automatic partition, classification, target detection and abstraction of other quantitative special information in the SAR image, so it is very meaningful to eliminate or furthest restrain the speckle noises when the spatial resolution of the image is not be reduced. In this paper, the FIR filter is used to remove the noise in the SAR image, and optimal filtering coefficient is selected through experiment and analysis in the filtering process. The results show that the FIR filter used to remove noises in the SAR image is better than other traditional filtering methods in keeping the radiolocation feature and restraining the speckle noises, and the filtering speed is quicker. At the same time, the selection of the filtering coefficient will largely influence the de-noising effect of the FIR filter.

Keywords: Speckle effect, SAR image, FIR filter, Noise removal

1. Introduction

Because the SAR image has the ability that can strongly penetrate through the unconsolidated sediment on the earth's surface and the advantage that can offer special information about the earth's surface through the microwave, so it is regarded as the important new generation remote sensing information source (Shu, 1997). The grainy noise on the SAR image is called the speckle, and it is the necessary result of the coherent imaging radar. The noise can largely impact the performance of the radar to judge the target, and when the speckle noise is serious, it will even make the geo-body feature lost. It is very important to remove the speckle noise for the SAR image processing and application, and it is always one of important research topics in the SAR image processing technology. From the 1980s, many filtering algorithms about the speckle noises begun to occur, but it is still very meaningful to remove or furthest restrain the speckle noises when the spatial resolution of the image is not be reduced (Zhang, 2006, P.4-6).

The usual methods to restrain the speckle noises of SAR image mainly include the average filtering, the mean filtering, the local filtering, the Lee filtering, the Sigma filtering, the Kuan filtering, the Frost filtering, the Gamma Map filtering and the threshold filtering based on the wavelet decomposition (Ding, 2008, P.390-394 & Cao, 2008, P.2862-2865 & Lu, 2008, P.1053-1055 & Yang, 2008, P.525-529 & Bai, 2008, P.1234-1241 & Zheng, 2008). Many of them have been adopted by the mainstream remote sensing image processing software. However, many algorithms such as the average filtering, the mean filtering, the local filtering, the Frost filtering and the wavelet de-noising filtering all mainly aim at the additive noise mode based on the Gaussian hypothesis, but the speckle noises in the radar image are just based on non-Gaussian distribution, so the de-noising effects are not ideal and the noises displayed in the edge disposed by the Lee filtering are not be smoothed.

FIR digital filter is a digital system which is used to filter the discrete time signals, and achieve the frequency domain filtering by mathematically processing the sample data, and its largest advantage is to realize the linear phase filtering, so it has been widely applied in the digital signal processing domain. At the same time, the FIR filter is a full-zero filter which does not need to consider the stability and can be implemented easily.

2. Implementation principle of the FIR filter

The FIR filter with the constant coefficient is called as the finite pulse response filter which is a kind of digital filter. Corresponding with the IIR digital filter, its unit pulse response f[n] only has finite data points. The relationship between the output with z-order or the length of z and the input time sequence x[n] is described as a kind of finite convolution quantity form as follows.

$$y[n] = x[n] * f[n] = \sum_{k=0}^{L-1} x[k] f[n-k] \tag{1}$$

Where, f[0]≠0 to f[L-1]≠0 are all the coefficients of z order of the filter. The formula (1) can be simplified as the form in the z domain.

$$Y(z) = F(z)X(z) \tag{2}$$

Where, F(z) is the transfer function of the FIR digital filter, and its form in the z domain is

$$F(z) = \sum_{k=0}^{L-1} f[k] z^{-k} \tag{3}$$

Figure 1 is the graphic analysis of the z order FIR filter. From the figure, the FIR filter is composed by one adder and one multiplier, and each operation number transferring to each multiplier is one FIR coefficient.

3. Approaches of designing FIR digital filter by the window function method

By the window function method, the design approach of the FIR digital filter can be described as follows.

First, confirm the required frequency response function $H_d(e^{j\omega})$;

Second, select the window function, and according to the allowed excess bandwidth $\Delta\omega$, evaluate the length of the h(n) sequence, and generally $N=A/\Delta\omega$ (where, A is the constant which is confirmed by the window function form, and $\Delta\omega$ approximately equals to main lobe width of the frequency $w(e^{j\omega})$ of the window function);

Third, compute the unit impulse response of the digital filter;

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(e^{j\omega}) e^{j\omega\left(n-\frac{N-1}{2}\right)} d\omega \tag{4}$$

Fourth, window the h(n) by the selected window function;

$$h(n) = h_d(n)\omega(n) \tag{5}$$

Fifth, compute the frequency response of the filter;

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} h(n) e^{-j\omega n} \tag{6}$$

Sixth, design a linear phase filter, and its inner extent of the pass-width is 1 and its inner extent of the block-width is 0, and the digital cut-off frequency is ω_c (Wei, 2008, P.6-7).

To maximize the orders and minimize the error of the filter, the Kaiser window function method is adopted to design and realize the FIR digital filter, and the Matlab software can be used to design the low-pass filter according with requirements. The parameters respectively are [110 140], [1 0], [0.01 0.1] and 1000, and the feature curve of the low-pass filter is seen in Figure 2.

4. Factor selection and effect evaluation

According to this low-pass filter, select different parameters and de-noise the SAR image.

4.1 Visual effect

The de-noising effect in the research is seen in Figure 3. Various methods are implemented by the current remote sensing image processing software such as ERDAS, and the FIR filter de-noising method is implemented by the wavelet image processing software developed independently.

From the visual effect (seen in Figure 3), comparing with traditional de-noising algorithm, the de-noising effect of the FIR filter used in the research is very obvious, and it can not only remove the speckle noises to the large extent and give prominence to the useful information of the original image, but keep the edge details of the image as possible.

The selection of the FIR filter coefficient will largely impact the de-nosing result, and from Figure 3, the bigger the filtering coefficient is, the better the de-noising degree is.

4.2 Quantitative evaluation

To evaluate the filtering effect, different factors can be selected. The usual factor is the smoothing index ***F*** which is the ratio between the average of all pixels in the processed image and its standard error, and it indicates the smoothing ability of the filter to the image. The higher the value of ***F*** is, the stronger the smoothing function is. The computation formula is:

$$F = M/S \tag{7}$$

The edge keeping index denotes the keeping ability of the filter along the level direction or the vertical direction after filtering, and the higher the value of ***E*** is, the stronger the keeping ability is. The computation formula is

$$E = \frac{\sum_{i=1}^{m} | G_{R1} - G_{R2} | \text{ after filtering}}{\sum_{i=1}^{m} | G_{R1} - G_{R2} | \text{ before filtering}} \tag{8}$$

Where, ***m*** is the amount of the pixel in the image, and *GR1* and *GR2* respectively denotes the grey degree values of the left-right neighboring pixels and the up-down neighboring pixels (Zhang, 1997). In fact, the edge keeping index is contrary with the de-noising intensity, and the de-noising intensity is stronger, and the edge keeping ability is worse. At the same time, the edge keeping index can be the evaluation standard of the keeping degree of effective information in the image.

Therefore, the evaluation factor ***Q*** is defined by the arithmetic average of the smoothing index and the edge keeping index.

According to above evaluation indexes, compute various filtering schemes, and the results are seen in Table 1. The curves of control factors corresponding with different parameters in Table 1 are seen in Figure 4.

From the analysis result, as parameters increase, the smoothing index rise continually, i.e. the soothing effect is better and better and the edge keeping index is lower and lower. The evaluation factor curve has three wave crests, i.e. 0.78, 0.6 and 0.2. According to the definition of the evaluation factor, when the coefficient is 0.78, the occurrence of the wave crest is because the smoothing index is the highest one, i.e. the smoothing effect is at the best moment, but the edge index is at the lowest moment, which indicates that the effective information has not been kept enough. When the coefficient is 0.2, the occurrence of the wave crest is because the de-noising is not enough and the edge keeping index is at the highest moment, so this value should be given up. Therefore, the FIR filter with the coefficient of 0.6 has the best de-noising effect for the SAR image.

5. Conclusions

This experiment shows that the FIR filter has better effect to de-noise the SAR image. The filtering coefficient will largely influence the filter, and as the coefficient increases, the de-noising degree will become stronger, but the kept effective information will accordingly become less, so it is very important to select proper coefficient for the FIR filter to de-noise the SAR image.

References

Bai, Jing, Hou, Piao, Wang, Shuang & Jiao, Licheng. (2008). SAR Image Denoising Based on Lifting Directionlet Domain Gaussian Scale Mixtures Model. *Chinese Journal of Computers.* No.31(7). P.1234-1241.

Cao, Zongjie, Min, Rui, Pang, Lingli & Pi, Yiming. (2008). A Variational Level Set SAR Image Segmentation Approach Based on Statistical Model. *Journal of Electronics & Information Technology.* No.30(12). P.2862-2865.

Ding, Xianwen, Han, Lin & Zhang, Wei. (2008). SAR Image De-speckling Based on Lifting Wavelet Transform. *Bulletin of Science and Technology.* No.24(3). P.390-394.

Lu, Xiaoguang, Han, Ping, Wu, Renbiao & Liu, Ruihua. (2008). An Approach for SAR Image Despeckling Based on 2D-Wavelet Transform and ICA. *Journal of Electronics & Information Technology.* No.30(5). P.1053-1055.

Shu, Ning. (1997). *Principle of Radar Remote Sensing.* Beijing: Surveying and Mapping Press.

Wei, Rui. (2008). Simulation and Implementation of FIR Digital Filter. *Science Mosaic.* No.5. P.6-7.

Yang, Xin & Huang, Shunji. (2008). SAR image segmentation method with multiple regions based on partial differential equations. *Chinese Journal of Radio Science.* No.23(3). P.525-529.

Zhang, Wei, Sun, Ronghua & Zhang, Xiaocan. (2006). Noise Removal of SAR Image Based on Improved Wavelet Soft-thresholding Method. *Remote Sensing Information.* No.4. P.4-6.

Zhang, Xiaocan, Huang, Zhicai & Zhao, Yuanhong. (1997). *Remote Sensing Image Processing.* Hangzhou: Zhejiang University Press.

Zheng, Wei & Lu, Gang. (2008). Wavelet-based SAR Image Denoising Research and Application. *Computer Knowledge and Technology.* No.15.

Table 1. Filtering image statistical features of various schemes

Filtered	Mean	Mean variance	Smoothing index F	Edge keeping index (level)	Edge keeping index (vertical)	Edge keeping index	Evaluation factor
Original image	104.35	63.85	1.63	1	1	1	1.315
0.2	90.5	36.81	2.46	0.84	0.85	0.845	1.6525
0.3	91.49	36.79	2.49	0.65	0.74	0.695	1.5925
0.4	91.06	37.1	2.45	0.64	0.73	0.685	1.5675
0.5	88.26	38	2.32	0.67	0.64	0.655	1.4875
0.6	95.84	37.29	2.57	0.63	0.49	0.56	1.565
0.7	92.37	36.79	2.51	0.63	0.44	0.535	1.5225
0.78	99.46	37.43	2.66	0.61	0.43	0.52	1.59
0.8	99.46	37.43	2.66	0.61	0.43	0.52	1.59

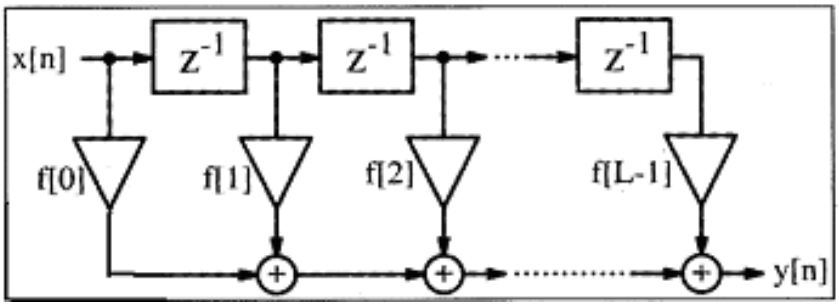


Figure 1. FIR Digital Filter with Direct Form

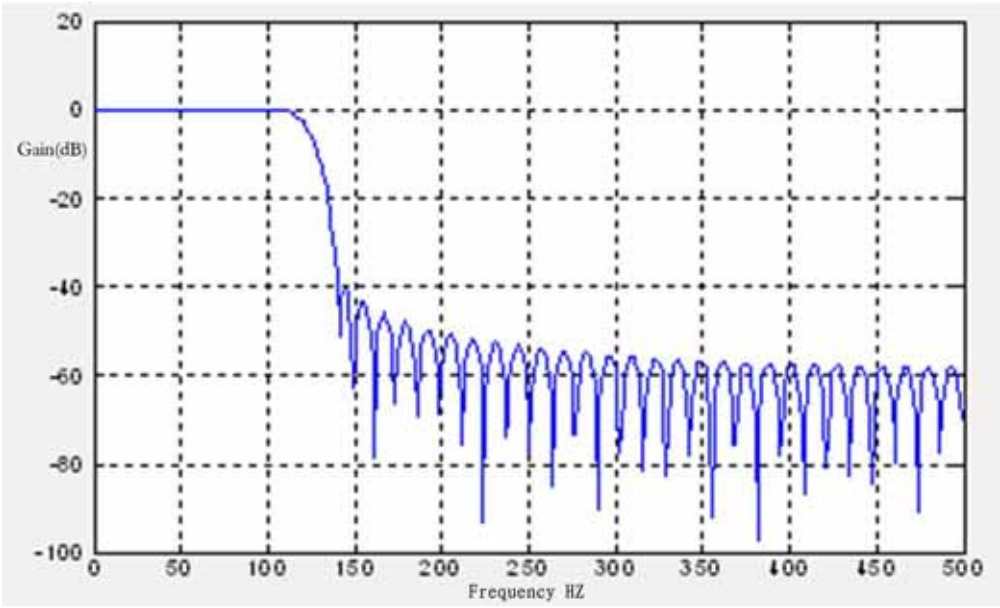
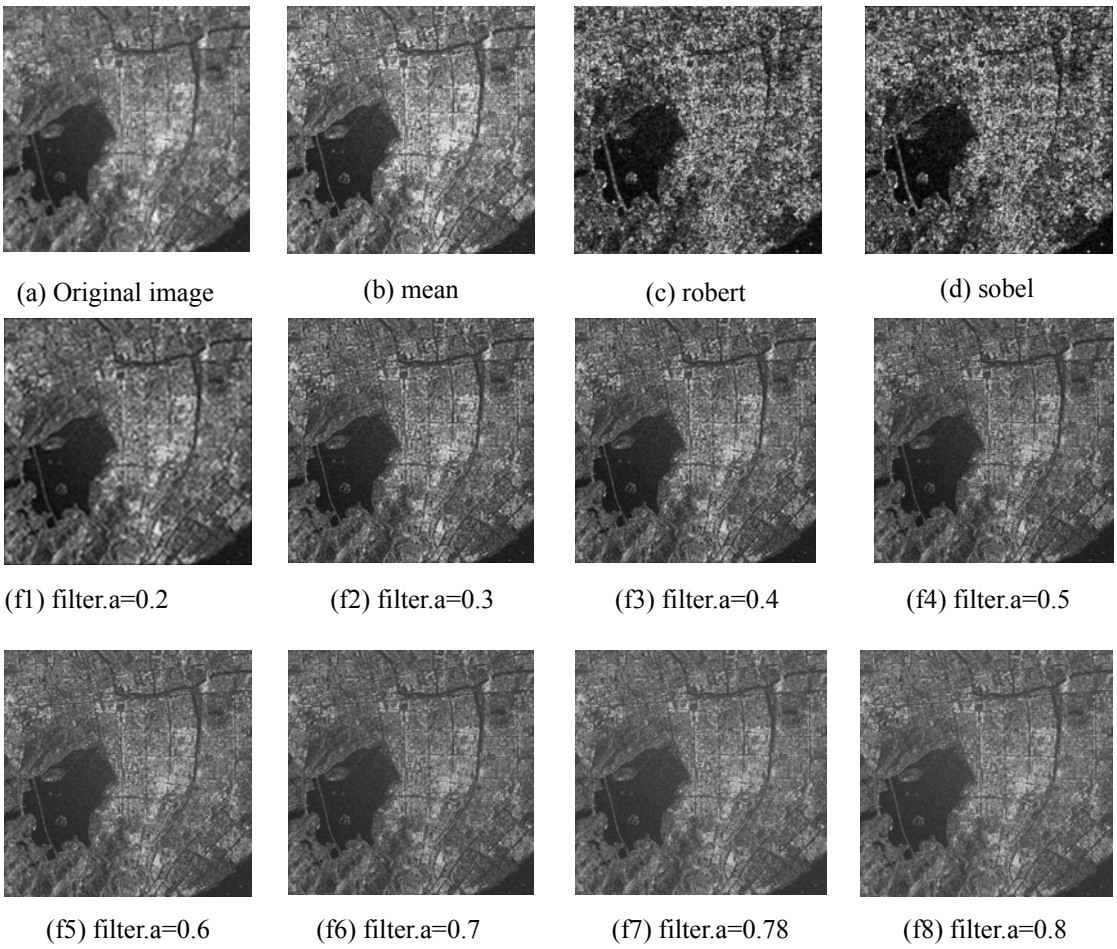


Figure 2. Feature Curves of the Low-pass Filter



(a) Original image (b) mean (c) robert (d) sobel

(f1) filter.a=0.2 (f2) filter.a=0.3 (f3) filter.a=0.4 (f4) filter.a=0.5

(f5) filter.a=0.6 (f6) filter.a=0.7 (f7) filter.a=0.78 (f8) filter.a=0.8

Figure 3. De-noising Results of Various Schemes

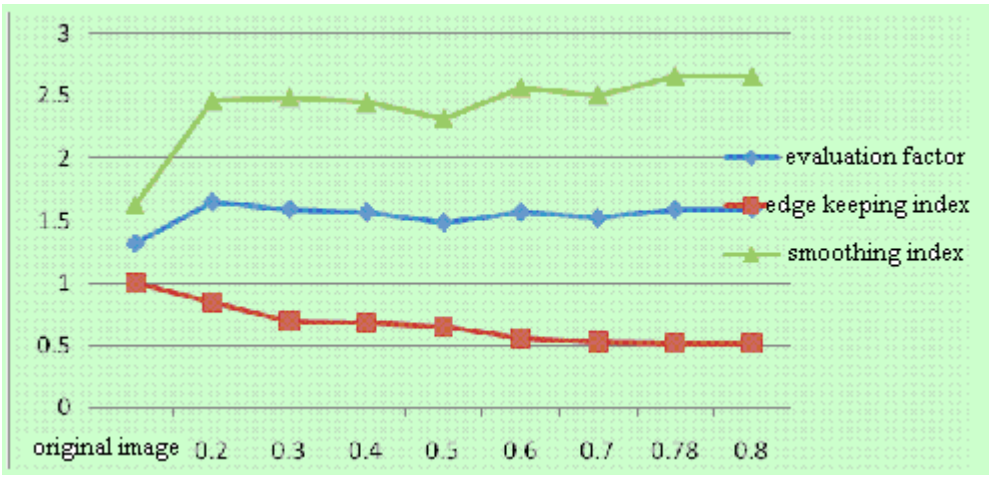


Figure 4. Control Factor Curve of Different Parameters

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Documenting Software Requirements Specification: A Revisit

Mingtao Shi
FOM Fachhochschule für Oekonomie & Management
University of Applied Science
Bismarckstr. 107, 10625 Berlin, Germany
Tel: 49-171-2881-169 E-mail: Consulting_Shi@yahoo.de

Abstract

Software Requirements Specification (SRS) is the key documentation, defining the functional and non-functional system requirements. By revisiting a number of in the literature much discussed key aspects related to SRS and extracting essential views from the author's daily work experience, this papers stresses the importance of the SRS and examines the process, which enables the emergence of a quality SRS. Business information must be acquired, discussed, analysed and digested, which forms the inputs for the documentation of SRS. A SRS typically consists of a set of documents, including the SRS itself, which is a written description of business requirements and system features, and analysis models.

Keywords: Requirements engineering, Software Requirements Specification (SRS), Requirements elicitation, Requirements modelling, SRS Tools

1. Software Requirements: An Overview

It is hardly contentious that requirements engineering as the functional interface between IT and the business domain has long become the initiating activity of the software development process, the importance of which cannot be overestimated (Paech & Rolland, 2008). What is equally uncontroversial is the notion contending that the success of software systems significantly depends upon the quality of the documentation in all key phases of software engineering, including requirements, design and construction, test and debugging, and release and delivery phases. This is particularly true for larger and more complex systems. Thus, meaningful requirements analysis and documentation is positively related to higher quality of software development projects. Importantly, Software Requirements Specification (SRS) is the key document in the requirements phase, laying the foundation for the realisation of functional and non-functional system requirements.

Many industrial firms today still tend to undervalue the necessity of requirements process in their daily practice, because the effects and benefits of related activities were allegedly "invisible". In fact, exactly the opposite is true. A number of authors have admirably written about the treatment of requirements in great detail (Pohl, 2007, Robertson & Robertson, 2006, Hull, Jackson & Dick, 2005). But to deal with this issue completely in compliance with theory would consume much firm-internal human resources and time capacity that is virtually non-existent, especially in small and medium-sized firms.

Therefore, this paper is intended to re-visit and streamline some of the key concepts related to the documentation of SRS from the perspective of an industry practitioner working in the area of requirements engineering, and by doing so, shows possible ways of effective and efficient SRS documentation. Efficiency is especially essential for firms with resource constraints.

2. Informational Sources of SRS

A formal SRS is normally created for more complex software system. It is often necessary to organise a launch meeting, while initiating the larger software project. It is vastly beneficial if the project members can enlist the support of the senior management and the top managers of the firm can show up in the meeting, in which analysts, designers, developers and testers can communicate with each other, responsibility areas are laid down and action plan, schedule, resources and budget are planned as precisely as possible.

SRS are based upon the needs stemming from the business domain. Rupp (2007, p. 115-134) discussed the methodologies of information elicitation extensively and differentiates among creativity, observation, questioning, history-oriented and other supplementary techniques such as Work-Shop and Mind-Mapping. Although all techniques have advantages and disadvantages, creativity techniques are generally suitable for advanced analysis and observation and questioning techniques are less time-consuming and also effective.

The information gathered in the first round does not represent meaningful or structured system requirements in most cases. Therefore, discussions afterwards within the system development team are necessary and highly recommended. To re-examine the "raw" information by using flipcharts or whiteboard in a group deepens the understanding of the nature of the elicited information and helps to restructure the available information in such a way that later documentation work will become easier and more effective. If uncertain or opaque information should arise during the group discussions, iterative interactions between the requirement engineer and the business is indispensible. This step taking place after information gathering through iteration is strikingly important to ensure the quality of the requirements analysis.

Once the required system features are seized and become intelligible to the analyst, the analysis work may begin. Business information needs to be converted into technical data, functionalities and system behaviour. Software authority Pressman (2009, p. 148-199) and Pressman (2004) recommended scenario-based, flow-based, class-based and behaviour-based techniques to analysts. A detailed delineation of these techniques is beyond the scope of this paper. However, it is worth mentioning that scenario-based modelling mainly includes use-cases and activity diagrams (swimlane diagrams). Although the requirement engineer can begin to conceive use-cases in the previous elicitation stage, individual user stories ought to be refined further and the relationship among the generated use-cases should be sorted out in this modelling stage of the requirement engineering. The use-cases themselves should become more rigour, optimally following a stringent template that includes the essential information for designer and programmer. Some authors emphasise other techniques such as Entity Relationship Diagram (ERD). Diagrammatical or graphical notations of modelling techniques are highly coveted in the daily business because of the ease of understanding, although the textual use-cases are also salient.

It is vital to point out that the modelled system is not necessarily always realistic for the software development. Negotiation between the future system users and system developers is always needed. Making efforts towards a wish-reality convergence and maintaining a friendly and motivating project environment is another challenge for the requirements analyst.

3. Creation of SRS

Firms that do carry out requirements engineering tend to only focus on models and diagrams, because the related activities are usually less time-consuming. However, modelling the user requirements is not sufficient for the software development team to design and construct a solid system. Experience shows that the requirement engineer should generally formulate a written document, the Software Requirements Specification. In fact, the models (use-cases and diagrams) build in the analysis/elaboration stage are appendices and attachments of SRS. SRS can be written while or after modelling the requirements.

A SRS must explicitly spell out what the system should precisely perform on the one hand and address the following issues on the other:

- Requirements should be labelled to establish linkages with use-cases and later test-cases. While *labelling the requirements*, it may be beneficial to do this in a hierarchical structure without using numerical structure. Numerical structure is less flexible for later structural changes. One non-numerical example would be: Banking.Retail.Accout.OpenAccount;
- *Open issues* that were not considered as user needs were elicited and elaborated and became visible in the modelling or specification stage must be registered in a list within the SRS and the business and technical implication of these issues be clarified gradually;
- Statements and notices that are not generally known to all participating parties (stakeholders) but are vital components of the SRS should be captured in a *reference system*, so that readers can clearly comprehend, from which authorised informational source certain requirements stem;
- In many cases, systems exist not only to support core businesses of the users but also their *ancillary processes*. A banking system, for example, must support not only core transactions such as standing orders and teller/till activities, but also reporting/controlling and accounting procedures. A SRS must mention and describe the needs of these ancillary processes that must be dealt with by the system and also establish corresponding requirements in additional chapters;
- Do not forget *interfaces towards external systems*. If these interfaces are based upon standards (e.g. ISO or IEEE interfaces), uncertainties exist but are not significant. On the contrary, if interfaces are proprietary, the degree of uncertainties rises exponentially. It is recommended to involve as less external interfaces as possible and standard-based interfaces only. The SRS should map the communication mode and content of these interfaces in great detail in a separated chapter;

- A successful SRS always includes a number of *important appendices*, such as use-cases, modelling diagrams, data dictionary, preliminary (or more precise) UI design, possibly test cases and test plan.

Wiegers (2003, p. 171-181) showed an effective SRS template, which can be used in the daily businesses immediately, although not necessarily all chapters must covered by the specification writers. Levels of details and areas to be covered in the specification vary from system to system. In order to reduce resources to be deployed, it is suggested that the analysts formulate the SRS as concisely as possible (KISS: Keep it Simple, Stupid) but as detailed as necessary. The capability of the analysts to write a powerful SRS grows as the experience of the analysts grows. In other words, the SRS must be seen as a routine task of the analysts, a viewpoint that is not shared by all analysts. Many of them are excel at analysing and modelling the requirements and delineating user scenarios but reluctant to write a SRS and frequently delegate this task to colleagues responsible for documentary tasks. This kind of modus operandi usually consume more but not less firm-internal resources, because of the different skills required and intensive communication necessary for complex information.

While writing a SRS, analysts also need to pay attention to the linguistic aspects. Rupp (2007, p. 140-174) intriguingly pointed out that deletion, generalisation and distortion cause substantive quality loss of an SRS and introduces a number of rules to effectively tackle this issue. A detailed discussion of this aspect is beyond the scope of this paper.

A completed SRS should be reviewed by the project team, especially by the requirements analyst. During this validation process, open issues must be clarified and the SRS updated. Furthermore, erroneous, inconsistent and conflicting statements in the SRS must be corrected. A SRS is baselined when the validation of the requirements is completed. An experienced analyst external to the project may review the SRS again to re-confirm the quality of the SRS.

4. SRS Tools

As stated, a SRS is usually not one document, but a set of documents. Firms adopting a SRS tool need to apply a multi-faceted selection strategy but should not only concentrate on one feature (SRS template) that the software may provide. There are plenty of software tools that support the creation and management of SRS. No matter whether open source (e.g.. BOUML) or licensed tools (e.g. IBM Rational Rose) are used, some key issues must be effectively dealt with. The tool should:

- enable the creation and management of textual use-cases;
- provide editable template for SRS;
- integrate diagrammatical, mathematical and logical illustration for requirements modelling (e.g. use-case diagram, activity diagram, entity-relationship-diagram, state diagram, class diagram, data dictionary);
- establish linkages among use-cases, software requirements and test-cases;
- support the management of documentation, including change orders, release and version management.

5. Conclusion

Documentation of SRS is the key activity to elicit and understand, elaborate and illustrate the business requirements and forms a crucial building block in the process of software engineering, linking business-oriented delineation with code-oriented modelling. In order to acquire the necessary information for a SRS, a requirements engineer may experience a number of procedural stages, in order to ensure quality. These stages typically include project launch, requirements elicitation, elicitation refinement, requirements modelling and negotiation. This paper, subsequently, discusses some of the most essential aspects during the documentation of a SRS and briefly looks at the tool requirements for supporting the creation and management of SRS.

References

Hull, E., Jackson, K., & Dick, J. (2005). *Requirements engineering.* (2nd ed.). Berlin: Springer Verlag.

Paech, B., & Rolland, C. (2008). REFSQ'08 international working conference on requirements engineering: Foundation for software quality. In B. Paech, & C. Rolland (Eds.), Requirements engineering: Foundation for software quality (pp. 1-5). Berlin/Heidelberg: Springer Verlag.

Pohl, K. (2007). *Requirements Engineering: Grundlagen, Prinzipien*, Techniken. Heidelberg: dpunkt.verlag GmbH.

Pressman, R. S. (2004). *Software engineering: A practitioner's approach*. (6th ed.). New York, NY: McGraw Hill.

Pressman, R. S. (2009). *Software engineering: A practitioner's approach.* (7th ed.). New York, NY: McGraw Hill.

Robertson, S., & Robertson, J. (2006). Mastering the requirements process. (2nd ed.). Westford, MA: Pearson Education Inc.

Rupp, C. (2007). Requirements-Engineering und Management: Professionelle, iterative Anforderungsanalyse für die Praxis. (4th ed.). München/Wien: Carl Hanser Verlag.

Wiegers, K. E. (2003). *Software requirements.* (2nd ed.). Redmond, Washington: Microsoft Press.

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Simulation Evaluation Algorithms of LSRP and DVRP in Bank Network

Feixue Huang
School of Management, Dalian University of Technology, Dalian 116024, China
Tel: 86-411-8470-7210 E-mail: software666@163.com

Yong Zhou
Dalian Branch, Bank of Communications, Co., Ltd, Dalian 116001, China
Tel: 86-411-86-8281-6791 E-mail: unite@sina.com

Zhijie Li
School of Computer Science & Engineering, Dalian Nationalities University, Dalian 116600, China
Tel: 86-411-8765-6179 E-mail: lizhijie@dlnu.edu.cn

Abstract

Because the computational complexity of high usability analysis is too high when selecting the routing protocol in the bank network plan, a simple measurement method of high usability for banks is proposed to solve this problem in the article. First, establish a simulation environment which is close to the real network of banks to offer the measurement environment and data. Second, based on the theoretical comparison of LSRP and DVRP, establish the simple simulation measurement algorithm by Shannon's information entropy theory. Finally, evaluate the degree of high usability of LSRP and DVRP in the bank network by this algorithm. The result of simulation shows that (1) the simulation is close to the periodic rule of the statistical data group in real environment, (2) the deviation ratio is less than 0.1, and (3) the covariance is unequal to 0. And the result indicates that the simulation is connected with the real environment and both are very close, and the simulation environment can offer effective data. The computation result of the simple simulation measurement algorithm shows that the time cost differences of LSRP and DVRP exist in the period of fault recovering, which indicates the simple measurement algorithm is effective.

Keywords: Network system, Distance vector routing protocol (DVRP), Link-state routing protocol (LSRP), High usability, Information entropy

1. Introduction

Up to the late of 2008, the development of network payment associating with network shopping develops quickly, and the using scale has achieved 52 million people at present, and the yearly increase rate achieves 57.6%. The people who use the web bank have achieved to 58 millions in 2008 from 40 millions in 2007, and the increase rate achieves 45% (China Internet Network Information Center, 2009). By the scientific and technological measures, various commercial banks create financial products actively, and the IT investment scale of Chinese bank industry in 2005 had achieved 23.82 billion Yuan. Up to now, the operation and management systems including office of various commercial banks have all entered into the times of network.

Facing the network faults which can not be avoided completely, people try to enhance the usability of network. The quantitative evaluation method of the usability can respectively define the measurement index based on network connection and the measurement index based on network communication ability (Zhang, 2004, P.103-105). The network usability includes four aspects including equipment, route, user, and application (XF Wang, 2002, P.521-531). Based on the measurement, people put forward the design approaches and design ideas of usable network (C Song, 2005, P.392-395). Aiming at two kinds of damages faced by the complex network, some scholars respectively defined the fault tolerance and the invulnerability for the network node and the network border (Wu, 2005, P.128-131). Some scholars defined the ratio of the possible route amount among network groups to the route amount which could be required to establish among network groups as the usability of network topology (Zhong, 2003, P.113-114). Some scholars also advised seeking the usability values of various nodes in the communication network and taking the mean square deviation of these values as the usability value of the whole network topology (Chen, 2006, P.6-7 & 24).

For the network with many nodes, the practice (Lin, 2008, P.15-16) or the space computational complexity (Li, 2008, P.311-318) of general algorithms is too high, so researchers begun to design a more effective algorithm with low computational complexity starting from other approaches, and the concrete ideas could be divided into following sorts. In decomposition method, Kennington and Nair designed the usability network by splitting the network into small loop networks (J L Kennington, 1999, P.219-230), and Kerivin put forward the splitting-solving algorithm based on a series of inequalities (H Kerivin, 2002, P.265-268). Ghashghai and Rardin designed a mixed gene algorithm to split the network into many networks including many trees (E Ghashghai, 2002, P.53-66). Borne and Gourdin et al established a multi-layer usability network design model and put forward the branch cutting algorithm based on the model (S Borne, 2006, P.41-73). In the increase limitation method, Jothi and Raghavachari put forward the minimum generation network problem solving algorithm containing the capability limitation (R Jothi, 2004, P.183-190). Some scholars put forward the algorithm to establish the protection chain (C Liu, 2006, P.301-311), or increase the transportation flux and the route bandwidth limitation in the network (P Broström, 2006, P.235-253) when solving the problem of usability network design, or the comprehensive method (S Soni, 2002, P.133-149). And many other researchers took the routing selection network and the confirmation of the communication chain capacity as the optimized objectives, and used the genetic algorithm (Yue, 2009, P.205-206 & 209) to solve this complex nonlinear plan problem with many restriction conditions (T. Lux, 2005, P.169-196).

By the research about the computer network usability (Oleszkiewicz J, 2006, P.66-77), three methods respectively had their advantages and disadvantages. The first method studied the bottom theory algorithm and proved that the routing protocol had higher usability than another protocol. This kind of proof is lack of events, which might be because the differences of the data types and the equipment state made the practice result to be different with the result of theoretical analysis. The second method used existing graph theory combining with network factors to compute new algorithm of usability measurement. And various measurement methods in this type should consider numerous nodes, and the measurement algorithm was too complex in actual application. The third method utilizes exiting algorithms and technologies to enhance the usability of existing network, which lacks in the researches and improvement about the used algorithms and technologies.

People try to maintain the usability by the excellent design before the network is established and the good operation after the network is established. In the initial term when the network is established, one or one set of routing protocol fitting for future network should be selected, and this method needs theoretical base and low computational complexity. Therefore, the simulation network can be used to statically compare the running of the usable routing, and establish a measurement algorithm with low computational complexity. First, the simulated network environment of commercial bank is established, and the established environment adopts the mainstream network topology structure, i.e. the star configuration, in Chinese commercial banks, and the connection mode adopts the mainstream double-chain and double-star crossing. And the variance and covariance method is used to validate the close relationship between the simulation environment and the real environment. Second, based on theoretical analysis, the time costs of maintainability generated by three factors including CPU processing, routing convergence and bandwidth occupation are selected (Samuel C.A. Pereira, 2008, P.321-325) when the network fails, and combining with Shannon's theory of information entropy (C.E. Shannon, 1948, P. 379-423 & 623-656 & Gianpietro Malescio, 2006, P.918-918) and the randomicity (Liu, 2007, P.350-351), the simple simulation measurement algorithm (with low computational complexity) could be established. By the comparison and analysis of simulation test data, the established simple simulation measurement algorithm can be used to confirm the routing protocol with low time costs is the routing protocol with high usability.

In the article, by the statistical data offered by the established simulation environment, the evaluation measurement algorithm and the theory base are established to select the routing protocol for the large-scale network construction and update alternation in certain bank.

2. Theoretical comparison of LSRP and DVRP

The dynamic routing means the router can automatically establish its own routing list and properly adjust it according to the change of actual situation. The operation of the dynamic routing selection algorithm depends on two basic functions, i.e. the maintenance of the routing list, and the proper routing information exchange among routers. The dynamic routing protocol mainly includes the link state routing protocol (LSRP) and the distance vector routing protocol (DVRP).

2.1 LSRP

LSRP is the inter-domain routing protocol which is used most widely, and it adopts the design strategy of "jigsaw puzzle", i.e. each router broadcasts its link states with neighbor routers to all other routers in the network. In this way, after one router accepts the routing information sent from other routers, it assembles these link states and generates a topology view of the whole network, so it can compute the shortest route to other routers by the shortest route algorithm.

The initial LSRP develops from multiple research results, and Dr. Radiap Perlman researching about the information fault tolerance broadcast was the early presenter. At the same time, IETF (Internet Engineering Task Force) issued the algorithm based on the link state in RFC1247, i.e. the OSPF (the routing selection protocol based on the SPF algorithm) (Yan, 2002, P.460-464). The OSPF message refreshing mechanism will generate many messages increasing the load of network, and Chen Zhibo and Xu Mingwei put forward the mechanism of ERR (enhanced randomized refresh). The method which randomly decentralizes the refreshing time of OSPF message and prolongs the refreshing interval of OSPF message can eliminate the abrupt flux brought by OSPF and reduce the amount of OSPF message (Chen, 2005, P.966-968). To control the amount of LSA when refreshing OSPF, Liu Kuixing and Wang Binqiang put forward the random refreshing mechanism of flux control. And this mechanism adopted many technologies such as the slow start and acceleration degression to adjust the amount of LSA and avoid the congestion according to the network flux (Liu, 2007, P.241-242 & 249).

2.2 DVRP

DVRP is based on R. E. Bellman, L. R. Ford and D. R. Fulkerson's works, so it is also called as Bellman-Ford algorithm or the Ford-Fulkerson algorithm.

In DVRP, each router maintains one sheet of routing vector list which lists known best distance (metric) and path to each destination. It refreshes the information in the list by exchanging information with the neighbor routers. DVRP is the route selection algorithm based on Bellman-Ford algorithm.

The core idea of this algorithm is to continually look for the extended routes in the map by the method of tab until the extended routes can not be found. The distance vector routing algorithm requires each router sends its whole routing list when refreshing each times only to its neighbors. The representative application protocols include RIP, IGRP, RTMP and so on.

This algorithm has three deficiencies, i.e. it can not reflect the deterioration, the routing exchange, and the infinite computation. Aiming at these deficiencies, Ding Qin and Cai Yuecai used the split horizon method (Ding, 2008, P.34-36) to improve the running performance of the routing protocol.

2.3 Comparison of LSRP and DVRP

Comparing with DVRP, LSRP has quicker convergence speed and higher flexibility, and it can send the link state information to neighbor routers only when the network changes, and store the topology of the whole network and observe the network from its own view. LSRP is a more complex routing protocol than DVRP.

3. Simulation and measurement algorithm

Except for prolonging the MTTF of the system, the method to enhance the network usability is to reduce the maintenance time of the system. Next, the problem which one of LSRP and DVRP can enhance the network usability by reducing the maintenance time of the system in the future network of banks will be discussed.

3.1 Simulation intension and environmrnt

The intension of the experiment is to establish the simulation environment which is equal to or close to the real environment, and by the analysis and comparison of the statistical data of the real network environment of certain bank with the data collected in the experimental environment, to validate whether the established simulation environment is close to the real environment.

For the experimental environment, the network with 20 nodes is selected to simulate the real network. The network topology adopts the double-star configuration, and the connections between each node and the core in the double-star configuration respectively adopts two 10/100M Ethernets, and two two-layer exchangers are used to simulate the local end of the telecom operator at the connections between the router and every nodes in the core route (seen in Figure 1). Operate the routing protocols including the OSPF (the representative routing selection protocol of LSRP) and the RIP (the representative routing selection protocol of LSRP) in the simulation environment successively. In the time period of 220 seconds, the experiment will collect the statistical data and analyze the time quantities respectively consumed by CPU, convergence and bandwidth.

3.2 Result and analysis of simulation test

Figure 2 reflects the statistical comparison of the CUP time consumptions of DVRP and LSRP. LSRP transmits the routing information to neighbor routers per 10 seconds to keep the consistence of the routing list. In the initiatory 63 seconds, CPU works continually to rebuild the neighbor relationship and the synchronous information. DVRP transmits relative routing information per 101 seconds to keep the information synchronization among routers. In the initiatory 25 seconds, CPU works continually to process the routing protocol and synchronous information. The cause to generate this result is that the computation of DVRP is less in the initial term of intermission, and the information of the routing

list is relatively simple, so the use of CPU is less than the use of LSRP no matter in the convergence process or in receiving data.

Figure 3 reflects the statistical comparison of the convergence time occupations of DVRP and LSRP. LSRP completes the convergence task in 25 seconds, and DVRP needs 63 seconds for the same task. The cause is that both DVRP and LSRP need using their own modes to complete the convergence and occupies the time in this process in the initial term that the network fails. When the convergence completes, DVRP and LSRP will enter into the stable state and won't occupy time because of the convergence process. Because DVRP only transmits the routing list, and the network scale of the experiment is appropriate, and the LSA information transmitted by LSRP is so complex that it needs computation of CPU, so in the experimental environment, the convergence time of LSRP is longer than the convergence time of DVRP.

Figure 4 reflects the statistical comparison of bandwidth times of DVRP and LSRP. The DVRP occupies wider bandwidth than LSRP. The cause to generate this result is that the routing list information synchronization can be completed per 100 seconds. But LSRP only synchronizes the changed information in the period, and if the information doesn't change, LSRP will directly confirm the survival of neighbors. If the data quantity is less, the bandwidth occupation time is less, so DVRP occupies less time than DVRP.

For the statistical analysis of the simulation and reality of LSRP and DVRP, there are two differences between the actual network environment and the simulation environment. The first one is that the network scales are different, and the scale of the real network is 2.5 times than the scale of the simulation network (the amount of network node increases from 20 to 53). The second one is that the network bandwidths are different. The wide area network connection in real network adopts the bandwidth line of 2M of telecom operator, and it is smaller than the Ethernet of 10/100M in the simulation network.

Figure 5 reflects the statistical comparison between the LSRP simulation and the actual network. The rules that two networks run LSRP are close, but not consistent completely. The cause to generate this result is that though the scale of the actual network is bigger than the scale of the simulation network, the networking equipment and the line bandwidth adopted whether in the simulation environment or in the real environment can completely fulfill the requirements of the network construction of this scale, so even the network scale increases, the change of comprehensive time costs is not obvious.

Figure 6 reflects the measurement results of the DVRP simulation and the actual network. The rules that two networks run DVRP are close, but not consistent completely. The cause to generate this result is that though the scale of the actual network is bigger than the scale of the simulation network, but the networking equipment and the line bandwidth adopted whether in the simulation environment or in the real environment can completely fulfill the requirements of the network construction of this scale, so even the network scale increases, the change of comprehensive time costs is not obvious.

From Figure 5 and Figure 6, the change tendencies of the data of the experimental simulation and the statistical data in the real environment are close which can be reflected in the chart of statistical data. And it can also be proved from exiting algorithm. Next, respectively utilize the variance method and the covariance method to analyze the deviation degree and the association degree of two groups of data from the quantitative angle.

The variance is the value to reflect the centralized degree of data, and it describes the discrete degree of data. For the group data of ($x_1, x_2, \cdots, x_n$), its average value is $\overline{x}$, so its variance is

$$S^2 = \frac{1}{n}\left[\left(x_1^2 + x_2^2 + \cdots + x_n^2\right) - n\overline{x}^2\right] \tag{1}$$

Supposed that the variances of the simulation data and the real data respectively are S_s^2 and S_r^2, and the deviation degree is S_Δ^2 , and the deviation rate is ρ , so

$$S_\Delta^2 = S_s^2 - S_r^2 \tag{2}$$

$$\rho = S_\Delta^2 \div S_r^2 \tag{3}$$

From Table 1, the deviation degree between the data obtained in the simulation experiment of LSRP and the experimental data in the reality is only 0.033716, and the deviation degree between the data obtained in the simulation experiment of DVRP and the experimental data in the reality is only 0.004238.

The covariance is to describe the association degree of the data. X and Y are two variables, and the covariance can be denoted as

$$Cov(X,Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right] \tag{4}$$

If two random variables X and Y are independent to each other, so Cov(X, Y) = 0, and if the mathematical anticipation is not zero, so X and Y are not independent to each other, i.e. they have certain relationship.

By the computation, the comparison analysis results about LSRP and DVRP can be obtained.

$Cov_{LSRP}(X,Y)$=3450.373 ≠ 0

$Cov_{DVRP}(X,Y)$=3970.117 ≠ 0

The result by the method of variance indicates that the deviation degrees between the simulation data and the statistical data are all less than 1 and even 0.1. Thus deviation degree shows that both are very close. The result computed by the method of covariance is not equal to zero. According to the definition of the covariance, two groups of data are not independent, and have certain associated relationship. By the statistics and computation, the result shows that the established simulation environment can replace the real environment completely.

3.3 Measurement algorithm

Except for the factors such as the hardware redundancy, the line redundancy, and the network structure, the factor to enhance the usability of network is the recovering time of network fault. So respectively in LSRP and DVRP the time costs of CPU, convergence, and bandwidth when network fails are compared. The routing protocol with the lowest time costs in above three factors (CPU, convergence, and bandwidth) more fits for the requirements of the future network plan.

Combining with the research result of the information entropy and the complexity degree, the relationship among these three factors is seen in the equation (5).

$$T_c = T_{cCPU} \times \mathrm{C}_{\mathrm{CPU}} + T_{cC} \times \mathrm{C}_C + \mathrm{T}_{\mathrm{c}B} \times \mathrm{C}_B \tag{5}$$

Where, T_c is the total time cost, T_{cCPU} T_{cC} and T_{cB} respectively denote the time costs of CPU, convergence, and bandwidth, and C can be defined as the equation (6).

$$C = -\sum_{i=1}^{k} n_i \log(\frac{n_i}{N}) \tag{6}$$

Where, N is the sum of $n_i (1 \le i \le k)$ numbers, and the probability that any one number occurs is $\frac{n_i}{N}$.

The equation (6) is the expression of the information entropy, and it denotes the uncertainty of the random data occurring in the data group.

3.4 Application of measurement algorithm

The intention of the measurement algorithm is to apply it into the engineering practice, i.e. the simulation experiment.

Use the equation (5) to compute the time costs of LSRP and DVRP in the simulation environment, and the result is seen in Table 2.

Table 2 lists the time costs of LSRP and DVRP in every time period.

In the initial time periods of 1-40, 41-80, 81-120, and 121-160, the time costs of LSRP are higher than the time costs of DVRP, because the computational complexity of LSRP is higher than that of DVRP, so LSRP will generate higher time costs than DVRP in large-scale network.

In the late time periods of 161-200 and 201-220, the time costs generated by DVRP are higher than LSRP, because after LSRP completes the convergence, it will not occupy the bandwidth of network if the routing changes little.

In the whole time period of 1-220, the time costs generated by DVRP are higher than LSRP, because LSRP sends less information quantity, and it will occupy less bandwidth in large-scale network.

Table 3 clearly lists the analysis results of evaluation of LSRP and DVRP and the causes to generate the result.

Through the analysis, LSRP more suits for the simulated bank network than DVRP in the simulated network structure.

4. Conclusions

Because the bank network plan lacks simple simulation method when selecting the routing protocol, the theoretical characters of the dynamic routing protocol are further studied based on former scholars' researches in the article. The measurement model taking the time costs of CPU, convergence, and bandwidth in the fault period by the method of information entropy, i.e. the equation (5), is established. By the experimental data in the simulation environment, the usability of dynamic routing protocol is evaluated.

First, the method of variance and the method of covariance indicate that the experimental environment established in the article can be realized. Second, the simple algorithm in the article is applied in the simulation data and the evaluation result is obtained. Above results all show that the application of the simulation environment established in the article and the simple measurement algorithm in the concrete engineering practices of large-scale network construction and updating alteration for commercial banks is feasible.

References

C.E. Shannon. (1948). A mathematical theory of communication. *Bell System Technical Journal.* No.27. P. 379-423 & 623-656.

Chen, Jianguo & Zhang, Yongjing. (2006). Study on Evaluation Algorithm for Topology Survivability of Communication Network. *Radio Communications Technology.* No. 32(1). P.6-7 & 24.

Chen, Zhibo, Xu, Mingwei & Cui, Yong. (2005). Improved OSPF Refresh Mechanism. *Journal of Tsinghua University (Science and Technology).* No.45 (7). P.966-968.

China Internet Network Information Center. (2009). *The 23rdStatistical Report on the Internet Development in China.* Beijing: China Internet Network Information Center.

C Liu, L Ruan. (2006). P-Cycle design in survivable WDM networks with shared risk link groups (SRLGs). *Photonic Network Communications*. No.11(3). P.301-311.

C Song, S Havlin & HA Makse. (2005). Self-similarity of complex networks. *Nature.* No. 433(7024). P.392-395.

Ding, Qin & Cai, Yuecai. (2008). Improvement Scheme of Distance Vector Routing Algorithm. *Journal of Sichuan University of Science & Engineering (Natural Science Edition).* No.21 (3). P.34-36.

E Ghashghai, R Rardin. (2002). Using ahybrid of exact and genetic algorithms to design survivable networks. *Computers & Operation Research.* No. 29. P.53-66.

Gianpietro Malescio. (2006). Coping with uncertainty. *Nature.* No.443 (7114). P.918-918.

H Kerivin, A R Mahjoub. (2002). Separation of partition inequalities for the (1, 2)-survivable network design problem. *Operations Research letters.* No.30 (4). P.265-268.

J L Kennington, VSS Nair, M H Rahman. (1999). Optimization Based Algorithms for Finding Minimal Cost Ring Covers in Survivable Networks. *Computational Optimization and Applications.* No.14 (2). P.219-230.

Li, Zhimin, Luo, Libo & Li, Xiang. (2008). Study on the Computational Complexity of the Complete Binary Tree. *Acta Mathematica Sinica (Chinese Series).* No. 51(2). P.311-318.

Lin, Xiaozhu, Wu, Yijuan & Fu, Dandan. (2008). Analysis of Computation Complexity on Euler Number of a Binary Image. *Microelectronics & Computer.* No. 25(7). P.15-16.

Liu, Haoguang & Cai, Shaohong. (2007). Message Entropy and Randomicity. *Journal of Guizhou University (Natural Sciences).* No.27(4). P.350-351.

Liu, Kuixing, Wang, Binqiang & Zhang, Jianhui. (2007). OSPF-LSA Refresh Mechanism with Traffic Control Function. *Application Research of Computers.* No.24 (9). P.241-242 & 249.

Oleszkiewicz J, Li X, Liu Y H. (2006). Effectively Utilizing Global Cluster Memory for Large Data-Intensive Parallel Programs. *IEEE Transactions on Parallel and Distributed Systems.* No.17 (1). P.66-77.

P Broström, K Holmberg. (2006). Multiobjective design of survivable IP networks. *Annals of Operations Research.* No. 147(1). P.235-253.

R Jothi, B Raghavachari. (2004). Survivable network design: the capacitated minimum spanning network problem. *Information Processing Letters.* No. 91(4). P.183-190.

Samuel C.A. Pereira, Paulo S.A. Sousa. (2008). Uncertainty and information accuracy in adverse selection. *Economics Letters.* No.100 (3). P.321-325.

S Borne, E Gourdin, B Liau, et al. (2006). Design of Survivable IP-over-optical networks. *Annals of Operations Research.* No. 146(1). P.41-73.

S Soni, H Pirkul. (2002). Design of Survivable Networks with Connectivity Requirements. *Telecommunication Systems.* No. 20(1-2). P.133-149.

T. Lux & S. Schornstein. (2005). Genetic learning as an explanation of stylized facts of foreign exchange markets. *Journal of Mathematical Economics*. No.41 (1-2). P.169-196.

Wu, Jun & Tan, Yuejin. (2005). Study on Measure of Complex Network Invulnerability. *Journal of Systems Engineering.* No. 20(2). P.128-131.

XF Wang & G Chen. (2002). Pinning control of scale-free dynamical networks. *Physica A: Statistical Mechanics and its Applications.* No. 310(3-4). P.521-531.

Yan, Binyu, Lu, Wei & Huang, Rui. (2002). A Study on OSPF Routing Protocol. *Journal of Sichuan University (Natural Science Edition).* No.39 (3). P.460-464.

Yue, Chengjun, Yin, Fengjie & Jing, Yuanwei. (2009). Multicast Routing Based on Information Entropy Multi-population Genetic Algorithm. *Computer Engineering.* No.35 (11). P.205-206 & 209.

Zhang, Yu & Yang, Xiaojing. (2004). Research on the Simulation Method of Local Communication Network in Field Zone. *Computer Engineering.* No. 30(8). P.103-105.

Zhong, Lianjiong & Xu, Feng. (2003). The Computer Network Algorithm of Invulnerability. *Fire Control & Command Control.* No. 28(z1). P.113-114.

Table 1. Deviation rates of LSRP and DVRP

	Variance of simulation data	Variance of reality data	Deviation degree	Deviation rate
LSRP	3407.471	3526.366	118.8948	0.033716
DVRP	3979.841	3996.7803	16.93931	0.004238

Table 2. Measurement summary of LSRP and DVRP

Time group	Tc_{LSRP}	Tc_{DVRP}
1-40	141483.3020	126945.2688
41-80	778866.0793	657273.7579
81-120	1584038.8700	1562649.7680
121-160	2694021.8310	2219259.6360
161-200	4136163.0950	4615592.7640
201-220	2112812.5600	2385727.9990
1-220	22189385.4300	24120264.1900

Table 3. Comparison and analysis of measurement results

Time period	1-160	161-220	1-220
LSRP	High time costs	Low time costs	Low time costs
DVRP	Low time costs	High time costs	High time costs
Causes	The computational complexity of LSRP is higher than DVRP, exceeding 1.52 times of convergence time in the simulation environment (seen in Figure 3).	The routing of LSRP doesn't change after the routing is convergent, and except for LSA information interaction, other routing information is not sent. DVRP should send the routing table to the neighbor routings in the period time, and this action occupies CPU and bandwidth time (seen in Figure 2 and Figure 4).	LSRP sends less information, occupies less bandwidth, and fits for large-scale network.

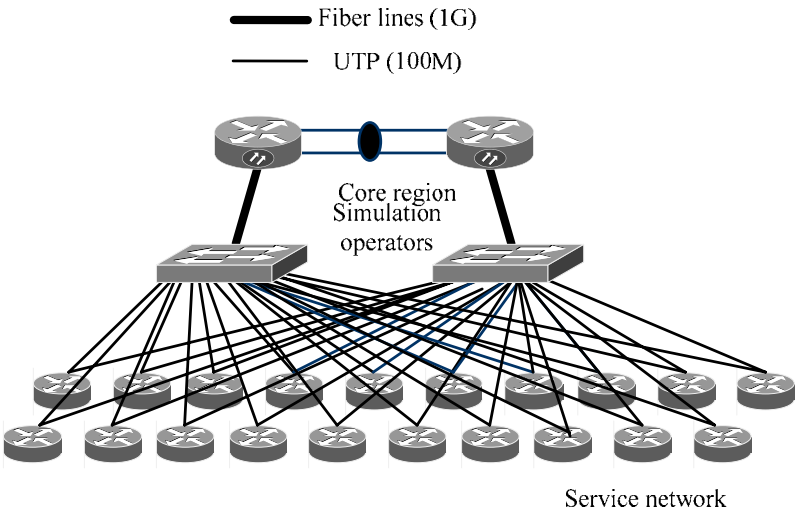


Figure 1. Simulation Network Topology

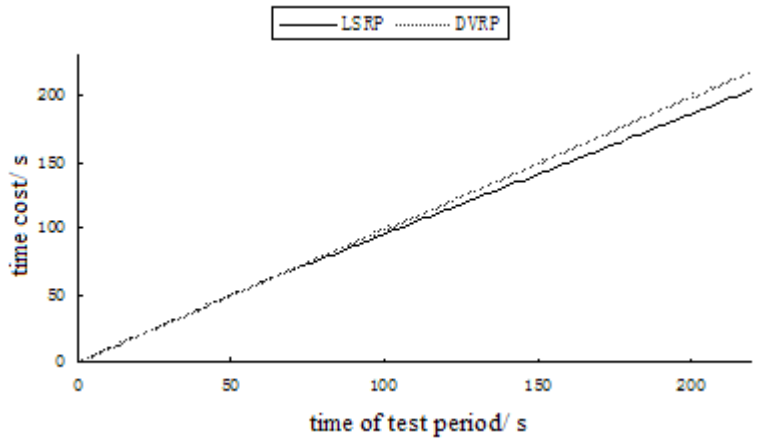


Figure 2. Comparison of Time Costs of LSRP and DVRP

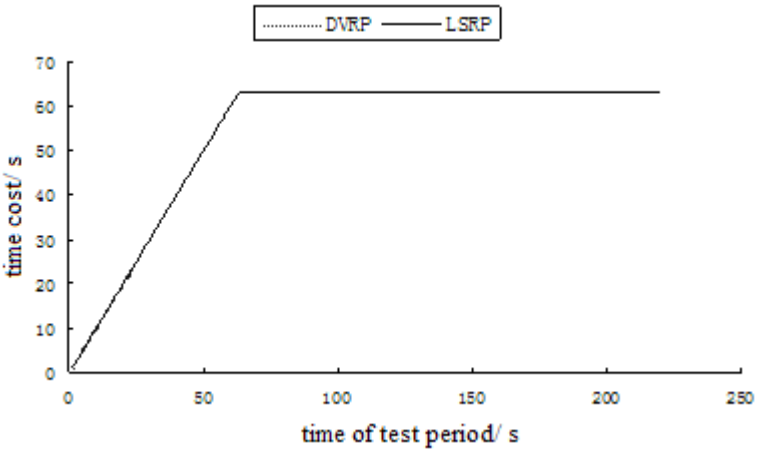


Figure 3. Comparison of Convergence Time Costs of LSRP and DVRP

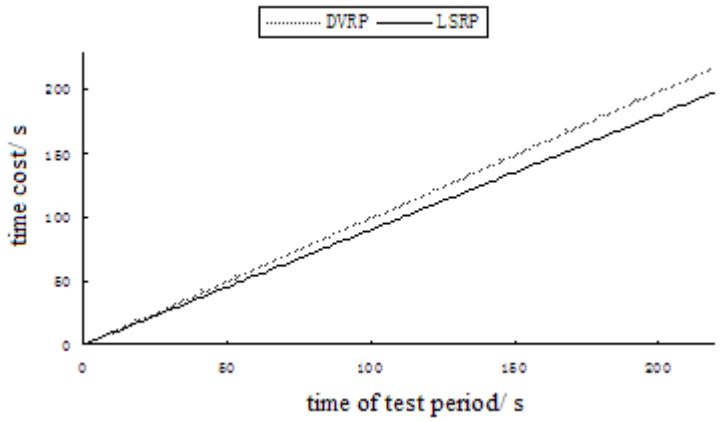


Figure 4. Comparison of Bandwidth Time Costs of LSRP and DVRP

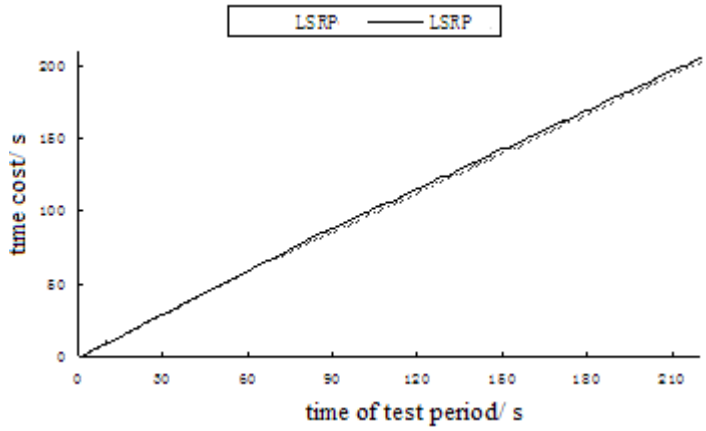


Figure 5. Comparison of Simulation and Reality Comprehensive Time Costs of LSRP

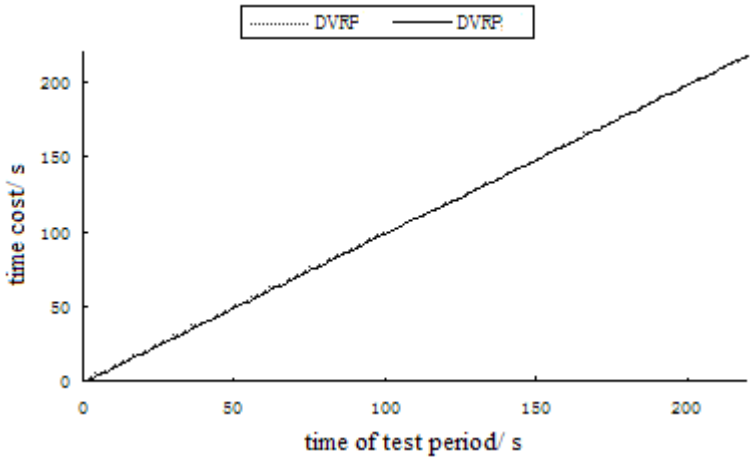


Figure 6. Comparison of Simulation and Reality Comprehensive Time Costs of DVRP

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Interacted Multiple Ant Colonies Optimization Approach to Enhance the Performance of Ant Colony Optimization Algorithms

Alaa Aljanaby (corresponding author)
Graduate Department of Computer Science
College of Arts and Sciences, Universiti Utara Malaysia
06010 Sintok, Kedah, Malaysia
E-mail: alaa.aljanaby@gmail.com

Ku Ruhana Ku-Mahamud
Graduate Department of Computer Science
College of Arts and Sciences, Universiti Utara Malaysia
06010 Sintok, Kedah, Malaysia
E-mail: ruhana@uum.edu.my

Norita Md. Norwawi
Faculty of Science and Technology
Universiti Sains Islam Malaysia
71800 Nilai, Bandar Baru Nilai, Malaysia
E-mail: norita@usim.edu.my

Abstract

One direction of ant colony optimization researches is dividing the ants' population into several colonies. These colonies work together to collectively solve an optimization problem. This approach offers good opportunity to explore a large area of the search space. This paper proposes a new generic algorithmic approach that utilized multiple ant colonies with several new interaction techniques. Computational test shows promising results of the new approach. The proposed approach outperforms the single colony ant algorithms in term of solution quality with the same computational effort.

Keywords: Ant colony optimization, Combinatorial optimization problems, Stagnation problem

1. Introduction

One of the successful applications of swarm intelligence is the application of Ant Colony Optimization (ACO). Swarm intelligence is a field of artificial intelligence that studies the intelligent behavior of groups such as the behavior of natural systems of social insects like ants, bees, wasps, and termites. ACO is inspired from the foraging behavior of real ants. Using a combination of priori information (heuristics) about the candidate solutions quality of and posteriori information (pheromone) about the goodness of the previously obtained solutions are the key elements of ACO success (Dorigo & Stützle, 2002; Blum, & Dorigo, 2005; Dorigo Birattari, & Stützle, 2006).

The problems tackled by ACO are called combinatorial optimization problems. These complex problems appear when the task is to get the best solution out of many possible solutions to a given problem. Traveling salesman problem (TSP), vehicle routing problem, quadratic assignment problem (QAP), sequential ordering problem and network routing problem are some well known examples of these problems (Dorigo & Stützle, 2002; Blum & Roli, 2003).

Ant System (Dorigo, Maniezzo, & Colorni, 1996), Ant Colony System - ACS (Dorigo & Gambardella, 1997), Max-Min Ant System - MMAS (Stützle &. Hoos, 2000), Ranked Ant System - RAS (Bullnheimer, Hartl, & Strauss, 1999) and Best Worst Ant System – BWAS (Cordon, Fernandez, Herrera, & Moreno, 2000) are several well known ACO

algorithms. These algorithms show interesting performance but are still far from being ideal. These algorithms can get a good solution at the early stages of the algorithm execution. However all ants speedily converged to a solution and the algorithm is unable to improve that solution. This is a common problem that all ACO algorithms suffer from regardless of the application domain. This is called search stagnation problem and the chance of stagnation proportionally increases with the increase of the problem size.

Multiple Ant Colonies Optimization (MACO) is an ongoing direction of ACO researches that aims to improve the performance of ACO algorithms. In this approach several colonies of ants are cooperatively working to solve a combinatorial optimization problem (Jong & Wiering, 2001; Kawamura, Yamamoto, Suzuki, & Ohuchi, 2000). MACO increases the chance of ACO algorithms to explore a large area of the search space and hopefully find (near-) optimal solution.

This paper considers the enhancement of ACO solution by proposing a new algorithmic approach that utilizes multiple ant colonies with certain techniques to organize the activities of these colonies. The rest of this paper is organized as follows. Section 2 reviews the related literature while section 3 proposes the new algorithmic approach. The computational results are presented in section 4. Concluding remarks and suggested future work are presented in the final section.

2. Related work

Jong and Wiering (2001) proposed a multiple ant colonies system for bus-stop allocation problem. In this problem there are n bus-stops and m bus-lines. A solution is to build m bus-lines, each one consisting of a sequence of n bus-stops that minimizes the average travel time. The results of the new algorithm were better than the results obtained from other meta-heuristic algorithms like greedy algorithm and simulated annealing.

Kawamura et al. (2000) proposed a MACO algorithm based on colony level interaction. A colony effects other colonies and these effects are different from one colony to another. These effects are determined by using an array $M \times M$ parameters, where M is the number of colonies. No exact way of choosing this large number of parameters was shown. The algorithm was tested on some TSP instances and the results were comparable with AS results but cannot outperform the results of the best known ant algorithms like ACS and MMAS.

Middendorf, Reischle, & Schmeck (2002) proposed the idea of using several ant colonies parallelized over several processors. The parallelized multiple colonies have to exchange information after completing a certain number of iterations. They tested their algorithm on one TSP instance and one QAP instance. The results showed that the algorithm with moderate number of colonies gave better results than AS for the TSP while the results for the QAP instance was worst than the results of ACS. Just like Kuwamura et al. (2000) the results of Middendorf et al. (2002) were also not compared with the results of ACS and MMAS. In fact the results presented in both papers were worst than those of ACS and MMAS.

3. The proposed interacted multiple ant colonies approach

The proposed algorithmic approach contains three mechanisms that are used to organize the work of the individuals in each colony and the work of all colonies. In other words there are two levels of interaction. The first is the colony level and the second is the population level.

The colony level interaction can be achieved through the pheromone depositing process within the same colony. The pheromone updating mechanism is responsible for the implementation of this kind of interaction. The population level interaction is achieved by evaluating the pheromones of different colonies using some evaluation function. The important aspect here is the pheromone evaluating mechanism. The degree of interaction of the different colonies is the role of the exploration / exploitation mechanism. This algorithmic approach will be called hereafter Interacted Multiple Ant Colonies Optimization (IMACO).

The work activities of a single colony in the proposed IMACO algorithm are based on ACS. Each colony has its own pheromone that is used as the interaction mechanism between the ants of the same colony. The interaction between ant colonies using pheromone can be organized in different terms. The IMACO algorithm is described as follows. M colonies of m ants each are working together to solve some combinatorial problem. The probabilistic decision of the ant k belongs to the colony v to move from node i to node j is defined as:

$$j = \begin{cases} \arg \max\limits_{l \in N_i^{kv}} \{ f(P_{il})\, H_{il}^{\beta} \} & \text{if } \; q \le q_0 \\ S & \text{otherwise} \end{cases} \tag{1}$$

where $f(P_{ij})$ is the evaluation function of pheromone on the edge (i,j) and H_{ij} is the problem dependent heuristic. N_i^{kv} is the set of remaining nodes to be visited by the k^{th} ant of colony v. β is a parameter that determines the relative

importance of pheromone versus heuristic, q is a random variable distributed in [0, 1] and q_0 is a parameter and $0 \le q_0 \le 1$. S is a random variable selected according to the following probabilistic rule.

$$S = \begin{cases} \dfrac{f(P_{ij})\, H_{ij}^{\beta}}{\sum_{l \in N_i^k} f(P_{il})\, H_{il}^{\beta}} & \text{if } j \in N_i^{kv} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

3.1 Pheromone evaluation mechanism

The proposed pheromone evaluation mechanism evaluates the pheromone as an average of the pheromone values of all colonies on some edge. This means that an ant will make its decision to choose an edge based on the average of the available experiences of ants of all colonies that visited this edge in the past. This variant of IMACO is referred hereafter as IMACO-AVG.

Given that for each edge there are M pheromone values each belongs to a single colony. Average pheromone evaluation function evaluates the pheromone on any edge as an average of the available M values. The pheromone evaluation function for IMACO-AVG is defined as:

$$f(P_{ij}) = \frac{\sum_{v=1}^{M} P_{ij}^{v}}{M} \tag{3}$$

where P^v_{ij} is the pheromone of colony v on the edge (i, j).

3.2 Exploration / exploitation control mechanism

Each ant makes a probabilistic decision when it needs to move to a new node. The probabilistic decision is based on heuristic information (cost) and pheromone information. Pheromone represents information about previous experiences of the ant's own colony and of the other colonies. While heuristic represent a priori information about the goodness of a solution. The relative importance of heuristic and pheromone information is determined by using the weighting parameter β.

Another parameter q_0 is usually used in ant's probabilistic decision as trade-off between exploitation (choosing the edge with the higher value of the multiplication of pheromone and heuristic values) and exploration (choosing the edge randomly according to some probability distribution).

In this paper β and q_0 are set to 2 and 0.9 respectively, these values are commonly used in ACS single colony algorithm. This study conduct experiments on cases using different values for the ant colonies for these two parameters. The goal of these tests is to reach a balanced form of exploitation / exploration that yields to the best algorithm performance.

3.3 Pheromone updating mechanism

The proposed pheromone updating mechanism encourages a balanced form of exploitation of previous experiences and exploration of new or improved paths. Basically the mechanism incorporates global and local pheromone updating. Local pheromone update encourages the exploration of new areas of the search space by reducing the importance of the visited edges. The global pheromone update encourages the exploitation of previously good solutions by giving extra weight to the edges of global best solutions.

Global pheromone updating includes that best ant of each colony deposits an amount of pheromone on its own path. After all ants of all colonies complete their tours (i.e., one algorithm iteration), the ant that finds the so far best solution in its colony will be allowed to deposit an amount of the colony's pheromone on the edges of its tour according to the following rule:

$$P_{ij}^{v} = (1-\sigma) P_{ij}^{v} + \sigma \Delta p_{ij}^{v.bs} \tag{4}$$

where σ is the trail evaporation such that (1- σ) represents the pheromone persistence. This parameter is used to avoid unlimited accumulation of pheromone trails and allows the algorithm to forget previously done bad choices. $\Delta P_{ij}^{v.bs}$ is the pheromone quantity added to the connection (i, j) belonging to the best solution of v^{th} colony $L^{v.bs}$ and is given by:

$$\Delta p_{ij}^{v.bs} = \begin{cases} 1/L^{v.bs} & if\ (i,j) \text{ belongs to the} \\ & \text{best tour of colony v} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The best solution used in the updating mechanism in ACS is the global best solution. This work considers the idea of using a combination global and iteration best solution. Using the iteration best solution after a certain number of using the global best solution in the pheromone updating mechanism helps to make a search diversion by directing the search process to different good solutions which avoids the quick convergence to a single solution at the early stages of the algorithm run.

This mechanism was implemented individually for each colony. The general updating policy used in this research is *xG 1I*, i.e., one update will be based on iterative best solution after x updating times of using global best solution. In this paper the results reported for x= 25, 50 and 75 along with the basic result without using this proposed updating policy.

Local pheromone update is then applied by each ant on the visited edges. It includes that each ants reduces the amount of pheromone on paths it uses in order to give a better chance to other paths to be chosen by future generations. It is a very important rule as it is performed during the solution construction which helps to yield different pheromone evaluation values for the same edge in the same iteration at different solution construction steps. The local pheromone update is given by:

$$P_{ij}^{v} = (1 - \gamma) P_{ij}^{v} + \gamma p_0 \tag{6}$$

where γ is another pheromone evaporation parameter and P_0 is the initial pheromone value.

4. Computational result

IMACO-AVG has been tested using lin318 TSP benchmark instances taken from TSP library (TSPLIB, 1995). The optimal solution for lin318 is 42029. Experiments were run using 1, 2, 3, 4, 5, 7, 9, 12, 15, 20 and 30 colonies where each colony has 10 ants. The results are averaged over 10 trials with 10000 iterations per trial. This is based on the assumption that the algorithm ran 3000 iterations for each 100 nodes of the problem instance. The parameter setting are β=2, $\sigma = \gamma = 0.1$ and $q_0 = 0.9$. The heuristic function used for TSP is the inverse of the distance, i.e., $H_{ij}=1/d_{ij}$.

Figure 1 shows the results of applying ACS, and IMACO-AVG lin318. For comparison purposes, ACS was run using 10, 20, 30, 40, 50, 70, 90, 120, 150, 200 and 300 ants. Note that since ACS has single ant colony, so in Figure 1 the number of ants used by ACS is equal to the number of colonies multiplied by 10. As shown in this figure, the results of ACS deteriorate as the number of utilized ants increases. This means that ACS cannot benefit from the increase in the number of utilized ants because the algorithm always get trapped in stagnation situation and cannot improve the solution quality. Better results are obtained when the number of ants is in the range of 20 to 30.

It is obvious that IMACO-AVG algorithm obtained better results in term of the overall average solution. The superior of IMACO-AVG is clear as this algorithm shows a stabilizing performance using the increased number of colonies. The average pheromone evaluation technique was a successful organizing technique of the ants' activities up to 30 utilized colonies.

Figure 2 shows the result of using IMACO-AVG with different pheromone updating techniques. A better result is obtained when updating the pheromone based on iterative-best solution once after each 50 times of pheromone updating based on global best solution.

Figure 3 shows a comparison between the trial average time of IMACO-AVG and ACS with 11 runs for each algorithm with a specified number of ants/colonies. The computational average time of both versions of IMACO was close to that of ACS when the number of colonies utilized is 10 or less. However, the difference between ACS and IMACO average time starts increasing when the number of utilized colonies was beyond 10. Based on these results, the rest of the experiments focus on testing the different versions of IMACO with no more than 10 colonies.

Experiment was also conducted using IMACO-AVG 50G 1I with 9 colonies on different symmetric and asymmetric instances of TSP. The results are reported in Table 1 which depicts the comparison of ACS and MMAS. In this table, the optimal solution for each TSP instance is shown in each column header under the instance name and all reported results of all algorithms represent the overall average solution. The results of ACS and MMAS are taken from the literature (Stützle &. Hoos, 2000). The number of iteration IMACO-AVG ran on each instance was according to the assumption made by Stützle and Hoos (2000) which is equal to k*10000*no of nodes/no of ants where k=1 for symmetric instances and k=2 for asymmetric instances.

The results show that IMACO-AVG outperformed ACS for all instances. In comparing IMACO-AVG with MMAS (which is the best known ant algorithm) for the first small instances, MMAS perform better. However, the results of IMACO-AVG were very close. For other bigger instances where the chance of stagnation increases IMACO-AVG outperformed MMAS for giving the best average solution.

5. Conclusion and future work

IMACO, the new algorithmic approach divides the ants' population into several colonies and effectively coordinates

their works. The results show that IMACO-AVG can outperform the ACS algorithm with similar number of ants. IMACO-AVG has been further tested on different TSP and compared with ACS and MMAS and its superior performance was obvious. Even though IMACO is computationally more expensive than other ACS and MMAS, it is still within the same computational class when number of colonies is ten or less.

Testing new pheromone evaluation mechanisms is a possible future direction. Another interesting future work is testing different values for the parameters involved in the exploration / exploitation control mechanism and investigating the best range for each parameter.

References

Blum, C. & Dorigo, M. (2005). Search bias in ant colony optimization: On the role of competition-balanced systems. *IEEE Trans. on Evolutionary Computation*, 9(2), 159-174.

Bullnheimer, B., Hartl, R. F., & Strauss, C. (1999). A new ranked-based version of the ant system: A computational study. *Central European Journal for Operations Research and Economics*, 7(1), 25-38.

Cordon, O., Fernandez, I., Herrera, F., & Moreno, L. (2000). A new ACO model integrating evolutionary computation concepts: The Best-Worst Ant System. In M. Dorigo, M. Middendorf, & T. Stützle (Eds.), *proceedings of ANTS2000-From Ant Colonies to Artificial Ants: A Series of International Workshops on Ant Algorithms* (pp. 22-29). IRIDIA, Université Libre de Bruxelles, Brussels, Belgium.

Dorigo, M. & Stützle, T. (2002). The Ant Colony Optimization Metaheuristic: Algorithms, Applications, and Advances. In: F. Glover and G. Kochenberger (Eds.), *Handbook of Metaheuristics*. Kluwer Academic Publishers.

Dorigo, M., & Gambardella, L. (1997). Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Trans on Evolutionary Computation*, 1(1), 53-66.

Dorigo, M., Birattari M. & Stützle T. (2006). Ant Colony optimization: Artificial Ants as Computational Intelligence Technique. *IEEE Computational Intelligence Magazine*, Nov. 2006.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics-Part B,* 26(1), 29-41.

Jong, J., & Wiering, M. (2001). Multiple ant colony system for the bus-stop allocation problem. In *Proceedings of the Thirteenth Belgium-Netherlands Conference on Artificial Intelligence (BNAIC'01)*, 141–148.

Kawamura, H., Yamamoto, M., Suzuki, K. & Ohuchi, A. (2000). Multiple ant colonies algorithm based on colony level interactions. *IEICE Trans. Fundamentals*, E83-A(2).

Middendorf, M., Reischle, F., & Schmeck, H. (2002). Multi Colony Ant Algorithms. *Journal of Heuristics* (special issue on Parallel Meta-heuristics), 8(3): 305-320.

Stützle T., & Hoos H. (2000). Max-Min Ant System. *Journal of Future Generation Computer Systems*, 16, 889–914.

TSPLIB(1995). http://www.iwr.uni-heidelberg.de /groups/comopt/software/TSPLIB95/tsp/.

