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

COMPUTER AND INFORMATION SCIENCE

Vol. 2, No. 3
August 2009

Canadian Center of Science and Education

Editorial Board

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Contents

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Contents

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

A Hybrid Resource Allocation Strategy with Queuing in Wireless Mobile Communication Networks

K. Venkatachalam (Corresponding author)

Velalar College of Engineering and Technology

Thindal post, Erode - 638 012, Tamilnadu, India

Tel: 91-997-665-5046 E-mail: venki_kv @yahoo.com

Dr. P. Balasubramanie

Kongu Engineering College, Perundurai, Erode- 638 052, Tamilnadu, India

E-mail: pbalu_20032001@yahoo.co.in

Abstract

The main objective of third and future wireless mobile communication networks is to provide services efficiently to the mobile users in all environments. In wireless mobile Communication the channel allocation and quality of service are the major factors and important issues to decide the system performance. Here in our work a Hybrid Channel Allocation(HCA)strategy for channel allocation and queuing technique applied to Hybrid Channel Allocation strategy for Quality of Service(QoS) provisioning are implemented. The proposed HCA strategy considers new calls in Fixed Channel Allocation (FCA) method and handoff calls in Dynamic Channel Allocation (DCA) method to reduce the call blocking and call dropping probabilities. The application of queuing technique applied to HCA strategy increases the efficiency of the cellular system performance especially in micro and pico cellular environments and effectively utilize the available allocated radio spectrum. The performance shows the decrease in call blocking and dropping and an increased number of users in the available channels.

Keywords: Fixed Channel Allocation (FCA), Dynamic Channel Allocation (DCA), Hybrid Channel Allocation (HCA), Quality of Service (QoS), Call Dropping Probability (CDP), Call Blocking Probability (CBP)

1. Introduction

The radio spectrum allocated to wireless cellular mobile communication system is fixed one. Efficient use of available spectrum will produce an increased number of users. The fundamental operational principle of wireless cellular mobile communication system (WCMCS) is the reuse of frequencies at different places within the areas of service. Reusing the same frequency will not interfere, as it will be attenuated enough. In doing this, the carrier to interference ratio (CIR) and the radio spectrum reuse efficiency are improved. The channel allocation strategy is the technique used to make the most efficient and intelligent way of utilizing the available radio spectrum by the way in which channels are allocated to mobile multimedia calls.

The third and future generation wireless communication networks will support heterogeneous traffic, consisting of voice, video (JPEG & MPEG) and data as multimedia. The Quality of service (QoS) is the major issue for these applications in any kind of communication networking environment. The wireless QoS is the complex problem due to the time varying characteristics of the channel and user mobility. In this work, an efficient channel allocation strategy with QoS provisioning framework is applied to third and future generation wireless networks. Here a Hybrid Channel Allocation (HCA) strategy [HCA is the combination of Fixed Channel Allocation (FCA) and Dynamic Channel Allocation (DCA) strategies] for multimedia mobile calls and queuing technique are considered. Because queuing technique is the effective and efficient way to reduce the new call blocking and handoff call dropping probabilities. Hence the implementation of HCA with queuing will effectively increases the channel utilization and thus providing QoS.

2. Overview on channel allocation strategies

The bandwidth or channel allocation strategies are developed to utilize the available radio spectrum effectively and efficiently. This will impact the performance of the system particularly as to how calls are managed when a mobile user making a new call or the user handoff from one cell to another cell.

2.1 Fixed Channel Allocation (FCA) strategy

In this method, each cell is allocated to predetermined set of channels, if all the channels in the cell are occupied, then the call is blocked, and the mobile subscriber does not receive the service. The frequency reuse distance (D) and the co-channel interference (Q) are the factors to be considered for allocating set of channels to the particular cell. This is a very simple method for allocating channels. Here there is no delay time for getting the channel or service, if the channels are free but at the same time once all channels are occupied, no one get the service from the network until any one of the user will disconnect their call.

Channel borrowing approach: In FCA, the channel borrowing approach is also used. In this the cell is allowed to borrow the channels from a neighboring cell, if all of its own channels are already occupied. The Mobile Switching Centre (MSC) supervises the borrowing procedure and ensures that, the borrowing of the channels does not disrupt or interfere with any of the call in progress in the donor cell.

2.2 Dynamic Channel Allocation (DCA) strategy

In this method, the channels are not allocated to different cells permanently. Instead, each time, a call request is made, the serving Base Transceiver Station (BTS) request a channel from the Mobile Switching Center (MSC), if centralized control is used or otherwise, sometimes for the decentralized control, the BTS request a channel from the Base Station Controller (BSC) itself. The switch then allocate a channel to the requested cell following an algorithm that takes into account, the likelihood of future blocking within the cells, the frequency of use of the candidate channel, the reuse distance of the channel, any other cost functions. Accordingly the MSC or BSC only allocates a given frequency or channel, if that frequency or channel is not presently in use in the cell or any other cell, which falls within the minimum restricted distance. The minimum restricted distance is given by $D = \sqrt{3K}\,R$ and $D/R = Q = \sqrt{3K}$ is the co-channel interference reduction factor. Where R is the radius of the cell coverage and K is the frequency reuse pattern i.e. cluster size. Dynamic channel allocation strategies will reduce the likelihood of blocking, which increases the trunking capacity of the system, since all the available channels in the system are accessible to all of the cells. DCA strategies require the MSC or BSC to collect the real-time data or information on channel occupancy, traffic distribution, and radio signal strength indicators (RSSI) of all channels on a continuous basis. This will increases the storage and computational load on the system, and increases the delay time for allocating channel to a particular mobile user but at the same time, it gives the advantage of increased channel utilization and decreased probability of blocked calls. Normally in MSC decentralized control method is used instead of centralized controller since decentralized control has

Less delay for allocating channels compare to centralized control method

Less computations load on the MSC which interns requires the less storage (Memory space) as compare to centralized control method.

Because of decentralized control, the fault identification and trouble shooting are easy.

2.3 Hybrid Channel Allocation (HCA) strategy

In HCA strategy, the total set of channels is divided into two subsets. The first subset of channels is assigned to the cells of the system according to the FCA strategy. The second set is kept in a central pool and assigned dynamically to the cells on demand to increase the flexibility. Therefore, there are two types of channel allocation strategy at the same time i.e., the combination of FCA and DCA strategy. Here for allocating channels, the fixed to dynamic ratio is very important one. This fixed to dynamic ratio depends upon the statistical nature of the cell i.e., depends on number of new calls and hand off calls initiated in peak hours and also in normal hours.

3. Steady state performance of cellular wireless networks

It is essential to know the steady state performance of the cellular wireless networks to evaluate the performance of our proposed HCA strategy with queuing. In the cellular wireless networks geographic region is divided into hexagonal cells with six neighbors each. It is the commonly used network model. A user enters the network in any cell, provided there is enough bandwidth available, otherwise the call is blocked. The users keep on moving from one cell to another i.e., neighboring cell provided the Mobile Unit (MU) finds the required bandwidth in the next cell. If it is not, then the call is dropped. The Steady State Utilization (SSU) is a parameter that gives an insight into the load, a system can support without loss. Take a fully loaded system i.e., maximum number of users in each cell and the users are move in and around the system frequently. Assume users are permanent and they do not leave unless they get dropped. The utilization at this condition is the steady state utilization of the system. The Steady State Arrival Rate (SSAR) is the arrival rate that will keep the system functioning at its steady state utilization. For a system with finite call durations, there is a corresponding arrival rate to keep the system functioning at the steady state utilization. The arrival rate at this condition is the steady state arrival rate and it is given by

$$SSAR = \left(\frac{Util_{ssu} \text{ x Max. users}}{T} \right) \quad (1)$$

Where

$Util_{ssu}$ is steady state utilization and 'T' is the average call duration time

3.1 SSAR analysis of cellular wireless networks

Consider a fully loaded system i.e. maximum number of users in each cell. Assume that the call duration of a user on average is 'T' and the arrival rate is 'λ'. The number of users that arrived in the system during the time interval 'T' is given by

$$N_{arrivals} = \lambda T \quad (2)$$

Some of the users that arrive are lost due to blocks and drops. Now the number of users that exist in the system is expressed as

$$N_{users} = (1 - loss)\,\lambda T \quad (3)$$

By dividing both side of the above equation by maximum number of users, then it becomes

$$\frac{N_{users}}{Max.users} = (1 - loss)\frac{\lambda T}{Max.users} \quad (4)$$

The left side of the above equation gives system utilization. Therefore

$$Util = (1 - loss)\frac{\lambda T}{Max.users} \quad (5)$$

From this

$$\lambda = \frac{Util}{(1 - loss)}\frac{Max.users}{T} \quad (6)$$

The above expression relates the arrival rate with utilization. The simulation for different call arrival rate with 10% loss is shown in figure 1. By rearranging the above expression we will get

$$T = \frac{Util}{(1 - loss)}\frac{Max.Users}{\lambda} \quad (7)$$

This expression relates the average time duration of the call (T) with utilization. The simulation for different call duration time with 10% loss is shown in figure 2. If the system with loss is equal to zero and utilization is equal to steady state utilization, then the arrival rate at this point the steady state arrival rate.

4. Related work

Variety of work has been done in the area of channel allocation strategies considering new calls and handoff calls. Minimizing the handoff occurrence is one possible solution. Virtual tree connection method (A.S. Acampora, & M. Naghshineh, 1994) is the example solution. The model in (C. Oliverira, J.B. Kim, & T. Suda, 1998) allocate bandwidth to a call in the cell where the call request originates and at the same time reserves the same amount of bandwidth in all neighboring cells for the call. A hierarchical cell model is proposed in (L.O.Guerrero , & A.H.Aghvami, 1998) in which macro cells and micro cells are taken, and it allocates the bandwidth to appropriate cell according to the velocity of the mobile unit (MU). The scheme in (S.Bajaj, L.Breslau, D. Estrin, K.Fall, & S.kumar, 1999) compares the model by assuming utilization and loss separately. Here the solution is to avoid the bandwidth shortage of the cell by reserving the resource, in which direction the mobile user moves. The model in (P. Ramanathan, K.M. Sivalingam, P. Agrawal, & S. Kishore, 1999) considers the acceptance of new call is, according to an approximately calculated probability of a mobile unit induces handoff. By reserving some amount of bandwidth for new calls and handoff calls in advance, so that, the blocking and dropping probabilities can be reduced (Tao Zhang, & Prathima Agarwal, 2001).Here the resource reservation is considered for reducing the dropping probability. To utilize the available spectrum effectively the 'Spectrum holes' i.e., prediction of resource availability in advance is discussed in (V.Syrotuik et..al., 2004).Resource prediction is also a method to reduce handoff call droppings.

In order to meet "anywhere and anytime" concept, the future wireless network is converge into heterogeneous cellular network to support multimedia calls. For this, the good and bad threshold state method is examined in (L.Xu.X.shen & J.W.Mark, 2005) to reduce the dropping probabilities. But for real time traffic (example voice or video) with delay constraint, if a mobile station is in bad channel state for relatively long period, the multimedia call quality will be very much decreased and probably the call will be disconnected from the network. Since it only considers the threshold condition of a multimedia call. An analytical model of cellular mobile communication networks with instantaneous movement is investigated in (Wei Li, & Xiuli Chao, 2004). This is very much useful to cost analysis for updating location and paging in wireless cellular mobile networks. In the bandwidth allocation, various requests of an MU for communication quality should be satisfied. Our work presents a new class of channel allocation method that overcomes

some of the critical limitations of existing methods by considering hybrid channel allocation with queuing. The proposed method is simple and especially suitable for micro and pico cellular multimedia wireless networks.

5. Objective and organization of the work

The main objective of our work is to present an efficient channel allocation strategy for multimedia calls used in future wireless cellular mobile communication networks formed by micro and Pico cellular structures.

Section 2 deals with introduction about different channel allocation strategies and section 3 discusses the steady state performance of the wireless cellular system with different call arrival rate and different call duration timings. Section 4 expresses the related work in this area in a short manner. Section 6 presents the proposed hybrid channel allocation (HCA) strategy with queuing technique applied to the same. Section 7 briefs the performance results with conclusion.