Table 1. Results Comparison for Symmetric and Asymmetric TSP Instances

Algorithm	TSP instance						ATSP instance			
	kroA100	***d198***	***lin318***	***att532***	***rat783***	***fl1577***	***ry48p***	***ft70***	***kro124***	***ftv170***
	21282	***15780***	***42029***	***27686***	***8806***	***22137***	***14422***	***38673***	***36230***	***2755***
ACS	21420.0	16045	43296.85	28522.8	9066.0	23136.0	14565.4	39099.0	36857.0	2826.5
MMAS	21291.6	15956.8	42346.60	28112.6	8951.5	NA	14523.4	38922.7	36573.6	2817.7
IMACO-AVG	21290.6	15953.5	42341.6	28075.7	8932.8	22520.3	14491.6	38871.4	36489.2	2791.7

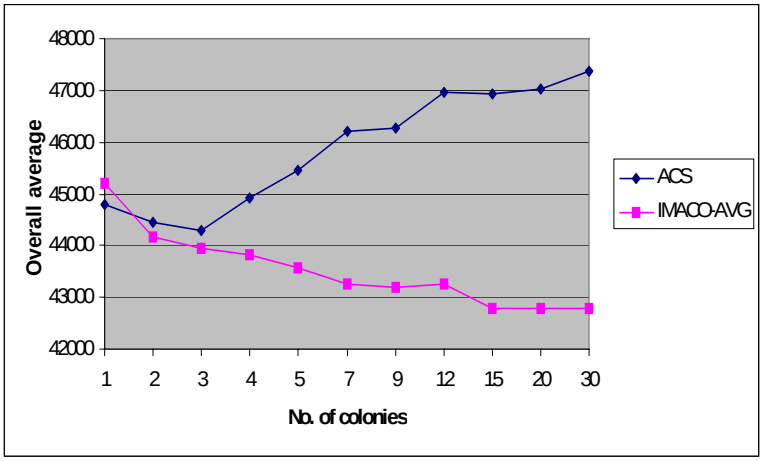


Figure 1. ACS and IMACO-AVG performance comparison

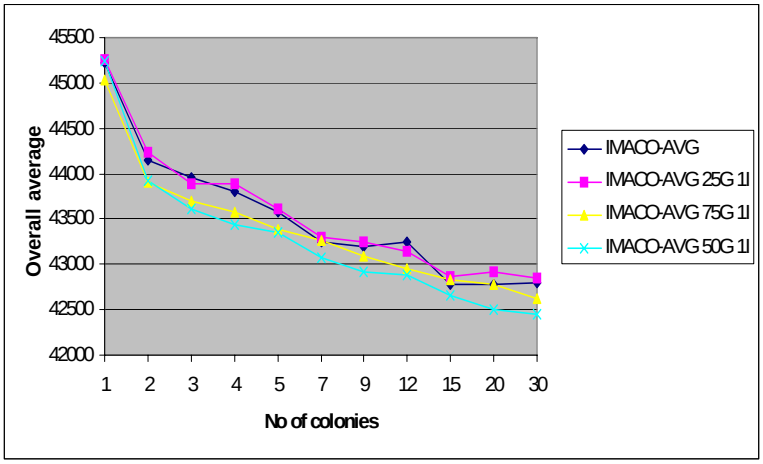


Figure 2. Different IMACO-AVG variants performance comparison

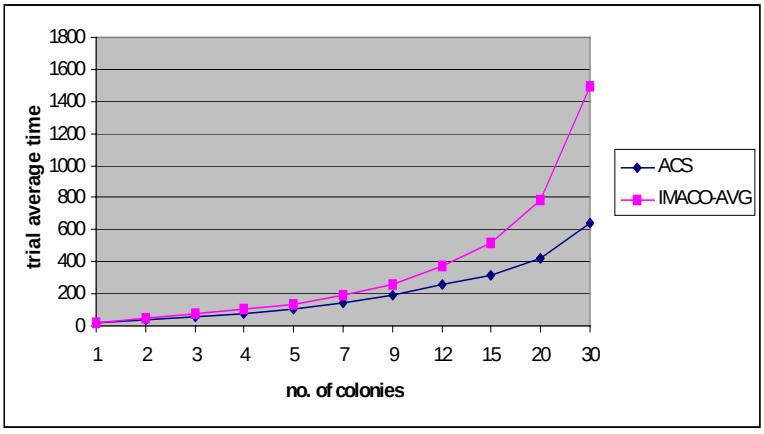


Figure 3. ACS and IMACO-AVG trial average time Comparsion

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

A New Method of Reducing Pair-wise Combinatorial Test Suite

Lixin Wang (Corresponding author)
College of Information Science and Technology, Donghua University, Shanghai, China
2999 Renmin Road North, Songjiang District, Shanghai, 201620, China
&
School of Electronics and Information Engineering
Anhui Institute of Architecture and Industry, Hefei, China
856, Jinzhai Road, Baohe District, Hefei, Anhui, 230022, China
E-mail: wlx@mail.dhu.edu.cn

Renxia Wan
College of Information Science and Technology, Donghua University, Shanghai, 201620, China
E-mail: xxrrww2002@163.com

Abstract

The biggest problem for combinatorial test is a numerous number of combinations of input parameters by combinatorial explosion. Pair-wise combinatorial coverage testing is an effective method which can reduce the test cases in a suite and is able to detect about 70% program errors. But, under many circumstances, the parameters in programs under test (PUTs) have relations with each other. So there are some ineffective test cases in pair-wise combinatorial test suites. In this paper, we propose a method of reducing ineffective combinatorial test cases from pair-wise test suite. The main ideas of the method is that we firstly analyzes the dependent relationships among input parameters, then use the relationships to reduce ineffective pair-wise combinations of input parameters, and lastly generate the pair-wise combinatorial coverage test suite. The experiments show that the method is feasible and effective, and considerably reduce the number of pair-wise combinatorial test cases for some programs under test.

Keywords: Pair-wise combinatorial test, Test suite, Dependent relationships

1. Introduction

The combinatorial test is one of main approaches in black-box tests. The combinatorial test creates test suites by selecting values of input parameters and combining these parameter values as test cases(Mandl, 1985). These parameters include program configuration parameters, internal events, user inputs, and other external events (Shiba, et al. 2004). Since there are often too many parameter combinations, so it is impossible to test all possible combinations of input parameters for many programs under test (PUTs). Pair-wise combinatorial testing is an effective method which can decrease the number of test cases in a suite and is able to detect about 70% program errors. Many researchers have presented a variety of methods (Cohen, et al. 1997) (Shiba, et al. 2004) (Tai, et al. 2002) (Schroeder, et al. 2000) to generate pair-wise combinatorial test suites and developed over ten tools to generate pair-wise test cases (Czerwonka.2009). But these methods do not consider the relationships of input parameters for a PUT.

In the definition of combinatorial test, a precondition is that the input parameters of a PUT are independent with each other. But, under many circumstances, these parameters have relations with each other and we often need test the PUT using combinatorial testing. So there are many invalid pair-wise combinations of input parameters due to the input parameters relationships so as to have some redundant and ineffective test cases in test suite, and degrade the performance of algorithm for generating test suite. So it is necessity to find some methodology to reduce the pair-wise combinatorial test suite.

In this paper, we present a new method to remove the ineffective pair-wise test cases from combinatorial test suite. The method is based on the using the dependent relationships of input parameters to achieve the aim of reducing pair-wise combinatorial test suite. This paper is constructed as follows: Section 2 analyzes the ineffective test cases in test suite.

Section 3 describes how to obtain the pair-wise combinations of input parameters. Section 4 elaborates the process of reducing the pair-wise combinatorial test suite. The experiment results and analysis are given in section 5, and the conclusions are presented in section 6.

2. Analysis of the ineffective pair-wise combinatorial test cases

2.1 The pair-wise combinatorial test

In the combinatorial test method, we generates test suite that cover all pair-wise, triple, or *n*-way combinations of test parameters specified in formal test requirements.

We usually generate pair-wise coverage test suite to test PUTs. Kuhn D. R. et al. found that about 70% field faults were caused by either incorrect single values or by an interaction of pairs of values in an empirical study of user interface software. David M. Cohen in their code coverage study also indicated that pair-wise coverage is sufficient for good code coverage (Cohen, et al. 1997).

See Table I, Cohen et al. tried several different models for the sort command. Table I shows the coverage data for two of them. The models differed from each other in two ways. First, the number of parameters were varied, which is called width, and the maximum number of values for a parameter, which is called height. In general, the number of pair-wise test cases is at least the product of the number of values for the two parameters with the most values. Consequently, rewriting a model to decrease the maximum number of values per parameter can decrease the number of tests, even if it increases the number of parameters.

Pair-wise combination is that for any two parameters p_1 and p_2 and any valid values v_1 for p_1 and v_2 for p_2, combine the v_1 and v_2 as a element into a set of combination. Covering all pair-wise combinations means that for any two parameters p_1 and p_2 and any valid values $v1$ for p_1 and v_2 for p_2, there is a test suite in which p_1 has the value v_1 and p_2 has the value v_2. We call the test suite as pair-wise combinatorial test suite. We use the pair-wise combinatorial test cases as input to test the PUT, which is called pair-wise combinatorial coverage test.

But, since there are dependent relationships between input parameters, the set of pair-wise combination often have some ineffective test cases. How do we find these ineffective test cases? And how do we reduce these ineffective test cases?

2.2 Ineffective test cases in pair-wise combinatorial test suite

Sometimes we need use combinatorial testing to test some PUTs with dependent relationships of input parameters. So there will be some ineffective pair-wise test cases in test suite.

For example, with respect to a PUT, we test an inquiry interface of the PUT, which has year, month, day, and other selections. After we select February (not other values of months), there are 1, 2, …, 28 or 29 in the day item list, but there will be not 30 and 31. In other words, we do not obtain some test cases such as (year-number, February, 30) and (year-number, February, 31), since there have been some relative restrictions among this input parameters in program codes, and some relative tests may have been conducted during white-box test or component test. So, in system or integrate test, we can find that the values of input parameter days are depend on the values of input parameter month in the PUT, and the combinations of (year-number, February, 30) and (year-number, February, 31) are ineffective.

Above example shows that a dependent relationship between two input parameters exists when the value of one input parameter implies some restrictions on the values of another input parameter. In this case, some input parameter combinations are not allowed and the total space of input parameter combination can be partitioned into the allowed and not allowed.

Definition 1: Suppose x_i and x_j are two input parameters, $i<j$, $a \in x_i$, $b \in x_j$. The pair-wise combination (a,b) is ineffective if and only if the (a,b) can not satisfy the user's demands, don't exist or is not able to be obtained.

For a pair-wise combination set with some ineffective combinations, when we generate the test suite which covers the combination set, the test suite will include some ineffective test cases. The definition of the ineffective test case is as Definition 2.

Definition 2: Suppose x_i and x_j are two input parameters, $i<j$, $a \in x_i$, $b \in x_j$ and (a,b) is a ineffective pair-wise combination, if a test case t denotes as $(v_1,v_2,\ldots, v_i,\ldots, v_j,\ldots,v_n)$, and $v_i=a$ and $v_j=b$, then we call the t is an ineffective pair-wise combinatorial test case.

In above-mentioned example, if a test case includes (year-number, February, 30) or (year-number, February, 31), then the test case is an ineffective combinatorial test case; if a test case includes (February, 30) or (February, 31), then the test case is an ineffective pair-wise combinatorial test case. For the pair-wise combinatorial coverage test, these ineffective pair-wise combinatorial test cases must reduce from test suite. But if we reduce these ineffective test cases in the manual analysis way, this will be very troublesome. In some complex cases, it may not be able to find all of these ineffective test cases. Therefore, we must adopt formal way to analyze and find all of the ineffective test cases.

3. Obtaining a pair-wise relationship set of input parameters

3.1 The dependent relationships among input parameters

Schroeder and Korel (Schroeder & Korel, 2000) used Input-Output relationships to reduce combinatorial test cases. Based on their idea, many researchers (Saraph, et al., 2003) (Cheng, et al., 2003) offer a variety of methods to reduce the combinatorial test suite via Input-Output relationships. But they all do not consider the dependent relationships among input parameters.

For many PUTs, the dependent relationships among some values of one input parameter and the values of another parameter are sometimes able to be obtained and useful. In these cases, not all combinations of values of input parameters are meaningful or valid for combinatorial testing, and the dependent relationships among input parameters should be systematically consider during modeling and generating combinatorial test suite.

If we use the dependent relationships of input parameters to reduce the combinatorial test suite, we must analyze and obtain these dependent relationships firstly. For the sake of the convenience of the discussion, we present two hypotheses as follows.

The dependent relationships of input parameters are defined and there is an existing set of dependent relationships for a PUT.

The dependent relationships of input parameters are one way and have transmissibility, but have no circulation.

There are also some dependent relationships between two parameters or among more than two parameters. There are some kinds of dependent relationship as follows (Silberschatz, et al., 2005):

One-one(1:1): The subset or all of values of one parameter dependent on the subset or all of values of another parameter, such as B dependents on A in Figure 1.

One-many(1:*n*): The subset or all of values of one parameter dependent on the subsets or all of values of many other parameters, such as D and E dependents on C in Figure 1.

Many-one(*n*:1): The subsets or all of values of many parameters dependent on the subset or all of values of one parameter, such as D dependents on B and C in Figure 1.

Many-many(*m*:*n*): The subsets or all of values of many parameters dependent on the subsets or all of values of other more than one parameter.

In these four types of relationships, 1:1 type of relationship is called two-parameter dependence relationship, which we can use the relationships of this type directly. The implication of 1:n type of relationship is one depended parameters decide n depending parameters, which sometimes we may transform some relationships of 1:n type into n relationships of 1:1 type and then use them. Whereas, n:1 and m:n type of relationships can not be used to obtain the pair-wise combinations of input parameters, but can be used to reduce the test suite.

We use the more succinct mark form to label the dependent relationships.

(a:b) represents parameter value b depends on parameter a(1:1 type).

($a \wedge b$:c) represents parameter value c depends on two parameter values a and b(n:1 type).

(a:$b \vee c$) represents parameter values b or c depends on parameter value a (1:n type). We may transform it into (a:b) and (a:c).

(a:$b \wedge c$) represents parameter values b and c depends on parameter value a (1:n type).

($a \wedge b$:$c \vee d$) represents parameter values c or d depends on parameter values a and b (m:n type). We may transform it into ($a \wedge b$:c) and ($a \wedge b$:d).

($a \wedge b$:$c \wedge d$) represents parameter values c and d depends on parameter values a and b (m:n type).

For example, date includes year, *month* and *day*. Suppose the values of *year* belong to {1999, 2000}; the values of *month* belong to {1, 2, …, 12}; the values of *day* belong to {1, 2, …, 31}. We analyze the dependent relationships for several special dates. The dependent relationships among *year*, *month* and *day* are as follows:

(1) The *day* value 29 in Feb. depends on *year* value 2000 and *month* value 2.

(2) The *day* value 30 depends on *month* value 1,3,4,5,6,7,8,9,10,11,12.

(3) The *day* value 31 depends on *month* value 1,3,5,7,8,10,12.

So we have the dependent set of input parameters as fellows.

(1){2000:2∧29}.

(2){1∨3∨4∨…∨12:30} or {1:30, 3:30, 4:30, …, 12:30}.

(3){1∨3∨5∨7∨8∨10∨12:31} or {1:31, 3:31, 5:31, 7:31, 8:31, 10:31, 12:31}.

In (2) and (3), the elements in dependent sets of input parameters are all two parameters dependent relationships. The element value in pair-wise combinations must be in accordance with the dependent relationships.

3.2 Obtaining input parameters relationships

This is one of the main problems for our method to gather input parameters relationships. There are many ways of gathering input parameters relationships. We can analyze program documentation (e.g., program specifications, or a cause-effect graph, should it exist) for the relationships and we also can analyze program codes to obtain the relationships. This is called as static analysis. Another way of analysis is dynamic, which analyze returning information during program execution so as to obtain some information we need.

The static analysis of program code includes program constraint code or glue code among components. For example, there is a program code which has date input parameters. There are several situations as fellows.

(1)Non any constrains among date input parameters in program codes. Then we can not obtain any information on relationships among input parameters.

(2)Having a piece code of constrains, seeing Figure 2(a). We can obtain some information on relationships among input parameters via analyzing the code.

(3)Having some output statements in program codes, seeing Figure 2(b). We can obtain some information on relationships among input parameters via analyzing the returning or output information during program execution.

We can analyze these program documentation or program codes via the manual way. Due to the diversification of form of program documentation, the manual way is only able to be adopted. But the way is complex and low efficiency. There already are automated techniques of analysis program codes to help us obtain the information of input-output parameter relationships. But automated analysis of input parameter relationships is difficult since codes sometimes are too complex. Whereas the method of analyzing returning or output information has both advantages and limitations. We hardly obtain all input relationships in a program. We had better adopt all kinds of methods to obtain more information of input parameters relationships as possible.

4. Reducing pair-wise combinatorial test suite by using the dependent relationships

After we obtain the pair-wise dependent relationship set, we can reduce pair-wise combinatorial test suite using dependent relationship set. Since existing methods of generating pair-wise combinatorial test suite are mature, we may use them to reduce test suite. There are two main ways to use the set of dependent relationship of input parameters.

(1)Remove the ineffective pair-wise combinations from existing set of pair-wise combinations, then cover the set without ineffective pair-wise combinations to generate test suite.

(2)Modify the existing methods so that examine a test case according to the set of dependent relationship of input parameters and judge whether the test case is a ineffective test case before put it into the pair-wise combinatorial test suite.

For the algorithms of PIO and AETG, the first way is easy to obtain test suite, since the PIO and AETG need a precondition i.e. the covered pair-wise combination set π. So we may remove the ineffective pair-wise combinations in advance and this do easily. The second way is more complex relatively.

In order to attain our aim which is to obtain a effective test suite, we firstly give the algorithm by which determine whether a given test case is invalid test cases. See the Figure 3.

In this section, we have a detailed account of reducing process of pair-wise combinatorial test suite using dependent relationships of input parameters. But, what demands notice is that if a PUT has ineffective *k*-wise combinations, we will also remove them from test suite.

5. Experiments

In our experiments, there are mainly two parts. One is the implementation of generating a pair-wise combination set without ineffective combinations. Another is generate the pair-wise combinatorial test suite which covers the pair-wise combination set.

The empirical study in this paper gives the results of several applied cases. All tests are performed on PC with a 2.2 GHz Intel CPU and 512 Mb of memory. Computer O.S. is Windows XP and the program code is edited, compiled and run on C++ Builder 6. Before the program running, an initialized text file must be built in advance, in which includes the number of input parameters, the symbolic value sets for each input parameters and the set of dependent relationships of pair-wise input parameters.

In the empirical study, there are six are practical applied cases at all. Details of the six cases used in the empirical study, and we can see the empirical results in Table II. Table II mainly shows the time which is taken in removing these

ineffective 2-way combinations from the set of 2-way combinations. Although the parameters numbers of our subjects are not large, it is acceptable in consumption of time.

From Table II we can see the comparison of generating pair-wise test cases between two algorithms, and between before and after reducing ineffective 2-way combinations of input parameters. In general, the number of after reducing ineffective 2-way combinations is less than the number of before reducing ineffective 2-way combinations.

In Table II and Table III, the meaning of *Subject* Col. is that m^n: n-the number of input parameters, m-the number of input parameter values; $m^n * q^p$: $n+p$-the number of input parameters; and the number of n input parameters is m and the number of p input parameters is q.

6. Conclusions

This paper presents a method to reduce the pair-wise combinatorial test suite and remove the ineffective test cases from test suite so as to make the testing of PUTs smoothly. The method is based on analysis the dependent relationships among input parameters. The main thing in the method is how to obtain ineffective pair-wise combinations of input parameters. In addition, how to modify these existing methods is a key too. Experiments data tell us our method will have good effects in the number of test cases.

Although the method we propose reduces pair-wise combinatorial test suite, but still has some limitations for some PUTs. Future research work mainly focuses on how to obtain the pair-wise dependent relationships of input parameters from program documents or codes.

References

Cheng C., Dumitrescu A. and Schroeder P. (2003). Generating Small Combinatorial Test Suites to Cover Input-Output Relationships. *Proceedings of the Third International Conference on Quality Software (QSIC'03)*, 76-82.

Cohen D. M., et al. (1997). The AETG system: an approach to testing based on combinatorial design. *IEEE Transactions on Software Engineering*. 23(7), 437-444.

Czerwonka J. (2009). Pairwise Testing. [Online] Available: http://www.pairwise.org/tools.asp.

Mandl R. (1985). Orthogonal Latin squares: An application of experimental design to compiler testing. *Communications of the ACM*, 28(10), 1054-1058.

Saraph P., Last M., & Kandel A. (2003). Test Case Generation and Reduction by Automated Input-Output Analysis. *Proceedings of 2003 IEEE International Conference on Systems, Man & Cybernetics (SMC 2003),* Washington, D.C., USA, 5–8.

Schroeder P. J. & Korel B. (2000). Black-Box Test Reduction Using Input-Output Analysis. *Proceedings of the 2000 ACM SIGSOFT International Symposium on Software testing and Analysis*, Portland, Oregon, United States, 173-177.

Shiba T., Tsuchiya T., & Kikuno T. (2004). Using artificial life techniques to generate test cases for combinatorial testing. *Proceedings of the 28th International Computer Software and Applications Conference (COMPSAC'2004)*, Hong Kong, China, 72-78.

Silberschatz A., Korth F.H. and Sudarshan S. (2005). *Database System Concepts (Fifth Edition),* New York, NY: The McGraw-Hill press.

Tai K. C., & Lei Y. (2002). A test generation strategy for 2-dimension testing. *IEEE Transaction Software Engineering*, 28 (1), 109-111.

Table 1. Coverage of pair-wise tests for sort command

	Height	Width	Avoids	Cases	Block(%)	Decision(%)	C-uses(%)	P-uses(5)
A	10	13	0	126	92	80	75	72
B	4	23	0	41	95	86	77	75
Avg.					93.5	83	76	73.5

Table 2. These results of algorithm of removing ineffective combinations

No.	Subject	Num. of 2-way combinations	Num. of ineffective 2-way combinations	time (s)
1	4^3	54	5	<0. 1
2	4^4	96	7	<0. 1
3	5^4	160	15	<0. 1
4	5^5	250	28	<0. 1
5	6^4	240	27	<0. 1
6	$3^3*4^4*2^5$	148	7	<0. 1

Table 3. The comparison of generating test cases between two algorithms, and between before and after reducing ineffective combinations

No.	Subject	Number of pair-wise test cases (PIO)		Number of pair-wise test cases(AETG)	
		Before reducing	After reducing	Before reducing	After reducing
1	4^3	9	9	9	9
2	4^4	10	10	10	10
3	5^4	11	11	10	10
4	5^5	14	13	12	11
5	6^4	12	11	11	10
6	$3^3*4^4*2^5$	11	10	10	10

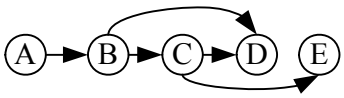


Figure 1. Dependent relationship illustration

```
if(month==2){
if(year%4==0&&(year%100!= 0||year%400== 0))
  {
day = 29;
  }
else{
 day = 28
  }
 }
else
if (month==4||month==6||month==9||month==11){
day = 30;
  }
else{
day = 31;
  }
```

(a)

```
if(month==2){
if(year%4==0&&(year%100!= 0||year%400== 0))
{ day = 29;
System.out.println(month+ "month value is :1-"+day);
}else{
day = 28;
System.out.println(month+ "month value is :1-"+day);
}
} else
if (month==4||month==6||month==9||month==11){
day = 30;
System.out.println(month+ "month value is :1-"+day);
}else{
 day = 31;
System.out.println(month+ "month value is :1-"+day);
}
```

(b)

Figure 2. Pieces of program codes

```
YNUselessTestCase(dependtR,t) //dependtR is the set of dependent relationships and t is a test case
{   r_i=1;dependtRNum;//the element number of dependent relationship set;
    lb1=0;lb2=0;//
    while (r_i<= dependtRNum)
    {   Obtain No. r_i dependent relationships in set of dependtR;
        Obtain the depending parameter value v_k in No. r_i dependent relationships;
        Obtain v_k subscript value k ;
        Obtain the No. k element value tv_k;
        lb1= compare(tv_k,v_k);
        If（lb1=0） {r_i=r_i+1;}
        Else
        {   n=Obtain the number of depended parameters in No. r_i dependent relationships;
            A=Obtain a vector which is made of the depended parameters in No. r_i dependent
                relationships;
            B=Obtain a vector whose elements is corresponding with A in test case t in subscript
                value;
            lb2=compareV(n,A,B);
            If（lb2=0） //
            {     removing test case t from test suite;
                  ri=ri+1;}
         }}
}
int compare(x,y) // It is a comparing function.
{   if(x==y) then return 1;
    else return 0;
}
int compareV(n,A,B) // It is a vector comparing function.
{   if(A==B) then return 1;
    else return 0;
}
```

Figure 3. The determining algorithm of ineffective test case

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Telecommunications Revolution: Implications on Criminality and Family Crisis in the South-South States of Nigeria

AGBA, A. M. OGABOH
Department of Sociology, University of Calabar, Calabar, Nigeria
Tel: 2-34-080-7272-7272 E-mail: ogabohagbagroup@yahoo.com

IKOH, MOSES
NIPOST, Benue State, Nigeria

USHIE, E.M. & BASSEY, A. O.
Department of Sociology, University of Calabar, Calabar, Nigeria

Abstract

In recent times telecommunications revolution especially the use of mobile phones is transforming the Nigerian society in many ways. The socio-economic and cultural impacts are enormous. As the spectrum of mobile phone usage increases, opinion differs among users on the impact of this revolution in Nigerian. This paper therefore set to investigate the impact of telecommunications revolution especially text based messages on criminality and family crisis in Nigeria. In doing this, the study elicited data through questionnaire from 7200 respondents who were randomly and purposively selected. Analysis of this data revealed that, most Nigerians were duped through text messages, while significant numbers of marriages are in crisis due to spouse assess to text messages that suggest suspicious of infidelity among partners. We recommended among others that Nigerian should scrutinize text messages before reacting to its content and that proper documentation of Sims pack users is necessary.

Keywords: Telecommunications revolution, Family crisis, Criminality information and communication technology

1. Introduction

The role of Information and Communication Technology (ICT) as a tool for development has received substantial attention since the turn of this century (Offurum, 2009; Keil & Johnson, 2005; Trevino, Webster & Stein, 2006). The term ICT has been used to encompass technological innovation and convergence in information and communication. The sector consists of a complex and heterogeneous set of goods, applications and services used to produce, process, distribute and transmit information (Nath, 2000). Its contribution to the emergence of information or knowledge societies(Marcelle,2000); changes in social interaction, economic and business practices, political engagement, education, health, leisure and entertainment, informed the United Nations Organization (UNO) declaration of May 24th every year as the World Communication Day (Offurum,2009).

The emergence of ICT has brought a new era in communication revolution to Nigeria. The internet, mobile phone and computer, all of which superseded hitherto used communication devices like cables and telegrams, have brought about a fundamental shift in patterns of communication and human relationships. These new technologies in communication satellites enable sound, images and messages, to reach target audience almost simultaneously. It is possible to record messages, stored them, retrieve and use at will. Communication revolution has brought about amazing social, economic, cultural and psychological transformation. It has reduced the globe into a village through reduction of time and space (Keil & Johnson, 2005). The crucial implications of this are numerous. It is promoting a culture of respect, dialogue and friendship (Offurum, 2009).

There is a growing understanding among researchers that Mobile-phone Mediated Communication (MMC) has made it possible for people to receive messages without having to meet face to face. Mobile phone thus becomes a powerful instrument for networking participation and advocacy within the society, advancing socio-economic development through the creation of new types of economic activities and employment opportunities. Text message is one type of MMC that allows for both one to one and one- to many textual communications without regard to an individual physical

location. It has provided the means to satisfy man's innate desires for connectedness and man's desire to reach beyond himself and commune with others.

In spite of the benefits of MMC in fostering contact, advancing economic and social development through enhanced network, various opinion polls suggest that MMC has been employed into criminal activities such as pornography, frauds, hates and obtaining goods by trick (OBT). Such ugly situation has received very little attention. There is equally no previous attempt to explore the impact of text based messages on family crisis. The family is the fundamental unit of society (Otte & Ogionwo, 2006; Giddens, 2006). It remains a vital ground for testing how people use objects (Carrier, 1990). Text based message through mobile- phone mediated communication (MMC) link families and individual members within the family, as well as the world beyond their front doors. How they do this or fail to do it involves a complex and often contradictory ways, which has threatened many marriages and the unity of many homes. Causal observation equally reveals the disruption of marriage engagements due to cheating discovered through MMC by either the fiancé or the fiacee. Such development is capable of generating ill feelings that can impact negatively on the well-being of society members. Accurate understanding of such developments is needed for dissemination so that innovative solutions that could enhance collaborative actions could be taken, to bring criminal tricks to the fore. It could help in developing statistics about emerging crime in a fast changing world. Specifically the paper intends to provide empirical examination the impact of text based messages through mobile-phones mediated communication (MMC) on criminality and family crisis in South-South Geo-political region of Nigeria.

2. Telecommunications revolution in Nigeria

Telecommunications revolution is transforming the Nigeria society in divers' ways since the dawn of the new millennium. The introduction of ubiquitous mobile phones that serve for chatting and text messaging is providing a major tool in this revolution. The emergence of the Global System Mobile (GSM) put into effective check the monopoly of the Nigerian Telecommunication Limited (NITEL) on telephones. The introduction of mobile phones provided the needed tools to bridge existing digital divide. While government deregulation and liberalization policies provide a conducive ground for the emergence of global system mobile network providers, it discouraged the monopoly hitherto enjoyed by the Nigeria Telecommunication Limited and thus encouraged a substantial private sector investment in the telecommunications industry.

Prior to the advert of telecommunication revolution low private sector involvement and poor technology in the telecoms industry between1960-1999, had left the country with only 40,000 connected telephone lines and about 25,000 analogues. Tele-density was 0.4 lines per 100 inhabitants. This period is often described as the "dark age" of telecommunications in Nigeria (Ndukwe, 2003). However, between 2000 and 2008, there was great change in the sector as more than 2000 licenses were given to digital service providers, fixed wireless operators, internet service providers, and long distance operators (Ndukwe, 2005[a]). Consequently, there was exponential increase in the numbers of telecommunication service providers. They include multinational companies like MTN, Vmobile now Zain, Globacom, Zoom, Starcom, Cisco, Intercellular, Mobitel, Multilinks, Mtel, Rainbownet, Reltel etc. This development led to annual growth rate of over one million lines and a market competition that reduces connection fees and increase access to telecommunication users. Teledensity also increased from 0.4 lines per 100 inhabitants in 1999 to 2.6 line per inhabitants in 2003. The number of connection on fixed lines equally increased from 450,000 to 1,120,000 line between 1999 and 2004 (Ndukwe, 2003, 2005[b&c]). A recent research conducted by Pyramid shows that Nigeria had overtaken South Africa as the biggest mobile phone telecom market in terms of subscribers. With 61 million subscribers in 2008, the market had grown by 23 per cent with about 8.4 billion Naira in overall telecom service revenue; and a penetration rate of 42 equally percent (Toure, 2009).

Beside the reduction in connection fee and or cost of lines, the need to bridge the digital divides equally pushed down the cost of mobile phones which were initially prohibited. As part of a project initiated by the Global Trade Association (GTA) for the world's GSM mobile operators (GSMMO), Motorola Company was selected in 2005 to supply mobile phones that cost less than thirty United State Dollars (US $30) to developing countries including Nigeria, India and South Africa (www.mongabay.com). The need to recognize Nigeria rural dwellers as information producers and users resulted in the installation of more satellite masks. With this development mobile phones have spread to the villages and are providing vital communication link to the rural people.

3. Theoretical and empirical survey

The advent of new technologies and the growing convergence of all media have major impact on information dissemination and communication. The new technologies on information and communication offer potentials for innovative social interaction, including peer and bottom-up communication (Marcelle, 2000). There exists equally creative opportunities for the creation, reproduction and dissemination of information relevant to economic development (Carrier, 1990). Three theories: the Technology Determinism Theory (Smith & Marx, 1994; Green, 2001), Actor-Network Theory (Latour, 1992, 1997 & Callon, 1999) and Social Presence Theory (Short, Williams &Christie,

1976), offer explanation for such creative opportunities. They equally seek to explain the relationship between technology and social change especially in the family, the basic unit of society.

According to Technology Determinism Theory (TDT) society's cultural values, social structure and history are all technology driven. The Theory posits that, rather than social context shaping technology, the uses made of technology determine the growth and development of the society. This implies that technology dictate users' behavior and action (Green, 2001). The implication of this postulates; is that cell phones (technology) exert large influence on the behavior of people including members of the family. The household factors such as marital status and relationship among couples could be improved or mar by mobile phone text-based messages; depending on the knowledge it exposes.

Actor-Network Theory (ANT) posits that, there is symbiotic relation between technology and its user. It holds that technology and human are equal interrelated actors in shaping society. Although, technology is created by humans its usage shapes human actions (Latour, 1992; Morphine & Potts, 2003). The implementation of ANT to this study is that, cell phone (technology) and human beings are not equal actors in shaping what happens, in society including the family. It is no surprise therefore, that societies with most innovative scientists, universities engineers and technologies are not able to solve complex problems. Meaning that though mobile phones influence what goes on within the family in term of relationship, the magnitude of the effect depends on spouse attitude towards one another.

Just like the ANT, Social Presence Theory (SPT) examines the social effect of communication technology on society. The theory posits that, the social consequences of communication technology dependence on the level of visual and physical contact the medium allows it users to have. It holds that the more social presence or personal contact, the greater the intimacy, immediacy and harmonious interpersonal rap pour, it creates among users. It argues that, communication devices that are text-based, e-mail oriented create less social contact and thus have little influence on society /users. Judging within the perimeter of this postulate, cell phone do not cause great changes in the family; this is because, it usage do not allow for great personal contact. However, its influence on society (family) cannot be relegated to the background.

Mobile telecommunication is a technology which allows electronic transfer of two-way voice, graphic and text data. While it maintains its portability and roaming capability, it equally offers a range of services such as call blocking and forwarding, packet and network data services. The compatibility of the mobile telecommunications network and terminal equipment with the information superhighway facilitates, internet browsing, short message services (SMS), mobile shopping, mobile banking and other cyber-mediated activities (Samuel, 2008).

The role of telecommunication in knowledge creation has severally been acknowledged (Friedman, 2006); though the raw supply of knowledge is booming through ICT, the fraction that is use productively is plummeting to ever-lower depths (Quah, 2006). And since knowledge cannot be added up arithmetically, the only way we can tell if knowledge is in excess supply is by asking whether it is underused. Such consideration influences the concern on the vulnerability of the knowledge society in both micro-and macro-knowledge perspective. The mobile phone is enriched with the faculties of creating databases, music downloads and financial transaction, among others. By macro-knowledge, it means that the great forces that move history, the impact of new ideas, mass movement and technologies, changes in political system, the evaluation of geopolitical relations as well as the transformation of cultures, reach us daily (Samuelson, 2006). Ability to use these for the benefit of the society becomes a compelling issue.

As the global use of mobile phones increases, there are mixed feelings about its impact on society. According to Lynn (2008), mobile phones have made their way into all aspects of our culture in an invasive and dominating way. As the society is becoming increasingly technological base, mobile phone has become a steady companion for both the young and the old. A study in Great Britain on mobile phone usage shows that 36 percent of the students surveyed were addicted to their phones. About 7 percent of them lost their jobs and relationship over mobile phone use (Steve, 2008).

Mobile phone use while driving is responsible for most accidents across the globe. Consequently countries are making laws that prohibit its usage while driving (Hall, 2007). Mobile phone is also seen as security risk. Wrong information and signals are easily spread via mobile phone, and this can easily escalate panic. It is for these reasons that Federal Officials in the United State of America after September 11 attacks, worked out plans to close mobile phone networks to everyone but government officials in the event of another scenario as September, 11 (Blair, 2001).

Mobile phones allow kids to reach their parents and vice versa in the event of emergency. Again during family crisis, mobile phone is a companion to the left alone child, that could be used to accessed others members of the family and friends. It also enhances medical care (Cohen, 2009). Unfortunately, despite the advantages of mobile phones in the medical field, they are responsible for a fivefold increase in brain cancer among children and teenagers in Great Britain (Lean, 2008).

In Nigeria, mobile phones have been used severally in the duping crime business, popularly known as 419. Victims have been asked to recharge phones to several amounts on the disguise of being neighbour or relatives, only to discover later that they have been duped.

4. Methodology

The data reported here were obtained from a field survey on the impact of telecommunication industry on socio-economic development of the Niger Delta Region (NDR) of Nigeria. A detailed description of the study is yet to be published. The NDR of Nigeria has nine states including the six-Akwa Ibom, Bayelsa, Cross River, Edo, Delta and Rivers – in the South- South Geo-Political Zone (SSGPZ) of the country. The six states have 3 senatorial district each and one hundred and twenty three (123) local government areas in all. The National Population Census (NPC, 2006) put the population at 21,014,655.

A cross –sectional survey which administered questionnaire on 7200 respondents elicited data for the study. Respondents whose age ranges from 18 to 58+ included men and women. Their marital status included those that were single, married, separated, divorced, widowed and widower. Other socio-economic characteristics of the respondents included type of occupation and educational attainment. Specific sections of the questionnaire elicited data on phone brand and phone line used by the respondents, the frequency of receiving and sending text based massages and experiences with text based massages and its implication.

The target population was residents of the six SSGPZ who subscribe to mobile telecommunications. We employed multistage random sampling techniques in the selection of respondents – with the help of the 1991 National Population Commission's (NPC) demarcated enumeration areas (EAs). From each state, we purposively selected one Local Government Area from each sensational district. We needed an urban based LGA where mobile phone subscribers are varied both in usage and network choice.

Beside, there is still a manifest digital derived between our rural and urban population. Urban dwellers still demonstrate the greatest affordability of access to mobile phones more than rural dwellers. Within each local government area, 10 enumeration areas were randomly selected, and within each enumeration area, 40 households were selected using a systematic random sampling technique for the administration of questionnaire. Consideration was given to gender equality in the administration of questionnaire such that if in the first household a man was interviewed, the next household would consider female respondents, except where such household was headed by a single parent, a bachelor or spinster. The study yielded a total sample size of 1,200 respondents a state and 7,200 respondents from the six states. Both descriptive and inferential statistics were employed in the analysis of data. We first presented the percentage distribution of responses for each variable and subsequently examined the bi-variants relationships between them. Prior to commencement of the study, the research protocol – including the study design, questionnaire, informed consent procedure and means of preserving respondents, confidentiality – was reviewed and approved by the research committee of AKMUM DEVELOPMENT CONSULTS LIMITED, Calabar, Nigeria.

5. Research findings

5.1 Socio-Demographic Back Ground of Respondents

As shown in Table 1, there was significant difference in the age range of the respondents [x^2, 8 (n=7,200) =2060.32, p=0.00]. More than a half of the respondents (N=3,736, 51.88%) were within the age bracket of 18-32 years. The overall means age was 34.96 (SD=5.91) years. The gender distribution was 3,694 (51.31%) men and 3,506 (48.69%) female. [X^2_3, (N= 7200) = 9.82, p = 0.00). Majority of them were married (N=3,818, 53.03%). The rest were either single (N=1,986, 27.58%), separated (N=437, 6.07%), divorced (N=898, 12.47%) or widowed/widower (N=61, 0.85%) [X^2, 4 (N=7200) = 7946, p=0.00]. One thousand nine hundred and eighty six thousand (27.58%) of the respondents earned their income from the civil service employment, while 2,192 (30.44%) were engaged in the private sector. There were more applicants (N=1740, 24.17%) than students (1,075; 14.93%) among the respondents. While 89 (1.24%) of the respondents were corpers serving the mandatory one year National Service, 647 (8.99%) of them were self-employed, [X^2, 5 (N=7500) = 2894.47, p=0.00].

Table 1 equally reveals the educational attainment of the respondents. The literacy level was high. Less than 4 percent of the respondents had the West African School Certificate. About 29 per cent (N=2,094) had Diplomas/OND/NCE, while 43.26 per cent (N=3115) of them held University first degrees and Higher National Diploma. About 24.06 percent (N=1,732) of them had higher degrees, [X^2, 3 (N=7,200) = 3107.38, p = 0.00]

5.1.1 Mobile phone ownership and phone lines

A significant difference in the type of phone used by the respondents was demonstrated during the study, [X^2, 7 (N=7200) = 3,945.71, P=0.00]. A large number of the respondents (N=2,256) representing 31.33 per cent of the populations of study owned mobile phones that were Nokia Brand. This was followed by Samsung (N=1253, 17.40%), Sagem (N=1095, 15.21%), Sunny Erikson (N=907, 12.60%) and Motorola (N=764, 10.61%). Acatel, Sendo and Simens brand were used by 449 (6.34%), 99 (4.15%) and 177 (2.46%) of the respondents respectively. This significant disparity is better demonstrated in Figure I. It suggests the popularity of the Nokia brand of mobile phones in the study area. Although several factors, ranging from price, design, distribution, etc (variable that were not considered in this study)

may contribute to the brands popularity. The study reveals that nearly all mobile phones brand available in the market were found in the study area.

When respondents choice of phone line was requested, more than 29 per cent of them (N=2107) were using MTN line. This was followed by Global Com network (N=1778, 24.69) and Zain (N=1368, 19%). 1204 (16.72%) of the respondents were using Etisalat, while Reltal, Starcom and Visafone were 339 (4.71%), 239 (3.32%) and 165 (2.29%) of the respondents, respectively [X^2, 6 (N=7200) = 4214.14, p=0, 00]. The disparity in line usage among respondents is demonstrated in Figure2. This finding suggests high patronage of MTN in the study area. Again this may be due to a number of factors, which were beyond the scope of this study.

5.1.2 Frequency of receiving and sending text based massages

Table 2 provides a comparative data of respondents' frequency of receiving and sending text based messages through mobile-phone mediated communication (MMC). A chi-square analysis of the difference, was significant [X^2, 4 (N=7200) = 81.27, P = 0.00]. Except on special occasions (occasionally) respondents frequency of receiving text based massages was higher than that of sending text based messages. For instance, on a daily basis, 2169 (30.13%) respondents acknowledged receiving text messages through their phone, while 2011 (29.93%) acknowledge sending text messages. In a similar manner, 5,729 (79.49%) received text based massages weekly while 4,819 (66.93%) send out text message per week. This finding was not different when the monthly frequency was examined. More receivers (N=6,249, 86.79%) than senders (N=5,315, 73.82%) of text based messages were recorded. This finding suggests the popularity of text based messages through mobile phones. Although nearly every text message received elicits a response, the findings suggest that many unsolicited text messages received are not responded to. However, occasion like birthday, festivities, etc, increases the frequency of receiving text messages than sending.

Surprisingly, there was no respondent who had never send nor received text message through his or her mobile phone. This suggests the multipurpose use of the mobile phone and the dependability of phone for receiving and sending text based messages.

5.1.3 Experiences with text based message through mobile phone mediated communication (MMC)

Respondents' experiences with mobile phone text messages were categories into three stages. At the first stage, we assessed respondents' experiences as criminal victims, through mobile encounter. At the second stage we assessed family/personal problem caused as a result of receiving and or sending mobile phone text based messages. Finally, we assessed the outcome of both stage 1 and 2 in terms of the lost suffered.

As shown in Table 3 respondents experiences significantly differed whether as criminal victims, or personal /family problem incurred through receiving and sending of text base messages [X^2,5 (N=7200) = 1405.32, P=0.00]. More than a half of the respondents (N=4,511, 62.65% experienced "attempts to defraud" through the mobile phone text message, while [X^2, 4 (N=7200, =8461.70, P=0.00] 1,005 (13.95%) were out rightly defrauded. The number of respondents that received text based messages that turned out to be lies (N=3,612, 50.17%) were higher than those who received false alarm (N=2, 135, 29.65%). A negligible number (N=98, 1.36%) received threats to their life through mobile phones.

When family/personal problems were assessed, 1,579 (21.93%) of the respondents said they discovered they were being cheated by their husbands through text based messages, while 1,243 (N=17.26%) of the respondents said they were being cheated by their wives through text based messages. Other respondents (N715; 9.93%) and (N271, 3.76%) believed they were being cheated by their fiancé and fiancé respectively through text based messages. The Table further reveals respondents experience of being cheated by their boy friends' (N=923, 12.82%) and "girl friends" (N=605, 8.4%) through text based message.

The outcome of falling victims to criminals through text based messages either through lies, frauds, threat to life and or false alarm; and being cheated by spouse, and or friends result in social problems that threaten peaceful co-existence. While 67 (0.93) of the respondents lost their life-saving as a result of mobile phone fraud, 125 (1.74%) of them have continued to live in fear. For majority of the respondent, (N=4,036, 56.06%), the discovery of being cheated through text based messages by their spouses and or friends had generated quarrel among them. Other had to resort to suspicion (N=1,336; 18.56%), fighting (N=918, 12.75%) and warning (N526; 7.31%). More than a third of the respondents (N=3,153, N=43.79%) had lost confidence in their relationship; while 78 (1.08%) of the respondents settled for divorce, 1007 (1.45%) of them called for separation.

6. Discussion

At the advent of information technology, media experts had predicted that it would usher in knowledge based society full of leisure (Smith, 2006). Emerging misinformation and distortion of information due to access by criminals has challenged that prediction. ICT is not only entertaining and informative but also a source of money for many users. And many users of ICT have been counting their lost. As revealed by this study text based message through mobile-phone

mediated communication (MMC) is popular among all mobile phone subscribers. There is hardly any mobile phone subscriber that has never received and or sent text based messages.

Despite the higher literacy level of the respondents in this study, many of them have been duped through text based messages sent through their phones. Others have received threat to life as well as false alarm. The implications of such text based messages are obvious. Many respondents who could not wait to verify the source of such information were defrauded of their savings. By conveying "threat to life" messages, mobile phone has become a source of fear, more so when the senders of such messages could not be traced.

Recently text based messages that warned subscribers about eating apple were received. The sender of the message even went on to emphasize that the National Agency for Food, Drug Administrator and Control (NAFDAC) was aware of the killer apple. In responses, the consuming public avoided apple. It was when apple sellers cried out that NAFDAC was aware of the text message, and inform the public that it was a false alarm. The implication of such alarm could result in ignoring genuine information in the future on the suspicion that it may be like the apple story.

Among young couples text based messages appear to have replaced letter writing. Love messages through text have received a prime of place. And because such messages could be stored and retrieved at will, other people who have access to the mobile phone could access such information. Others have stored pornographic photographs and six movies which they could watch at their convenience. Others have stored photographs of love ones, and dating alarm. In this study, there were counter accusations of cheating between husbands and wives, fiancé and fiancée as well as boyfriends and girl friends such accusations originate from access to and or outright suspicious that the other partner has been dating someone else or making love advances through text messages in mobile phones for many couples such discovery resulted in divorce (N=78, 1.08%), while other marriages ended in separation (N=107, 1.49%). Such could be traumatic and call into question basic human relationships.

Where families are constantly quarrelling (N=4,036, 56.06%) and suspicious (N=1,336; 18.56%), there is a tendency for continuous misunderstanding that could result in fighting (N=918, 12.75%). When families deteriorate to such level, child rearing and upbringing suffer. This has a serious implication for the society.

7. Conclusion

Telecommunications revolution especially the use of mobile phones is accelerating significant change in Nigeria society. There is a fundamental shift in the pattern of communication and human relationship with the introduction of text based messages through Mobile-phone Mediated Communication (MMC). This medium made it possible for people to receive and send messages without physical contact, and these messages can be retrieved at will. This revolution also fosters socio-economic development through the creation of new types of economic activities, employment opportunities, telecom- medical- care and human empowerment. Despite these advantages and more, findings from this study revealed that MMC has been employed into criminal activities such as frauds, threat to life, false alarm, lies etc. Text-based messages that suggest spouse cheating on the other has also been a source of family crisis and ill relationships among fiancé and fiancée, boyfriends and girlfriends etc. We therefore recommended that, legislature should be put in place to regulate and ensure that every sim card user in Nigeria is properly registered. And that spouses and the general public should scrutinize text messages before reacting to its contents.

References

Blair, J. (2001). A nation challenged: Telecommunications U.S considers restricting cell phone use in disasters. The New York Times. [Online] Available: Http//query.nytimes.com/gst/fullpage.html (August 25, 2009).

Callon, M. (1999).Some elements of a sociology of translation: Domestication of the scallops and the fishermen of saint brieuc bay. In M. Biagioli (Ed.), *The Science Study Reader (pp.*67-83). New York: Routledge.

Carrier, J. (1990). Reconciling commodities and personal relations in industrial society. In R. Silverstone E. Hirsch & D. Morley (Eds.), C*onsuming technologies (*pp. 1-29). London: Rodledge.

Cohen, M. (2009). Cell phones at school: should they be allowed? [Online] Available: http://life.familyeducation.com.cellular-telephones/school/51264.html (August 25, 2009).

Friedman, T. L. (2006, December -February). The exhausting race for ideas. *Newsweek (Special Edition.)* Pp. 11-12.

Giddens, A. (2006). *Sociology* (5th ed.) Malden: Polity Press.

Green, L. (2001). *Techno-culture.* Crows Nest: Allen and Unwin.

Hall, G. (2007). How to avoid the downsides of using cell phones. In L. Steve (Ed.), [Online] Available: http://usacellsearch.com/tag/cell-phone-use (August 25, 2009).

Keil, M., & Johnson, R.D. (2005).Feedback channels: Using social presence theory to compare voice mail to e-mail. *Journal of Information Systems Education, 13* (4), 295-302.

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W.Bijiker & J.Law (Eds.), *Shaping Technology/Building Sociology.* Cambridge: MIT Press.

Latour, B. (1997). On actor network theory: A few clarifications. [Online] Available: http://en Wikipedia.org/wiki/actor network Theory *(*November 22, 2007).

Lean, G. (2008). Mobile phone use raises children's risk of brain cancer fivefold. Independent News*.* [Online] Available:http//www.independent.co.uk./news/science/
mobile-phone-use-raises-children-risk-of-brain-cancer-fivefold-937007.html(August 25, 2009).

Lynn, K. (2008). In L. Steve (Ed.), *Cell phone etiquette*. http://usacellsearch.com/ tag/cell-phone-sue (August 25, 2009).

Marcelle, G (2000). The economics of ICT: Challenges and practical strategies of ICT used for women's economic empowerment*.* Kuala Lumpur: VAM.

Murphie, A. & Potts, J. (2003).*"I", culture and technology.* London: Palgrave

Nath, V. (2000). Empowerment and governance through information and communication technologies: Women's perspective. paper presented at the international conference on women in the new ict era: challenges and opportunities. kuala lumpur, malaysia, 22-24 october, 2000. [Online] Available:http:www.cddc.vt.edu/knownet/articles/women and ICT.htm(August 25, 2009).

National Population Commission. (2006). *National census report*. Abuja: Government Printing Press.

Ndukwe, E. (2003). The role of telecommunications in national development. 19thomolayole annual management lecture at Chartered Institute of Banks' Auditorium, Victoria Island, Lagos, Nigeria.

Ndukwe, E. (2005[a]). Three years of gsm revolution in nigeria. [Online] Available: http//www.ncc.gov.ng/speeches_presentations/Evc's(August 20, 2005).

Ndukwe, E. (2005[b]).Telecoms regulatory environment. [Online] Available: http//www.ncc.gov.ng/speeches_presentations/Evc's(August 20, 2005).

Ndukwe, E. (2005[c]). Developing telecommunications infrastructure. [Online] Available: http//www.ncc.gov.ng/ speeches_presentations EVc's(August 20, 2005).

Offurum, S.C. (2009, May 31). Communication revolution. The Leader, 1, 19, p. 2.

Otite, O. & Ogionwo, W. (2006). *An introduction to sociological studies.* Ibadan: Heinemann Educational Books.

Quah, D. (2006, December -February). Knowledge glut. Newsweek (Special edition). Pp. 43.

Samuel, K. J. (2008). Can mobile telecommunication reduce intra-city work travel? Empirical evidence from a third world city. *Ibadan Journal of the Social Sciences,* 6(1), 45-54.

Samuelson, R. .J. (2006- December, -February). The stealth factor. Newsweek (Special edition). Pp. 74-75.

Short, J.A., Williams, E. & Christie, B. (1976). *The social psychology of telecommunications.* New York John Wiley and Sons.

Silverstone, R., Hirsch, E., & Morley, D. (1992). *Consuming technologies.* London: Rodledge.

Smith, M. R. & Marx, L. (1994). Does technology drive? *The dilemma of technological determinism.* Cambridge: MIT Press.

Steve, L. (2008). The growing problem of cell phone addiction. [Online] Available: http://usacellsearch.com/tag/cell-phoneuse(August 20, 2009).

Toure, H, (2009, March). Mobile phone banking: View point. *Africa Global Perspective*. Pp. 56-57.

Trevino, L.K, Webster, J., & Stein, E.W. (2006). Making connections; Complementary influence on communication media choice, attitude and use. *Organization Science*, 2 (2), 163-182.

Table 1.

Variables	**N**	**%**	**M**	**SD**	**X^2**	**P**
Age			**34.96**	**5.91**	**2060.32_8**	**0.0000**
18-22	1019	14.15				
23-27	1433	19.90				
28-32	1284	17.83				
33-37	638	8.86				
38-42	489	6.79				
43-47	1039	14.43				
48-52	784	10.89				
53-57	335	4.65				
$\geq$ -58	179	2.49				
Gender						
Male	3,694	51.31			**9.82**	**0.0017**
Female	3.506	48.69				
Marital status					**7946_4**	**0.000**
Single	1986	27.58				
Married	3818	53.03				
Separated	437	6.07				
Divorced	898	12.47				
Widow/widower	61	0.85				
Occupation					**2897.47_5**	**0.0000**
Civil servant	1457	20.24				
Private sector	2192	30.44				
Self-employed	647	8.99				
Students	1075	14.93				
Corper	89	1.24				
Applicants	1740	24.17				
Education					**3107.38_3**	**0.0000**
$\leq$ WASC	259	3.60				
Diploma/OND/NCE	2094	29.08				
Degree/HND	3115	43.26				
Higher Degree	1732	24.06				

Socio-demographic background of respondents, n=7200

Source: Field work (2009)

Table 2.

Frequency	Receiving (%)	Sending (%)	X^2	P
Daily	2169 (30.13)	2011 (29.93)	81.27_4	0.000
Weekly	5723 (79.49)	4819 (66.93)		
Monthly	6249 (86.79)	5315 (73.82)		
Occasionally	6810 (94.58)	6950 (96.53)		
Once in a while	479 (4.79)	516 (7.17)		
Never	-	-	-	-

* p = 0.05 *MMC = Mobile-phone Mediated Communication

Frequency of receiving and sending text based messages through MMC

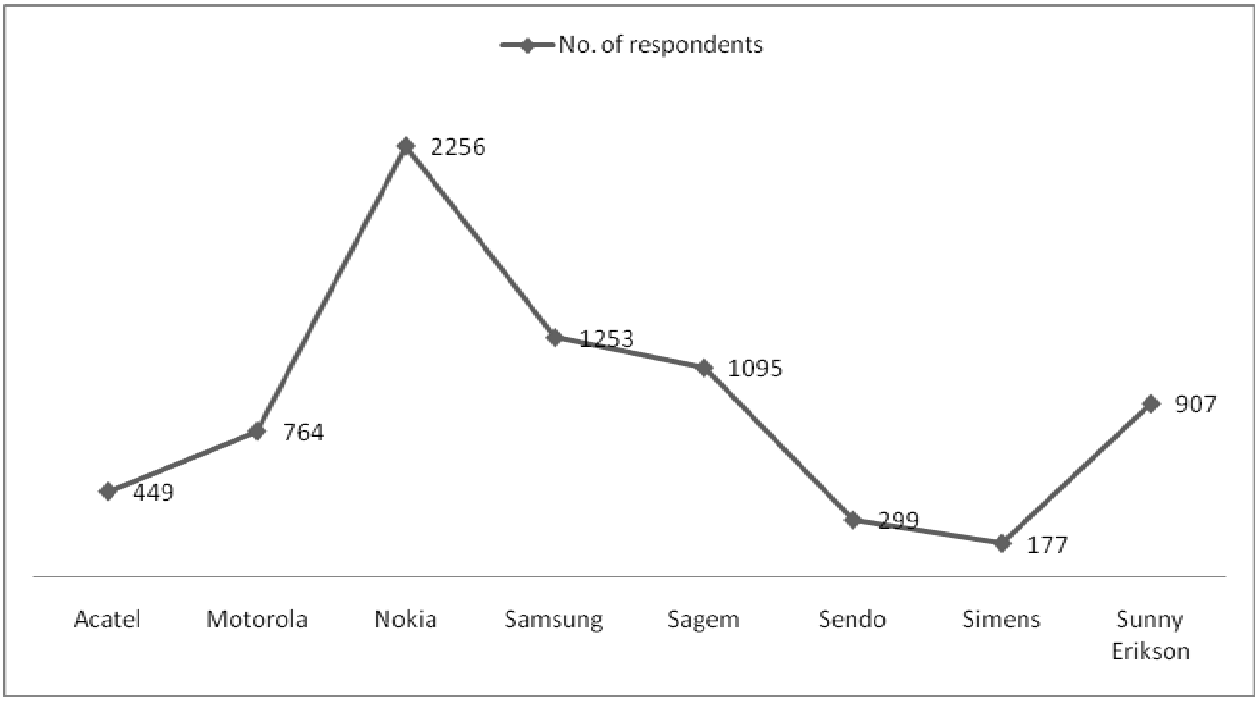
Source: Field work (2009)

Table 3.

	Experiences	N	%	X^2	P
A	**As victims of:**			**8461.70_4**	**0.0000**
	Lies	3,612	50.17		
	Frauds	1,005	13.95		
	Attempts to defraud	4,511	62.65		
	Threats to life	98	1.36		
	False alarm	2,135	29.65		
B	**Being cheated by your:**			**1405.32_5**	**0.0000**
	Husband	1,579	21.93		
	Wife	1,243	17.26		
	Fiancé	715	9.93		
	Fiancée	271	3.76		
	Boy friend	923	12.82		
	Girl friend	605	8.40		
C	**Outcome of A & B**				
	Loss a life saving	67	0.93		
	Live in fear	125	1.74		
	Quarrel	4,036	56.06		
	Suspicion	1,336	18.56		
	Fighting	918	12.75		
	Warning	526	7.31		
	Loss confidence	3,153	43.79		
	Call for divorce	78	1.08		
	Call of separation	107	1.49		

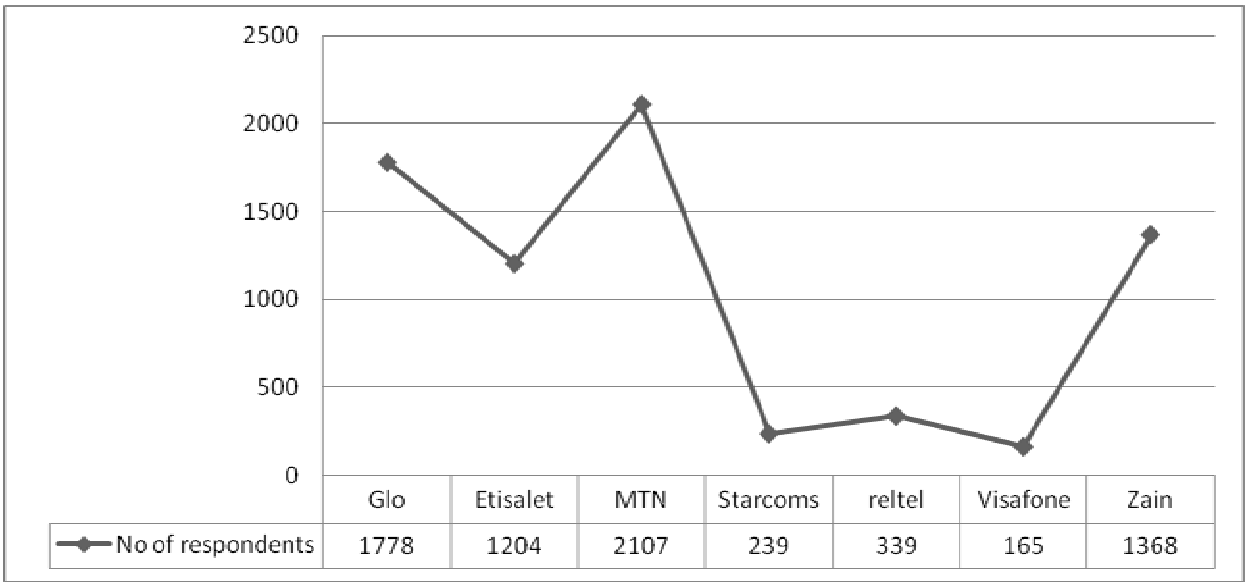
Respondents experiences with text based message through MMC; N= 7200

Source: Field work (2009)



Source: Field work (2009)

Figure 1. Distribution of phone brand by respondents



Source: field work (2009)

Figure 2. Distribution of phone lines by respondents

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Infer the Semantic Orientation of Words by Optimizing Modularity

Weifu Du (Corresponding author)
School of Computer Science, Harbin Institute of Technology
92 Xi Da Zhi Street, Harbin 150001, China
Tel: 86-10-6260-0941 E-mail: duweifu@software.ict.ac.cn

Songbo Tan
Institute of Computing Technology, Chinese Academy of Sciences
6 Ke Xue Yuan Nan Street, Beijing 100190, China
Tel: 86-10-6260-0941 E-mail: tansongbo@software.ict.ac.cn

This work was mainly supported by two funds, i.e., 0704021000 and 60803085, and one another project, i.e., 2004CB318109)

Abstract

This paper proposes a novel algorithm, which attempts to attack the problem of word semantic orientation computing by optimizing the modularity of the word-to-word graph. Experimental results indicate that proposed method has two main advantages: (1) by spectral optimization of modularity, proposed approach displays a higher accuracy than other methods in inferring semantic orientation. For example, it achieves an accuracy of 88.8% on the HowNet-generated test set; (2) by effective usage of the global information, proposed approach is insensitive to the choice of paradigm words. In our experiment, only one pair of paradigm words is needed.

Keywords: Sentiment analysis, Opinion mining, Information retrieval

1. Introduction

In the Web2.0 era, the Internet turns from a static information media into a platform for dynamic information exchanging, on which people can express their views and show their individualities. More and more people are willing to record their feelings (blog), give voice to public affairs (news review), express their likes or dislikes on products (product review), and so on. In the face of the increasing volume of sentimental information available on the Internet, there is a growing interest in helping people to better find, filter, and manage these resources.

Automatic sentiment analysis could play an important role in a wide variety of flexible and dynamic information management tasks. For example, with the help of sentiment analysis system, in the field of public administration, the administrators could receive the feedback on one policy in a timelier manner; in the field of business, manufacturers could perform more targeted updates on products to improve the consumer's experience.

Sentiment analysis can be considered as texts classification according to different opinions they hold. A frequently used method is to label the sentiment words manually (Turney and Littman, 2003). The essential idea is to manually label the semantic orientation of words that in common use, such as "good" labeled as "positive" and "bad" labeled as "negative". When classifying, count directly the numbers of subjective words, then the text was sentenced to positive if it contained more positive orientation words, or negative otherwise. Therefore, it is a fundamental and important task to infer the semantic orientation of words (i.e., for a list of words, partition it into two disjoint sub-lists with semantic orientation: one is positive and another is negative). This paper aims to automatically construct such sub-lists from glosses in a lexicon, as well as from a corpus.

Most of previous methods use word-to-word similarity and some paradigm words (i.e. some representative words with pre-labeled semantic orientation, positive or negative) to infer the semantic orientation of words. The basic observations underlying these methods are quite different from each other. However, these methods could roughly be classified into two categories in terms of the manner of using the word-to-word similarity.

The first kind of approaches uses local information to infer the semantic orientation of words (Turney and Littman,

2003). When computing, only the relationship between the word and paradigm words is taken into account, while the relationship between the words and other words in the test set is ignored, which makes they are sensitive to the choice of the paradigm words.

The second kind of approaches can avoid this drawback by the usage of global information (the relationship of the word with not only paradigm words, but also other words in the test set) (Hatzivassiloglou and McKeown, 1997; Kobayashi et al., 2001). The graph based approaches are the typical ones (Kamps et al., 2004; Hu and Liu, 2004; Andreevskaia and Bergler, 2006; Esuli and Sebastiani, 2006; Pang et al., 2004; Takamura et al.,2005). The essential idea of these approaches is 'minimum cut' that is to look for divisions of the vertices into two subgroups so as to minimize the number of edges running between the subgroups. However, if subgroup sizes are unconstrained then we are, for instance, at liberty to select the trivial division of the network that puts all of the vertices in one of the two subgroups and none in the other, which guarantees we will have zero intergroup edges. This division is, in a sense, optimal, but clearly it does not tell us anything of any worth.

Several approaches have been proposed to get around this problem. For instance, the ratio cut method (Wei and Cheng, 1989) minimizes not the simple cut size but the cut ratio. The ratio cut method does allow some leeway for the sizes of subgroups to vary around their specified values, which makes it more flexible than the simple minimum cut method, but at its core it still suffers from the same drawbacks that they require in advance the sizes of subgroups, which would be determined after the computing.

The study of community finding is the extension of the graph partition and many algorithms had been proposed to find more nature community, such as edges density (Palla et al., 2005), betweeness (Newman and Girvan, 2004), information centrality (Fortunato et al., 2004), random walk (Pons and Latapy, 2005), spectral analysis (Newman, 2006) etc. Modularity optimization based methods (New-man, 2004; Reichardt and Bornholdt, 2006) are the typical ones. Different from conventional graph-partitioning based algorithms' manner of counting edges directly, modularity based algorithm takes the hypothesis that a good division of a network into subgroups is not merely one in which there are few edges between subgroups; it is one in which there are fewer than expected edges between subgroups. This makes the algorithm successful in finding the communities that the sizes of which are unknown in advance.

The efficiency of modularity based algorithm motivates us to take it as the fundamental framework of our method, as far as we know, this approach has not been employed in inferring the term semantic orientation yet.

As a result, we propose an algorithm based on spectral optimization of modularity matrix to infer the semantic orientation of terms. Modularity (Newman, 2004) is a measure to evaluate the goodness of a partition of matrix. Therefore we construct a quality function in the use of modularity (Q value) and then partition the modularity matrix by the manner of making the objective function (Q value) have the maximum value.

2. Proposed algorithm

The proposed word semantic orientation inferring method consists of two steps: (1) the modularity matrix is built to reflect the semantic relationship between words; (2) based on the modularity matrix, an spectral optimization based algorithm is imposed to obtain the semantic orientation of every word.

2.1 Matrix building

Given the word collection $T = \{t_j \mid 1 \le j \le m\}$ of a document, the semantic similarity between any two words t_i and t_j can be computed using approaches that are either knowledge-based or corpus-based.

In this study, we simply choose the mutual information to compute the semantic similarity between word t_i and t_j as follows:

$$sim(t_i, t_j) = \log \frac{N \times p(t_i, t_j)}{p(t_i) \times p(t_j)}$$

which indicates the degree of statistical dependence between t_i and t_j. Here, N is the total number of words in the corpus and $p(t_i)$ and $p(t_j)$ are respectively the probabilities of the occurrences of t_i and t_j, i.e. $count(t_i) / N$ and $count(t_j) / N$, where $count(t_i)$ and $count(t_j)$ are the frequencies of t_i and t_j. $p(t_i, t_j)$ is the probability of the co-occurrence of t_i and t_j within a window with a predefined size k, i.e. $count(t_i, t_j) / N$, where $count(t_i, t_j)$ is the number of the times t_i and t_j co-occur within the window.

We use an adjacency matrix $A=[A_{ij}]_{m \times m}$ to describe the initial word-to-word relationship, where $A_{ij}=sim(t_i, t_j)$, if $i \neq j$ and $A_{ij}=0$ if $i=j$. Then A is normalized to make the sum of each row equal to 1.

Let k_i be the degree of the vertex i and $m = 1/2 \sum_i k_i$, which denotes the total number of the edges in the matrix. Then, we can build the modularity matrix $B=[B_{ij}]_{m \times m}$, where $B_{ij} = A_{ij} - k_i k_j / 2m$.

2.2 Semantic orientation inferring

Based on the modularity matrix B, we take a spectral optimization based algorithm to infer the semantic orientation of words.

In computing, we find the single eigenvector of the modularity matrix B corresponding to the most positive eigenvalue firstly. This is most efficiently achieved by the direct multiplication or power method. Starting with a trial vector, we repeatedly multiply by the modularity matrix and—unless we are unlucky enough to have chosen another eigenvector as our trial vector—the result will converge to the eigenvector of the matrix having the eigenvalue of largest magnitude. In some cases this eigenvalue will be the most positive one, in which case our calculation ends at this point. In other cases the eigenvalue of largest magnitude may be negative. If this happens then, denoting this eigenvalue by β_n, we calculate the shifted matrix $B-\beta_n I$, which has eigenvalues $\beta_i-\beta_n$ (necessarily all nonnegative) and the same eigenvectors as the modularity matrix itself. Then we repeat the power-method calculation for this new matrix and this time the eigenvalue of largest magnitude must be $\beta_1-\beta_n$ and the corresponding eigenvector is the one we are looking for.

When come here, we only get an approximate division, and there is room for improvement of the solution. Then we move single vertices between the subgroups so as to increase the value of the modularity as much as possible, with the constraint that each vertex can be moved only once. Repeating the whole process until no further improvement is obtained, we find a final value of the modularity, and get the semantic orientation of every word in the test set.

3. Datasets and experimental setup

We download texts from the Internet, which including comments on education (from http://blog.sohu.com/learning/), electronics (from http://detail.zol.com.cn/) and stock (from http://blog.sohu.com/stock/). The detail information is illustrated in table 1.

We use ICTCLAS, a Chinese word segmentation software, to extract words (including adjectives, adverbs, nouns, and verbs) from these texts. For each word, if it also occurs in *HowNet*, it is inserted into termset1.

After scanning the words in this set, we find many words either have no sentiment at all or will show distinct orientation in different context, so we ask three people to select words that are considered full of sentiment and as definite as possible by them. Finally, we label each word with the semantic orientation that agreed by the most of people and use the similarity provided by *HowNet* to build the word-to-word matrix. Termset2 and termset3 are constructed in this process. The detail information of individual test set generated by *HowNet* is illustrated in table 2.

There are other three test sets generated by the means of co-occurrence based term similarity computing method. In general, the scale of the corpus and the size of the co-occurrence window will affect the result of term similarity. To test the impact of the variation of corpus's scale, we use the whole corpus to generate termset4; and then followed by decreasing the corpus size by 10%, we generate termset5 and termset6.

For the sparsity of the corpus (nearly almost all the terms occur in one text only one time), we set the co-occurrence window with the range of a whole text. After removing the isolate terms (not co-occur with any other terms at all), we get these three term sets. The detail information of individual test set is illustrated in table 3.

We later compare our approach with Turney and Littman's PMI approach.

4. Experimental results and conclusions

4.1 Performance comparison

The selection of the paradigm words is a pivotal step in the PMI method, and the accuracy of solutions is affected greatly by the choice of the paradigm words. To illustrate this, we ask four people to select some pairs of representative words respectively as the candidate paradigm words. Then put these words into the search engine, google, and sort these words by the related page count returned by the search engine, finally, we take the top 20 pairs of words as the paradigm words.

Table 4 shows the detail information of the paradigm words. Column 1 is the word pair ID, then for positive words and negative words, the word (in English) and related page count (unit is million) are listed respectively.

Then we conduct experiment to observe the fluctuation of solution in the PMI method and the proposed method. As we mentioned, the paradigm words were removed from the testing words for our experiments. The detail information is illustrated in figure 1.

Six curves are plotted in figure 1, one for each of the performance of the two approaches on the three term sets generated by the *HowNet* similarity. The three blue curves are PMI approaches, and the three red curves are our proposed approaches.

Seeing the figure1, we can find that the choice of paradigm words affect the accuracy of the PMI solutions very much, though the rise in accuracy correlates with the increase of the paradigm word set, which is one of the motivations of us

to find a novel method to solve the problem of identifying the semantic orientation to reduce the dependence on the paradigm words. From this figure, we can find the proposed approach is insensitive to the choice of paradigm words.

We evaluate the performance of our method in terms of the comparison with PMI method with 20 pairs of paradigm words which is mentioned in table 4. These comparisons indicate the validity of the proposed method. Table 5 shows the performance comparison on test sets generated by *HowNet*, table 6 shows the performance comparison on test sets generated by co-occurrence.

From table 5, we can find that in different term sets, with the exclusion of noise, the three approaches all have the enhancement in the accuracy. In the three test sets, the proposed approach outperforms the PMI algorithm. Our proposed method outperforms the baseline method in termset2 and termset3 while is exceeded by the baseline method in the termset1. For the reason that the test set2 and test set3 are refined by people from the test set1, we consider that the words in them display more strong semantic orientation, and therefore they show community structure more evidently, which make the proposed method work efficiently.

From table 6, we can find that the proposed method outperform the baseline approach in all term sets generated by co-occurrence. In this experiment, we find the accuracy of proposed method is stable, which indicate that the proposed approach is relatively insensitive to small scale of corpus.

4.2 Discussions

Seen from experiments above, in termset2 and termset3, which are generated by *HowNet* similarity (see in table 2), the proposed method outperforms both the baseline approach and graph partitioning approach. The high performance achieved by our method benefits from the effective utilization of the global information in the term graph.

In the test sets generated by the co-occurrence information of words in corpus (see in table 3), the performance of the three approaches decline sharp, we consider it is because that the co-occurrence relation of words is more the 'relatedness' than the 'similarity'. Incorporating more information, the similarity provided by lexicons is more reasonable than the co-occurrence information of words in relative small corpus.

Because of the consideration of the expected edges within a group, modularity optimization based method can find more nature community, which contributes to the performance of our approach greatly. In the proposed approach, the expected edges were quantified by the probability manner, if it is computed with a more exact manner, the better solution will be attained.

5. Conclusions and future work

Term semantic orientation computing aims to identify the semantic orientation, commendatory or derogatory, of terms; it is the foundation of text sentiment analysis. In this paper, we present a novel modularity optimization based method to identify the term semantic orientation. The proposed approach attains an accuracy of 88.8% on the *HowNet*-generated test set. The experimental results suggest that our algorithm is effective.

Our proposed method has two main advantages: (1) by spectral optimization of modularity, proposed approach displays a higher accuracy in identifying the term semantic orientation. (2) by effective usage of the global information, only one pair of paradigm words is needed in proposed method.

As future work, for modularity computing, we plan to explore in how to find more exact quantity of the expected edges within a group to substitute the current probability manner. Furthermore, we use the co-occurrence information of words in corpus and similarity function provided by *HowNet* in this paper, if more essential relation between words can be detected, our algorithm could be further improved. Thus, in the further work, we will study how to find more essential relation between words.

References

Alina Andreevskaia and Sabine Bergler. (2006). *Mining WordNet For a Fuzzy Sentiment: Sentiment Tag Extraction From WordNet Glosses*. In Proc. EACL-06, Trento, Italy.

Andrea Esuli and Fabrizio Sebastiani. (2005). *Determining the semantic orientation of terms through gloss classification*. In Proceedings of CIKM-05, the ACM SIGIR Conference on Information and Knowledge Management, Bremen, DE, 2005.

Andrea Esuli and Fabrizio Sebastiani. (2006). SentiWordNet: *A Publicly Available Lexical Resource for Opinion Mining*. In Proceedings of LREC 2006, 5th Conference on Language Resources and Evaluation, Genova May 2006.

Andrea Esuli and Fabrizio Sebastiani. (2006). *Determining term subjectivity and term orientation for opinion mining*. In Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, 2006.

Bo Pang and Lillian Lee. (2004). *A sentimental education: sentiment analysis using subjectivity summarization based*

on minimum cuts. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, July 21-26, 2004, Barcelona, Spain.

Fortunato S, Latora V, Marchiori M. (2004). *A method to find community structures based on information centrality*. Phys. Rev. E, 2004, 70: 056104.

H Takamura, T Inui, and M Okumura. (2005). *Extracting Semantic Orientations of Words using Spin Model*. In Proceedings of the 43rd Annual Meeting of the ACL, Ann Arbor. 2005: 133-140.

Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten de Rijke. (2004). *Using WordNet to measure semantic orientation of adjectives*. In Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation, volume IV. 2004: 1115-1118.

Michael Gamon and Anthony Aue. (2005). *Automatic identification of sentiment vocabulary exploiting low association with known sentiment terms*. In Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in NLP. 2005: 57-64.

Newman M E J, Girvan M. (2004). *Finding and evaluating community structure in networks*. Phys. Rev. E, 2004, 69: 026113.

Newman M E J. (2006). *Finding community structure in networks using the eigenvectors of matrices*. Phys. Rev. E, 2006, 74: 036104.

Newman M E J. (2004). *Fast algorithm for detecting community structure in networks*. Phys. Rev. E, 2004, 69: 066133.

Nozomi Kobayashi, Takashi Inui, and Kentaro Inui. (2001). *Dictionary-based acquisition of the lexical knowledge for p/n analysis* (in Japanese). In Proceedings of Japanese Society for Artificial Intelligence,SLUD-33. 2001: 45-50.

Palla G, Derényi I, Farkas I, and T. Vicsek T. (2005). *Uncovering the overlapping community structure of complex networks in nature and society*. Nature. 2005, 435: 814-818.

Peter D. Turney and Littman M.L. (2003). *Measuring Praise and Criticism: Inference of Semantic Orientation from Association*. ACM Transactions on Information Systems, 2003, 21(4): 315-346.

Pons P, Latapy M. (2005). *Computing communities in large networks using random walks*. Proceedings of the 20th International Symposium on Computer and Information Sciences, volume 3733 of Lecture Notes in Computer Science, Springer, New York. 2005: 284-293.

Reichardt J, Bornholdt S. (2006). *Statistical mechanics of community detection*. Phys. Rev. E, 2006, 74: 016110.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. (1997). *Predicting the semantic orientation of adjectives*. In Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and the Eighth Conference of the European Chapter of the Association for Computational Linguistics. 1997: 174-181.

Y.-C. Wei and C.-K. Cheng. (1989). *Toward efficient hierarchical designs by ratio cut partitioning*. In Proceedings of the IEEE International Conference on Computer Aided Design, pp. 298–301, Institute of Electrical and Electronics Engineers, New York (1989).

Table 1. The detail information of the text sets

TextSet ID	Positive	Negative	Total
Education	254	1012	1266
Stock	364	683	1047
Electronics	1054	554	1608

Table 2. The detail information of the term sets generated by *HowNet*

TermSet ID	Positive	Negative	Total
1	2365	2923	5288
2	1881	2481	4362
3	1098	2039	3137

Table 3. The detail information of the term sets

TermSet ID	Percent of full corpus	Positive	Negative	Total
4	100%	512	664	1176
5	90%	498	647	1145
6	80%	474	641	1115

Table 4. Paradigm words used in the PMI method

Positive		Negative	
Word	Freq (mil.)	Word	Freq (mil.)
Good	2,400	Mistake	214
Active	220	Badness	190
Excellent	219	Agony	96.4
Beautiful	203	Depressed	68.8
Proficient	142	Conservative	44
Mature	127	Worry	43.1
Nice	114	Falsity	41.7
Harmonious	113	Lousy	37.5
good luck	79.7	Terrific	36.5
Peace	78.2	Collapse	34.5
Energy	77.2	Maze	29.4
Comfortable	69.6	Shortcoming	26.2
Fineness	53.7	Misery	24.4
Grateful	50.5	Contort	18.1
Summit	48.6	Phony	15.1
Goodness	43.6	Freaky	12.8
Honest	27	oafish	10.5
Allowance	23.8	Sad	10.4
Glary	23.4	Shame	9.99
Decency	21.5	Asperity	8.50

Table 5. the performance of the two methods on test sets generated by *HowNet*

Approach	Testset ID	Accuracy	Average Accuracy
PMI	1	0.642	0.726
	2	0.694	
	3	0.843	
Proposed Approach	1	0.617	0.741
	2	0.718	
	3	0.888	

Table 6. the performance of the two methods on test sets generated by co-occurrence

Approach	Testset ID	Accuracy	Average Accuracy
PMI	4	0.457	0.453
	5	0.456	
	6	0.446	
Proposed Approach	4	0.618	0.604
	5	0.598	
	6	0.596	

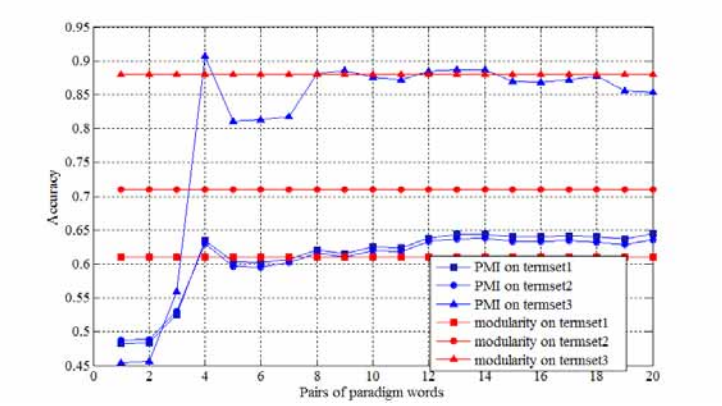


Figure 1. the accuracy of the PMI method and the proposed method on different term sets with the variety in the selection of paradigm words

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Distributed University Registration Database System Using Oracle 9i

Sana J. Alyaseri
Department of Computer Engineering
College of Engineering
University of Basrah
E-mail: sana_jabbar@yahoo.com

Abstract

In database and information systems, increasing attention has been focused lately on parallel and distributed database systems. The future of large database systems lies into the realm of distributed computing. The main reason for this is that distributed computing can be constructed at a low cost without the need for any specialized technology, using existing sequential computers and relatively cheap computer networks. The basic motivations for distributed databases are improved performance, increased availability, share ability, expandability, and access flexibility. Oracle9i is one of the most famous database management systems that provide distributed features. Oracle9i supports transparent distributed gains to access data from multiple databases. It also provides many other distributed features, such as transparent distributed transactions and transparent automatic two-phase commit. This paper presents a distributed registration database system using Oracle9i distributed features such as replication and fragmentation. The analysis and design steps of the distributed registration database system are shown as well as to the practical steps required for implementing the distributed database system using Oracle9i.