6. Proposed Hybrid Channel Allocation strategy

Hybrid Channel Allocation (HCA) is the combination of Fixed Channel Allocation (FCA) and Dynamic Channel Allocation (DCA). Here a portion of total frequency channels will use FCA and the rest will use the DCA. In HCA the channel allocation is done as follows, when a new or handoff call arrives at the cell BTS, the first attempt to serve, it is by a nominal or fixed channel, if there is no fixed free channel, then a channel from a dynamic set is assigned to the calls. If this also fails, then the call is blocked. The ratio of fixed to dynamic channel, is a significant parameter that determines and defines the performance of the system in much, the same manner that the ratio of nominal to borrowable channels defines the performance of a strategy with channel borrowing. In general, the fixed to dynamic ratio of the channel is a function of the traffic load and would vary over time according to the offered load distribution estimations. In HCA strategy, different methods are followed for channel allocation to multimedia mobile calls.

The normal method is, when a new call or a handoff call arrives to the Base Transceiver Station, the channels are allocated first in FCA strategy, if there is no free channels available in FCA, then channels are taken from DCA pool and allocated in DCA strategy, The second is, our proposed strategy, in which whenever a new call arrival to the BTS, the channels are allocated in FCA, and for the handoff call arrivals the channels are allocated in DCA strategy. The simplified model of our proposed HCA strategy is shown in figure 3. The first subset of channels are assigned to a cell, in FCA strategy for new calls, and the second subset of channels are assigned to the calls as DCA strategy for handoff calls. Here we consider FCA strategy for new calls because as far as new calls are considered delaying the call to provide the service or connection is better than blocking the call. But at the same time, for handoff calls, providing continuous connection with minimum acceptable call quality is better than dropping a call in the middle of the service during handoff. Hence DCA strategy is proposed for handoff calls.

6.1 Proposed HCA strategy with queuing technique

The proposed hybrid channel allocation with queuing technique is shown in figure 4. The flow chart is self explanatory one. Whenever a new call arrives to BTS, the channels are allocated in fixed manner i.e. FCA method and for handoff call arrival the channels are allocated dynamically i.e. DCA method. Here the fixed to dynamic ratio is the important factor and it varies according to the traffic conditions and cell site nature. All the new and handoff calls are queued before channel allocation and then they are added. It gives the present traffic condition of a particular cell. Before allocating the calls of new or handoff type to particular channel allocation strategy i.e., DCA or FCA, queuing the calls are effective way to reduce the blocking and dropping probabilities, if the calls are arrived in bundles. But for sequential arrival of calls, queuing is not effective. Generally, in practical situation the calls are arrived in bundles during the busy or peak hours. Hence queuing is needed one at all times. If the prediction method is implemented in BSC for distributed control or MSC in centralized control then which predicts the future bandwidth requirement by using present traffic condition. This prediction will also considerably reduces the new call blocking and handoff call dropping probabilities.

6.2 Call admission procedure in our proposed HCA strategy

In our proposed HCA strategy, the call admission procedure is shown in figure 5. In this for new calls, if the system having capacity to accommodate the call, then the call is admitted otherwise the call is blocked. For handoff calls, the capacity is estimated and then according to the constraints in DCA, the call is admitted.

6.3 Channel allocation for new calls and handoff calls in HCA strategy

The FCA method for new calls and DCA method for hand off calls in HCA strategy are shown in figure 6 & 7 respectively. In the fixed channel allocation for new calls, if the base station in the cell having capacity, then the call is accepted otherwise the call is blocked. In the dynamic channel allocation for handoff calls, the frequency reuse distance (D) and co-channel interference reduction factor (Q) are considered and then the channel is allocated to a call. In this both handoff reservation and handoff release calls are considered. In case, the above constrains are not satisfied, then the call is dropped. If the call is handoff release then the channel is released and kept in MSC or BSC pool for allocating channels to other handoff reservation request.

6.4 Queuing technique applied to our proposed HCA strategy

In order to improve the performance and efficiency of our system, first we consider HCA strategy for channel allocation. Secondly to avoid the call blocking of new calls and dropping of handoff calls, queuing technique is applied to HCA strategy for still better performance of the cellular wireless networks. In this, the MSC will queue the new call requests and handoff call requests before allocating the channel instead of rejecting them, if the cell site is busy. The queuing technique is effective only when the new call requests and handoff call requests are arrived at the MSC in batches or bundles. If the call requests are arriving uniformly, then queuing technique is not needed. In practical situations, calls are arrived to MSC or BSC in bundles during the busy hours. Hence queuing is the essential process to avoid call blocking and dropping in the cellular wireless networks.

Consider λ_1 and λ_2 are the arrival rate of new calls and handoff calls per second respectively. Take M_1 and M_2 are the queue size of new calls in FCA and handoff calls in DCA respectively of the HCA strategy. Assume the cell is having totally 'N' number of channels. Queuing analysis for new calls and handoff calls are shown below.

6.4.1 Queuing of calls not applied to new calls and handoff calls

In this case, the blocking probability (p_b) of the handoff call and new call is given by

$$p_b = \frac{a^N}{N!} P(0) \tag{8}$$

Where

$$P(0) = \left(\sum_{n=0}^{N} \frac{a^N}{n!} \right)^{-1} \tag{9}$$

This is nothing but an erlang 'B' formula to calculate the blocking and dropping probability of new and handoff calls.

6.4.2 Queuing of calls applied only to new calls and not on handoff calls

In this case, the blocking probability of the new call (p_{nq}) with queue is given by

$$p_{nq} = \left(\frac{b_1}{N} \right)^{M_1} Pq(0) \tag{10}$$

Where

$$Pq(0) = \frac{1}{N! \sum_{n=0}^{N-1} \frac{a^{n-N}}{n!} + \frac{1 - \left(\frac{b_1}{N} \right)^{M_1+1}}{\left(1 - \frac{b_1}{N} \right)}} \tag{11}$$

Now the call dropping probability (p_{nd}) of the handoff call is given by

$$p_{nd} = \frac{1 - \left(\frac{b_1}{N} \right)^{M_1+1}}{1 - \left(\frac{b_1}{N} \right)} \times \frac{1}{N! \sum_{n=0}^{N-1} \frac{a^{n-N}}{n!} + \frac{1 - \left(\frac{b_1}{N} \right)^{M_1+1}}{1 - \left(\frac{b_1}{N} \right)}} \tag{12}$$

6.4.3 Queuing of calls applied only to handoff calls, and not on the new calls

Here the call dropping probability (p_{hd}) of the handoff call is given by

$$p_{hd} = \left(\frac{b_2}{N} \right)^{M_2} \times \frac{1}{N! \sum_{n=0}^{N-1} \frac{a^{n-N}}{n!} + \frac{1 - \left(\frac{b_1}{N} \right)^{M_1+1}}{1 - \left(\frac{b_1}{N} \right)}} \tag{13}$$

and the call blocking probability (p_{hb}) of the new call is given by

(14)

$$p_{hb} = \frac{1-\left(\frac{b_2}{N}\right)^{M_2+1}}{1-\left(\frac{b_2}{N}\right)} \times \frac{1}{N!\sum_{n=0}^{N-1}\frac{a^{n-N}}{n!} + \frac{1-\left(\frac{b_1}{N}\right)^{M_1+1}}{1-\left(\frac{b_1}{N}\right)}}$$

Where $b_1 = \frac{\lambda_1}{\mu}$, $b_2 = \frac{\lambda_2}{\mu}$ and $a = \left(\frac{\lambda_1 + \lambda_2}{\mu}\right)$

$1/\mu$ is the average calling time in seconds, including new calls and handoff calls in each cell.

7. Performance evaluation and Conclusion

The performance analysis of our proposed hybrid channel allocation with queuing of new calls and handoff calls are given in section 6. For simulation purpose, each cell having 78 channels (N) and the call holding time of 101 seconds = 0.028 hours are taken. From the total available channels N, in our proposed HCA strategy the fixed to dynamic ratio of the channels are taken as 7: 3 for new and handoff calls. The arrival rate of new calls (λ_1) is taken as 2270 and handoff calls attempted (λ_2), per hour is taken as 80. From the analysis of queuing and simulation results in HCA strategy, it is seen that, with queuing of new calls in HCA, the probability of blocking (p_b) is reduced compare to without queuing in FCA method. The response result is shown in figure 8. But at the same time, queuing of new calls only resulting an increased call dropping probability (p_d) on the handoff calls shown in figure 9. This situation should be avoided. However, at the same time, queuing of handoff calls only, shown in figure 10 will give, reduced call dropping probability (p_d) of the handoff calls with single space queue. Hence it is very worthwhile to implement a simple single space queue for handoff calls to reduce call dropping probability. Since in a single space queue method of adding queues in the handoff calls does not affect the call blocking probability (p_b) of the new calls. From the analysis of with queuing and without queuing, as shown in figure 11 for a HCA and DCA strategy respectively, for handoff calls it is seen that by applying a single space queue to the handoff calls will reduce the call dropping probability in a considerable amount, as compared with without queuing method. In wireless environment resource allocation for a call is very complex and complicated process. This will directly affect the overall system performance and the mobile users. Here our proposed HCA strategy with queuing shows that, as compared with normal FCA and DCA strategies, the new call blocking and handoff call dropping probabilities are very much reduced in a reasonable amount and thus this will increase the micro and pico cellular wireless system performance. It also satisfies the mobile user's requirements during busy peak hours and in normal environments. The simulation of HCA strategy with queuing shows that, it should always be aware that, the queuing of the handoff calls are more important than the queuing of new calls. Because the call drops during the handoff will disturbs the mobile users more than getting a new call connection with delay.

References

A.S. Acampora, & M. Naghshineh. (1994). An architecture and methodology for mobile-executed handoff in cellular ATM network. *IEEE J. Sel Areas in communication*, vol.12, No.8, 1365-1375.

C. Oliverira, J.B. Kim, & T. Suda. (1998). An adaptive band-width reservation scheme for high-speed multimedia wireless networks. *IEEE J. Sel. Areas in Communication*. vol.16, No.6, 858-874.

L.O.Guerrero, & A.H.Aghvami. (1998). A distributed dynamic resource allocation for a hybrid TDMA/CDMA system. *IEEE Trans. Veh. Technology*. Vol. 47, No.4, 1162-1179.

L.Xu.X.shen & J.W.Mark. (2005). Fair resource allocation with guaranteed statistical Qos for multimedia traffic in W-CDMA cellular network. *IEEE Trans. on mobile computing.* Volume 4, No. 2, 166 - 177

P. Ramanathan, K.M. Sivalingam, P. Agrawal, & S. Kishore. (1999). Dynamic resource allocation scheme during handoff for mobile multimedia wireless networks. *IEEE J. Sel. Areas Communication*. vol.17, No.7, 1270-1283.

S.Bajaj, L.Breslau, D. Estrin, K.Fall, & S.kumar. (1999). Improving simulation for network research. USC computer science department report.

Tao Zhang, & Prathima Agarwal. (2001). Local predictive resource reservation for handoff in multimedia wireless networks. *IEEE J. SAC*,Vol. 19, No.10, 1931-1941

V.Syrotuik et..al. (2004). Dynamic spectrum utilization in Ad Hoc Networks. *The International Journal of Computer and Telecommunications Networking*. Vol.46 , No.5, Special issue: Military communications systems and technologies. 665 - 678

Wei Li, & Xiuli Chao, (2004). Modeling and performance evaluation of a cellular mobile network. IEEE/ACM Transactions on Networking, Vol. 12 , No. 1, 131-145.

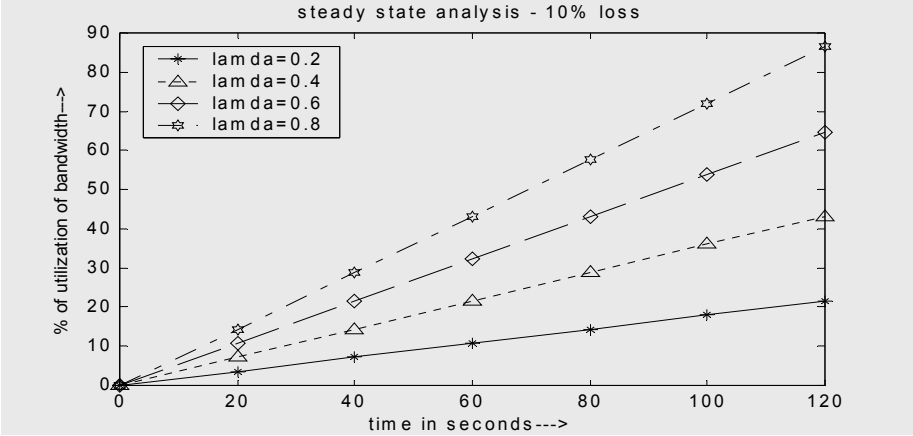


Figure 1. Steady state response with 10% loss for different call arrival rates

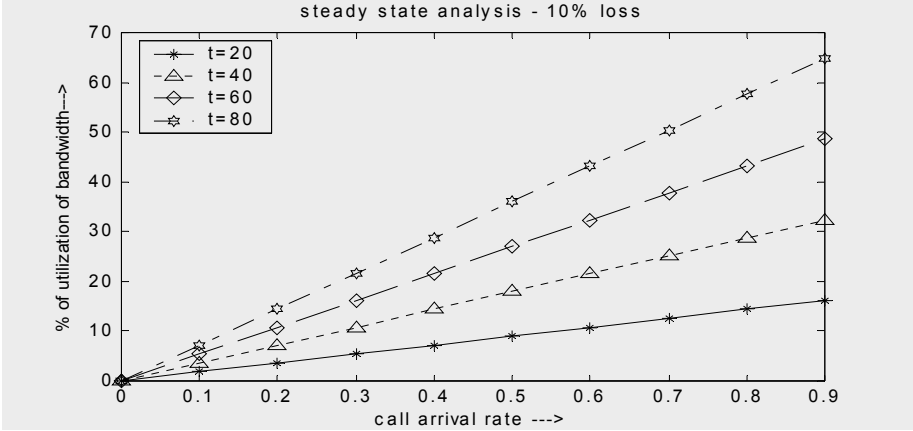


Figure 2. Steady state responses with 10% loss for different average call duration time