Keywords: Distributed database, Oracle9i distributed database, Replication, Fragmentation

1. Introduction

Distributed computing is one of the most recent and important development in the computing era. The next decade of research in database era will focus toward delivering widely available access to unprecedented amounts of constantly expanding data that is distributed all over the net (Loose & Church, 2004; Abiteboul et al., 2003; Ohta & Ishikawa, 2003). Users will benefit from new machine learning technologies that mine new knowledge by integrating an analyzing huge amounts of widely distributed data to uncover and report upon subtle relationships and patterns of events that are not immediately recognized by direct human inspection (Aljanaby et al., 2005; Kossman, 2000).

Distributed Database (DDB) technology emerged as merger of two technologies; database technology and data communication technology. The maturation of Database Management System (DBMS) technology has coincided with significant development in distributed computing and parallel processing technology. The end result is the emergence of Distributed Database Management System (DDBMS) and parallel database management system. These systems have started to become the dominant data management tools for highly accessed data. (Ozsu, 2007) A distributed database is a collection of multiple logically interconnected databases distributed over a computer network, or a single logical database that is spread physically across computers in multiple locations that are connected by a data communications link and a DDBMS as a software system that managed a distributed database, while making the distribution transparent to the users. There are two main types of distributed databases. The first type is homogeneous database that has same DBMS at each node and the other type is heterogeneous database that has different DBMSs at different nodes (Ozsu, 2007).

Reliability and availability are the most common potential advantages cited for a distributed database. Reliability is defined as the possibility that a system is running at a certain time point. Availability is the probability that the system is continuously available during a time interval. When the data and DBMS software are distributed over several sites, one site may fail while other sites continue to operate. This improves both reliability and availability (Loose & Church, 2004).

Oracle introduced inter-database connectivity with SQL*Net in Oracle Version 5 and simplified its usage considerably with the database links feature in Oracle Version 6, opening up a world of distributed possibilities. Oracle9i now supplies a variety of techniques that can be used to establish inter-database connectivity and data sharing. It provides a

robust and complete solution that addresses all the needs that may arise when operating in a distributed environment. It includes communication between users on the Oracle database using queues, data replication, and distributed data access in both homogenous and heterogeneous environments (Oracle Corporation, 2002a; Oracle Corporation, 2002c). Oracle9i even moved beyond that and was built using the Grid structure. This will dramatically lower the cost of computing, extend the availability of computing resources, deliver higher productivity, and higher quality (Oracle Corporation, 2002b).

The remainder of this paper is organized as follows. In section 2, some of the concepts and terminology related to the distributed features of Oracle9i are introduced. The analysis and design of the distributed registration database system is presented in section 3. Development and implementation is presented in section 4. Finally, section 5 concludes the paper.

2. Oracle9i Distributed Features

There is a great deal of confusion surrounding the various products and terminology from Oracle. It's worthwhile to clarify some of these terms up front so the reader gets the most benefit from this paper. This section introduces Oracle9i terminology and distributed features related to the work presented in this paper.

2.1 Distributed Terminology

The terms database and database instance are often used interchangeably, but they are not the same. In Oracle parlance, a *database* is the set of physical files containing data. These files comprise table spaces, redo logs, and control files. A *database instance*, or simply *instance*, is the set of processes and memory structures that manipulate a database. A database may be accessed by one or more database instances, and a database instance may access exactly one database (Oracle Corporation, 2002c). *Oracle Parallel Server (OPS)* is a technology that allows two or more database instances, generally on different machines, to open and manipulate one database. In other words, the physical data files in a database can be seen, inserted, updated, and deleted by users logging on to two or more different instances; the instances run on different machines but access the same physical database (Oracle Corporation, 2002a)

The Parallel Query Option (PQO) is a technology that can divide complicated or long-running queries into several independent queries and allocate separate processes to execute the smaller queries. A coordinator process collects the results of the smaller queries and constructs the final result set. Parallel queries are effective only on machines that have multiple CPUs. Oracle also introduced the parallel DML feature in Oracle9i. Parallel DML is similar to parallel query, except that the independent processes perform DML (Baumgartel, 2002). For example, an update of several hundred thousand rows can be doled out to several processes that execute the update on separate ranges of the table.

Figure 1 illustrates an environment in which data in two or more database instances is accessible as though this data were in a single instance. This access may be read-only, or it may permit updates to one or many instances. The referenced data may be real time, or it may be seconds, hours, or days old. Generally, the different database instances are housed on different server nodes, and communication between them is via Net8 (for Oracle9i). In addition to database servers, a distributed database system usually includes application servers and clients. Application servers, like database servers, typically are high-capacity machines that run intensive utilities such as web applications, Oracle's application cartridges, report generators, and so forth (Oracle Corporation, 2002a).

All Oracle databases in a distributed database system use Oracle's networking software, Net8, to facilitate inter-database communication across a network. Just as Net8 connects clients and servers that operate on different computers of a network, it also allows database servers to communicate across networks to support remote and distributed transactions in a distributed database. Net8 makes transparent the connectivity that is necessary to transmit SQL requests and receive data for applications that use the system. Net8 performs all processing independent of the underlying network operating system (Oracle Corporation, 2002c).

Optionally, an Oracle network can use Oracle names to provide the system with a global directory service (Oracle Corporation, 2002a). When an Oracle network supports a distributed database system, you can use Oracle names servers as central repositories of information about each database in the system to ease the configuration of distributed database access. Database links are the invisible glue that makes location transparency possible (Oracle Corporation, 2002a). In more technical terms, a database link defines a connection from one database instance to another, and this definition is stored in the Oracle data dictionary. Since database link connections log in to a normal account in the remote database instance, this provides a complete control over its privileges and quotas.

2.2 Transparency

Users of a distributed database system need not to be aware of the location and functioning of the parts of the database with which they work. The goal of transparency is to make a distributed database system appear as though it is a single Oracle database (Oracle Corporation, 2002c). The DBA and network administrators can ensure that the distributed nature of the database remains transparent to users. An Oracle distributed database system has features that allow

application developers and administrators to hide the physical location of database objects from application and users. Location transparency exists when a user can universally refer to a database objects such as a table. Local views can provide location transparency for local and remote tables in a distributed database system. For example, assume that table EMP is stored in a local database. Another table, DEPT, is stored in a remote database. To make the location of, and relationship between, these tables transparent to users of the system, a view named company can be created in the local database that joins the data of the local and the remote servers.

```
CREATE VIEW company AS
SELECT empno, ename, dname
FROM emp a, dept@Basrah_link b
WHERE a.deptno = b.deptno;
```

When users access this view, they do not know, or need to know, where the data is physically stored, or if data from more than one table is being accessed. Thus, it is easier for them to get required information. The following example provides data from both the local and remote database table.

```
SELECT * FROM company;
```

Synonyms are very useful in both distributed and non-distributed environments because they hide the identity of the underlying object, including its location in a distributed database system. If the underlying object must be renamed or be moved, only the synonym needs to be redefined; applications based on the synonym continue to function without modification. A synonym can be created for any table, type, view, snapshot, sequence, procedure, function, or package. All synonyms are stored in the data dictionary of the database in which they are created. To simplify remote table access through database links, a synonym can allow single-word access to remote data, isolating the specific object name and the location from users of the synonym (Oracle Corporation, 2002a).

The syntax to create a synonym is:

```
CREATE [PUBLIC] synonym_name
FOR [schema.] object_name [@database_link_name]
```

Oracle allows the following standard DML statements to reference remote tables:

```
SELECT (queries), INSERT, UPDATE, DELETE
SELECT... FOR UPDATE
LOCK TABLE
```

Queries including joins, aggregates, sub queries, and SELECT ... FOR UPDATE can reference any number of local and remote tables and views.

2.3 Replication

Oracle offer two modes of replication; the choice of mode will depend on the reason for replication deployment. In some cases, a hybrid is appropriate. Multi-master replication (also called peer-to-peer or n-way replication) enables multiple sites, acting as equal peers, to manage groups of replicated database objects. Each site in a multi-master replication environment is a master site, and each site communicates with the other master sites. Asynchronous replication is the most common way to implement multi-master replication. Other ways include synchronous replication and procedural replication. Using asynchronous replication, information about a DML change on a table is stored in the deferred transactions queue at the master site, where the change occurred. These changes are called deferred transactions. The deferred transactions are pushed (or propagated) to the other participating master sites at regular intervals. The amount of time in an interval can be controlled. Using asynchronous replication means that data conflicts are possible, because the same row value might be updated at two different master sites at nearly the same time. However, some techniques can be used to avoid conflicts and, if conflicts occur, Oracle provides rebuilt mechanisms that can be implemented to resolve them. Information about unresolved conflicts is stored in an error log (Baumgartel, 2002; Pratt, 2001).

A materialized view contains a complete or partial copy of a target master from a single point in time. The target master can be either a master table at a master site or a master materialized view at a materialized view site. A master materialized view is a materialized view that functions as a master for another materialized view. A multi-master materialized view is one that is based on another materialized view, instead of a master table. Materialized views provide the following benefits:

- Enable local access, which improves response time and availability.
- Offload queries from the master site, because users can query the local materialized view instead.

- Increase data security by allowing you to replicate only a selected subset of the target master's data set.

A materialized view may be read-only, updatable, or writeable, and these types of materialized views provide benefits in addition to those listed previously.

2.4 Database Link

A database link is a pointer that defines a one-way communication path from an Oracle database server to another database server. The link pointer is actually defined as an entry in a data dictionary table. To access the link, you must be connected to the local database that contains the data dictionary entry. A database link connection allows local users to access data on a remote database. The great advantage of database links is that they allow users to access another user's objects in a remote database so that they are bounded by the privilege set of the object's owner.

3. Design of the Distributed Registration System

In this section, the main steps for analyzing and designing distributed registration database system will be described. The distributed registration system is for a university has two campuses, the first campus in Basrah governorate and the other in Mesan governorate. The central campus will be in Basrah campus and it is responsible for defining all colleges and departments that are allowed to be added in both campuses. The instructor can be added from any site. The central campus is responsible for producing courses for both sites. Each site is responsible for producing its schedule. Student can be added or registered from any site regardless of his study location. The tables of colleges, departments, instructors and courses are the same as in both sites, that's why the tables are replicated in both sites. The information about the students in each campus is stored in its site, so the student table; teach table (schedule), and the registration table are fragmented. Table 1 shows the distributed technique used for each table and the corresponding technique in Oracle9i (Fodor, 2007; Hoffer et al., 2008). The entity relationship diagram for the distributed registration system is shown in Figure 2. The architecture of Oracle distributed registration database system is shown in Figure 3.

4. Development and Implementation

In this section, the implementation steps of the distributed registration database system using Oracle9i will be described. Oracle9i DBMS was installed with its global name in both sites, Basrah_db and Mesan_db. After setting a connection between the two sites, the parameters that must be set or added to the initSID.ora file are initialized as in Table 2. The recommended initial table space Requirements are shown in Table 3.

To setup Net8, an Oracle Net listener must be running on each of the servers involved in the replication environment and the listener.ora file must contain an entry for the instance. Net8 is then set on each site using Net Configuration Assistant tool by filling the service name and the IP address. For a successful transaction process, a database link is created for each site (Basrah link and Mesan link). Before setting up database links, it is important that all replicated databases have unique global name.

The next step is creating all tables for Basrah, the centre site. Using the Enterprise Manager Console in Oracle9i, Mesan_db is added from Basrah site, and Basrah_db from Mesan site. After all the previous steps have been completed at all sites to be involved in a master replication, the OEM replication management setup wizard is used to complete the configuration. The Enterprise Manager Console is used to setup the master sites. Through the master sites setup, the master sites, Basrah_db and Mesan_db are defined. The default user admin (repadmin) is also defined during this process. This user is responsible of the multi-master replication process and he is allowed to be the propagator and receiver for transaction. Then, a default link scheduling (refresh interval) is set. To create a master group, it must to login to the Enterprise Manager Console as a repadmin. Then, setting the refresh type which is synchronous and specified the database link on which the multi-master replication was made on. The multi-master tables is then specified in the object of the created master group and generate replication supports on multi master tables to other sites. During that actually two process occurred

1) Setup the deferred queue to be pushed automatically to run the fallowing at each master site.

```
Connect repadmin/<password>
Begin
 Dbms_defer_sys.schedule_push(
    Destination      =>  '<destination databases global name>.world',
    Interval         =>  '/10:mins*/sysdate+10/(60*24)',
    Next_date        => sysdate,
    Stop_on error    => false,
    Delay_seconds  => 0,
    Parallelism      => 1);
```

```
End;
```

2) Setup the deferred queue to be purged automatically to run the fallowing at each master site.

```
Connect repadmin/<password>
Begin
  Dbms_defer_sys.schedule_purge(
    Next_date          => sysdate,
    Interval           =>  '/1:hr*/sysdate*1/24',
    Delay_seconds      => 0,
    Rollback_segment => ' ');
      End;
```

Completing the above steps will reach the finishing of multi-master replication. Through the same tool, that is the Enterprise Manager Console, the materialized view sites are then setuped. Through the setup process the materialized view master site-Basrah and materialized view site (Mesan) is defined. A schema to be created in the materialized view sites is selected which is a registrar schema. The default user admin (MVadmin) which is responsible of the materialized view replication and he is the propagator and receiver for transactions. The next process is to specify the interval of refresh for each materialized view and its name and query, the fast refresh type for some materialized view and the complete refresh for the others were then selected. There were some problems in the materialized view refresh, so some of the materialized views were manually created as shown below:

Example about fast refresh:

```
create materialized view log on Department
create materialized view Department
    refresh fast
    next sysdate+1/1440
    as
    select * from Department @Basrah_db
```

Example about complete refresh:

```
create materialized view Course
    refresh complete
    next sysdate+1/1440
    as
  select * from Course @Basrah_db
```

After determining the fragmented tables, the same tables into the two sites (Basrah_db and Mesan_db) are created. In each site a remote synonym for each fragment table has been created. The insert, delete and update for DML commands are done as shown in the following algorithm:

5. Conclusion

The future of huge databases relies on the network based parallelism. Distributed and parallel databases are becoming one of the hot topics in the last decade. The need for more distribution of data is increasing as we rely more on the net. The internet and the World Wide Web significant developments are encouraging DBMS vendors on making their products web-enabled. This is leading to distributing data in many servers over the net.

Distributed database management systems are commercially available. Oracle is one of the leading database corporations who moved early toward distribution. Oracle9i is one of the most common database management systems that supports distributed features. It supports transparent data access across multiple databases. Also, it provides many other distributed features, such as transparent distributed transactions and transparent automatic two-phase commit. This paper presented a distributed database registration information system using Oracle9i distributed features such as replication and fragmentation. In addition to that, the paper shows the practical steps in implementing a distributed database using Oracle9i in a local area network.

References

Abiteboul, S., Bonifati, A., Cobena, G., Manolesco, I., & Milo, T. (2003). Dynamic XML documents with distribution and replication. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data* (pp. 527

– 538), San Diego, California, USA.

Aljanaby, A., Abuelrub, E., & Odeh, M., (2005). A survey of parallel and distributed query optimization". *International Arab Journal of Information Technology (IAJIT)*, 2(1), 48-57.

Baumgartel, P. (2002). *Oracle Replication: An Introduction*. New York, USA: Addison Wesley.

Fodor, A.. (2007). Aspects of data allocation in distributed database systems. In *Proceedings of the 7th European Conference of Young Research and Science Workers (TRANSCOM 2007)*, pp. 51-53.

Hoffer, J., Gearge J., & Valacich J. (2008). *Modern Systems Analysis and Design* (5th ed.). New Jersey, USA: Prentice Hall.

Kossman, D. (2000). The state of the art in distributed query processing. *ACM Computing Surveys*, 32(4), 422-469.

Loose, R., & Church, L. (2004). Information retrieval with distributed databases: Analytic models of performance. *IEEE Transactions on Parallel and Distributed Systems*, 15(1), 18-27.

Ohta, M., & Ishikawa, H. (2003). *An Active Web-based Distributed Database System for E-Commerce*. Technical Report, Department of Electronics and Information Engineering, Tokyo Metropolitan University, Japan.

Oracle Corporation. (2002a). *Oracle9i Database Administrator's Guide Release 2 (9.2).* Oracle Corporation: California, USA.

Oracle Corporation. (2002b). *Oracle and the Grid (an Oracle White Paper).* Oracle Corporation, World Headquarters, California, USA, November 2002.

Oracle Corporation. (2002c). *Oracle's Solutions for the Distributed Environment* (an Oracle White Paper). Oracle Corporation, World Headquarters, California, USA, June 2002.

Ozsu, T. (2007). *Principles of Distributed Database Systems* (3rd ed.). New Jersey, USA: Prentice Hall.

Pratt, M. (2001). *Oracle9i Replication*. California, USA: Oracle Corporation.

Table 1. Distributed database techniques used for different tables

Table Name	Distributed Database Technique	Oracle9i Distributed Database Technique
Location	Replication Technique	Materialized View (Readable)
College	Replication Technique	Materialized View (Readable)
College_Location_Department Intersection	Fully Replication Technique	Multi-master Replication
Department	Replication Technique	Materialized View (Readable)
Student	Horizontal Fragmentation	Manually (because there is no Oracle tool that supports this technique)
Instructor	Fully Replication	Multi-master Replication
Course	Replication Technique	Materialized View (Readable)
Student_Course (Teach) Intersection	Horizontal Fragmentation	Manually (because there is no Oracle tool that supports this technique)
Instructor_Course (Registration) Intersection	Horizontal Fragmentation	Manually (because there is no Oracle tool that supports this technique)

Table 2. Parameters initialization values

Parameter Name	Initial Value
COMPATIBLE	9.0.1.0.0(minimum 9.X) OSS recommended this equate to the server release
SHARED _POOL_SIZE	30M (for basic)
PROCESSES	Added 12 to the current value
GLOBAL_NAMES	TRUE
DB_DOMAIN	extension component of the local Database global name
OPEN_LINKS	4 was added
DISTRIBUTED TRANSACTION	5 was added
REPLICATION_DEPENDENCY_TRACKING	TRUE
JOP_QUEUE_INTEREVAL	10 seconds
JOP_QUEUE_PROCESSES	3 was added
PARALLEL_MAX_SERVERS	10
PARALLEL_MIN_SERVERS	2

Table 3. Recommended initial table-space Requirements

Table-space	Initial Value
SYSTEM	At least 80 Mb free
ROLLBACK SEGMENT	At least 60 Mb free
TEMPORORY	At least 40 Mb free

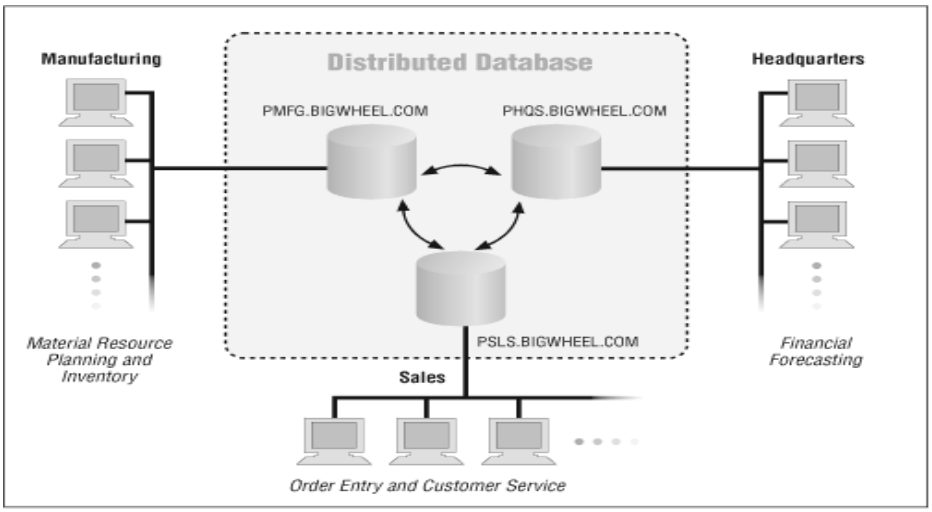


Figure 1. Distributed Database System

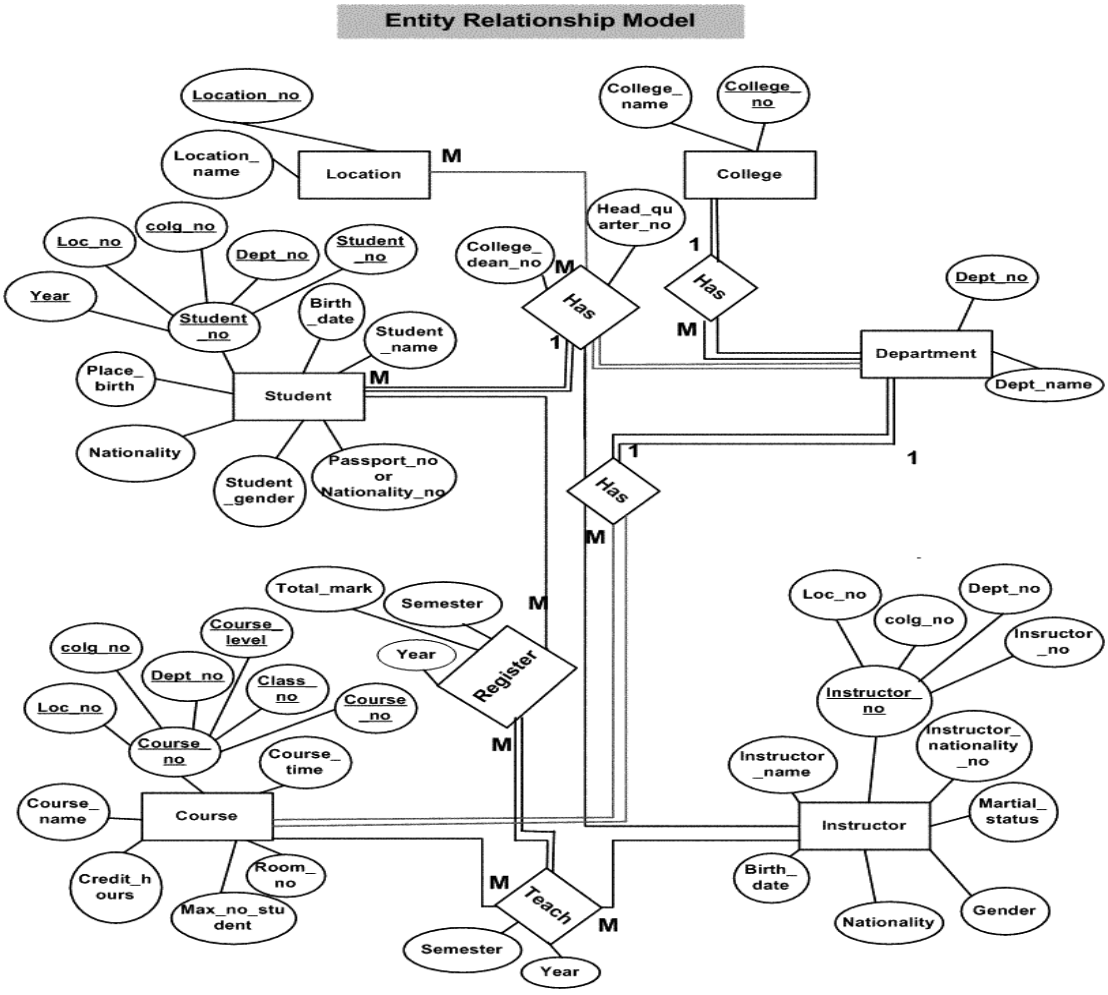


Figure 2. Entity Relationship Model

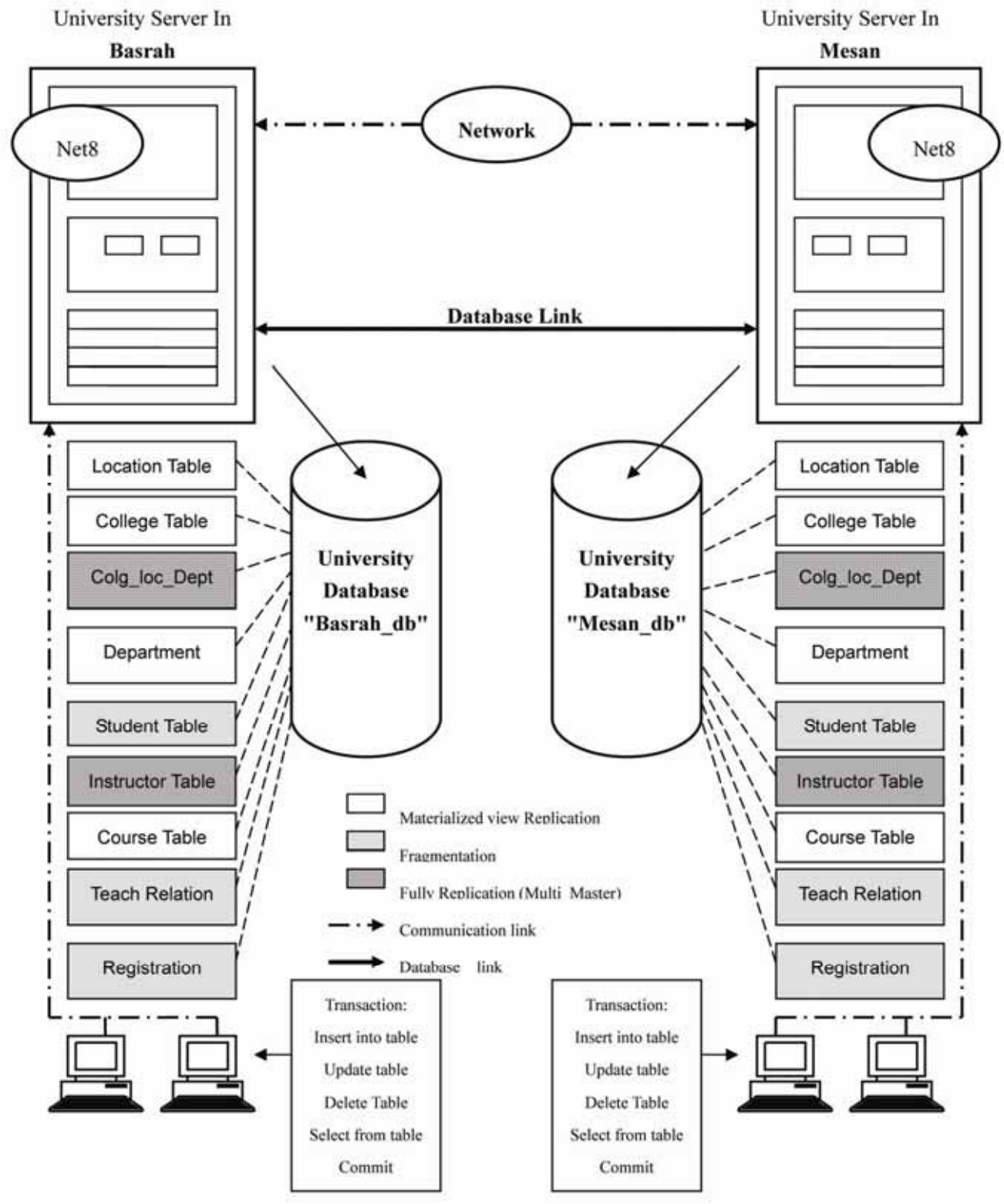


Figure 3. Oracle distributed registration database system

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Developments of the Research of the Formant Tracking Algorithm

Gang Lv & Heming Zhao
School of Electronic Information, Soochow University, Suzhou 215006, China
Tel: 86-512-6219-0898 E-mail: lvgang@suda.edu.cn

The research is financed by the National Natural Science Foundations of China (No. 60572076) and the University Natural Science Research Project of Jiangsu Province of China (No. 05KJB510113). (Sponsoring information)

Abstract

The formant is the important part of the phonetic characters, and reliable formant tracking algorithm is the base to study the phonetics. Based on the development course of the phonetic formant tracking algorithm, the linear prediction coding (LPC) and the model matching method are introduced emphatically, and there own advantages and disadvantages are analyzed, and the model matching method based on the hidden dynamic model will be the development direction of the future formant tracking technology.

Keywords: Formant, Tracking, Linear prediction coding, Model matching

1. Introduction

When pronouncing, the air currents pass the vocal tract, which will induce the resonance of the channel, and generate a group of resonance frequency that is called as the formant frequency, i.e. the formant for short. The formant is the important parameter to differentiate different vowels The algorithm to position and mark the tracks of the change of the formant frequency with the time is called as the formant tracking algorithm. The formant tracking is the reflection of speaker's individual character. The acquirement of the formant parameter and the tracking algorithm have been widely used in the speaker recognition, the speech synthesis and the speech coding transfer, and they are the important research topics in the speech signal processing domain.

Based on the important meaning and the wide application foreground of the formant to the speech signal processing, many scholars have applied themselves to the study of the acquirement of the formant parameter and the tracking algorithm in recent years, and new algorithms are continually pushed. By the research and analysis of literatures, these algorithms can be classified as two sorts, i.e. the LPC method and the model matching method. These two methods have their own advantages and disadvantages. The separation of the linear prediction equation can exactly confirm the central frequency and bandwidth of the formant, but the ascriptions of pecks are difficult to be judged in LPC, which can be avoided by the model matching method, but the model needs to be trained by large numbers of data, and the training result depends on the quantity and kind of the training data. These two methods will be briefly introduced and analyzed as follows.

2. Linear prediction coding method

S S McCandless first used the LPC method to acquire three front formant frequencies and extents in the phonetic fragment of vowel (S.S.McCandless, 1974, p.135-141). By the estimated characters of LPC spectrum, in the region that the energy of signals is strong, i.e. the region closing to the peak value of the spectrum, the LPC spectrum is closing to the signal spectrum, but in the region that the energy of signals is weak, i.e. the region closing to the vale of the spectrum, both spectrums are significantly different. So to check the peak values of the LPC spectrum can confirm the formant. In ideal situation, three front formants of the speech are three front formants of the LPC spectrum.

D Broad et al improved the LPC analysis algorithm, and the new algorithm adopted the cepstral spectrum coefficient of LPC to acquire the parameters of formant (D.J.Broad, 1989, p.2013-2017). Comparing with S S McCandless' LPC spectrum estimation algorithm, the robustness of the improved algorithm was better when acquiring the formant of the fragment of vowel.

A.M.De Lima Araujo et al combined the Mel frequency scale according with human ear's hearing character with the LPC analysis to estimate the first formant F1 and the second formant F2 of speech signals (A.M.De Lima Araujo, 1998, p.207-212). Traditional LPC algorithm needed to confirm the orders of the linear predictor according to the amount of the acquired formant, but this algorithm could acquire F1 and F2 by setting up a fixed order of the predictor, and needed

not to change the orders of the linear predictor by changing the amount of the acquired formant.

D. Talkin put forward an automatic tracking method of phonetic formant track. This method adopted the dynamic programming method to realize the tracking of the formant by introducing the continual limited conditions of frequency (D.Talkin, 1987, p.S55). First, acquire the candidate values of the formant frequency by seeking the roots of the linear predicted equation, then, establish a stationarity function as the limited condition of the frequency continuity, finally, by an improved Viterbi algorithm, compute the minimum value between the mapping value of the formant frequency in the current frame and the mapping value of the formant frequency in the last frame in the limitation of stationarity function to realize the track connection among frames of the formant. The key of this algorithm is to design a proper frequency continual limited stationarity function. However, the experiments showed that it was very difficult to design reasonable stationarity function (G..Kopec, 1986, 709-729).

In conclusion, because the orders of the coefficient predicted by the LPC method are confirmed beforehand, so the amount of the acquired complex conjugate peak pair is the half of the orders at most. Generally, the bandwidth of the extra peak is bigger than the bandwidth of the formant, so to acquire the formant means to judge the ascription of the peak. To compute simply, for the standard vowel signals, the LPC method can exactly confirm the central frequency of the formant and the bandwidth by separating the linear predicted equation. But if the voice signals are interfered by the noise source, the fake peak and the combined peak will occur on the frequency spectrum, which will bring large difficulty to judge whether formant the peak points belong to, and influence the tracing nicety of the formant track, that is the essential deficiency of the LPC method in the formant tracking analysis.

3. Model matching method

The model matching method avoids the problem that the LPC method is easy interfered by the fake peak and the combined peak, and it is the research hot in the acquirement of formant parameters and the tracking study in recent years. The model matching method experienced a development process from the HMM model to the HDM model.

In 1975, Baker put forward the idea to adopting the hidden Markov model (HMM) to trace the formant track (J.K.Baker, 1975, p.24-29), but the experiment failed. Ten years ago, G. Kopec first successfully used the formant tracking method based on HMM (G.Kopec, 1985, p.1113-1116). He divided the formant tracing problem into two independent problems including checking and estimation, and the formant checking was to judge whether each frame speech signal had the formant or not, and the formant estimation is to endow certain frequency value for the checked formant. The checking and estimation of formant all adopted the Viterbi algorithm to search the optimal status sequence of HMM. This algorithm adopted the statistical method to realize the tracking of formant, but it could only realize the tracking of one formant.

After that, G. Kopc put forward the improved formant tracking algorithm based on HMM and the vector quantization (VQ) technology (G.Kopec, 1986, p.709-729). Comparing with the method in the literature (G.Kopec, 1985, p.1113-1116), two aspects were improved in the new algorithm. First, by set up two tracing modes, i.e. the single formant tracing mode and the multi-formant tracing mode, multiple formant tracks could be traced simultaneously, and the problem which could trace only one formant track was solved. Second, adopt the forward-backward algorithm (F-B algorithm) to replace the Viterbi algorithm to check and estimate the formant, because the Viterbi algorithm could generate one single status sequence, not a probability distribution, which would produce two problems. The first one is the problem that the formant checking is difficult, and the formant checking based on the Viterbi algorithm is to control the probability of the error checking and the peak value omission by setting up the thresholds, and it can not directly adjust the thresholds in the tracing process, so it is not flexible enough. And if the checking and the estimation of formant are implementing simultaneously, the formant checking performance will depend on the quantity which is used to denote the status of the formant parameters. With the increase of the status density of the frequency field space, the probability that the single status is checked will decrease. Therefore, with the increase of the status quantity, the probability that the real status is checked will gradually reduce. The HMM formant tracking algorithm based on F-B algorithm will avoid this problem. The second problem is that the formant track traced by the Viterbi algorithm is a group of discrete frequency values defined by the model status, but the formant track obtained by the F-B algorithm is the weighted average value in the status of discrete model, so the track obtained by the latter will be more smooth than the track obtained by the former.

P. Zolfaghari (P.Zolfaghari, 1996, p.1229-1232) and J. Darch (J.Darch, 2005, p.941-944) put forward the phonetic formant tracking method based on the Gaussian mixture model (GMM). Because GMM is a continually distributed HMM which status is 1, this method still can be regarded as a formant tracking method based on the HMM model.

There are two deficiencies to adopt the HMM model to solve the formant parameter tracking problem. First, when the algorithm is used to estimate the status of certain time of the formant track, it only takes the continuity of the track as the restriction condition to select the formant, which will easily induce the tracing error. Aiming at this deficiency, Minkyu Lee et al put forward an improved method (M.Lee, 2005, p.741-750). By estimating the status of certain time

of the formant track, they combined the phoneme information based on the speech signal text with the continuity of the track to be the restriction condition of selecting the formant, which could enhance the precision of the tracking. D. T. Toledaro et al also put forward similar improved method (D.T. Toledano, 2006, p.511-522). These methods could significantly enhance the tracing precision when they were used to trace the formant track of special people's sound which is related with the text, but for the tracking of non-special people's sound which is not related with the text, the had not obviously improved effects. Second, the algorithm needed large numbers of data to train the model, and the result of the final tracing was decided by the kind and the quantity of the training data. But in different using environments, whether the training data have sufficient representative quality or not could not be confirmed. This deficiency is instinctive for HMM. HMM is general statistical model which is widely used in many different domains. If it is applied in some special domain, special data will be needed to train the model. That is to say, the HMM is a data-driven model, and it doesn't involve any mechanism about the generation of data. So in the actual environment with noises and interferences, the formant tracking method based on HMM can not fulfill the actual requirements. To overcome this deficiency, the new established model should not only consider the observation data but also the pronunciation mechanism of the speech signals to describe the speech signals.

In the late of 1990s, L. Deng put forward a dynamic speech modeling method combining the metrics characters and the phonetic characters (L.Deng, 1998, p.299-323). This method considered the conversion of the co-articulation and the neighboring phones in the pronunciation for the modeling process of the speech signals. It regards the pronunciation system of sound as a hidden dynamic system, and in which, each phone corresponds with one vector objective, i.e. when certain phone is pronounced, the muscles of the vocal cords and the track will approach certain objective status or shape according to the "program". This modeling method which is specially used for sound considers the generation mechanism of the speech signals, and gets rid of the modeling mechanism which only is driven by the data. Almost in the same term, Richards et al put forward similar speech modeling method which was named by the hidden dynamic model (HDM) (H.B.Richards, 1999, p.357-360). To describe the dynamic structure of the sound, Richards mapped the hidden space on the phonetic character space by the nonlinear multiplayer perceptron (MLP), and trained the parameters of the model (target vector and the weight value of MLP) by the selected algorithm.

HDM was successfully applied in the speech recognition (L.Deng, 1998, p.299-323 & L.Deng, 2000, p.3036-3048), and many new phonetic formant tracking algorithms were developed based on that. For example, I. Bazzi put forward a formant tracking algorithm based on the expectation maximization (EM) (I.Bazzi, 2003, p.464-467). This algorithm is composed by two parts, and one is the acquirement of the mapping relationship between the formant and the phonetic observation information, which maps the formant parameters on the Mel frequency cepstral coefficient by a parameter-free nonlinear predictor, and establishes the prediction code text, and the other is the acquirement of the residual information of speech signals, which adopts the EM algorithm to train the residual coefficients of the speech signals and search the optimal format parameters in the prediction code text. Combining with the restriction condition with the target orientation, L. Deng et al put forward a nonlinear predictor which could be used in tracking of VTRs (L.Deng, 2006, p.425-434). This nonlinear predictor maps the formant parameters on LPCC, not on MFCC, and because the LPCC has good separation character, so it can enhance the computation efficiency.

Above two algorithms all first quantifies the parameter space of formant, and maps the quantified formant parameters on MFCC or LPCC to form the prediction code text, and finally selects the optimal formant parameters by training the residual coefficients. To quantify the parameter space, the quantifying dimension should be selected, and too big quantifying dimension will produce large computation, and too small quantifying dimension will influence the tracing precision. Aiming at this problem, L. Deng also put forward an improved method. He regarded the formant parameters as the continual values of variables in the hidden status, and adopted the Kalman filtering and the smoothing technology to trace the track of VTRs, which could solve the problem induced by the quantification of frequency field space (L.Deng, 2007, p.13-23). This method introduced extra prior information by the form of VTRs in the VTRs tracking process. Because the prior information can capture the timing character of VTRs track in the generation process of sound, so this method can exactly trace not only the VTRs with obvious frequency spectrum peak value, but also the VTRs phonetic segments (such as stops, spirants) without obvious formant structure. But this method has a deficiency, i.e. it needs to linearly process the nonlinear predictor, because the Kalman filtering is an implementation of Bayes filtering, and it is the optimal linear filter under the rule of the minimum square error, so it can not be used in the nonlinear occasions. The linear processing of nonlinear speech model will not only increase the computation, but the linearized model can not often represent real nonlinear model.

To overcome the deficiency that the nonlinear predictor needs linear processing, foreign scholars applied the particle filtering technology in the formant tracking based on the HDM model in recent years. The particle filtering is another method to realize the Bayes filtering, and it adopts a group of randomly weighted particles to approach the posterior probability distribution, and because it is not limited by the linear Gaussian conditions, so it is widely applied in the control domain. Yanli Zheng et al first applied the particle filtering technology in the tracking of phonetic formant (Yanli, 2004, p.565-568). This method can process the nonlinear model, and needs not implement the linear processing

to the nonlinear speech model, and it is easy to be implemented. But Yanli Zheng et al only offered a developmental idea, and the speech model they used was the simplified HDM without target orientation, so the tracing precision still needs to be enhanced.

4. Conclusions

The formant tracking technology of speech signals is continually developing, and many foreign and domestic scholars are applying themselves to the research about it and have put forward many methods and algorithms. The formant tracking method developed from the LPC analysis method to the HMM model matching method and to the HDM model matching method. Now, the formant tracking method based on HDM is more and more emphasized by researchers. Of course, the HDM model matching method still has some deficiencies, for example, whether the established model can exactly describe the character of speech signals enough, and how to enhance the precision of the model tracing when simplifying the computation. As the research develops, these problems will be solved gradually, and the HDM model matching method will certainly exert important function in the domain of the speech signal formant tracking.

References

A.M.De Lima Araujo, F.Violaro. (1998). Formant frequency estimation using a mel scale LPC algorithm. *Telecommunications Symposium, ITS '98 Proceedings.* Vol.1. p.207-212.

D.J.Broad, F.Clermont. (1989). Formant estimation by linear transformation of the LPC cepstrum. *Journal of the Acoustical Society of America.* No.86(5). p.2013-2017.

D.Talkin. (1987). Speech formant trajectory estimation using dynamic programming with modulated transition costs. *Journal Acoustical Society of America*. No.82(S1). p.S55.

D.T.Toledano, J.G..Villardebo, L.H.Gomez. (2006). Initialization, training, and context-dependency in HMM-based formant tracking. *IEEE Transactions on Acoustics, Speech, and Signal processing.* No.14(2). p.511-522.

G..Kopec. (1986). Formant Tracking Using Hidden Markov Models and Vector Quantization. *IEEE Transactions on Acoustics, Speech, and Signal processing.* No.34(4). p.709-729.

G.Kopec. (1985). Formant tracking using hidden Markov models. *Proc. ICASSP.* Vol.10. p.1113-1116.

H.B.Richards, J.S.Bridle. (1999). The HDM: A segmental hidden dynamic model of coarticulation. *Proc. ICASSP.* Vol.1. p.357-360.

I.Bazzi, A.Acero, L.Deng. (2003). An expectation maximization approach for formant tracking using a parameter-free non-linear predictor. *Proc. ICASSP.* Vol.1. p.464-467.

J.Darch, B.Milner, S.Xu, S.Vaseghi and Y.Qin. (2005). Predicting formant frequencies from MFCC vectors. *Proc. ICASSP.* Vol.1. p.941-944.

J.K.Baker. (1975). The Dragon System-An overview. *IEEE Transactions of Acoustics, Speech, and Signal Processing.* No.23(1). p.24-29.

L.Deng, A.Acero, I.Bazzi. (2006). Tracking vocal tract resonances using a quantized nonlinear function embedded in a temporal constraint. *IEEE Transactions on Acoustics, Speech, and Language processing.* No.14(2). p.425-434.

L.Deng. (1998). A dynamic, feature-based approach to the interface between phonology and phonetics for speech modeling and recognition. *Speech Communication.* No.24(4). p.299-323.

L.Deng, J. Ma. (2000). Spontaneous speech recognition using a statistical coarticulatory model for the vocal-tract-resonance dynamic. *Journal of the Acoustical Society of America.* No.108(6). p.3036-3048.

L.Deng, L.J.Lee, H.Attias, A.Acero. (2007). Adaptive kalman filtering and smoothing for tracking vocal resonances using a continuous-valued hidden dynamic model. *IEEE Transactions on Audio, Speech, and Language processing.* No.15(1). p.13-23.

M.Lee, J.vanSanten, B.Mobius, J.Olive. (2005). Formant Tracking Using Context-Dependent Phonemic Information. *IEEE Transactions on Speech and Audio Processing.* No.13(5). p.741-750.

P.Zolfaghari, T.Robinson. (1996). Formant analysis using mixtures of Gaussians. *Proc. ICSLP.* Vol.2. p.1229-1232.

S.S.McCandless. (1974). An algorithm for automatic formant extraction using linear prediction spectra. *IEEE Transactions on Acoustics, Speech, and Signal Processing.* No.22(2). p.135-141.

Yanli.Zheng, M. Hasegawa-Johnson. (2004). Formant tracking by mixture state particle filter. *Proc. ICASSP.* Vol.1. p.565-568.

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Recovery of Software Architecture Using Partitioning Approach by Fiedler Vector and Clustering

Shaheda Akthar
Sri Mittapalli College of engineering
Affiliated to JNTU, Kakinada, India
E-mail: shaheda.akthar@yahoo.com

Sk.MD.Rafi
Sri Mittapalli Institute of Technology for Women
Affiliated to JNTU, Kakinada, India
E-mail: rafi_527@rediffmail.com

Abstract

Software Architecture Recovery includes the extraction of design patterns. Patterns may be found using many techniques such as fielder vectors, using clustering methods, query languages etc. In this chapter, for evaluating design patterns clustering methods and the general notion of fielder vector are used.

Keywords: Fiedler Vector, Laplacian matrices, Eigen Vector, Cohesion, Coupling

1. Introduction

A Software system is comprised of modules (B.W. Kernighan, S.Lin, 1970)(R.J. Wilson, J.J. Watkins, Graphs, 1990) (which includes procedures, files, functions etc.). In the beginning these modules should be classified into subsystems. For this purpose, construct a graph G= (V, E) such that each vertex is consists of the modules and each edge shows the relationship between these modules. After that we decide the classifications of the nodes into subsets, there by the cohesion between the nodes of class may be maximized and the coupling between the nodes of different classes be minimized.

2. Modules Identification

We cannot construct a graph, without identifying modules and relations. In this way the first step is to find the modules and its relations among them.

The following ways can be adopted to find modules

1) The easiest way is to treat each file as a module because the functions in the file are semantically related.

2) We can consider groups of files as a module. But here the question is which files should be grouped?

3) We can also consider procedure as a module, as we are following the easier approach i.e. considering each file as a module.

3. Relations

There will be three types of relations among files.

f1 Useproc f2: shows that there is one function in f1 which calls the other function in f2.

f1 Usevar f2:shows that there is one function in f1 which uses a variable defined in f2.

f1 Implementby f2:shows that there is one header function in f1 which is implemented in f2.

In this way, we can construct a graph, (B.W. Kernighan, S.Lin, 1970)(R.J. Wilson, J.J. Watkins, Graphs, 1990) using these modules and relations among modules. The next step is to classify these modules into subsystems. So as to divide the graph into sub graphs for using the concept of Fiedler vector. First this we have to know what a Fiedler Vector (M. Fiedler) is.

For the input graph G= (V, E)

1) calculate the Adjacency matrix

2) Calculate the Degree matrix

3) Calculate the Laplacian matrix

The adjacency matrix is

$$A(i,j) = 1 \text{ if } (i,j) \in E$$
$$= 0 \text{ otherwise}$$

The degree matrix of the graph is the diagonal matrix of the row sum of the adjacency matrix

$$D = \text{diag}(\deg(i);\ i \in V)$$
$$\deg(i) = \sum_{j \in V} A(i,j)$$

The Laplacian matrix L is the difference between diagonal matrix and Adjacency matrix (B. Mohar, 1997)

$$L = D - A$$
$$L(i,j) = \sum_{(i,k) \in E} A(i,k) \quad \text{if } i=j$$
$$= -A(i,j) \text{ if } i \neq j \text{ and } (i,j) \in E$$
$$= 0 \text{ otherwise}$$

Now the matrix available is symmetric. The Eigen vector $(1, 1 \ldots.1)^T$ corresponds to trivial zero, Eigen values. With these Eigen values now, we get the Fiedler vector. We have to arrange first these Eigen values in the ascending order. The largest Eigen value and the second smallest Eigen value, whose corresponding Eigen vector is referred to as the Fiedler vector (M. Fiedler, 1975).

In this way Fiedler vector is known to us. Now decompose the graph into sub graphs. In this, the path sequence for the nodes has to be calculated using a permutation π. The sequence is that the elements of the edge weight matrix W decrease as the path is traversed **if $\pi(i) < \pi(j) < \pi(k)$ then $W(i,j) > W(i,k)$ and $W(j,k) > W(i,k)$**

Consider a vector $x^- = (x_1, x_2, \ldots\ldots x_{|v|})$ of continuous variables x_i . Calculate the penalty function

$$g(x)^- = \sum_{i=1}^{|V|} \sum_{j=1}^{|v|} W(i,j)(x_i - x_j)^2.$$

The constraints of this function are

$$\sum_{i=1}^{|V|} (x_i)^2 = 1 \text{ and } \sum_{i=1}^{|V|} x_i = 0.$$

4. Decomposing a graph into sub graphs

By making use of Fiedler vector concept, the graph has to be divided into sub graphs (R.J. Wilson, J.J. Watkins, 1990). The neighbourhood of the node i consist of its center node together with its immediate neighbours connected by edges in the graph.

$$N_i\hat{} = \{i\} \cup \{u ;(i,u) \in E\}$$

Assign each node measure of significance as the center of the neighbourhood. After assigning, it traverses the path defined by the Fiedler vector. Select center nodes on the basis of this measure to traverse the path. Assign weights to nodes based on the rank-order in the permutation. The weight assigned to the node is i€ V is wi=**Rank (π (i)).** After assigning weight to each node calculate the score function. The significance of the node will be known after giving score to each node. Score can be calculated by using a function

$$S_i = C_1(\deg(i) + |N_i \Lambda P|) + C_2 / Wi$$

Where C_1 and C_2 are threshold values that were detected heuristically. P is the set of nodes on the perimeter of the graph. The first term depends on the degree of the node and its proximity to the perimeter. In this way the nodes will be sorted according to their distance from the perimeter. This is proposed as it is better to decompose first from the outermost layer. The second term says that the first ranked nodes in the Fiedler vector are visited first. To locate the non overlapping neighbourhoods of the graph G, we use the scoring function. We traverse this list until we find a node K which is neither in the perimeter and also the calculated score should not exceed of its neighbours. If this condition gets satisfied then the node K together with its neighbours represents the first sub graph. This process will be repeated for all the nodes that satisfies the condition. Then we have to find out the sub graphs which are overlapping with its neighborhood. By doing this step the sub graphs are found. Thus Input the overlapping sub graphs to a clustering algorithm. There are two major approaches for subsystem classification.

1) Top down approach

2) Bottom up approach.

In the Top down approach, creation of subsystem includes all modules and then iteratively decomposes the current subsystem to create them at lower levels. In the bottom up approach, consider each module as a subsystem and then iteratively merge them to create those at higher levels.

Top down approaches suffer from exponential complexity as in A* algorithm. So, follow the bottom up approach.

1) For clustering, calculate the similarity between two nodes.
2) Identify a set of nodes that are pair wise & most similar. After identifying, create a cluster by taking the union of the most similar cluster or creation of more than one cluster is also possible by taking the union of some of pairs of this set.

5. Similarity/Dissimilarity measures

Two nodes are said to be similar if they have either the highest similarity measure or lowest dissimilarity.

If the component of the system is entirely connected to just other component, that connection should be computed as a lower dissimilarity than any other connection that is not complete. It is based on the percentage degree of vertices & common neighbours of the two vertices. That is, let p be the dissimilarity matrix and is defined by

p(i,j)=deg(i)+deg(j)-2*b(i,j)/deg(i)+deg(j)-b(i,j)

where deg(i)is the degree of the vertex i in the graph and b(i,j) is the number of common neighbour of vertices i and j. since deg(i)+deg(j)-b(i,j) is the number of all vertices connected to exactly one of the i and j. note that if deg(i)=deg(j)=b(i,j) then p(i,j)=0 and so i and j are completely similar. Note also if i and j have no common neighbor then p(i,j)=1 and so i and j are completely dissimilar.

After clustering now there is a need to optimize the solution. Then consider the measurement of intra-connectivity and inter-connectivity.

5.1 Intra-connectivity

It is a measure of connectivity between the two components that are grouped together in the same cluster. The degree of intra-connectivity should be high for good sub system partitioning , because many software level features are shared by the modules grouped with in a common subsystem.

$$\mathbf{A_i = m_i / N_i^2}$$

Where Ai is the intra-connectivity measurement
N_i is number of components
m_i intra-edge dependencies.

5.2 Inter-connectivity

It is a measure of connectivity between two distinct clusters. Inter connectivity should be very less. It is denoted by E_{ij}

$$\mathbf{E_{ij} = m_{ij} / 2N_iN_j \quad if\ i \neq j}$$
$$\mathbf{= 0\ if\ i=j}$$

i and j are clusters consisting of $N_i N_j$ components. m_{ij} is inter dependency.

The clusters will be derived. To apply the clustering techniques to software architecture recovery and reengineering, the object-attribute data matrix should be converted to object-object data matrix, so that the input reflects the interconnectivity of components. The clustering techniques are then used to minimize interconnections among components. Here we explain how the clustering technique could be used to support the identification of a pattern.

There are some client classes that are accessible to some subsystem classes. With the existing software architecture recovery assistants, especially file names based approaches; the result may look perfect for the subsystem. In other words, the architecture recovered through this type of technique is close to or the same as the modules that are partitioned by the designer.

Certainly, architecture capture is important and valuable. But we are also concerned with the ways to improve the architecture rather than simply capture it. Besides, in reality, the directory structures already often reveal the high level components of a system. Simply capturing software architecture at a higher level abstraction often has limited benefits. We get very different partitions by applying the clustering techniques to this example. In fact, the subsystem does not exist anymore, since many subsystem classes are directly accessed by or related to client classes. In other words, the clustering technique reveals that some classes in the subsystem are more closely related to client classes, which contradicts the design concept. Ideally, the subsystem classes should be grouped together as one unit. Clustering techniques could be used in this type of situations to enforce the architect to reason ways to keep the subsystem classes in a more cohesive manner. Facade pattern provides common interfaces to subsystem classes and facilitates separation of concerns. The subsystem classes in the new pattern-based design are grouped in the same unit according to the

clustering method, which is consistent with the original design. In this example, the clustering technique helps the adoption of a design pattern to reduce the coupling between the subsystem and the clients.

6. Conclusion

With respect to graph matching there is exponential complexity, however we have proven the complexity is linear in certain situations not for all the problems. Due to this problem, to decompose the matching problem into subunits (smaller graphs). On this subunits investigate suing edit distance method and use the Fiedler matrix for the partition of graph. This process may be a hierarchical framework which is suitable for parallel computation.

References

A.Marzal, E. Vidal. (1995). Computation of normalized edit distance and applications, IEEE Trans. Syst. Man Cybern. 25 (1995) 202-206.

B. Mohar. (1997). Some applications of Laplace eigenvalues of graphs, Graph Symmetry: Algebraic Methods and Applications, NATO ASI Series C, vol. 497, 1997, pp. 227–275.

B.W. Kernighan, S.Lin. (1970). An efficient heuristic procedure for partitioning graphs, Bell Syst. Tech. J. (1970) 291-307.

M. Fiedler. (1975). A property of eigenvectors of non-negative symmetric matrices and its application to graph theory, Czech. Math. J. 25 (1975) 619–633.

R.J. Wilson, J.J. Watkins. (1990). Graphs: An Introductory Approach: A first Course in Discrete Mathematics, Wiley International Edition. New York, etc., Wiley, 1990.

W.H. Haemers. (1995). Interlacing Eigen values and graphs, Linear Algebra Appl. 1995, pp. 593–616.

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

An Improved Wide-band Noise Signal Analysis Method

Wenyan Wang, Didi Liu & Xiaohua Wang
College of Electronic Engineering, Guangxi Normal University, Guilin 541004, China
Tel: 86-773-5826-559 E-mail: wwy@mailbox.gxnu.edu.cn

The research is financed by the National Nature Science Fund of China (No. 30973418) and the fund of Guangxi Education Department (No. 200808LX019) and the fund of young outstanding teacher of Guangxi Normal University (No.20097). (Sponsoring information)

Abstract

The representation form of most signals in practice is non-stationary signal, and the main analysis methods for these signals are centralized in the field of time frequency (or time scale), and a series of results have been acquired. A reversible transformation method based on time sequence bit is proposed from another angle, and this method can be used to analyze non-stationary wide-band noise signals according to the difference of some characters of noises and speech signals (such as the randomness degree). The experiment result indicates that this method can reversibly convert random sequences with unordered time sequence into approximately ordered time sequences, and the difference between the wide-band noise character and the speech character is relatively obvious. The method and the principle are simple, and the algorithm can be easily realized real time.

Keywords: Non-stationary signals, Reversible transformation, Wide-band noise

1. Introduction

Noises come from actual application environment, so their characters change largely. Noises can be additive or non-additive. For non-additive noises, some of them can be transformed as additive noises. For example, the multiplicative noises or convolution noises can be changed to additive noises by the isomorphic transformation. The noises concerned in the article can approximately include the voice interference and periodic noises, the impulse noise and the wide-band noise.

The speech interference is mainly induced by other speeches acquired by the mikes or the overhearing in the transfer.

The character of the periodic noise is that it contains many discrete narrow pecks which always come from the periodic rotation of the mechanisms such as engine. Electric interference, especially 50 or 60 Hz AC noises, also will induce periodic noises. Problems induced by periodic noises may be less, because they can be found by the power spectrum or removed by the filtering technology or the transformation technology. However, it is difficult to restrain the AC noise, because its frequency component is not the fundamental tone (it is blow the effective frequency of speech signals), but the harmonic wave component (it may cover the whole audio frequency spectrum by the pulse form).

The impulse noise is embodied as the narrow pulse suddenly occurring in the time domain wave, and it generally is the result of discharge. To eliminate this noise can be implemented in the time domain, i.e. the threshold value can be confirmed according to the average value of extents with speech noise signals. When the extent of noise exceeds this threshold value, the noise signal is judged as the pulse noise which can be depressed and even eliminated completely. If the interference pulses are not too close, they can be simply eliminated in the time function by the interpolation method according to neighboring sample values of signals.

The wide-band noise generally can be supposed as the Gauss noise and white noise. It comes from wind, breathing noise and common random noise source. The quantizing noise generally can be treated as white noise, and it also can be regarded as the wide-band noise. Because the wide-band noise completely superposed with speech signal in time domain and frequency domain, it is very difficult to eliminate it (Hu, 2000).

For the non-stationary signals such as wide-band noise, the traditional analysis measure is not sufficient in time domain or frequency domain, so it needs to develop the united time frequency analysis method which can optimize the non-stationary signals in the time domain and the frequency domain simultaneously. At present, the time frequency analysis has been a very active and important research task in modern signal processing domain.

The wide-band noise signal analysis is to reflect certain character of the noise signals differentiating with other signals by above various methods. With the development of the research of non-stationary signals, new signal analysis processing methods need to be developed continually, and these methods are making us more really cognize the essentials of the nature step by step.

A new reversible transformation method based on time sequence bit is proposed in the article, and this method can be

used to analyze the wide-band noises. By the noise experiment, this method can reversibly convert random sequences with unordered time sequence into approximately ordered time sequences, and the difference between the wide-band noise character and the speech character is relatively obvious. The method principle is simple, and the algorithm can be easily realized real time.

2. The reversible transformation based on time sequence bits

The reversible translation method based on the time sequence bit can be described as follows.

(1) Quantify the sampling analog signals as the "binary time sequence A";

(2) Supposed that the length of "unprocessed binary sequence B" is N, inquire whether the binary sequence represented by former n bits (n=1-N) in the sequence B occurs in the "processed binary sequence set C". And if it doesn't occur, stop the inquiry, and add the former n binary sequence into the sequence set C;

(3) Circulate step 2 until traverse the sequence A to end;

(4) Treat the final sequence (the final sequence may be same with certain sequence in the sequence set C);

(5) Transform each sequence X in the sequence set C into the decimal number x, and store it into the sequence set D in turn.

Supposed the length of sequence X is n, and X_j is the number denoted by the j'th number in the sequence X (0 or 1). So the transformation function is

$$x = \sum_{i=1}^{n} 2^i + \sum_{j=0}^{i} X_{j+1} \times 2^j$$

3. The analysis method of non-stationary wide-band noise signals

Apply above reversible transformation method based on the time sequence bit in the wide-band noise analysis.

Experiment environment: Windows 2008, Matlab 2008a, Cooledit2.1.

Experiment data: Sampling rate 22050 Hz, 16 bits quantification.

(1) Respectively generate 1s white noise and 1s brown noise by the Cooledit2.1 software, and the format is the Windows wave voice file (.wav). The time domain figure and the transformed time domain are respectively seen in Figure 1 and Figure 5.

(2) Record 1s fan noise, and the format is the Windows wave voice file (.wav). The time domain figure and the transformed time domain are respectively seen in Figure 2 and Figure 6.

(3) Record 1s flute music signals, and the format is the Windows wave voice file (.wav). The time domain figure and the transformed time domain are respectively seen in Figure 3 and Figure 7.

(4) Record 1s male voice of "twelve" and the format is the Windows wave voice file (.wav). The time domain figure and the transformed time domain are respectively seen in Figure 4 and Figure 8.1 and Figure 8.2.

Experiment method: Use Matlab 2008a to realize the programming of new method. Test the sampling data by the Matlab 2008a Simulation Platform. And the experiment figures are seen in Figure 1 to Figure 8.

4. Result analysis and theoretical analysis

In short time, the speech signals can be approximately regarded as time domain stationary signals. In practice, wide-band noise mainly includes white noise and color noises (pink noise and brown noise), and both of them are representative non-stationary random signals, and even in the time domain analysis of signals in short time, the wide-band noise still shows unpredictable random character, i.e. the randomness degree of wide-band noise is higher than the speech signals. The new transformation method in the article just can accord with this character. By large numerous of experiment, after new reversible transformation, the voice signal, the music signal and the wide-band noise could present completely different characters, and the random noise in unordered time sequence could show the character of rule.

Being similar with the ergodicity of random signals, in the wide-band noise analysis, the new method also embodies the tendency of "ergodicity" that the transformed sequence set data increase approximately and gradually.

5. Conclusions

For non-stationary signals, the most widely analysis methods are all based on the time frequency domain (or time scale) at present, and a series of result has been acquired. With the development of the research of non-stationary signals, new signal analysis processing methods should be developed continually. The non-stationary signals can be transformed from another angle in the article, and the reversible transformation method based on the time sequence bit is proposed. Though relative theories have not been studied deeply, but many experiments have showed that this method can

transform more unordered wide-band noise signals in time sequence into ordered sequences. The deeper research in the application of normal speech signal processing is been implemented.

References

Hu, Hang. (2000). *Speech Signal Processing.* Heilongjiang: Harbin Institute of Technology Press. May of 2000.

Zhang, Xianda & Bao, Zheng. (2001). *Non-stationary Signal Analysis and Processing.* National Defense Industry Press. July of 2001.

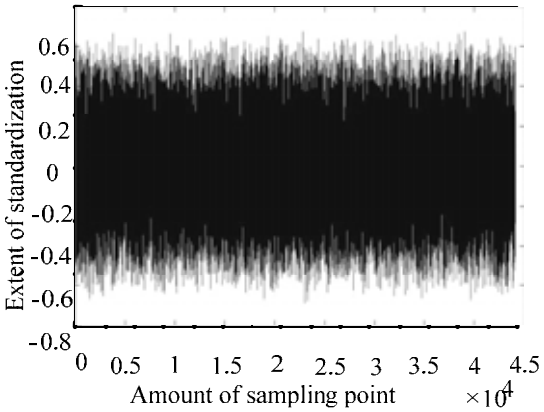


Figure 1.1 Second White Noise in Time Domain

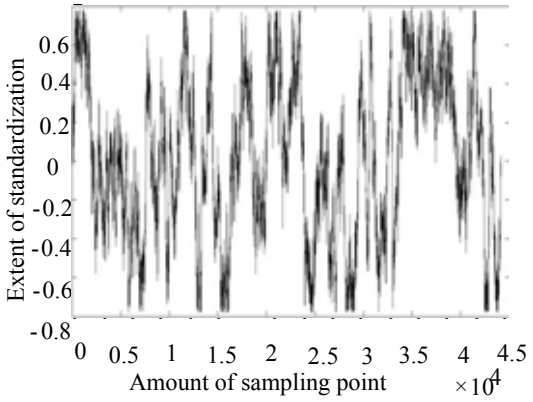


Figure 2.1 Second Brown Noise in Time Domain

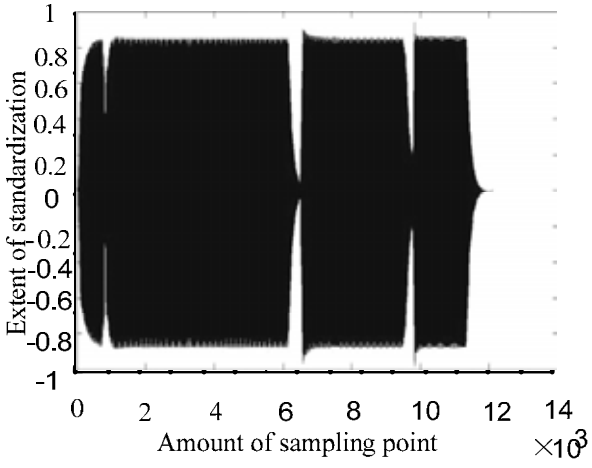


Figure 3.1 Second Flute Music in Time Domain

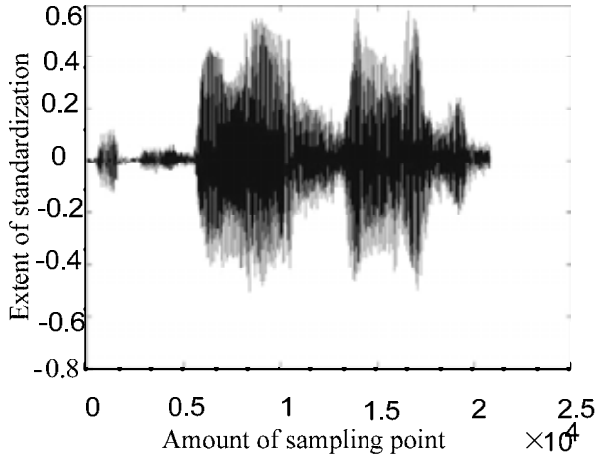


Figure 4.1 Second Male Voice "Twelve" in Time Domain

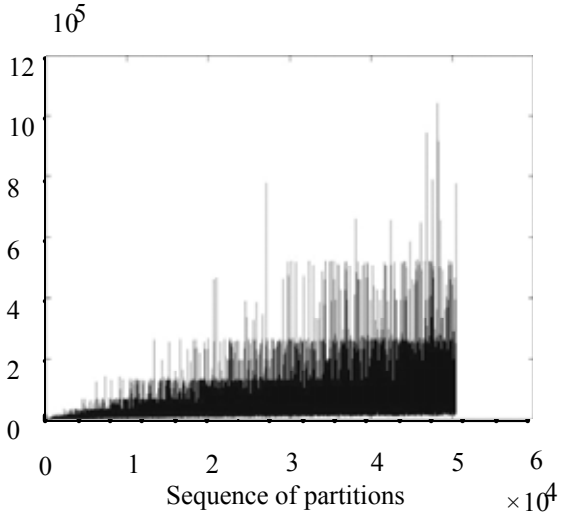


Figure 5.1 Second White Noise after New Transformation

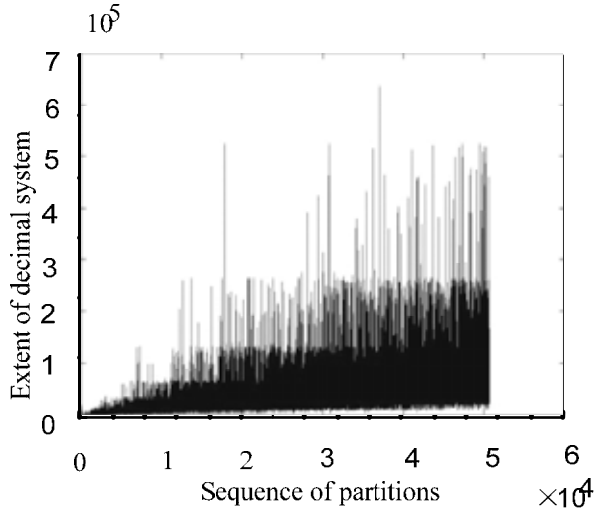


Figure 6.1 Second Brown Noise after New Transformation

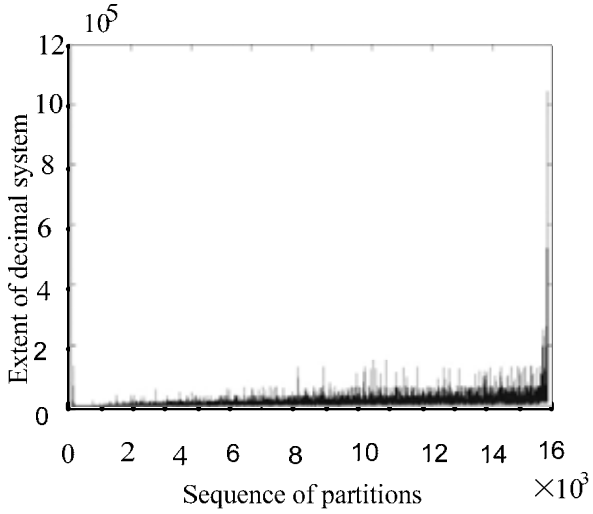


Figure 7.1 Second Flute Music after New Transformation

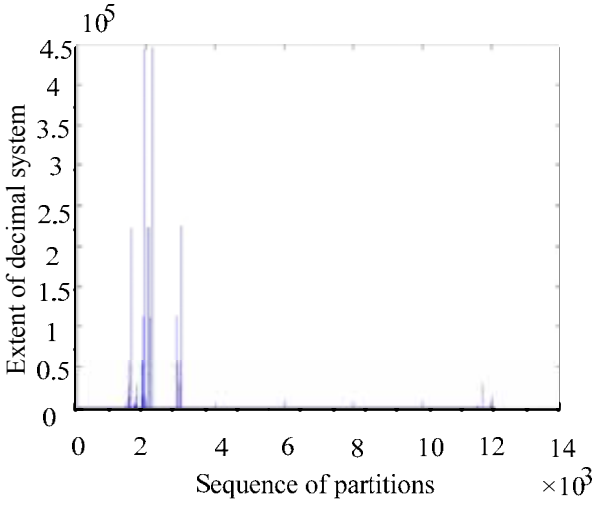


Figure 8.1.1 Second Male Voice “Twelve” after New Transformation (Overall View)

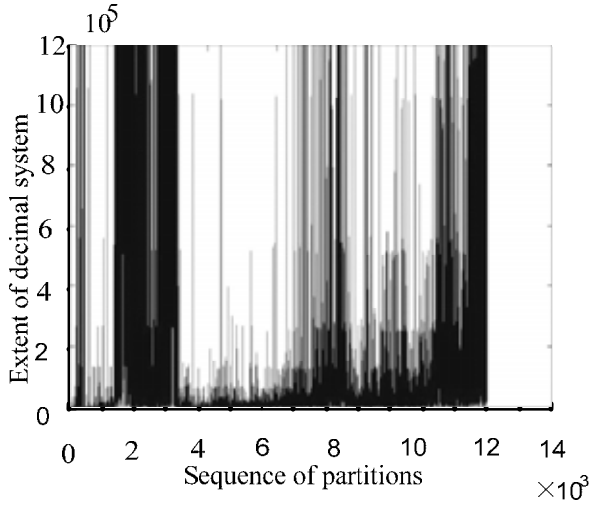


Figure 8.2.1 Second Male Voice “Twelve” after New Transformation (Local View)

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

A Framework of Collaborative Knowledge Management System in Open Source Software Development Environment

Modi Lakulu (Corresponding author)
Faculty of Information Technology and Communication, Sultan Idris University of Education
35900, Tg Malim, Perak, Malaysia
Tel: 60-19-2894-877 E-mail: modi@ftmk.upsi.edu.my

Rusli Abdullah, Mohd Hasan Selamat, Hamidah Ibrahim & Mohd Zali Mohd Nor
Faculty of Computer Science and Information Technology, University Putra Malaysia
43400, Serdang, Selangor, Malaysia
Tel: 60-3-8946-6557 E-mail: rusli, hasan, hamidah {@fsktm.upm.edu.my}, mohd.zali@my-newstar.com

Abstract

The global economy crisis reveals the advantages of Open Source Software (OSS). Software developers benefit not only from reduced cost of acquisition, but also access to source code and components. In this aspect, knowledge sharing among developers are immensely important in all facets of System Development Life Cycle (SDLC). Feller and Fitzgerald (2000) raised the critical questions on what life-cycle underpins the OSS model and what is the best methodology to support the OSS as well as what toolkit support OSS methodology. This paper shall discuss the formulation of Knowledge Management System (KMS) framework for sharing knowledge in OSS using SDLC from the planning phase until the maintenance phase. An initial fact finding survey was conducted on selected OSS developers in Malaysia to analyze the current usage and acceptance of OSS. The results are quite unexpected, with many OSS developers are still not fully using OSS tools in SDLC stages. The proposed KMS model is envisaged to allow OSS Community-of-Practice to share the OSS knowledge for the whole SDLC.

Keywords: Open source software, Knowledge, Knowledge management, Community of practice, KMS framework

I. Introduction

The OSS has become increasingly important and has attracted developers from both public and private sectors. Open source model, as a radically new software development model, begins in the mid-90s. Since then, many software created by open source model have been widely adopted and used by various industries. OSS is released under license conforming to the Open Source Definition (OSD) as articulated by the Open Source Initiative (OSI) [Feller and Fitzgerald, 2000]. OSS source codes are published and made available to the public, allowing everyone to copy, modify, and redistribute the source codes without paying royalties or fee [Jennex, 2006]. The most important requirements of an OSS system is that the source codes must be freely available to anyone who wish to examine or change for their own purpose [Godfrey and Tu, 2001]. This creates software which are reliable, high quality, inexpensive and timely[Feller and Fitzgerald, 2000].

Managing knowledge in OSS development is crucial to allow developers to capture, locate and share the knowledge on codes and methods. Knowledge is defined as a fluid mix of frame experience, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experience and information. It often embed not only in documents or repositories but also in organizational routines, processes, practice; and norm [Davenport and Prusak, 1998]. There are two types of knowledge; explicit and tacit. Explicit knowledge is clearly formulated or defined, easily expressed without ambiguity or vagueness, and codified and store in a database. Tacit knowledge is the unarticulated knowledge that is in a person's head that is often difficult to describe and transfer [Bollinger and Smith, 2001].

Knowledge Management (KM) in technical perspective, refers to the methods and tools for capturing, storing, organizing, and making accessible the knowledge and expertise within and across communities [Canfora, Cimitile, Visaggio, 2002]. The idea of a KM system is to enable employees to have ready access to the organization's based documented facts, sources of information, and solutions for a specific Community-of-Practice (CoP) members such as software engineer, database administrator, programmer and users.

In this paper we focus on System Development Life Cycle (SDLC) as a method in software development, using

information available throughout SDLC, as the main source of knowledge. Hence, the CoP can share knowledge throughout the development process such as the OSS tools they used in each phase, documentation, source code, support and best practice. Besides, to increase the participation of CoP, the communication infrastructure is very important as open source projects rely heavily on tools as modes of communication [Barnet and Schwaber, 2004].

Why we choose software development in Open Source? Based on the preliminary survey we found most of open source practitioners do not fully use OSS tools throughout the processes of software development. For example in planning phase only 8.3% of respondents use OSS tools, in design phase only 16.7% and in testing phase only 8.3% use OSS tools. They prefer to use OSS tools for web server, programming language and database, but do not prefer to use OSS tools in planning phase, designing phase and testing phase. (Refer to section 4, Result and Discussion for more details.)

The main motivation of this paper is to formulate a Knowledge Management System Framework for Open Source in Software Development. The function of this framework is to enable the CoP to share the knowledge of Open Source especially in Software Development. The proposed framework is based on Hahn and Subramani (2000), Meso and Smith (2000) Tiwana' s KMS Architecture (2002), Abdullah, Sahibuddin, Alias, Selamat (2006) and Shankar, Acharia, Baveja (2009) and consist of five main parts; layers, functionality, process, knowledge and Community-of-Practice.

The strategies of managing knowledge according to ontology give the opportunity to CoP members to be knowledgeable in all facets of System Development Life Cycle. Knowledge sharing and collaboration among the CoP create the knowledge networks and can be considered to be part of the third generation knowledge management initiatives. The benefits of the above framework is it's ability to provide a mechanism to share the knowledge throughout the development processes, from the planning phase until the maintenance phase in terms of OSS tools, experience, best practice cases, research, documentation, source code, support and many others. Why users should change his/her mind to use the OSS? There are many reasons why the open source model has been successful and popular with developers and the benefits for the users/organizational in terms of Access to source code, broad rights, encourage software re-use, can increase code quality and security, decrease vendor lock-in, reduce cost of acquisition, increase customizability and community [Olla and Crider, 2006]. The other of the benefits of open source first is the low cost or free access to knowledge, technologies, and/or other product, second is a counter to quality issue and finally the initiative cut across organizational boundaries and improve and/or create communication paths and protocol leading to improved cooperation [Jennex, 2006].

The remainder of the paper is structured as follows: Section 2 discusses related works. Section 3 reviews the methodology, and Section 4 presents result and discussion, and finally the conclusion is presented in Section 5.

2. Related Works

In this section, the discussion shall focus on Software Development and Existing KMS Frameworks. The Software Development is focus on the phases in software development, while Existing KMS Framework is describing and evaluating five KMS frameworks. The details of software development and existing KMS framework are in section 2.1 and 2.2 respectively.

2.1 Software Development (SD)

Software engineering knowledge are acquired and generated in order to produce high quality software, shorter development period and lower development costs. There are many software engineering methods used in industry. Among the well known methodologies are SDLC, Rapid Application Development (RAD), Agile Methods, OOAD and RUP. In this paper, we choose SDLC starting from planning, Analysis, design, implementation and finally to maintenance [Hoffer, George, Valacich, 2005](refer to Figure 1), the activities for each phase of SDLC are shown in table 1.

2.2 Existing KMS Frameworks

Five existing KMS frameworks were evaluated and analyzed, as follows:

- KMS architecture by Meso and Smith (2000).
- Framework for KM Support by Hahn and Subramani (2000).
- Tiwana's KMS Architecture (2002).
- KMS framework for Higher Learning Institution (HLI) by Abdullah, Sahibuddin, Alias and Selamat (2006).
- Soft-system Knowledge management framework by Shankar, Acharia and Baveja (2009).

Meso and Smith (2000) proposed a KMS architecture which comprises of technology, function and knowledge (see Figure 2). They proposed that the technology should include computer-mediated collaboration, electronic task management, and messaging. The functions utilize KM processes in using, finding, creating and packaging knowledge. Finally, the knowledge content includes how to answer the questions such as the know-how, know-what, know-why and so on.

Hahn and Subramani (2000) proposed a framework for KM Support (see Figure 3) which suggests two important considerations in managing knowledge, as follows:

- Where the knowledge resides and
- The extent to how the knowledge is structured.

The locus of the knowledge determines how the KMS connects a user (with a problem or question regarding an artifact) to an expert. Meanwhile, the level of a priori structure determines the extent to which KMS usage imposes the burden of a translation (or transformation of the problem or question) to a form that corresponds to implicit logic underlying the priori structure.