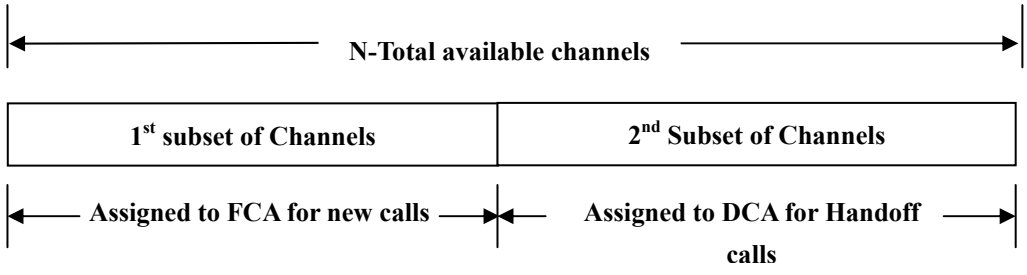


Figure 3. Channel Allocation in HCA Strategy

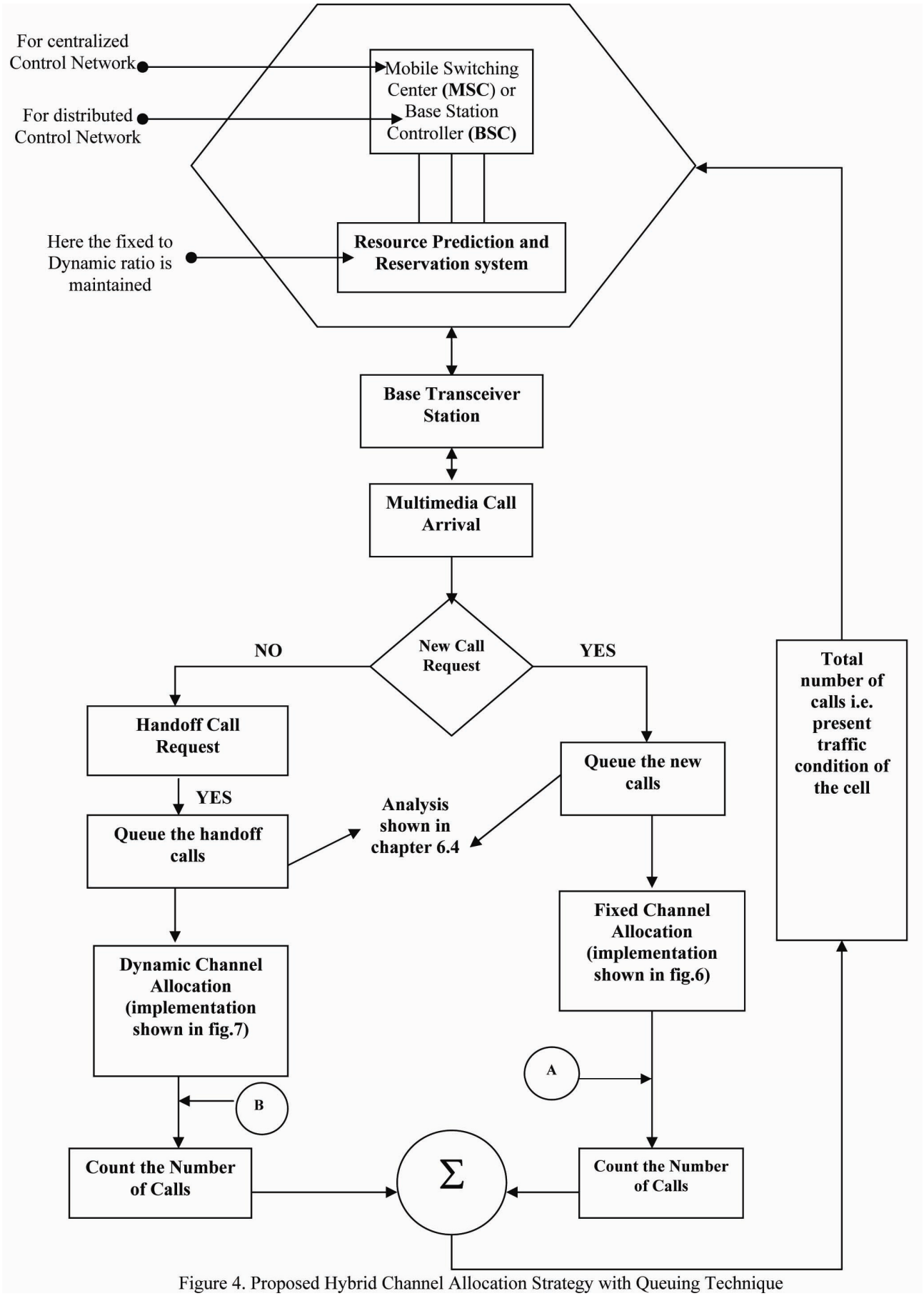


Figure 4. Proposed Hybrid Channel Allocation Strategy with Queuing Technique

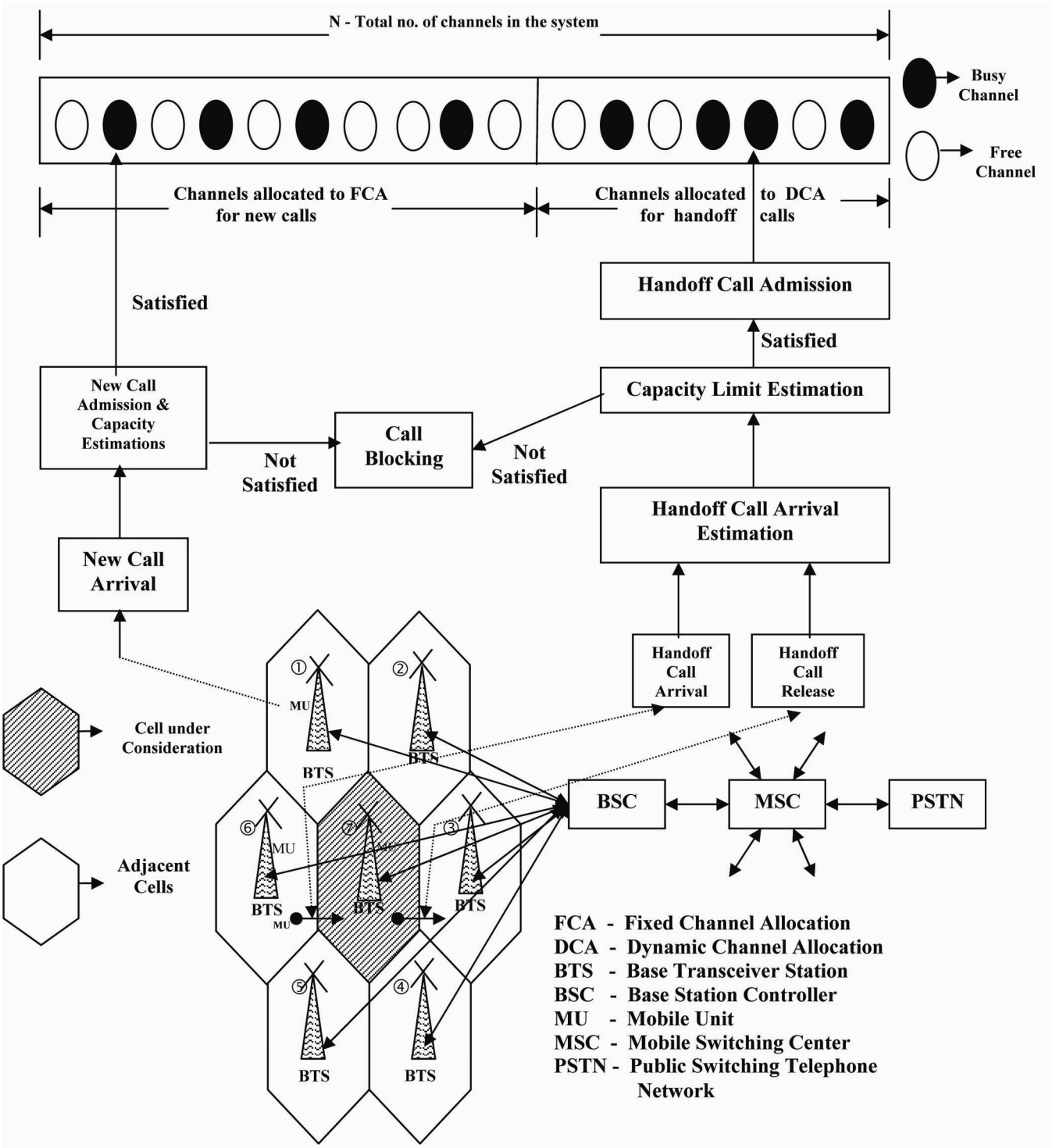


Figure 5. Call Admission Procedure in our Proposed HCA Strategy

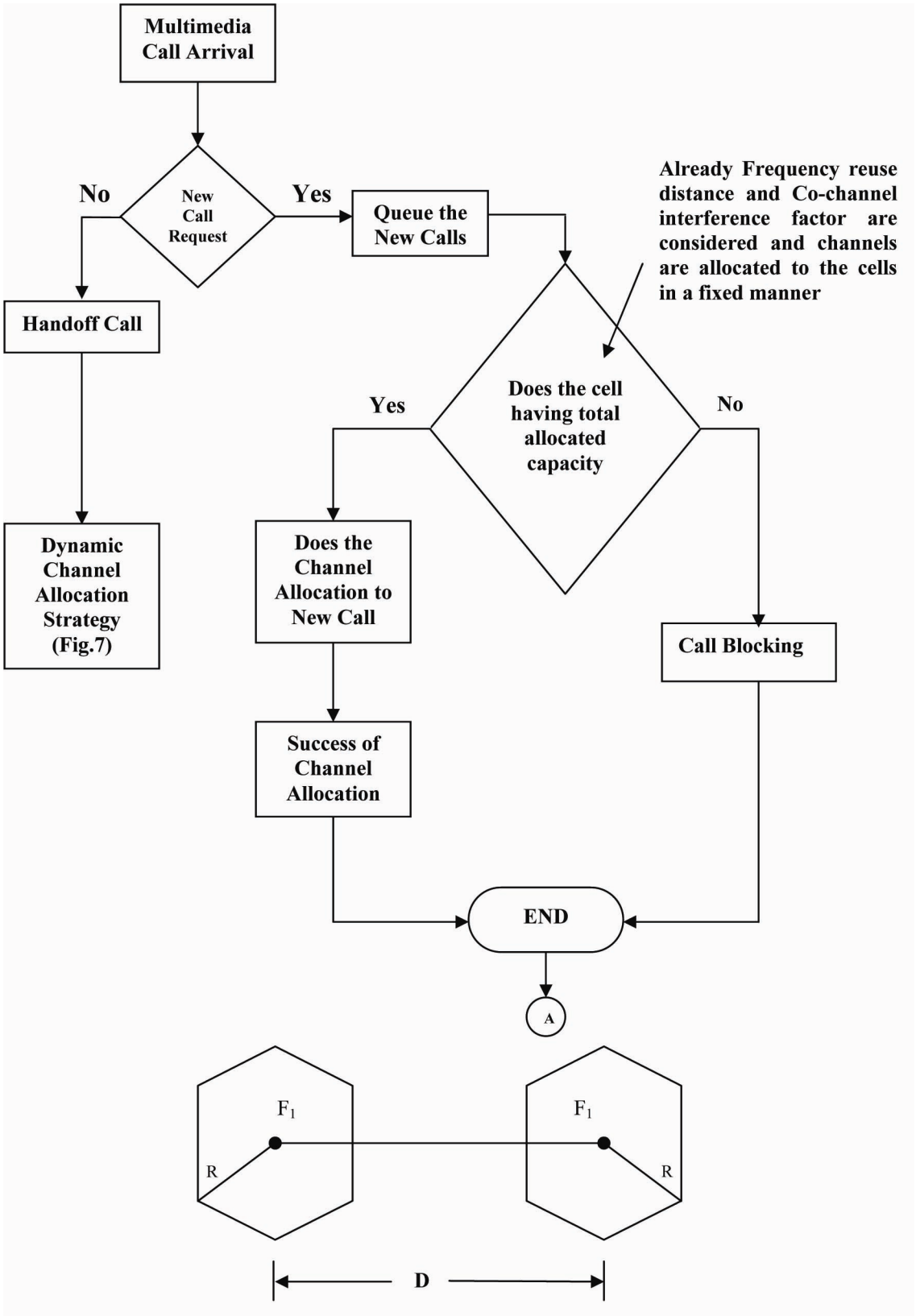


Figure 6. Fixed Channel Allocation for New Calls in HCA Strategy

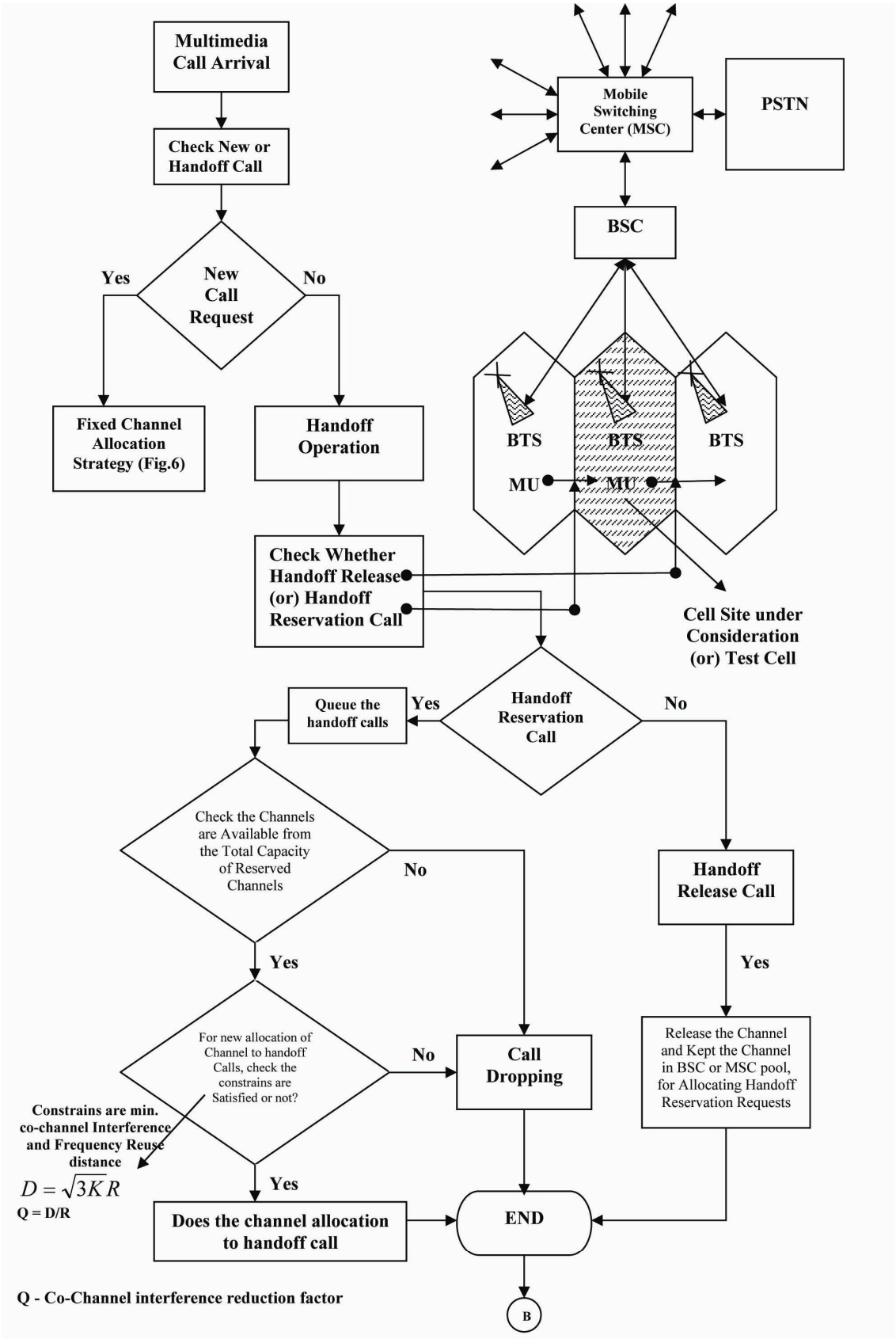


Figure 7. Dynamic Channel Allocation for Handoff Calls in HCA Strategy

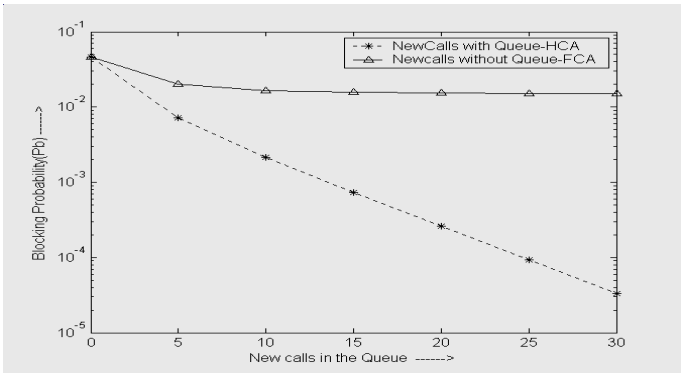


Figure 8. New call blocking probability (Pb) comparison between FCA and HCA strategies

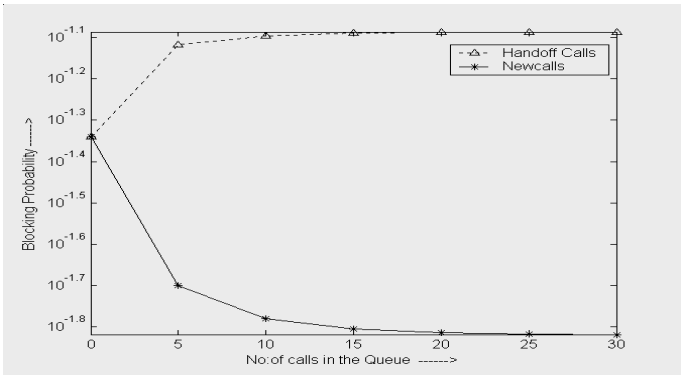


Figure 9. Call blocking probability comparison between new calls and handoff calls with new calls queued

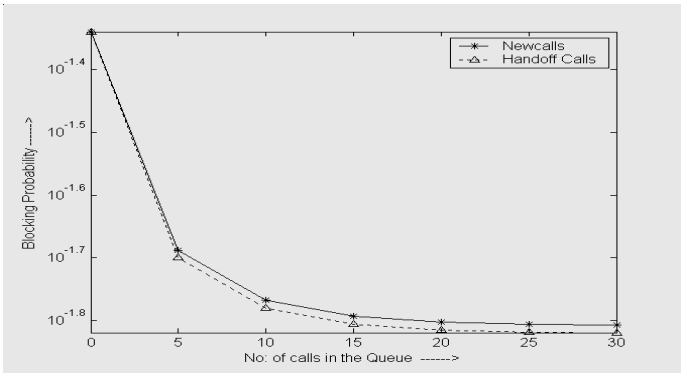


Figure 10. Call blocking probability comparison between new calls and handoff calls with handoff calls queued

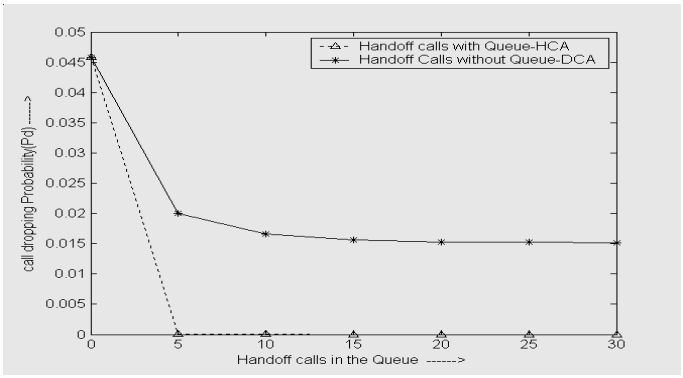


Figure 11. Handoff call dropping probability (Pd) Comparison between DCA and HCA strategies

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Machine Learning Emulation in Nature-inspired Computation Systems

Jianhao Tan (Corresponding author)

Electrical and Information Engineering College, Hunan University

Changsha 410082, China

E-mail: tanjianhao96@sina.com.cn

Jing Zhang

Electrical and Information Engineering College, Hunan University

Changsha 410082, China

Fu Guo

Electrical and Information Engineering College, Hunan University

Changsha 410082, China

The research is financed by National Natural Science Foundation (Approval No. 60634020) and Hunan Province Natural Science Foundation (Approval No. 08JJ3132).

Abstract

The whole frame of nature_inspired computation systems is inquired into, the characteristics of machine learning in nature_inspired computation systems are researched, and a particular scheme on machine learning in nature_inspired computation systems is designed with environment being gathered present data; study unit adopting fuzzy optimizatio algorithm based on genetic algorithm; knowledge base adopting fuzzy optimization BP neural networks; executive unit being complicated industry process. The fuzzy optimizatio learning algorithm of fuzzy optimization BP neural networks is built, the flow chart of the algorithm is constructed, and the emulation test is made. At last, the design criteria of flash metal comsuption are obtained, and the stability of the algorithm is verified through this example. The result shows that machine learning makes nature_inspired computation systems be able to gain know; edge automatically, their quality improved, their intelligent level advanced, and machine learning will greatly influence the memory mode, information input mode and system structures of nature_inspired computation systems.

Keywords: Nature_inspired computation, Machine learning, Emulation, Genetic algorithm, Fuzzy optimization, Neural networks, Flash metal consumption

1. Introduction

Based on the natural world, particularly on the functions, characteristics and mechanism of the organism, nature_inspired computation researches various information processing mechanisms contained in them, takes out the relevant calculation models, and designes the relevant algorithms which are applied in various fields (Cai Zixing, Xu Guangyou, 2004; Cheung L C Y, Holden T S I, 1991, pp499-508; Chien Steve, Decoste Dennis, Doyle Richard, et al, 1997, pp103-121; Mccormark D M, Day R, 1993, pp 6; Ng E Y K, Fo K S C, Peh Y C, et al, 2002, pp152-157). Nature_inspired computation is usually a class of self-adaptive, self-organized, self-learning algorithms, including these researching fields such as evolutionary computation, neural computation, ecological terms, quantum computing and complex adaptive systems and so on which base the design of algorithms on the natural world mechanism, having the features of imitating the natural world and being able to solve many complex problems that the traditional method is hard to do. It has a very good use in such fields as optimizing designing, optimal controlling, computer network security, creative designing in large-scale complex systems.

According to the view that the nature_inspired computation is the internal mechanism in the natural world and the basis of intelligence, a general framework for nature_inspired computation is made (Chien Steve, Decoste Dennis, Doyle

Richard, et al, 1997, pp103-121). However, whether the natural_inspired computation is the internal mechanism of nature phenomenon needs waiting for future experiment and natural phenomena verification. Besides, the model which mapps the computation mode exsiting in natural world broadly to the new nature_inspired computation mode has brought forward, and then the structure, principle and characteristic of the model have been analysed (Gong Tao, Cai Jingfeng, Cai Zixing, 2003, pp727-732). Nature_inspired computation can enhance many characteristics in general systems and give them new life. Because nature_inspired computation is an innovative one, the development of the calculation will promote the fundamental development of systematics including the generalization of system structure and the expansion of system domain. How to coordinate the relationship between the traditional computations and such emerging computations as nature_inspired computation and how to deal with the relationship among the branches of these emerging computations are pending further study.

On the basis of describing the whole framework of nature_inspired computation systems, the article studies emphatically the mechanism of action of machine learning in them. A particular scheme on machine learning in nature_inspired computation systems is designed with environment being gathered present data; study unit adopting fuzzy optimizatio algorithm based on genetic algorithm; knowledge base adopting fuzzy optimization BP neural networks; executive unit being complicated industry process. The principle, process and characteristics of the machine learning in nature_inspired computation systems are described and exemplied.

2. Whole Frame of Nature_inspired Computation Systems

Nature_inspired computation makes the high_level decision_making in the computer, the advanced mathematical modeling and synthetic approach in system theory, as well as linguistics methods dealing with inaccurate and incomplete information combined together, forming an agreed method to meet the project. It consists of such three levels as organization, coordination and implementation.

The organization level functions imitating human behavior, having highest-level intelligence in nature_inspired computation systems. It acts as reasoning, planning, decision-making and exchanging long-term memory information, as well as learning under the information from external environment and lower level feedback; Being an interface between the organization level and the implementation level, the coordination level functions coordinating these tasks in accordance with the instruction of the organization level; Being the lowest level, the implementation one completes various detailed tasks.

The intelligence of nature_inspired computation is mainly represented in high level. These problems encountered in the high level are often uncertain, and so adopting the knowledge_based organization level is just about right because the knowledge-based organization level can facilitly process information and use the instinct reasoning logic and experience of human being. The working process of the system can be described in two ways: from the horizontation, dividing a complex system into a number of interrelated sub_systems, each of which is equiped a controller in order to be easy to be directly controlled, as so to make the complex issue simplified in a large extent; from the vertication, breaking down the complex system from high to low according to the quantity of the knowledge or the level of intelligence needed for the system, which means that intelligence controller systems have two_fold meaning in the structure of the multi_level: on one hand, the controlled object has many levels, some of which have many uncertain and unknown factors. The more the levels are, the harder the accorrding control is; On the other hand, there exsits the control in many levels, which is related with the designing, development and running of nature_inspired computation systems. The whole framework of nature_inspired computation systems is shown in Figure.1 (a), and these parts instructed by the rough arrows are the machine learning sub_systems abstracted from nature_inspired computation systems, which constitute the main feedback loops in nature_inspired computation systems and whose diagram is shown in Figure.1(b).

3. Machine Learning in Nature_inspired Computation Systems

Knowledge,knowledge representation and reasoning algorithm using knowledge are the core of the nature_inspired computation systems, and the machine learning is the key among them. For centuries, philosophers and psychologists have believed: the basic mechanism on learning is trying translating successful performance action in one situation to another similar new situation. Learning is the process of acquiring knowledge, experiencing, improving performance, finding regularity and adapting environment. Figure.1 (b) gives a simple model on machine learning in nature_inspired computation systems.The model includes four basic sections of machine learning systems. The environment provides external information, similar to the role of the teacher, the study unit deals with the information provided by the environment, equivalent to a variety of learning algorithms, the knowledge base stores information by means of some knowledge representation mode, the executive unit completes some using the knowledge in the knowledge base.

A particular scheme on machine learning in nature_inspirde computation system is given with environment being gathered present data, knowledge base adopting artificial neural network, study unit adopting fuzzy optimization algorithm based on genetic algorithm, and executive unit being complicated industry process. The diagram of the

system is shown in Figure.2.

3.1 Neural Network Model

The knowledge base of this article adopts fuzzy optimization BP neural networks (Horikawa S, 1992, 25-29). The input_output relationship of the neural network can be described as follows.

To Node i of the input layer, its input is

$$I_i = x_i \quad i = 1,2,\cdots,M \tag{1}$$

Its output is

$$O_i = I_i \tag{2}$$

To Node j of the hidden layer, its input is

$$I_j = \sum_{i=1}^{M} w_{ij} O_i \quad i = 1,2,\cdots,M; j = 1,2,\cdots,Q \tag{3}$$

Its output is

$$O_j = f(I_j) = \frac{1}{1+(I_j^{-1}-1)^2} = \frac{1}{1+[(\sum_{i=1}^{M} w_{ij} O_i)^{-1}-1]^2} \tag{4}$$

The output layer has only a node, its input is

$$I = \sum_{j=1}^{Q} w_j O_j \tag{5}$$

Its output is

$$y = O = f(I) = \frac{1}{1+(I^{-1}-1)^2} = \frac{1}{1+[(\sum_{j=1}^{Q} w_j O_j)^{-1}-1]^2} \tag{6}$$

The input-output relationship of the neural network can be written as $y = F(\boldsymbol{X}, \boldsymbol{w})$, where $\boldsymbol{X} = (x_1, x_2, \cdots, x_M)$, $\boldsymbol{w} = (w_{ij}, w_j \,|\, i = 1,2,\cdots,M; j = 1,2,\cdots,Q)$ .The process of the weight estimation of the neural network is:The N instances $(\boldsymbol{X}_k, Y_k^*), k = 1,2,\cdots,N$ which are known constitute a group of study samples, where the input x of the instance k can be represented as a vector $\boldsymbol{X}_k = (x_{1k}, x_{2k}, \cdots, x_{Mk})$ of M variables, and the expected output of the instance K is the single output $Y_k^* = y_k^*$, $y_k = F(\boldsymbol{X}_k, \boldsymbol{w})$.

Then the problem of the weight estimation of the neural network is converted into the fuzzy optimization problem as discribed below (Jimenez F, Sanchez G, Cadenas J M, Gomez-Skameta A F. A, 2004, 23-26).

$$\max \; y_k = F(\boldsymbol{X}_k, \boldsymbol{w}), \quad k = 1,2,\cdots N \tag{7}$$

Where y_k, $k = 1,2,\cdots N$ is the fuzzy goal, whose expected value is y_k^*, whose symmetry tolerance is ε, and whose changing zone is $\left[y_k^* - \varepsilon, y_k^* + \varepsilon \right]$. In Domain A, we create fuzzy goal set $\boldsymbol{G}_k, k = 1,2,\cdots,N$ corresponding to y_k, whose membership function $\mu_{\boldsymbol{G}_k}(\boldsymbol{w})$ is defined as follows:

$$\mu_{\boldsymbol{G}_k}(\boldsymbol{w}) = \begin{cases} 0 & y_k \le y_k^* - \varepsilon \\ 1 - \dfrac{y_k^* - y_k}{\varepsilon} & y_k^* - \varepsilon \le y_k \le y_k^* \\ 1 - \dfrac{y_k - y_k^*}{\varepsilon} & y_k^* < y_k < y_k^* + \varepsilon \\ 0 & y_k \ge y_k^* + \varepsilon \end{cases} \tag{8}$$

If the fuzzy goal set $\boldsymbol{G}_k, k = 1,2,\cdots,N$ is given in Domain A, the intersection $\boldsymbol{G} = \bigcap_{k=1}^{N} \boldsymbol{G}_k$ is called fuzzy superior set.The basic idea of the weight estimation of fuzzy optimization based on neural network is: in the decision space $\boldsymbol{A}$, finding $\boldsymbol{w}^*$ which makes the membership function $\mu_{\boldsymbol{G}_k}(\boldsymbol{w})$ of the fuzzy superior set $\boldsymbol{G}$ get the maximal value. $\boldsymbol{w}^*$ is

called the fuzzy optimum solution. $\mu_G(\boldsymbol{w})$ can be calculated by the following equation (Jimenez F, Sanchez G, Cadenas J M, Gomez-Skameta A F. A, 2004, 23-26).

$$\mu_G(\boldsymbol{w}) = \bigwedge_{k=1}^{N} \mu_{G_k}(\boldsymbol{w}) = \min(\mu_{G_k}(\boldsymbol{w}) \big| k = 1, 2, \cdots, N) \tag{9}$$

The mathmatical model of fuzzy optimization can be described as follows: solving w^* in order to make

$$\mu_G(\boldsymbol{w}^*) = \max(\mu_G(\boldsymbol{w})) = \max(\min(\mu_{G_k}(\boldsymbol{w}) \big| k = 1, 2, \cdots, N) \tag{10}$$

3.2 Learning Algorithm

1) Weight Estimation of Neural Networks

The issue (10) is the unconstrained optimization problem, but its goal is not continuously derivative. This problem can not be solved by traditional methods but can use genetic algorithms to optimize, This article try to solve the problem by adopting the particular genetic algorithm using mutating along the weighted gradient direction, which was brought up by Wang Dingwei and Tang Jiafu (Tang Jiafu, Wang Dingwei, 2000).The algorithm regardes the mutation as the main operator, then adoptes the arithmetic combination crossover operator in the later period, The basic idea of the algorithm is: firstly, randomly generating initial population containing *pop_size* individuals, which are chosen to generate descendants whose membership $\mu_G(\boldsymbol{w})$ on the fuzzy superior set $\boldsymbol{G}$ gets increased. With the genetic algorithm executed, these individuals whose membership is less than α_0 (acceptable membership) get less chance of generating offsprings than others. With the genetic generation number increasing, these individuals whose membership is less than α_0 will eventually die,while others get survival. After a few of generations, the membership of all individuals are greater than α_0, and most of individuals will be close to optimum solution.

To the individual $\boldsymbol{w}$, let $\mu_{\min}(\boldsymbol{w}) = \min(\mu_{G_k}(\boldsymbol{w}) \quad \big| k = 1, 2, \cdots, N)$, if $\mu_{\min}(\boldsymbol{w}) \le \mu_{G_k}(\boldsymbol{w}) < 1$, move along the gradient direction of $\mu_G(\boldsymbol{w})$ so as to improve the value of $\mu_G(\boldsymbol{w})$. The smaller $\mu_G(\boldsymbol{w})$ is, the more improvement it gets. Based on the above idea, the weighted gradient direction can be constructed as follows.

$$D(\boldsymbol{w}) = \sum_{k=1}^{N} \alpha_k \nabla \mu_{G_k}(\boldsymbol{w}) = \sum_{k=1}^{N} \alpha_k (\nabla \mu_{G_k}(w_j), \nabla \mu_{G_k}(w_{ij})) = (D(w_j), D(w_{ij})) \tag{11}$$

Where α_k is the weight along the gradient direction, which is defined as follows.

$$\alpha_k = \begin{cases} 0 & \mu_{G_k} = 1 \\ \dfrac{1}{\mu_{G_k} - \mu_{\min} + e} & \mu_{\min} \le \mu_{G_k} < 1 \end{cases} \tag{12}$$

Where *e* is a small-enough positive number, $1/e$ is the maximal weight.

$$\nabla \mu_{G_k}(w_j) = \frac{1}{\varepsilon} \operatorname{sgn}(y_k^* - y_k) \nabla y_j \quad k = 1, 2, \cdots, N \tag{13}$$

$$\nabla \mu_{G_k}(w_{ij}) = \frac{1}{\varepsilon} \operatorname{sgn}(y_k^* - y_k) \nabla y_{ij} \quad k = 1, 2, \cdots, N \tag{14}$$

Now ∇y_j and ∇y_{ij} are derivated as follows:

$$\nabla y_j = \frac{\partial y}{\partial w_j} = \frac{\partial y}{\partial I} \frac{\partial I}{\partial w_j} = \frac{\partial O}{\partial I} \frac{\partial (\sum_{j=1}^{Q} w_j O_j)}{\partial w_j} = f'(I) O_j \tag{15}$$

$$\nabla y_{ij} = \frac{\partial O}{\partial O_j} \frac{\partial O_j}{\partial I_j} \frac{\partial (\sum_{i=1}^{M} w_{ij} O_i)}{\partial w_{ij}} = \frac{\partial O}{\partial O_j} f'(I_j) O_i \tag{16}$$

Whereas

$$\frac{\partial O}{\partial O_j} = \frac{\partial O}{\partial I} \frac{\partial I}{\partial O_j} = \frac{\partial O}{\partial I} \frac{\partial (\sum_{j=1}^{Q} w_j O_j)}{\partial O_j} = f'(I) w_j \tag{17}$$

So

$$\nabla y_{ij} = f'(I) f'(I_j) w_j O_i \tag{18}$$

Therefore

$$\nabla \mu_{G_k}(w_j) = \frac{1}{\varepsilon}\operatorname{sgn}(y_k^* - y_k) f'(I) O_j \tag{19}$$

$$\nabla \mu_{G_k}(w_{ij}) = \frac{1}{\varepsilon}\operatorname{sgn}(y_k^* - y_k) f'(I) f'(I_j) w_j O_i \tag{20}$$

$$D(w_j) = \frac{1}{\varepsilon}\sum_{k=1}^{N} \alpha_k \operatorname{sgn}(y_k^* - y_k) f'(I) O_j \tag{21}$$

$$D(w_{ij}) = \frac{1}{\varepsilon}\sum_{k=1}^{N} \alpha_k \operatorname{sgn}(y_k^* - y_k) f'(I) f'(I_j) w_j O_i \tag{22}$$

Where

$$f'(I) = \frac{2I^{-2}(I^{-1}-1)}{(1+(I^{-1}-1)^2)^2} \tag{23}$$

$$f'(I_j) = \frac{2I_j^{-2}(I_j^{-1}-1)}{(1+(I_j^{-1}-1)^2)^2} \tag{24}$$

According to Formula (10), to an individual w, its adaptive value is just the minimum membership in the fuzzy goal set. The aim of the optimization is to make the minimum membership increasing, so the priority should be given the goal with the minimum membership.