Tiwana (2002) has proposed KMS architecture consists of seven layers, which are interface, access, collaborative, application, transport, integration and repositories.

Meanwhile, Abdullah, Sahibuddin, Alias and Selamat (2006) proposed a KMS framework for HLI which includes 6 components: Architecture, Functionality and Application, Taxonomy and Processes, Culture, Psychological and Audit, as depicted in figure 4.

Likewise, Shankar, Acharia and Baveja (2009) proposed a Soft-System Knowledge Management framework (see figure 5), which propose an approach for KM system development to ensure a fit between the organizational needs on new product development (NPD) and KM initiatives. This framework consists of three main components, as follows:

- Knowledge Sharing Methods
- Organizational Level
- Key Enablers

The comparison on the major components in all of the above frameworks is depicted in Table 2. In summary, the existing frameworks overlaps on some general components such as infrastructure, technology, repository and collaboration. However, these combined frameworks are still not sufficient to support the Open Source Software Knowledge Management System (OSSKMS) requirements.

3. Methodology

There are three steps taken in the methodology, first conducted a survey and analyzed, second analyzed KMS framework and lastly the process of the formulation of the proposed framework. We choose this method because we stated with identified the problems and then analyzed several KMS frameworks as a reference to formulate a proposed framework. The formulation of the proposed framework is take into account the problems identified from the survey result and existing KMS framework. This is very important to make sure the proposed framework is based on the objective and the limitation.

A survey was conducted in selected software development organizations in Klang Valley Malaysia to gauge the developer's awareness of OSS and to ascertain the usage of OSS tools in OSS development. Fifty questionnaires were distributed to the respondents, but only twelve questionnaires were returned. The questionnaire data were verified and was analyzed using Rasch Method and SPSS. The result of the survey contributed to the formulation of the proposed framework The main components derived from the questionnaire are open source infrastructure, knowledge and CoP. Finally we analyze five existing KMS frameworks and formulate a proposed KMS framework. The details of the formulation of the framework is discussed in section 3.1.

3.1 Formulation of the Framework

A new framework shall be synthesized from previous frameworks by Hahn and Subramani (2000), Meso and Smith (2000), Tiwana's architecture (2002), Abdullah *et al.*(2006), and R. Shankar *et. al.*(2009) as follows:

- Structured knowledge, which includes document repository, data warehousing, databases, repositories and data mining are derived from all from frameworks. This knowledge shall be mapped as OSS knowledge onto each phase of software development.
- The collaborative environment elements are derived from Meso and Smith (2000), Hahn and Subramani (2000), Tiwana (2002), and Abdullah *et. al.*(2006).
- Functions or process are derived from Meso and Smith (2000) and Abdullah *et. al.*(2006).
- Knowledge are derived from Meso and Smith (2000), and Shankar *et. al* (2009).

The result of proposed KMSOS^{2}oD Framework is shown in Figure 6. The proposed framework is formulated especially to support SDLC using OSS and consists of five main components: layers, components, process, knowledge, and CoP (see Figure 6).

The layers consists of application, collaboration tools, open source infrastructure, technologies, and repositories.

Meanwhile, the functionality layer includes portal, collaboration tools, OS infrastructure, technologies, and repositories. Basically the OS infrastructures are to act as a bridge between repositories, technologies, collaboration tools and application to CoP.

The Processes are similar to KM processes for acquiring or collecting, organizing, disseminating or sharing knowledge, using knowledge and store the knowledge. The knowledge content focuses on all facets of SDLC, starting from planning, analysis, design, implementation and finally maintenance. There are two categories of knowledge tacit and explicit. Based on the SECI model by Nonaka and Takeuchi (1995), there are four conversion patterns of knowledge are Socialization, Internalization, Combination, and Externalization. Finally, the CoP consist of students, academicians, programmers, system analyst and users.

OSS, Software Development (SD) and Knowledge Management (KM) are the main conceptual foundation for the proposed framework. These three foundations are collaborating among each other during KM activities. The OSS will provide the knowledge taxonomy as knowledge content either as a tacit or explicit knowledge. This knowledge will be map and manage into SDLC phases and activities. Finally, KM will act as tools for capturing, storing, organizing, and making accessible knowledge and expertise within and across communities.

How the proposed framework can support the knowledge and use OSS tools? Basically, in this case the CoP can only disseminate/acquire the OSS knowledge and mapped into software development process either in planning, analysis, design, implementation or maintenance through the portal. In this case, software development process refers to SECI model. All the OSS knowledge will be stored according to the ontology.

4. Result and Discussion

Figure 7 shows the summary of Person and Item data, where the Persons are (developers) and Items are the OSS tools in software development. For person reliability, the Cronbach Aplha of 0.87 logit and the item reliability of 0.91 logit are good and very good (Fisher, 2007), respectively. The persons summary indicates a good separation of G=1.05 logit and good item spread of 2.54 logit. Meanwhile, the item summary reveals a good separation of G=2.81 logit and but a poor item spread of 8.28 logit. The poor item spread (2.54 logit) could be contributed by item Q5, Q11, Q6, and Q7.

Based on Person-Item MAP (Figure 10), the two person scoring the highest point are E2134 and L2132. Both of them score 3.06 logit. Both of them are male and their age group is between 31 to 35 years old. There are one person (K4124) score the lowers point. For items, the most difficult item is Q18, measured at 4.17 logit. This means 25% strongly do not agree and 75% do not agree that products of OSS are very hard to commercialize. Nevertheless, the easier item is Q5 and the min measure is -4.09 logit, where 8% agree and 92% strongly agree that OSS tools are easy get in the market / in the Internet. There is a big gap between item Q18 and Q15 and Q12. Q15 is about OSS tools are difficult to learn and Q12 is about OSS tools are suitable for non-profit company. For both questions, most of respondents are not agree that OSS tools are difficult to learn and suitable for non-profit company. Besides that, there are also exist big gap below the item mean, there are Q5 and Q11, Q6, Q7, Q13 and Q10 with contributes a poor item spread of 8.28 logit. Q11 is about OSS tools should be taught in the Higher Learning Institution in Malaysia, Q6 is regarding OSS tools are totally reducing cost of software development, Q7 is about OSS tools are always available in the internet, Q13 is about OSS tools should be transfer to knowledge for sharing purpose and Q10 is about companies/organizations should use OSS tools in software development. Based on the responds from the respondents, according to items Q11, Q6, Q7, Q13 and Q10, most respondents agree to the questions asked.

Nevertheless, 91.7% (see Figure 8) of the OSS developers do not use OSS tools in planning phase. Surprisingly, 83.3% did not use OSS tools in design phase. This situation same in the testing phase, where 91.7% did not use OSS tools in testing.

Figures 9 show that the OSS developers fully utilized OSS tools in Operating System, database, programming language and web server. Linux (75%) is the popular operating system compared to the others. MySQL (75%) is the most popular database and PHP (75%) is the most well known programming language. While, Apache (83.3%) is the highest usage for web server.

As a conclusion, usage of OSS tools in software development in planning, design and testing phases are very low. However, the majority of OSS developers fully utilize OSS tools in database, programming language and web server.

5. Conclusion

Knowledge sharing and collaboration among the community-of-practice create the knowledge networks and can be considered to be part of the third generation knowledge management initiatives. In a survey carried out to gauge on usage of OSS tools in OSS development reveals that OSS developers not fully utilized OSS tools throughout the system SDLC. Hence, information and knowledge passed down from various stages of SDLC are not captured nor flow smoothly between stages. As such, we propose the KMSOS2oD framework to provide a mechanism to share the knowledge throughout the development processes, from the planning phase until the maintenance phase in terms of OSS tools,

experience, best practice cases, research, documentation, source code, support and many others. This framework attempts to adapt the software development life cycle to develop application system. This framework gives the opportunities and guidelines for CoP to share their knowledge and encourage people to use OSS in software development throughout the development phase.

References

Abdullah, R, Sahibuddin, S, Alias, R. A., & Selamat, M.H. (2006). Knowledge Management System Architecture For Organizational Learning With Collaborative Environment, *International Journal of Computer Science and Network Security*, 6(3A), pp. 237-246.

Barnett, L., & Schwaber, C.E. (2004). Applying Open Source Process in Corporate Development Organizations, Forrester Research, Inc, Technology Square Cambridge, MA, USA..

Bollinger, A. S., & Smith, R.D. (2001). Managing Organizational Knowledge as a Strategic Asset, *Journal of Knowledge Management*, 5(1), pp. 8-18.

Canfora, G, Cimitile, A., & Visaggio, C. A. (2002). From Knowledge Management Concepts Towards Software Engineering. Product Focused Software Process Improvement, 4th International Conference PROFES 2002, Rovaniemi, Finland, December 9-11, 2002, pp. 407-422.

Davenport, T.H., & Prusak, L. (1998). *Working Knowledge: How Organisations Manage What They Know,* Boston, MA: Harvard Business School Press.

Feller, J., & Fitzgerald, B. (2000). A Framework Analysis of The Open Source Development Paradigm, Proceedings of the twenty first international conference on Information systems, December 2000, ACM, Brisbane, pp. 58-69.

Fisher, W.P. (2007). Rasch Measurement Transactions. [Online] Available:http://www.rasch.org/rmt/rmt211m.htm(8 April 8, 2009)

Godfrey, M.W., & Tu, Q. (2001). Evolution in Open Source Software: A Case Study, University of Waterloo. [Online] Available: http://plg.uwaterloo.ca/~migod/papers/2001/iwpse01.pdf (April 8, 2009)

Hahn, J., & Subramani, M. (2000). A Framework of Knowledge Management System: Issues and Challenges For Theory and Practice, Proceedings of the twenty first international conference on Information systems, ACM, Brisbane.

Hoffer, J. A., George, J. F., & Valacich, J. S. (2005). *Modern system Analysis and Design: Fourth Edition*, Pearson Prentice Hall, New Jersey.

Jennex, M. E. (2006). Open Source Knowledge Management. [Online] Available: http://www.igi-global.com/files/prefaces/IJKM%20preface%202(4).pdf (8 April 2009)

Meso, P., & Smith, R. (2000). A Resource-Based View of Organizational Knowledge Management Systems, *Journal of Knowledge Management*, 4(3), pp.224-234.

Mettam, G. R., & Adams, L. B. (1994). How to prepare an electronic version of your article. In B. S. Jones, & R. Z. Smith (Eds.), *Introduction to the electronic age* (pp. 281-304). New York: E-Publishing Inc.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-creating Company*. New York, Oxford: Oxford University Press.

Olla, P., & Crider, R. (2006). Open-Source Online Knowledge Portals for Education": Benefits and Risks. In A. Tatnal, *Encyclopaedia of Portal Technology and Applications.* IDEA. In pres

Shankar, R., Acharia, S., & Baveja, A. (2009). Soft-system Knowledge Management Framework for New Product Development, *Journal of Knowledge Management*, 13(1) , pp.135-153.

Tiwana, A, (2002). *The Knowledge Management Toolkit: Orchestrating IT, Strategy and Knowledge Platforms.* USA: Prentice Hall.

Figures and Tables

Table 1. Phases and activities of SDLC (SD using OSS)

Phases	Activities
Planning	• Project Identification and Selection • Project Initiation and Planning
Analysis	• Requirements Determination • Requirements Structuring
Design	• Files and Databases • Forms and Reports • Interfaces and Dialogues • System and Program Structure • Distributed Systems
Implementation	• Coding • Testing • Installation • Documentation • Training • Support
Maintenance	• Obtaining Maintenance Requests • Transforming Requests into Changes • Designing Changes • Implementing Changes

Table 2. Comparisons Between the Frameworks

Item/ Framework	Knowledge Management Support (Hahn and Subramani, 2002.)	KMS framework for HLI (Abdullah, 2008)	KMS Architecture (Meso & Smith, 2000)	TIWANA'S KMS Architecture (2002)	Soft-system Knowledge Management framework (R. Shankar, 2009)	KMSOS^{8}oD (Suggestion)
Application	NA	Y	NA	Y	Y	Y
Technology	Y	Y	Y	Y	NA	Y
Infrastructure	Y	Y	Y	Y	Y	Y
Repositories	Y	Y	NA	Y	Y	Y
Knowledge	NA	NA	Y	NA	Y	Y
Functionality/KM Process	NA	Y	Y	NA	NA	Y
Collaboration Tools	Y	Y	Y	Y	NA	Y
Open Source Infrastructure	NA	NA	NA	NA	NA	Y

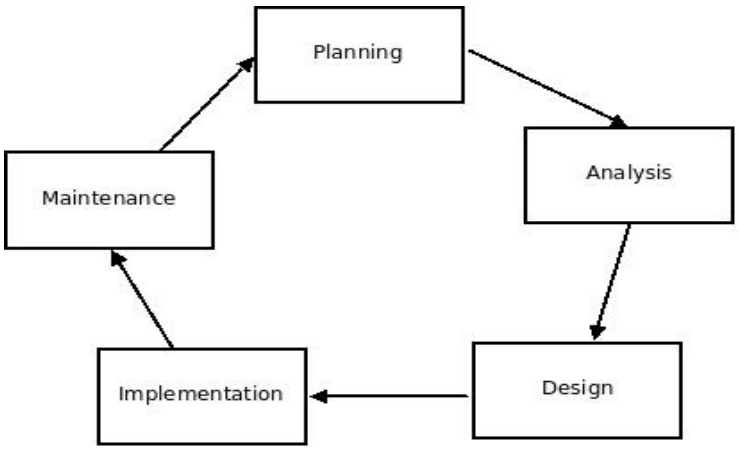


Figure 1. The System Development Life Cycle

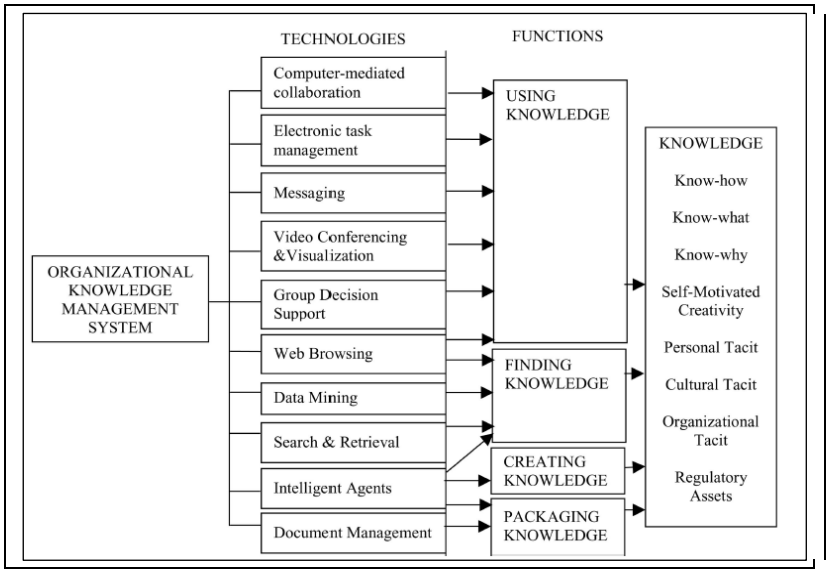


Figure 2. The Technical Perspective of a KMS

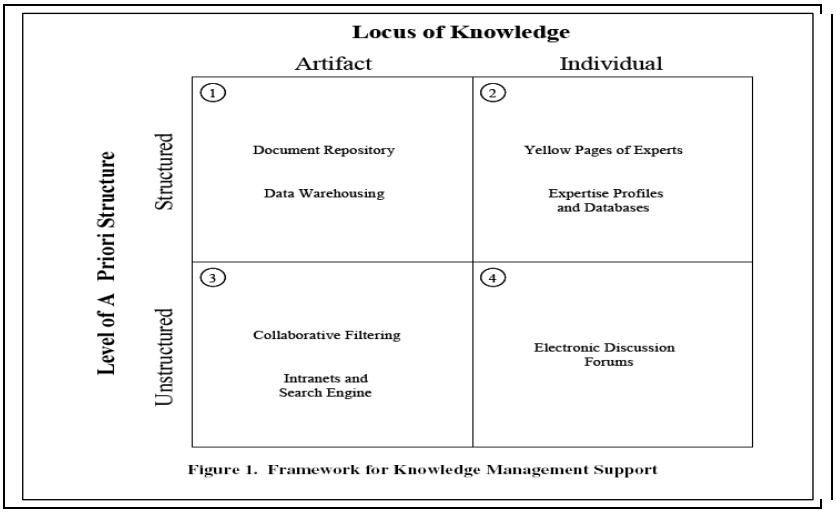


Figure 3. Framework for Knowledge Management Support

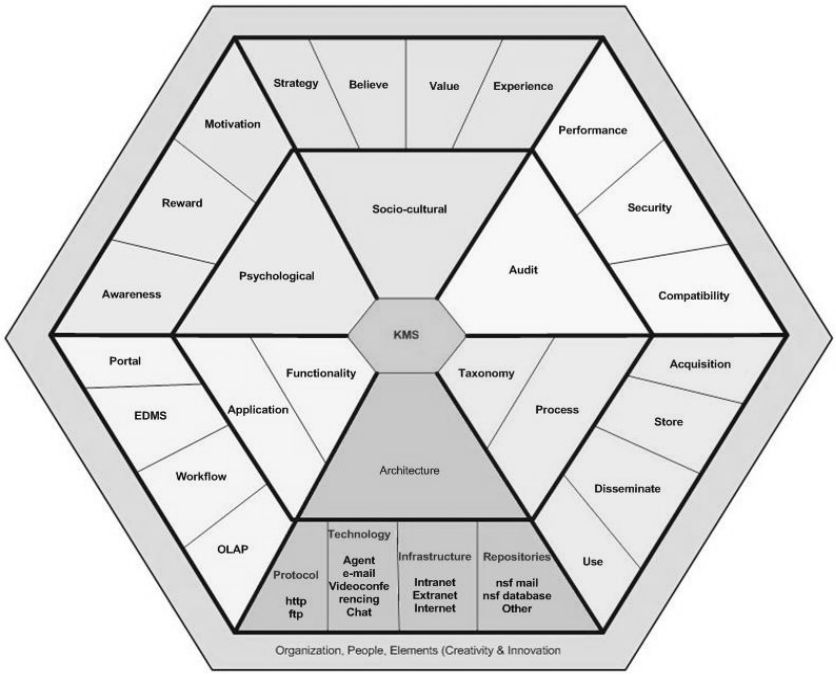


Figure 4. KMS framework for HLI

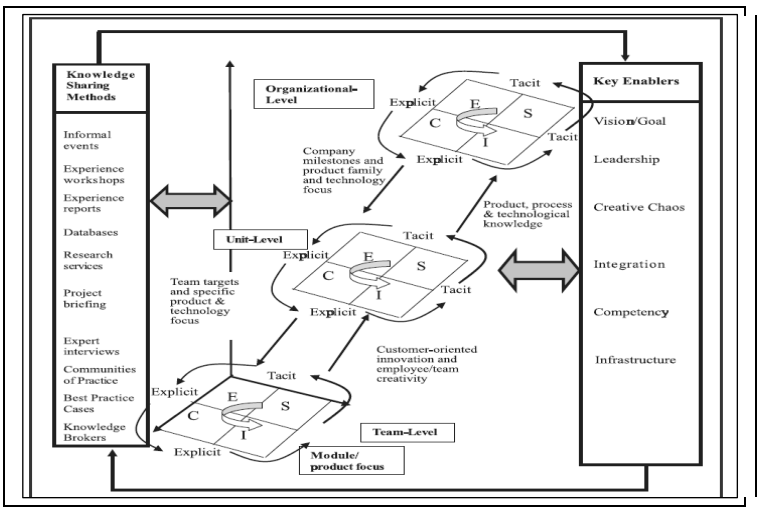


Figure 5. Soft-system Knowledge Management Framework

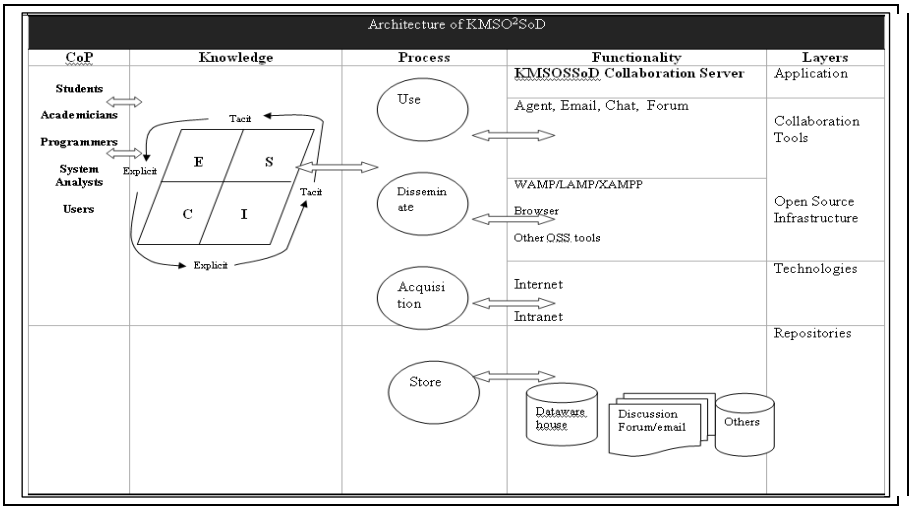
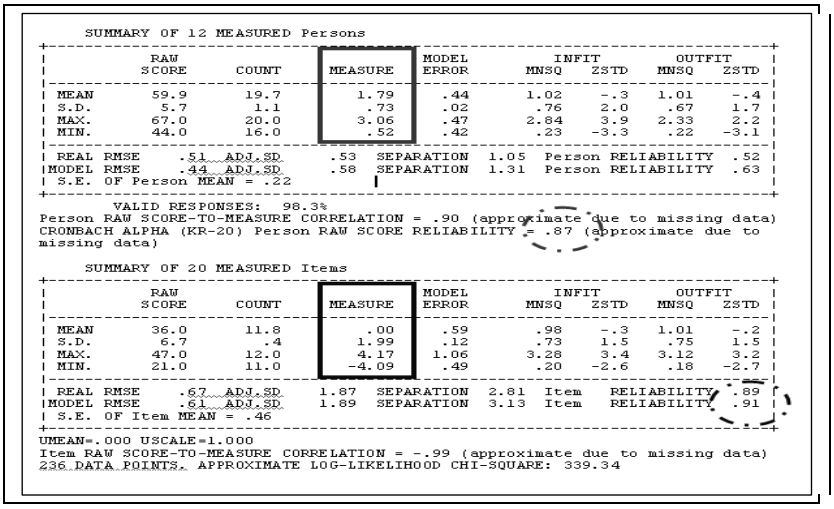


Figure 6. Proposed KMSOS²oD Framework



SUMMARY OF 12 MEASURED Persons

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	59.9	19.7	1.79	.44	1.02	-.3	1.01	-.4
S.D.	5.7	1.1	.73	.02	.76	2.0	.67	1.7
MAX.	67.0	20.0	3.06	.47	2.84	3.9	2.33	2.2
MIN.	44.0	16.0	.52	.42	.23	-3.3	.22	-3.1

REAL RMSE	.51	ADJ.SD	.53	SEPARATION	1.05	Person	RELIABILITY	.52
MODEL RMSE	.44	ADJ.SD	.58	SEPARATION	1.31	Person	RELIABILITY	.63
S.E. OF Person MEAN = .22								

VALID RESPONSES: 98.3%

Person RAW SCORE-TO-MEASURE CORRELATION = .90 (approximate due to missing data)

CRONBACH ALPHA (KR-20) Person RAW SCORE RELIABILITY = .87 (approximate due to missing data)

SUMMARY OF 20 MEASURED Items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	36.0	11.8	.00	.59	.98	-.3	1.01	-.2
S.D.	6.7	.4	1.99	.12	.73	1.5	.75	1.5
MAX.	47.0	12.0	4.17	1.06	3.28	3.4	3.12	3.2
MIN.	21.0	11.0	-4.09	.49	.20	-2.6	.18	-2.7

REAL RMSE	.67	ADJ.SD	1.87	SEPARATION	2.81	Item	RELIABILITY	.89
MODEL RMSE	.61	ADJ.SD	1.89	SEPARATION	3.13	Item	RELIABILITY	.91
S.E. OF Item MEAN = .46								

UMEAN=.000 USCALE=1.000

Item RAW SCORE-TO-MEASURE CORRELATION = -.99 (approximate due to missing data)

236 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 339.34

Figure 7. Reliability of Person-Item

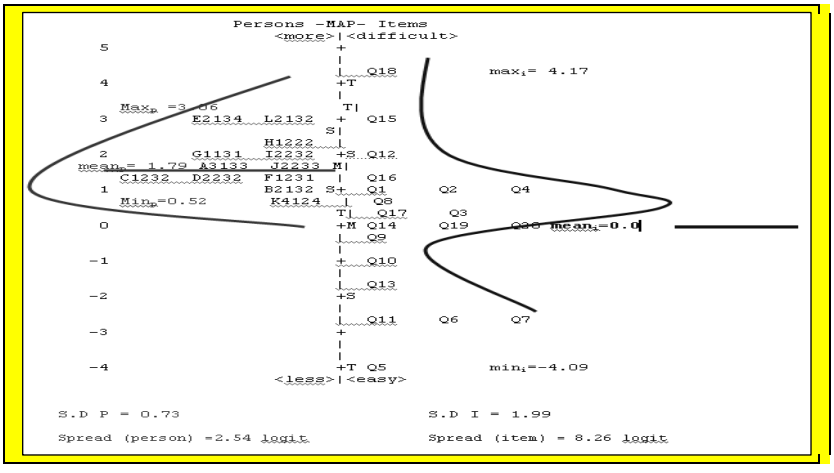


Figure 8. Person-Item MAP

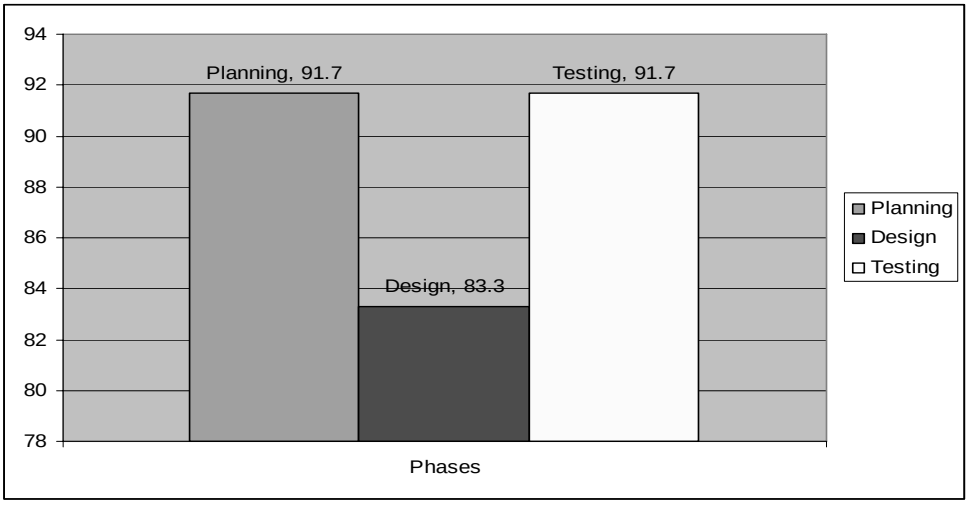


Figure 9. Did not used OSS tools

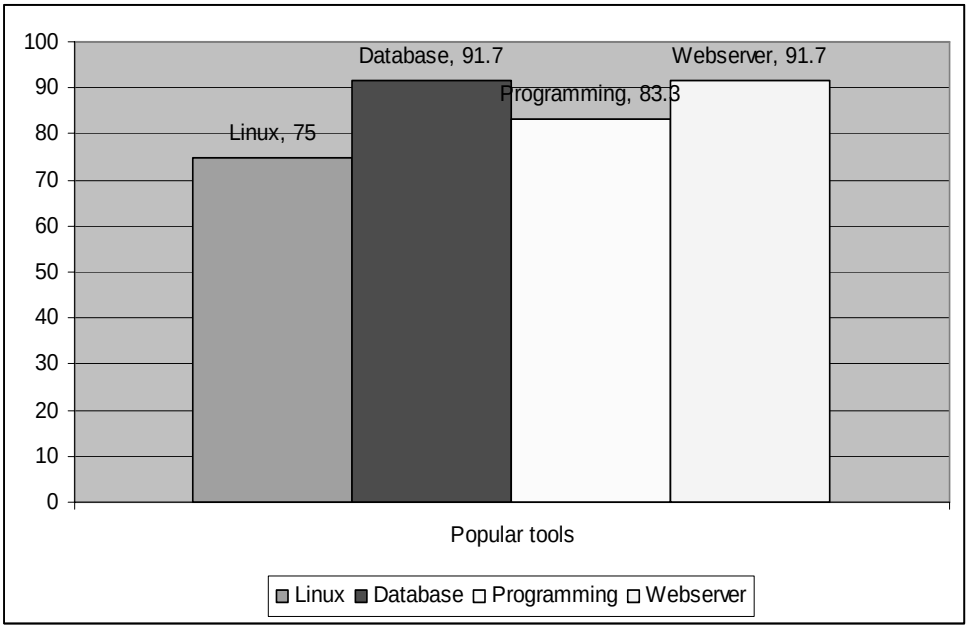


Figure 10. Popular OSS tools

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

CLCL-A Clustering Algorithm Based on Lexical Chain for Large-Scale Documents

Ming Liu (Corresponding author)
School of Computer Science and Technology, Harbin Institute of Technology
PO box 319, Harbin, Heilongjiang Province, 150001, China
E-mail: mliu@insun.hit.edu.cn

Xiaolong Wang
School of Computer Science and Technology, Harbin Institute of Technology
PO box 319, Harbin, Heilongjiang Province, 150001, China
Tel: 86-451-8641-3322 E-mail: xlwang@insun.hit.edu.cn

Yuanchao Liu
School of Computer Science and Technology, Harbin Institute of Technology
PO box 319, Harbin, Heilongjiang Province, 150001, China
Tel: 86-451-8641-3322 E-mail: lyc@insun.hit.edu.cn

The research is financed by National Natural Science Foundation of China No.60435020, National High Technology Development 863 Program of China No. 2007AA01Z172. (Sponsoring information)

Abstract

Along with explosion of information, how to cluster large-scale documents has become more and more important. This paper proposes a novel document clustering algorithm (CLCL) to solve this problem. This algorithm first constructs lexical chains from feature space to reflect different topics which input documents contain, and documents also can be separated into clusters by these lexical chains. However, this separation is too rough. So, idea of self organizing mapping is used to optimize cluster partition. For agglomerating documents with semantic similarities into one cluster, influences from similar features are also considered. Experiments demonstrate that because effects of semantic similarities between different documents are considered, CLCL has better performance than traditional document clustering algorithms.

Keywords: Lexical chain, Large-scale document clustering, Self organizing mapping, Neuron adjustment

1. Introduction

Along with evolvement of technology on internet, how to cluster large-scale information from website has become more and more important. After some tags are removed from webpage which are coded as HTML or XML, information from website is just document. This is why document clustering has become an effective method to analyze information from website (Niu et al., 2007; Luo et al., 2009; Xu & Wunsch, 2005). Recently, there are many document clustering algorithms, such as K-means (Guo et al., 2006), FTC (Beil et al., 2002), FCM (Wang et al., 2004), Hierarchical clustering (Wu et al., 2009), WEBSOM (Azcarraga et al., 2004). These algorithms all use Vector Space Model (VSM) to organize documents. In this model, all the features from feature space are used to construct vectors to represent documents and clusters. This method not only imports many useless features, but also augments dimension number of vectors to increase running time. This problem is called as “disaster of dimensionality”. When dimension number of feature space becomes larger, this problem is worse. That will make similarities among most documents close to 0, and will decrease partition ability of document clustering algorithms (Jennifer & Carla, 2004).

There is another problem of traditional clustering algorithm. That is traditional algorithms often use Euclidean distance

as similarity computation. This similarity is decided by minus between weights of same feature in different vectors (Tuomo et al., 2007). In document clustering, words are often used as features, and different words may have semantic similarity (Liu et al., 2009). That means different words reflect similar meaning. This situation will cause that even if feature vectors of different documents don't share same features, these two documents will also reflect similar meaning because of features which have semantic similarities among them. There have been proposed some clustering algorithms which import semantic similarity, such as ConSOM (Liu et al., 2008). It uses WordNet to compute semantic similarity between different features, and combine semantic similarity and Euclidean distance together to form a novel similarity computation method. However, it neglects importing semantic similarity to optimize partition of clusters.

In order to solve previous problems, a novel document clustering algorithm (CLCL) is proposed in this paper. This algorithm constructs lexical chains to remove features which are irrelative to topic of document. Lexical chains are also used to construct initial clusters. After construction of initial clusters, idea of self organizing mapping is imported to optimize partition of clusters and influences from similar features are considered in similarity computation and neuron adjustment. Experiments demonstrate that because semantic similarity is imported, precision and time complexity of CLCL are better than those of traditional document clustering algorithms.

2. Construction of lexical chain

Lexical chain is first proposed by Hirst in literature (Jane & Graeme, 1991). It is constructed by linearly scanning feature space, and each chain may reflect profile of topic information which document reflects. This technique has been applied in many fields, such as text analysis and abstract extraction (Chan, 2004; Kumar et al., 2003). In order to construct lexical chain, the first thing is to know how to compute similarity between different features. In this paper, co-occurring word vectors are constructed to compute semantic similarity between different features.

2.1 Semantic similarity computation

The linguist indicates: "the syntax function of a feature is the distribution of this feature" (Sven et al., 1998). The context of a feature is a typical distribution. So, if the contexts of different features are almost the same, the semantic similarity between these features is close. The word before the feature and the word after the feature mostly determine the semantic meaning of this feature. So, in this paper, co-occurring word and co-occurring word probability are used to construct feature's co-occurring word vector. Each dimensionality of this vector corresponds to one co-occurring word. The value of this dimensionality is the co-occurring probability between the feature and its co-occurring word. By computing the Kullback-Leibler divergence (Frans, 2005) between different co-occurring word vectors, semantic similarity between different features can be gotten.

$$SSim(F_p, F_q) = 1 - H(\boldsymbol{FV}(\boldsymbol{F_p}), \boldsymbol{FV}(\boldsymbol{F_q})) \tag{1}$$

Formula (1) describes how to compute similarity between tow features. $H(\boldsymbol{FV}(\boldsymbol{F_p}), \boldsymbol{FV}(\boldsymbol{F_q}))$ is the Kullback-Leibler divergence between two co-occurring word vectors-$\boldsymbol{FV}(\boldsymbol{F_p})$ and $\boldsymbol{FV}(\boldsymbol{F_q})$-of features F_p and F_q. Main meaning of this formula can be gotten from formula (3).

$$P_p(CoW_k) = \frac{Fre(F_p, CoW_k)}{Fre(F_p) * Fre(CoW_k)} \tag{2}$$

Formula (2) shows how to compute co-occurring probability of feature-F_p and its co-occurring word-CoW_k. This probability can reflect the semantic relation between the feature and its co-occurring word at certain degree (Rishi & David, 2006).

$$H(\boldsymbol{FV}(\boldsymbol{F_p}), \boldsymbol{FV}(\boldsymbol{F_q})) = -\sum_{i=1}^{n} \frac{(p_i + q_i)}{2} \log_2 (p_i + q_i) + \frac{1}{2}\left[\sum_{i=1}^{n} (p_i * \log_2 p_i) + \sum_{i=1}^{n} (q_i * \log_2 q_i) \right] \tag{3}$$

Formula (3) describes the Kullback-Leibler divergence between F_p and F_q. p_i is ith co-occurring word probability in $\boldsymbol{FV}(\boldsymbol{F_p})$, and q_i is ith co-occurring word probability in $\boldsymbol{FV}(\boldsymbol{F_q})$. n represents the size of co-occurring word vector. From formula (3), it can be gotten that, the larger the difference between the contexts of different features is, the bigger the value of Kullback-Leibler divergence is. So, Kullback-Leibler divergence can compute the semantic similarity between different features.

2.2 Construction of lexical chain

After semantic similarity is computed, lexical chains can be constructed to represent topic which each document reflects. Besides, lexical chains also can be used to separate feature space to construct initial clusters. In this method, each chain represents one cluster, and this cluster includes the documents which all reflect the topic which this chain describes. Construction steps of lexical chain are shown as follows.

[1] Assume input document set as D, and ith document in it as D_i. Assume FS_i as feature set of D_i. Assume LC_i as chain

set of D_i, and L_k as kth lexical chain in LC_i.

[2] Scan FS_i from top to down. Assume F_j as the feature which is being scanned.

[3] Use formula (4) to compute similarity between F_j and each lexical chain in LC_i. Assume L_k as the lexical chain which has the max similarity to F_j, and insert F_j in L_k.

[4] Repeat steps [1] ~ [3] until all the features in FS_i have been scanned.

$$Sim(L_k, F_j) = \sum_{LF_d \in L_k} SSim(LF_d, F_j) \Big/ |L_k| \tag{4}$$

Formula (4) shows how to compute similarity between lexical chain and feature. In this formula, LF_d means dth feature which is includes by L_k. This formula uses average similarity between F_j and each feature in L_k to be the similarity between L_k and F_j as literature (Gonenc and Ilyas, 2007) shows.

After previous construction, one set can be used to include lexical chains of each document, and each chain in it reflects one subtopic which this document describes. Formula (5) shows how to compute the weight of lexical chain L_k. By this formula, the lexical chain which has the largest weight can be regarded as the representation of topic which this document emphasizes.

$$w(L_k, D_i) = \sum_{LF_d \in L_k} w_d(LF_d, D_i) \times \log(|L_k|) \tag{5}$$

In formula (5), $w_d(LF_d,D_i)$ represents weight of LF_d in document D_i. This weight is computed by classical TF/IDF (Akiko, 2004) method as formula (6) shows.

$$w_d(LF_d, D_i) = fre(LF_d, D_i) \Big/ fre(LF_d) \tag{6}$$

In formula (6), $fre(LF_d,D_i)$ represents frequency of appearances of LF_d in D_i. It is TF value of LF_d. $fre(LF_d)$ represents number of documents which includes LF_d. It is DF value of LF_d.

Previous steps also can be used to partition feature space into some lexical chains. Each chain among them represents one cluster, and this cluster includes the documents which all reflect similar information to this chain. The approach which constructs lexical chains from feature space is similar to previous approach which constructs lexical chains from document. It is shown as follows.

[1] Assume input document set as D. Assume FS as feature space of D. Assume LS as lexical chain set of FS, and L_k as kth lexical chain in LS.

[2] Scan FS from top to down. Assume FS_j as the feature which is being scanned.

[3] Use formula (4) to compute similarity between F_j and each lexical chain in LS. Assume L_k as the lexical chain which has the max similarity to F_j, and insert F_j in L_k.

[4] Repeat steps [1] ~ [3] until all the features in FS have been scanned.

2.3 Initial cluster partition by lexical chain

After previous construction, each lexical chain can represent one cluster, and we can compute semantic similarity between document and lexical chain to map documents into clusters.

Formula (7) shows semantic similarity between document and cluster. In this formula, $L(D_i)$ represents lexical chain which is constructed from document D_i and has the largest weight. This chain is used as the representation of topic which D_i reflects. L_k represents lexical chain which is constructed to represent cluster C_k. $w_d(f_t,D_i)$ represents weight of f_t in document D_i as formula (6) shows. $w_c(f_t,C_k)$ represents weight of f_t in cluster C_k. It can be computed by formula (8).

In formula (7), two parameters are combined to compute similarity between document and cluster. They are intersection between lexical chains of document and cluster and weights of features in this intersection.

$$Sim(D_i, C_k) = Sim(L(D_i), L_k) = \frac{\sum_{f_t \in L(D_i) \&\& f_t \in L_k} w_d(f_t, D_i) + w_c(f_t, C_k)}{\sum_{f_p \in L(D_i)} w_d(f_p, D_i) + \sum_{f_q \in L_k} w_c(f_q, C_k)} \tag{7}$$

$$w_c(f_t, C_k) = \sum_{D_l \in C_k} \sum_{f_i \in L(D_l)} w_d(f_t, D_l) \Big/ |D_l| \tag{8}$$

In the following section, idea of self organizing mapping is imported to optimize cluster partition. From experiments, we can see, after lexical chains are used to get initial clusters, CLCL only needs little running time to get convergence. This is because, initial clusters, which are partitioned by lexical chains, include documents which reflect distinct topics. So, there are few documents which are not correctly partitioned, and it only needs few iterative steps to perform adjustments.

3. Training approach

Previous cluster partition is not exact. There are some documents which are not correctly partitioned. The one reason to this situation is that previous lexical chains are constructed linearly. They may include some features which reflect irrelative meanings. The other reason is that there are some features which have several different meanings. That means they will be included by more than one lexical chain, whereas, previous construction can't satisfy this situation.

In order to solve previous problems, idea of self organizing mapping (SOM) is imported. There are two parameters of SOM (Kohonen, 1997; Alahakoon & Halgamuge, 2000). They are neuron topology and initial neurons. In CLCL, each cluster which is constructed from lexical chain is set as one initial neuron. Assume N_k as the neuron which is constructed from lexical chain L_k. Features in N_k are separated as two parts. They are features which are included by L_k and features which aren't included by L_k. Weights of features which are included by L_k are set as their weights in L_k. Weights of features which aren't included by L_k are set as 0. After previous operations, initial neurons are gotten. In CLCL, neurons are organized as square topology as Figure 1 shows (Andreas et al., 2002). This topology has two layers. The upper layer is competitive layer. It includes neurons. The lower layer is input layer. It includes vectors of input documents.

SOM and its varieties often use the following approach to perform clustering.

[1] Initialize neuron topology of SOM. Evaluate each dimensionality of each neuron vector with a small random value.

[2] Randomly select a datum as input of SOM network. Assume this datum as D_k.

[3] Use formula (1) to compute similarity between D_k and each neuron in neuron structure. The neuron which has the max similarity is the winner.

[4] Use monotonous anneal algorithm to adjust vector of winner neuron and other vectors of neurons which are in the neighborhood of winner neuron.

[5] Estimate whether clustering algorithm achieves convergence condition or not. If achieve, stop running. If not, repeat [2] ~ [5] until it achieves convergence condition.

Traditional SOM algorithms use Euclidean distance as similarity computation method (Kohonen et al., 2000). Only the weights of features which are both included by neuron and document are considered by this method. However, it neglects semantic similarity between documents. That means documents which don't share same features may reflect similar meanings. This is caused by some features which have semantic similarity between them, such as "internet" and "network". If they are respectively included by different documents, these documents may reflect similar meanings.

Semantic similarity is considered in CLCL. When one feature in neuron is adjusted, the features which reflect similar meaning to it are also adjusted. If previous method isn't performed, the documents, which have semantic similarities, may be partitioned into different clusters. This assumption is proved in the following paragraph.

Assume features f_1 and f_2 have semantic similarity, such as "internet" and "network". It is possible to find two initial neurons, such as n_1 and n_2. Weight of f_1 is large in n_1, and weight of f_2 is large in n_2. Assume d_1 and d_2 are two documents which respectively include f_1 and f_2. If documents d_1 and d_2 are selected to adjust neurons in training approach, d_1 will map to n_1 and weight of f_1 is adjusted larger to make d_1 and n_1 more similar. Certainly, d_2 will map to n_2 and weight of f_2 is adjusted larger to make d_2 and n_2 more similar. When algorithm converges, weight of f_1 will be large in n_1, and weight of f_2 will be large in n_2. This situation makes the gap between weights of f_1 and f_2 augment in the same neuron. As literature (Kohonen, 1997) shows, if feature in neuron has larger weight, the documents which are mapped to this neuron will reflect similar meaning to this feature. We know f_1 represents "internet", and f_2 represents "network". So, the documents which map to n_1 reflect similar meaning to "internet" and the documents which map to n_2 reflect similar meaning to "network". Because "internet" and "network" reflect similar meaning, the clusters which are gotten from n_1 and n_2 include the documents which reflect similar meanings. That means previous method partitions similar documents into different clusters.

In order to solve previous problem, weights of similar features are considered in similarity computation and neuron adjustment in training approach. From section 2, we know, lexical chain includes the features which reflect relative meanings. So, in CLCL, the features which are in the same lexical chain are regarded as similar features.

Because pervious method adjusts more features in training approach than traditional SOM algorithms. That will obviously reduce number of adjustment steps to get convergence, and also reduces clustering time. Figure 2 in experiments also proved it.

3.1 Similarity computation

Formula (9) shows how to adjust neuron vector according to input document. This formula considers influences from similar features on similarity computation.

$$Dist(D_i, N_k) = \begin{cases} \sum_{f_t} (W(f_t, D_i) - W(f_t, N_k))^2 & If\ f_t \in D_i; \\ \sum_{f_s} (SSim(f_t, f_s) \times W(f_t, D_i) - W(f_s, N_k))^2 & \\ \qquad If\ f_s \notin D_i \&\& f_t \in D_i \&\& \exists p, f_s, f_t \in L_p; & \\ \sum_{f_s} (\frac{\sum_{t=1}^{m} SSim(f_t, f_s) \times W(f_t, D_i)}{m} - W(f_s, N_k))^2 & \\ \quad If\ f_s \notin D_i \&\& f_t \in D_i (t = 1 \sim m) \&\& \exists p, f_s, f_t \in L_p; & \end{cases} \tag{9}$$

Previous formula has three parts. The first part is the same as Euclidean distance to compute similarity between document and neuron. This part disposes the features which appear in the document, such as feature f_t in the document D_i. The second part disposes the features which don't appear in the document, whereas, there are other features which appear in the document and they have semantic similarity to the features which don't appear in the document. Let's make it clearer. Assume f_s as the feature which doesn't appear in D_i. Assume f_t as the feature which appears in D_i, and it has semantic similarity to f_s. That also means f_t and f_s in the same lexical chain such as L_p. The second part of formula (9) uses semantic similarity between f_t and f_s to import influences from similar features on adjustment. There is another problem in previous adjustment. That is it exists more than one feature which appears in D_i and has semantic similarity to f_s. Assume these features as $f_1, f_2, \ldots f_m$. The third part of formula (9) disposes this situation. This part uses average semantic similarity among features to perform adjustment.

3.2 Neuron adjustment

Formula (10) shows neuron adjustment according to input document. This adjustment is similar to previous similarity computation. It not only considers the features which appear in the document D_i such as f_t, but also considers the features which don't appear in the document D_i but have semantic similarity to f_t such as f_s.

$$\begin{cases} W(f_t, N_k)(t+1) = W(f_t, N_k)(t) + a(t)h(t)(W(f_t, D_i) - W(f_t, N_k)(t)) & \\ \qquad If\ f_t \in D_i; & \\ W(f_s, N_k)(t+1) = W(f_s, N_k)(t) + a(t)h(t)(SSim(f_t, f_s) \times W(f_t, D_i) - W(f_s, N_k)(t)) & \\ \qquad If\ f_s \notin D_i \&\& f_t \in D_i \&\& \exists p, f_s, f_t \in L_p; & \\ W(f_s, N_k)(t+1) = W(f_s, N_k)(t) + a(t)h(t)(\frac{\sum_{t=1}^{m} SSim(f_t, f_s) \times W(f_t, D_i)}{m} - W(f_s, N_k)(t)) & \\ \qquad If\ f_s \notin D_i \&\& f_t \in D_i (t = 1 \sim m) \&\& \exists p, f_s, f_t \in L_p; & \end{cases} \tag{10}$$

Formula (10) also has three parts to adjust weights of features. The first part adjusts weights of features as traditional SOM algorithm. In this formula, $a(t)$ is learning rate, which decreases along with training approach. $h(t)$ is adjusted function, and Gauss function is often used as adjusted function (Kohonen et al., 2000). The second part is used to adjust f_s which doesn't appear in the document D_i but has semantic similarity to f_t. The third part disposes the situation that there is more than one feature which appears in D_i but has semantic similarity to f_s.

4. Experiments and analysis

Ten hundred thousand articles are selected from China Daily of 1998 as the first testing corpus. However, this corpus is too large, and each document from it can't be manually marked with certain cluster index. Precision and recall also can't be computed in this situation. So, a smaller corpus is constructed. Five thousand articles are randomly selected from website Yahoo as the second testing corpus. This corpus is manually classified into thirty classes. They include sports, entertainment, medicine, education, military, and so on.

In this paper, purity is used to test clustering precision on the smaller corpus (Gu et al., 2001).

$$P(S_r) = \frac{1}{n_r} \max_{q=1}^{z} (n_r^q) \tag{11}$$

In this formula, z represents cluster number or neuron number. S_r represents rth cluster after clustering algorithm. n_r represents data number of S_r. We know each datum in the smaller testing corpus is already marked with certain cluster index. Then, C_q is used to represent the cluster which includes the documents that are marked with qth cluster index in testing corpus. n^q is used to represent the number of documents in C_q. n_r^q is used to represent the number of documents, which belong to C_q in testing corpus and belong to S_r after clustering algorithm.

The average purity of different clusters can be used to represent precision of clustering algorithm.

$$Purity = \sum_{r=1}^{z} \frac{n_r}{n} P(S_r) \tag{12}$$

In Table 1, the performance of algorithm which imports semantic similarity in similarity computation and neuron adjustment is tested. In this table, we call the algorithm, which imports semantic similarity in similarity computation and neuron adjustment, as CLCL_S. We call the algorithm, which doesn't import semantic similarity, as CLCL_N. Table 1 shows clustering precision and running time of CLCL_S and CLCL_N on smaller corpus.

From Table 1, we can see, running time and clustering precision of CLCL_S are better than those of CLCL_N. The reason is that the clusters which are constructed from CLCL_S can agglomerate the documents which have semantic similarities into one cluster. This situation will obviously improve precision. Besides, when CLCL_S performs neuron adjustment, more features are adjusted. So, it obviously needs smaller iterative steps to get convergence.

Because the larger testing corpus is too large, we can't manually separate it and then compute precision and recall. So, we regard the clusters which are generated from clustering algorithm as C_1, C_2,…,C_m, and each cluster among them is separated into some sub-clusters by their meanings. Assume the sub-clusters which are separated from C_i as SC_{i1}, SC_{i2},…,SC_{iz}. The sub-cluster which has the max number of documents is used to represent C_i, and this sub-cluster is assumed as SC_{imax}. Regard other sub-clusters are irrelative to C_i. Then, clustering precision on C_i can be represented as: $|SC_{imax}|/|C_i|$. $|C_i|$ is the number of documents in C_i. The average precision on all the clusters is used to represent precision of clustering algorithm.

Running time and clustering precision of CLCL-S on larger testing corpus is respectively compared with those of CLCL-N in Figure 2 and Figure 3.

From Figure 2 and Figure 3, we can see that the method which imports semantic similarity in similarity computation and neuron adjustment can effectively improve clustering precision and reduce running time. When number of documents increases, this situation is more obvious. The reason is that, when number of documents increases, there are more features which have semantic similarities. So, there are more documents which have semantic similarities, and CLCL_S can better cluster them than CLCL_N. Besides, when more features have semantic similarities, CLCL_S needs to adjust more features in neuron adjustment. That will greatly decrease number of iterative steps, and will reduce running time.

Lexical chains which are constructed by linearly scanning feature space are shown in Table 2.

From Table 2, we can see, different lexical chains reflect distinct meanings. So, initial clusters which are constructed from these chains will have clear inter-cluster distinctness. However, there are some irrelative features in lexical chains which will decrease clustering precision, such as "city" in lexical chain 1 and "brand" in lexical chain 3.These features will be adjusted by training approach of CLCL.

Clustering precision and running time of different algorithms on smaller testing corpus are shown in Table 3. They include K-means, FTC, Hierarchical clustering, WEBSOM, ConSOM.

From Table 3, we can see, CLCL has the least running time among all the testing algorithms. This is because CLCL uses semantic similarity to construct lexical chains to partition documents into initial clusters. This partition is close to convergent partition. So, it needs few iterative steps to get convergence. This is one reason why CLCL has the lowest time complexity. In training approach of CLCL, similar features are also adjusted. This method increases number of features which are adjusted in training approach, and will decrease number of iterative steps. This is the other reason why CLCL has the lowest time complexity. Besides, similar features are considered in training approach. That will make cluster agglomerate documents which have semantic similarities, and will increase clustering precision greatly. From Table 3, we also can see, two algorithms based on SOM (WEBSOM and ConSOM) have better performance on

time complexity and clustering precision. This is because, SOM uses neurons to represent clusters, and iteratively adjusts neuron vector to get convergence. Between these two algorithms, ConSOM has better precision. This is because ConSOM imports semantic similarity (Liu et al., 2008). However, this algorithm neglects influences from similar features on neuron adjustment. So, precision of ConSOM is lower than that of CLCL. Because, ConSOM has more steps to compute semantic similarity than WEBSOM, running time of WEBSOM is lower than that of ConSOM. FTC measures overlapping scale between feature sets to compute similarity between different documents (Beil et al., 2002). This computation certainly can't find semantic similarity between different documents. Besides, FTC doesn't perform any optimization on cluster partition. So, precision of FTC is little lower. FTC needs to linearly scan feature space to cluster documents. So, its time complexity is linear with the scale of feature space. In our experiments, running time of FTC is in the middle position among testing clustering algorithms. Running time of K-means is short and it is close to CLCL. This is because K-means has linear time complexity. However, because K-means uses Euclidean distance to compute similarity, precision of K-means is lower. Time complexity and clustering precision of Hierarchical clustering are both worst. This is because it has $O(n^2)$ time complexity. Besides, when cluster is formed, documents which belong to this cluster can't be moved to different clusters.

Running time and clustering precision of CLCL are also compared with those of other clustering algorithms on larger testing corpus. They are shown in Figure 4 and Figure 5.

From Figure 4, we can see, precisions of the algorithms (CLCL, ConSOM, and WEBSOM) are stable, when number of documents increases. This is because these algorithms are based on the idea of self organizing mapping. They map high dimensional space into low dimensional plane, which can avoid the interferences from the problem of "disaster of dimensionality". We also can see that rank of precisions from low to high is WEBSOM, ConSOM and CLCL. This is because semantic similarity is considered in CLCL and ConSOM. So, precisions of them are better than that of WEBSOM. In comparison with ConSOM, CLCL not only imports semantic similarity but also considers influences from similar features in adjustment approach. So, precision of CLCL is better than that of ConSOM. Precision of FTC is not greatly affected by number of documents. This is because FTC uses overlapping scale between feature sets to measure similarity. This method can avoid the problem aroused by high dimension number. Precisions of K-means and Hierarchical clustering drop greatly, when number of documents increase. This is because they don't perform any optimization to cluster documents in high dimensional space. So, when number of documents increases, high dimensional feature space will greatly decrease clustering precision.

From Figure 5, we can see, when number of documents increases, running time of CLCL, ConSOM, and WEBSOM increases stably. This is because they map high dimensional feature space to low dimensional plane. This method can avoid interferences aroused by high dimension number. When number of documents increases, running time of FTC increases greatly. This is because, time complexity of FTC is sensitive to dimension number of feature space. Running time of K-means and Hierarchical clustering also increases greatly, when number of documents increases. This is because they use Vector Space Model (VSM) to organize documents. When number of documents increases, it needs much time to compute similarity by this model.

5. Conclusions

A novel clustering algorithm based on lexical chain (CLCL) is proposed in this paper for large-scale documents. This algorithm first constructs co-occurring word vectors to compute semantic similarity between different features. After computation, lexical chains are constructed to partition documents into some initial clusters. Experiments demonstrate that these initial clusters reflect distinct meanings and need few iterative steps to get convergence. In order to agglomerate documents which have semantic similarities into one cluster, similar features are considered in similarity computation and neuron adjustment. This method not only can improve clustering precision but also can reduce running time. Experiments demonstrate running time and clustering precision of CLCL are both better than those of traditional clustering algorithms.

References

Akiko A. (2004). An information-theoretic perspective of tf–idf measures. *Information Processing and Management*, 39, 45-65.

Alahakoon D, & Halgamuge S. K. (2000). Dynamic self-organizing maps with controlled growth for knowledge discovery. *IEEE Transactions on Neural Networks*, 11, 601-614.

Alam H. Kumar, A. Nakamura, M. Rahman, F. Tarnikova, & Y. Che Wilcox. (2003). Structured and unstructured document summarization: design of a commercial summarizer using Lexical chains. *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, pp. 1147-1152.

Andreas Rauber, Dieter Merkl, & Michael Dittenbach. (2002). The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. *IEEE Transactions on Neural Networks*, 13, 1331-1341.

Arnulfo P. Azcarraga, Teddy N. Yap, Jonathan Tan, & Tat Seng Chua. (2004). Evaluating keyword selection methods for WEBSOM text archives. *IEEE Transactions on Knowledge and Data Engineering*, 16, 380-383.

Chan S. W. (2004). Extraction of salient textual patterns: synergy between lexical cohesion and contextual coherence. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, 34, 205-218.

Congnan Luo, Yanjun Li, & Soon M. Chung. (2009). Text document clustering based on neighbors. *Data and Knowledge Engineering*, 68, 1271-1288.

Florian Beil, Martin Ester, & Xiaowei Xu. (2002). Frequent term-based text clustering. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 436-442.

Frans M. Coetzee. (2005). Correcting the Kullback–Leibler distance for feature selection. *Pattern Recognition Letters*, 26, 1675-1683.

Gonenc E, & Ilyas C. (2007). Using lexical chains for keyword extraction. *Information Processing and Management*, 43, 1705-1714.

Hai-xiang Guo, Ke-jun Zhu, Si-wei Gao, & PTing Liu. (2006). An improved genetic k-means algorithm for optimal clustering. *Proceedings of the Sixth IEEE International Conference on Data Mining*, pp. 793-797.

Jane Morris, & Graeme Hirst. (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17, 21-48.

Jennifer G. Dy, & Carla E. Brodley. (2004). Feature selection for unsupervised learning. *Journal of Machine Learning Research*, 5, 845–889.

Junjie Wu, Hui Xiong, & Jian Chen. (2009). Towards understanding hierarchical clustering: a data distribution perspective. *Neurocomputing*, 72, 2319-2330.

M. Gu, H. Zha, C. Ding, X. He, H. Simon, & J. Xia. (2001). *Spectral relaxation models and structure analysis for k-way graph clustering and bi-clustering*. Technical Report, CSE-01-007, Penn State University.

Min Liu, Weiming Shen, Qi Hao, & Junwei Yan. (2009). A weighted ontology-based semantic similarity algorithm for web service. *Expert Systems with Applications*, 36, 12480-12490.

Rui Xu, & Wunsch D. II. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16, 645-678.

Rishi Jobanputra, & David A. Clausi. (2006). Preserving boundaries for image texture segmentation using grey level co-occurring probabilities. *Pattern Recognition*, 39, 234-245.

Sven M., Jorg L., & Hermann N. (1998). Algorithms for bigram and trigram word clustering. *Speech Communication*, 24, 19-37.

T. kohonen. (1997). *Self-organizing maps*. Berlin: Springer, (Chapter 3).

Teuvo Kohonen, Samuel Kaski, Krista Lagus, Jarkko Salojarvi, Vesa Paatero, & Antti Saarela. (2000). Self organization of a massive document collection. *IEEE Transactions on Neural Networks*, 11, 574-585.

Tuomo Korenius, Jorma Laurikkala, & Martti Juhola. (2007). On principal component analysis, cosine and Euclidean measures in information retrieval. *Information Sciences*, 177, 4893-4905.

X. Z. Wang, Y. D. Wang, & L. J. Wang. (2004). Improving fuzzy c-means clustering based on feature weight learning. *Pattern Recognition Letters*, 25, 1123-1132.

Yuanchao Liu, Xiaolong Wang, & Chong Wu. (2008). ConSOM: A conceptional self-organizing map model for text clustering. *Neurocomputing*, 71, 857-862.

Zheng-Yu Niu, Dong-Hong Ji, & Chew Lim Tan. (2007). Using cluster validation criterion to identify optimal feature subset and cluster number for document clustering. *Information Processing and Management*, 43, 730-739.

Table 1. Clustering time and clustering precision of CLCL_S and CLCL_N on smaller corpus

Methods	CLCL_S	CLCL_N
Time (s)	83	127
Precision (%)	84.13	77.36

Table 2. Lexical chains constructed by linearly scanning feature space

List	Feature chains
1	champion, competition, game, train, match, almightiness, athlete, Olympic game, judgment, score, goal, coach, club, world cup, city, torch, gym, race, trials, contest,
2	economy, price, commodity, company, technique, garden, enterprise, industry, agriculture, labor, product, supply, order goods, trade, market, shop, check,
3	computer, crash, PC, machine, cursor, hypertext, pure text, icon, link, space, brand, byte, internet, off-line, memory, website, notebook, processor, keyboard, mouse,
4	land, earth, location, mountain, sea, coast, port, delta, furrow, bog, highland, shore, seaport, dock, cattle farm, cliff, peak, river bank, grotto, latitude, geography,
5	professor, research, investigate, teacher, doctor, bachelor, student, building, grade, school, certification, education, diploma, suspend classes, dissertation, college,

Table 3. Clustering precision and running time of different algorithms

Methods	K-means	FTC	Hierarchical	WEBSOM	ConSOM	CLCL
Time (s)	85	176	273	126	149	83
Precision (%)	69.51	73.07	68.49	76.41	79.98	84.13

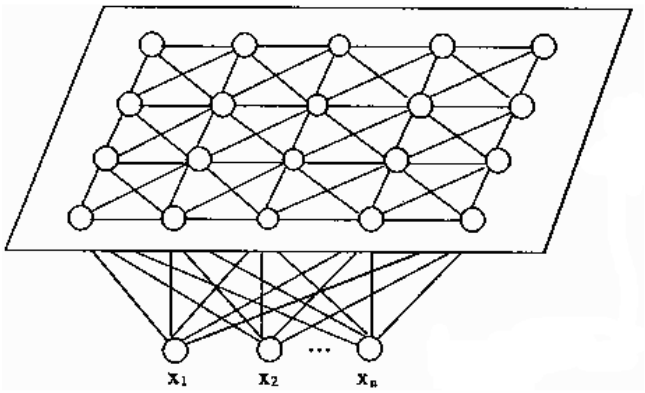


Figure 1. Square neuron topology of SOM

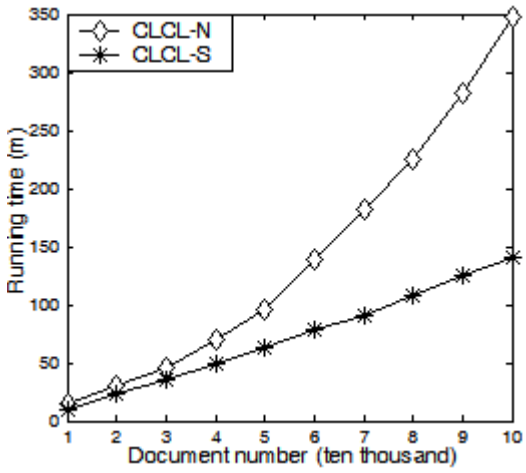


Figure 2. Running time of CLCL_S and CLCL_N

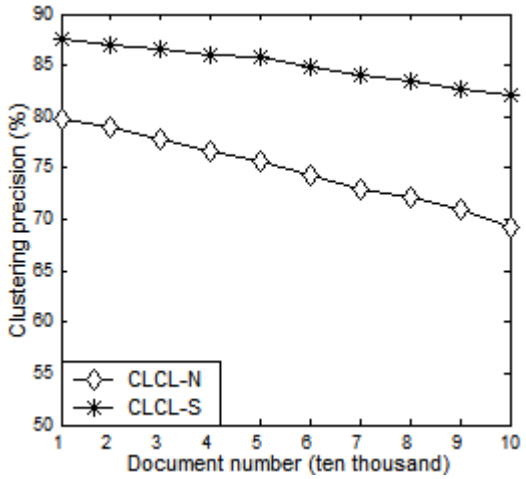


Figure 3. Clustering precisions of CLCL_S and CLCL_N

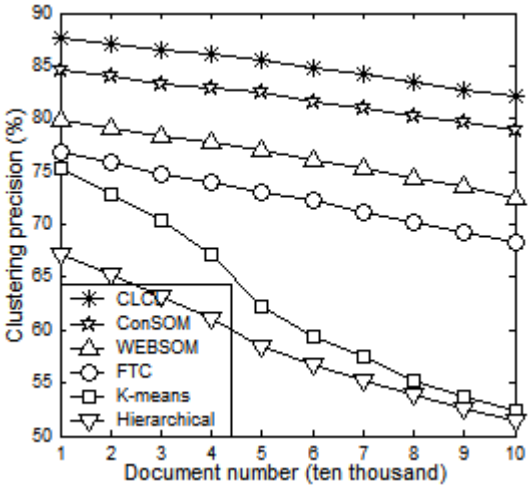


Figure 4. Clustering precisions of different algorithms

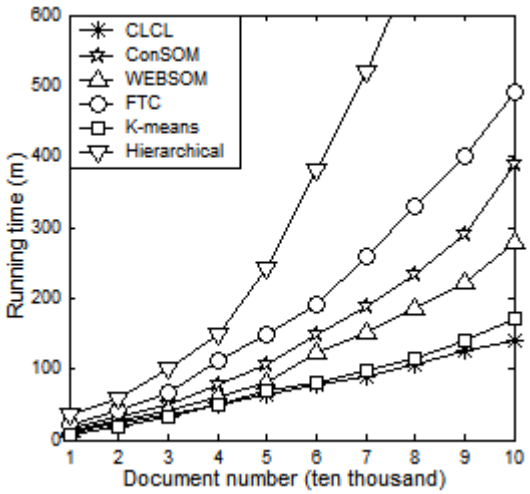


Figure 5. Running time of different algorithms

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Utilizing Usability Model with Multi-agent Technology to Facilitate Knowledge Sharing of Software Maintenance Process Knowledge Environment

Amir Mohamed Talib (Corresponding author)
Faculty of Computer Science & IT, University Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia
Tel: 60-1-7323-3051 E-mail: ganawa53@yahoo.com

Rusli Abdullah
Faculty of Computer Science & IT, Information System Department
University Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia
Tel: 60-1-9315-9820 E-mail: rusli@fsktm.upm.edu.my

The research is a part of the master thesis of the first Author (Amir Mohamed Talib)

Abstract

This paper described a system to manage knowledge generated during the software maintenance process (SMP). Knowledge Management System (KMS) is utilizing to help employees build a shared vision, since the same codification is used and misunderstanding in staff communications may be avoided. The architecture of the system is formed from a set of agent communities each community of practice (CoP). The agents can learn from previous experience and share their knowledge with other agents or communities in a group of multi-agent system (MAS). This paper also described on the theoretical concept and approach of (MAS) technology framework that could be implemented (SMP) in order to facilitate knowledge sharing among the maintainers as well as to demonstrate it into the system wise, on how the (MAS) technology could be utilized in (SMP) system model for serving the maintainer that is developed by using groupware such as Lotus Notes software. This paper applied the definition of (ISO 9241-11, 1998) that examines effectiveness, efficiency, and satisfaction. The emphasis will be given to (SMP) activities that may concern with (MAS) technology that to help the maintainers in order to work collaboratively including critical success factor ensuring that SMP initiatives would be delivered competitive advantage for the (CoP) as well as users of the organization.

Keywords: Multi-agent system, Knowledge Management, Software Maintenance, Lotus Notes, Knowledge Sharing and Usability

1. Introduction

Knowledge comes not only from the expertise of the professionals involved in the process, but it is also intrinsic to the product being maintained, and to the reasons that motivate the maintenance (new requirements, user complains, etc.) and processes, methodologies and tool used in the organization. In addition, during software maintenance (SM) may occur substantial changes? One such example are changes in the maintenance staff (which could mean that the people's expertise changes as well), or the frequency with which each type of maintenance (corrective, perceptive, adaptive or preventive) is carried out. Using a KMS a new knowledge might be produced, thus obtaining the maximum performance from the current information. By reusing information and producing relevant knowledge the high costs of SM could also be decreased (De Looff L, 1990).

KM defines as a discipline that promotes an integrated approach to identifying managing and sharing of all of an enterprise’s information assets. These information assets may include database documents, policies procedures as well

as previously unarticulated expertise and experience resident in individual workers. KM issues include developing, implementing and maintaining the appropriate technical and organizational infrastructure to enable knowledge sharing (GartnerGroup, 2005).

SMP-Based KMS communities contain the software engineers, software developers, workers' knowledge or (system's users) and maintenance engineers in order to facilitate knowledge sharing among CoP. SMP involve many activities in which different people intervene. Each person has partial information that is necessary to other members of the group. If the knowledge only exists in the software engineers and there is no system in charge of transferring the tacit knowledge (contained in the employees) to explicit knowledge (stored on paper, in files, etc) when an employee abandons the organization part of the intellectual capital goes with him/her. When this occurs in an organization involved in SM the end effect is a loss in intellectual capital and increased maintenance effort and costs. Unfortunately, this is often the case. Another well-known issue that complicates the maintenance process is the scarce documentation that exists related to a specific software system, or even if detailed documentation was produced when the original system was developed, it is seldom updated as the system evolves. For example, legacy software from other units often has not documentation which describes the features of the software (De Looff L, 1990).

For an organization that deals with SM, an interesting alternative is to have KMS which stores explicit knowledge and enables the organization to own its intellectual capital and share it with the sub-units. Otherwise, the software developers own this information and the company depends on them. Another advantage of using a KMS is that it reduces the time that a person needs to mature professionally because it favors the professional development of employees and in this way increases the intellectual capital of the organization. With the passing of time, workers acquire knowledge which they do not normally pass on to his/her peers in the same area (let alone to workers in different areas (De Looff L, 1990).

For this reason, different solutions are often used in order to solve the same problem. Using a KMS which acquires workers' knowledge and transmits it, the above mentioned situation would decrease since all workers could benefit from other employees' experience and the organization would increase its expertise and coherence of information.

With a KMS the staff may also be informed about the location of information. It is critical for maintenance engineers to have access to the knowledge the organization has carried out a study which found that the number one barrier to knowledge sharing was "ignorance": the sub-units are ignorant of the knowledge that exists in the organizations, or the sub-units possessing the knowledge are ignorant of the fact that another sub-unit needs such knowledge. Sometimes the organization itself is not aware of the location of the pockets of knowledge or expertise (Orton, J.D., & Weick, K.E, 1990).This fact has been summarized by management practitioners as "the left hand not only does not know what the right hand is doing, but it may not even know there is a right hand" (Szulanski, G, 1994).

KMS also help employees build a shared vision, since the same codification is used and misunderstanding in staff communications may be avoided. Several studies have shown that a shared vision may hold together a loosely coupled system and promote the integration of an entire organization (International Standards Organization ISO 9241-11, 1998).

The above explained issues motivated us to design a KMS for capturing, managing, and disseminating knowledge in a SM organization, thus increasing the workers' expertise, the organization's knowledge and its competitiveness while decreasing the costs of the SMP. To have a shared vision of the maintenance process it is advisable to define a conceptualization of the domain. An explicit specification of such conceptualization is ontology (Hoffer , J. George, J. Valchich, J, 2005).

Normally, programmers will start creating the knowledge that being proposed for a certain project in the organization. When he or she has finished depositing knowledge into the knowledge repositories, the system will trigger the event and pass it to any member specified through e-mail system. This notification will be done based on previous record in order to make sure alerts could be done to those who are interested in the particular knowledge in order to make a decision. Otherwise, this knowledge will be un-meaningful for the other member (Ginsawat, R., Abdullah, R. & Nor, M. Z, 2009). When the system analysts or supervisors also want to make a decision, they should open their mailboxes and look on the subject matter. If they are willing to know about the detail of the knowledge created, they are asked to enter the username and password for security purposes. At the same time another agent will work by updating the status of accessing document as users who are interested with the subject matter (Ginsawat, R., Abdullah, R. & Nor, M. Z, 2009)

CoP is groups of people who share a concern, set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis (Wenger, E, 2002).

The changeable character of the SMP requires that the information generated be controlled, stored, and shared. We proposed in order to manage the knowledge generated during maintenance a MAS formed of three agents are under the client agents implementation. One agent, called the send and receive mail agent, is in charge of organizing the information sent and received from the group. The second agent is scheduler agent and the third agent for the security. The rest of the agents are also communicated, thus enabling them to interchange information. The main goal of this

paper is to design and applying MAS techniques-based KMS in a collaborative environment of lotus notes to facilitate knowledge sharing of SMP among the users of the community of practice. This techniques was inspired by the "there is lack of model of MAS used in SM in order to product the sharing of knowledge in SMP" and also" there is inconsistency of MAS using in test of its functionality"

An experiment is setup setting based on the proposed usability testing model is discussed. The testing is specially designed to verify the significant of the send and receive mail agent, decryption and decryption file agent and file transferring schedule agent and the algorithms used, and to get user satisfaction on the overall system. The success of system is evaluated through user satisfaction survey which covers (i) File sent (ii) File encrypted/ decrypted (iii) File extracted from SMP data warehouse, and (iv) SMP data warehouse technique.

The contributions of the paper are three-fold. First, this research identifies components for agent- techniques-based knowledge sharing system of SMP. Second, finding suitable approach for MAS techniques to be used in knowledge sharing system of SMP. Third, provide reasonable background for applying existing usability testing of MAS techniques-based KMS to facilitate knowledge sharing of SMP and describe whether MAS techniques is significant in knowledge sharing of SMP through user satisfaction experiment.

2. Literature Review

Knowledge about agent concept alone is not sufficient to build a good agent system. There are some fundamental issues needed to drive the design of an agent (Bigus, J. P., Bigus, J., 2001). The first is to view the agents as adding value to a single standalone application, or as a freestanding community of agents that interact with each other and other applications. The first type views the agent from the perspective of application-centric, where the agents are helpers to the application, while the second is more agent-centric, where the agents monitor and drive the application.

In recent years, Multi-Agent System (MAS) has been an active research topic. Due to the difficulties in solving process planning and production scheduling problems using traditional centralized problem solving methodology, MAS approach – a distributed problem-solving paradigm is used as another attempt to solve the planning and scheduling problems. As a distributed problem-solving paradigm, MAS breaks complex problems into small and manageable sub-problems to be solved by individual agents co-operatively (Vermeulen, S. Bohte, D. Somefun &Poutré J. L, 2006).

Agent paradigm lets users think in term of agents rather than objects / functions. The agent exhibits presents high dependencies compared with an object-oriented approach. Such a software application needs an appropriate software development method. An analysis and design methodology is intended to assist first in gaining understanding of a particular system, and secondly in designing it (Wooldridge, M, 2004). There are few choices of agent-oriented methodologies to help software engineers to specify, design and build agents to achieve the system's goals.

(Dignum, V., 2006) proposed Operation per Organizations (OperA), a model for agent's organization, society and interaction model. The Organizational Model implements the desired organizational structure of an agent society, the description of an agent population that will enact the roles described in the structure is detailed in the Social Model, and the specification of agent interactions to achieve the desired society global objectives is described in the Interaction Model. However, this model needs other agent oriented methodology to help designing the system.

(Park, S., Sugumaran, V., (2005) introduced a framework of multi-agent system (MAS) development that considers both functional (services to solve complex problems in distributed environments) and non-functional service (capability to reuse, easy to extend, adapt and process uncertain data) of the system. They also suggested that, in order to develop MAS in a systematic way, system should be analyzed in terms of its ultimate goals and the system should be designed both in the abstract as well as concrete by mapping the goals and the sub-goals to software agents.

(Elst,L. V., Dignum V., & Abecker A., 2004) asserted a three-dimension overview on agent-mediated knowledge management which includes (i) understanding the stage in a system's development process where agents are used (analysis, conceptual design, or implementation) (ii) analyzing the architecture / topology of the agent system, and (iii) identifying KM functionality / application focused on.

MAS developed for job shop scheduling problems in which standard operating procedures are combined with a look-ahead coordination mechanism that should prevent 'decision myopia' on part of the agents. Using their approach, system performance is said to improve in tightly-coupled, real-time job-shop scheduling environments. However, their coordination mechanism is not appropriate for competitive, self-interested agents, which makes it an undesirable choice for coordination in a de-icing setting (Liu and K. P. Sycara, 1996).

Analyzing the structural and representational aspects of large application software, explanation and/or predictions can be made about its maintainability. In that study, maintenance improved significantly when the language used supported (Rombach, H. D, 1987).

Usability testing is a multidimensional construct and can be assessed using various criteria. This paper applies the definition of (Nielsen J, 1993) that examines effectiveness, efficiency, and satisfaction. Usability testing model is to

explain how to identify the information which is necessary to take into account when specifying or evaluating usability testing of a visual display terminal in terms of measures of user performance and satisfaction. Guidance is given on how to describe the context of use of the product (hardware, software or service) and the relevant measures of usability testing in an explicit way. The guidance is given in the form of general principles and techniques, rather than in the form of requirements to use specific methods (Nielsen J, 1993).

The guidance in ISO 9241-11 (1998) can be used in procurement, design, development, evaluation, and communication of information about usability. ISO 9241-11 (1998) includes guidance on how the usability of a product can be specified and evaluated. It applies both to products intended for general application and products being acquired for or being developed within a specific organization (International Standards Organization ISO 9242-11, 1998). ISO 9241-11 (1998) also explains how measures of user performance and satisfaction can be used to measure how any component of a work system affects the whole work system in use. The guidance includes procedures for measuring usability but does not detail all the activities to be undertaken. Specification of detailed user-based methods of measurement is beyond the scope of ISO 9241-11 (1998) (International Standards Organization ISO 9242-11, 1998). According to the benefits and importance of ISO 9241-11 (1998), this paper proposed a testing model for assessing usability of SMP. As reflected in the definition, three central criteria for usability are the effectiveness, efficiency and satisfaction with which users can achieve specified goals (International Standards Organization ISO 9242-11, 1998) as shown in Figure 1 (Note 1).

3. Methodology

As shown in Figure 2 (Note 1) the methodology discusses about the usability testing model used as the methodology to describe that the MAS applied in collaborative environment will help the users according to their needs to support communities in the organization and performance aspects, as well as any other aspects that suggested by the respondents during the survey. Also in these following sections, each criteria of the usability testing model with domino designer, system configuration and others software used in development of the system will be further elaborated in order to achieve its objectives.

Our methodology composed of four main phases as followed:

3.1 Phase 1 - literature review and analysis of MAS based KM in SM

This phase involves evaluation of existing software maintenance processes and activities and literature reviews on academic journals, software maintenance books, and information search on groupware and our Cop tools, MASs applications, usability testing model and other tools supporting software maintenance activities (described in section 2 above).

3.2 Phase 2 - systems design

Systems design is the process or art of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements (Gruber, T, 1995).

System design includes user interface design, designing database, designing form and reports and designing dialogues. In software engineering context, design focuses on four major areas of concern: data, architecture, interface, and components.

In this phase the actual coding or development of the system is done. The system designs are translated into code. Based on the system design, the work is divided in modules/ units. With respect to the type of application, the right programming language is chosen. In this stage, resources such as programming language, database and compiler will work together to produce a system that can function effectively.

User interface design is the design of computers, software application, and websites with the focus of the user's experience and interaction. It's also creates and effective communication medium between a human and computer. An effective user interface design must able to understand how users think and needs. Any user interface contains the Name, Mail, Phone Number, Year of Experience and user Expertise as shown in Figure 5 (Note 1). The author followed the steps above in order to design the all user interfaces design that exist in the system.

Logical and physical database design is the efficient guidelines for design database and file. File and database design occurs in two steps, we begin by developing the logical database model , which describes data using notation that corresponds to a data organization used by data management system. This is the system responsible for storing, retrieving, and protecting data as shown in Figure 6 (Note 1).

3.3 Phase 3 - system development

As shown in Figure 3 (Note 1) the process of sharing knowledge start when a user willing to share its knowledge (file) of SMP application. The user login to his/her email and if the file available with the user then the encryption and decryption agent will encrypt/decrypt the file, and send and receive email will activate directly after that the file will be shared (sent) to the requester user, or if the file exist into SMP warehouse then the file transferring agent will activate

and the encryption and decryption agent will encrypt/decrypt the file and send and receive email will activate then the file will be shared (sent) to the requester user.