To w_j, the offspring $w_j^l(t+1)$ which is produced by the mutation of $w_j^l(t)$ along the weighted gradient direction ,where l means the number of individuals and t means the generation number, can be described as follows:

$$w_j^l(t+1) = w_j^l(t) + \beta^t D(w_j^l(t)) \tag{25}$$

Where β^l is the random step length which has declining mean value with Erlang distribution. The Erlang distribution is generated by randomizer.

Similarly, to w_{ij}, there is :

$$w_{ij}^l(t+1) = w_{ij}^l(t) + \beta^t D(w_{ij}^l(t)) \tag{26}$$

From the weight and variation formulas (11) ~ (26), it is seen that the goals unable to be satisfied have the lest membership and can obtain the maximum weight $1/e$, so the variation will guide the individual to the feasible region. As $\mu_G(\boldsymbol{w}) > 0$, the goal with the minimum membership obtains the maximum weight. Therefore, the weighted gradient direction will improve the minimum membership value so as to make $\mu_G(\boldsymbol{w})$ improved. As a result, the weighted gradient direction will guide all individuals to approach the precise optimal resolution, and these individuals constitute one neighborhood including the precise optimal resolution.

In the genetic algorithm of the article, the variation along the weighted gradient direction is the main operator, the arithmetic combination crossover is only used in the later period, and the new population is selected in each generation by means of the proportional selection strategy.

2) Design of Fuzzy Optimization Algorithms

[Algorithm1: The Tans Algorithm]

Input: training samples

Output :neural network weight

According to the basic strategy discribed above, the algorithm is constructed as follows.

(1) Scanning the data base and reading the samples $(\boldsymbol{X}_k, Y_k^*), k = 1, 2, \cdots, N$.

(2)Constructing the objective function $y_k = F(\boldsymbol{X}_k, \boldsymbol{w})$ whiose expected value is $y_k^* = Y_k^*$ and giving the symmetry tolerance ε.

(3)Giving the population scope *pop_size* and making *t*=0, then initializating the population $w_j^l(0)$, $w_{ij}^l(0), l = 1, 2, \cdots, pop\_size$.

(4)To each individual $w_j^l(t)$ and $w_{ij}^l(t)$,

①Using Eq.'s(2)-(7), computing the value $y_k^l(t)$ of each objective function.

②Using Eq.(9), computing the membership $\mu_{G_k^l(t)}(w^l(t))$ of the fuzzy goal set.

③Using Eq.(10), determing the membership $\mu_{G^l(t)}(w^l(t))$ of the fuzzy superior set.

④Using Eq.(19), Eq.(20) and Eq.(12), respectively determing the gradient $\nabla\mu_{G_k^l(t)}(w_j^l(t))$ and $\nabla\mu_{G_k^l(t)}(w_{ij}^l(t))$, and the weight $\alpha_k(t)$ of each goal, and using Eq.(21) and Eq.(22), determing the weighted gradient direction $D(w_j^l(t))$ and $D(w_{ij}^l(t))$.

(5)By the scale-selecting algorithm, choosing new parameter population $w_j^l(t+1)$ and $w_{ij}^l(t+1)$.

(6)According to the arithmetic combination method, executing crossover operation.

(7)Using Eq.(25) and Eq.(26), executing the mutation operation along the the weighted gradient direction.

(8)Whether to meet the conditions for termination ?

① If not, making t=t+1, then jumping to (4).

② If so, exporting $w_j^l(t)$ and $w_{ij}^l(t)$ which makes $\mu_{G^l(t)}(w^l(t))$ the largest as BP fuzzy neural network weight.

4. Emulation Example

Machine learning systems can solve many engineering problems. This article only introduces the application example of the algorithm in determinging flash metal consumption design criteria in a nature-inspired computation system.

4.1 Relationship among the Variables in Flash Metal Consumption Design Criteria

According to the former Soviet Union scholar Jiejielin's research result about the factors afffecting flash metal consumption and the analysis of the field data collected from some factory, the relationship among the varibles in flash metal consumption design criteria can be described as (Xiao Jingrong, Li Dequn trans, 1983)

$$\begin{cases} Q_{FY}/Q_Y*100\% &= f_1(Q_Y^{-0.2}, S_Y, (D_0/D_Y)^2, (B_Y/H_Y), (D_0/D_Y)^2*S_Y) \\ Q_{FZ}/Q_Z*100\% &= f_2(Q_Z^{-0.2}, S, (D_Y/D_Z)^2, (B_Z/H_Z), (D_Y/D_Z)^2*S) \end{cases} \tag{27}$$

Where Q_{FY} and Q_{FZ} are the flash metal consumption of pre-forging and finish_forging phases, Q_Y and Q_Z are the weight of pre-forging and finish_forging pieces, D_0 , D_Y and D_Z are the maximum diameter of rough, preforging piece and finish_forging piece, and S_Y and S are the shape complex coefficient of pre-forging and finish_forging phases.

4.2 Example

The samples tables responding to all sub-formula of Eq. 27 are respectively written in Table.1 and Table.2.

In the machine learning emulation system, the specific expression forms of all sub-formula of Eq.27 are minied using the Tans Algorithm 1. The neural network operating parameters are set as follows: input layer adopting 5 nodes, hidden layer adopting 15 node and output layer adopting 1 node. The GA parameters are set as follows: *MIND*(the number of individuals)=40, the number of excellent individuals being 5, *MAXGEN*(the maximum genetic number)=100, chromosome using real-code, *GGAP*(generation gap)=0.9, P_c (crossover probability)=0.7, P_m (variation probability)=0.05, and ε (symmetry tolerance)=0.1.

According to the running result, finally get

Weight matrix of Connecting the input layer with the hidden layer: to the first sub-formula of Eq.(27)

$$(w_{ij}^l)=\begin{bmatrix} 0.21 & 0.34 & 0.35 & 0.26 & 0.28 & 0.32 & 0.45 & 0.42 & 0.28 & 0.30 & 0.55 & 0.12 & 0.32 & 0.11 & 0.58 \\ 0.12 & 0.22 & 0.24 & 0.18 & 0.27 & 0.23 & 0.11 & 0.32 & 0.53 & 0.21 & 0.32 & 0.18 & 0.63 & 0.38 & 0.23 \\ 0.14 & 0.78 & 0.25 & 0.55 & 0.34 & 0.21 & 0.39 & 0.75 & 0.42 & 0.45 & 0.19 & 0.24 & 0.37 & 0.23 & 0.41 \\ 0.58 & 0.34 & 0.48 & 0.65 & 0.18 & 0.65 & 0.23 & 0.21 & 0.11 & 0.13 & 0.46 & 0.15 & 0.29 & 0.29 & 0.27 \\ 0.24 & 0.18 & 0.16 & 0.19 & 0.72 & 0.34 & 0.14 & 0.22 & 0.21 & 0.27 & 0.13 & 0.36 & 0.13 & 0.54 & 0.14 \end{bmatrix} \tag{28}$$

To the second sub-formula

$$(w_{ij})=\begin{bmatrix} 0.32 & 0.24 & 0.34 & 0.27 & 0.25 & 0.31 & 0.43 & 0.41 & 0.68 & 0.33 & 0.56 & 0.32 & 0.34 & 0.15 & 0.18 \\ 0.14 & 0.32 & 0.25 & 0.15 & 0.17 & 0.24 & 0.12 & 0.12 & 0.51 & 0.42 & 0.31 & 0.16 & 0.61 & 0.48 & 0.43 \\ 0.24 & 0.68 & 0.17 & 0.54 & 0.33 & 0.32 & 0.23 & 0.45 & 0.32 & 0.15 & 0.29 & 0.25 & 0.36 & 0.72 & 0.21 \\ 0.56 & 0.54 & 0.38 & 0.65 & 0.17 & 0.45 & 0.43 & 0.51 & 0.18 & 0.33 & 0.36 & 0.17 & 0.28 & 0.24 & 0.57 \\ 0.34 & 0.23 & 0.26 & 0.29 & 0.62 & 0.64 & 0.17 & 0.42 & 0.27 & 0.26 & 0.11 & 0.32 & 0.16 & 0.53 & 0.24 \end{bmatrix} \quad (29)$$

Weight vector connnecting the hidden layer with the output layer: to the first sub-formula of Eq.(27)

$$(w_j)=[0.23\ \ 0.42\ \ 0.53\ \ 0.11\ \ 0.72\ \ 0.25\ \ 0.39\ \ 0.43\ \ 0.64\ \ 0.13\ \ 0.27\ \ 0.13\ \ 0.61\ \ 0.29\ \ 0.41] \quad (30)$$

To the second sub-formula

$$(w_j)=[0.13\ 0.52\ 0.23\ 0.42\ 0.35\ 0.65\ 0.18\ 0.25\ 0.72\ 0.15\ 0.38\ 0.23\ 0.11\ 0.59\ 0.38] \quad (31)$$

5. Conclusion

This article discusses the whole framework of nature_inspired computation systems, deeply researches the action mechanism of machine learning in nature_inspired computation systems, and gives a particular scheme on machine learning which is emulated. The results show that machine learning makes natue_inspired computation systems be able to automatically get knowledge, their performance improved, them given more intelligence. Machine learning will have a significant impact on the storage mode, information input mode and architecture of nature_inspired computation systems. In the scheme, because the initial weight of neural networks and the initial groups of genetic algorithm are randomly generated, the results of the algorithm will be more or less unstable. But ,if the algorithm model is reasonably choosed and the algorithm parameters are appropriatly setted , the results will be relatively stable.

References

Cai Zixing & Xu Guangyou. (2004). *Artificial Intelligence and Application (Version III)*. Beijing: Tsinghua University Press.

Cheung L C Y, Holden T S I. (1991). Survey of artificial intelligence impacts on information system & engineering. *Information and Software Technology*, 1991, 33(7): 499-508.

Chien Steve, Decoste Dennis, Doyle Richard, et al. (1997). Making an impact artificial intelligence at the jet propulsion laboratory. *A.I.Magazine*, 1997, 18(1): 103-121.

Gong, Tao, Cai, Jingfeng, Cai, Zixing. (2003). A Coding and Control Mechanism of Natural Computation [C] / / Yen G Gary, L IU Derong. *Proceedings of the 2003 IEEE International Symposium on Houston, Piscataway*: IEEE Society Press, 2003, 727-732.

Guimarse F G, Campelo F, Saldanla R R. (2006). A hybrid methodology for fuzzy optimization of electromagnetic devices. *IEEE Transaction on Magnetics*, 2006,15(4):15-19.

Horikawa S. (1992). On Fuzzy modeling using fuzzy neural networks with BP Algorithm. *IEEE Trans on Neural Networks*, 1992, 39(5): 25-29.

Jimenez F, Sanchez G, Cadenas J M, Gomez-Skameta A F. (2004). A multi_objective evolutionary approach for nonlinear constrained optimization with Fuzzy costs. *Systems, Man And Cybernetics*, 2004, 12(2):23-26.

Jin Y W, Shen H, Li K Q, et al. (2005). Solution to multi-objective fuzzy optimization dynamic programming with uncertain information. *Parallel and distributed computing-Application and Technologies*, 2005, 32(2):53-57.

Liu Y K. (2006). Convergent results about the use of fuzzy simulation in fuzzy optimization problems. *Fuzzy Systems*, 2006, 23(1):34-37.

Mccormark D M, Day R. (1993). How artificial intelligence impacts E&P productivity. *World Oil*, 1993, 214 (10): 6.

Ng E Y K, Fo K S C, Peh Y C, et al. (2002). Computerized detection of breast cancer with artificial intelligence and thermograms. *J of Medical Engineering and Technology*, 2002, 26 (4): 152-157.

Shi, Zhongzhi. (2006). *Advanced artificial intelligence*. Beijing: Science Press.

Tang, Jiafu, Wang, Dingwei. (2000). Fuzzy optimization theory and methods. *Control Theory and Applications*.

Xiao, Jingrong, Li, Dequn trans. (1983). *Optimization and automation principles of hot volume die-forging process design*. Beijing: National Defense Industry Press.

Yi, Jikai, Hou, Yuanbin. (2003). *Intelligent control technology*. Beijing: Beijing Technology University Press.

Table 1. Q_{FY}/Q_Y training samples of fuzzy optimization BP neural networks

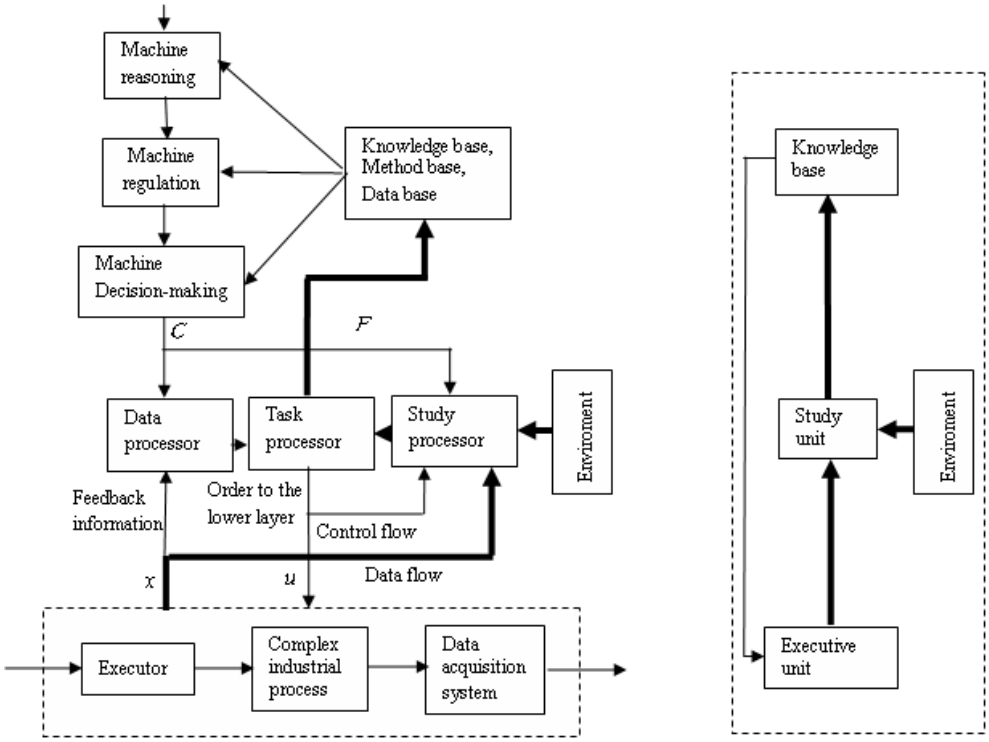
$Q_Y^{-0.2}$	S_Y	$(D_0/D_Y)^2$	B_Y/H_Y	$S_Y*(D_0/D_Y)^2$	Q_{FY}/Q_Y
0.75390	1.62942	0.88054	2.93900	1.43477	0.17662
0.80757	2.69403	1.01835	3.00000	2.73980	0.13091
0.78038	3.36251	0.85031	4.26471	2.85919	0.15356
…					
0.79921	3.10378	0.78632	3.00000	2.44056	0.15168

Table 2. Q_{FZ}/Q_Z training samples of fuzzy optimization BP neural networks

$Q_Z^{-0.2}$	S	$(D_0/D_Z)^2$	B_Z/H_Z	$S*(D_Y/D_Z)^2$	Q_{FZ}/Q_Z
0.76829	1.01682	0.99431	3.78000	1.01103	0.09915
0.93519	0.86370	0.79549	4.26670	0.68706	0.09338
1.00778	0.99236	0.95413	4.00000	0.94684	0.15615
…					
0.67096	1.16473	0.99598	3.00000	1.16005	0.16811

Table 3. Test results of flash metal consumption based on fuzzy optimization

NO.	Sample No	Predictive value Q_{FY}/Q_Y	Measured value Q_{FY}/Q_Y	Relative error (%)	Average Relative Error (%)	Computation time (s)	Genetic algorithm iteration
1	1	0.19186	0.17662	8.626	3.542	158	95
	2	0.13315	0.13091	1.712			
	3	0.15400	0.15356	0.287			
2	1	0.18442	0.17662	4.420	1.732	142	92
	2	0.13105	0.13091	0.105			
	3	0.14326	0.15356	0.671			
…							
5	1	0.16326	0.17662	7.562	4.484	161	88
	2	0.13375	0.13091	2.173			
	3	0.15927	0.15356	3.718			



(a)Whole frame (b)Machine learning sub-systems

Figure 1. Nature_inspired computation systems

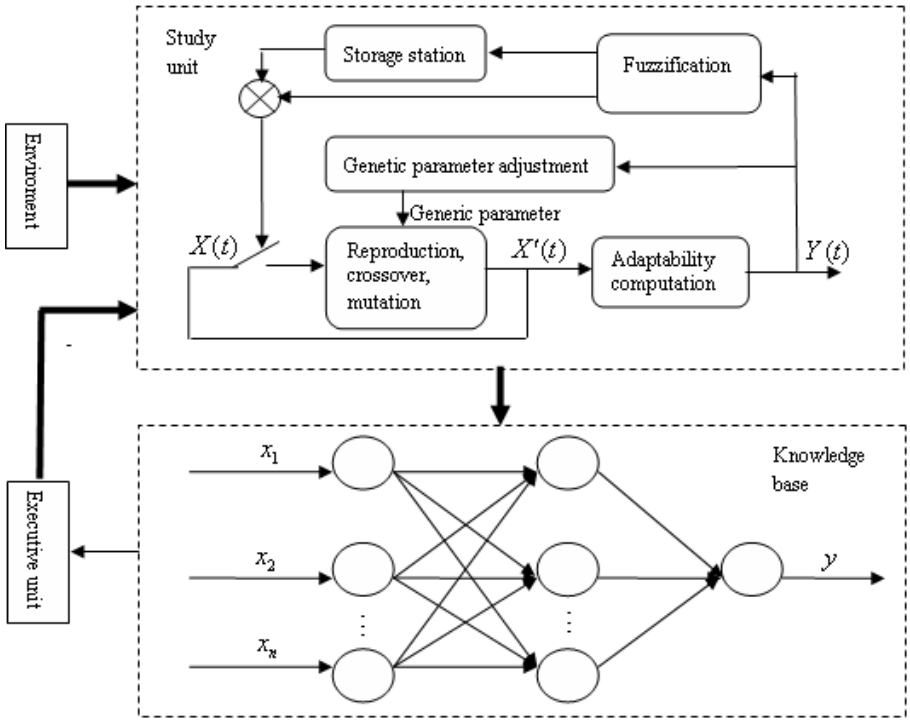


Figure 2. A particular scheme on machine learning in nature-inspired computation systems

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Fuzzy Evaluation of Weapons System

Mahmod Othman (Corresponding author)

Faculty of Science Computer and Mathematics

Universiti Teknologi Mara, Perlis Branch, 02600 Arau, Malaysia

Tel: 60-4-987-5157 E-mail: mahmod135085@perlis.uitm.edu.my

Shaiful Annuar Khalid

Faculty of Business Management

Universiti Teknologi Mara (UiTM), 02600, Arau, Perlis, Malaysia

Tel:60-12-5140436 E-mail: shaiful@perlis.uitm.edu.my

Mohammad Ismail

Faculty of Business Management

Universiti Teknologi Mara (UiTM), 02600, Arau, Perlis, Malaysia

Tel:199-544-900 Tel: mohammadismail@perlis.uitm.edu.my

Norshimah Abdul Rahman

Faculty of Business Management

Universiti Teknologi Mara (UiTM), 02600, Arau, Perlis, Malaysia

Mohamad Fadhili Yahaya

Academy of Language Studies

Universiti Teknologi MARA, 02600 Arau, Perlis, Malaysia

Tel: 60-12-455-5164 E-mail: mohdfadhili@perlis.uitm.edu.my

Abstract

This paper proposed to apply fuzzy sets and approximate reasonings to evaluate the weapons system. The objective of the study is to determine the ranking of the weapons system in a subjective environment. The proposed model based on fuzzy sets has initiated the idea of membership set score value evaluation of each criterion alternative. This enables the inclusion of requirements which are incomplete and imprecise. The approximate reasonings of the method allows the decision maker to make the best choice in accordance to human thinking and reasoning processes. The proposed model is based on fuzzy multi-criteria decision-making that consists of fuzzy rules. The use of fuzzy rules, which are extracted directly from input data in making evaluation, contributes to a better decision in selecting the best choice and is less dependent on the domain of expert. The dataset from previous research was used to validate the fuzzy evaluation model. Results from numerical examples are comparable to other fuzzy evaluation approaches.

Keywords: Fuzzy sets, Multi-criteria decision making, Approximate reasoning

1. Introduction

A reliable evaluation method which is of quality is a necessary process in decision making environments. In practice, the evaluation of performance usually uses the subjective criteria. In doing so, one has to depend on one's wisdom, experience, professional knowledge and information, which is difficult to define and/or describe accurately. When analysing using incomplete data, a lot of uncertainties will arise and this will confuse not only decision-makers but also complicate decision-making as it is made under unknown situations. The application of fuzzy sets theory in evaluation systems can improve the evaluation results (Turban, et al. 2000). Several researchers have tried to solve this problem using analytical hierarchy process (AHP), for example in personnel selection (Liang & Wang, 1992; Sonja, 2001) and shipping performance evaluation (Chou & Liang, 2001), whereby evaluation was done by aggregating all the fuzzy sets. However, the presence of imprecision, vagueness and subjectivity at each level further accumulates greatly the undesired elements in aggregating the marks.

In the literature, various concepts have been proposed focusing on the combination of fuzzy logic model with multi objective decision that can assist in reducing errors in making judgment (Pedrycz & Gomide, 1982; Liang & Wang, 1992). The research provides approaches to judgment procedure on personnel selection through the development of AHP fuzzy multi criteria. It is cited as being able to minimise subjectivity. Some research in fuzzy evaluation methods is discussed in Othman (2004a; 2004b; 2004c). The authors have proposed algorithms based either on fuzzy similarity function or fuzzy synthetic decision and ranking procedures through satisfaction function. Fuzzy sets membership enables the interpretations of linguistic variables in a very natural and plausible way to formulate and solve various problems. However, expressing the linguistic variable using the singleton fuzzy sets such as in Capaldo and Weon (2001) could result in the loss of much important information and would additionally complicate the course of action. Although many evaluation methods for selecting or ranking have been suggested in the literature, there is yet a method which can give a satisfactory solution to every situation. For this reason, a fuzzy evaluation method is proposed by combining the concepts introduced by Othman (2003) and integrating it with a fuzzy rule (Othman, 2004a) that is derived automatically from input data. This research makes its contribution by introducing the bridging and linking of these two methods. Previous studies on fuzzy evaluation methods evaluate (Tseng & Yeh, 2000; Chang & Yeh, 2002; and Kuo & Chen, 2002) the use of the number of respondents who answered the survey questions to represent fuzzy set in forms of membership function. However, these methods have a drawback, whereby they are unable to produce a generalised fuzzy evaluation method to evaluate various types of data. Hence, this research introduced the membership set score where various input data are transformed and are not predetermined by the expert. This is important to ensure the consistency in generalising the proposed framework.

The paper proceeds as follows. The proposed model is introduced in Section 2. Section 4 presents the algorithm of the proposed model and numerical results. It is followed by the concluding in Section 5.)