The system architecture includes MAS architecture design, steps followed to create the agents, and applied MAS description.

Architecture design is the design process for applying MAS in a collaborative environment and the framework for MAS control and communication. Usually it involves applying major system components and communication.

Figure 4 (Note 1), describes the communication between the agent and the whole system among the users mails and also demonstrate the agents into the system.

SMP requires that the information generated be controlled, stored, and shared. We propose in order to manage the knowledge generated during maintenance a MAS formed of three agents are under the client agents implementation. This part contains the description for the applied MAS and describes the name, job, and the language used. First agent, called the send and receive mail agent, is in charge of organizing the information sent and received from the group. The second agent is encryption and decryption agent and the third agent is file transferring scheduled agent. The send and receive agent communicates with the other two agents. When it receives information, it processes it in order to determinate to what agents should the information be sent to. Since specific information may influence the knowledge managed by different agents, the send and receive agent must know the relationship that exists between all the agents. They also mediate communication among people, and this is of prime importance. In this case, agents will act as a communicator for the user that is based on the direction given and produces the result when it is required to do so. Agents also could be categories in terms of its roles in knowledge searching, knowledge monitoring, and others.

3.3.1 Applied MAS description

The changeable character of the SMP requires that the information generated be controlled, stored, and shared. We propose in order to manage the knowledge generated during maintenance a MAS formed of three agents are under the client agents implementation.

This part contains the description for the applied MAS and describes the name, job, and the language used.

First agent, called the send and receive mail agent, is in charge of organizing the information sent and received from the group. The second agent is encryption and decryption agent, use to protect the SM files that existed into SMP warehouse and the third agent is file transferring scheduled agent, use to extract the file from SMP warehouse and apply it to the user that requested to share it with other users.

The send and receive agent communicates with the other two agents. When it receives information, it processes it in order to determinate to what agents should the information be sent to. Since specific information may influence the knowledge managed by different agents, the send and receive agent must know the relationship that exists between all the agents.

3.4 Phase 4 - system evaluation

3.4.1 Participants

The respondents including System Analyst, System Developer, Software Engineer, and User and will be chosen to fill the questionnaire of this study. The respondents should be applying the system before solving the questionnaire to be situated.

3.4.2 Procedures

In the beginning, the respondents will receive a short, scripted verbal orientation explaining the purpose of the usability testing. Then they will be asked to complete a short background questionnaire to collect their demographic characteristics. The respondents will be asked to perform a set of information about how to share knowledge using the usability test as a kind of multi-agent technology for SMP. The tasks were written on a sheet of paper that included a space where respondents will be asked to indicate their answers. Once the tasks are completed, respondents will be asked to complete a short participant satisfaction questionnaire to collect and test their own perceptions towards SM.

3.4.3 Tasks

Respondents will complete three tasks:

1. They will complete a background/experience questionnaire that *including name, gender, age, education level, Major/Department, and years of experience.*

They will perform tasks using the questionnaire's sheet.

There is also a post-survey questionnaire that specifically examines MAS techniques. After completing a task, the respondents will ask to rank satisfaction and to write down comments.

3.4.4 Data collection

This evaluation model considers both quantifying elements of performance (experience and experiment) as well as subjective empirical. If the answer is wrong, or he/she not familiar with this question then skip to the second question until all the question will be solved. We will, however, record whether respondents are able to complete tasks successfully. The criteria for successful task completion are:

- Participant is able to give a correct answer based on his own information about the system. Any guessed or assumed answers, whether correct or not, are not record as successfully completed tasks.
- Participant is able to give a definite answer to the question. Where respondents indicated they are unsure about the answer or would seek clarification, the task will record as not successfully completed.

3.4.5 Questionnaires

The purpose of the questionnaire is to prove:

- ✓ Handle the interpretation of the term KM and the company's key objective in SM.
- ✓ Handle the aspects that come into play in KM, such as the existence of a strategy, the processes of quality control of data, the content that is being managed, and the functioning of communities of practice.
- ✓ Identify the Multi-Agent technique of willingness of cooperation for research work.
- ✓ Identify the Multi-Agent technique for helping the user according to his needs.

Basically, there are two types of questionnaire that we prepared as part of usability testing for the respondents for the level of the questions is shown in (Appendix B). (Note 4)

3.4.5.1 Pre-Survey questionnaire (background)

A series of questions designed to collect demographic information about the respondents to assess their level of his information about the system is shown in table 1 (Appendix A). (Note 4)

3.4.5.2 Post-Survey questionnaire

After the test subject completed each scenario, he/she should answer a specific questions related to the tasks. To indicate whether the tasks was clear and completed successfully.

After the test subjects complete all the scenarios, he/she will answer thirty six points, eighteen related to KMS and also eighteen related to MAS questionnaires to record user satisfactory.

4. The Integrated Development Environment (IDE) tools of Lotus Notes software

By using the groupware of Lotus Notes (Lotus company, 2007), the best agent technology capability that could be developed is used Java Script programming that comes along with this package. The examples of the IDE and scripting development of interface are shown in Figure 7(a), Figure 7(b), Figure 7(c) and Figure 7(d) as stated (Note 2).

5. Result and Discussions

The result was conducted according to the methodology described in the previous section. It is starts with an overview of data collected by analyzing trends. The Pre-Survey Questionnaire for the respondents shown in (Appendix A). The Post-Survey Questionnaire for both Quantitative and Qualitative MAS questions. Satisfaction is a multi-dimensional construct. This study applies MAS technology to support knowledge sharing of SMP. (Note 4)

5.1 Usability testing for the system

Several types of data were collected to assess user's performance and user's perceptions of negotiating the system. In addition, we examined selected features of the normal lotus notes system to determine their effectiveness. Tasks were deemed to be either completed or not completed.

5.1.1 Effectiveness

As we can notice all respondents were able to complete all the tasks. Effectiveness was measured by the number of tasks successfully completed. For other tasks respondents were able to complete in success percent ranging between (50% & 100%). The Successful task completion for the individual tasks is summarized in Figure 8 and Figure 9 below for the respondents. As we can see the average of successfully completion task are high, according to this results the successfully completion task that presented the effectiveness, achieved correctly. Moreover these results of the tasks successfully completed are high.

Figure 8 (Note 3) shows that Q2, Q6, Q8 and Q9 are completed answered successfully (100%) and the others questions got more than 70%.

Figure 9 (Note 3) shows that Q1, Q5 and Q9 are completed answered successfully (100%) and the others questions got more than 60%.

5.1.2 Efficiency

Completion time is the one factor used for measuring efficiency in this paper. Efficiency was measured by the amount of time taken to complete all tasks. An average of 44 minutes and 3.5 seconds per respondent was taken to complete the all tasks. However, there was much variation among the respondents, for example, the fastest respondent took only 18 minutes and the slowest took 37 minutes and 9 seconds which are about three times longer. Pearson's product-moment correlation analysis was conducted to see if the respondents' completion time is related. The results showed that total completion time is independent, (see Figure 10 and Figure 11) (Note 3).

As a result, efficiency was measured by evaluating completion time used in this survey by each respondent. Respondents who they were familiar with the systems in general tended to use less time to complete their tasks. When the respondent knows how to get the answer, it takes them fewer time while when they don't know how to use the system, they take more time.

5.1.3 Satisfaction

Respondents Satisfaction measured by using the two scales (YES= respondent agreed, NO= respondent not agreed). Satisfaction was measured by a rating scale for several satisfaction elements.

According to the result below, the satisfaction for the respondents were in moderate level.

Figure 12 (Note 3) shows the respondent satisfaction for Quantitative analysis for MAS. The Q2 and Q8 give high satisfaction (Mean=5.10) out of (Maximum =6) and Q6 give high satisfaction (Mean=4.70), and Q7 also give high satisfaction (Mean=4.40) out of (Maximum=5). All respondents feel satisfied with the system when they fail to perform the task correctly.

Figure 13 (Note 3) shows the respondent satisfaction for Qualitative analysis for MAS. Q8 give high satisfaction (Mean=5.40) and Q9 give high satisfaction (Mean=5.50) out of (Maximum =6) and Q7 give high satisfaction (Mean=4.20) out of (Maximum=5). 30% of respondents feel less satisfied with the system when they fail to perform the task correctly.

Groupware software has been used which is Lotus Notes (Lotus company, 2007). This software provides various types of services for a CoP. The most important service offers by this software or product is particularly in term of SMP warehouse. This service will serve the CoP to share their knowledge and any things that are stored in SMP warehouse. Another common service is agent technology, which it allows people to share the knowledge or notification regarding the latest knowledge of SM in the SMP warehouse at any time and any place. The set of the agents are communicated to show how the system running as shown in section 4 (system development). Usability testing model cafeterias (effectiveness, efficiency & satisfaction) are supporting this idea based to the (Appendix B). (Note 4)

6. Conclusion

SM is one of the most important stages of the software life cycle. This process takes a lot of time and effort. Besides, it generates a huge amount of different kinds of knowledge that must be suitably managed. MAS in charge of managing this knowledge might improve the maintenance process since agents would help developers find information and solutions to problems and to make decisions, thus increasing organization's competitiveness. KMS is a good place where people could share their knowledge between the CoP. In this case, agent's technology is a tool that could be used in order to act on behalf of CoP of SM to do something repetitively and time based system especially in maintaining various types of maintenance such as Adaptive, Perfective, Corrective and Preventive. The agent techniques describes send and receive agent use to enable the user to share their knowledge among their emails and file transferring scheduled agent use to extract the file from the SMP warehouse and encryption/decryption agent use as the security agent to protect the file. Usability testing model use the three main criteria effectiveness was measured by the number of tasks successfully completed, efficiency was measured by amount of time taken to complete the tasks, and satisfaction was measured by a rating scale for several satisfaction elements.

References

Bigus, J. P., Bigus, J. (2001). "Constructing Intelligent Agents Using Java". Second Edition 2001. Wiley Computer Publishing, John Wiley & Sons, Inc.

De Looff, L. (1990). "Information Systems Outsourcing Decision Making": *a Managerial Approach. Hershey, PA: Idea Group Publishing.*

Dignum, V. (2006). "A Model for Organizational Interaction: Based on Agents, Founded in Logic". PHD Thesis, Utrecht University

Elst,L. V., Dignum V., & Abecker A. (2004). "Towards Agent-Mediated Knowledge Management". Achmea & University Utrecht

GartnerGroup. (2005). "Available at: http://www.gartner.com". Accessed on 23/10/08.

Ginsawat, R., Abdullah, R. & Nor, M. Z. (2009). "Applying Knowledge Management System Architecture in Software Maintenance Environment" *Journal of Computer and Information Science.* Vol. 2, No. 4

Gruber, T. (1995). "Twards Principle for the design of ontologies used for Knowledge Sharing". *International Journal of Human-Computer Studies*, 43(5/6). pp 907-928.

Hoffer , J., George, J., & Valchich, J. (2005)."*Modern Systems Analysis and Design*". Fourth edition. Upper Saddle River, NJ: Prentice Hall. Design, pp 327- 436.

International Standards Organization. (1998). ISO document 9241- Ergonomic requirements for office work with visual display terminals.

Liu & Sycara. K. P. (1996). "Multiagent coordination in tightly coupled task scheduling". Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pages 164-171.

Lotus company. (Access on August 2007). available at:www.lotus.com

Nielsen J. (1993). "Usability Engineering", Academic Press, London.

Orton, J.D., & Weick, K.E. (1990). "Loosely coupled systems": *A reconceptualization. Academy of Management Review, 15-2*

Park, S., & Sugumaran, V. (2005). "Designing Multi-agent Systems: A Framework and Application". Expert System with Applications 28 259-271. Elsevier Ltd.

Rombach, H. D. (1987). "A Controlled Experiment on the Impact of Software Structure on Maintainability", IEEE Trans. Software Eng. 13, 344-355.

Szulanski, G. (1994). "Intra-Firm Transfer of Best Practices Project" American Productivity and Quality Centre, Houston, Texas

Vermeulen, S. Bohte, D., Somefun, & Poutré J. L. (2006). "Improving patient activity schedules by multi-agent Pareto appointment exchanging".

Wenger, E. (2002). "Communities of Practice: learning, meaning and identity" New York: Cambridge University Press.

Wooldridge, M. (2004). "An Introduction to Multi-Agent Systems" John Wiley & Sons Ltd. (Chapter 2)

Notes

Note 1

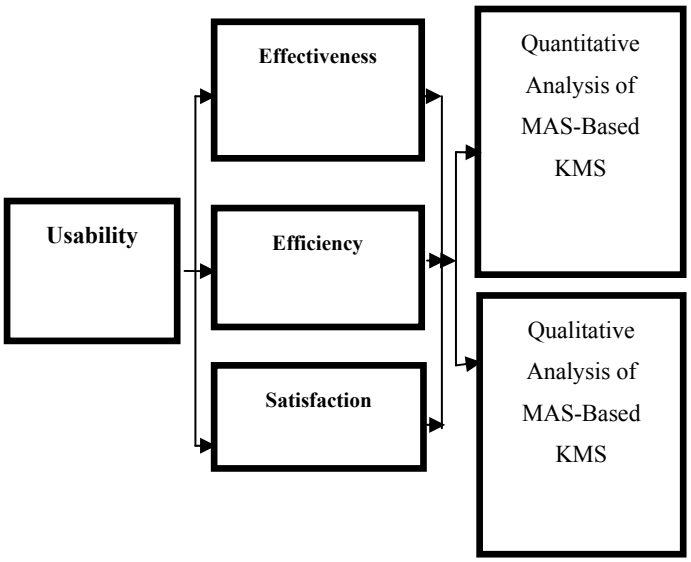


Figure 1. Utilized Usability Model

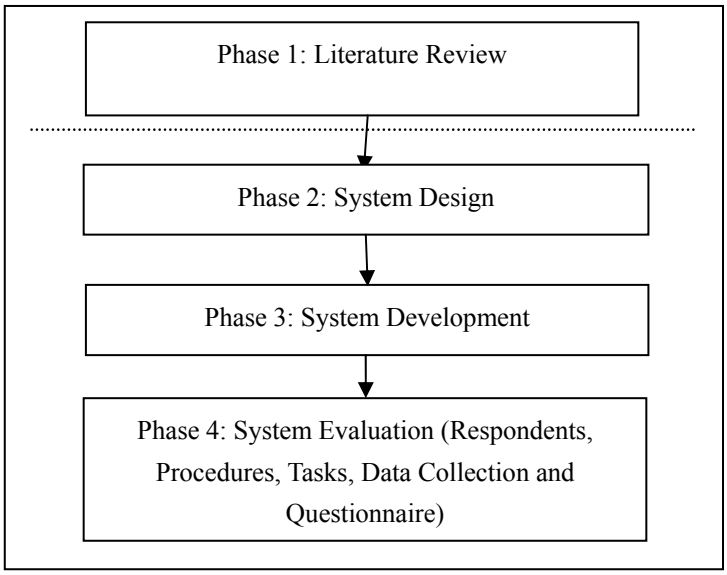


Figure 2. Methodology Diagram

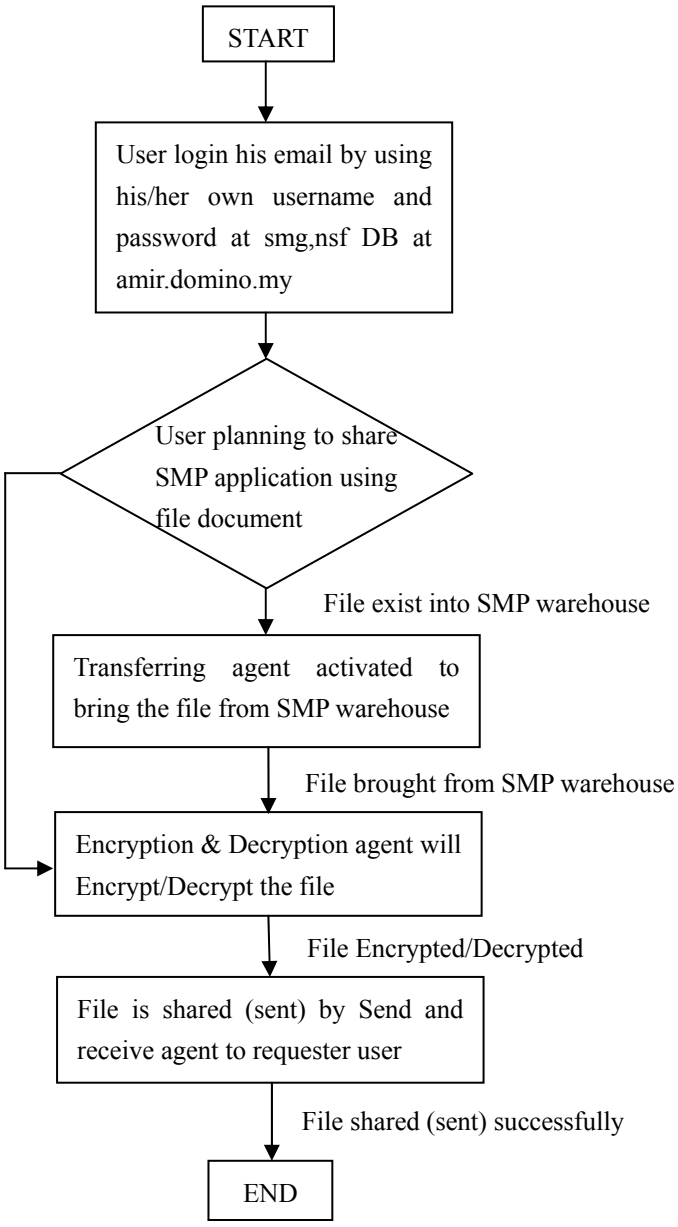


Figure 3. System Flowchart

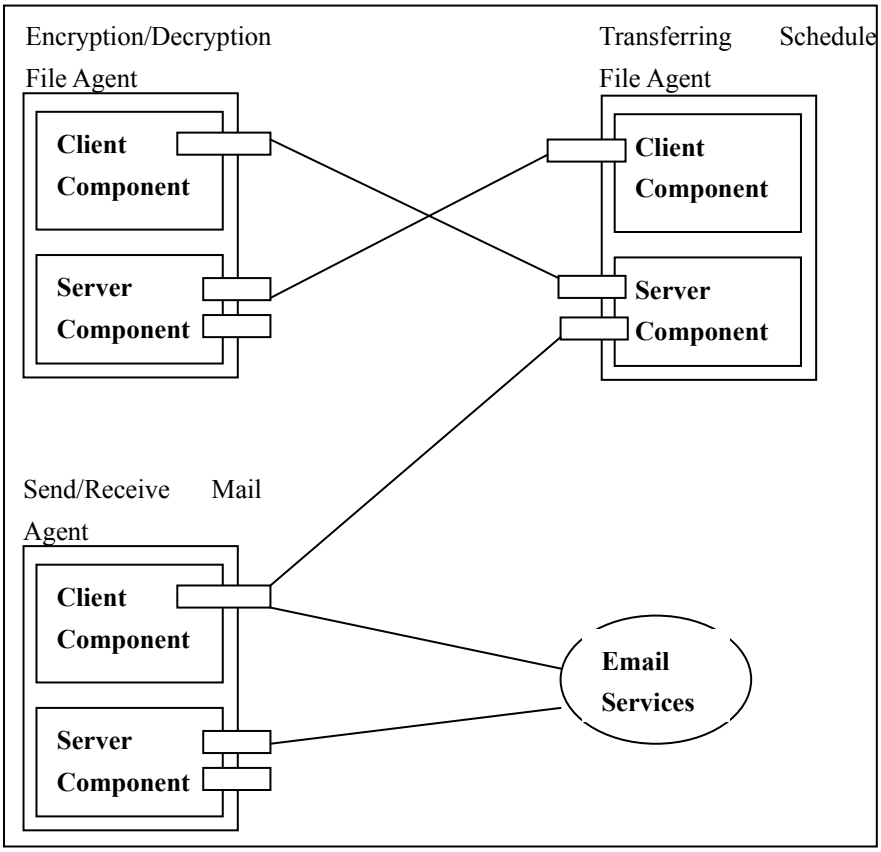


Figure 4. MAS Architecture

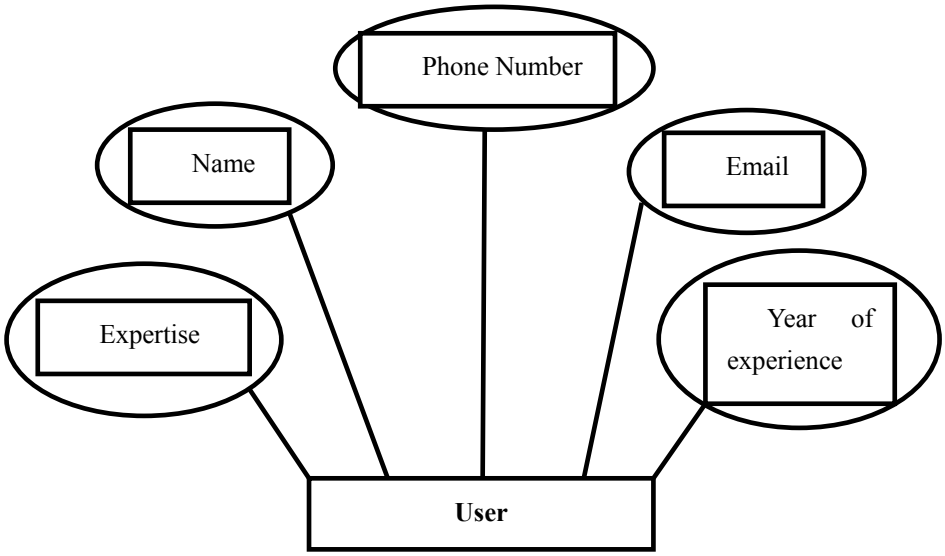


Figure 5. User Interface Design

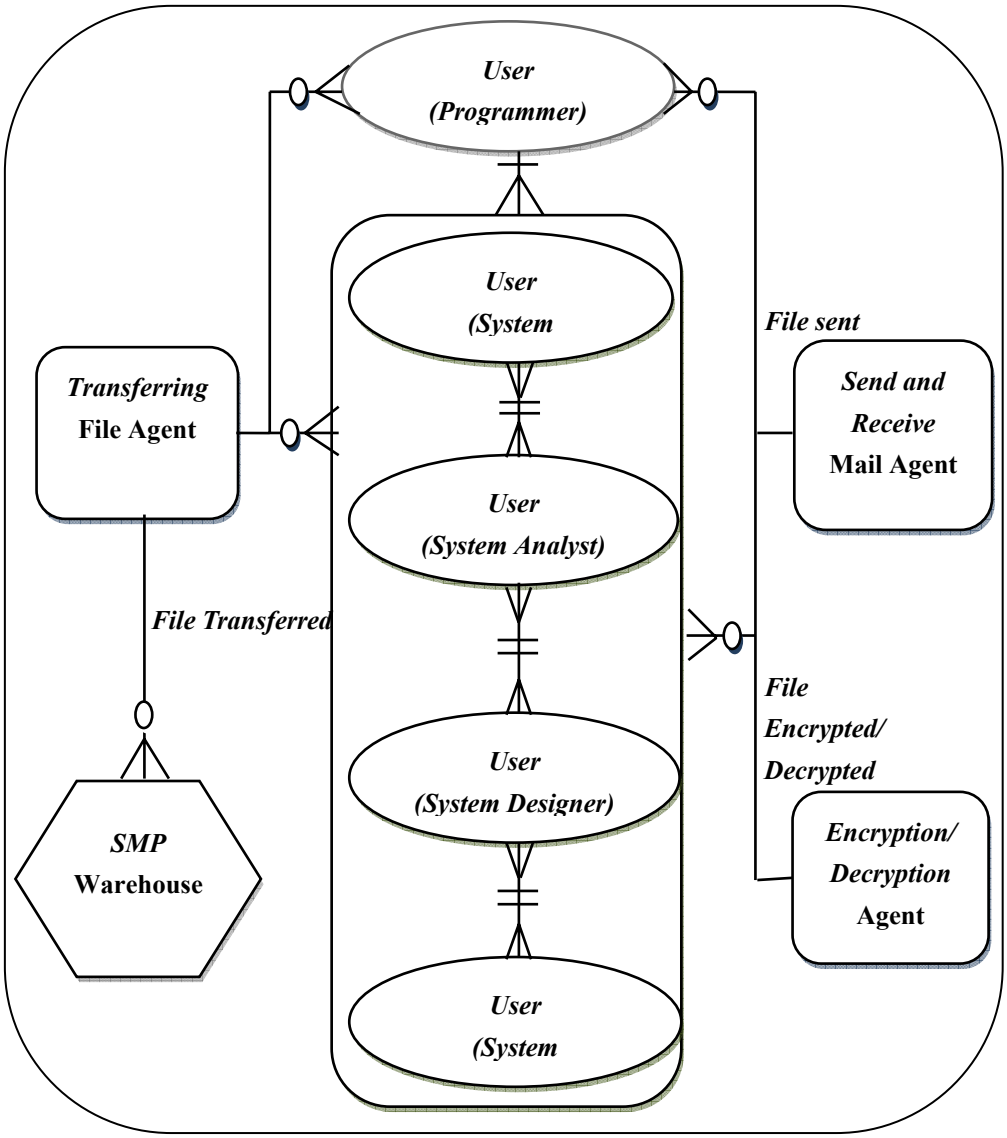


Figure 6. Database Design

Note 2

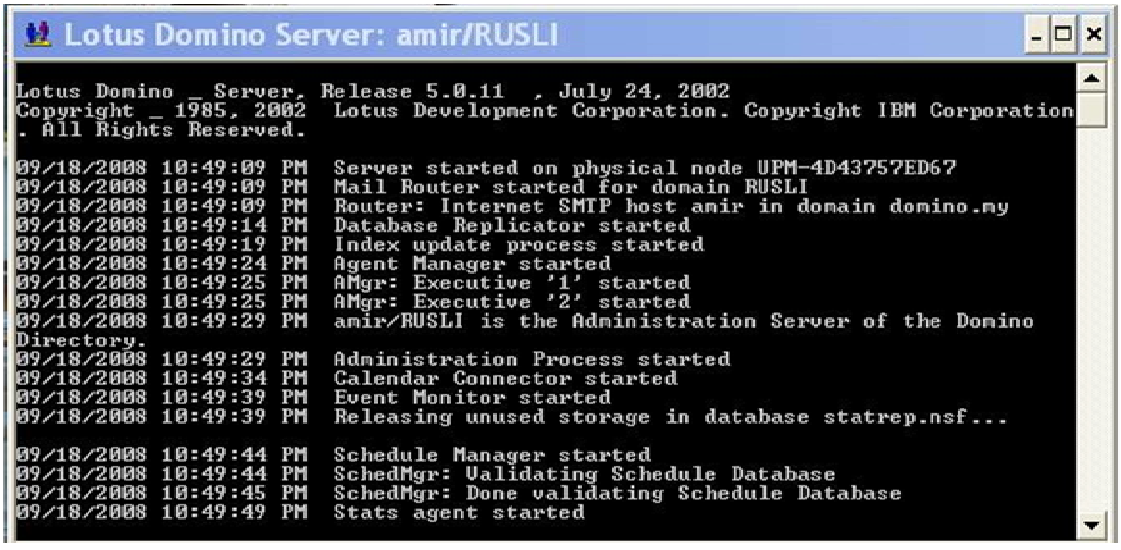


Figure 7 (a). Lotus Domino Server amir/RUSLI

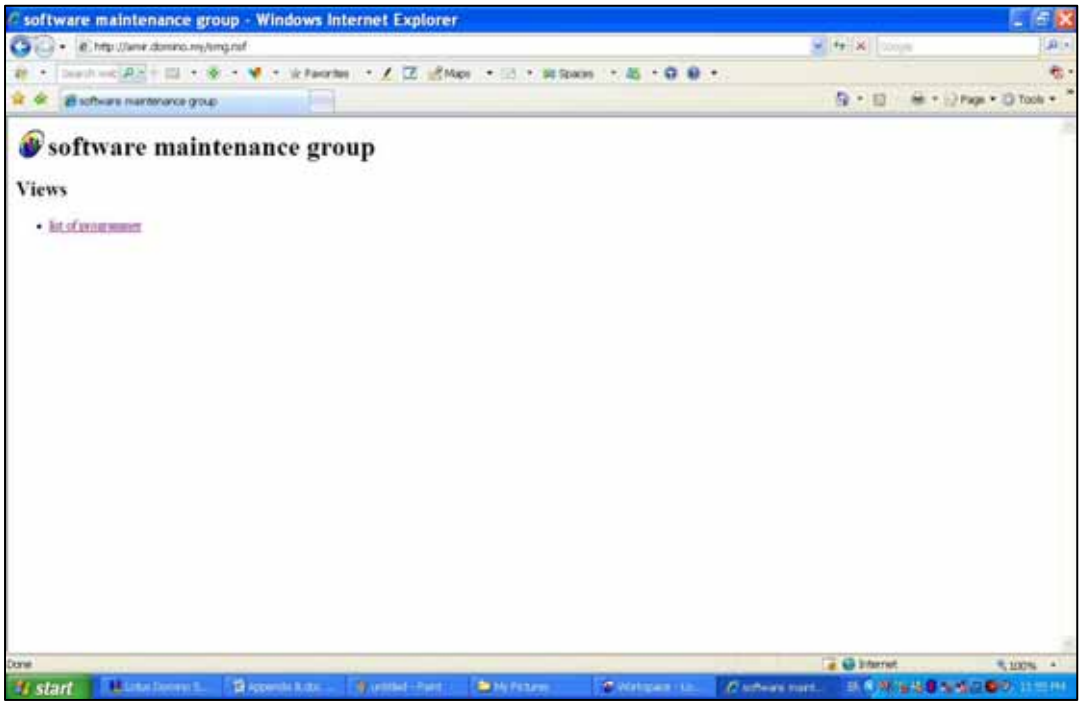


Figure 7 (b). Software Maintenance Group

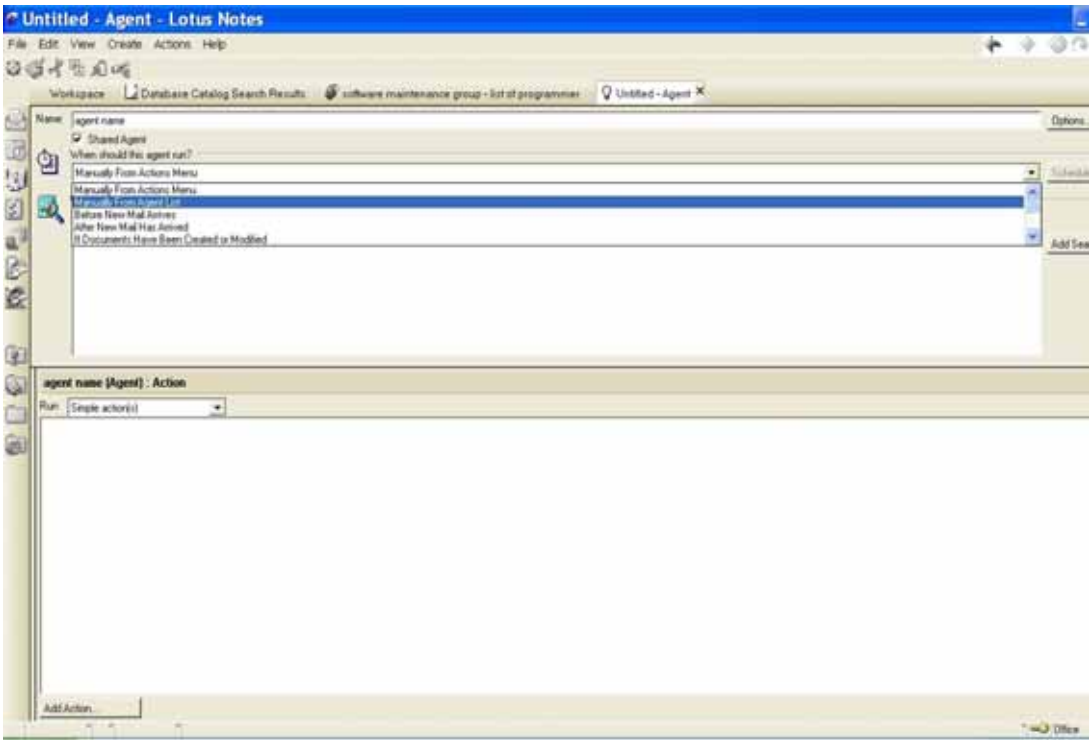


Figure 7 (c). Selecting trigger to run the agent

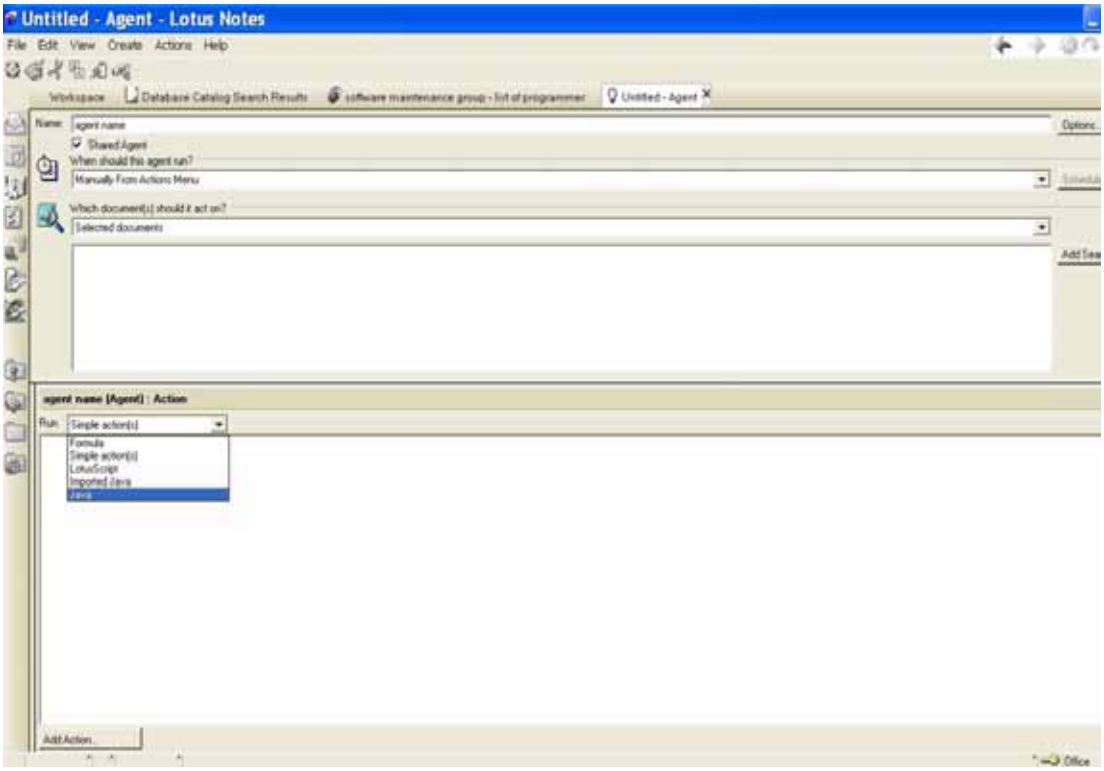


Figure 7 (d). Selecting java to write the agent code

Note 3

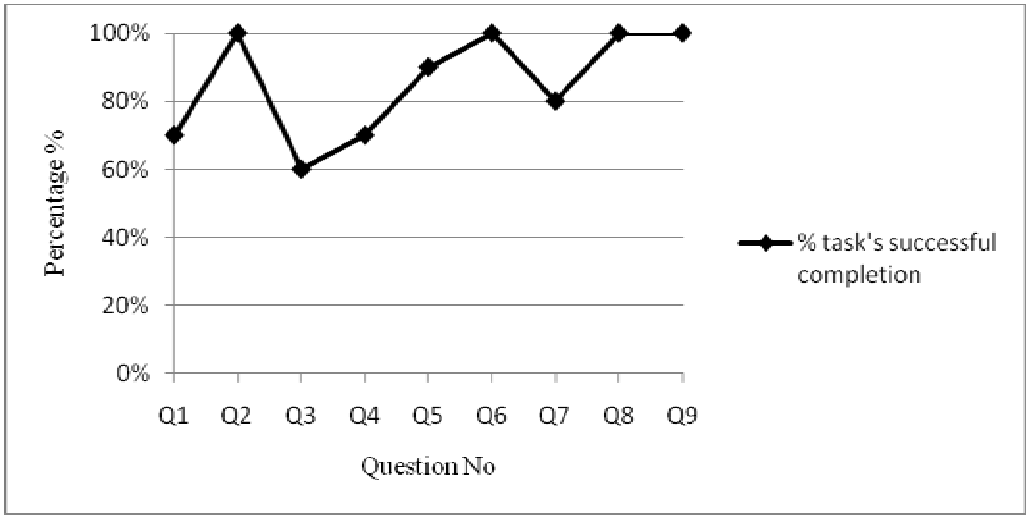


Figure 8. Completed Tasks Successfully for Quantitative analysis for MAS

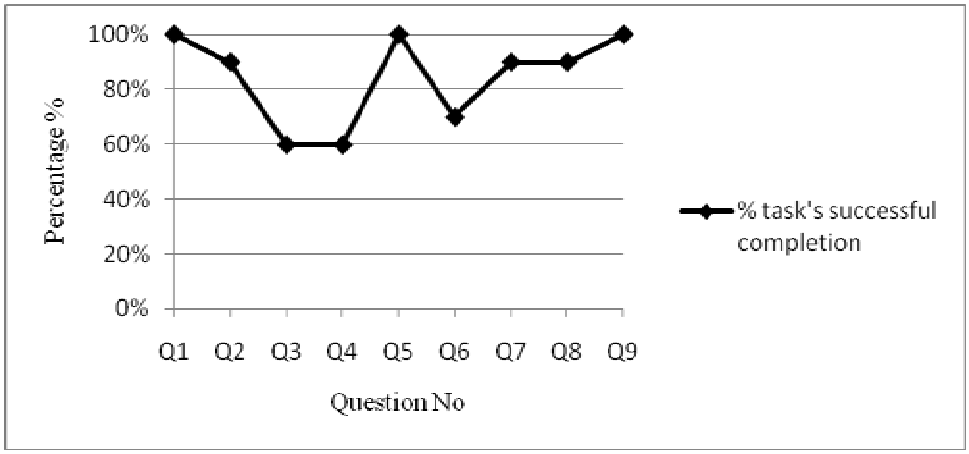


Figure 9. Completed Tasks Successfully for Qualitative analysis for MAS

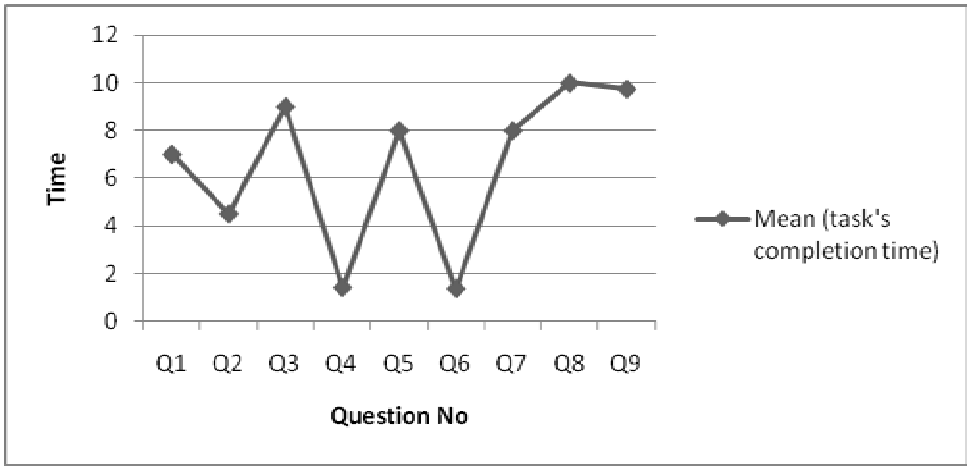


Figure 10. Time used to completed for Quantitative analysis for MAS

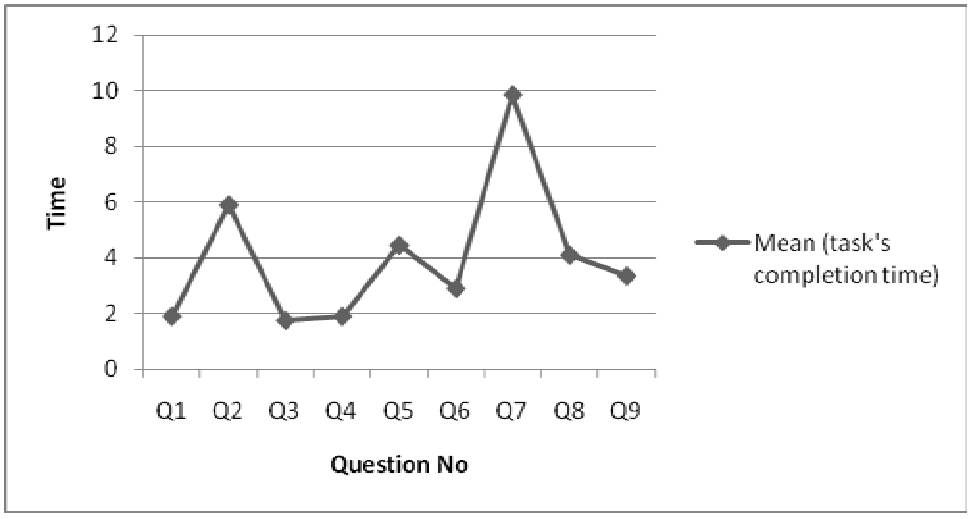


Figure 11. Time used to completed for Qualitative analysis for MAS

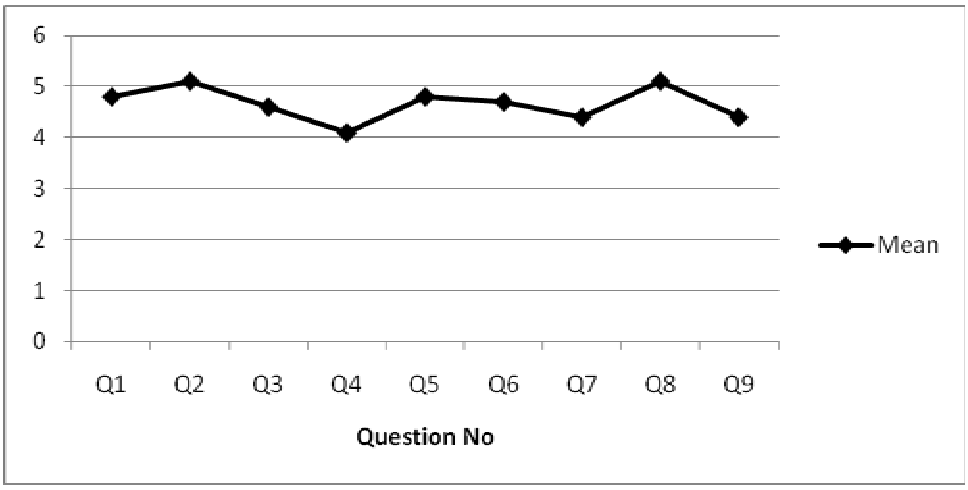


Figure 12. Respondent satisfaction for Quantitative analysis for MAS

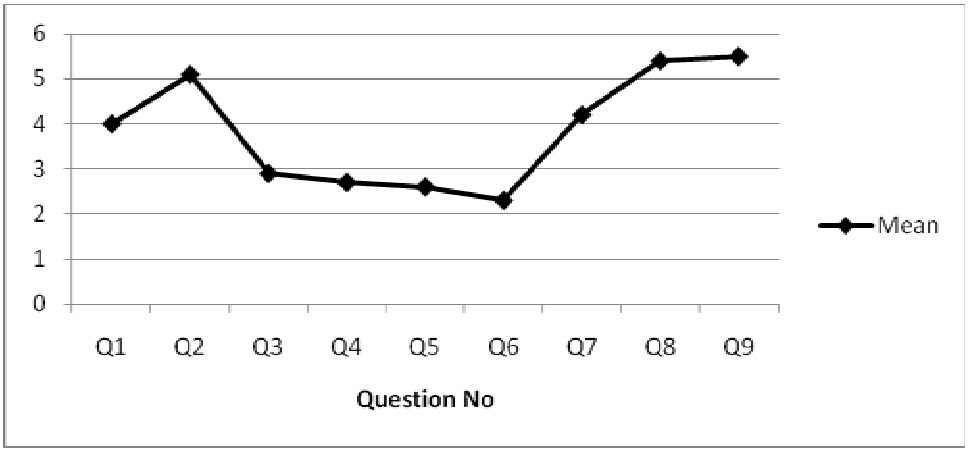


Figure 13. Respondent satisfaction for Qualitative analysis for MAS

Note 4

APPENDIX A

Respondent demographics

Ten individuals participated in this study. After collecting the background questionnaire we calculate the demographic characteristic of the respondents.

Table A.1: Sample representative of the respondents

		Sample N=10	**Percentage %**
Gender	Male	8	80%
	Female	2	20%
Education	Degree (Bachelor, Diploma)	5	50%
	Postgraduate (Master, PhD)	5	50%
Age group	20 to 29	2	20%
	30 to 39	5	50%
	40 to 49	3	30%
	50 to 59	1	10%
System Usage	Daily	4	40%
	From time to time	6	60%
System Shown	First time	6	60%
	Familiar	4	40%

Table A.1 summarized the basic characteristics of the initial sample as well as those of the respondent. As can be noted from the table above, the sample is rather skewed towards males in the age group between 23 and 40 years old and of higher education. Most participants were experienced and familiar with lotus notes software: forty percent (40%) used the system on a daily basis and were familiar with the lotus notes system and its applications; moreover sixty percent (60%) have shown and described the system for them.

APPENDEX B

QUESTIONNAIRE SHEET

Appendix B.1 Pre-Survey Questionnaire

Thank you very much for agreeing to participate in this experiment. All of your personal data that we collect will be entirely confidential. I would like to gather a bit of background information about u.

Participant Name_____

Gender: _____Male _____Female

Date_____

How old are you? 20-29 30-39 40-49 50-59 60 or above

Level of education:

_____Certification Bachelor _____ Certification Diploma

_____Degree Postgraduate

Race: _____Malaysian (Local) _____International

Years of Experience _____

Appendix B.2 Usability Testing Questions

The goal of this Survey to evaluate the KMS by using Usability testing questions and prove the KMS is a useful support system.

I will ask you a series of questions and would like you to think out loud while you look for the answer. Please remember that we are testing the effectiveness of the KM and this is not a test of you. The whole test should take less than one hour. Thank you

Description for How to Answer the Question:

Evaluation of the matrix: Assign yourself the following points for each

NA = 0, where 0 is doing nothing at all = NONE and

1 = Don't Know, Not Sure or Can't Say = NO

2 = Not Important or as Not been Addressed = MINIMALLY

3 = Partially Beneficial or somewhat Effective or Less Scope for Overall Improvement =

PARTIALLY

4 = Important or May not be effective but other associated necessary actions being taken =SUBSTANTIALLY

5 = Critical or already in place and effective = FULLY

Also, the scale can generally be summarized as follows for majority situations

'NA 1 2 3 4 5' is calibrated as in

'5 (Always) 4 (Often) 3 (Sometimes) 2 (Occasionally) 1 (Never)'

NA (Not Applicable), (Note: "NA" and "1" scale values are equivalent.)

QUESTIONNAIRE - Part One (Quantitative Analysis)

1. Is recording and sharing knowledge a routine and like any other daily habits for the employees?

NA 1 2 3 4 5

2. Are the employees co-operative and helpful when asked for some information or advice?

NA 1 2 3 4 5

3. Is Knowledge sharing seen as strength and knowledge hoarding as a weakness?

NA 1 2 3 4 5

4. Is good knowledge management behavior like sharing, reusing knowledge actively promoted on a day-to-day basis?

NA 1 2 3 4 5

5. Are people in the organization aware of the need to proactively manage knowledge assets?

NA 1 2 3 4 5

6. Do people at all levels in the organization participate in some kind of a community or communities of practice?

NA 1 2 3 4 5

7. Is there top management representation for KM?

NA 1 2 3 4 5

8. Is knowledge management a formal function area in the organization?

NA 1 2 3 4 5

9. Are the teams in the organization effective? Are self managed teams composed of individuals capable of learning from each other?

NA 1 2 3 4 5

QUESTIONNAIRE- Part Two (Qualitative Analysis)

1. Do the employees share their knowledge?

Yes No

2. Is the intranet used to share knowledge in an informal manner (non-routine, personal and unstructured way)?

Yes No

3. Do workplace settings and format of meetings encourage informal knowledge exchange?

Yes No

4. Are there incentives given for knowledge contribution, exchange or on knowledge sharing in your firm?

Yes No

5. Is the support from executive management to KM (Knowledge Management)\ knowledge sharing VISIBLE?

Yes No

6. Are there specific knowledge roles identified and assigned?

Yes No

7. Are all senior managers and professionals trained in knowledge management techniques?

Yes No

8. Is knowledge validated through peer or superior review or, is there some kinds of librarians or information management staff that coordinate knowledge repositories.

Yes No

9. Is knowledge sharing across departmental boundaries actively encouraged? (Not similar to ‘’incentives’’)

Yes No

Appendix B.3 Post-Survey Questionnaire

Thanks again for participating in this experiment. This questionnaire gives you an opportunity to tell us your reactions to the system you used. Please circle a number on the scale to indicate your reactions. Thank you

The goal of this part to evaluate the MAS that applying into the Lotus Notes Domino and to prove the MAS will help the users according to their needs.

QUESTIONNAIRE - Part One (Quantitative Analysis)

1. Is it possible to change the send and receive agent schedule.

NA 1 2 3 4 5

2. We can run the send and receive agent "After new mail arrives" and "Before new mail arrives".

NA 1 2 3 4 5

3. Send and receive agent option will appear in the current mail file.

NA 1 2 3 4 5

4. One of our users left the office without enabling the send and receive agent. We can enable it for him or her.

NA 1 2 3 4 5

5. I sent to someone multiple e-mails while that person is out of the office. So I will receive only one e-mail notification.

NA 1 2 3 4 5

6. To customize the "Welcome Back" message, the "Disable Reminder" message, or the default wording of the e-mail notifications sent to all senders of e-mail.

NA 1 2 3 4 5

7. In order to notice the Domino Designer 5 client has new agent properties, such as "Allow user activation" and "Run on behalf of." The both of these we need to set in the mail template (on the server) or in the individuals' mail files for the send and receive agent to work properly.

NA 1 2 3 4 5

8. The send and receive agent work in a clustered environment.

NA 1 2 3 4 5

9. We can enable the scheduler agent for leaving "Today" instead of the recommended "Tomorrow" or another date in the future.

NA 1 2 3 4 5

QUESTIONNAIRE- Part Two (Qualitative Analysis)

1. We can set the scheduler agent for an absence period of a half day or a few hours.

Yes No

2. Whenever we receive a warning in Designer while attempting to save an agent
"You do not have execution access privileges for this agent on server ". This indicates one of two things: either the agent signer does not have the rights on the scheduled server, or that server is not reachable to check the signer rights. Running agent "test" in the Designer will give you a better indication.

Yes No

3. "Do you know why I get 'Object variable not set'?" This is a result of a logic error in the code. The problem should become clear if you single step through the code in debugger (File - Tools –Lotus Script debugging). Server might be configured to delay execution of your agents.

Yes No

4. If these tips don't help you figure it out on your own, when you post in the forum please include in your post screen shot of server log output with agent manager debug flags set to '*' (best) and/or diagnostic output of "agent test" (a good second choice when you don't have access to the server log).

Yes No

5. It is possible to pass parameters between agents.

Yes No

6. It is easy to sign an agent with a server.id For Lotus Notes 5.

Yes No

7. It is easy to console commands from send and receive and scheduler agent.

Yes No

8. Agents runs but mail is not being sent. If our agent runs to completion (i.e. no run time errors that stop the agent before it gets to the send logic) this symptom usually means that it is configuration issue, not an agent problem.

Yes No

9.Does the agents that applied will help the users of the system?

Yes No

Comment about the system:

_____
_____
_____.

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Comparative Research on Particle Swarm Optimization and Genetic Algorithm

Zhijie Li, Xiangdong Liu & Xiaodong Duan
Institute of Nonlinear Information Technology, Dalian Nationalities University
Dalian 116600, China
Tel: 86-411-8123-2976 E-mail: lizhijie@dlnu.edu.cn

Feixue Huang
Department of Economics, Dalian University of Technology, Dalian 116024, China

The research is financed by the National Natural Science Foundation of China (No. 60573124) and the Ph. D. Foundation of Dalian Nationalities University (No. 20086205). (Sponsoring information)

Abstract

Genetic algorithm (GA) is a kind of method to simulate the natural evolvement process to search the optimal solution, and the algorithm can be evolved by four operations including coding, selecting, crossing and variation. The particle swarm optimization (PSO) is a kind of optimization tool based on iteration, and the particle has not only global searching ability, but also memory ability, and it can be convergent directionally. By analyzing and comparing two kinds of important swarm intelligent algorithm, the selecting operation in GA has the character of directivity, and the comparison experiment of two kinds of algorithm is designed in the article, and the simulation result shows that the GA has strong ability of global searching, and the convergence speed of PSO is very quick without too many parameters, and could achieve good global searching ability.

Keywords: Evolvement computing, Genetic algorithm (GA), Particle swarm optimization (PSO)

1. Introduction

Generally, single natural biology is not intelligent, but the whole biotic swarm can have the ability to treat complex problems. The swarm intelligence (Kennedy, 2001 & Peng, 2003, P.1982-1988 & Jiang, 2009, P.26-28 & Tarasewich P, 2002, P.62-67) is a method which simulates the natural biotic swarm to realize artificial intelligence, and it emphasizes the simple character of individual behavior, the avalanche character of swarm, and the research strategy from bottom to top. The swarm intelligence has shown better optimization in the existed application domains, so the extensive attentions have been produced in relative domains. At present, the research about swarm intelligence is still in the initial stage, so it has large development potential. Some swarm intelligent algorithms such as GA and PSO all belong to the category of evolvement computation, so they have many similar aspects in principle mechanism. In this article, GA and PSO will be deeply studied, and the principal mechanism, flow implementation and optimization result of these two algorithms will be analyzed and compared in detail. GA (Ge, 2008, P.2911-2916 & Zhang, 2003) was proposed by professor Holland of American Michigan University in 1975 and gradually developed based on Darwin' theory of evoluation. This method is a kind of random optimization algorithm to simulate natural selection and genetic mechanism, and it only needs the objective function to be the optimization information, and find the optimal individual by the continual selection, crossing, and aberrance of initial swarms, and it has strong global optimization ability and the character to treat discrete variables. But GA has low computation efficiency and slow optimization speed. PSO was proposed by Dr. Kennedy and Eberhart in 1995 under the elicitation of artificial life research result when they simulated the migration and swarm behaviors in simulation of bird swarm, and it can optimize many complex problems such as nonlinear problems, non-differentiable problems and multiple-peak problems (Kennedy, 1995, P.1942-1948 & Parsopoulos, 2002, P.235-306 & Duan, 2005 & Grosan C, 2005, P.623-632). The individual of particle swarm presents the possible solution of a problem. Each particle has two description variables including position and speed, and the corresponding objective function value of the particle position coordinate can be regarded as the fitness degree of this particle. PSO measures the quality of particle by the fitness degree, so one significant character is that its parameter quantity is less, but the setup of these key parameters will obviously influence the precision and efficiency of the

algorithm. Comparing with GA, the information sharing mechanism of PSO is very different. In GA, chromosomes share information with each other, so the transfer of the whole swarm evenly moves to the optimal region. In PSO, only the optimal particle will transfer the information to other particles, which is single-directional information flow. The whole searching updating process is to follow the optimal solution, and compared with GA, all particles may more quickly be convergent to the optimal solution under most situations.

2. Principle and implementation of GA

2.1 Basic principle

When solving problems by GA, the solution of the problem can be denoted as chromosome, i.e. the string of binary coding in the algorithm, and many chromosomes can be generated by the random mode to form the initial swarm. Then these chromosomes are put in the environment of problem, and the chromosomes which can better adapt the environment according to the principle of survival of the fittest can be selected and copied, and new-generation chromosome swarm which can more adapt the environment by crossing and aberrance. In each generation swarm, each individual can be evaluated by one numerical value by the fitness function, so the individual with high fitness can join in the inheritance operation and the individual with low fitness will be eliminated. In this way, more generations can evolve, and finally one chromosome which can adapt the environment most will be convergent, and it is the optimal solution of the problem. That is the basic principle of GA.

2.2 Algorithm flow

The running parameters of basic GA include N, T, Pc and Pm, which are related with the dimension amount of the variable and the geometric character of the function. Commonly, N is the size of the swarm, i.e. the amount of individual in the swarm, T is the terminal evolvement generation of GA, Pc is the crossing probability, and Pm is the aberrance probability. These four running parameters will influence the solution result and the solving efficient of GA. The concrete flow of GA can be described as follows.

(1) Code and generate initial swarm. Because the disposal objective of GA is the chromosome coding string, and GA can not directly dispose the solution data in the solution space, and the solution space must be denoted as the genetic string structured data of genetic space by the mode of coding. The basic GA uses binary symbol string with fixed length can be used to denote the individuals in the swarm, and its allele (genes in same gene position) is composed by binary symbol set {0, 1}. The gene value or individual of various individuals in initial swarm can be generated by the evenly distributed random values. Aiming at two kinds of usual coding of GA, the advantages and disadvantages of real number coding and binary coding can be combined to code. The implementation of independent variable in real number coding and the operation of binary coding are convenient, so the binary-real number combined coding is used in the article.

(2) Compute the fitness value of each individual in swarm. In the searching process, GA generally needs not other exterior information, and it only uses the fitness function to evaluate the each individual's fitness ability to the environment, and regards it as the reference of heritance operation. The standard GA adopts roulette-wheel-selection mechanism which decides the opportunity degree that each individual in the present swarm is inherited to next generation swarm according to the probability of the positive proportion with the individual fitness. Because the probability that each individual can be copied should be computed, so the fitness of individual must be kept as positive number or zero. Especially, when the objective function is negative, the conversion rule between the objective function to individual fitness must the confirmed beforehand.

(3) Select next generation individual according to certain rule decided by individual fitness value. The essential of selection is the copy of chromosome, and it is the most important cause to keep the properties and achieve the stability of species. In GA, the function of selection is like a boult, and according to the fitness of individual to the objective function, the individual chromosome with high fitness can be selected, so its gene can be inherited and copied to next generation, but the individual chromosome with low fitness can be eliminated. So the essential of selection is filtration, and the function means directional evolvement. By multiple directional accumulations of selection operators, individuals in the swarm will quickly close to the region with high value of objective function, and form the swarm of high-quality solution. The implementation method is to use the roulette-wheel-selection mechanism, i.e. selecting the operator according to the proportion.

The roulette-wheel method is an evaluation method of directional selection based on probability. Suppose that the fitness values of individual in swarm respectively are f(i) ($i \in (1, 2, \ldots,$ pop size)), and the chip in the roulette-wheel is c(i), if i>begin.

$$\mathrm{rf}(i)=\frac{\sum_{i=begin}^{i} f(i)}{\sum_{i=1}^{n} f(i)} \tag{1}$$

If (i <begin)

$$\mathrm{rf}(i)=\frac{\sum_{j=begin}^{n} f(j)+\sum_{j=1}^{i} f(j)}{\sum_{j=1}^{n} f(j)} \tag{2}$$

Where, "begin" is the subscript of the particle selected last time, and the fitness and the selected probability in the roulette-wheel method is

$$p(i)=\frac{rf(i)}{\sum_{i=1}^{popsize} rf(i)} \tag{3}$$

Where, *p(i)* denotes the probability that the i'th individual is selected, and it obviously seems the probability that the individual with big fitness value can be selected is also higher. When the popsize time roulette-wheel is executed, the father swarm of next generation can be generated. The essential of selection is the copy of chromosome, and it is the most important cause to keep the properties and achieve the stability of species. One important stage in the roulette-wheel method is to compute the fitness of each particle after each selection in the roulette-wheel method.

(4) Implement the crossing operation according to the probability Pc. The essential of crossing operation in GA is that the son-generation individual can be obtained from father-generation individual in small range by large-probability aberrance, and this range is decided by two individuals though "or" or "and" operation. The crossing operator is the core of the GA operation, and it includes two basic contents, and one is to randomly match the individual in the matching base formed by the selection operation, and decide whether each pair needs the crossing operation according to the pre-established crossing probability Pc. And the other is to set up the crossing point of individual to exchange part structures before and after these crossing points. The random method is used in the whole process, and it will produce the crossing individuals, i.e. continually generating random number temp$\in$[0,1] in the circulation, and if temp is in the interval of [0, Pc], this individual is the matching individual. So the size of Pc is the aberrance probability. The random method produces matching individuals, match=floor(rand*i), match$\in$[0,i], i.e. it is crossed with former individuals, and this way can simplify the algorithm, and it would not the exchange quality by testing. The random method produces the exchange bit, n=ceil(codel*D*rand), and the single-point crossing method is adopted.

(5) Implement the aberrance operation according to the probability Pc. The introduction of the aberrance operator is to help the swarm to search the whole solution space. In the evolvement process of algorithm, the aberrance plays very important function to recover the diversity of swarm and ensure the global constringency of the algorithm. The aberrance operator means gene value on some gene positions in individual chromosome coding string can be substituted by other alleles on this gene position, and new individual will be generated. The aberrance is the change of son-generation gene generated according to small-probability interference. The intention is to add the diversity of swarm, promote the performance of algorithm, and avoid the immature convergence. The operation uses the basic bit aberrance operator or the even aberrance operator, and it is implemented according to the bit, i.e. the values of certain bit are exchanged. When the aberrance happens to certain gene bit, the potential of new same-order mode will exist. Therefore, the aberrance is also the important part of GA to implement the global searching. Aiming at the strong aberrance of GA, to increase the global searching ability, the dynamic aberrance factor can be adopted because the GA is difficult to be convergent, which makes the aberrance factor to be gradually reduced with the increase of iterative times, and avoid the deficiency of global optimization solution induced by too many mutations. To be similar with crossing operator, the aberrance operator decides the aberrance probability, and the implementation method is similar with the crossing algorithm. The dynamic binary aberrance factor is adopted in the article, because with the maturation of the evolvement of particle swarm, to reduce the aberrance degree, the good individuals should be saved to avoid the loss of global optimal solution. With the running of each evolvement, the experience function of Pm with the change of the evolvement time is

$$\mathrm{p_m}=\mathrm{p_{m0}}+\mathrm{k}^4\times(\mathrm{p_c}/3-\mathrm{p_{m0}})/G^4 \tag{4}$$

Where, P_{m0} is the initial value, K is the evolvement time, Pc is the crossing factor, and G is the maximum iterative time.

(6) If certain stop condition has not been satisfied, turn to the step (2), or else, enter into the step (7).

(7) Output the chromosome with optimal fitness in swarm as the satisfactory solution or the optimal solution of the problem.

3. Principle and implementation of PSO

3.1 Basic principle

PSO originated from the simulation of bird swarm, and it is one of main algorithm in the domain of swarm intelligence research (Kennedy, 1995, P.1942-1948 & Parsopoulos, 2002, P.235-306). similar with other evolvement algorithms, PSO moves the individuals in the swarm to better region according to the fitness to the environment, and the difference is that it doesn't use the evolvement operator to the individual like other evolvement algorithms, but regard the potential solution of each optimization problem as a "particle" without volume and quality in the *d* dimensional searching space, and all particles have one their own fitness value decided by the objective function, and one speed to decide the flying direction and distance, so particles can fly in the solution space following the current optimal particle, and dynamically adjust the flying speed according to the experiences of individual and collectivity, and finally find the approximately optimal solution by the searching iterative. Concrete mathematical description can be described as follows.

Suppose the i'th particle in the swarm is $X_i = (x_{i1}, x_{i1}, ..., x_{id})$, and its position is $P_i = (present_{i1}, present_{i1}, ..., present_{id})$, and the best position is pBest. The best position that all particles go through in the swarm is gBest. The speed of the particle i can be denoted as $V_i = (v_{i1}, v_{i1}, ..., v_{id})$. For the iteration each time, the updating formulas about the speed and position of the particle i in d (1≤d≤D) dimensional space are

$$V(t+1) = \omega * V(t) + c_1 \cdot rand \cdot (pBest(t) - present(t)) + c_2 \cdot rand \cdot (gBest(t) - present(t)) \quad (5)$$

$$present(t+1) = present(t) + V(t+1) \quad (6)$$

Where, ω is the inertia weight, and its range is [0.1, 0.9], and it can keep the inertia of the particle to make it to explore new region, and the updating of ω accords with the formula (7), and the maximum inertia weight $\omega_{\max}$ descends to $\omega_{\min}$ with the iteration linearly, and $iter_{\max}$ is the maxim iterative time established for the algorithm, and $iter$ is the present iterative time.

$$\omega = \omega_{\max} - iter \cdot \frac{\omega_{\max} - \omega_{\min}}{iter_{\max}} \quad (7)$$

c_1 and c_2 are acceleration constants, and they make each particle to move to the positions of pBest and Gbest more quickly. Rand is the random number in [0, 1]. In addition, the speed of particle V_i is limited by the maximum speed $V_{\max}$ which decides the searching precision of the particle in the solution space. The formula (8) and the formula (9) show the searching processes of the positions of pBest and gBest. Present is the present position of the particle, and the position with the best fitness is appointed as the best position pBest of this particle, so it is appointed as the global optimal position gBest.

$$present(t+1) = \begin{cases} pBest(t) & if \quad f(present(t)) \le f(pBest(t)) \\ presetn(t+1) & if \quad f(present(t)) > f(pBest(t)) \end{cases} \quad (8)$$

$$gBest(t+1) = \arg\max_i f(pBest_i(t+1)) \quad \forall i \in N \quad (9)$$

The terminal condition of PSO generally is to achieve the maximum iterative time or search the optimal fitness value to satisfy the threshold condition.

3.2 Algorithm flow

The concrete flow of PSO is shown in Figure 4, and there are many parameters to be adjusted.

(1) The particle swarm scale, i.e. the particle amount N (swarm size): The amount is generally in 20-40, and 10 particles can acquire good result, and for the complex problem or special problems, and amount of particle can be 100 or 200, and the amount of particle in the searching space each time in the article is 80 to keep consistent with GA and control the comparison condition possibly.

(2) The length of particle D (space dimension): It is decided by the testing problem, and it is the length of the problem solution. The testing function in the article is Schaffer function, and the length of this function D is 2.

(3) The coordinate range of particle: It is decided by the optimization problem, and each dimension can establish different ranges such as [-2.0, 2.0].

(4) The maximum speed V_{max}: It decides the maximum movement distance of the particle in one circulation, and it is set as the range width of the particle.

(5) The learning factor: c1 and c2 are all in 0-4.

(6) Inertia factor ω: To add inertia into initial PSO is to seek the optimal balance between global optimization and local optimization, and big inertia can help to search the global optimization, and small inertia can help to search the local optimization. The researches show that with the increase of iterative time, the inertia descending gradually in [1.4, 0] can obtain good effect. In this article, $\omega'=\omega*u^{-k}$ ($\omega\in[0,1]$; $u\in[1.000\ 1,1.005]$), ω is the inertia weight, and the representative initial value is in [0, 1], and big ω is propitious to global searching, and small ω is propitious to local searching.

(7) Terminal conditions: The conditions include the maximum circulation times and the minimum warp requirement, and the maximum circulation time is the terminal condition in the article.

4. Experiment result and analysis

According to above two algorithms, corresponding program is developed in the article, and the performances of GA and PSO are studied and researched by the representative testing function Schaffer. Following factors such as easy implementation, good testing effect and variable exchangeability are mainly considered, and closely related with the characters of GA and PSO. Two functions all code variables in one matrix, so variables have same running tendency and numeric area. The maximum of 3D Schaffer function is 1, and the global maximum point is at [0, 0]. Schaffer is the testing function because there are infinite local extreme values around the global optimal values. The intensive shocking character and the above character of the Schaffer function make general algorithm to be difficult to find the global optimization of this function. Therefore, this function can be used to test the influence of parameters on the optimization performance of algorithm in PSO. The expression of the Schaffer testing function is the formula (10), and its curved surface of 3D function is shown in Figure 1.

$$f(x)=0.5-(\sin\sqrt{x_1^2+x_2^2}-0.5)/(1.0+0.001(x_1^2+x_2^2))^2$$
$$-2.5\le x_i\le 2.5,\quad i=1,2 \tag{10}$$

To ensure that the testing result can objectively reflect the characters of two algorithms as much as possibly, both irrelative parameters should be controlled strictly, such as same iterative time (MaxDT), same searching dimension (D), and same searching range [-2, 2]. Both common maximum iterative time MaxDT=500, and the variable dimension D=2, and the particle or individual amount of one time iteration size=80. For the PSO algorithm, $c_1=c_2=2$, ω= 0.7298. For the GA, pr=0.6, pc=0.8, pm0=0.01. From Figure 2, the algorithm of PSO first enters into the convergence state. From the particle distribution of Figure 3 and Figure 4, the distribution range of PSO particle is centralized like the nebula state, but the distribution of GA is discrete without obvious convergence. As viewed from the principle of algorithm, the cause is that the PSO is the algorithm of directional convergence, and new particles all will compare with global values and historical values to more effectively confirm the selection direction.

Table 1 includes the data of comparison experiments ten times of GA and PSO. $\overline{N(PSO)}=225.6$, $\overline{N(GA)}=279.8$, $\overline{fitness(PSO)}=0.99$, $\overline{fitness(GA)}=0.98$. When obtain the maximum, the average iterative time of PSO is less, and the searching effect is better.

5. Conclusions

In this article, two important swarm intelligence algorithms including GA and PSO are deeply analyzed and compared. GA is mainly evolved by the form of swarm through four operations including coding, selection, crossing and aberrance to seek the optimal solution. Only the selection operation in four operations has the character of direction, and other operation operators are mainly to ensure the diversity of swarm. So the selection operation is selected as the core to implement and promote GA. The simulation experiment result of two algorithms shows that GA has high global searching ability, but the computation efficiency is low and the optimization speed is slow, and it is hard to be convergent, and the complexity of algorithm is higher. PSO can be implemented simply, and the convergence speed is quick without too many parameters, and it has good global searching ability, because the information of particle is single-directional, each particle would remember the past position, and the convergence is very quick, but at the same time, it is easily to fall into local optima. At present, the selection operation in GA is mainly to apply the probability mechanism without the judgment of convergence condition and the selection process of memory, which is the

combination point of GA and PSO.

References

Duan, Xiaodong, Wang, Cunrui & Liu, Xiangdong. (2005). *Particle Swarm Optimization and Application.* Shenyang: Liaoning University Press.

Ge, Jike, Qiu, Yuhui & Wu, Chunming. (2008). Survey on Genetic Algorithm. *Computer Applications and Software.* No.2. P.2911-2916.

Grosan C, Abraham A & Han SY, et al. (2005). Hybrid Particle Swarm- evolutionary Algorithm for Search and Optimization. *Proceedings of the 4th Mexican International Conference on Artificial Intelligence, Springer-Verlag, London.* P.623-632.

Jiang, Jiuchuan & Xia, Zhengyou. (2009). The Strategy Control Mechanism Based on Swarm Intelligence. *Control & Automation.* No.16. P.26-28.

Kennedy J & Eberhart R C. (1995). *Particle Swarm Optimization.* Proceedings of the IEEE International Conference on Neural Networks IV, IEEE computer society, Washington DC. P.1942-1948.

Kennedy J & Eberhart R C. (2001). *Swarm Intelligence.* USA: Academic Press, 2001.

Parsopoulos K E, Vrahatis M N. (2002). Recent approaches to global optimization problems through Particle Swarm Optimization. *Natural Computing.* No.1. P.235-306.

Peng, Xiyuan, Peng, Yu, Dai, Yufeng. (2003). Swarm Intelligence Theory and Applications. *Acta Electronica Sinica.* No.z1. P.1982-1988.

Tarasewich P & McMullen P R. (2002). Swarm Intelligence: Power in Numbers. *Communication of ACM*. No.45(8). P.62-67.

Zhang, Wenxiu & Liang, Yi. (2003). *Mathematical Base of Genetic Algorithm.* Xi'an: Xi'an Jiaotong University Press.

Table 1. Comparison of searching results by GA and PSO

Experiment times	1	2	3	4	5	6	7	8	9	10
N (PSO)	128	222	232	240	238	261	2	255	230	252
N (GA)	456	323	419	343	427	68	244	179	235	5
Fitness (PSO)	0.90	1	1	1	1	1	0.75	1	1	1
Fitness (GA)	0.98	0.99	0.98	0.99	0.91	0.98	0.96	0.997	0.98	0.97

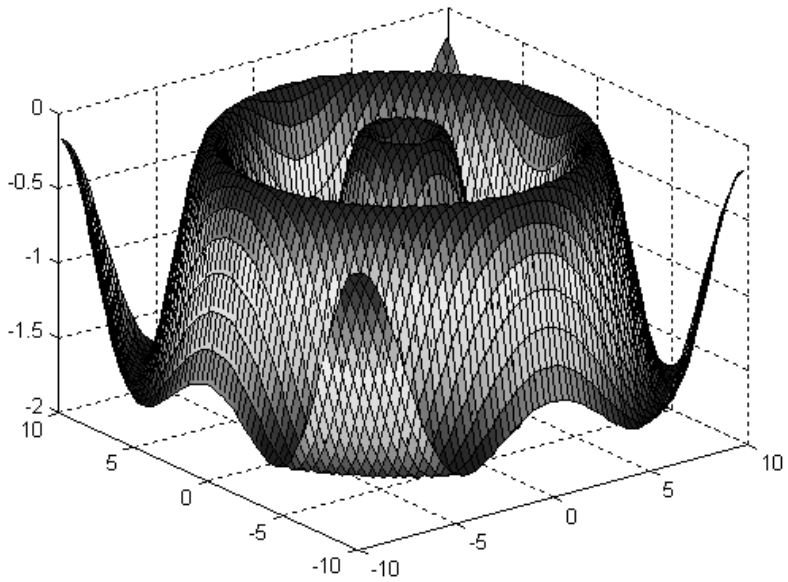


Figure 1. 3D Schaffer Function

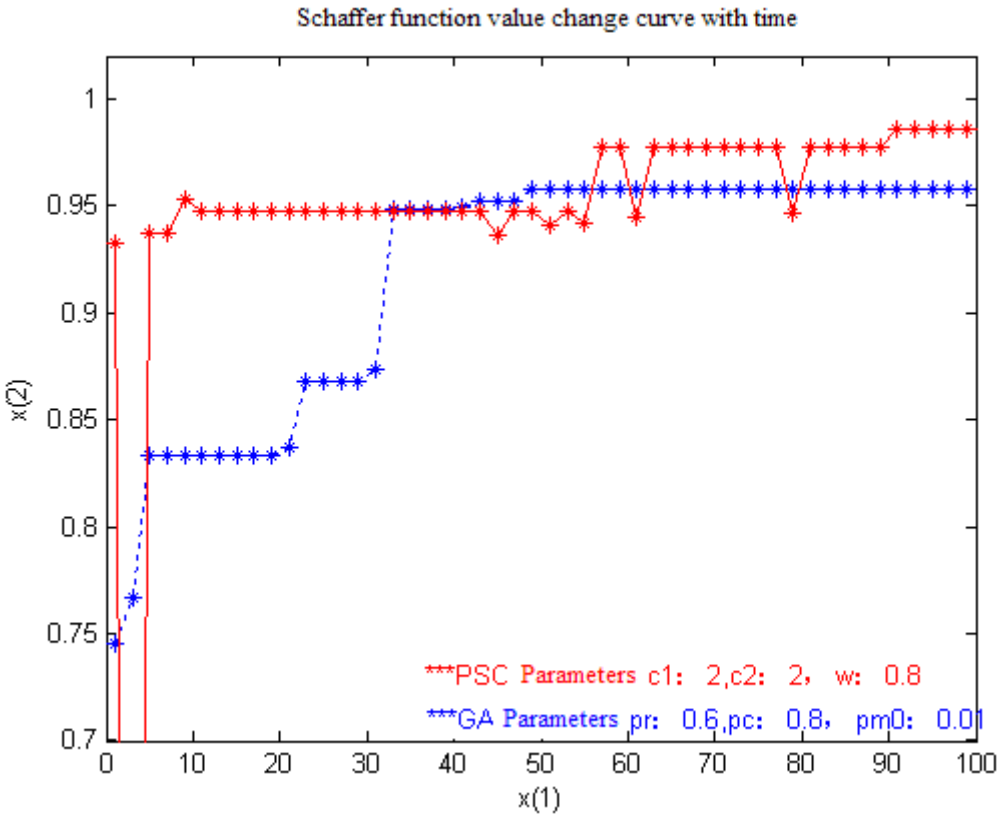


Figure 2. Comparison of Iterative Processes of GA and PSO

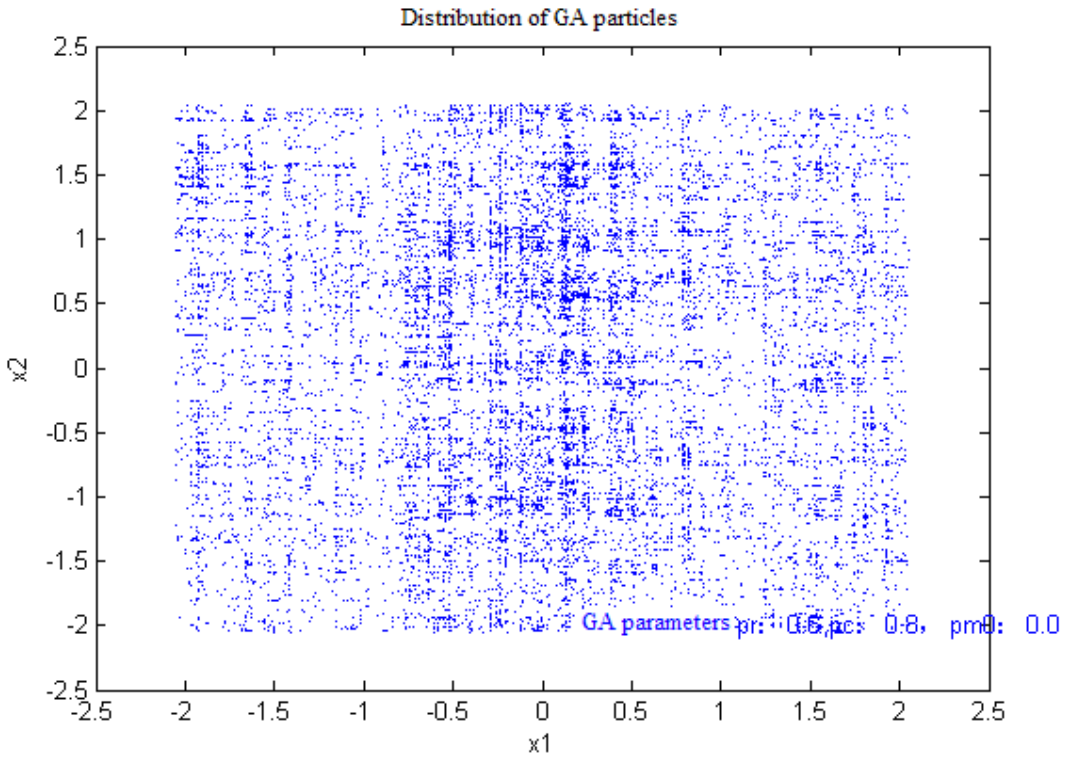


Figure 3. Chromosome Convergence Result of GA

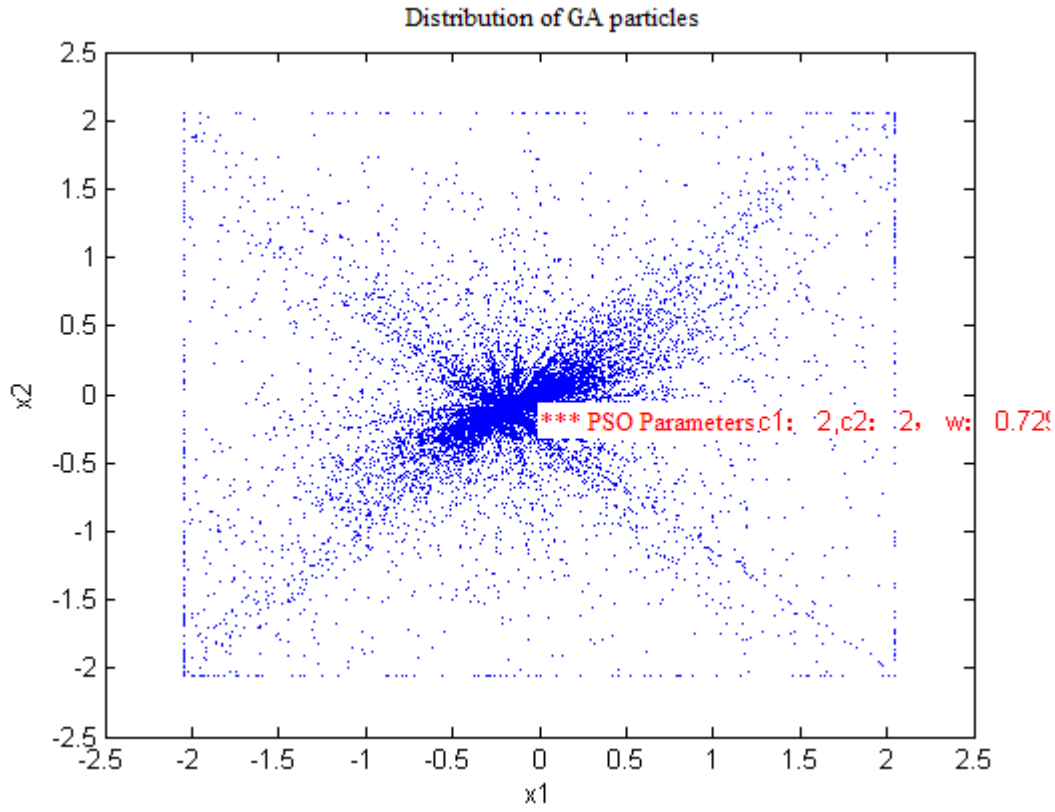


Figure 4. Chromosome Convergence Result of PSO

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Collaborative Supply Chain in the Digital Age: A Case Study of its Extent of Adoption by Indigenous Organizations in Building Inter-and Intra-firm Alignments

Hart O. Awa (Corresponding Author)
Department of Marketing, University of Port Harcourt
P. M. B. 5323, Port Harcourt, Nigeria
E-mail: sundayeze@yahoo.com
Nsobiari F. Awara
Department of Marketing, University of Calabar, Calabar, Nigeria
Bartholomew C. Emecheta
Department of Management, University of Port
Harcourt, P. M. B. 5323, Port Harcourt, Nigeria

Abstract

The term collaborative supply chain or SCM is a social software package that promises inter-and intra-firm alignments and information sharing to assure outstanding performance of the whole. It involves integrating resources and automating transactions across traditional boundaries to build mutually benefiting competitive advantage. SCM in the digital age encompasses accessibility of internal activities and measurement metrics by external parties as well as optimization of processes for mutual improvement in customer service, and reduction of inventory level and associated cost; good business relationships built on trust and minimized selfishness; and long term flexibility and adaptability evidenced by co-creation of products along the value chain. Apparently, experience has shown too many firms avoid engagement in this extensive supply chain integration, perhaps for fear of breeding conflicts. The paper viewed the current practice of SCM by studying the experiences and insights of 108 managers/officers, drawn from eight indigenous firms. The benchmark for selection was that the subject must be directly involved in supply chain initiatives and case study interviews were used for the survey. The finding unravelled discrepancy between the ideals of theories of SCM and actual practice; no consensus about its meaning, multiple supply chain collaboration compounds the problem even the more, and most times it may be difficult to meet the laid down conditions. Firms easily embrace internal cross functional integration and shun inter-organizational integration, perhaps because they want to build unique competitive advantage. Therefore more orientations should be done to educate managers on the ideals of SCM, especially on the area of dissuading selfish maximization of profits.

Keywords: Supply chain, Supply chain management, Collaboration, Optimization, Information technology, Innovation, Integration, Network effects

1. Introduction

Thomas Friedman's The World is Flat reported that the main effects of the Internet is people learning to adapt to a world where everybody is connected, everyone contributes, and everyone is in zero distance (or close enough) from everyone else (Israel, 2007).

The digital age is right here; everything is revolutionized. Not only has it changed and/or reshaped how the entire enterprise communities communicate and conduct businesses but specifically how the supply chains are operated and managed. The theories of communities of interest/group of companies (Power and Sohal, 2002), collaborative commerce (c-commerce), strategic alliance, B2B (Boone and Kurtz, 2004), social software (McAfee, 2006), enhanced network effects (Koch, 2001), buyer-supplier interface (Humphreys et al, 2006), Vertical Marketing System (VMS) and/or operating system (Ridgeway, 1957) are clear manifestations of the nascent digital force that encourages appropriate inter-firm alignments and relationships to assure outstanding performance. Implicit is that advances in IT are endless (Awa, 2003); they have currently helped supply chains to link their activities into networks enabling them to communicate, collaborate and co-operate through Internet connection (Cook, 2008; Mason et al, 2008). Organizations have relentlessly restructured and re-engineered their processes (Fawcett and Magnan, 2002) often by integrating resources and automating transactions across their traditional boundaries (Ballou et al, 2000; Bartholomew, 1999; Dell

and Fredman, 1999; Kaplan and Sawhney, 2000) to improve network relationships, to cut costs and create exceptional but difficult to copy customer values (de Burca et al, 2005; Blackwell, 1997).

The objectives and resources of teams of suppliers, finished goods producers, and service providers (transporters and dealers) are integrated to cost-effectively deliver customer value. Abell (1999) opined that this requires the alignment of functional and supply-chain partner activities with company strategy and harmonizing such with organizational structure, processes, culture, incentives, and people. Further, organizations now view markets as social conversation involving participative architecture through which Internet platforms are used to enable online communities (manufacturers, dealers, suppliers, and customers) to relate, to interact and to collaborate. Such social software platforms allow for overtime creativity, communications, secured information sharing in real time, and collaboration between members of on-line communities (McAfee, 2006). Internet-based information transfer attracts more interactivity/interconnectivity built among supply chain partners by replacing traditional supply chain with supply web (Deloitte and Touche, 1998; Kehoe and Boughton, 2001) and by conducting business in a global village (Chou et al, 2004). SCM-related information and decisions are integrated into a complex but well-defined Web of relationships with multiple channels and an open flow of information (Koch, 2001) thereby breaking the old paradigms of inter-organizational boundaries. Supply chains change from an order-driven-lot-sizing approach to one likened to a capacity-availability-booking approach supported by appropriate search engines (Kehoe and Boughton, 2001). The economic potential of this integration in terms of improving both profitability and competitive position was first identified by Forrester (1958) and later operationalized by Wood (1997) and perhaps other theorists who demonstrated that a 10 percent reduction in supply chain's cost structure may yield 40 percent to 50 percent improvement in pre-tax profit.

Collaborative environment permits collaborative structure to determine authority; sharing of risk and rewards; long-term and shared commitment and goals; dividing of cognitive processes into intertwined layers; and mutual participation architectures in a co-ordinated effort to solve a problem. All steps in the supply chain from design to after-sales service are integrated flows (Kennerley and Neely, 2001). Rather than viewing supply chain traditionally (shaving suppliers' margin) (de Burca et al, 2005), the needs of all stakeholders can be best served through more strategic approach of optimization (Power and Sohal, 2002). Networking information for access to all partners in the supply chain provides strong basis for building sustainable competitive advantage (SCA). For instance, Symantec Organization invested in Website containing several bulletin boards, whereupon dissatisfied customers e-mail, or post automatically, their problem(s) to the appropriate bulletin board and expect solution(s) from Symantec's technicians or customer support representatives. Awa (2003) viewed customer satisfaction as not being one-man show, rather it is a complex activity involving inputs from vendors and other independent firms and continues even by the manner the dealers handle and explain customers' complaints and doubts. Such relationship has increasingly attracted primary focus from organization worldwide, perhaps owing to rapid changes in technology, demography, globalization of businesses, product customization, and others, which combine to refocus awareness on optimizing performance based on McCammon and Little's (1965) view of centrally programmed networks rather than individual organizations chasing profit maximization selfishly.

Although technologies such as EDI, barcoding, and product numbering have been in use for over thirty years (particularly EDI) (Johnston and Mak, 2000) to create inter-organizational information flows to the physical movement of goods (Power and Sohal, 2002), Kehoe and Boughton (2001) are of the view that the role of individual organizations in the supply chain need be re-evaluated to reflect current developments in communications and information technologies. Rapid change in technology has positive and negative economic impacts; it provides further opportunities for improved supply chain performance and promotes obsolescence. The Internet, Intranet, and, in particular, the use of Extranet offer opportunity for improvements in supply chain (Kehoe and Boughton, 2001) in order to realize the opportunities of streamlined logistics flows; costs reduction and improved operational efficiencies, typically in procurement, communication, inventory holding, network relationship and search activities (Bakos, 1991; Fernie, 1995; Cottrill, 1997; Chan and Swatman, 2000; Davenport and Brooks, 2004); improved customer service, better quality goods and services and consistency via transparency, value-added information and new levels of innovation from network externalities and knowledge sharing (Raisch, 2001; de Burca et al, 2005).

Despite the apparent opportunities of eSCM; adoption by firms remains slow, especially amongst SMEs. Only a few firms have adopted and disseminated a formal eSCM definition (Fawcett and Magnan, 2002). Many firms show resistance to information sharing for want of building unique competitive advantage. Though much desire to be done, ample research has been done on how eSCM contributes to different business sectors. And therefore the purpose of this paper is to make our humble scholarly contribution in this fascinating area and to specifically report on its current state of adoption in Nigeria. This will provide knowledge on the extent of understanding of how eSCM assists in optimization of the value chain through strategic responses to its challenges. However, this will timely bridge knowledge gap now many Nigeria-based organizations long to re-brand and enhance performance and competitiveness, and to identify the barriers inhibiting them from harnessing the potentials of Internet technologies.

2. IT and Change in Supply Chain

IT brought fundamental changes (Walton and Gupta, 1995; Ferguson, 2000; Koch, 2001) that create competitive advantage in the supply chain (Chou et al, 2004) via cost cutting, bridging of new markets, and protecting of intellectual assets (McKinnon, 2009). Newman and Thomas (2009) noted that IT constantly throws at us new buzz phrases; the thresholds of dramatic shift in ways firms organize, innovate, and create values. The disruptive effects of Internet question tried-and-tested business models, ways of working, organizational structures, and accepted truths (Stamer, 2008) and fundamentally changes the rules of the game since the said changes are exponential (Stamer, 2008; Mancini, 2009). The effects of Internet on supply chain and supply chain management are evident in their operation, at least to the extent of transparency and information visibility. Given that recent states of art may become outdated legacy quite early (Hewitt, 1999; Froehlich et al, 1999), at least for now and in the near future, Internet technology permits optimization of the entire supply chain through inventory cost reduction, total cycle time compression and accessibility of demand data and supply capacity data (Kehoe and Boughton, 2001) for informed decision. The Internet changes how supply chain is managed, planned, and controlled.

2.1 Supply Chain and Supply Chain Management (SCM)

The term supply chain has varying connotations ranging from cross-functional process integration within the firm, backward integration with valued first-tier suppliers, and forward integration with valued first-tier customers, to complete forward and backward integrations (Fawcett and Magnan, 2002). Whichever way, confusion still besieges its meaning and definition, perhaps because of its expanding scope and differences in functional views. Its common focus rests on demand and supply planning and forecasting, manufacturing scheduling, and transportation planning though uncommon focus rests on graphical supply chain modeller and supply chain optimizer. Supply chain encompasses all activities associated with the flow and transformation of goods from the raw materials stage (extraction), through to the end user(s) as well as all information flows (Handfield and Nichols, 1999). The bottom line is the creation of mutual values requiring not only managing materials integrated into the flows across enterprise traditional boundaries but also relationships and information in order to minimize conflicts.

This definition contrasts Supply Chain Management (SCM), which is frequently associated with materials management environment as effort to align objectives and integrate resources across company boundaries to deliver greater values. Du (2007) and Fawcett and Magnan (2002) noted that SCM involves advanced information technologies, rapid and responsive logistics service, effective supplier management, and increasing customer relationship management. It is a managerial philosophy, the implementation of a managerial philosophy, and as a set of managerial process (McKeown, 2000). Managerial philosophy connotes SCM integrating the total flow of a distribution channel from suppliers to ultimate users (seeing the individual firms as integrated by common goals); implementation of a managerial philosophy extends integrated behaviour to incorporate customers and suppliers through external integration; and a set of managerial processes involves managing relationships, information, and material flows across enterprise borders to deliver enhanced customer service and economic values through synchronized management of the flow of physical goods and associated information from sourcing to consumption (McKeown, 2000; Chou et al, 2004).

The last being more encompassing was further improved by Mentzer et al (2001) and later by de Burca et al (2005). Mentzer et al viewed SCM as the systemic, strategic co-ordination of the traditional business functions within a particular company and across businesses within the supply chain, for the purpose of creating a sustainable competitive advantage (SCA) for the individual firms and the supply chain as a whole. Whereas de Burca et al (2005) saw it as the movement of products through the supply chain, the management of associated information flows and business relationships, and ultimately the creation of customer value. Ridgeway (1957) and McCammon and Little's (1965) likened SCM to Vertical Marketing System and operating system when they noted that the economic process, right from the acquisition of resources and running through manufacturing to the ultimate consumption, is a continuous process involving, in many cases, the economic flow of, and co-ordination of, activities of many independent firms.

The implication is Supply Chain Management (SCM) serves as the back-end application that links suppliers, manufacturers, distributors, and resellers in a cohesively collaborative production and distribution networks (Chou et al, 2004) that permit individual firms to exploit its unique competence fully and to outsource non-core activities from other channel members (Cox, 1999; Quinn, 2000), who perhaps possess superior capabilities. They are all independent firms working cohesively in collaboration with one another to form one extended system that must be administered as a whole rather than a part; decision by each sub-system/firm must be consistent and mutually supporting (Awa, 2003). The links involve moving products through the supply chain and creating customer values via effective management of information and relationships (de Burca et al, 2005). Thus the premise of SCM is profoundly based on the postulates of Relationship Marketing and Customer Relationship Marketing (CRM) that customer satisfaction or creating competitive advantage(s) does not depend solely on matching production to audiences' needs (JIT); rather it transcends to a great deal of other relationships with other corporate partners in the value chain (Boone and Kurtz, 2004). Previously and traditionally, relationship in supply chain focused on viewing suppliers and perhaps distributors as adversaries against

whom to fiercely negotiate conditions and play one off games to shave margins. SCM has changed this awkward attitude as supposedly independent firms in the chain discovered the direct and/or indirect benefits of optimization in collaborative relationships (Cooper et al, 1997) perhaps subject to the harmonization of the interests of the individual firms.

2.2 Internet Technology and SCM

The Internet is a technological and business innovation with new methods of communications (Chellappa et al, 1996), business transactions (de Kare-Silver, 1998), market structure (Giaglis et al, 1999), education (Murison-Bowie, 1999), and work (Doukidis et al, 1998). Lai et al (2001) viewed Internet in SCM as an Inter-Organizational Information Systems (IOIS), which facilitates and enhances communications, transactions, productivity, and building of competitive advantages. It connotes digitalized networks connecting a business and its trading partners (e.g.; suppliers, dealers, customers, etc.) in ways to enhance operational effectiveness through out the value chain particularly with respect to real time information exchange and commercial transactions. Shuen (2008) opined that information sharing on the Web benefits business managers and industry analysts. The former can easily analyze internal business models to determine a company's or supply chain's profit engine; and the latter can compare a firm's performance relative to competitors and use that as a basis for assessing its stock market capitalization value, acquisition value, or total enterprise value. Exchange of information relates to the theory of bullwhip effects (Lee et al, 1997). Forrester's theory of bullwhip effects states that irregularities and unpredictability in order quantities increase with the number of layers in the chain and the implication is that access to demand data and supply capacity data within the supply chain makes for proactive and perhaps reactive responses to demand fluctuations. This bullwhip theory may not be wholly validated because collaborative relationship demands rationalization and consolidation of supply base (Matthyssens and Van den Builte, 1994) as reduction in number of suppliers or dealers will improve and deepen relationships (Humphreys et al, 2006). Reduction in the immediate supply chain leads to change in structure and the number of tiers in it. In manufacturing context, Humphreys et al (2006) observed that operationalizing this involves original equipment manufacturers (OEMs) buying up assembled systems or complete sub assemblies rather than individual components, thereby bringing another level into the supply chain. These policies permit first-tier suppliers to supply assembly systems and to co-ordinate the inputs and other activities of second-tier suppliers.

The Internet has revolutionized the supply chain (de Burca et al, 2005). Its strengths over EDI e.g.; ubiquitous and cost effective connectivity, speedy network transmission (Chou et al, 2004) and simplicity of use are well stimulated across the value chain (Davenport and Brook, 2004) to minimize conflicts (McCommon and Little 1965; Awa, 2003) and to enhance simultaneous improvement in customer service and network relationships (Kehoe and Boughton, 2001). eSCM manages information flows and represents a philosophy of managing technology and processes to optimize product delivery attributes, and information from suppliers through customers. An organization's Enterprise Resource Planning (ERP) must be integrated with CRM, e-procurement, and others, and implemented correctly before the advantages of eSCM capabilities can be fully exploited (Norris et al, 2001). Changes reflecting on management practices, performance standards, and business processes must be radically instituted in the entire supply chain to accommodate the new trend. Deloitte and Touche (1998) described Internet within the infrastructure of supply chain as an integral and one of the key drivers of change. Specifically, two major assets leading to the success of eSCM are often emphasized (Norris et al, 2001); they are viewing collaboration as a strategic and operational priority to foster trust (Morgan and Hunt, 1994), promise (Calonius, 1988), commitments (Dwyer et al, 1987), and mutuality (Czeipal, 1990) among trading partners; and management of information in the value chain with strict disciplines, policies, and monitoring.

The latter builds trusted relationship between parties and ultimately competitive advantage because the network effects of best practices are shared from the exchange of accurate and up-to-date information. The former viewed collaboration in terms of connoting some measure of social software involving mutual engagement and relationship building. It encompasses accessibility of internal activities and metrics by external parties (de Burca et al, 2005) as well as optimization of processes for mutual improvement in customer service, and reduction of inventory level and associated cost (Chou et al, 2004; Kehoe and Boughton, 2001); good business relationships built on trust and minimized selfishness even as perfect control over partners rarely exists (Scalet, 2001); and long term flexibility and adaptability evidenced by co-creation of products along the value chain (Sarkis and Sundarraj, 2000; van Hoek, 2001). The growth of e-commerce shifts decision power to buyers/users (Chou et al, 2004) and permits greater customization and user collaboration as well as faster organizational turnaround (Van Hoek, 2001), and ultimately improvement in relationships to build more customer loyal behaviour. Merono-Cerdan et al (2008) classified such collaborative tools into electronic communication systems (ECS) (e.g.; wikis and human-based computation) and teamwork system.

For instance, wiki is a metaphor for a new era of collaboration and participation (Tapscott and Williams, 2006) that encourages the entire supply chain accessing and editing websites; and human-based computation relates to technology that allows for collaborative problem solving within the supply chain (Cook, 2008). The ECS aims at facilitating information and opinion exchange, and enables relationships to be established among workgroups, customers, and

institutions/supply chain members. The shared database updates information, keeps the organizations' systems memory up to date and prevents data from being repeated while being modified by authorized persons within the team or organization. The teamwork systems pre-define work processes and integrate information, which is classified into two-repositories and workflow. Repositories are important documents incorporating both unspoken and spoken knowledge in form of pictorial, textual and diagrammatic formats from project specialists; and workflow shows the sharing of the task among value chain members.

Porter (2001) opined that eSCM will build and deliver better sustainable competitive advantage (SCA) when strategically synchronized with such traditional sources as scale, human resources, and investments in physical assets, which also play prominent role. Knowledge management in eSCM system involves not technology alone but also its management by people. It espouses not only technology philosophies (ease-of-use, web-based, true multi-media, the use of broad-band, and mobile technologies) but also suggested behaviours (transparency, immediacy, participation, responsiveness, etc) (Mason et al, 2008; Coleman and Levine, 2008). eSMC is recognized for its characteristics of encouraging collaboration, knowledge sharing and management that culminate into enhanced productivity and building of long lasting inter-and intra-firm relationships, which ultimately reflect on customer satisfaction. It implies strategic alliance since integrated significant digital data generated through data-base marketing and relationship marketing may be archived and utilized successfully and collectively by an organization and its strategic/trading partners to build competitive advantage (Porter and Millar, 1985; Bibas, 1994; Gray and Watson, 1998).

3. Methodology and Supply Chain Model

In the light of the research agenda, authenticated information on current industry practice of SCM was based on case-study approach, involving the investigation of experience and knowledge of relevant industry managers, drawn across eight Port Harcourt-based indigenous organizations. This objective of this qualitative survey was to develop and report on the plausible broad-based view of how managers perceive the use of SCM as strategic for building competitive advantage. Specifically, the study unravelled whether the real perception of SCM varies across internal functions or inter-organizational positions in the channel structure. The cross-functions and inter-organizational characteristics of SCM signify boundary-spanning activity (Bowersox, 1999). Yin (1994) recognized case-study as an empirical inquiry, which investigates a contemporary phenomenon within its real life context when the boundaries between phenomenon and context are not clearly understood. In-depth interview(s) was conducted for more meaningful follow-up questioning and more extensive findings and insights.

The in-depth interviews lasted for 4 months and the target was 160 middle and senior level managers of purchasing, logistics, and manufacturing departments of the chosen organizations. The questions were structured and standardized to ensure uniformity. Where such positions do not exist in any of the chosen organizations, equivalents were used provided the occupants of such positions were directly involved in the implementation of SCM. The firms chosen for the study as reported in table 1 below are strategic in terms of being industry pace-setters and specifically they possess progressive reputation in supply chain practice. Literature review, pre-survey interviews, and advisory inputs from scholarly colleagues served to sharpen the survey instrument. The different sizes of the organizations informed the different number of respondents talked to. Only 108 (or 67.5 per cent) of the supposedly respondents actually held talks with the researchers; the rest, even when repeatedly visited were either not found on sit, too busy to attend to surveys, or inundated by surveys.

Figure 1 below presents a simplified model of supply chain integration as involving the independent firms and employees working collaboratively as one entity to achieve common goals of customer satisfaction and profitability. Often because of differences in tasks, the materials suppliers and service providers are managed differently in a co-ordinated manner perhaps by different functional areas of an organization. While managing materials suppliers is the responsibility of the purchasing department in most organizations, distribution and transportation (service providers) tasks are often handled by logistics, marketing, and perhaps purchasing departments. Except where streamlined standards exist and adhered to, this often breeds conflicts because each of these functional areas views organizational success, profitability and of course inter-and intra-organizational integration differently.

4. Discussion

The discussion of findings of this paper reflects the emerging nature of IT in Nigeria. Integrative attempts by firms surveyed loudly suggest some measure of substantive ambiguity and inconsistency over its meaning and actual practice because of varied functional views of SCM and of integrative nature. This supports Chou et al (2004), Kehoe and Boughton (2001), Deloitte and Touche (1998) and Fawcett and Magnan (2002). The organizations surveyed rated SCM on a seven-point scale and indicated it a critical strategy. For O7 and O8, the logistics managers are better disposed to integration and collaboration because they build a *front-office system*; purchasing managers are sceptical, *back-office system,* hesitant to integrative endeavours, and more comfortable with adversary practices, and the manufacturing managers appear on the fence, often responding more to the ideals of *front-office* than to *back-office systems*. Purchasing managers recognize SCM and eSCM as important competitive advantage building strategies but they are

rarely sure of their ability to stand the tests of time. They (purchasing managers) trade price for traditional philosophy of non-collaborative behaviour perhaps because of the multiple relationships between a firm and its suppliers. Best suppliers may equally be suppliers to stiffest competitors, some suppliers may simultaneously serve as our competitors in other business areas, and sometimes it is difficult to enforce strict disciplines, policies, and monitoring conditions of Norris et al (2001) when key suppliers delay deliveries.

For O2, O3, O4, O5, and O6, the stock controller/manager, the ordering/purchasing officer, and the sales manager have similar understanding of SCM and eSCM. Though hampered by trading partners' and customers' readiness, especially other retailers, these firms view SCM and eSCM in terms of reducing inventory holding and ordering costs, and improving customer service and collaborative relationships. This finding validated previous studies (e.g.; Chan and Swatman, 2000; Davenport and Brooks, 2004; Raisch, 2001; de Burca et al, 2005). Despites these benefits of inter-organizational integration, this study identified internal cross functional process integration as the key of supply chain initiatives in the organizations surveyed. Backward integration with valued first-tier suppliers and forward integration with valued first-tier customers are the most common forms of supply chain integration (see the first two options in figure 2 below. Downward and upward streams integration (see option three), though recorded by authorities as being practised by few firms, was never practised by the firms surveyed. Finally, complete forward and backward integration (the last option), expressed from suppliers' supplier to consumers' to consumers, is more of theoretical framework.

Asking the respondents to identify any of these types of integration in relation to their extent of use by their employers, it was revealed that organizations prefer internal cross-functional integration to inter-organizational integration. Fawcett and Magnan (2002) validated this with the clause of difficulties of tackling issues as aligning measures, meshing information systems, and sharing risks and rewards. Function-to-function interactions produce the least conflicts in the value chain. Agrawal and Pak (2001) attributed firms' resistance to information sharing across the value chain to using it to build unique competitive edge. Resistance to information sharing contrasts the view that eSCM provides opportunities that are largely unexploited (Ramdani et al, 2009); particularly it offers small firms plausible ways to strengthen competitive capability against their larger counterparts (Gengatharen and Standing, 2005; Urwin, 2000; Raymond, 2001). For inter-organizational integration, forward integration is operated more than backward integration and complete integration rarely operates. Logisticians are more concerned with customer satisfaction (front-face), which often attracts more of top management's interests than the purchasers' back-face systems.

All the organizations studied are at the preliminary stage of their inter-organizational collaborative relationships given the emerging nature of IT in Nigeria; each firm is at different stages in the journey, evidencing that the firms surveyed strongly know about the strategic implications of such integration. The major barrier to investment in collaborative integration is that of harmonizing the interests of front-face and back-face systems.

5. Conclusions and Managerial Implications

The world is now a global village and perhaps, less predictable. This is predominantly evidenced in the mode of business operations, which have vastly changed in favour of firms that are strategic and entrepreneurial in embracing the changing trends. Synchronizing SCM with IT is a welcome development for building inter-and intra-organizational cohesive teams that operate to compete across borders to increase productivity and market share amidst globalization, technological break-through, customer dynamism, competition, and other environmental demands. Several operations-based models (e.g.; Just in Time, Business Process Re-engineering, Total Quality Management, etc) have recently flooded the business environment, many of which have no staying power. Today SCM is smoke-balled a better model responsible for inter-and intra-organizational alliance in the supply chain though its actual practice is at variance with its theoretical frameworks. Its purpose is to optimize performance and competitive advantage in the value chain subject to having the broad-based mindset and the underlying collaborative infrastructures. The practice of SCM demands a set of fundamental criteria such as optimized collaborative relationship, inter-organizational processes, organizational structure, shared risks and rewards, aligned objectives and goals, consistent measurement metrics, and corporate cultures. These building blocks must be assessed, harmonized and aligned to form an integrative strategy that will be mutually benefiting. SCM as a social platform and/or a social system directs that once a sub-system is weak, the entire whole becomes weak. Having a profile page on the sites and linking online with other chain members for reasons of accessing their operations, eSCM permits proactive and perhaps, reactive handling of problems.

Authenticated and real time information capabilities deliver significant improvements in productivity and responsiveness in the supply chain and above all customer satisfaction and retention. The practice of collaborative alliance transcends information sharing and exchange to include information co-ordination to ensure their real time use in supply chain decision-making. SCM and of course eSCM have their staying power yet to be confirmed for it may be difficult to believe that one day competition against rivals will disappear and competition between supply chains for market supremacy fills the vacuum. Firms form vertical integration team that competes across boundaries in attempts to improve network effects and ultimately to capture larger global market share. The ideals of SCM appear seemingly

fascinating and rosy perhaps because it is a paradigm shift from selfishness to mutually consistent operations but in actual practice, much tension exists between SCMs' competitive potential and the inherent difficulty of collaboration, especially in the area of inter-organizational collaboration. Internal cross-functional, function-to-function, and front-face integrative relationships are more adopted and sometimes traditional non-collaborative philosophy still obtains. Supply chain is rarely managed from suppliers' supplier to customers' customer; and even among great users of SCM, integration spans a triad of companies, involving the company and one-tier up (front-office) and one-tier down (back-office) streams.

Lack of perceptual commonality besieges SCM and multiple supply chain collaborations often complicate supply chain design and management because boundaries and specifics of relationships may be vaguely defined. Differences exist in the front-office and back-office systems. The front-office enjoys somewhat measure of collaboration and integration because of its huge interest in building customer satisfaction and the back-office believes more in non-collaborative moves with suppliers.

Based on our findings, the following recommendations appear to encourage more successful applications by the firms surveyed and related ones.

- Harmonization need be done and everything spelt out before hand to avoid conflicts in the value chain.
- Precise and all encompassing formal definition of SCM as well as metrics of focus and measurement should be thought out and visibly posted and communicated through out the value chain to ensure balance between theory and practice.
- Trust and mutuality should be built into the value chain and a broad-based mindset developed to show commitment to the underlying collaborative infrastructures of SCM in an effort to guarantee its staying power.
- The front-office and back-office systems focus on customers and suppliers respectively and each has different views about SCM. Handling the doubts of front-office system and integrating and harmonizing its philosophies with those of back-office system will improve the adoption and results of SCM.
- A balanced behaviour by manufacturing manager will undoubtedly encourage more inter-organizational collaborative relationship.

References

Abell, D. (1999). Competing today while preparing for tomorrow. *Sloan Management Review,* 40(3), 73-81.

Agrawal, M. & Pak, M. (2001). Getting Smart about Supply Chain Management. *The McKinsey Quarterly*, 2, 22-27.

Awa, H. (2003). *Modern Marketing Theory: An Integrated Approach to Managerial Decisions.* Owerri: Kolleey Publishers.

Ballou, R; Stephen, M; & Mukherjee, A. (2000). New Managerial Challenges from Supply Chain Opportunities. *Industrial Marketing Management,* 29(1), 7-18.

Bartholomew, D. (1999). What's Really driving Apple's Recovery? *Industry Week,* 15(March), 34-40.

Bibas, S. (1994). A Contractual Approach to Data Privacy. *Harvard Journal of Law & Public Policy,* 17(2), 591-622.

Blackwell, R. (1997). *From Mind to Market: Reinventing the Retail Supply Chain.* New York: Harper Business.

Boone, L. & Kurtz, D. (2004). *Contemporary Marketing* 11th ed. Ohio, USA: South-Western.

Bowersox, D; Closs, D; & Stank, T. (1999). 21st Century Logistics: Making Supply Chain Integration a Reality. *Council of Logistics Management*, Oak Brook 11.

Brown, D. & Kaewkitipong, L. (2009). Relative Size and Complexity: E-Business Use in Small and Medium Sized Tourism Enterprises in Thailand. *Journal of Enterprise Information Management,* 22(1/2), 212-231.

Brown, D. & Lockett, N. (2004). Potential of Critical E-Applications for Engaging SMEs in E-Business: A Provider Prospective. *European Journal of Information Systems,* 13(1), 21-34.

Buchanan, R. & Gilles, C. (1990). Value Managed Relationships: The Key to Customer Retention and Profitability. *European Journal of Management* 8(4), 205-215.

Buonanno, G; Faverio, P; Pigni, F; Ravarini, A; Sciuto, D; & Tagliavini, M. (2005). Factors affecting ERP Systems Adoption: A Comparative Analysis between SMEs and large companies. *Journal of Enterprise Information Management. 18(4), 384-426.*

Calonius, H. (1988). A Buying Process Model in Innovative Marketing: An European Perspective. Proceedings of the XVIIthe Annual Conference of the European Marketing Academy. Blois, K. & Parkinson, S. (ed.), University of Bradford, 86-103.

Chan, C. & Swatman, P. (2000). From EDI to Internet Commerce: The BHP Steel Experience. *Internet Research-Electronic Networking Applications & Policy,* 10(1), 72-82.

Chellappa, R; Barua, A; & Whinston, A. (1996). Looking beyond Internal Corporate Web Server. In Kalakota, R. &Whinston, A. (eds). *Readings in Electronic Commerce.* Reading, MA: Addison-Wesley.

Chou, D; Tan, X; & Yen, D. (2004). Web Technology and Supply Chain Management. *Information Management and Computer Security,* 12(4), 338-349.

Chuang, T; Rutherford, M. & Lin, B. (2007). Owner/Manager Characteristics, Organizational Characteristics and IT Adoption in the SMEs, *Journal of Management & Enterprise Development,* 4(6), 619-634.

Chuang, T; Nakatani, K; & Zhou D. (2009). An Exploratory Study of the Extent of Information Technology Adoption in SMEs: An Application of Upper Echelon Theory. *Journal of Enterprise Information Management***,** 22(1/2), 183-196.

Cook, N. (2008). *Enterprise 2.0: How Social Software will Change Future of Work.* Farnham Gower Publishing Ltd.

Cooper, M; Lambert, D; & Pagh, J. (1997). Supply Chain Management: More than aNew Name for Logistics*. The International Journal of Logistics Management,* 8(1), 1-14.

Cottrill, K. (1997). The Supply Chain of the Future. *Distribution,* 96(11), 52-54.

Cox, A. (1999). Power, Value and Supply Chain Management*. Supply Chain Management: An International Journal,* 4(4), 167-175.

Czeipal, J. (1990). Service Encounters and Service Relationships: Implications for Research. *Journal of Business Research,* 20, 13-21.

Dabholkar, P. (1992). The Role of Prior Behaviour and Category-based Affect in On-site Service Consumers. *In Sherry, T. & Sternthal, B. (eds.) Diversity in Consumer Behaviour.* Association for Consumer Research, Provo, UT. 563-569.

Davenpart, T. & Brooks, J. (2004). Enterprise Systems and the Supply Chain. *Journal of Enterprise Information Management*, 17(1), 8-19.

de Burca, S; Fynes, B; & Marshall, D. (2005). Strategic Technology Adoption: Extending ERP Across the Supply Chain. *Journal of Enterprise Information Management,* 18(4), 427-440*.*

de Kare-Silver, M. (1998). *E-Shock: The Electronic Shopping Revolution: Strategies for Retailers and Manufacturers.* Basingstoke: Palgrave Macmillian.

Dell, M. & Fredman, C. (1999). *Direct from Dell: Strategies that Revolutionized an Industry*. New York: Harper Business.

Deloitte and Touche (1998). 1998 Vision in Manufacturing (Global Report). Global Manufacturing Survey by Deloitte and Touche & Deloitte Consulting.

Dixon, D. (2004). The Truce between Lean and IT. *Industrial Engineer,* 36(6), 42-45.

Doukidis, G; Poulymenakou, A; Terpsidis, I; Themisticleous, M; & Miliotis, P. (1998). *The Impact of the Development of Electronic Commerce on the Employment Situation in European Commerce,* Athens University of Economics and Business, Athens.

Du, L. (2007). Acquiring Competitive Advantage in Industry through Supply Chain Integration: A Case Study of Yue Yuen Industrial Holdings Ltd; *Journal of Enterprise Information Management,* 20(5), 527-543.

Dyer, J. (1996). Specialized Supplier Networks as a Source of Competitive Advantage: Evidence from the Auto Industry. *Strategic Management Journal*, 17(4), 271-291.

Esteves, J. (2009*).* A Benefits Realization Road-Map Framework for ERP Usage in Small and Medium-Sized Enterprises. *Journal of Enterprise Information Management, 22(1/2), 25-35.*

Fawcett, S. & Magnan, G. (2002). The Rhetoric and Reality of Supply Chain Integration. *International Journal of Physical Distribution &Logistics, 32(5), 339-361.*

Federici, T. (2009). Factors Influencing ERP Outcomes in SMEs: A Post-Introduction Assessment. *Journal of Enterprise Information Management, 22(1/2), 81-98.*

Ferguson, B. (2000). Implementing Supply Chain Management. *Production and Inventory Management Journal*, 2(2), 64-67.

Fernie, J. (1995). International Comparisons of Supply Chain Management in Grocery Retailing. *Service Industries Journal***,** 15(4), 134-147

Filley, A; House, R; & Kerr, S. (1971). Relation of Leader Consideration andInitiating Structure to R & D Subordinates' Satisfaction. *Administrative Science Quarterly, 6, 19-30.*

Fillis, I; Johanson, U; & Wagner, B. (2003). A Conceptualization of the Opportunities and Barriers to E-Business Development in the Small Firms. *Journal of Small Business and Enterprise Development,* 10(1), 336-344.

Forrester, J. (1958). Industrial Dynamics: A Major Breakthrough for Decision-Makers. *Harvard Business Review,* July-August, 37-66.

Frechlich, G; Hoover, H; Liew, W. & Sorenson, P. (1999). Application Framework Issues when evolving Business Applications of Electronic Commerce*, Information Systems,* 24(6), 457-473.

Giaglis, G; Klein, S; & O'Keefe, R. (1999). Disintermediation, Reintermediation, or Cybermediation? The Future of Intermediaries in Electronic Marketplace.

Handfield, R. & Nichols, E. (1999). Introduction to Supply Chain Management**.** Prentice-Hall, Englewood Cliffs, New Jersey.

Hewitt, F. (1999). Information Technology mediated Business Process Management: Lesson from the Supply Chain. *International Journal of Technology Management.* 17(1/2), 37-53.

Humphreys, P; Mclvor, R; & Cadden, T. (2006). B2B Commerce and its Implications for the Buyer-Supplier Interface. *Supply Chain Management: An International Journal.* 11(2), 131-139.

Gengatharen, D. & Standing, C. (2005). A Framework to assess the Factors Affecting Success or Failure of the Implementation of Government-Supported Regional E-Marketplaces for SMEs. *European Journal of Information Systems,* 14, 417-433.

Gray, P. & Watson, H. (1998). *Decision Support in the Data Warehouse.* A Simeon & Schuster Co; Englewood Cliffs, New Jersey: Prentice-Hall.

Johnston, R. & Mak, H. (2000). An Emerging Vision of Internet-enabled Supply Chain Electronic Commerce. *International Journal of Electronic Commerce,* 4(4), 43-59.

Kaplan, S. & Sawhney, M. (2000). E-Hubs: The New B2B Marketplaces. *Harvard Business Review,* 78(3), 97-103.

Kehoe, D. & Boughton, N. (2001). New Paradigms in Planning and Control Across Manufacturing Supply Chains: The Utilization of Internet Technologies. *International Journal of Operations & Production Management,* 21(5/6), 582-593.

Kennerley, S. & Neely, B. (2001). Enterprise Resource Planning: Analyzing the Impact. *Integrated Manufacturing Systems,* 12(2), 103-113.

Koch, C. (2001). ERP and ERP: Realizing a Vision of Process with IT. *Business Process Management Journal,* 7(3), 285-296.

Kowath, N. & Choon, T. (2001). Determinants of Website Development: A Study of Electronic Commerce in Singapore*. Information & Management,* 39(3), 227-242.

Lai, M; Humphreys, P. & Sculli, D. (2001). The Implications of Western Electronic Commerce for Chinese Business Networks. *Industrial Management & Data Systems,* 101(6), 281-289.

Lee, H; Padmanabhan, V; & Whang, S. (1997). Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science,* 43(4), 546-558.

Mason C; Castleman, T. & Parker, C. (2008). Communities of Enterprise: Developing Regional SMEs in Knowledge economy. *Journal of Enterprise Information Management,* 21(6), 571-584.

Matthyssens, P. & Van den Builte, C. (1994). Getting Closer and Nicer: Partnership in the Supply Chain. *Long Range Planning,* 27(1), 72-83.

McAfee, A. (2006). Enterprise 2.0: The Dawn of Emergent Collaboration. *MIT Slogan Management Review,* 47(3), 21-28.

McCommon, B. & Little, R. (1965). Marketing Channels: Analytical Systems and Approaches. In Schwartz, G. (ed).*Science in Marketing*. John Wiley & Sons, Inc; New York.

McKeown, P. (2000). *Information Technology and the Networked Economy.* Boston, M.A: Course Technology Publishing.

Mentzer, J; DeWitt, W; Keebler, J; Min, S; Nix, N; Smith, C; & Zacharia, Z. (2001) Defining Supply Chain Management. *Journal of Business Logistics,* 22(1), 1-25.

Merono-Cerdan A. L., Soto –Acosta P. & Lopez Nicolas C. (2008). Analysing Collaborative Technologies Effect on Performance through Internet Use Orientations. *Journal of Enterprise Information Management,* 21(1), 39-51.

Morgan, R. & Hunt, S. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing,* 58(July), 20-38.

Murison-Bowie, S. (1999). Forms and Functions of Digital Content in Education. In Leer, A. (ed). *Masters of the Wired World.* London: Financial Times, Pitman Publishing.

Norris, G; Hurley, J; Hartley, K; & Dunleavy, J. (2001). *E-Business and ERP: Transforming the Enterprise.* London: John Wiley.

Power, D. & Sohal, A. (2002). Implementation and Usage of Electronic Commerce in Managing the Supply Chain: A Comparative Study of Ten Australian Companies. *Benchmarking: An International Journal,* 9(2), 190-208.

Porter, M. (2001). Strategy and the Internet. *Harvard Business Review,* 79(2), 63-78.

Porter, M. & Millar, V. (1985). How Information gives you Competitive Advantage. *Harvard Business Review,* 63(4). 149-160.

Quinn, F. (2000). The Clock-speed Chronicles. *Supply Chain Management Review,* 3(4), 60-64.

Raisch, W. (2001). *The E-Marketplace: Strategies for Success in B2B E-Commerce.* New York: McGraw-Hill.

Ramdani, B; Kawalek, P; & Lorenzo, O. (2009). Knowledge Management and Enterprise Systems Adoption by SMEs: Predicting SMEs' Adoption of Enterprise Systems. *Journal of Enterprise Information Management,* 22(1/2), 10-24.

Ridgeway, V. (1957). Administration of Manufacturer-Dealer Systems. *Administration Science Quarterly*. March, 464-467.

Sarkis, J. & Sundarraj, R. (2000). Factors for Strategic Evaluation of Enterprise Information Technologies. *International Journal of Physical Distribution & Logistics Management,* 30(3/4), 196-220.

Scalet, S. (2001). The Cost of Secrecy. *CIO Magazine*, July, www.cio.com.

Shiau, W; Hsu P; & Wang, J (2009). Development of Measures to Assess the ERP Adoption of Small and Medium Enterprises. *Journal of Enterprise Information Management.* 22(1/2), 99-118.

Shuen, A. (2008). *Web 2.0: A Strategic Guide.* First Edition, Sebastopol: O' Reilly Media, Inc.

Tapscott, D. & Williams, A. (2006). *Wikinomics: How Mass Collaboration Changes Everything*. New York: Portfolio. *Businesses. **Journal of Management Information Systems**. 15(4), 27-31*

Van Hoek, S. (2001). *E-Supply Chains-Virtually Non-Existing. Supply Chain Management: An International Journal,* 6(1), 21-28.

Walton, S. & Gupta, J. (1995). Electronic Data Interchange for Process Change in an Integrated Supply Chain. *International Journal of Operations and Production Management,* 19(4), 372-388.

Wood, A. (1997). Extending the Supply Chain: Strengthening Links with IT. *Chemical Week.* 159(25), 26.

Yin, R. (1994). *Case Study Research: Design and Methods.* Newbury Park, CA: Sage Publications.

Table 1. Sample Size and Response rates

	Sample Size	Response	Response Rate (%)
Oil Tools Retailers:			
1. Kaycee & Sons Ltd.	27	15	55.6
2. Ghani Int'l.	23	17	73.9
Superstores:			
3. Spot One	20	16	80.0
4. Everyday Supermarket	25	18	72.0
Electronics & Home Appli.:			
5. Value City Electrical	15	9	60.0
6. Obiekwe Electrical	15	7	46.7
Bakery Firms:			
7. Tea-Mate Bread	18	12	66.7
8. Lulus Bread	17	14	82.4
Total	**160**	**108**	

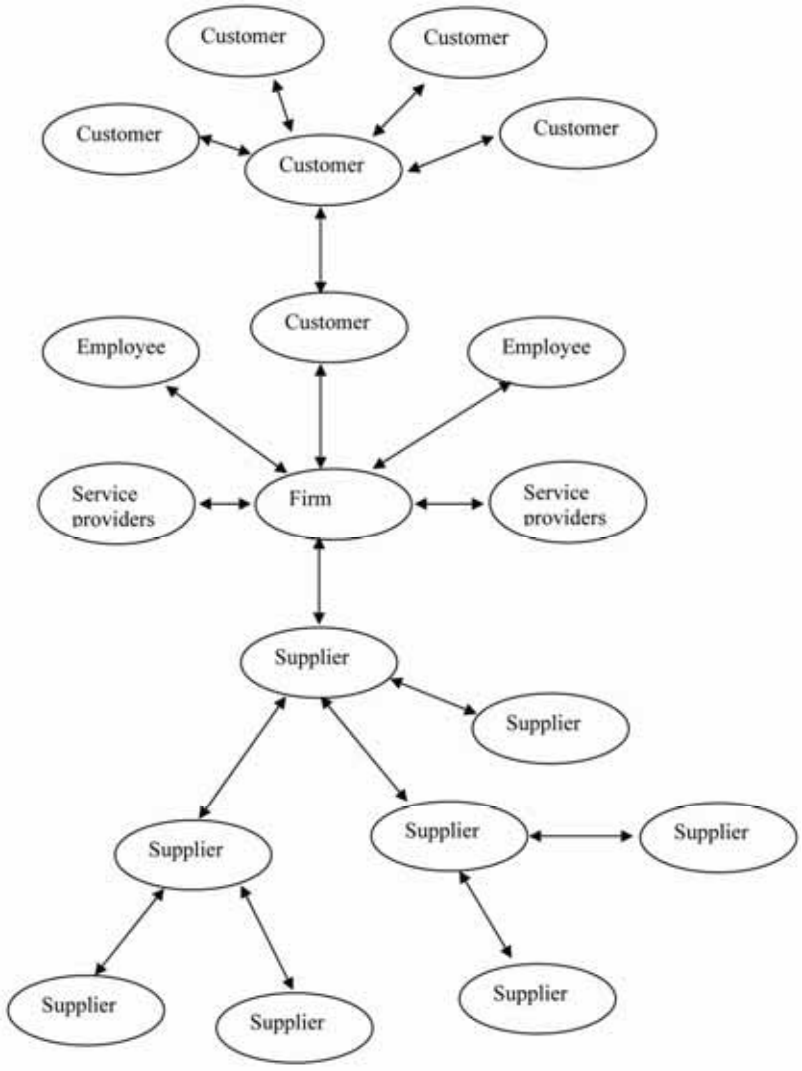


Figure 1. Model of Relationship in Supply Chain

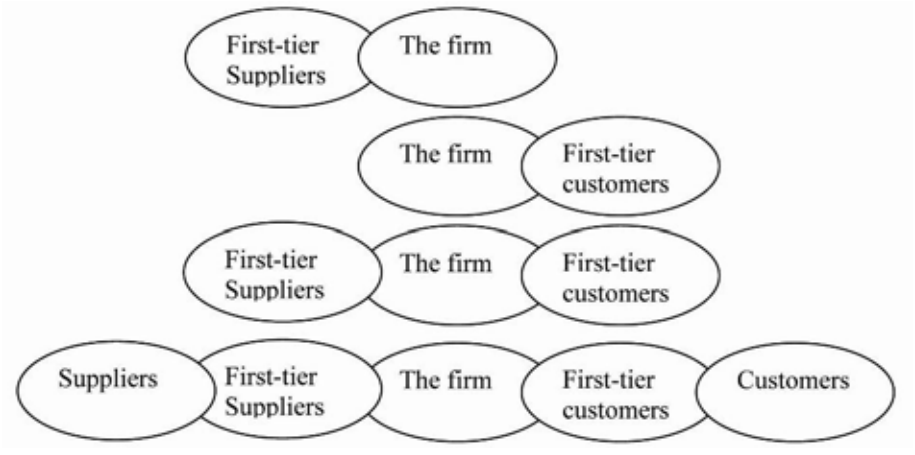


Figure 2. Strategic options in supply chain indicating the status of integration

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Study on the Route Optimization of Military Logistics Distribution in Wartime Based on the Ant Colony Algorithm

Hongtao Shi, Yucai Dong, Lianghai Yi, Dongyun Zheng, Hong Ju & Erchang Ma
Institute of Nonlinear Science, Academy of Armored Force Engineering, Beijing 100072, China

Weidong Li
Institute of Department of Science Research, Academy of Armored Force Engineering, Beijing 100072, China

Abstract

Because of the impacts from the road conditions and the hostile attacking, the military logistics distribution in wartime should consider the pressure degrees of various demand points to the materials, so it is much more complex than the common logistics. Aiming at the actual situation of the military logistics distribution in wartime, the basic ant colony algorithm is improved in the article, and the software of MATLAB is used to simulate the improved algorithm. The result indicates that the algorithm possesses higher solving efficiency, and can fulfill the requirement of the route optimization of the military logistics distribution in wartime.

Keywords: Ant colony algorithm, Pheromone, Route optimization

The military logistics distribution in wartime directly influences the result of the war, and it is the focus to optimize the vehicle routes and enhance the speed and the security of the military logistics distribution for the military logistics industry. In this article, the actual requirements of the military logistics distribution in wartime are comprehensively considered, and the state transfer probability function in the ant colony algorithm is modified, and the unobstructed factor of the route and the importance factor of the demand point are added to make the algorithm be fit for the problem.

1. Description of the route optimization model of the military logistics distribution in wartime

The rout optimization of the military logistics distribution in wartime means to use the minimum time to carry the materials to the appointed place and ensure the safety of the materials by establishing reasonable logistics distribution route. Comprehensively considering various factors, this problem can be described as follows (Xie, 2008).

Use many vehicles to carry materials to the appointed place from the distribution center, and the position and the demand quantity of each demand point are certain, and the load of each vehicle is certain, and the requirements include that the vehicle routes are reasonably arranged to make the transportation routes to be minimum or the transportation time to be minimum, and each demand point should obtain the distribution service in time, and one vehicle can interview the place once, and one vehicle can only serve for one route, and the distribution vehicles start from and return to the distribution center.

In addition, because of some special situations such as vehicle condition, road condition and hostile attacking, the military logistics distribution in wartime must consider the compulsory places, the prohibitive places, the compulsory routes and the prohibitive routes to ensure the safety of the materials. After fully considering the restriction conditions and the optimized objects of above problems, the mathematical model of the route optimization of the military logistics distribution in wartime is established as follows (Zhu, 2006 & Zhang, 2008, P.83-86). First, define variables as follows.

$$y_{k_i} = \begin{cases} 1 \text{ the distribution of the demand point i is accomplished by the vehicle k} \\ 0 \qquad\qquad\qquad\qquad\qquad \text{or else} \end{cases}$$

$$x_{ijk} = \begin{cases} 1 \text{ vehicle k runs from the demand point i to the demand point j} \\ 0 \qquad\qquad\qquad\qquad\qquad \text{or else} \end{cases}$$

The objective function is

$$\min z = \sum\sum\sum d_{ij} x_{ijk} \tag{1}$$

$$\text{s.t.} \sum_i g_i v_{k_i} \le q \qquad \forall k \tag{2}$$

$$\sum_i y_{k_i} = 1 \qquad i = 1,\cdots,l \tag{3}$$

$$\sum_i x_{kijk} = y_{ki} \qquad j = 0,1,\cdots,l \qquad \forall k \tag{4}$$

$$\sum_j x_{kijk} = v_{ki} \qquad i = 1,\cdots,l \qquad \forall k \tag{5}$$

Where, d_{ij} denotes the route length from the demand point i to the demand point j, g_i denotes the transportation quantity of the task i, and q is the vehicle load. The formula (2) is the capacity restriction of the vehicle, the formula (3) ensures that the transportation task of each client point is accomplished by only one vehicle, and all transportation tasks are accomplished by m vehicles together, and the formula (4) and the formula (5) limit that there is one and only one vehicle to arrive at and depart from certain one client point.

2. Solving the problem of the military logistics support route optimization in wartime by the ant colony algorithm

2.1 Introduction of the ant colony algorithm

The ant colony algorithm is the simulated evolution algorithm based on the research result of the real ant colony collective behaviors in the nature, and it has many features such as positive feedback, parallel computation and strong robustness, and it is effective to use this algorithm to solve the problem of route optimization (Li, 2004). This algorithm is mainly decided by two formulas, i.e. the state transfer probability and the pheromone updating. Suppose that *m* is the total amount of ant, *n* is the amount of the route node, and $p_{ij}^k(t)$ is the state transfer probability that the ants transfer from the demand point *i* to the demand point *j* at the time of *t*.

At the initial time, the amount of pheromone on each route is equal. Suppose that $\tau_{ij}(0) = C$, and C is the constant, and the ant $k(k = 1,2,\ldots,m)$ decides its transfer direction according to the pheromones on various routes in the movement, and the tabu list $tabu_k(k = 1,2,\ldots,m)$ is used to record the demand points that the ant goes across, and the set is dynamically adjusted with the evolution process of $tabu_k$. In the searching process, the ant computes the state transfer probability according to the pheromone and the illumination information on various routes. At the time of *t*, the state transfer probability $p_{ij}^k(t)$ of the ant *k* transferring from the demand point *i* to the demand point *k* is

$$p_{ij}^k(t) = \begin{cases} \dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum\limits_{s \in allowed_k} \tau_{is}^{\alpha}(t)\eta_{is}^{\beta}(t)}, & s \in allowed_k \\ 0, & s \notin allowed_k \end{cases} \tag{6}$$

Where, $allowed_k = \{N - tabu_k\}$ denotes the set of the demand points that the ant is allowed to select next, α is the information illumination factor which denotes the relative importance of the track and reflects the function of the accumulated pheromones to the movement for the ant, and its value is bigger, and the ant is more inclined to select the routes that other ants have gone across, and β is the anticipated illumination factor which denotes the relative importance of the visibility, and reflects the importance of the illumination information to the route selection in the movement for the ant, and its value is bigger, and the probability that the ant selects the demand point near it is higher, and the expression of the illumination function $\eta_{ij}(t)$ is $\eta_{ij} = \dfrac{1}{d_{ij}}$, and d_{ij} denotes the distance between two neighboring demand points. For the ant *k*, d_{ij} is smaller, $\eta_{ij}(t)$ is bigger, and $p_{ij}^k(t)$ is bigger.

To avoid much more residual pheromones submerge the illumination information, after each ant goes across one step or accomplishes the traverse of all *n* demand points, i.e. one cycle ends, the residual pheromones should be updated. At the time of $t + n$, the amount of pheromone on the route (i, j) can be adjusted according to following rules.

$$\tau_{ij}(t + n) = (1 - \rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}(t) \tag{7}$$

$$\Delta\tau_{ij}(t)=\sum_{k=1}^{m}\Delta\tau_{ij}{}^{k}(t) \tag{8}$$

Where, ρ denotes the dispersing factor, so $1-\rho$ denotes the residual coefficient of pheromone, and to prevent the infinite accumulation of information, ρ is limited in $\rho\in(0,1)$, and $\Delta\tau_{ij}(t)$ denotes the increment of pheromone on the route (i,j) in this cycle, and the initial time is $\Delta\tau_{ij}(0)=C$, and $\Delta\tau_{ij}^{k}(t)$ denotes the amount of pheromone that the ant *k* leaves on the route (i,j) in this cycle.

The Ant-Cycle model is adopted to deal with the route optimization in this article, and in this model,

$$t_{ij}^{k}=\begin{cases}\dfrac{Q}{L_k}, & \textit{if the ant k goes across (i, j) between t and t + 1}\\ 0, & \textit{or else}\end{cases}$$

Where, Q denotes the intension of the pheromone, and it is constant, and it can impact the constringency speed of the algorithm to certain degree, and L_k denotes the total length of the route that the ant *k* goes across in this cycle, and it equals to the sum of various routes d_{ij} that the ant *k* goes across.

2.2 Improvement of the state transfer probability function

The state transfer function of the basic ant colony algorithm only considers the illumination factor and the anticipation factor, and combining with the actual situation of the urban traffic condition, Xie Min et al (Xie, 2008) added two factors in the function, but the route optimization of the military logistics distribution in wartime must consider the compulsory places, the prohibitive places, the compulsory routes and the prohibitive routes because of some special situations such as vehicle condition, road condition and hostile attacking, and the importance degrees of various demand points are different, so the demand points with higher importance degree should be first fulfilled.

Comprehensively considering above factors, two illumination factors, w_{ij} and z_{ij}, are added in the state transfer probability function, and a improved state transfer rule can be obtained as follows.

$$p_{ij}^{k}(t)=\begin{cases}\dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)w_j^{\gamma}z_{ij}^{\theta}}{\sum_{s\in allowed_k}\tau_{is}^{\alpha}(t)\eta_{is}^{\beta}(t)w_s^{\gamma}z_{is}^{\theta}}, & s\in allowed_k\\ 0, & s\notin allowed_k\end{cases} \tag{10}$$

Where, $w_j(t), j=0,1...n$ denotes the importance degree of the next demand point j at the time of t, and it shows the demand degree of the demand point j to the materials at the time of t, and $w_j(t)\in[0,1]$, and when the value of $w_j(t)$ is bigger, the demand of the demand point j to the materials is more urgent, and the materials should be carried to the demand point as soon as possible, so the selected probability is higher, and $z_{ij}(t)$ denotes the unobstructed degree of the road (i,j) at the time of t, and $z_{ij}(t)\in[0,1]$, and when $z_{ij}(t)=1$, the road (i,j) has not been impacted by the war at the time of t and the road condition is in the optimal state, and when $z_{ij}(t)=0$, the road (i,j) is destroyed very badly, or is forbade to entry because of hostile firepower, and in *(0,1], $z_{ij}(t)$ is bigger, the road (i,j) is wider and the probability that this road is selected is higher.

2.3 Improvement of the object function

The route optimization of the military logistics distribution in wartime should not only require the minimum route length, but use the minimum time to accomplish the task when ensuring the safety of materials. Based on the established route optimization model of the military logistics distribution, the time factor is added into the objective function to obtain the improved objective function model as follows.

$$\mathrm{F}=\min(a\sum_{\mathrm{i}=tabu_k(0)}^{\mathrm{tabu_k(s\text{-}1)}}\mathrm{d}_{\mathrm{i,i+1}}+b\sum_{i=tabu_k(0)}^{tabu_k(s-1)}t_{i,i+1}) \tag{11}$$

Where, $\sum_{i=tabu_k(0)}^{tabu_k(s-1)}d_{i,i+1}$ is the length of the planed route, $\sum_{i=tabu_k(0)}^{tabu_k(s-1)}t_{i,i+1}$ is the passing time of the planned route. And t_{ij}

denotes the time required to pass the road (i, j), and in the article, $t_{ij} = \frac{v \cdot d_{ij}}{z_{ij}}$, and v is the vehicle speed when the road condition is good and there are not dangerous conditions such as hostile attacking, and it is a constant. $tabu_k(0)$ denotes the initial point, and $s = length(tabu_k)$ denotes the amount of the factor in the tabu list, i.e. the amount of the demand point, a and b are the balance of the route length and the passing time, and $a + b = 1$. The route length and the passing time are anticipated to be short. In practice, the planned route is a optimal combination combining above two factors.

3. Example of the algorithm

3.1 Basic steps of the algorithm

According to the above model, the route optimization model of the military logistics distribution in wartime includes following steps.

Step 1. Initialize the pheromones on each side as a very small constant. The start point of each ant is the distribution center, and set the distribution center into the tabu list at the same time.

Step 2. Judge whether the ant offer service for all clients, and if the service is accomplished, the ant returns to the distribution center and complete the construction of the route. Or else, select the next client, and judge whether the remnant capacity of the vehicle can fulfill the demand of the selected client. If it can be fulfilled, the ant serves for the selected client, and set the selected node into the tabu list. Then, update the pheromones on the corresponding side. If it can not be fulfilled, give up the selected client and return to the distribution center, and send another vehicle to serve again.

Step 3. Update all pheromones. Find the optimal route in this traverse and update all pheromones.

Step 4. If the setting search times have not used up, clear up the tabu list, and repeat above steps.

3.2 Experiment and analysis

Suppose that one oil warehouse needs transport oils to 19 demand points which are distributed in the square with the border length of 10 kilometer. The oil warehouse is in the midpoint of this region, and its coordinate is (0, 0). Suppose that the distribution center is the No.1 point, and the demands of various demand points are seen in Table 1, and the capacity of the vehicle is 9 tons. And the unobstructed factor of the road adopts the random data.

The adopted parameters include $m = 10$, $NC = 100$, $\tau_{ij}(0) = 0.01$, $\alpha = 1$, $\beta = 1$, $Q = 10$, $\rho = 0.6$, $\gamma = 0.7$ and $\theta = 0.3$. Run all steps ten times, and the optimal solution is 44.49km, and the corresponding route is

Route 1: 1→13→8→7→4→20→2→10→1;

Route 2: 1→16→6→14→17→12→11→3→19→1;

Route 3: 1→9→15→18→5→1.

It is obvious that the computation result of the ant colony algorithm is stable and avoids dangerous roads. So for the route optimization of the limitary logistics distribution in wartime, the improved ant colony algorithm can obtain ideal result.

4. Conclusions

The ant colony algorithm is a new bionic algorithm in the optimization domain. In this article, it is applied in the military logistics distribution, and the transfer probability function and the object function are improved. The experiment result indicates that this algorithm can quickly find the optimal solution and avoid the dangerous routes, and the calculation result is stable, and it can fulfill the requirements of the route optimization of the military logistics distribution in wartime.

References

Ge, Panpan, Wang, Jirong & Li, Jun. (2008). Research of the Routing Optimization for Container Trucks in Dock with Ant Colony Algorithm. *Logistics Sci-Tech.* No.12.

Li, Shiyong. (2004). *Ant Colony Algorithm and Its Application*. Harbin: Harbin Institute of Technology Press.

Xie, Min & Gao, Lixin. (2008). Ant Algorithm Applied in Optimal Path Planning. *Computer Engineering and Applications.* No.44(8).

Zhang, Lifeng, Zhao, Fanggeng & Sun, Jiangsheng. (2008). Vehicle Scheduling Problem of Military Distribution Based on the Ant Colony Algorithm. *Computer and Communications.* No.26(6). P.83-86.

Zhu, Jianrong. (2006). *Study on the Vehicle Optimization Transfer of Logistical Distribution Based on the Ant Colony*

Algorithm. Thesis of Changsha University of Science and Technology.

Table 1. Coordinates and demands of various nodes

No. of node	1	2	3	4	5	6	7	8	9	10
Abscissa	0	0	0	-2	-3	3	-4	-4	1	1
Ordinate	0	-1	3	-2	-3	-1	0	-1	-2	-1
Distributing quantity/t	0	1.5	1.8	2.0	0.8	1.5	1.0	2.5	3.0	1.7
No. of node	11	12	13	14	15	16	17	18	19	20
Abscissa	1	3	-3	2	1	2	2	1	-3	-1
Ordinate	3	4	0	0	-3	-1	1	-4	2	-1
Distributing quantity/t	0	0.2	2.4	1.9	2.0	0.7	0.5	2.2	3.1	0.1

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Exposure of Computer Games among IHL Students in Malaysia: Case Study of Computer Science Students in UiTM Terengganu

Hasiah Mohamed @ Omar
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA (UiTM),
Dungun Campus, 2300 Dungun, Terengganu, Malaysia
Tel: 60-9-840-3835 E-mail: hasia980@tganu.uitm.edu.my

Nora Yanti Che Jan
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA (UiTM)
Dungun Campus, 2300 Dungun, Terengganu, Malaysia
Tel: 60-9-840-3957 E-mail: noray084@tganu.uitm.edu.my

Nik Marsyahariani Nik Daud
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA (UiTM)
Dungun Campus, 2300 Dungun, Terengganu, Malaysia
Tel: 60-9-840-3770 E-mail: nikma944@tganu.uitm.edu.my

Abstract

A survey on users' exposure of computer games among students of Institute of Higher Learning (IHL) has been done to gather useful information and relevant data regarding this study. This exploratory study focused on how exposure to computer games influenced students' academic performance in two computer science related subjects from the student's point of view. These two subjects were Fundamental of Computer Science (CSC119) and Introduction to Problem Solving (CSC125). The data was gathered from the questionnaire distributed among the whole population of Computer Science students from October–July 2007 intake (67 respondents) and then the data was analyzed using Statistical Package (SPSS version 12). From the analysis done, it is shown that respondents with early exposure score higher when compared to those with the late exposure.

Keywords: Exposure, Computer games, Computer related subject, Opinion

1. Introduction

Computer games have been a preference to most of computer users. Nowadays, it does not only focus on computer that we normally used in our office or house, but handheld devices such as handphone and Personal Digital Assistant (PDA) have also been a target platform for users to play computer games. Computer games come in various types to suit different user's requirement such as role playing game, educational game, sport game and entertainment game. Researches reveal that the focal group in computer games is adolescents (Griffiths & Hunt, 1995; Subrahmanyam, Greenfield, Kraut, & Gross, 2001).

There are many arguments regarding advantages and disadvantages of computer games towards users especially teenagers (Griffiths & Hunt, 1995; Subrahmanyam et al., 2001). List of the advantages starts with computer games play an important role in education and it is adequate enough to assist many people, namely students, with their learning. Learning process in computer games can be achieved in formal and informal approach. Research suggests that computer games can actually be educative, beneficial and does help in human learning and development activities. Researchers have found that people who play computer games have better vision and computer games also assist students in improving hand-eye coordination and improve reflexes (Dorman, 1997; Emes, 1997; Tazawa & Okada, 2001; Tazawa,

Soukalo, Okada, & Takada, 1997). Computer games also improve logical thinking and problem solving skills (Benton, 1995; Colwell, Grady, & Rhaiti, 1995; Gentile & Walsh, 2002; Pillary, Brownlee, & Wilss, 1999). In the years that have followed, researchers found that educational software and games can indeed have several very positive effects on students' academic skills (Banks & Miller, 2001; Greenfield, DeWinstanley, Kilpatrick, & Kaye, 1996). As for the disadvantages, computer games have been discovered to cause eyestrain and chest pain (Tazawa et al., 1997) and sleep deprivation associated with black ring under the eyes and muscle stiffness in the shoulders (Tazawa & Okada, 2001) to the players. Heavy computer gamers are also less sociable (Cowell & Stanney, 2005; Roe & Muijs, 1998). The increment in numbers of computer gamers is contributed by two main factors, which are the (1) accessibility of the computer games itself in terms of the software and platform and (2) variation of games categories offered in the market. Students are exposed to computer games not just as an entertainment but it has become a type of social tools. Derived from the trends and exposure are the curiosity that are very interesting to be explored in order to classify details on how students view it. Curiosities in our focus are interest, effects and exposure. However, in this paper only one curiosity will be highlighted, which is the exposure to computer games.

1.1 Computer games

The use of computer games nowadays has become popular among users despite their age and gender. This type of entertainment has become important to the users because of its functions – to release stress, learning process tools and there are various categories that can be chosen based on player's preferences, social life medium and popular leisure time activities (Poole, 2000; Vorderer & Bryant, 2006). Computer games are defined as a set of activities involving one or more players. It has goals, constraints and consequences (Dempsey, Lucassen, Haynes, & Casey, 1997). Elaborating further on that, Crawford (1984) said that a game is rule-guided and artificial in some aspects, and also involves some aspects of a contest or a trial of skill or ability, even if that contest is with oneself. According to him, there are four common factors that present in all games. These factors are representation, interaction, conflict and safety. So when the computer is used to present a game and to act as an opponent or as a referee, we have a computer game.

1.2 Exposure to computer games

Interactive media, such as video and computer games, have redefined today's leisure activities. These widely available games may be played on dedicated consoles systems, any computer and over the Internet, handheld devices and various technology toys. Several recent reports confirm that computer games have a significant present in the lives of children and adolescents (Buchman & Funk, 1996; Gentile & Walsh, 2002; Roberts, Foehr, Rideout, & Brodie, 1999) & Frunk, 1996; Gentile & Walsh, 2002; Roberts, Foehr, Rideout & Brodie, 1999; (Wright et al., 2001) and boys are avid consumers of computer games, playing nearly twice as frequently as girls (Roberts et al., 1999) also reported that children are spending an increasingly amount of time playing computer games. Computer games units are now present in 83% of homes with children. In 2004 alone, children spent 49 minutes per day playing computer games and on any given day, 52% of children aged 8 – 18 years play games. Games use peaks during childhood, with an average of 65 minutes per day for 8 – 10 years old and declines to 33 minutes per day for 15 – 18 years old. A study conducted by (Gentile & Walsh, 2002) found that the average American child aged 2-17 years plays games for 7 hours a week. This average marks wide differences between boys and girls and children in different ages. For example, in this study it is found that adolescent girls played games for an average of 5 hours a week whereas boys averaged 13 hours a week.

2. Methodology

In this exploratory study, 67 respondents of Part One Computer Science students in UiTM Terengganu were asked their opinions towards how the exposure to the computer games relate to their performance in two computer science related subjects (CSC119 and CSC125). The first phase in this exploratory study is planning and analysis where the problem mentioned earlier is to investigate study's significance. Analysis phase is being executed to find a suitable focus group and the best way to represent influences, effects and exposure of computer games in students' lifestyle. To ensure that the right methods are used, a thorough literature review was done during this phase.

The second and third phase are about designing and developing specific questionnaire to obtained data from students. In order to design the questionnaire, a preliminary survey has been done to find types of computer games that are played by the students. Developing questionnaire involved a few processes where the questionnaire is divided into five parts. Part A involves in collecting demographic data, while part B is to seek for categories of computer games that students prefer to play. Part C is dedicated for exposure and frequencies. Part D focuses on students' interest towards computer games and part E seeks for students' opinion on effects of computer games towards their lives. The questionnaires were distributed to students by the end of the semester. The distribution process is done with assistance from colleagues. Grades for each student in two computer related subjects, Fundamental of Computer Problem Solving (CSC125) and Fundamental of Computer Science (CSC119), are then collected to find any correlations. The data analysis phase involves analyzing data where the data obtained from the questionnaire and students' grade in the computer science related subjects by using statistical package which is SPSS version 12. This exploratory study is later documented as a finalized report.

2.1 Research tool

Statistical methods used in this study to evaluate student response were based on descriptive technique. This technique was used to fit in with the requirement of the study which was categorized as an exploratory study. Sample for this study was taken from the whole population of Computer Science students from July-October 2007 intake. This was based on their level of exposure on academic life and to ensure that their knowledge toward computer usage was not influenced by their studies in UiTM. Questionnaire was used as the primary data collection tool in this exploratory study despite the other research methods such as interview and observation. The main reason for using questionnaire is because it can cover a wide range of sample and it is relatively cheap compared to other methods. Preliminary survey was conducted prior to the questionnaire. In our preliminary survey, a list of frequent play computer games was distributed to the students.

2.2 Premilinary study

A preliminary survey had been distributed to students at the beginning of semester July-October 2007. The survey was to seek for types of computer games that students frequently played. Students were asked to list down 10 of their most favorite's computer games. Based on the given answers, a list of computer games in Part B of the questionnaire is constructed. In this survey, all respondents are required to list 10 computer games that they often play. The list of the collected games was then divided into a few categories, which were based on literature review where different types of computer games categories had been discussed. All the computer games were divided into six main categories, which are action games, card and arcade games, role-playing and adventure games, sport games and strategy games.

2.3 Questionnaire

Questionnaire is divided into five main parts. Questions in Part A are prepared using open-ended question while Part B is constructed by listing out computer games based on category. Part B in the questionnaire was constructed using a list of computer games that are frequently played by this focus group. These computer games were selected based on the preliminary survey which has been explained in previous section. Not all games were selected as only 21 out of 49 listed games are listed in Part B. The total number of computer games in each category was varied as it was based on the students' survey result. From the survey, it seemed that most of the games picked out can be categorized as role-playing games. For each category, half of the listed games were taken to be included into the actual questionnaire. Part B was constructed using this technique to ensure that data collected can be easily analyzed. The list of computer games that are played by students can be controlled, and yet it is still based on the students' choice of computer games. Blank spaces are also provided for students to list any specific computer games if the game is not listed. Questions in part C were designed to be closed-ended questions.

Question in Part C focuses mainly in finding information regarding respondents' exposure and how frequent they play computer games. Question 2 in Part C asks the respondents to state their age when they are first expose to computer games. Age then being categorized based on their level and the categories are as follows: Kindergarten (4-6 years old), Lower Primary (7-9 years old), Upper Primary (10-12 years old), Lower Secondary (13-15 years old) and Upper Secondary (16-18 years old). Questions in Part D refers to students' interest towards computer games while Part E refers on effects of computer games based on students' opinion. Both questions in Part D and E are designed using Likert Scale.

Data that has been collected from the questionnaires were then analyzed using Statistical Packages for Social Sciences (SPSS) version 12. Statistic description used in this research was done based on percentage, frequencies and cross tabulation between students' achievements and their exposure in computer games. Data collected from all part of questionnaire are presented based on their percentage and frequencies while cross tabulation is applied to see the correlation between exposure's age and their results in computer related subjects.

2.4 Data analysis

Data that has been collected from the questionnaires were then analyzed using Statistical Packages for Social Sciences (SPSS) version 12. Statistic description used in this research was done based on percentage, frequencies and cross tabulation between students' achievements and their exposure in computer games. Data collected from all part of questionnaire are presented based on their percentage and frequencies while cross tabulation is applied to see the correlation between exposure's age and their results in computer related subjects.

3. Result and discussion

From the 67 questionnaire distributed, only 65 are returned. This gives the respond rate of 96.8%. Most of the respondents prefer to play games that can be categorized into card and arcade games (80%), followed by action games (69.2%), role playing games (63.1%), puzzle games (50.8%), strategy games (26.2%) and the least played games is in sport games category (21.5%). This clearly shows that respondents have tendency towards computer games that challenge minds but in a more relaxing situation. A game that needs concentration and focus is more likely to be

avoided.

3.1 Exposure and frequency

From the 65 returned questionnaire, it was found that 100% of the respondents are exposed to computer games. Table 1 shows in details percentage of age in which the respondents started to play computer games. From the table, majority of the respondents (36.8%) were first introduced to computer games during their upper primary year which is 10 to 12 years old. Another 21.1% of the respondents came from a younger age level which is 7-9 years old. Only 7% of the respondents started playing computer games at the age of 16 to 18 years old. This showed that most of the respondents have been exposed into computer games since their early age which is referring to 4-12 years old. When asked about their source of exposure, 67.2% of the respondents said they were introduced to computer games by friends. 23.4% mentioned there were exposed to computer games by the family members, 3.1% said through school and 6.3% mentioned other sources. This clearly shown that their environment is quite supportive on encouraging the students to play computer games. Table 2 shows the details.

3.2 Result analysis

Comparison was made based on relation between respondents' result with their exposure age toward computer games. The cross tabulation can only be made using 57 respondents' data due to missing data on exposure age for 8 respondents (Part C question 2). Table 3 shows the cross tabulation between the exposure age and grade in CSC125 subject. Though early exposure in playing computer games cannot assure that the respondents will achieve a better grade in this subject, it seemed that respondents that are exposed to computer games at a later age tend to score lower. This can be observed based on the result where respondents that are exposed during their kindergarten, junior primary and senior primary (23 out of 57) get good grades while only 6 respondents from junior secondary and senior secondary manage to get good grades.

Table 4 shows the cross tabulation between the exposure age and grade in CSC119 subject. Respondents' marks seemed to be scattered all over the age range, but from an observation made on the table, 48 out of 57 respondents get good grades when they were exposed during their kindergarten, junior primary and senior primary phase. In contrast, there are only 9 respondents that manage to get good grades when exposed during their junior secondary and senior secondary phase.

4. Conclusion

Figure 1 shows the questionnaire that is used to collect data related to computer games exposure towards students in IHL. The questionnaire is divided into five parts. Part A is a section focusing on collecting data related with demographic. These data can be used to provide information such as gender and age and how these variables can be related to students' tendency on playing games. Part B or Category of Games is introduced to provide information regarding types of games that students prefer to play. Part C or Exposure and Frequency is prepared to find information on students' background and how can this be related to students' tendency in playing games. Part D is prepared to search for information on student interest toward computer games. This part is provided to validate on theory that students are influenced by computer games. Part E is the last section that is used to find out on students opinion on computer games.

This exploratory study is conducted to answer our curiosity about the relation between exposure to computer games and academic performance among adolescents. Scope of this study involved 67 Part One Computer Science students in UiTM Terengganu and two computer science related subjects which are CSC119 and CSC125. Student's opinions were gathered in order to answer the curiosity mentioned using research instrument which is questionnaire. At the final phase of the study, students' results for the two subjects were being analyzed to identify the relationship between students' opinion and their academic performance. It is proved that early exposure to computer games does help in students' academic performance.

References

Banks, J. S., & Miller, N. R. (2001). *Signaling games in political science Fundamentals of pure and applied economics*. London: Routledge.

Benton, P. (1995). Conflicting cultures: Reflections on the reading and viewing of secondary-school pupils. *Oxford Review of Education, 21*(4), 457-470.

Buchman, D. D., & Funk, J. B. (1996). Video and computer games in the '90s: children's time commitment and game preference. *Children Today, 24*(1), 12.

Colwell, J., Grady, C., & Rhaiti, S. (1995). Computer games, self-esteem and gratification of needs in adolescents. *Journal of Community & Applied Social Psychology, 5*(3),195-206.

Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human-Computer Studies, 62*(2), 281-306.

Crawford, C. (1984). *The Art of Computer Game Design.* Berkely, USA: Osborne/McGraw-Hill.

Dempsey, J. V., Lucassen, B. A., Haynes, L. L., & Casey, M. S. (1997). An exploratory study of forty computer games (COE Technical Report No 97-2). *Mobile, AL. University of South Alabama*.

Dorman, S. M. (1997). Video and computer games: Effect on children and implications for health education. *Journal of School Health, 67*(4), 133-138.

Emes, C. E. (1997). Is Mr Pac Man eating our children? A review of the effect of video games on children. *Canadian journal of psychiatry. Revue canadienne de psychiatrie, 42*(4), 409.

Gentile, D. A., & Walsh, D. A. (2002). A normative study of family media habits. *Journal of Applied Developmental Psychology, 23*(2), 157-178.

Greenfield, P. M., DeWinstanley, P., Kilpatrick, H., & Kaye, D. (1996). Action video games and informal education: Effects on strategies for dividing visual attention. *Interacting with video, 11*, 187-205.

Griffiths, M. D., & Hunt, N. (1995). Computer game playing in adolescence: Prevalence and demographic indicators. *Journal of Community & Applied Social Psychology, 5*(3),189-193.

Pillary, H., Brownlee, J., & Wilss, L. (1999). Cognition and recreational computer games: Implications for educational technology. *Journal of Research on Computing in Education, 32*(1), 203-216.

Poole, S. (2000) *Trigger Happy: The Inner Life of Videogames.*London: Fourth Estate.

Roberts, D. F., Foehr, U. G., Rideout, V. J., & Brodie, M. (1999). Kids & media@ the new millennium. *Menlo Park, CA: Kaiser Family Foundation*.

Roe, K., & Muijs, D. (1998). Children and computer games: A profile of the heavy user. *European Journal of Communication, 13*(2), 181.

Subrahmanyam, K., Greenfield, P., Kraut, R., & Gross, E. (2001). The impact of computer use on children's and adolescents' development. *Journal of Applied Developmental Psychology, 22*(1), 7-30.

Tazawa, Y., & Okada, K. (2001). Physical signs associated with excessive television-game playing and sleep deprivation. *Pediatrics International, 43*(6), 647.

Tazawa, Y., Soukalo, A. V., Okada, K., & Takada, G. (1997). Excessive playing of home computer games by children presenting unexplained symptoms. *Journal of Pediatrics, 130*(6), 1010.

Vorderer, P., & Bryant, J. (2006). *Playing Video Games: Motives, Responses, and Consequences.*NJ: Lawrence Erlbaum.

Wright, J. C., Huston, A. C., Vandewater, E. A., Bickham, D. S., Scantlin, R. M., & Kotler, J. A., et al. (2001). American children's use of electronic media in 1997: A national survey. *Journal of Applied Developmental Psychology, 22*(1), 31-47.

Table 1. Age starting to play computer games

Level	Frequency	Percentage
Kindergarten (4-6 years old)	8	14.0
Lower Primary (7-9 years old)	12	21.1
Upper Primary (10-12 years old)	21	36.8
Lower Secondary (13-15 years old)	12	21.1
Upper Secondary (16-18 years old)	4	7.0

Table 2. Source of computer games

Source	Percentage
Friends	67.2
Family	23.4
School	3.1
Others	6.3

Table 3. Relation between exposure age and CS125 grade

		Exposure age				
		Kindergarten	Junior Primary	Senior Primary	Junior Secondary	Senior Secondary
CSC125 Grade	A+	0	1	0	0	0
	A	2	0	0	1	0
	A-	2	1	1	0	0
	B+	1	3	4	2	0
	B	0	1	3	0	0
	B-	1	1	2	3	0
	C+	0	1	5	1	0
	C	0	2	2	0	3
	C-	1	0	1	3	0
	D+	0	1	0	1	0
	D	1	1	1	1	0
	E	0	0	2	0	1

Table 4. Relation between exposure age and CS119 grade

		Exposure age				
		Kindergarten	Junior Primary	Senior Primary	Junior Secondary	Senior Secondary
CSC119 Grade	A+	0	1	0	0	0
	A	0	4	3	2	0
	A-	5	2	3	1	1
	B+	1	1	3	2	1
	B	0	2	5	3	1
	B-	0	1	2	3	1
	C+	1	0	4	0	0
	C	1	0	1	0	0
	C-	0	1	0	0	0
	D	0	0	0	1	0

Figure 1 Questionnaire: Students Opinion Towards Influence Of Computer Games: Case Study Of Part One Computer Science Students In UiTM Terengganu

Part A (Demographic Data)

Please fill in the information required below.

Student ID : 2007 _____ Group : _____

Gender : ☐ Male ☐ Female Age : _____

Hometown : _____

Part B (Category of Games)

Select (√) 5 games that you often play from the list below. If the game is not listed, write it in the space provided.

	1. Need for Speed		8. Winning Eleven		15. Half Life
	2. Diner Dash		9. Hangaroo		16. Ragnarok
	3. King of Fighter		10. Minesweeper		17. Grand Theft Auto
	4. Lord of The Ring		11. Bejewelled		18. Daytona
	5. Pinball		12. Zuma		19. FIFA World Cup
	6. Mario		13. Final Fantasy		20. The Sims
	7. Solitaire		14. Counter Strike		21. Command and Conquer

Part C (Exposure and Frequency)

Please tick (√) in the appropriate box.

1. Do you play computer games? If not, go to **Part E**

   ☐ Yes ☐ No

2. When did you start playing computer games? (State your age)

   _____ years old

3. Do you use computer at home?

   ☐ Yes ☐ No

4. Where do you usually play computer games?

   ☐ Home ☐ Cyber Café ☐ Friend's home ☐ Others (Please specify) _____

5. What type of computer games do you usually play?

   ☐ PC games ☐ Online games ☐ Games Console (PS2/PS3/Xbox)

6. Where do you usually get the software for the computer games?

   ☐ Buy ☐ Friends ☐ Download from the Internet ☐ Others (Please specify)_____

7. Who introduced you to computer games?

   ☐ Friends ☐ Family members ☐ School (computer class) ☐ Others (Please specify) _____

8. How often do you play computer games?

   ☐ Everyday ☐ 1 – 2 times/week ☐ 1– 2 times/month ☐ less than once a month

9. How many hours do you normally spend each time you play computer games?

   ☐ Less than 1 hour ☐ 1 < hours < 2 ☐ 2 < hours < 3

   ☐ 3 < hours < 4 ☐ 4 < hours < 5 ☐ more than 6 hours

Part D (Students Interest towards Computer Games)

Circle your answer.

Questions	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. I like to play computer games.	1	2	3	4	5
2. I like to spend most of my time playing computer games.	1	2	3	4	5
3. I do not realize how time goes by while playing computer games.	1	2	3	4	5
4. If possible, I want to play computer games everyday.	1	2	3	4	5
5. I like to try out new computer games.	1	2	3	4	5

Part E (Students Opinion on Computer Games)

Circle your answer.

Questions	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. Playing computer games will help me improve my academic performance.	1	2	3	4	5
2. Playing computer games will increase my critical and logical thinking.	1	2	3	4	5
3. Playing computer games will increase my knowledge on computers.	1	2	3	4	5
4. Playing computer games will help me release stress.	1	2	3	4	5
5. Playing computer games is more beneficial than other leisure activities.	1	2	3	4	5
6. Playing computer games helps me get good results in programming subject (CSC125)	1	2	3	4	5
7. Playing computer games will help me get good results in computer subject (CSC119).	1	2	3	4	5
8. My academic performance will decrease because I play computer games.	1	2	3	4	5
9. I do not have many friends because of computer games.	1	2	3	4	5
10. Playing computer games makes me passive	1	2	3	4	5

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

A Knowledge Innovation Algorithm Based on Granularity

Taishan Yan (Corresponding author)
School of Computer Science and Engineering, Xi'an University of Technology
Xi'an, Shaanxi 710048, China
E-mail: yantaishan163@tom.com

Duwu Cui
School of Computer Science and Engineering, Xi'an University of Technology
Xi'an, Shaanxi 710048, China
E-mail: Cuidw@xaut.edu.cn

The research is supported by the National Natural Science Foundation of China under Grant No. 60873035 and No.60743009, the Natural Science Foundation of Shaanxi of China under Grant No.2006F43.

Abstract

The structure of human knowledge is regarded as granule state by rough sets theory. Granularity is used to denote this structure of knowledge. Knowledge itself evolves ceaselessly as creatures. The mechanism of knowledge evolution includes the productive mechanism for new knowledge and the natural choice mechanism for the selection of the superior and the elimination of the inferior. Knowledge innovation is an important step of knowledge evolution course. Based on knowledge granularity, a knowledge innovation method is proposed in this paper. The main idea of this method is to constitute the partition granularity of knowledge base space ceaselessly depend on the measure consistency of attribute, till the sort of every granules in the granularity is only one. In the algorithm, the only one computation work is to measure the consistency of attributes in knowledge base space. So the numerical calculation is little and the time complexity is low. Experiments were taken on the imperfect knowledge base space of day weather classification by this algorithm. The working course of the algorithm was explained in the example. The successful results show that this algorithm is valid and feasible.

Keywords: Knowledge innovation, Imperfect knowledge base, Granule, Granularity, Consistency

1. Introduction

Knowledge is one of human being's own concept which belongs to epistemology. It is the accumulation or induction of experience and contains almost all the meaning of data, information, knowledge and intelligence (Zhu, 2000; Zhong, 2007). In the course of human evolution, knowledge plays a vital role. Humans have to obey the natural law, at the same time, formed a knowledge system gradually. The knowledge system had a great influence on human evolution and made human evolution rise rapidly. The famous science-philosopher Karl · Popper (Liu *et al*., 2007; Shu *et al*., 2005; Fu *et al*., 2005) thought that the selection of scientific theory is similar to Darwin's natural selection. Our knowledge is always constituted by hypothesis, and passed in the struggle to survive down and showed its relative adaptability. The hypothesis which didn't adapt was eliminated in the competition. Karl · Popper put biological evolution together with scientific development, explained how the science evolve and created his knowledge evolutionism. Knowledge evolutionism tells us that knowledge can be regarded as a changing objective knowledge world, in which there is no eternal, immutable knowledge. That is to say, knowledge is incomplete, knowledge itself evolves ceaselessly. There are two mechanisms in knowledge evolution, one is the productive mechanism for new knowledge, and the other is the natural choice mechanism for the selection of the superior and the elimination of the inferior. Knowledge evolution has its own characteristics and rules. In the course of knowledge evolution, knowledge innovation is a very important aspect.

At present, the research on natural computing theory largely focused on the biological natural selection level ((Liu *et al*.,

2007; Fogel, 1994). However, the research on knowledge evolution theory and algorithm is bound to become a new research direction of natural computation. In order to provide a preliminary exploration for knowledge innovation and support the research on knowledge evolution algorithm, a knowledge innovation method based on knowledge granularity was proposed in this paper.

2. Granularity Concept

Knowledge is examined from a new view in Z Paw Lak's rough set theory (PAWLAK, 1991), and it is generally regarded as a granular structure which is expressed by granularity. Granularity is a very important conception in intelligent information processing. With the help of granularity, knowledge mining and knowledge processing can be carried through from a different perspective, knowledge innovation and knowledge evolution can be realized.

Definition 1 A granule can be defined as any subset, class, object or cluster of a universe (Yao, 2008). A basic granule is considered to be indivisible or there is no need to divide. Elements in a granule are drawn together by indistinguishability, similarity, proximity, functionality, etc. Different fields have different units corresponding to granules (Yao, 2007; Yi *et al.*, 2008). That is to say the elements of a granule are regarded as a cluster, which is accordant with the basic viewpoint of rough set theory (PAWLAK, 1982,1991; Dun *et al.*, 2006; Yang *et al.*, 2006; QUINLAN, 1987).

Definition 2 Giving a granule set, and all the elements of a given field can be found in a element of the granule set, then the set can be called granularity of the given field (Dun *et al.*, 2006). Granularity calculation is knowledge-oriented, which is different from traditional numerical calculation. Traditional numerical calculation is data-oriented (Bernard *et al.*, 2006).

How to construct the granularity of a given knowledge base space? Combine the view of rough set and decision-making tree, the granularity of a given knowledge base space can be constructed according to the attribute value. Given a knowledge base space *(U, C, D, V, f)*, where, *U* is the set of objects, *C* is the set of condition attributes, *D* is the set of decision attributes. $V = \bigcup_{a \in A} V_a$, V_a is the value region of attributes. $f : U \times (C \bigcup D) \to V$ is a evaluating function, it gives each attribute of all objects a value, namely, $\forall a \in C \bigcup D, x \in U, f(x,a) \in V_a$. Select attribute $a \in C \cup D$ randomly, $| a_v |$ delegates the set formed by the objects with attribute *a* whose value is *v*. If $v(a) = \{ v_a^1, v_a^2, \cdots, v_a^k \}$, then set$\{ | a_{v_a}^{a_i} | : i = 1, 2, \cdots, k \}$ forms a granularity of the knowledge base space.

For instance, table 1 denotes a Knowledge base space of weather classification (Shi, 1998; Miao *et al.*, 2002), in which weather is sorted N or P according to condition attribute set { a_1(Outlook),a_2(Temperature),a_3(Humidity),a_4(Windy)}.

In table 1, the values of condition attribute *Outlook(a_1)*are *Sunny, Overcast* or *Rain*, their corresponding granule is $|a1_{Sunny}| = \{1,2,8,9,11\}$, $|a1_{Overcast}| = \{3,7,12,13\}$, $|a1_{Rain}| = \{4,5,6,10,14\}$. The values of condition attribute *Temperature(a_2)*are *Hot, Mild* or *Cool*, their corresponding granules are $|a2_{Hot}| = \{1,2,3,13\}$, $|a2_{Mild}| = \{4,8,10,11,12,14\}$, $|a2_{Cool}| = \{5,6,7,9\}$. The values of condition attribute *Humidity(a_3)*are *High, Normal* or *Low*, their corresponding granules are $|a3_{High}| = \{1,2,3,4,8,12,14\}$, $|a3_{Normal}| = \{5,6,7,9,10,11,13\}$, $|a3_{Low}| = \{\}$. The values of condition attribute *Windy(a_4)*are *True* or *False*, their corresponding granules are $|a4_{True}| = \{2,6,7,11,12,14\}$, $|a4_{False}| = \{1,3,4,5,8,9,10,13\}$. The values of decision-making attribute *Class(d)*are *N* or *P*, their corresponding granules are $|d_N| = \{1,2,6,8,14\}$, $|d_P| = \{3,4,5,7,9,10,11,12,13\}$. So, the granularities formed by condition attribute Outlook(a_1), Temperature(a_2),Humidity(a_3),Windy(a_4)and decision-making attribute Class(d)are:

$$G_{a_1} = \{\{1,2,8,9,11\},\{3,7,12,13\},\{4,5,6,10,14\}\}$$

$$G_{a_2} = \{\{1,2,3,13\},\{4,8,10,11,12,14\},\{5,6,7,9\}\}$$

$$G_{a_3} \{\{1,2,3,4,8,12,14\},\{5,6,7,9,10,11,13\}\}$$

$$G_{a_4} = \{\{1,3,4,5,8,9,10,13\},\{2,6,7,11,12,14\}\}$$

$$G_d = \{\{1,2,6,8,14\},\{3,4,5,7,9,10,11,12,13\}\}$$

Definition 3 Given a knowledge base space *(U, C, D, V, f)*, $X \subseteq C$ is a sub-set of condition attribute. $CON(X \to D)$ denotes the consistency of $X \to D$. Its definition is:

$$CON(X \to D) = |X \cup D| / |X|$$

Where, $|X| = |ND(X)|$ denotes the base of $ND(X) \subseteq U \times U$.

$CON(X \to D)$ is a kind of measurement for expressing or forecasting *D* by *X* .Generally, $CON(X \to D)$ is more big, *X* is more excellent in the forecasting attribute set of *D*.

For instance, in table 1, the combined granularities of each condition attribute $a_i (1 \le i \le 4)$ and decision attribute d are:

$G_{a_1 \cup d} = \{\{1,2,8\},\{9,11\},\{3,7,12,13\},\{4,5,10\},\{6,14\}\}$

$G_{a_2 \cup d} = \{\{1,2\},\{3,13\},\{4,10,11,12\},\{8,14\},\{5,7,9\},\{6\}\}$

$G_{a_3 \cup d} = \{\{1,2,8,14\},\{3,4,12\},\{5,7,9,10,11,13\},\{6\}\}$

$G_{a_4 \cup d} = \{\{1,8\},\{3,4,5,9,10,13\},\{2,6,14\},\{7,11,12\}\}$

And the consistency $CON(a_i \to d)(1 \le i \le 4)$ is:

$$CON(a_1 \to d) = |a_1 \cup d| / |a_1| = 0.6364$$

$$CON(a_2 \to d) = |a_2 \cup d| / |a_2| = 0.5588$$

$$CON(a_3 \to d) = |a_3 \cup d| / |a_3| = 0.6327$$

$$CON(a_4 \to d) = |a_4 \cup d| / |a_4| = 0.5800$$

3. Knowledge Innovation Algorithm Based on Granularity

3.1 Connotation of Knowledge Innovation

Knowledge is expressed by the following formula in the knowledge engineering field (Zhang *et al.*, 2005):

$K = F + R + C$

Where, *K* denotes knowledge, *F* denotes fact, *R* denotes rule, *C* denotes concept. Fact indicates the description for the state, attribute and feature of things. Rule indicates the causal relationship between premise and conclusions. Concept indicates the semantic description of facts and rules. Knowledge innovation indicates a new combination or a new transformation of knowledge factors. This is new knowledge production course using old knowledge fully. In other words, the emergence of new knowledge is not isolated, new knowledge is based on relevant existing knowledge.

3.2 Basic Idea of Algorithm

In knowledge base space *(U, C, D, V, f)*, if there is at least an object $x \in U$ and it has at least one attribute $a \in C \cup D$ with value $f(x,a)$ which is not in this knowledge base space, the knowledge base space can be called imperfect knowledge base space. Knowledge base space *(U, C, D, V, f)* may be abbreviated as *(U, C, D)*. Generally, knowledge base spaces are imperfect and perfect knowledge base space is only a special form of imperfect knowledge base space. Knowledge may be innovated and common knowledge base space can evolve.

For a given imperfect knowledge base space (*U , C, D)*, if a granularity *G* is consistent with the following condition: $\forall g \in G, |d(g)| = 1$. Namely, the values of decision-making attributes of every granule in granularity *G* are same. That means all granules in granularity *G* belong to a same class. So, after describing the granularity properly, the new knowledge rules can be gained according to the granularity. The algorithm uses consistency of attribute as a measure, constitute the partition granularity of knowledge base space ceaselessly, till the sort of every granules in the granularity is only one. This is a coarse-to-fine partition process for knowledge base space, and a tree structure is formed at last.

3.3 Steps and Working Flow of Algorithm

The working flow of knowledge innovation algorithm based on granularity can be described as follows.

Input- Imperfect knowledge base space (*U , C, D)*

Output-New knowledge which is not included in the old knowledge base space.

Step1: Calculate the consistency of each attribute $x_i \to D$ in knowledge base space (*U , C, D)*.

Step2: Choose the attribute with the biggest consistency value according to the result of Step1. Take this attribute as the

most important attribute a_m. If all attributes have a same consistency value, the coverage should be chosen as the most important attribute a_m.

Step3: Constitute the partition granularity of knowledge base space depend on the most important attribute a_m.

Step4: Judge if all the objects in a granule belong to a same class or not. For every granule, if all the objects belong to a same class, turn to Step7; otherwise turn to Step5.

Step5: Delete attribute a_m from attribute set.

Step6: Repeat the course of Step1~ Step4 till all the objects in a granule belong to a same class.

Step7: Describe every granule and output the new knowledge rules.

Step8: Stop.

The working flow can be described as Figure 1.

3.4 Time Complexity of Algorithm

In knowledge innovation algorithm based on granularity, the only one calculation object is attribute consistency. The time complexity is mainly decided by attributes combination in knowledge base space. In the worst case, the number of attributes in each calculation is $|C|, |C-1|, \ldots, 1$. So, the most total calculation number is:

$$|C| + (|C| - 1) + \cdots + 1 = |C|(|C| + 1)/2$$

If the influence of objects number is not considered, in the worst case, the time complexity of algorithm is $O(|C|^2)$.

4. Example

We take the imperfect knowledge base space of day weather classification shown in table 1 as an example, test the validity of above-mentioned knowledge innovation algorithm based on granularity.

The calculation results of example 2 tell us that consistency $CON(a_1 \to d)$ is the biggest one. So, attribute a_1 is selected to constitute the partition granularity of knowledge base space, as Figure 2 shows.

We can see from Figure 2 that there are two different conclusion attribute values in the first granule and the third granule. And there is only one conclusion attribute value in the second granule. So, it is only need to partition the first and the third granule again and not need to partition the second granule any more. After attribute a_1 is deleted, the objects in tow granules are shown in table 2 and table 3.

In the sub-space shown in table 2, each consistency is:

$$CON(a_2 \to d) = |a_2 \cup d| / |a_2| = 0.7778$$
$$CON(a_3 \to d) = |a_3 \cup d| / |a_3| = 1.0000$$
$$CON(a_4 \to d) = |a_4 \cup d| / |a_4| = 0.5385$$

As $CON(a_3 \to d) = 1.0000$ is the biggest one, so, attribute a_3 is selected to constitute the partition granularity of sub-space.

In the sub-space shown in table 3, each consistency is:

$$CON(a_2 \to d) = |a_2 \cup d| / |a_2| = 0.5385$$
$$CON(a_3 \to d) = |a_3 \cup d| / |a_3| = 0.5385$$
$$CON(a_4 \to d) = |a_4 \cup d| / |a_4| = 1.0000$$

As $CON(a_4 \to d) = 1.0000$ is the biggest one, so, attribute a_4 is selected to constitute the partition granularity of sub-space.

At last, a tree structure is formed, as Figure 3 shows.

In Figure 3, the leaf node set is the namely granularity with a knowledge space in which each granule only has one conclusion attribute value. This granularity is:

$$G = \{\{1,2,8\}, \{9,11\}, \{3,7,12,13\}, \{6,4\}, \{4,5,10\}\}$$

Then, describe all granules according to the path from base node to leaf nodes and the new knowledge rules can be gained. As table 4 shows.

The results in table 4 tell us that knowledge innovation is realized by the algorithm. The correctness and validity of algorithm is proved. There is only one computation namely the measure consistency of attribute in the algorithm, the time complexity is low. So, the feasibility of algorithm can be ensured.

5. Conclusion

In this paper, the concept of granularity was expounded and a knowledge innovation method based on granularity was proposed. The numerical calculation work of this algorithm is little. It is entirely feasible. Examples show that the algorithm can realize knowledge innovation successfully using knowledge of imperfect knowledge base. So, the algorithm is correct and valid. Future work is to study how to evaluate and test the existing knowledge and new knowledge generated by the algorithm. The excellent new knowledge will be used to update the original knowledge base, and consequently realize knowledge evolution.

Acknowledgments

Support from the National Natural Science Foundation of China under Grant No.60873035 and No.60743009 and the Natural Science Foundation of Shaanxi of China under Grant No.2006F43 is gratefully acknowledged.

References

Bernard, C., Tai, P.C., Harrison, R., & Pan, Y. (2006). *FIK Model: Novel Efficient Granular Computing Model for Protein Sequence Motifs and Structure Information Discovery*. Sixth IEEE Symposium on Bio-Informatics and Bio-Engineering, 27(2), 1-7.

Dun, Y., Zhang, X., Sun, H., &Zhao, L. (2006). *A rules mining method based on granularity*. Journal of Lanzhou University of Technology, 32(1),105-108.

Fogel, D. B. (1994). An introduction to simulated evolutionary optimization. *IEEE Trans. on Neural networks,* 5(1), 3-14

Fu, J., Zhou, C., &Jiang, G. (2005). *Speculation and Refution--Growth of Scientific Knowledge.* Shanghai: Shanghai Translation Publishing House.

Liu, C., Yang S., &Zhang Y. (2007). Knowledge evolution strategies. *Systems Engineering and Electroni,*. 29(6), 1017-1020.

Miao, D., &Fan, S. (2002). *The Calculation of Knowledge Granulation and Its Application*. Systems Engineering Theory and Practice, 1, 48-55.

PAWLAK, Z. (1982). *Rough sets* . Int J Comput Inform Sci, 11(5), 341-356.

PAWLAK, Z. (1991). *Rough sets—theoretical aspects of reasoning about data*. Boston: Kluwer Academic Publishers.

QUINLAN, J. R. (1987). *Simplifying decision trees* . International Journal of Man-Machine Studies, 27, 221-234.

Shi, Z. (1998). *Advanced Artificial Intelligence*. Beijing: Science Press.

Shu, W., &Zuo R. (2005). *Objective knowledge--Research on A Theory of Evolution*. Shanghai: Shanghai Translation Publishing House.

Yang, W., Ye, X., Tang, Y., & Wei, P. (2006). *Granularity of Knowledge Computed by Genetic Algorithms Based on Rough Sets Theory*. Journal of Nanchang Institute of Technology, 25(2), 97-121

Yao, J.T. (2008). *Granular Computing: A New Paradigm in Information Processing*. North American Simulation Technology Conference 2008 (NASTEC'08), 5-6.

Yao, Y.Y. (2007). *The art of granular computing*. Proceedings of the International Conference on Rough Sets and Emerging Intelligent Systems Paradigms, 101-112.

Yi, Z., &Ning, Z. (2008). *On Granular knowledge Structures*. Progress of Advanced Intelligence: Proceedings of 2008 International Conference on Advanced Intelligence, 28-33.

Zhang, F., &He, C. (2005). *The Theory and Path of Knowledge innovation*. Journal of Chinese Academy of Sciences Institute, 20(5), 389-394.

Zhong, Y. (2007). Knowledge concepts and measures in views of information – knowledge - intelligence ecosystem. *Journal of Frontiers of Computer Science and Technology*, 1(2), 129-137.

Zhu, Z. (2000). Research on the mechanism of knowledge evolution and knowledge innovation. *Journal Research and Development Management,* 12(6), 16-19.

Figures and tables

Table 1. Knowledge base space of weather classification

No	Condition attributes				Class
	a_1	a_2	a_3	a_4	(d)
1	Sunny	Hot	High	False	N
2	Sunny	Hot	High	True	N
3	Overcast	Hot	High	False	P
4	Rain	Mild	High	False	P
5	Rain	Cool	Normal	False	P
6	Rain	Cool	Normal	True	N
7	Overcast	Cool	Normal	True	P
8	Sunny	Mild	High	False	N
9	Sunny	Cool	Normal	False	P
10	Rain	Mild	Normal	False	P
11	Sunny	Mild	Normal	True	P
12	Overcast	Mild	High	True	P
13	Overcast	Hot	Normal	False	P
14	Rain	Mild	High	True	N

Table 2. Subspace(*value*(Outlook)=*Sunny*)

No	Condition attributes			Class
	a_2	a_3	a_4	(d)
1	Hot	High	False	N
2	Hot	High	False	N
8	Mild	High	False	N
9	Cool	Normal	False	P
11	Mild	Normal	True	P

Table 3. Subspace(*value*(Outlook)=*Rain*)

No	Condition attributes			Class
	a_2	a_3	a_4	(d)
4	Mild	High	False	P
5	Cool	Normal	False	P
6	Cool	Normal	True	N
10	Mild	Normal	False	P
14	Mild	High	True	N

Table 4. New knowledge rules

No	Condition attributes				Class
	a_1	a_2	a_3	a_4	(d)
1	Sunny	Random	High	Random	N
2	Sunny	Random	Normal	Random	P
3	Overcast	Random	Random	Random	P
4	Rain	Random	Random	True	N
5	Rain	Random	Random	False	P

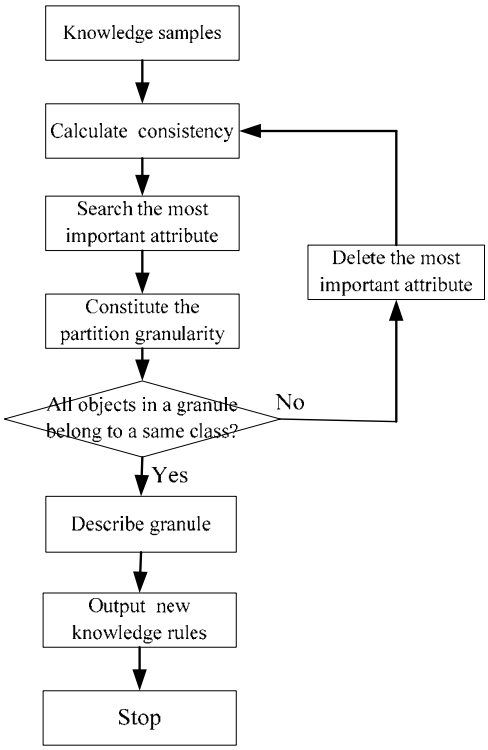


Figure 1. Working flow of Knowledge Innovation Algorithm Based on Granularity

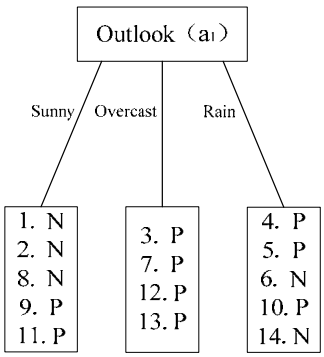


Figure 2. Granularity of knowledge space formed by a_1

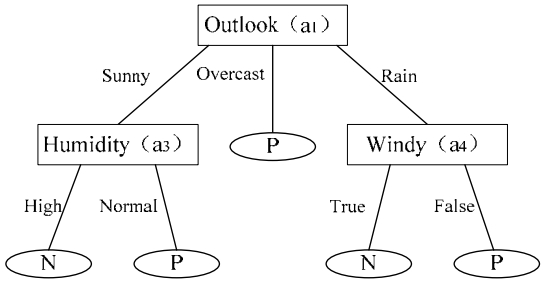


Figure 3. A tree structure gained by granularity calculation

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

E-Community System towards First Class Mentality Development: An Infrastructure Requirements Analysis

Rusli Abdullah
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
43400 UPM Serdang, Malaysia
Tel: 60-3-8946-6518 E-mail: rusli@fsktm.upm.edu.my

Musa Abu Hasan, Siti Zobidah Omar, Narimah Ismail & Jusang Bolong
Faculty of Modern Languages and Communications, Universiti Putra Malaysia
43400 UPM Serdang, Malaysia

Abstract

E-Community portal can be classifed as an extension of normal type of knowledge management system (KMS) development towards first class mentality. It servers varities of expects in term of capabilities and services especially for the benefits of community. Most of the community today are looking on this matter as a very important issue and try to search the best way to manage or organize this community system for sustain a high rate of continuous improvement. While e-community system (ECS) or portal is a system that related to the process of knowledge capture, re-use, searching and representation to the user in a variety of form. The role of system could be determined by looking on the issues on how knowledge can be applied at the right time in the faster ways that based on the simplest command or agent given to the system in order to get the relevant knowledge from the portal. Besides that, system also could be looked on how the best element of infrastructure requirement could be used for, in the benefits of users in order to stored and captured as well as presenting the knowledge portal. The paper presents the analysis of the ECS infrastructure requirement, and its system implementation in a community of practise (CoP) especially towards first class mentality development as well as discussing a variety issues that related to its involvement, so that it will help CoPs to increase their productivity and quality as well as to gain return on investment (ROI).

Keywords: E-Community, Community of Practise (CoP), Knowledge Management System, Information Technology

1. Introduction

The important of knowledge in a community is something that we have to think carefully. This is because of the successful of utilization of it may help community of practice (CoP) to become more powerful in term of services and products delivered to its communities mobilizing knowledge towards the first class mentality. Knowledge includes experience, values, insights, and contextual information in organization and helps them evaluate and incorporate new experiences and information. Knowledge originates and is applied by knowledge workers. People use their knowledge in making decisions. During the last several years, organizations realized they own a vast amount of knowledge and this knowledge needs to be managed. Davenport and Prusak (1998) defined knowledge as a fluid mixture of experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. They argue that knowledge originates and is applied in the minds of people who know.

In a CoP, it becomes embedded in documents and repositories, in processes, practices, and norms. There is a slightly different definition is given by Alavi and Leidner (1999). They see that knowledge as a justified personal belief that increases an individual's capacity to take action. They used Churchman's idea that 'knowledge resides in the user and not in the collection of information'. In their definition, action refers to physical skills and competencies, cognitive/intellectual activity or both (e.g., surgery involves both). Knowledge is an asset with four characteristics (McKinsey 1996); (1) *Extraordinary leverage*. It is not subject to diminishing returns. Fixed cost to create but not to manufacture or distribute. (2) *Fragmentation, Leakage*. Over time, knowledge assets become less valuable as they become more widely known. To be successful, knowledge must be refreshed to keep it as a source of competitive advantage. (3) *Uncertain Value*. Value is difficult to estimate and steady growth in knowledge may suddenly halt. (4) *Uncertain Value Sharing*. It can't predict in alliances that are will capture the lion's share. Knowledge management (KM) is a concept where it is could be used for creating knowledge repositories, improves knowledge access and

sharing as well as to communicate through collaboration, enhancing the knowledge environment and its management as an asset for an organization (Bertztiss, 2001; Satyadas *et.al.,* 2001) or in e-community system (ECS) or portal environment. Therefore, the paper presents the analysis of the ECS infrastructure requirement, and its system implementation in a CoP especially towards first class mentality development as well as discussing a variety issues that related to its involvement, so that it will help CoPs to increase their productivity and quality as well as to gain return on investment (ROI).

2. E-Community System and Its Functionality

E-Community system (ECS) is an extension of knowledge management system (KMS) that needs to be developed in the community. There are a lot of perspectives in describing the KMS. Among them is shown in Figure 1 below. The figure shows from the technical perspective has been proposed by Meso, & Smith (2000) that consists of three components that are *technology, function and knowledge* by itself. This ECS involved the processes for knowledge acquiring or collecting, organizing, disseminating or sharing knowledge among peoples in certain institution.

- *Data warehousing and data mining* – Data warehouses are centralized repositories of information. Data mining refers to specialized tools that allow the organization to convert increasingly complex sets of data into useful information.
- *Document management system* – A collection of tools that facilitate electronic document management, including storage, cataloguing, search, analysis and routing.
- *Push Technologies* – Delivering of appropriate information to individual based on specific criteria.
- *Collaboration* – expert modelling and decision making analysis that lead to more collaboration, information expertise and insight sharing among knowledge workers.
- *Visualization and navigation system*- Relationship between knowledge elements and holders of knowledge.

3. Infrastructure Requirement for E-Community System

In the process of developing e-community in the country, particularly towards the developing first class mentality where all the knowledge should be accessed very quickly and easily, and also based on the research worked that has been done, we found that there are some elements and issues could be considered especially in setting up the infrastructure requirement for the ECS. These elements of requirement could be highlighted in term of network, processors, software and database. The relationship between these elements can be shown as at Figure 2 below.

3.1 Network

The networking capabilities enable us to access and process information from unknown sources around the world combined with human interaction and collaboration on a scope and scale previously not achievable. Intranets and groupware are use for knowledge sharing, from best knowledge and expertise to the point of action, from those who have it, to those who need it. It provides a foundation for the sharing of information or 'explicit knowledge' as encoded in databases and 'tacit' knowledge, as partially transmitted in email and other conversations. According to Skryme (1999), and Seufert *et al.,* (1999), information and communication technology (ICT) enhance and support knowledge processes. In the evolution of the contribution of ICT in the workplace, the focus has change over time. In 1960's to the 70's the focus was more on automating the procedures. The 1980's, it was edge towards communications notably through electronic mail. There is an increase in efficiency in the office automation of standard procedures and the software development has shifted to support less structured activities those of professionals and managers, this is what termed as the cognitive computing focus, which has the most profound effect from the internet related activities.

The Internet is an information source that has brought information directly to the end-users, without involving an intermediary such as a librarian or information professional. An improved intelligent search engines, and intelligent software agents will roam the net and bring back relevant information to the desk-top (Kulakov *et. al.,* 2001: Wagner and Prasarnphanich, 2007). Internet has been almost entirely to do with its information role. But by the time knowledge becomes encoded in a database, it is 'explicit' knowledge. In many respects effective sharing of tacit knowledge needs the face-to-face socialization process as described by Nonaka and Takeuchi (1995). This is difficult over a distance although new technologies such as video conferencing are helping. The Internet here pays the role of enhancing remote access, through making the necessary connections and enhancing global communications.

Users congregate their areas of shared interests, and start electronic conversations. Electronic mail is the main way that this happens, but computer conferencing, such as with Lotus Notes (Same-Time), is popular in a corporate setting. However, knowledge sharing across organizational boundaries is increasingly required. It is ease of use that had made the use of Internet technology, such as browsers and search engines, of interest to organization wanting to share information. The advantages in a corporate setting of using Intranets (internal Internets) are similar to those that make use of the Internet attractive in external information access and communications. End-users are familiar with browser interfaces, information can be shared across different local area networks and computer platforms, and published

information is instantly available over the whole network. Furthermore this information need not just be HTML (the Web mark-up language) documents, but can be in any number of common formats, such as word processing for documents. Increasingly Intranets are also hosting transaction and database applications with the Web browser being the universal interface to different 'back-end' systems. Information and communications technologies enhance knowledge processes and support knowledge workers in several ways (Majchrzak *et. al*, 2002: Whelan, 2007) such as:

- Ready access to organised information.
- Better communications and interaction with fellow knowledge workers (either individually or in groups)
- Access to personal knowledge support tools (such as cognitive mapping tools)
- Use of specific point solutions (e.g. risk analysis in lending)
- Group decision support system that to facilitate decision making processes.

For the purpose of development and implementation, the ECS/KMS functionality could be as shown in Figure 3 below.

3.2 Processors

These processors are aim to support the distributed of knowledge workers or experts that stay at anywhere in the campus life. This supporting could be done using the connectivity between client, servers and other peripherals like routers, gateways and intelligent hubs. This processor also used for to make sure that application of KMS could be running on time or at different time in virtual environment.

3.3 Software

Software or sometimes called tools of e-community can be categorized into four groups: tools that support knowledge sharing examples are the groupware, intranets, and the internet; tools that support knowledge distribution for example the electronics calendars, desktop databases and desktop publishing; tools that support knowledge capture and codification example are the expert systems, neural networks and intelligent agents; and tools that support knowledge creation examples are the workstation, CAD and virtual realities. Ruggles (1997) suggested three categories of KM tools that represent primary knowledge activities of most organization:

- Knowledge Generation – the creation of new ideas, recognition of new patterns, the synthesis of separate disciplines and the development of new processes.
- Knowledge Codification – the auditing and categorization of knowledge
- Knowledge Transfer – the forwarding of knowledge between individuals, departments and organizations.

According to Laudon and Laudon (2002), there is also intelligence-based of KM Tools, which can be organized into multivariate content classification schema:

- MIND: Assimilation and Interpretation Tools – Includes Mapping, Mining, Summarization, Pattern Discovery, Decision support
- CONTENT: Network and Communications – Includes Conversing, Workflow, Information Sharing, Resource Sharing (intranet)
- MEDIA: Storage and Form – Includes text-bases, image-bases, multimedia

3.4 Database

This database is also called knowledge database of ECS/KMS is used for storing knowledge that contribute by the knowledge workers. This database is also a place where data or any kind of knowledge that could generate by the people or regenerate by itself in the system. The relationship between the acquiring knowledge and knowledge disseminating of for a k-base from the ECS as KMS system environment is shown in Figure 3.

4. Infrastructure Requirements and Related Issues of Ecs/Kms

a) **Technologies ICT Involvement –**

Another useful perspective on the role of ICT in KM is the role of technology infrastructure. Figure 4 shows the layers that build together in enterprise-wide and inter-enterprise collaborative infrastructures.

Layers of functionality:

Connections - the ability to connect anyone into the network at any time.

- Communications - establishing communications facilities such as electronic distribution lists, electronic meeting places (forums and discussion lists).
- Conversations - developing techniques and skills in conversing electronically; extracting meaning from ongoing threads of conversation. There is a role here for what is increasingly called a 'knowledge editor'.

- Collaboration - developing a tool case of collaborative tools, or knowledge collaboration 'archetypes'; supplementing these with moderators who nurture the development of new knowledge through a wide range of contributions

b) **Management and User Commitment**

The management and user commitment could be looked into their role and responsibilities. This is to ensure that they could be more supported and contributed for the knowledge wealth in term knowledge repository, and knowledge sharing.

c) **Easy of use or User friendly -** The objective of this issue is to make sure that the user will use the ECS not only for the first time but also following time. This element also at the some time they will be an agent to motive the others in order to use the ECS.

d) **System Stability & Usability -** This job basically is a part of webmaster responsibilities in order to make sure that system running at anytime and allowed people to do their work also at anywhere.

e) **Support & Training -** Computer support can make the input processes more effective. This includes selecting information and knowledge that is relevant. Text summarising for, example extracts the key parts from a document, so that the reader can gain most of the sense in a fraction of the original. Data mining will be extracting the new knowledge from existing data. It can find patterns that humans cannot, but considering many more dimensions and variables. In terms of the knowledge base, there is an increasing emphasis of adding some context to the information. This might be a fuller description of the application of the information, an indication of the quality of the source, and many other little human touches that are often not found in formal databases. In dissemination and use of knowledge, Intranets and groupware have most impacts on KM. Communication method such as video conferencing or face-to-face, allow experts and knowledge workers to communicate effectively and share 'tacit' knowledge without the need to travel.

f) **Security -** This is very importance component in ECS to ensure that people can use the knowledge in a trusty environment.

5. Discussion and Recommendation

In this paper, we are going to look at the extension of IT infrastructure contributes to the implementation of successful EC towards developing the first class mentality. As a practitioner of KM, one should do non-stop discovery on the latest issues of EC, the perspectives of the field, including other essential aspects such as technology, business, organization, and information technology. Our interest is on the role of IT as one of the core enabler towards the implementation of the KM in the community of practice (CoP). But still the role it plays is not really sufficient to direct to the sense-making enabling processes as well as to increase the organizations' collective intellect, although it is a necessity. We have seen the rapid growth of IT evolution since early 1990's until today. The development of IT has been spreading its wings when technology such as email, groupware, web technology emerged as powerful tools, which empower the IT environment (Beckett *et. al.,* 2000; Durcikova and Everard, 2002). We have also witnessed the convergence of IT to link knowledge workers and individuals by using these tools. There are ongoing debates about the role the IT can play for EC.

According to Borghoff and Pareschi (1998), IT is used pervasively in organizations, and thus qualifies as a natural medium for the flow of knowledge. Either we realized it or not, the most important key to make the organizations work is the knowledge. Thus, a comprehensive understanding of the tacit and explicit knowledge is required in order to understand the knowledge creation of EC at the individual level. As we have mentioned previously, Nonaka and Takeuchi (1995), have proposed four EC of knowledge interactions that build on the distinctions between tacit and explicit knowledge, which was described by Polanyi (1966). Tacit knowledge is that which is implied, is not actually documented; something an individual 'knows' from experience, other people, or from a combination of sources. Whereas, explicit knowledge is externally visible; it is documented tacit knowledge. The IT facilities provided in each interaction is summarized in Figure 5.

5.1 Tacit to tacit knowledge

The most basic element in tacit knowledge is sharing experiences. GroupWare is known for its capabilities to facilitate users with its application software, which helps to provide a 'place' where people can work together in groups or teams. This technology has a great impact, which is to some extent, it may support all four of the facets of interactions shown above. The 'place' here means the virtual space or a synthetic environment, within which participants of the interaction activities can share their experiences, documents, or even conduct meetings and discussions. Another richer form of shred experiences (such as online meetings) can be greatly supported by more sophisticated technology like video and text-based conferencing, as well as synchronous communications and chats. We can see that most of the organizations today have already applied these kinds of technologies and the numbers are keeping increasing day by day. And so with

the 'village wells' where the existence of this technology has embarked to a more virtual environment, where all conversations take place virtually, and even synchronously, both inside organizations and totally outside the walls of organizations. This leads to the existence of virtual community. But there are also some limitations of groupware for tacit knowledge. Based on the literatures, videoconferencing was almost as food as face-to-face meetings, whereas audio conferencing was less effective and text chat least so. This result might suggest us that video conferencing is the most suitable knowledge exchange medium in the socialization process so far.

5.2 Tacit to explicit knowledge

To convert tacit into explicit knowledge, it means documenting shared experiences into visible mediums such as white paper, reports and tapes. This conversion is called the externalisation process. According to Nonaka and Takeuchi (1995), the conversion of tacit to explicit knowledge involves forming a shared mental model, then articulating through dialog. This kind of interaction can be supported by technologies such as the collaboration systems, and GroupWare software. Online discussions are a method of interaction which allows a group of people to communicate to each other virtually, share experiences and knowledge and other group activities. Unlike typical team discussions, newsgroup is one form of technology, which is slightly different. It differs from the aspect of participants of the newsgroup, which they are typically strangers. But still, they are willing to share advices, knowledge, and assistance and so on. We should also consider some barriers experienced on Internet newsgroups such as faming, personal abuse, and irrelevant postings.

5.3 Explicit to explicit knowledge

According to Marwick (2001), there can be little doubt that the phase of knowledge transformation best supported by IT is combination, because it deals with explicit knowledge. Although the most common way to capture knowledge by far is to write a document, technology has made the use of other forms of media feasible. Technologies such as digital audio and video recordings are now easily made and available. Unlike the conventional methods of disseminating knowledge, the experts are no longer have to document or write the knowledge, they only have to speak to a camera or microphone instead. Furthermore, it is also now relatively easy to distribute audio and video over networks. Speech recognition is another technology used in the combination interaction type. Based on the journal written by the same author mentioned above, he has also highlighted several other technologies, which are important in dealing with explicit knowledge. They are search, portals, and meta-data, summarization and taxonomies and document classification approaches and techniques.

5.4 Explicit to tacit knowledge

From documentation type of knowledge to the understanding the knowledge, is really a challenge to the KM practitioners. A good KM should have the ability to facilitate the understanding of the information in the organizations. Practically speaking, this conversion type of knowledge is not as easy as it seen. This process requires a set of technology, which is called the visualization application. Methods include text-based category trees, exemplified by the current Yahoo! user interface. There are also several graphical visualizations for better understanding of the explicit knowledge. The best way of applying information technology to KM is probably a combination of two factors: on the one hand, the awareness of the limits of IT, and of the fact that any IT deployment will not achieve much, if it is not accompanied by a global cultural change toward knowledge values; on the other hand, the availability of information technologies that have been expressly designed with KM in view (Burghoff and Pareschi, 1998). Whatever it is, the application or the use of IT possesses quite a great impact on KM. It is assumed to be an essential enabler or key to the successful implementation of KM although some claimed it is just a part of it.

6. Conclusion

E-community is a good concept which can be used for creating knowledge repositories, improves knowledge access and sharing as well as to communicate through collaboration, enhancing the knowledge environment and managing knowledge as an asset. In term of the infrastructure requirement, the highest facilities that provided for ECS functionality is much more better. By using the capabilities of infrastructure such network, processor, and others, we could produce the best work and deliver it in any time and at the right place, while achieving the good quality and productivity as well as the return of investment (ROI) in an organization.

Of course, in the process of developing and implementing ECS towards the first class mentality in the community, there are a lot of issues that might be considered. Among the issues that could be considered are as follows:

- To determine the best practice for approaching and managing knowledge effectively including motivating members in community to share knowledge and access through the system.
- To determine the good metrics for evaluating efficient EC.
- To determine the best way to perform a knowledge audit.
- To determine how people create, communicate and use knowledge.

- To determine more inclusive, integrated ECS software packages.

Knowledge is a valuable asset to any individual and organizations. It is a key to enabling the existence of knowledge world. Its value should be treasured and shared by all levels. Trusting culture in knowledge environment should be built upon one purpose, which is to get to know how much value the knowledge possesses, is to know how much we trust the person whom we shared the knowledge with. It is also proved by researches that successful companies reward employees for seeking knowledge. Less successful communities by contrast is pushing knowledge to where it is needed. This kind of way is known as top –down approach. We have been to several seminars, which talked about the reasons why some communities tend to fail in implementing the knowledge environment. According to their surveys and studies, they identified that top-down approach is one of the causes to the failure. Goals, incentives and participations are identified as most reliable approaches to attract people seeking, creating and sharing knowledge. Each approach has a significant role in the aspect of ECS/KMS. They have great potential as essential enablers to implement successful ECS/KMS for CoPs to support knowledge sharing processes.

References

Alavi, M. and Leidner, D. (2001). "Knowledge Management Systems: Issues, Challenges, and Benefits" *Communication of AIS*, Vol. 1, Article 14

Beckett A. J., Wainwright, C. E. R., Bance, D. (2000). "Knowledge management: strategy or software?," *Management Decision,* vol. 38, no. 9, pp. 601-606.

Bertztiss A. T. (2001). "Dimensions of the Knowledge Management Process," *IEEE*, pp. 437-441.

Borghoff, U., & Parechi, R. (1998). *Information technology for knowledge management*. Berlin, Germany: Springer Verlag.

Davenport, T.H. and L. Prusak. (1998). *Working Knowledge: How Organizations Manage What They Know* Boston, MA: Harvard Business School Press

Durcikova, A. and Everard, A. (2002). "An Employee Typology: A Knowledge Management Perspective," in *Proceedings of Proceedings of the Eighth Americas Conference on Information Systems*, pp. 2042-2048.

Kulakov A., Trajkovik, V., Davcev, D. (2001). "An agent based model as a marketing consultant of companies," *IEEE*, pp. 3113-3117.

Laudon, J. P.; Laudon, K. C. (2002). *Management Information Systems. Managing the Digital Firm (7th edition).* Prentice Hall

Lotus Company (www.lotus.com). Accessed on December 2007.

Majchrzak A., Malhotra, A., & Chai, L. (2002). "Different Strokes for Different Folks: Managing and Enabling Virtual Teams," in *Proceedings of Proceedings of the Eighth Americas Conference on Information Systems*, pp. 2091-2096.

Marwick A. D. (2001). "Knowledge management technology," *IBM Systems Journal*, vol. 40, no. 4, pp. 814-831.

McCampbell A. S., Clare, L. M., & Gitters, S. H. (1999). "Knowledge management: the new challenge for the 21st century," *Journal of Knowledge Management,* vol. 3, no. 3, pp. 172-179.

McKinsey, (1998). "Best Practice and Beyond: Knowledge Strategies" *The McKinsey Quarterly* Number 1, pp. 19-25

Meso, P. and Smith, R. (2000). "A Resources-based View of organizational knowledge Management Systems", *Journal of Knowledge Management*, Vol. 4 No. 3, pp. 224-234.

Nonaka, I and Takeuchi, H. (1995). "*The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*". New York: Oxford University Press.

Numamaker, Jr., Romano Jr. & Owen R. (2001). "A Framework for Collaboration and Knowledge Management", *Proceeding of 34th Hawaii International Conference on System Sciences, IEEE.*

Polanyi, M. (1966). *The Tacit Dimension*, Doubleday, New York, NY.

Ruggles, R. L. (1997). *Knowledge Management Tools.* Butterworth Heinemann,

Satyadas A., Harigopal, U., Cassaigne, N. P. (2001). "Knowledge Management Tutorial: An Editorial Overview," *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications And Review*, vol. 31, no. 4, pp. 429-437.

Seufert A., von Krogh, G., Bach, A. (1999). "Towards knowledge networking," *Journal of Knowledge Management,* vol. 3, no. 3, pp. 180-190.

Skyrme, D.J. *(*1999), Knowledge Networking: Creating the Collaborative Enterprise, , Butterworth-Heinemann.

Wagner, C. and P. Prasarnphanich, P. (2007). "Innovating Collaborative Content Creation: The Role of Altruism and Wiki Technology", *Proceedings of the 40th Hawaii International Conference on System Sciences*.

Whelan, E. (2007). "Exploring Knowledge Exchange in Electronic Networks of Practice", *Journal of Information Technology*, 22(1), pp. 5-12.

List of Figures

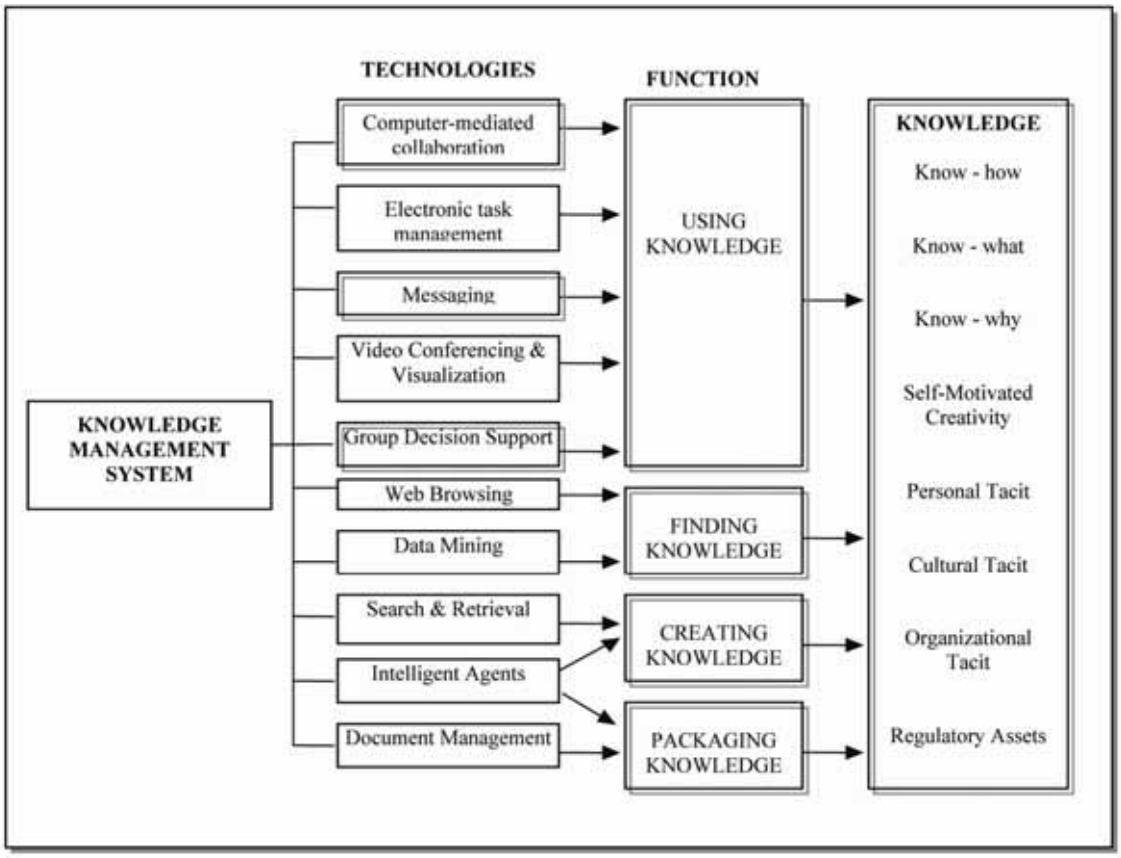


Figure 1. The technical perspective of KMS for E-community system

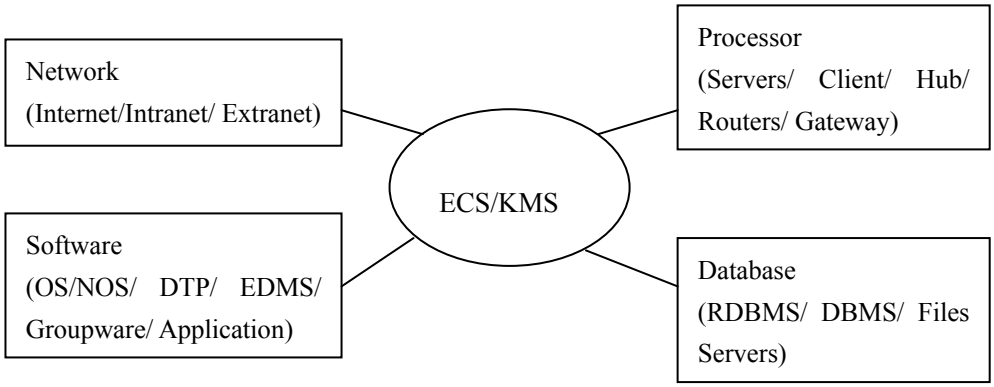


Figure 2. The infrastructure requirement of ECS

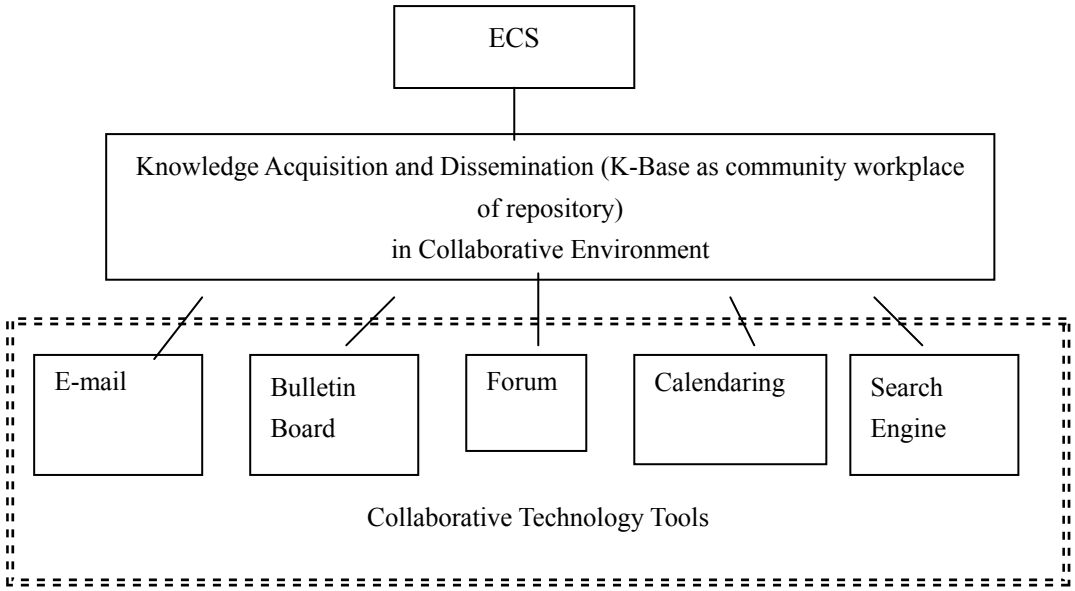


Figure 3. The EC System Functionality

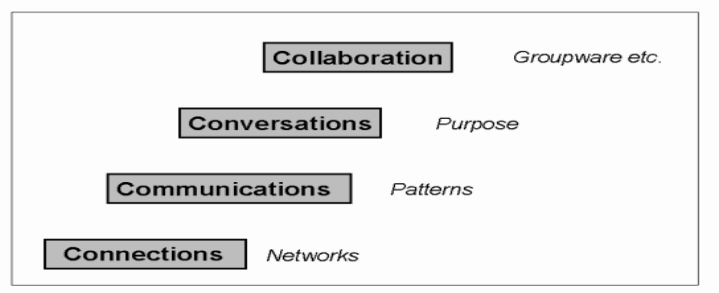


Figure 4. Layered elements of a collaborative technological infrastructure

Tacit to tacit knowledge	**Tacit to explicit knowledge**
Knowledge exchange: one-to-one, one-to-many, many-to-many Traditional knowledge exch. Medium: same place/same time, face to face meetings Technologies: teleconferencing, desktop video conferencing tools, E-meetings, chatting, synchronous collaboration	Knowledge exchange: one-to many Traditional knowledge exch. medium: created periodic reports, white papers Technologies: Electronic mail (E-mail), broadcasting information via distribution lists, answering questions, annotation
Explicit to tacit knowledge	**Explicit to explicit knowledge**
This form of knowledge creation depends on an individual's ability to make sense out of explicit information Today technologies: visualization	Today technologies: E-mail, GroupWare, Homepages

Figure 5. Summarization of the EC Infrastructure for knowledge Interactions

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Automatic Recognition of Focus and Interrogative Word in Chinese Question for Classification

Zhichang Zhang
School of Computer Science and Technology, Harbin Institute of Technology
Harbin 150001, China
&
School of Mathematics and Information Science, Northwest Normal University
Lanzhou 730070, China
E-mail: pangzhang@gmail.com
Yu Zhang, Ting Liu & Sheng Li
School of Computer Science & Technology, Harbin Institute of Technology
Harbin 150001, China

Abstract

Question classification is one of the most important components in a question answering (QA) system. When there are fewer features in a question can be used for classification, the interrogative word and focus in question are critical features. Most previous studies in question classification used heuristic rules to identify the focus and interrogative word in question. In this paper, a statistical method is explored to automatically label them for Chinese question using condition random fields (CRFs) model. The features for CRFs model are extracted from word segmentation, part-of-speech (POS) tagging, named entity recognition, and dependency parsing results. A knowledge base *HowNet* is also used. The experimental results show that the precision for interrogative word recognition is 98.97% and 90.85% of focus can be correctly recognized in a free available Chinese question data set.

Keywords: Question answering, Question classification, Interrogative word, Focus, Condition random fields

1. Introduction

Question Answering, as one of the important directions in information retrieval (IR) and natural language processing (NLP) research, is the task of locating the answer to a natural language question in large collection of documents responding. A typical question answering system consists of four central components including question analysis, document retrieval, passage retrieval, and answer extraction, where question analysis is to attain the expected answer type of a question. For example, "*What is the population of China?*" expects a number as answer, and "*Which country has the largest population?* " expects a country name. Thus deciding the expected answer type of a question can be seen as classification problem. The goal to classify the expected answer type is to provide the constraint condition for answer extraction. Results of the error analysis of an open domain QA system showed that 36.4% of the errors were generated by the question analysis module (Moldovan et al., 2003).

Question classification is a special kind of text classification. Compared with text documents, questions are generally short in content and there are fewer features available in them than in text for classification. Thus selecting important features can take significant effects on classification performance. Among these features, interrogative word and focus are critical, and in many cases the question can be correctly classified just using these two features.

Interrogative word and focus in Chinese question are more flexible in expression form and location compared with English question. Interrogative words are not stable in Chinese and their location can be at the start, end, or middle of a Chinese question. While many previous studies used heuristic methods to recognize interrogative word and focus in question, we use Condition Random Fields model to label them employing the dependency relations and other syntactic information in question.

The rest of this paper is organized as follows. Section 2 introduces related work about focus and interrogative word identification in question classification. Section 3 describes the method to label the interrogative word and focus in Chinese question automatically. Section 4 details the experimental results and analysis. Section 5 concludes the paper and provides some future directions.

2. Related Work

In an earlier work (Li and Roth, 2002 and 2004) about English question classification, the focus and interrogative word in question are not explicitly extracted as features and all words in question are not distinguished. The key features would be recognized by automatic learning process, so the classifier used many types of features to classify questions.

(Donald Metzler and W. Bruce Croft, 2004) viewed the phrase containing focus in question as the main noun phrase and applied simple heuristics based on POS tags to it, then extract the headword from this phrase as important feature for question classification. This headword is really identical as focus used in this paper.

(Lu and Zhang, 2004) studied the problems of Chinese question understanding in question answering system. They also used rules to recognize the focus in Chinese question although the definition of question focus they gave is not completely same as ours in this paper.

(Sun et at, 2007) used HowNet to classify Chinese question. They named the focus in question as question intent word and extract it from noun words surrounding interrogative word. In this paper, question focus might be not only noun word, but also adjective.

3. Identification of Focus and Interrogative Word

3.1 *Interrogative Word in Chinese Question*

Interrogative words in English are *what*, *when*, *when*, *why*, and *who*, *which* and *how*. However, there are more interrogative words in Chinese than in English. Table 1 lists some interrogative words in Chinese. All these words contain a special character which can be used alone as an interrogative word, such as "几", "多", etc

百分之几, 第几, 多, 多长, 多久, 多少, 多重, 干什么, 何, 何处, 何地, 何年, 何时 何谓 何在 几 几分之几 几十 哪 哪儿
哪个 哪家 哪里 哪些 哪类 如何 啥 啥样 什么 什么样 谁 为何 为什么 怎么 怎么办 怎么样 怎样

Figure 1. The Chinese interrogative words

Interrogative word in Chinese is very flexible. The number of interrogative words in Chinese is not stable and it is difficult to list all of Chinese interrogative words. While an interrogative word in Table 1 is embedded as a substring in a more long word, this long word itself can be used as a new interrogative word. For example, the word "多少度" embeds the usual interrogative word "多少", and the word "多少度" might be regarded as a single word by Chinese segmentation tools, then this word can also be an interrogative word.

When one of these words exists in a question, it might be just a modifier as adverb or number instead of interrogative word, and sometimes it might be in a named entity. Furthermore, there might be multiple interrogative words in a Chinese question.

Below are some examples to show these different situations.

- 江主席与克林顿的几次会谈分别在哪年进行的？
- "谁是最可爱的人"是哪个作家写的？
- 诸葛亮在哪几年出兵讨伐曹魏？
- 朱镕基从哪年到哪年在清华大学学习？

In the first example, the word "几" is not real interrogative word even if it can be used as an interrogative word. In the second example, the first interrogative word "谁" is in a named entity and not the interrogative word of the whole question. In the third and fourth example, there are two interrogative words in a single question. While the two interrogative words "哪" and "几" are different in the third example, the interrogative word "哪年" occurs two times in the fourth example.

In addition, unlike English interrogative words which generally occur in the start or end of a question clause, Chinese interrogative words can occur in the middle of a question clause besides the start and end positions. The four examples above show this case.

Thus there should be some ambiguities as described above to detect and remove when automatically recognizing the interrogative words in a Chinese question. While there might be multiple interrogative words in a Chinese question, there is priority level difference among them to become a real interrogative word.

3.2 *Focus in Chinese Question*

Similar to interrogative word, the focus in question is also a kind of critical feature for question classification, no matter

the classifier is based on statistical methods or rules. The focus in a Chinese question is generally a noun, quantity, adjective or their pair which often expresses the expected answer type of the question. For example, the question "2002年诺贝尔奖的货币价值是多少？(*What was the monetary value of the Nobel Peace Prize in 2002?*)" expects a monetary number answer which is expressed by focus "货币|价值 (*monetary | value*)". For the question "世界上最高的山是什么山？(*What is the highest mountain in the world?*)", the focus is "山(*mountain*)" and expresses that the answer of this question should be a mountain name. While the focus in the second example is a single noun word, the focus in the first question is a noun pair "*monetary | value*" for the reason that the word "价值 (*value*)" is an abstract concept and can be an attribute of any entity, then it cannot convey the concise expected answer type alone and a modifier word should be given to narrow and make the focus semantic concise. For the question "北京和天津之间有多远 (*How far is between Beijing and Tianjing*) ?", the focus is an adjective word "远 (far)" and expresses a distance concept.

For some complex questions, their answers are often long descriptions which can not be constrained on concise named entity or phrase types. When the interrogative words in questions will give clearly the expected answer type, there not exist appropriate focuses corresponding to the answer type.

Question focus has usually tight syntactic relations with interrogative word. The interrogative word in Chinese question can be expressed in very flexible syntactic format. Thus when the function of focus in a Chinese question is similar to that in English, the expression of them is more flexible than in English.

The syntactic role of focus in Chinese question can be: 1) the head of interrogative word modifier; 2) the subject of the question while the object is the phrase containing the interrogative word; 3) the object of the question while the subject is the phrase containing the interrogative word; 4) For example, in the question "五个联合国常任理事国中面积最小的是哪个？", the focus "理事国" . These conditions can be utilized as features to recognize the question focus.

3.3 *Using CRFs model to recognize interrogative word and focus*

We view recognition of interrogative word and focus as a sequence labeling task in an ordered question words given part of speech (POS) of these words and dependency relations between them. For a label set L={question_word, focus_word，other}, the task is to label the class of every word in a question based on the features of the word, i.e., given the observation sequence X, to find the output random variables which can lead to that the random probabilities in the formulation (1) have the maximum value, these random variables will be the labeling results.

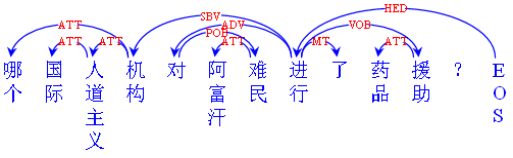
$$Y = \arg\max_{Y} P(Y \mid X), \tag{1}$$

where

$$X = \{x_1,\ldots,x_n \mid x_i \in D, i = 1,\ldots,n\}$$
$$Y = \{y_1,\ldots,y_n \mid y_i \in L, i = 1,\ldots,n\} \tag{2}$$

Because of excellent performance of CRFs model reported in many research works in sequence labeling task, we select it to recognize the interrogative word and focus in a Chinese question.

Table 1. Examples of lexical and syntactical analysis results for a Chinese question

question	哪个国际人道主义机构对阿富汗难民进行了药品援助？
Segmentation and POS tagging	哪个/r 国际/n 人道主义/n 机构/n 对/p 阿富汗/ns 难民/n 进行/v 了/u 药品/n 援助/n ?/wp
Dependency parsing	
Interrogative word/focus	哪个/机构 (which / organization)

A very important factor for CRFs is to select apt feature set according to the specific labeling task. For our identification task of focus and interrogative word, we first segment Chinese question into words and tag their POS, then syntactically parse the total question before extracting feature set. For example, the question "哪个国际人道主义机构对阿富汗难民进行了药品援助？(*Which international humanistic organization aided drugs to Afghan refugees ?*) ", its lexical and syntactical analysis results are shown in table 1.

CRF model can utilize overlapped features among words in a word window. We set the sliding window length as 5 and design features in the following types according to current word.

1) Word N-*grams*, POS N-*grams* (including *unigram* and *bigram*). They are used to get the context word and POS information of the current word.

2) *Unigram* of dependency modifier word, N-*gram* of dependency modifier POS, N-*gram* of dependency relation between head and modifier. They are used to attain the dependency structure information of the current word.

3) Combination of 1) and 2). CRFs model can use sufficient overlapped features to enrich the description ability of context. Hence we are to combine the types described above as new features.

4) Other conditions about current word and question.

All features and their expression patterns used by CRF model are listed in table 2.

Table 2. Features used by CRF model

Feature class	**pattern(** k = -2, -1, 0, 1, 2 **)**
Word N-gram	W_k W_k/W_{k+1}
POS N-gram	P_k P_k/P_{k+1} $P_k/P_{k+1}/P_{k+1}$
Modifier word unigram	Dep_k
Modifier POS N-gram	DP_k DP_k/DP_{k+1} $DP_k/DP_{k+1}/DP_{k+2}$
Dependency relations N-gram	Rel_k Rel_k/Rel_{k+1} $Rel_k/Rel_{k+1}/Rel_{k+2}$
Combination of word and POS	W_k/P_k
Combination of POS, POS of modifier, and dependency relation	$P_k/DP_k/Rel_k$
Combination of POS of modifier, dependency relation	DP_k/Rel_k
Whether the word is a part of a named entity	NE_k
The hypernym of word in HowNet	H_k

Table 3. The example of features

Pattern (k = -2, -1, 0, 1, 2)	Example （current word: 机构）
W_k W_k/W_{k+1}	[国际 人道主义 机构 对 阿富汗] [国际/人道主义 人道主义/机构 机构/对 对/阿富汗]
P_k P_k/P_{k+1} $P_k/P_{k+1}/P_{k+1}$	[n n n p ns] [n/n n/n n/p p/ns] [n/n/n n/n/p n/p/ns]
Dep_k	[人道主义 机构 进行 进行 难民]
DP_k DP_k/DP_{k+1} $DP_k/DP_{k+1}/DP_{k+2}$	[n n v v n] [n/n n/v v/v v/n]
Rel_k Rel_k/Rel_{k+1} $Rel_k/Rel_{k+1}/Rel_{k+2}$	[ATT ATT SBV ADV ATT] [ATT/ATT ATT/SBV SBV/ADT ADT/ATT]
W_k/P_k	[国际/n 人道主义/n 机构/n 对/p 阿富汗/ns]
$P_k/DP_k/Rel_k$	[n/n/ATT n/n/ATT n/v/SBV p/v/ADV ns/ v/ADV]
DP_k/Rel_k	[n/ATT n/ATT v/SBV v/ADV v/ADV]
NE_k	[0 0 0 0 1]
H_k	[属性值 精神 群体 - 物质]

To decide the hypernym of word, we use the knowdelge base *HowNet* (Dong and Dong, 1999) as the taxnomomy tree. *HowNet* divides all concepts into the following classes: 'physical|物质', 'mental|精神', 'fact|事情', 'group|群体', 'time|时间', 'space|空间', 'componet|部分', 'Appearance|外观', 'Measurement|量度', 'Property|特性', 'Relationship|关系', 'Situation|状况', 'QuantityProperty|数量特性', 'Quantity|数量', 'AppearanceValue|外观值', 'MeasurementValue|量度值', 'PropertyValue|特性值', 'RelationshipValue|关系值', 'SituationValue|状况值', 'QuantityPropertyValue|数量特性值', 'QuantityValue|数量值'. The hypernym of a word will be one of these classes or a single-bar "-" if the word is not in the dictionary of *HowNet*.

With the question example ("哪个国际人道主义机构对阿富汗难民进行了药品援助？") in table 1, we explain the features in detail in table 3.

In these features, some may have negative effect to overall performance of system. We should find and eliminate those features with experiments.

4. Experimental Evaluation

4.1 *Experiment setup*

IR Lab of Harbin Institute of Technology(Note 1) provides an open available Chinese question data set for Chinese question classification research, which consists of the training and test set. We use the training question data set (4981 questions) as our total experimental data set of recognition of interrogative word and focus, while the 70% (3528 questions) of it is used as training set and the 30% (1453) is used for test. The interrogative words and focuses in questions of all training and test set were labeled manually.

The free available CRF++(Note 2) tool is used to label the interrogative word and focus in Chinese question words. Before training and test, words segmentation, POS tagging, and dependency parsing for all questions will be performed with the open and free available IR LTP tool. We use three metrics to evaluate the achieved performance, including *QP*, *FP*, and *F_Score*, they are defined as follows.

$$QP = \frac{\#\ of\ tagged\ correct\ interrogative\ words}{\#\ of\ total\ interrogative\ words} \tag{3}$$

$$FP = \frac{\#\ of\ tagged\ correct\ focuses}{\#\ of\ total\ focuses} \tag{4}$$

$$F\_score = \frac{2 * QP * FP}{QP + FP} \tag{5}$$

QP is used to evaluate the precision of interrogative word labeling and *FP* is for evaluation of focus labeling. These two precision are then combined using F measure with equal weight given to them.

4.2 *Experimental results and analysis*

In ten features listed in Table 2, some might have negative effect for overall performance of CRFs model. We check and evaluate every feature with same training and test data.

Table 3 list the interrogative word labeling precision *QP*, focus labeling precision FP and F Score while all ten features are used or one of them is eliminated from the feature space. We incorporated word hypernym feature in the model using *HowNet* knowledge base. Unfortunately, this fails to yield improved precision.

From the results we can see that eliminating the feature "*Word hypernym*", "*Whether the word is a part of a named entity*", or "*Combination of POS of modifier, dependency relation*" in the model will lead to the increase of F score, thus these three features have negative effect for labeling performance. When the feature "*Word hypernym*" and "*Combination of POS of modifier, dependency relation*" have negative effect on interrogative word labeling and focus labeling as well, the feature "*Whether the word is a part of a named entity*" only declines recognition performance of focus in question.

Thus we discarded two features "*Word hypernym*" and "*Combination of POS of modifier, dependency relation*" and selected remained features for final training and test. The recognition performance and the contribution to overall performance of every feature are shown in Table 4.

Table 3. The performance effect of single feature before feature selection (%)

Features	QP	FP	F Score
All	98.76	90.78	94.60
All – *Word hypernym*	**98.97**	**90.85**	**94.73**
All - Whether the word is a part of a named entity	98.69	**90.92**	**94.64**
All - *Combination of POS of modifier, dependency relation*	**98.76**	**90.85**	**94.64**
All - Combination of POS, POS of modifier, and dependency relation	98.69	90.50	94.42
All - Combination of word and POS	98.15	90.50	94.17
All - N-gram of dependency relation	98.69	90.64	94.49
All - N-gram of dependency modifier POS	98.76	90.43	94.41
All - Unigram of dependency modifier word	98.69	90.57	94.46
All - POS N-gram	98.42	90.43	94.26
All -Word N-gram	98.34	89.88	93.93

Table 4. The performance effect of single feature after feature selection (%)

Features	QP	FP	F Score
All	98.97	90.85	94.73
All - Whether the word is a part of a named entity	98.90 (-0.07)	90.71 (-0.14)	94.63 (-0.10)
All - Combination of POS, POS of modifier, and dependency relation	98.90 (-0.07)	90.50 (-0.35)	94.51 (-0.22)
All - Combination of word and POS	**98.42 (-0.55)**	**90.36 (-0.45)**	**94.22 (-0.51)**
All - N-gram of dependency relation	**98.90 (-0.07)**	**90.09 (-0.76)**	**94.29 (-0.44)**
All - N-gram of dependency modifier POS	98.97 (-0.00)	90.23 (-0.62)	94.40 (-0.33)
All - Unigram of dependency modifier word	98.97 (-0.00)	90.43 (-0.42)	94.51 (-0.22)
All - POS N-gram	**98.49 (-0.48)**	**90.43 (-0.42)**	**94.29 (-0.44)**
All - Word N-gram	**98.35 (-0.62)**	**89.88 (-0.97)**	**93.93 (-0.80)**

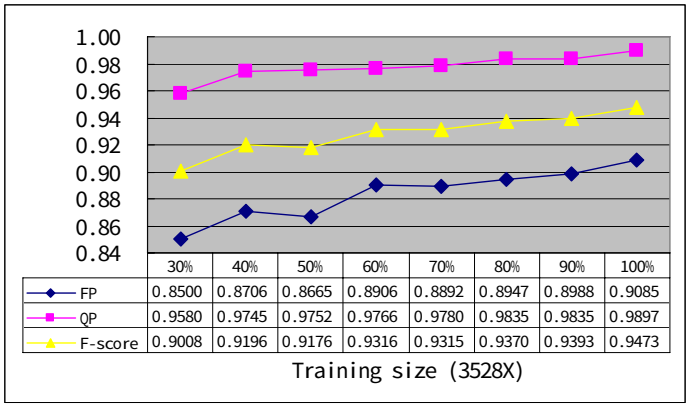


Figure 2. The impact of training corpus size on recognition performance

In eight features, four of them (including "*Word N-gram*", "*Combination of word and POS*", "*N-gram of dependency relation*", "*POS N-gram*") have the largest contributions to overall system performance.

Furthermore, we tested the relation between accuracy of the CRFs model and training corpus size. We extract different proportions of training questions randomly on every question class from the original total training questions to form different scale of new training data sets representing 30% (1022 questions), 40% (1378 questions), 50% (1743 questions), 60% (2082 questions), 70% (2430 questions), 80% (2788 questions), 90% (3135 questions) of the original training set. When labeling model is trained on these training data, the results are judged on the same test data for comparison. The impact of corpus size on recognition performance of focus and interrogative word is shown in Figure 2.

The results are presented in Figure 2. The figure indicates that system performance for recognition of focus and interrogative is directly related to training corpus size. While performance of focus recognition does improve with corpus size obviously, the corpus size has very slight impact on recognition performance of interrogative word. This can be explained that interrogative words in Chinese are relatively stable and have less variation in format and expression compared to focus word.

Those questions their focus or interrogative words are wrongly labeled should be analyzed. Most errors of interrogative word labeling are caused by wrong lexical or syntactic analysis. For example, in the question "究竟 该 购买 多 高 频率 的 CPU 呢", the POS of word "多" is falsely tagged as "a" (adjective) while its right POS is "d" (adverb). Some are caused by data sparseness of training data. For example, in the question "笔记本 电脑 重 多少克", the real interrogative word is "多少克", but it does not occur in the training data as an interrogative word, then the model cannot correctly tag it as interrogative word. Some errors are produced by learning process itself. For focus recognition, most errors are caused by feature selection method.

5. Conclusion

Concerning the automatic recognition of interrogative word and focus recognition in Chinese question for classification, this paper reports the experimental results given by sequence labeling model CRFs, which is trained using lexical and syntactic analysis results as features. The performance effects of selected features are tested and the impact of training

corpus size on recognition performance is also evaluated. It shows that the features "*Word N-gram*", "*Combination of word and POS*", "*N-gram of dependency relation*", and "*POS N-gram*" have the largest contributions to overall system performance of recognition. And the performance of focus recognition can be affected by corpus size observably while the corpus size has very slight impact on recognition performance of interrogative word.

References

Diekema, A. Liu, X., Chen, J., Wang, H., McCracken, N., Yilmazel, O., and Liddy, E. D. (2000). Question Answering: CNLP at the TREC-9 Question Answering Track. In: *Proceedings of the 9th Text REtrieval Conference (TREC-9)*, National Institute of Standards and Technology, Gaithersburg, MD.

Dong, Z. D. and Dong Q. (1999) Introduction to HowNet - Chinese Message Structure Base. [Online] Available: http://www.keenage.com/html/e_index.html (October 29, 2009).

Li X. and Roth D. Learning Question Classifiers. (2002). In: *Proceedings of the 19th International Conference on Computational Linguistics (COLING'02)*. Taipei, Taiwan: Association for Computational Linguistics, 1-7.

Li X. and Roth D. (2004). Learning question classifiers: The role of semantic information. *Natural Language Engineering*, 1(1).

Lu Z. J. and Zhang D. M. (2004). Question Interpretation for Chinese Question Answering. *Computer Engineering*, 30(18): 64-65.

Metzler, D. and Croft W. B. (2005). Analysis of Statistical Question Classification for Fact-based Questions. *Information Retrieval*, vol. 8, 481-504.

Moldovan D., Pasca M., Harabagiu S., and Surdeanu M. (2003). Performance issues and error analysis in an open-domain question answering system. *ACM Transactions on Information Systems*, 21(2): 133-154.

Sun J. G., Cai D. F., Lv D. X. and Dong Y. J. (2007). HowNet Based Chinese Question Automatic Classification. *Journal of Chinese Information Processing*, 21(1): 90 - 95.

Notes

Note 1: http://ir.hit.edu.cn/demo/ltp/Sharing_Plan.htm
Note 2: http://crfpp.sourceforge.net/

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Simulation of Radiation Characteristics for Cylindrical Conformal Phased Array

Rongcang Han
School of Informatics, Linyi Normal University, Linyi 276005, China
E-mail: hanrongcang@lytu.edu.cn

Abstract

FEKO is a 3D simulation tool for the analysis of electromagnetic fields. FEKO can solve the problem of any structural type with a very comprehensive MoM code or other tailored code. In this paper, a radiation pattern for conformal phased array antenna on the surface of a PEC finite cylinder is calculated using FEKO 4.0. The simulation results accord with the true characteristics of conformal array antenna on a finite length cylinder. The results also can meet with which in related reference. So, the simulation can be proved to be correct. Furthermore, the patterns of different scanning angle for this conformal phased array are simulated and some important parameters of the array patterns are given.

Keywords: Conformal Phased Array, Radiation Characteristics, FEKO 4.0

1. Introduction

A conformal antenna is an antenna that conforms to something; in our case, it conforms to a prescribed shape. The shape can be some part of an airplane, high-speed train, or other vehicle (J.A. Ferreira and F. Ares, 1997, 1188). The purpose is to build the antenna so that it becomes integrated with the structure and does not cause extra drag. The purpose can also be that the antenna integration makes the antenna less disturbing, less visible to the human eye; for instance, in an urban environment. A typical additional requirement in modern defense systems is that the antenna not backscatter microwave radiation when illuminated by, for example, an enemy radar transmitter (Lars Josefsson, Patrik Persson, 2006, chapter1).

Conformal antennas are of great interest due to their ability to blend into curved surfaces. Antennas of cylindrical geometry can be analyzed in a number of ways, among which are the finite element method, the auxiliary sources method, the method of moments (MOM), etc. The method of moments can be employed in both spectral and spatial domains. Both solutions work well until the cylinder is electrically small (Alexander Svezhentsev, 2004, 629). Here, we use MOM to simulate the conformal phased array by FEKO 4.0.

2. Radiation Pattern of A Conformal Element Antenna

The specific coordinate and model of a conformal antenna are shown in Figure 1, which is a microstrip antenna described as an element of a conformal phased array. The center of rectangular microstrip antenna on the cylinder at the point (according to cylindrical coordinate system) $(a, 0°, 0)$, of which the radius of cylinder is a=0.5λ, the length of cylinder is 4λ. The length of rectangular patch is $2L$ and width is $2\varphi_o a$. The relative dielectric constant for dielectric is ε_r, $\varepsilon_r = 2.33$, the thickness of the substrate is h, $h = 0.05\lambda_g$, where λ_g is the medium wavelength, $\lambda_g = \lambda / \sqrt{\varepsilon_r}$. The microstrip antenna is probe feed; the probe is not in the patch center, but at the point which is $0.125\lambda_g$ away from the center along the direction φ cylindrical coordinate system.

The simulation of the array element is carried out in FEKO 4.0. The simulation results of element patterns are shown in Figure 2 when the cylinder with a radius of 0.5λ. There two radiation patterns at plane E and plane H for conformal element. The behavior of the two patterns shows a good coherence with theoretical results and engineering experiences.

3. Radiation Pattern of A Cylindrical Conformal Phased Array

In order to study the radiation patterns of a conformal phased array antenna, an example in Figure 3 is carried out. The locations of each element are shown in Figure 3. The structure in WinFEKO surroundings of KEKO 4.0 is shown in Figure 4. There are 8 elements of an array on the surface of a cylinder in line. The distance between different elements is 0.5λ. The detailed geometric parameters of the cylinder and the array are shown in Table 1.

When calculating the current element, other elements are regarded as metal scatterers. The point numerical of radiation pattern is the sum of which 8 elements contributed. The result of the array radiation pattern is shown in Figure 5 (a). A numerical result from the reference (Lei XIAO, 2004, 30) is shown in Figure 5 (b). The two curves in Figure 5 coincide

with each other. Because of its structural symmetry of the conformal phased array, the antenna patterns in the opposite direction to the same point of view, the pattern shows symmetrically.

Furthermore, the patterns of different scanning angle for this conformal phased array are simulated (see Figure 6). The element excitations of the phased array are identical currents and linear phases, and some important parameters of the array patterns when scanning are shown in Table 2.

Several significant results can be induced from Figure 6 and Table 2. The first side lobe level (SLL1) of the array pattern is higher when the scanning angle rises. With increasing scanning angle, the main lobe of the pattern will be wider; the sidelobe of the pattern will rise gradually. By the inter-element mutual coupling, as well as the combined effect of carrier, the direction of the main beam will appear to deviation while scanning angle equal to 45 degrees. When scanning large angle the array pattern have begun to deteriorate, the main sidelobe has gradually blurred the boundaries of the main beam when scanning angle equal to 60 degrees.

4. Conclusions

In general, a good radiation pattern of a conformal array antenna is simulated with FEKO 4.0. The results show that FEKO 4.0 is a good tool in the simulation of conformal array antennas. The results simulated with FEKO 4.0 can meet with which in related reference and some significant parameters from conformal phased array patterns by simulation are given. This study has great significance to conformal phased array in radar engineering.

References

Alexander Svezhentsev. (2004). RADIATION CHARACTERISTICS OF CYLINDRICAL MICROSTRIP ARRAY ANTENNA WITH TWO SYMMETRICALLY PLACED ELEMENTS. *MSMW'04 Symposium Proceedings*, Kharkov. Ukraine, June 21-26, 629-631.

J.A. Ferreira and F. Ares. (1997). Pattern synthesis of conformal arrays by the simulated annealing technique. *IEE Electronics Letters*, Vol. 33 No.14, 1187-1189.

Lars Josefsson, Patrik Persson. (2006). *CONFORMAL ARRAY ANTENNA THEORY AND DESIGN. A WILEY-INTERSCIENCE PUBLICATION*, Chapter 1.

Lei XIAO. (2004). *Analysis of Mirostrip Conformal Antenna Based On Cylindrical Carrier*. Master's thesis of Electronic Science University of China, 2004, 20-51.

Table 1. The parameters of the conformal phased array

Cylinder length	Cylinder radius	Patch dimension	Adjacent distance	Center frequency
4λ	0.5λ	0.5λ*0.5λ	0.5λ	2G Hz

Table 2. Parameters of the array patterns with different scanning angles

Scanning angles (degree)	SLL1 (dB)	3dB beam-width (degree)	Warp of main-beam direction (degree)
0	-13	13	0
15	-12.4	13.5	0
30	-11.8	15	0
45	-10.8	18	1.5
60	-9	22	4.2

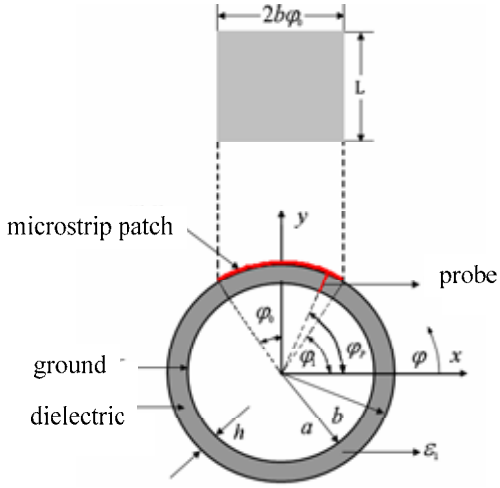
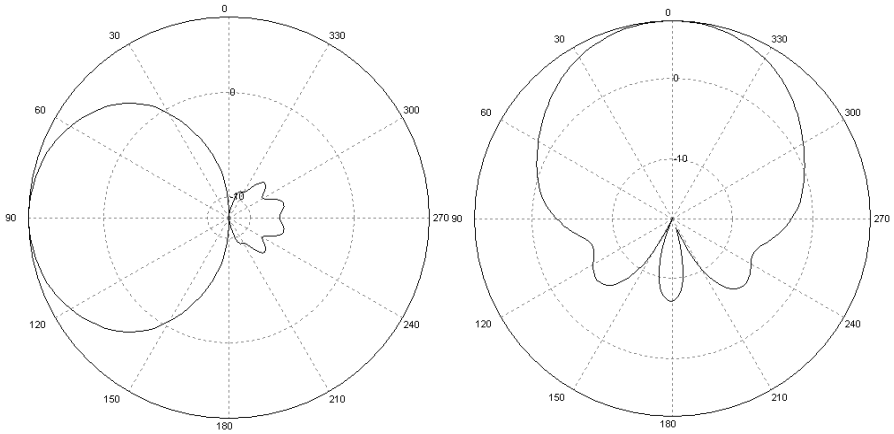


Figure 1. Cross-section diagram of the conformal microstrip antenna



(a) Plane E (b) Plane H

Figure 2. The radiation patterns for conformal element of an array

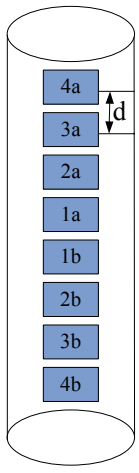


Figure 3. Conformal array of 8 array elements placed on finite length cylinder

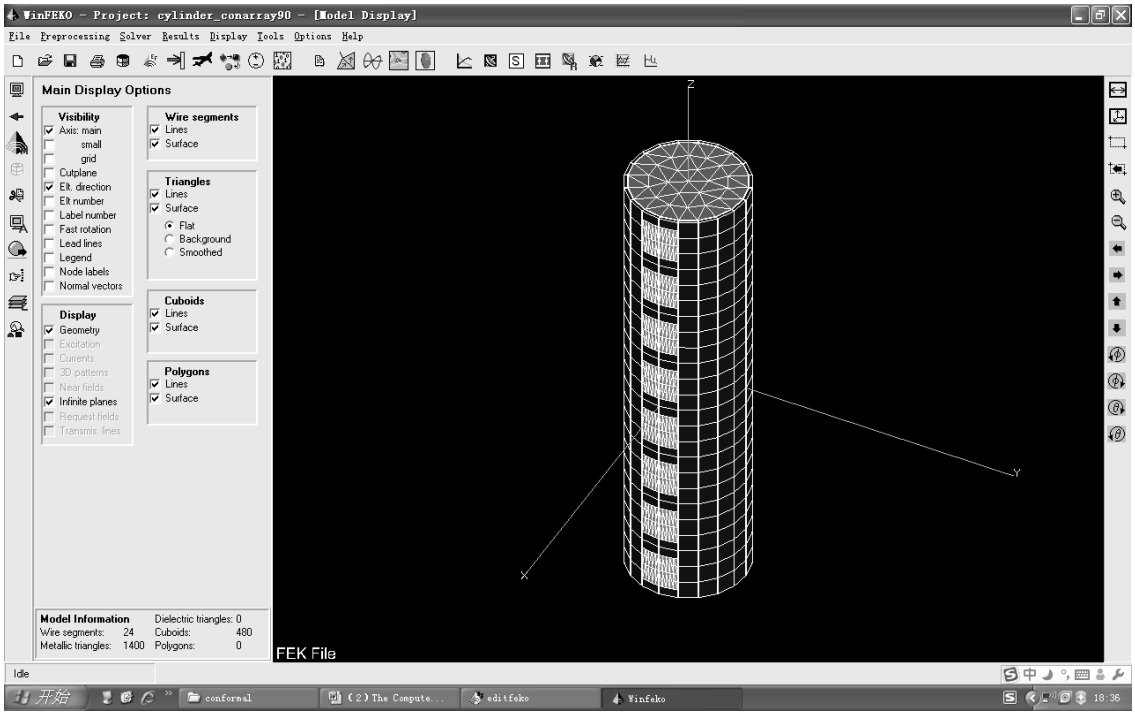
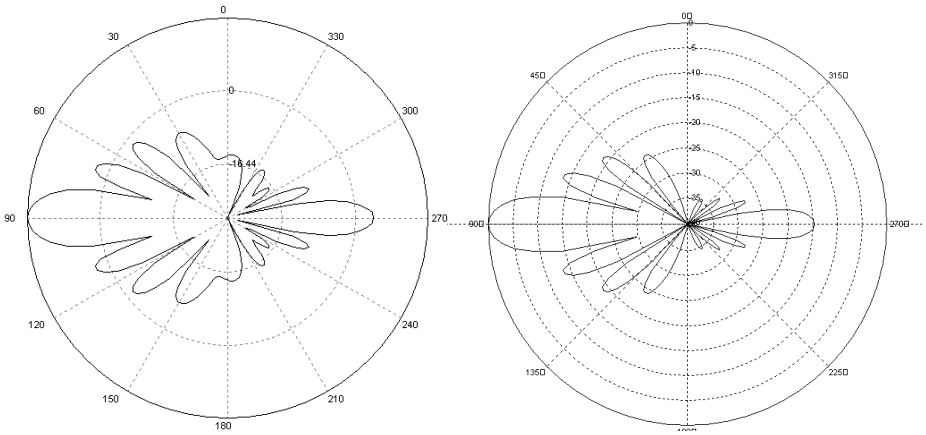
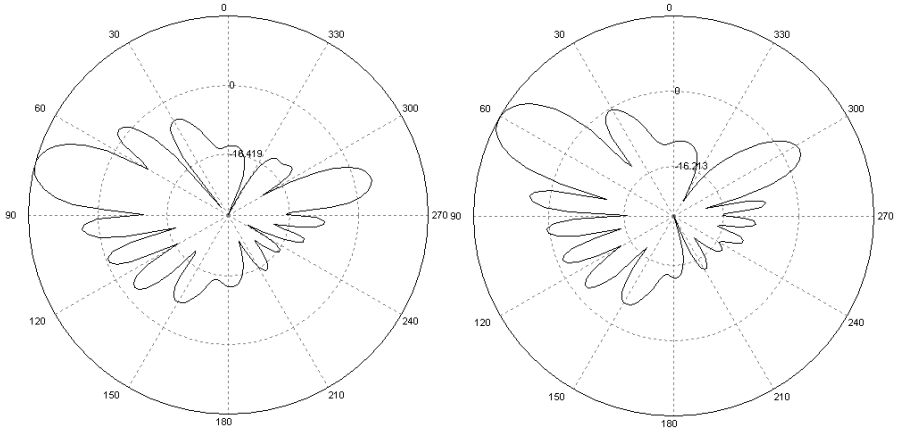


Figure 4. The structure in WinFEKO surroundings of KEKO 4.0

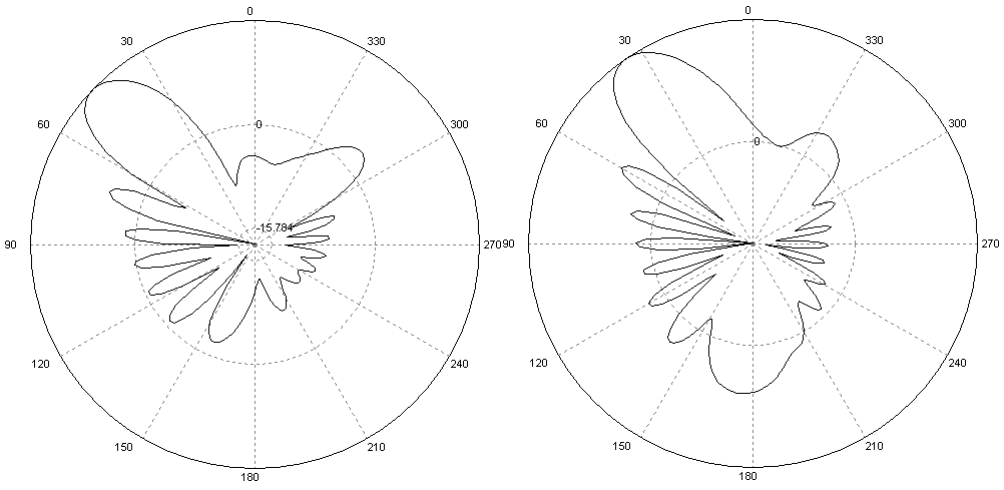


(a) (b)

Figure 5. The results of the array radiation pattern



(a) $SA = 15^{\circ}$ (b) $SA = 30^{\circ}$



(c) $SA = 45^{\circ}$ (d) $SA = 60^{\circ}$

Figure 6. The patterns of different scanning angle (SA) for this conformal phased array

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Analysis of Particle Swarm Optimization Algorithm

Qinghai Bai
College of Computer Science and Technology
Inner Mongolia University for Nationalities
Tongliao 028043, China
Tel: 86-475-239-5155 E-mail: baiqh68@163.com

Abstract

Particle swarm optimization is a heuristic global optimization method and also an optimization algorithm, which is based on swarm intelligence. It comes from the research on the bird and fish flock movement behavior. The algorithm is widely used and rapidly developed for its easy implementation and few particles required to be tuned. The main idea of the principle of PSO is presented; the advantages and the shortcomings are summarized. At last this paper presents some kinds of improved versions of PSO and research situation, and the future research issues are also given.

Keywords: Particle swarm optimization Algorithm, Swarm intelligence, Heuristic

1. Introduction

Particle swarm optimization is a heuristic global optimization method put forward originally by Doctor Kennedy and E berhart in 1995(Kennedy J,Eberhart R,1995;Eberhart R,Kennedy J,1995) It is developed from swarm intelligence and is based on the research of bird and fish flock movement behavior. While searching for food, the birds are either scattered or go together before they locate the place where they can find the food. While the birds are searching for food from one place to another, there is always a bird that can smell the food very well, that is, the bird is perceptible of the place where the food can be found, having the better food resource information. Because they are transmitting the information, especially the good information at any time while searching the food from one place to another, conduced by the good information, the birds will eventually flock to the place where food can be found. As far as particle swam optimization algorithm is concerned, solution swam is compared to the bird swarm, the birds' moving from one place to another is equal to the development of the solution swarm, good information is equal to the most optimist solution, and the food resource is equal to the most optimist solution during the whole course. The most optimist solution can be worked out in particle swarm optimization algorithm by the cooperation of each individual. The particle without quality and volume serves as each individual, and the simple behavioral pattern is regulated for each particle to show the complexity of the whole particle swarm. This algorithm can be used to work out the complex optimist problems.

Due to its many advantages including its simplicity and easy implementation, the algorithm can be used widely in the fields such as function optimization, the model classification, machine study, neutral network training, the signal procession, vague system control, automatic adaptation control and etc(Zheng Jianchao,Jie Jing,Cui Zhihua,2004,(In Chinese)).

2. Basic Particle Swarm Optimization Algorithm

In the basic particle swarm optimization algorithm, particle swarm consists of "n" particles, and the position of each particle stands for the potential solution in D-dimensional space. The particles change its condition according to the following three principles:

(1) to keep its inertia (2) to change the condition according to its most optimist position (3) to change the condition according to the swarm's most optimist position.

The position of each particle in the swarm is affected both by the most optimist position during its movement (individual experience) and the position of the most optimist particle in its surrounding (near experience). When the whole particle swarm is surrounding the particle, the most optimist position of the surrounding is equal to the one of the whole most optimist particle; this algorithm is called the whole PSO. If the narrow surrounding is used in the algorithm, this algorithm is called the partial PSO.

Each particle can be shown by its current speed and position, the most optimist position of each individual and the most optimist position of the surrounding. In the partial PSO, the speed and position of each particle change according the following equality (Shi Y,Eberhart R C,1998):

$$v_{id}^{k+1} = v_{id}^{k} + c_1 r_1^k \ (pbest_{id}^k - x_{id}^k) + c_2 r_2^k \ (gbest_d^k - x_{id}^k) \quad (1)$$

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \quad (2)$$

In this equality, v_{id}^{k} and x_{id}^{k} stand for separately the speed of the particle "i" at its "k" times and the d-dimension quantity of its position; $pbest_{id}^{k}$ represents the d-dimension quantity of the individual "i" at its most optimist position at its "k" times. $gbest_{d}^{k}$ is the d-dimension quantity of the swarm at its most optimist position. In order to avoid particle being far away from the searching space, the speed of the particle created at its each direction is confined between $-v_{dmax}$, and v_{dmax}. If the number of v_{dmax} is too big, the solution is far from the best, if the number of v_{dmax} is too small, the solution will be the local optimism; c1 and c2 represent the speeding figure, regulating the length when flying to the most particle of the whole swarm and to the most optimist individual particle. If the figure is too small, the particle is probably far away from the target field, if the figure is too big, the particle will maybe fly to the target field suddenly or fly beyond the target field. The proper figures for c1 and c2 can control the speed of the particle's flying and the solution will not be the partial optimism. Usually, c1 is equal to c2 and they are equal to 2; r1 and r2 represent random fiction, and 0-1 is a random number.

In local PSO, instead of persuading the optimist particle of the swarm, each particle will pursuit the optimist particle in its surrounding to regulate its speed and position. Formally, the formula for the speed and the position of the particle is completely identical to the one in the whole PSO.

3. An Analysis on the Advantages and Disadvantages of the Basic Particle Swarm Optimization Algorithm.

Advantages of the basic particle swarm optimization algorithm:

(1)PSO is based on the intelligence. It can be applied into both scientific research and engineering use.

(2)PSO have no overlapping and mutation calculation. The search can be carried out by the speed of the particle. During the development of several generations, only the most optimist particle can transmit information onto the other particles, and the speed of the researching is very fast.

(3)The calculation in PSO is very simple. Compared with the other developing calculations, it occupies the bigger optimization ability and it can be completed easily.

(4) PSO adopts the real number code, and it is decided directly by the solution. The number of the dimension is equal to the constant of the solution.

Disadvantages of the basic particle swarm optimization algorithm:

(1)The method easily suffers from the partial optimism, which causes the less exact at the regulation of its speed and the direction.

(2)The method can not work out the problems of scattering and optimization(Chen Yonggang,Yang Fengjie,Sun Jigui ,2006,(In Chinese)).

(3)The method can not work out the problems of non-coordinate system, such as the solution to the energy field and the moving rules of the particles in the energy field

4. The Present Research Situation of the Particle Swarm Optimization Algorithm

The PSO method is based on swarm intelligence. The research on it is just at the beginning. Far from the Genetic algorithm (GA) and the simulated annealing (SA) approach, the POS has no systematical calculation method and it has no definite mathematic foundation. At present, the method can only be used successfully in the aspect of Evolutionary neural network, and its other applications are still being explored. By the national documents on it, the research on PSO concerns mainly the mathematic foundation and application research. The mathematic foundation includes the mechanical principle of PSO itself, the prove of its convergence and Robustness and etc. In the publicly published documents, there are fewer documents about the study on its mathematic foundation, the prove on the convergence and the estimate of the speed of the convergence has not been found., which demands the research on the PSO should be perfected; The application research involves continuing its advantages, overcoming its shortcomings and developing its application ranges. The study on PSO should be concentrated on the following :some modern technologies should be applied to PSO to design the improved PSO; PSO can be combined with the other intelligent optimization methods to design several compound optimization methods; PSO can be also led into scattering system, compound optimist system, non-coordinate system to develop PSO's application ranges.

5. The Improvement of Particle Swarm Optimization Algorithm

5.1 Inertia weights

Inertia weights is put forward by Shi and others (Eberhart R C,Shi Y,1998; Eberhart R C,Shi Y,2000). An Inertia weight ω is a proportional agent that is related with the speed of last time, and the formula for the change of the speed is the following:

$$v_{id}^{k+1} = \omega v_{id}^{k} + c_1 r_1^{k} (pbest_{id}^{k} - x_{id}^{k}) + c_2 r_2^{k} (gbest_{d}^{k} - x_{id}^{k})$$

The influence that the last speed has on the current speed can be controlled by inertia weights. The bigger ω is, the bigger the PSO's searching ability for the whole is, and the smaller ω is, the bigger the PSO's searching ability for the partial. Generally, ω is equal to 1, so at the later period of the several generations, there is a lack of the searching ability for the partial. Experimental results show that PSO has the biggest speed of convergence when ω is between 0.8 and 1.2. While experimenting, ω is confined from 0.9 to 0.4 according to the linear decrease, which makes PSO search for the bigger space at the beginning and locate the position quickly where there is the most optimist solution. As ω is decreasing, the speed of the particle will also slow down to search for the delicate partial. The method quickens the speed of the convergence, and the function of the PSO is improved. When the problem that is to be solved is very complex, this method makes PSO's searching ability for the whole at the later period after several generation is not adequate, the most optimist solution can not be found, so the inertia weights can be used to work out the problem.

5.2 Increase Convergence Factor

A particle swarm optimization algorithm with convergence agents is introduced in paper (Clerc M,1999), and the following is the formula for its position and speed changing:

$$v_{id} = \chi \{v_{id} + c_1 \text{ rand}( )(p_{id}-x_{id})+c_2\text{rand}( )(p_{gd}-x_{id})\}$$

$$\chi = \frac{2}{|2-\varphi-\sqrt{\varphi^2-4\varphi}|}$$

is called the convergence factor, $\varphi = c_1+c_2 > 4$. Generally, φ is equal to 4.1, so χ is equal to 0.729. The experimental result shows compared with the particle swarm optimization algorithm with inertia weights, the convergence speed in the particle swarm optimization algorithm with the convergence agent is much quicker. In fact, when the proper ω, c_1 and c_2 is decided, the two calculation methods are identical. So, the particle swarm optimization algorithm with convergence agent can be regarded as a special example of the particle swarm optimization algorithm with inertia weights. Meanwhile, the properly selected parameters in the algorithms can improve the function of the methods.

5.3 Selection

The compound PSO put forward by Angeline is based on the basic mechanism and the selection mechanism created during the development of the computers (Angeline P J, 1999). Due to POS's depending on pbest and gbest during its searching, the area to be searched will be confined greatly. The introduction of the selection mechanism will solve the problem gradually. The test result shows although selection has better effect than basic PSO in the most tested functions, the result is less satisfying as far as the function "Griewank"is concerned. As a result, this method improve PSO's searching ability for the partial, meanwhile, it makes the searching for the whole area less powerful.

5.4 The Blending of the PSO Algorithm and the Other Intelligent Algorithms

The main process of the particle swarm optimization algorithm put forward by Gaoying based on depends on the main process of basic particle swarm optimization algorithm (Gao Ying,Xie Shengli ,2004). To introduce the simulated annealing (SA) approach, hybrid algorithm in the hybrid particle swarm optimization algorithm and mutation algorithm in the mutation particle swarm optimization algorithm are adapted to regulate further the optimized swarm. Angeline introduces the selection particle and the better particles selected after each generation is reproduced into the next generation to ensure the particle swarm has the better property. This algorithm has a better effect on the single peaks function. Higash(Higashi N,Iba H ,2003) and the other persons put forward their own mutation algorithms. By introducing the mutation particle, the algorithms escape the attraction of the optimized point in the partial area to improve the searching ability for the whole area. Inspired by the ants' behavioral pattern in their searching food, Italian Colorni and Dorigo(Colorni A,Dorigo M,Maniezzo V,et al,1991;Dorigo M,Maniezzo V,Colorni A,1996),ACO put forward originally Ant Colony Optimization, another kind of intelligent optimization algorithm. The algorithm is based on the research on the behavior of the ant colony's searching for food, and the algorithm simulates the virtual ant colony's cooperation. The behaviors of the several ants consist the routes of the solution together; the optimization can be achieved by leaving and exchanging information in the routes to find the more exact solution. Duan Haibin(Duan Haibin,2005, (In Chinese)) makes a further exploration into the ant colony's algorithm.

6. The Conclusion and the Future Research on PSO

Particle swam optimization is a new heuristic optimization method based on swarm intelligence. Compared with the other algorithms, the method is very simple, easily completed and it needs fewer parameters, which made it fully developed. However, the research on the PSO is still at the beginning, a lot of problems are to be resolved. The research on PSO will be mainly concentrated on the following:

(1)The math's basic theory of the Algorithm

Although PSO's application has been proved to be effective, its theoretical foundation is rather weak. Clerc and Kennedy(Clerc M,Kennedy J ,2002)make a analysis on the convergence of the method from the point of math's. By analyzing the stability of the condition transmitting matrix, they find the limited conditions where the particle can move stably. Based on this, Bergh makes the further analysis on it. Lebesgue and Borel explore the effect of casualty on the locus of the particle, and analyze the convergence from the point of measuring space.

Still, there is no mathematically proved about the convergence and the speed of the convergence. The most optimist solution of PSO can not be ensured in theory.

(2) Topology of the particle swarm

Research on the topology of the new pattern particle swarm which has a better function can be carried out. The neighboring topology of the different particle swarms are based on the imitation of the different societies. It is meaningful to the use and spread of the algorithm to select the proper topology to enable PSO have the best property and do the research on the suitable ranges of different topologies.

(3) The blending with the other intelligent optimization algorithm

Blending PSO with the other intelligent optimization algorithms means combining the advantages of the PSO with the advantages of the other intelligent optimization algorithms to create the compound algorithm that has practical value. For example, the particle swarm optimization algorithm can be improved by the simulated annealing(SA) approach ; It can be connected with the hereditary agents, the algorithm of a colony of ants, vague method and etc.

(4)Develop the application area of the Algorithm

The effect can be found out in the practical application. Although the PSO algorithm has been used widely, it will be very meaning to explore the developing area further. At present, the most research on PSO aim at the coordinate system. Although in practical usage, it is used in non-coordinate system, scattered system and compound optimization system, there is less research on the PSO algorithm application in these systems.

References

Angeline P J. (1999). Using selection to improve Particle Swarm Optimization.Proceedings of the 1999 Congress on Evolutionary Computation.Piscataway. NJ:IEEE Press,1999:84-89.

Chen Yonggang,Yang Fengjie,Sun Jigui. (2006). A new Particle swam optimization Algorithm. *Journal of Jilin University*, 2006, 24(2):181-183. (In Chinese)

Clerc M, Kennedy J. (2002). The particle swarm-explosion stability,and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 2002, 6(1):58-73.

Clerc M. (1999). The swarm and the queen:towards a deterministic and adaptive particle swarm optimization.Proceedings of the Congress on Evolutionary Computation. Piscataway, NJ:IEEE Service Center,1999:1951-1957.

Colorni A, Dorigo M, Maniezzo V, et al. (1991). Distributed optimization by ant colonies. Proceedings of the 1st European Conference on Artificial Life,1991:134-142.

Dorigo M,Maniezzo V, Colorni A. (1996). Ant system:optimization by a colony of cooperationg agents. IEEE Transactions on Systems. Man, and Cybernetics-Part B,1996,26(1):29-41.

Duan Haibin. (2005). Ant Colony Optimization theory and Application. Science Publishing Company of Beijing,2005. (In Chinese)

Eberhart R C, Shi Y. (1998). Comparison between genetic algorithms and Particle Swarm Optimization.Porto V W,Saravanan N,Waagen D, et al.Evolutionary Programming VII.[S.l.]:Springer,1998:611-616.

Eberhart R C, Shi Y. (2000). Comparing inertia weights and constriction factors in Particle Swarm Optimization.Proceedings of the Congress on Evolutionary Computating, 2000: 84-88.

Eberhart R, Kennedy J. (1995). A New Optimizer Using Particle Swarm Theory.Proc of 6th International Symposium on Micro Machine and Human Science, Nagoya, Japan. IEEE Service Center Piscataway NJ,1995:39-43.

Gao Ying, Xie Shengli. (2004). Particle swam optimization Algorithm based on simulated annealing(SA) approach .Computer Engineering and Application, 2004,40(1):47-49. (In Chinese)

Higashi N, Iba H. (2003). Particle Swarm Optimization with Gaussian mutation.Proceedings of the 2003 Congress on Evolutionary Computation.Piscataway,NJ:IEEE Press,2003:72-79.

Kennedy J, Eberhart R. (1995). Particle Swarm Optimization.Proc of IEEE International Conference on Neural Network,Perth,Australia,IEEE Service Center Piscataway NJ,1995:1942-1948.

Shi Y, Eberhart R C. (1998). A modified particle swam optimizer. IEEE Word Congress on Computational Intelligence, 1998: 69-73.

Zheng Jianchao, Jie Jing, Cui Zhihua. (2004). *Particle swam optimization Algorithm*[M].Science Publishing Company of Beijing, 2004.(In Chinese)

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Ontologies Acquisition from Relational Databases

Shufeng Zhou (Corresponding author)
College of Mathematics and Science, Liaocheng University
No.34 Wenhua Road, Liaocheng 252059, Shandong, China
E-mail: zhousflc@sohu.com

Guangwu Meng
College of Mathematics and Science, Liaocheng University
No.34 Wenhua Road, Liaocheng 252059, Shandong, China

Haiyun Ling
College of Computer Science, Liaocheng University
No.34 Wenhua Road, Liaocheng 252059, Shandong, China

The research is financed by Natural Science Foundation of Liaocheng University(X051034).

Abstract

Ontologies play an important role in solving the problem of semantic heterogeneity of heterogeneous data sources. The ontology acquisition from the relational database is one of the fundamental technologies on the information integration field. Ontology acquisition from relational database (OARDB) and transformation rules used are presented and discussed. The key technologies and model are given.

Keywords: Ontology, Relational Database, Rules, Semantic Web

1. Introduction

The Semantic Web proposed by Tim Berners-Lee has been viewed as the next generation of the current Web. And with the development of Semantic Web research, people have realized that the success of Semantic Web depends on the proliferation of ontologies and pay more attention to the construction of ontologies. The popularity of ontologies is rapidly growing. However, most of the world's data today are still stored in relational databases. Therefore, it is necessary to acquire ontologies from relational databases.

Though ontology construction tools have become mature over the years, the manual development of ontologies still remains a tedious and cumbersome task. So this paper proposes how to generate ontology automatically or semi-automatically from relational databases.

The paper is organized as follows. Section 2 introduces related works. Section 3 analyzes the ontology acquisition from relation database. Section 4 introduces the implementation of OARDB in detail. And section 5 gives a summary and the future works.

2. Related works

A number of researchers have made contribution to this problem, such as defining semantics in database schema, extracting semantics out of database schema, and serve as a foundation for our work. But there does not exist hierarchies and cardinality about properties in those previous approaches. So these approaches cannot be used to describe the ontology from relational database directly and correctly. In this paper, we analyze the database schema information firstly. Then systematically present how relational database can be transformed into ontologies.

3. Ontologies Acquisition from Relational Databases

In this section, the paper explains the ontologies acquisition from relational databases. First, the paper lists the fundamental condition to express the ontologies acquisition. Then, explains steps to acquire the ontologies from relational databases.

3.1 Fundamental condition

In order to acquire ontologies from relational databases, the paper makes the assumption: The relational database schema is normalized, at least up to third normal form. In fact, the third normal form is the most common normal form in relational database schema. So if some databases might not be well normalized, it is possible to automate the process of finding functional dependencies within data and to algorithmically transform a relational database schema to third normal form.

3.2 Steps to acquire the ontologies

The process of acquiring ontologies can be divided into two stages: Extracting Relational Database Schema Information and Acquiring Ontology. In the first stage, the paper adopts the reverse engineering methods to extract relational database schema information. It is the key stage to acquire the ontology, and then based on the relational database schema information ontological structure can be constructed. After that, the relational database tuples can be retrieved and formed as the ontological instances. So the OARDB generally consists of four steps.

Step 1: Extracting relational database schema information. Such as relation names, attribute names, primary keys, foreign keys and integrity constraints.

Step 2: Analysis of primary keys, foreign keys and attributes information, and then construct ontology and ontological concept.

Step 3: Retrieving tuples from relational database.

Step 4: Mapping the tuples to ontological instances and forms knowledge base.

In the four steps, the step 1 and step 2 are the transformation of relational schema, which can be extracted using Java API. Based on the relational schema information, the ontology and ontological concept are constructed. And the step 3 and step 4 are the transformation of relational data. After step 1 and step 2, an ontological structure is formed. Then we can use the relational tuples as the ontological instances according the transformation rules.

4. Implementation of OARDB

According to schema information extracted from relational database, the paper firstly analyzes the relationship between relational schemas, and then classifies the relational schema. Finally, the rules for acquiring ontology are given below.

4.1 Transformation from relational database schema information to ontology

Schema 1: For relation R, suppose that there is no foreign key. Such as, Course(courseId, courseName, credits, creditHours). This is the most basic relational schema, the corresponding transformation rules are as follows.

Rule 1: A Class can be created, using the relation name as the name of Class.

Rule 2: All attributes domain in relational schema are mapped into xsdDatatype.

Rule 3: All attributes in relational schema are mapped into DatatypeProperty.

Schema 2: For relations R_1, R_2 in database, suppose that the primary key of R_1 consists of only one foreign key referring to R_2, such as PhdStudent(phdId, memo), phdId referring to Student(id, name, birthday,…). Then, in addition to abide by the first schema, we must express the subclass relationship between the two relations. So the rule is as follows.

Rule 4: The class corresponding to R_1 is a subclass of the class corresponding to R_2.

Schema 3: For relations R_1, R_2, R_3, if the primary key of R_1 is consisted of only two foreign keys A_{11}, A_{12} referring to R_2, R_3 respectively, and there is no other attribute except the primary key of R_1, such as Teaching(teacherId, courseId). In this case, the relation R_1 is used to describe the many-to-many relationship between two relations. Then, two inverse object properties can be created based on semantics of the relation. The rule of transformation is as follows.

Rule 5: Two object properties P_{23} and P_{23}' are created based on the semantics of R_1. Suppose that the classes corresponding to R_2, R_3 are C_2, C_3 respectively. The domain and range of P_{23} are C_2 and C_3, and the domain and range of P_{23}' are C_3 and C_2. At the same time, P_{23} and P_{23}' are two inverse object properties.

Schema 4: All schemas with the exception of above schemas. These relational schemas are divided into four cases.

Case 1: For a relation R, the number of the foreign key is equal or greater than 3.

Case 2: For relations R_1, R_2, R_3, if the primary key of R_1 is consisted of only two foreign keys A_{11}, A_{12} referring to R_2, R_3 respectively, and there are other attributes except the primary key of R_1.

Case 3: For a relation R, there is a foreign key. But the foreign key is not integral part of the primary key.

Case 4: For a relation R, there is a foreign key. And the primary key of R is consisted of the foreign key and other attributes of R.

In above four cases, all relations can be transformed into classes according Schema 1. In addition to describe the relationship between relations, inverse object properties can be created applying rule 5.

In addition, the following rules can be created to describe the cardinality constraint on properties.

Rule 6: For a relation, if attribute A (except foreign key) is declared as NOT NULL, then the cardinality of the property corresponding to A is 1.

Rule 7: For a relation, if attribute A (except foreign key) is declared as UNIQUE, then the maximal cardinality of the property corresponding to A is 1.

Rule 8: For a relation, if foreign key F is declared as NOT NULL, then the minimal cardinality of the object property corresponding to F is 1.

Rule 9: For a relation, if foreign key F is declared as NULL sometimes, then the minimal cardinality of the object property corresponding to F is 0.

4.2 Transformation from relational database tuples to ontological instances

Applying above rules, an ontological structure can be extracted from a relation schema. Then, the process of transformation from relation database tuples to ontological instances can start. The rules are as follows.

Rule 10: For a relation R, suppose that class C is corresponding to R, then every tuple t can be mapped to a ontological instance associated with unique identifier by appending the value of the primary key to the name of the relational name.

Rule 11: For a relation R, the values of the tuple can be mapped to the values of the corresponding property of ontological instance.

Rule 12: For a relation R, suppose that there a foreign key F, then the value of F can be mapped to ontological instance I, and the I can be mapped to the value of the object property corresponding to the foreign key.

5. Summary and Future Works

In summary, the main contributions of this paper are listed as follows. Firstly, we have presented a new approach to acquire ontology from relational database. It captures semantic information contained in the structures of the entities. Secondly, we have experimentally evaluated our approach on several data sets from real world domains. The results demonstrate that our approach performs well as compared to some existing approaches.

In the future work, we look forward to comparing our approach with some intermediate approaches. We also hope to consider some machine learning techniques for mining some other interesting and useful semantic mappings.

References

An, Y., Borgida, A., Mylopoulos, J. (2005). Inferring complex semantic mappings between relation tables and ontologies from simple correspondences. In Proceedings of International Conference on Ontologies, Databases and Applications of Semantics (ODBASE'05).1152-1169.

Astrova, Irina. (2004). Reverse Engineering of Relational Databases to Ontologies. *The Semantic Web: Research and Applications*, 327-341.

Berners-Lee, T., Hall, W., Hendler, J., (2006). Creating a science of the Web. Science.769-771.

Chen,H., Wu, Z., Wang, H., Mao, Y. (2006). RDF/RDFS-based relational database integration. In Proceedings of the 22nd International conference on Data Engineering (ICDE'06). 94.

Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Synthesis of Shaped Beam for Base-station Antennas in Cellular Mobile Communications

Ruying Sun
School of Informatics, Linyi Normal University, Linyi 276005, China
E-mail: srysd@163.com

Abstract

Genetic Algorithms is used to synthesize shaped beam and optimize base-station antenna coverage in the article. Two cases have been studied: one is to change the feeding amplitude and phase of each element; the other is to change the feeding phase of each element and the distances between them. And then we combine the Genetic Algorithms with FEKO to synthesize antenna patterns. Comparing to the Woodward method, the results achieved by Genetic Algorithms are better than Woodward method because it is nearly to the ideal square-cosecant patterns.

Keywords: Shaped Beam, Antenna Synthesis, Base-station Antennas, Genetic Algorithms, FEKO 4.0

1. Introduction

The object functions for synthesis of array antennas usually have the characteristics of multi-parameters, non-differentiable even discontinuities. The optimization of pattern function is non-linear problem. Traditional optimization techniques search for the best solutions, using gradients or random searching. Gradient methods are efficient, but have disadvantages of getting stuck in local minima, requiring gradients calculations, working only on continuous parameters. Random-search methods don't require gradient calculations, but tend to be slow, and susceptible to getting stuck in local minima. Genetic-algorithm optimizer is robust global search method. Its research is based on probability, having advantage of avoiding getting stuck in local minima (Holland J.H, 1992, chapter3).

In mobile communications system, the radiation pattern of base-station antenna should have weak interference to adjacent areas, and generate intense and uniform radiation Intensity in cellular area (Rongcang HAN, 2006, 34). According to the system request, a particular type of elevation pattern is achieved by technology of beam forming

2. Methods

Here, there is a quick overview of genetic algorithms, much more detail on genetic algorithms is found in reference (Randy L, Haupt, 1994, 995).Genes are the basic building blocks of genetic algorithms. A gene is binary encoding of a parameter. In computer algorithm, a chromosome is an array of genes, a number of chromosomes make up one population. Each chromosome has an associated fitness function, assigning a relative merit to that chromosome.

The algorithm begins with a large list of random chromosomes. Fitness functions are evaluated for each chromosome. The chromosomes are ranked from the most-fit to the least-fit, according to their respective fitness functions. Unacceptable chromosomes are discarded, leaving a superior species-subset of an original list, which is the process of selection. Genes that survive become parents, by crossing over some of their genetic material to produce two new offspring. The parents reproduce enough to offset the discarded chromosomes. Thus, the total number of chromosomes remains constant after next iteration. Mutations cause small random changes in a chromosome. Fitness functions are evaluated for the offspring and mutated chromosome, and the process is repeated. The algorithm stops after a set number of iterations, or when an acceptable solution is obtained. Figure.1 is a flow chart of Genetic Algorithms (K.Markus, L.Vaskelainen., 1998, 461).

Suppose we consider an array of antenna elements uniformly spaced in a straight line along the z axis. The far-field radiation pattern produced by such an array may be expressed as

$$E(\theta,\varphi)=\sum_{n=0}^{N-1} I_n EP_n(\theta,\varphi) e^{j(nkd\cos\theta+\beta_n)} \tag{1}$$

where I_n, β_n are the element excitation current amplitudes and phases, $EP_n(\theta,\varphi)$ are the individual array element patterns, k=2π/λ is the free-space wave number, d is the separation distance between elements.

If the common assumption is made that all of the elements have identical element patterns, then equation (1) may be written in the form of

$$E(\theta,\varphi) = EP(\theta,\varphi) AF(\theta) \tag{2}$$

$$AF(\theta) = \sum_{n=0}^{N-1} I_n e^{j(nkd\cos\theta+\beta_n)} \tag{3}$$

Where $AF(\theta)$ is the associated array factor (Diógenes Marcano, and Filinto Durán, 2000, 14).

Now the main process of antenna synthesis in genetic algorithms is given below.

A. Establish decision variable and constraint condition, then encoding them.

$$I_n \in [0,1], \beta_n \in [0, 2\pi], n = 0,1,2,\cdots N-1 \tag{4}$$

B. Create optimization pattern:

$$\min E_{mt} = \left[\frac{1}{Q} \sum_{i=1}^{Q} |e_i|^2 \right]^{\frac{1}{2}} \tag{5}$$

where

$$e_i = \frac{T_i - F_i}{T_i}, i = 1,2,\cdots,Q \tag{6}$$

T_i is the level of the desired radiation pattern at the point Q, and F_i is the level of the pattern generated by genetic algorithms.

C. Define fitness function:

$$F_m = \frac{1}{1+E_{mt}^{\alpha}} \qquad \alpha \in (0,1] \tag{7}$$

D. Define object function according to the characteristics of desired radiation pattern.

3. Numerical Examples

Now, we define an object function according to the request of base-station antenna in order to achieve intense and uniform radiation intensity in cellular area. The object function is defined as equation 8.

$$F_{\circ} = \begin{cases} \sec\theta, \theta \in [0^{\circ}, 90^{\circ}) \\ 0 \quad , \quad \text{else} \end{cases} \tag{8}$$

Here $N = 8$, $d = \lambda/2$, array elements are 8 half-wave dipole, which structure in WinFEKO surroundings is described in Figure 2. Combining GA with FEKO 4.0 can take into account both the radiation pattern of array elements and the mutual coupling among them when synthesizing antenna array patterns.

Figure 3 shows the result of base-station antenna pattern using Genetic Algorithms in solid curve and the result of Woodward in dashed curve. The major lobe of both array factor patterns direct ground. The radiation of electromagnetic energy is more intensive and uniform in array factor produced by genetic algorithms than Woodward.

4. Conclusions

Anyway, Genetic Algorithms is a good method to synthesis of shaped beam. The synthesis of shaped beam for the base-station array antenna with 8 elements is a good case. The result achieved by Genetic Algorithms is nearly to the ideal square-cosecant patterns, which is better than Woodward method. Genetic Algorithms also can be used to synthesize arrays with unequal distances between elements. It can be applied in other topics of antenna synthesis.

References

Diógenes Marcano, and Filinto Durán, (2000). Synthesis of Antenna Array Using Genetic Algorithms. *IEEE, Antennas and Propagation Magazine*, vol.42, no.3, pp12~20.

Holland J.H. (1992). *Adaptation in Nature and Artificial Systems*. MIT Press, chapter3.

K.Markus, L.Vaskelainen. (1998). Optimization of synthesized array excitations using array polynome complex root swapping and genetic algorithms. *IEE Proc-Microw Antennas Propag*, vol.145, no.6, pp460~464.

Randy L, Haupt. (1994). Arrays Using Genetic Algorithm. *IEEE Transactions on Antennas and Propagation*, vol.42, no.7, pp993~999.

Rongcang HAN. (2006). Synthesis of Array Antenna Based on Genetic Algorithms. *Master's thesis of University of Electronic Science and Technology of China (UESTC)*, 34

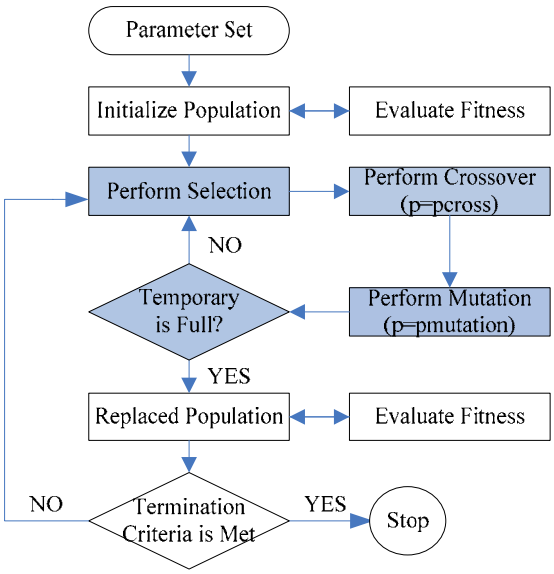


Figure 1. A flow chart of Genetic Algorithms

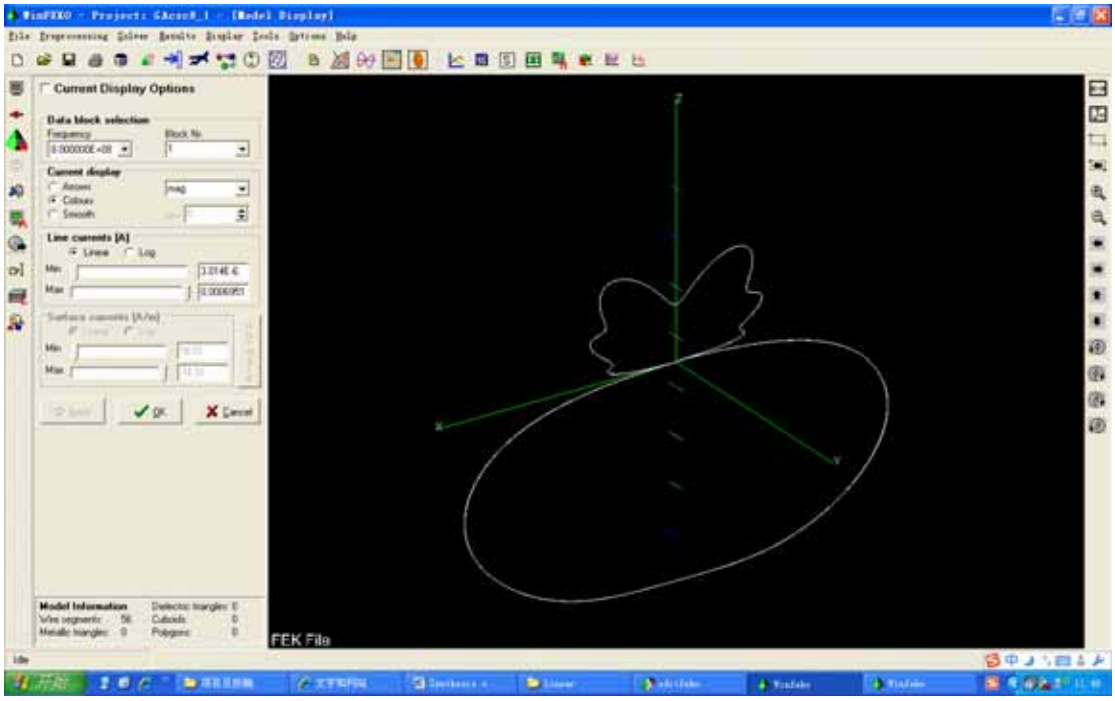


Figure 2. The structure of base-station antenna in WinFEKO surroundings

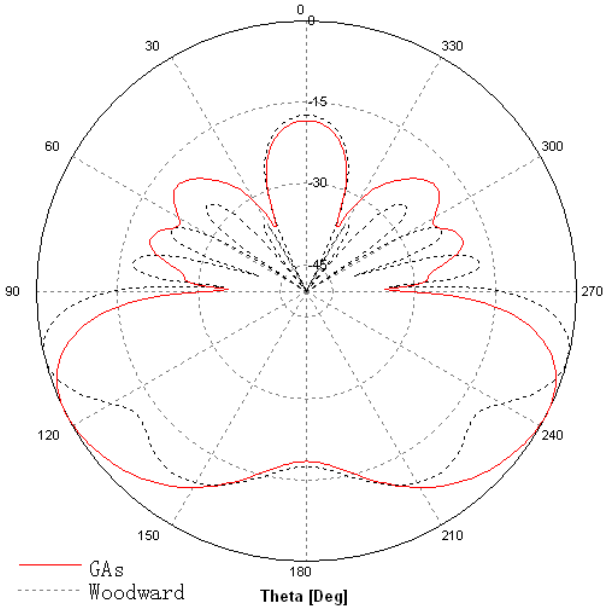


Figure 3. Shaped beam of base-station antenna generated by Genetic Algorithms and Woodward

Computer and Information Science

www.ccsenet.org/cis

Vol. 3, No. 1

February 2010

Call Admission Control for Next Generation Wireless Networks Using Higher Order Markov Model

Ramesh Babu H.S

Department of Information Science and Engineering, Acharya Institute of Technology

E-mail: rameshbabu@acharya.ac.in

Gowrishankar

Department of Computer Science and Engineering, B.M.S. College of Engineering

E-mail: gowrishankar.cse@bmsce.ac.in

Satyanarayana P.S

Department of Electronics and Communication Engineering, B.M.S. College of Engineering,

Bangalore, INDIA

E-mail: pssvittala.ece@bmsce.ac.in

Abstract

The Next generation wireless networks (NGWN) will be heterogeneous which will have different radio access technologies (RATs) operating together. The Radio Resource Management (RRM) is one of the key challenges in NGWN. The Call admission control (CAC) mechanism is one of the Radio Resource Management technique plays instrumental role in ensuring the desired QoS to the users working on different applications which are having the diversified nature of QoS requirements to be fulfilled by the wireless networks. One of the key challenges to be addressed in this prevailing scenario is the distribution of the available channel capacity amongst the multiple traffic with different bandwidth requirements so as to guarantee the QoS requirements of the traffic .The call blocking probability is one such QoS parameter for the wireless network and for better QoS it is desirable to reduce the call blocking probability. In this customary scenario it is highly advantageous to bring about an analytic Performance model. In this paper we propose a call admission control framework based on higher order Markov chains to effectively handle the call blocking probability in NGWN and to provide optimal QoS for the mobile users. In the proposed algorithm we have considered three classes of traffic having different QoS requirements. The results obtained from the Performance model are encouraging and optimistic and indicates the need of an intelligent decision making system for CAC.

Keywords: Radio Resource Management, Call admission control, Call blocking probability, QoS

1. Introduction

The recent advances in the wireless networks and mobile devices are inclined towards emerging of ubiquitous computing where the users and applications running in the mobile terminal (MT) can enjoy seamless roaming. It is well known that the basic problem in the wireless networks is the scarce of the radio resources. The efficient radio resource management is very essential. The admission control is one of the radio resource management technique which plays dominant role in effectively managing the resources. The admission control in the wireless networks will reduce the call blocking probability in the wireless networks by optimizing the utilization of the available radio resources. The mobile communication environment is featured by moving terminals with different QoS requirements and in this current scenario the need of guaranteed QoS is more sought for.

The future users of mobile communications look for always best connected (ABC) networks at anywhere and anytime among the complementary access technologies like wireless local area networks (WLAN), and worldwide inter operability for microwave access (Wi-MAX), Global Systems for Mobile Communications(GSM), General Packet Radio Service(GPRS), Universal mobile telecommunication systems(UMTS) etc.The mobile communication networks are evolving into adaptable Internet protocol based networks that can handle multimedia applications. When the multimedia data is supported by wireless networks, the networks should meet the quality of service requirements. One of the key challenges to be addressed in this prevailing scenario is the distribution of the available channel capacity among the multiple traffic that are with different bandwidth requirements.

The existing admission control strategies can handle the resource management in homogeneous wireless networks but are unable to effectively handle the issues in heterogeneous wireless environment. The mobility of the terminals in the mobile communication environment makes the resource allocation a challenging task when the resources are always in scarce. The efficient call admission control policies should be in place which can take care of this contradicting environment to optimize the resource utilization.

The design of call admission control algorithm must take into consideration packet level QoS parameters like minimum delay, jitter as well as session level QoS parameters like call blocking probability (CBP) and call dropping probability (CDP). The CBP is the probability of denial of accepting the new call and CDP the likelihood of dropping the call by a new access network due to decline of the network resources to an unacceptable level in other words the networks is exhausted with the available resources at which it drops the handover calls. In mobile networks the admission control traffic management mechanism is needed to keep the call blocking probability at a minimal level and another RRM strategy vertical handovers plays crucial role in reducing the and call dropping probability in an heterogeneous wireless networks.

In further sections of the paper is organized as follows. The section II discusses on the motivation and related work. Section III focuses on the proposed system model for the call admission control based on higher order Markov chains. The section IV represents the simulation results and conclusion and future work is indicated in section V.

2. Related Work

At present, dissimilar wireless access networks including 2.5G , 3G , Bluetooth , WLAN and Wi-Max coexist in the mobile computing environment, where each of these Radio access technologies offer complementary characteristics and features in terms of its coverage area, data rate, resource utilization , power consumption etc.. With all these there are constant improvements in the existing technologies offering better performance at lesser cost. This is beneficial in both the end users and service provider's perspective. The idea of benefiting from integrating the different technologies has lead to the concept of beyond International mobile telephony 2000(IMT-2000) wireless networks known as the next generation wireless networks (NGWN). In this heterogeneous environment, the end user is expected to be able to connect to any of the different available access networks. The end user will also be able to roam seamlessly within these access networks through vertical handover mechanisms. The global roaming is supplemented by the existence of IP networks as the backbone which makes the mobile computing environment to grow leaps and bounds and can effectively address the issue with regard to converge limitations is concerned. In this multifaceted wireless radio environment the radio resource management plays major role. The effective utilization of the limited available resources is the challenge. The admission control is one such challenge a network service provider face to achieve better system utilization face in handling this complex scenario to provide the best QoS to the users of the network.

Call admission control schemes can be divided into two Categories, local and collaborative schemes [1]. Local schemes use local information alone (e.g. local cell load) when taking the admission decision. Examples of these schemes are(T. Zhang, E.v.d. Berg, J. Chennikara, P. Agrawal, J.-C. Chen, T. Kodama, 2001)(C.-T. Chou, K.G. Shin, 2002)(C.W. Ahn, R.S. Ramakrishna, 2004). Collaborative schemes involve more than one cell in the admission process. The cells exchange information about the ongoing sessions and about their capabilities to support these sessions (T. Zhang, E.v.d. Berg, J. Chennikara, P. Agrawal, J.-C. Chen, T. Kodama, 2001).The fundamental idea behind all collaborative admission control schemes is to consider not only local information but also information from other cells in the network. The local cell, where the new call has been requested, communicates with a set of cells that will participate in the admission process. This set of cells is usually referred to as a cluster. In general, the schemes differ from each other according to how the cluster is constructed, the type of information exchanged and how this information is used. In (M. Naghshineh, M. Schwartz, 1996) for example, the cluster is defined as the set of direct neighbors. The main idea is to make the decision of admission control in a decentralized manner.

There are good amount of work reported for homogenous wireless networks and single service wireless networks and few works in the heterogeneous wireless networks. The Call admission control in Heterogeneous wireless networks is a real challenge. The varied QoS requirements of multimedia applications and the coexistence of different RATs, facade major challenges in designing CAC algorithms for next generation heterogeneous wireless networks. The challenges are heterogeneous networking, multiple service classes, flexibility in bandwidth allocation and cross layer issues based design(Ramesh Babu H.S., Gowrishankar, Satyanarayana P.S. 2009).

2.1 Heterogeneous networking

4G networks will have different types of RATs different from each other by air interface technology, cell size, services, price, access method, coverage, so CAC schemes must be able to handle new type of handoff called vertical handoff.

2.2 Multiple service classes

The B3G networks should be able to accommodate the applications and user with different QoS requirements, so the CAC algorithms should be designed to handle different classes of service to meet the QoS needs of all types of

applications.

2.3 Flexible in bandwidth allocation

The diversity is in multimedia applications and mobile users QoS requirements in NGWN. The resource utilization and QoS performance can be improved by adaptive bandwidth allocation. This clearly indicates that the CAC should be designed taking into consideration the flexible bandwidth allocation, where, more resources can be allocated when the there is less traffic and the allocated bandwidth can be revoked when there is congestion.

2.4 Cross layer issues based design

The traditional CAC schemes were based on the call level QoS only and few of them have considered the physical layer QoS like SIR as QoS criteria. Unlike the tradition voice oriented circuit switched network, the next generation network predicted to be pure packet based network and the QoS needs to be addressed both at call level as well as at packet level. This mandates that the new call has to be admitted only if both call level QoS metrics like call blocking and dropping probabilities and the packet level QoS measures like packet transmission delay and packet dropping probability are maintained at some desired levels.

The other important solution for the decision making of call admission control is by multi criteria decision making (MCDM). This is an optimization technique used to analyze the contradicting decision making parameters. The MCDM based decision making systems are generally used in the fields of reliability, financial analysis, social and political related analysis and environmental impact analysis etc. The NGWN has different RATs coexisting which are with different capabilities and they should cater the varied QoS requirements of multimedia applications admission control with single criteria may be too trivial, in this prevailing scenario the admission control decision should be based on Multi criteria such that the optimization user satisfaction and selection of optimal RAT is achieved. There are several algorithms proposed on handling the admission control decision making using MCDM in heterogeneous wireless networks.

There are Different admission control algorithms based on multiple criteria decision making. They are categorized as *Utility- function based CAC* and *computation Intelligence CAC*. In the *Utility -function based CAC* the incoming calls are admitted based on some utility or cost function based on multiple criteria. These algorithms are very optimal algorithms and in most of the case are complex in nature and pose high computational overhead.

The *computation-Intelligence-based CAC* use evolutionary approaches like Genetic Algorithm (GA), fuzzy logic and Artificial Neural Networks(ANN) (R.T.Marler and J.S. Arora, 2004).Majority of the computational-intelligence-based CAC algorithms incorporate fuzzy logic(P.M.L. Chan, R.E. Sheriff, Y.F. Hu, P. Conforto, C. Tocci, 2001), fuzzy neural(R. Agusti, O. Sallent, J. Pérez-Romero, L. Giupponi, 2004) and fuzzy MCDM(W. Zhang, 2004)(A.L. Wilson, A. Lenaghan, R. Malyan, 2005) methods. There are very few works reported on the usage of Artificial Neural Networks in CAC.

This paper is pitched upon the call admission control mechanism in heterogeneous networks scenario handling different types of calls in the system.

3. System Model

In this paper we propose a novel analytical model for admission control for the call blocking probability there by increasing the resource utilization. This would achieve the Objective of guaranteeing the user QoS requirements. The proposed model is able to handle three types of applications which are complementary in nature with respect to their QoS requirements are considered. The applications considered for the study involves conversation traffic, interactive traffic and background traffic. The representative applications could be voice calls, Web browsing and file transfer applications respectively. We have considered a heterogeneous network which comprises a set of RATs *Rn* with co-located cells in which radio resources are jointly managed. Cellular networks such as Wireless LAN and Wi-Max can have the same and fully overlapped coverage, which is technically feasible, and may also save installation cost. H is the set of heterogeneous wireless networks coexisting is given as H= {RAT 1, RAT 2, RAT k} and where K is the total number of RATs in the heterogeneous wireless network. The heterogeneous wireless network supports n-classes of calls, and each RAT in set H is optimized to support certain classes of calls.

The Analytical model for Call admission control mechanism in heterogeneous wireless networks is modeled using higher order Markov chain as shown in figure2.The study considers that, whenever a new user enters the network will originate the network request at the rate λ_i and is assumed to follow a *Poisson process*. The service time of the different class of traffic and types of calls is μ_i.The mean service time of all types of users were assumed to follow negative exponential distribution with the mean rate 1/μ. Since Voice traffic is Erlang distributed and the condition that is considered for simulation is Negative Exponential distribution. The total number of virtual channels in the system are N. When the numbers of available channels are below the specified threshold the system will block/drop the calls. The threshold limit is determined by three positive integers A_1, A_2 and A_3. When the number of available channels falls

below the threshold A_3 the proposed system will accept only the voice calls and web browsing. When the number of available channels falls below the threshold A_2 the proposed system will accept only the voice calls and when the available number of channels falls below the threshold A_1 the proposed system will not accept any calls as it reaches the stage where there will be no channels available to allocate to the incoming calls and leads to system blocking. The $P(0)$ is probability that there are no allocated channels in the designated system.

For the system model shown in figure1 the system state can be defined for the lower and upper boundaries as in (1)-(3) and (4)-(6) respectively.

The equations (1) - (3) are lower boundary equations represents the system states P_0, P_1, P_2

$$\lambda_1 P_0 + \lambda_2 P_0 + \lambda_3 P_0 - \mu_2 P_2 - \mu_1 P_1 - \mu_3 P_3 = 0 \quad (1)$$

$$\lambda_1 P_1 + \lambda_2 P_1 + \lambda_3 P_1 - \mu_1 P_2 - \mu_2 P_3 - \mu_3 P_3 = 0 \quad (2)$$

$$\lambda_1 P_2 + \lambda_2 P_2 + \lambda_3 P_2 + \mu_1 P_2 + \mu_2 P_2 - \mu_1 P_3 - \mu_2 P_4 - \mu_3 P_5 = 0 \quad (3)$$

The equations (4) - (6) are upper boundary equations for the system states P_n, P_{n-1}, P_{n-2} and are expressed as

$$P_{n-3}(\lambda_1 + \lambda_2 + \lambda_3 + \mu_1 + \mu_2 + \mu_3) - \lambda_1 P_{n-4} - \lambda_2 P_{n-5} - \lambda_3 P_{n-6} - \mu_1 P_{n-2} - \mu_2 P_{n-1} - \mu_3 P_n = 0 \quad (4)$$

$$P_{n-2}(\lambda_1 + \lambda_2 + \mu_1 + \mu_2 + \mu_3) - \lambda_1 P_{n-3} - \lambda_2 P_{n-4} - \lambda_3 P_{n-5} - \mu_1 P_{n-1} - \mu_2 P_n = 0 \quad (5)$$

$$P_{n-1}(\lambda_1 + \mu_1 + \mu_2 + \mu_3) - \lambda_1 P_{n-2} - \lambda_2 P_{n-3} - \lambda_3 P_{n-4} - \mu_1 P_n = 0 \quad (6)$$

The repeated states are those which are in-between these upper boundary and lower boundary states based on figure1. The repeated states of the system are represented in a generic form as.

$$P_4 (\lambda_1 + \lambda_2 + \lambda_3 + \mu_1 + \mu_2 + \mu_3) - \lambda_1 P_3 - \lambda_2 P_2 - \lambda_3 P_5 - \mu_1 P_6 - \mu_2 P_6 - \mu_3 P_7 = 0 \quad (7)$$

The equation that can be presumed as the general equation for call blocking probability for traffic type 1 is

$$P_n = \frac{\lambda_1 P_{n-1} + \lambda_2 P_{n-2} + \lambda_3 P_{n-3}}{(\mu_1 + \mu_2 + \mu_3)} \quad (8)$$

Assuming that the arrival time of all the types of traffic are equal i.e. $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ and the service time for the types of traffic are equal i.e. $\mu_1 = \mu_2 = \mu_3 = \mu$, the call blocking probability for type1 traffic could be expressed as

$$P_n = \frac{a}{3}(P_{n-1} + P_{n-2} + P_{n-3}) \quad (9)$$

Where $a = \lambda / \mu$ is called as *Utilization rate* which should be generally less than one for the system stability.

Similarly, the call blocking probability for type2 traffic P_{n-1} is

$$P_{n-1} = \frac{a}{3}(P_{n-2} + P_{n-3} + P_{n-4}) \quad (10)$$

And the call blocking probability for type3 traffic P_{n-2} is represented as

$$P_{n-2} = \frac{a}{3}(P_{n-3} + P_{n-4} + P_{n-5}) \quad (11)$$

The call blocking probability for the overall system traffic P_{nb} can be expressed as

$$P_{nb} = \frac{a}{3}(P_n + P_{n-1} + P_{n-2}) \quad (12)$$

4. Simulation Results and Discussion

In this section, we present the numerical results and compare the call blocking probabilities of the different types of traffic and MATLAB is used for the simulation. The proposed performance model for call admission control mechanism is analyzed for the call blocking probability by making the variation in the number of channels. .The experiment setup is conducted by varying the traffic intensity of Type1 traffic and the blocking probability of type1, blocking probability of the type2, and type3 traffic of the system is plotted. The Figure 2 shows call blocking probability for all three types of traffic. The horizontal axis shows the number of users with type 1 traffic while the vertical axis shows the call blocking probability of all types of traffic.

The parameters of analytic performance model are also called as Performance model parameters. The Performance model parameters are number of virtual channels (N), user arrival rate (λ), arrival rate of type1 call (λ1), arrival rate of type2 call (λ2.) arrival rate of type3 call (λ3) and the service time of the user (μ) , service time of the call of type1(μ_1) , Service time of the call of type2 (μ_2), Service time of the call of type3(μ_3).The simulation results shows that the call blocking probability of the different types of traffic will increase with the increase in the intensity of type1 traffic. The simulation results with increase the intensity of type2 traffic and simulation results with increase the intensity of type3 also showed the similar kind of results. The simulation results indicate that at particular state the call blocking probability of all three types of traffic will be reduced.

The important observations from the performance model are: Increase in the traffic of Type1 call and reduce in the type2 and type3 call will reduce the Call blocking probability of type1 call. The results indicate reduction in call blocking probability of type2 calls When reduce the type1 and type2 traffic in the system. Similarly when we reduce the type1 and type2 traffic, the call blocking probability of type3 traffic is minimal. In a next generation wireless system, it is desirable to uphold guaranteed QoS to the all types of users and optimum Utilization of system resource needed. In order to provide superior QoS to the user/application it is required to maintain the call blocking probability minimal. The bandwidth reservation for each type of traffic may reduce the call blocking probability. But this is possible when we have non varying traffic for all the types of traffic and there is always a possibility of wastage of the bandwidth if one of the three traffic intensity is lower than the bandwidth of the allocated channels. It is predicted that the next generation network is heterogeneous and obviously we have the time varying traffic as iterated in the in the traffic model in section 4.

In this prevailing condition it is very difficult and tricky situation to satisfy the needs of different types of traffic. Hence it is enviable to use a multi constraint optimization technique in achieving better tradeoff between set of system performance metrics. The performance parameters such as BER, ABW and network traffic are the input to the optimization technique for achieving better tradeoff between set of performance metrics. Some of the well known techniques in solving multi constraint optimization problem are Game theory, Markov Decision Process(MDP), Genetic Algorithm(GA), Goal programming and Multi attribute Decision Making (MADM) technique. Among these set of techniques MADM is a most widely used method (K.P.Yoon and C.Hwang, 1995).

5. Conclusion and Future Work

In this paper, we have proposed a performance model for call admission control mechanism in the heterogeneous RATs and analyzing the call blocking probability keeping the variation in the number of channels. In order to measure the call blocking probability of the analytical model the simulation study was made and following observations were made. In the experiment setup all the types of traffic was varied and Firstly, increase in the number of type1 users will increase the call blocking probability of type2 and type3 calls and vice versa. Second, Increase in the traffic intensity of one type of traffic will increase the system blocking probability.

The concept of minimizing the call blocking probability is an optimization technique to provide fair QoS to the set of users in the wireless network and there is also a need of intelligent call admission control strategy in the admission control mechanism to make the decision of accepting are rejecting a call keeping the blocking probability minimal in a heterogeneous RATs based network working under dynamic network condition. The future work of this research is pitched upon using Intelligence to the call admission control decision process .The future work includes application of Fuzzy neural technique for the decision making .It is evident from the above discussion that that the application of intelligent techniques with multiple criteria for decision making is worthwhile.

References

A.L. Wilson, A. Lenaghan, R. Malyan. (2005). Optimizing wireless network selection to maintain QoS in heterogeneous wireless environments, in: Proceedings of World Communication Forum, Denmark, September, 2005.

C.-T. Chou, K.G. Shin. (2002) Analysis of combined adaptive bandwidth allocation and admission control in wireless

networks, in: IEEE Conference on Computer Communications (INFOCOM), 2002 pp. 676–684.

C.W. Ahn, R.S. Ramakrishna. (2004). QoS provisioning dynamic connection admission control for multimedia wireless networks using a Hopfield neural network, IEEE Transactions on Vehicular Technology 53 (1)(2004) 106–117.

K.P.Yoon and C.Hwang, (1995). "Multiple attribute decision making", Sage,1995.

M. Naghshineh, M. Schwartz. (1996). Distributed call admission control in mobile/wireless networks, IEEE Journal on Selected Areas in Communications (JSAC) 14 (4) (1996) 711–717.

P.M.L. Chan, R.E. Sheriff, Y.F. Hu, P. Conforto, C. Tocci. (2001). Mobility management incorporating fuzzy logic for a heterogeneous IP environment, IEEE Communications Magazine 39 (12) (2001) 42–51.

R. Agusti, O. Sallent, J. Pe´rez-Romero, L. Giupponi. (2004). A fuzzy-neural based approach for joint radio resource management in a beyond 3G framework, in: 1st International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (Qshine'04), Dallas, USA, October, 2004.

R.T.Marler and J.S. Arora. (2004). "Survey of multi-objective optimization methods for engineering", Structural and multidisciplinary optimization, Vol.26.No.6,pp 369-395,2004.

Ramesh Babu H.S., Gowrishankar, Satyanarayana P.S. (2009). Call admission control approaches in beyond 3G networks using multi criteria decision making , 2009 First International Conference on Computational Intelligence, Communication Systems(CICSYN2009).

T. Zhang, E.v.d. Berg, J. Chennikara, P. Agrawal, J.-C. Chen, T. Kodama. (2001). Local predictive resource reservation for handoff in multimedia wireless IP networks, IEEE Journal on Selected Areas in Communications (JSAC) 19 (10) (2001) 1931–1941.

T. Zhang, E.v.d. Berg, J. Chennikara, P. Agrawal, J.-C. Chen, T. Kodama. (2001). Local predictive resource reservation for handoff in multimedia wireless IP networks, IEEE Journal on Selected Areas in Communications (JSAC) 19 (10) (2001) 1931–1941.

W. Zhang. (2004). Handover decision using fuzzy MADM in heterogeneous networks, in: Proceedings of IEEE WCNC'04, Atlanta, GA, March, 2004.

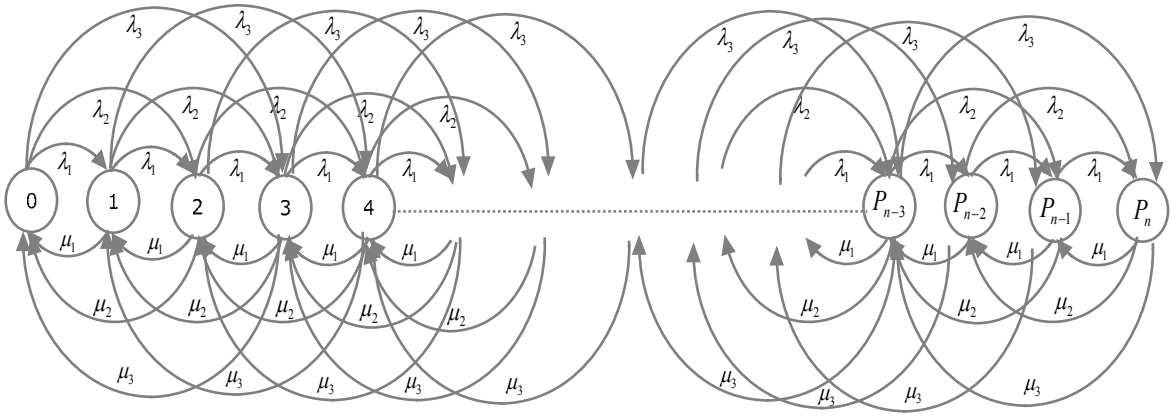


Figure 1. System Model for Call Admission Control

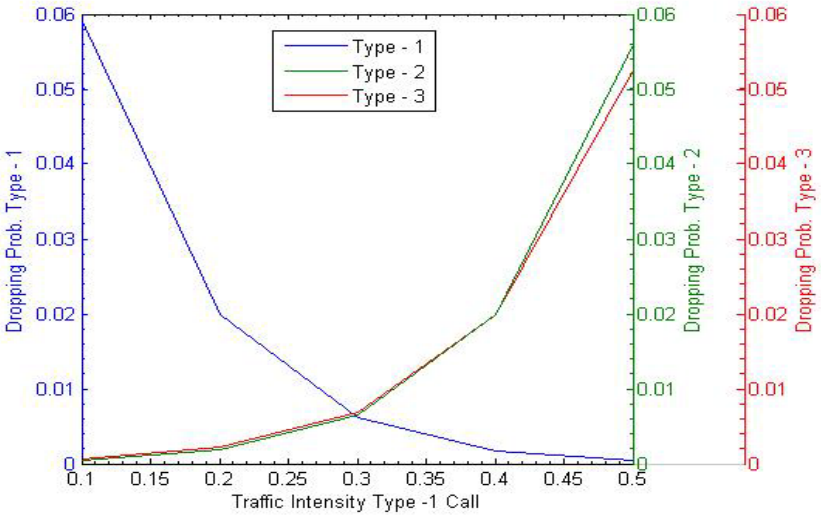


Figure 2. Call blocking probablity for varying traffic

Call for Manuscripts

Computer and Information Science is a peer-reviewed journal, published by Canadian Center of Science and Education. The journal publishes research papers in the fields of computer, information science, e-business, internet, electronic and relevant subjects. The journal is published in both printed and online versions. The online version is free access and download.

We are seeking submissions for forthcoming issues. The paper should be written in professional English. The length of 3000-8000 words is preferred. All manuscripts should be prepared in MS-Word format, and submitted online, or sent to: cis@ccsenet.org

Paper Selection and Publication Process

a). Upon receipt of paper submission, the Editor sends an E-mail of confirmation to the corresponding author within 1-3 working days. If you fail to receive this confirmation, your submission/e-mail may be missed. Please contact the Editor in time for that.

b). Peer review. We use single-blind system for peer-review; the reviewers' identities remain anonymous to authors. The paper will be peer-reviewed by three experts; one is an editorial staff and the other two are external reviewers. The review process may take 2-3 weeks.

c). Notification of the result of review by E-mail.

d). The authors revise paper and pay publication fee.

e). After publication, the corresponding author will receive two copies of printed journals, free of charge.

f). E-journal in PDF is available on the journal’s webpage, free of charge for download.

Requirements and Copyrights

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other languages, without the written consent of the Publisher. The Editors reserve the right to edit or otherwise alter all contributions, but authors will receive proofs for approval before publication.

Copyrights for articles published in CCSE journals are retained by the authors, with first publication rights granted to the journal. The journal/publisher is not responsible for subsequent uses of the work. It is the author's responsibility to bring an infringement action if so desired by the author.

More Information

E-mail: cis@ccsenet.org

Website: www.ccsenet.org/cis

Paper Submission Guide: www.ccsenet.org/submission

Recruitment for Reviewers: www.ccsenet.org/reviewer.html

A journal archived in Library and Archives Canada
A journal indexed in CANADIANA (The National Bibliography)
A journal indexed in AMICUS
A journal included in DOAJ (Directory of Open-Access Journal)
A journal included in Google Scholar
A journal included in LOCKSS
A journal included in PKP Open Archives Harvester
A journal listed in Ulrich's

Computer and Information Science

Quarterly

Publisher Canadian Center of Science and Education
Address 4915 Bathurst St. Unit 209-309, Toronto, ON. M2R 1X9
Telephone 1-416-208-4027
Fax 1-416-208-4028
E-mail CIS@ccsenet.org
Website www.ccsenet.org
Printer Paintsky Printing Inc.
Price CAD.$ 20.00

ISSN 1913-8989