2. The Fuzzy Evaluation Model

The algorithm for the evaluation model consists of 10 steps as listed below:

Step 1	**:**	Calculate membership set score
Step 2	**:**	Determine grade range, mid-points and mid-intervals
Step 3	**:**	Construct fuzzy set membership for each criterion
Step 4	**:**	Define fuzzy sets for the grades
Step 5	**:**	Calculate maximum similarity value and determine grade
Step 6	**:**	Construct the similarity curve and map the grade to mid-point or mid-interval mark
Step 7	**:**	Calculate the normalised synthetic score value
Step 8	**:**	Determine multi-criteria rules combination and calculate factor rule value
Step 9	**:**	Calculate appraisal fuzzy value and the appraisal product value
Step 10	**:**	Compute satisfaction value and ranking

The performance evaluations of weapons system (WS) datasets are taken from Mon et al. (1994). The objective of this evaluation is to choose the best weapon system among a finite number of alternatives. The weapons system has three tactical missile systems (TMS) alternatives. The model started with the calculation of the membership set of score. A fuzzy number was used to evaluate the fuzzy weapons system. The data in terms of fuzzy triangle number had to be transformed into degree of membership set. Degree of membership $\mu_A(x)$ was defined as the degree of belonging fuzzy set score grade to the universe of discourse X for each criterion, as in

$f_{ij} = \{ \mu_{f_{ij}}(x) / x, x \in X \}$

Where f_{ij} = the degree of membership of fuzzy evaluation mark (i = 1, 2, 3, weapon systems and j = 1, 2, …m, the criteria environment), $\mu_{f_{ij}}(x)$ = fuzzy set of average fuzzy performance rating of 3 TMS according to the criteria given by experts in terms of fuzzy number (for example $\tilde{N}$, where N = 1, …, 9). Fuzzy numbers $\tilde{1}$, $\tilde{x}$, $\tilde{9}$ are defined as (1, 1, 3), (x – 2, x, x + 2) for x = 3, 5, 7 and (7, 9, 9). The data in terms of fuzzy numbers could be transformed into membership set score by using the following membership function $\mu_A(T)$ defined as

$$\mu_A(x) = \begin{cases} 0, & x_i \le a_1 \\ \dfrac{x_i - a_1}{T - a_1}, & a_1 < x_i \le T \\ \dfrac{a_3 - x_i}{a_3 - T}, & T < x_i \le a_3 \end{cases} \quad (1)$$

$\mu_A(x)$ describes the degree of membership of x X in fuzzy set A. The generated fuzzy set characterises the membership values $\mu_A(x)$ E [0, 1]. Table 1 depicts part of the membership set score of the first criteria. For example, to obtain fuzzy set score C_1 in Table 1, the element in the third row and in the eighth column was computed by taking the input as a fuzzy number. Let the fuzzy number be $\tilde{5}$, expressed by three parameters of the symmetric triangular fuzzy number as, (x – 2, x, x + 2) = (5 – 3, 5, 5 +2) = (2, 5, 7). The fuzzy number was transformed into the fuzzy membership set score as in Table 1 using eqn. 1. The values μ_{13}, μ_{14} and μ_{15} were computed as $\mu_{13} = \frac{4-3}{5-3} = 0.5$; $\mu_{14} = \frac{5-3}{7-5} = 1$; $\mu_{15} = \frac{7-2}{7-5} = 0.5$ (using eqn. 1). The same method was applied to calculate all the fuzzy membership set scores for each alternative evaluated by the three experts, and the results were tabulated in Table 1.

The second step determined the range of marks for each grade and then calculates the mid-point or mid-interval mark for each criterion. The standard fuzzy sets grade had to be defined so as to feed the model with knowledge. The standard fuzzy sets grade for the weapons evaluation was then defined as shown in Table 2. The values are the enhancement of the values defined as practised in (Biswas, 1995).

The construction of the fuzzy set membership was undertaken in the third step. For example, the fuzzy set of Λ_1 for C_1 can be written as {0/0, 0/10, 0/20, 0.5/30, 1/40, 0.5/50, 0/60, 0/70, 0/80, 0/90, 0/100} to represent the degree of belonging of the score to each mark. The results of the calculated fuzzy set membership were tabulated in Table 3.

In the fourth step, the fuzzy sets grade was defined in Table 4. The fifth step involved the calculation of the maximum similarity value and determination of the grade for each criterion. The normalisation operation process used the fuzzy similarity function as discussed in Biswas (1995). The grade for each criterion of the three TMS was accorded by solving the fuzzy similarity function as in eqn. 2.

$$S(F,M) = \frac{\hat{F}.\hat{M}}{max\ (\hat{F}.\hat{F}, \hat{M}.\hat{M})}, \quad (2)$$

where $\hat{F} = (\mu_F(x_1), (\mu_F(x_2), \ldots), \hat{M} = (\mu_M(x_1), (\mu_M(x_2), \ldots)$ were the vectors and $\hat{M}$ denoted the transpose vectors A^T, B^T, C^T, D^T and E^T. F represented the transpose vector of fuzzy set f_{ij} where i = 1, 2, 3 and j = 1,2 3, 4, 5. Set X = $(x_1, x_2, \ldots, x_n)$ symbolised the set of universe of discourse and '•' is the dot product. For example, if S(A, C_1) is the similarity value between grade A and criteria C_1, then the computation of $S(A, C_1)$ = ((0, 0, 0.5, 1, 0.5, 0, 0, 0, 0, 0) • (0, 0, 0, 0, 0, 0, 0, 0, 0.33, 0.67, 1)^T)/ max((0, 0, 0.5, 1, 0.5, 0, 0, 0, 0, 0)^2) • (0, 0, 0, 0, 0, 0, 0, 0, 0.33, 0.67, 1)^T) = 0.67. Next, the maximum similarity values for each factor were identified from the highest similarity values. Table 5 displays the similarity values obtained for the three TMS for criterion 1.

The sixth step was constructed based on the calculated similarity value. The similarity curve was developed for course Λ_1 from Table 5. The maximum similarity value was determined by identifying the maximum of the similarity values in Table 5. Next, the grade was mapped to the appropriate mid-interval mark. In this step, the similarity value and the similarity curve were used to map the mark. The results of allocating an appropriate mid-point and mid-interval mark to each criterion for the first course are shown in Table 6.

The normalised synthetic score was then calculated as shown in Table 7 using eqn. 3.

$$\text{Normalised synthetic score} = \frac{1}{N} r \quad (3)$$

where N = 100 and r is fuzzy mark.

Table 8 shows the fuzzy rules generated by the proposed model from the weapons system data in terms of rules properties, number of rules, maximum length, minimum length are 3, 3 and 3 respectively.

The computational results of factor rule values are shown in Table 9. For example, the value 0.4000 in row two and column two of Table 9 was obtained by using the antecedent of decision criteria DC_1, that is $C_2 \cap (C_3 \cup C_4)$ in Table 8. Therefore, when the values of C_2, C_3 and C_4 of Λ_1 from Table 7 were substituted into $C_2 \cap (C_3 \cup C_4)$ the result became $0.4000 \cap (0.8000 \cup 0.4000) = 0.4000$.

Then the appraisal fuzzy value, $(d_j(m,l))$, of Table 10 was computed as follows (Othman et al., 2004d):

$d_j(m,l) = 1 \wedge (1 - \widetilde{c}(u_m) + A_k(v_l))$, where j = 1, 2, 3, m = 1, 2, 3, l = 1, 2, ..., 11 and $\widetilde{c}(u_m)$ is the factor rule value.

The appraisal fuzzy values for decision criteria DC_1 was tabulated in Table 10.Therefore, the appraisal product value D was calculated by multiplying all elements of the appraisal fuzzy value, Dj obtained earlier. The formula is given in eqn. 5.

$$\mathrm{D} = \left(\prod_{j=1}^{3} d_j(m,l) \right) = \left(\widetilde{E}_1, \widetilde{E}_2, ..., \widetilde{E}_F, ..., \widetilde{E}_M \right) \in M_{M\times 1} \tag{5}$$

Assuming that $E_{m\alpha}$ is the α level of $\widetilde{E}_m$, $\alpha \in [0, 1] = \mathrm{I}$, it should be noted that the sets $E_{m\alpha}$ were ordinary subsets of V. For each $E_{m\alpha}$, $H_l(E_{m\alpha})$ = mid-point could be calculated. The calculated appraisal product value is shown in Table 11.

The calculated values of the range of appraisal product value (α), the different of range of appraisal product value ($\Delta\alpha_l$), and mean value of $E_{m\alpha}$, ($H_l(E_{m\alpha})$) were tabulated in Table 12.

The calculated values of the range of α, $\Delta\alpha_l$, and $H_l(E_{m\alpha})$ were substituted in the following Equation (6) to calculate the satisfaction value in the final step of the method.

$$SV(m) = \frac{1}{\alpha_{\max}} \sum_{l=1}^{11} H_l(E_{m\alpha}) \Delta\alpha_l \tag{6}$$

where α = degree of appraisal product value D; $\Delta\alpha_l = \alpha_l - \alpha_{l-1}$; $\alpha_0 = 0$; $H_l(E_{m\alpha})$ = mid-point of V_l (l = 1,2,3…,L); and $\alpha_{\max}$ = maximum degree of appraisal product value.

3. Numerical Results

The results of evaluating the ranking of weapons system were tabulated in Table 13. Columns 2, 4 and 3, 5 of Table 13 illustrate the performance value and ranking order for measuring TMS alternatives Λ_1, Λ_2, and Λ_3 respectively. The satisfaction values calculated by using the fuzzy evaluation model represent the performance values which are used to rank the TMS alternatives. The satisfaction values in column 4 were 0.7006, 0.6100, 0.0.6043 and in column 5 the rankings were 1, 2, 3 respectively. The Mon *et al*. (1994) method produced the performance values and ranking as listed in columns 2 and 3 as 0.3392, 0.3368, 0.3241 and 1, 2, 3, respectively. Clearly, it shows that the satisfaction values were higher than the values obtained from Mon *et al*.'s method. The higher value indicates that the reliable experts are satisfied with the TMS alternatives offered. From these results, the fuzzy evaluation model shows outstanding performance when compared to Mon *et al*.'s method with 100% accuracy in ranking three TMS alternatives, Λ_1, Λ_2, and Λ_3. Again the subjective evaluation method showed advantage with simpler rules properties with a smaller number of rules and maximum and minimum length.

4. Conclusion

A new fuzzy evaluation model has been proposed for the evaluation of the weapons system. The model was implemented using C++ programming language and is suitable for various fuzzy environments. Experimental results produced are comparable to results obtained from the model by Mon et al. The main contribution of the research model was the use of a fuzzy expert system consisting of a set of rules in the form of IF (antecedent) THEN (Conclusion). The model could be used as an alternative approach in solving problems that involve uncertainties. The evaluation output would become more precise if the combination factors were accurately defined. The rule properties were also analysed to judge the strength of the subjective evaluation method. The results of the experiments showed remarkable ranking performance even with the use of small-sized rule properties.

References

Biswas, R. (1995). An Application of Fuzzy Sets in Student's Evaluation, *Fuzzy Set and Systems*, 74: 187-194.

Chou, T.Y and Liang, G.S., (2001). Application of a Fuzzy Multi-criteria Decision-making Model for Shipping Company Performance Evaluation, MARIT. POL. MGMT., 28(4): p. 375 – 392.

Chu, F. (1995). Fuzzy Multicriteria Decision-Making in Distribution of Factories: An Application of Approximate Reasoning, *Fuzzy Sets and Systems*, 71: 197-205.

Hong, T.P. and Lee, C.Y. (1996). Induction of Fuzzy Rules and Membership Functions from Training Examples, *Fuzzy Sets and Systems*, 84: 33-47.

Klir, G. J. and Folger, T. A. (1988). *Fuzzy Sets, Uncertainty, and Information,* New York: Prentice-Hall International Inc.

Laarhoven and Pedrycz. (1983). A Fuzzy Extension of Saaty's Priorty Theory, *Fuzzy Sets and Systems*, 11: 229-241.

Lee, K.M., Cho, C.H., and Kwang, H.L. (1994). Ranking Fuzzy Values with Satisfaction Function, *Fuzzy Sets and Systems*, 64: 295-309.

Li, H.C. and Ilacqua, J.A. (1994). Job Search and Employment, *Fuzzy Sets and Systems*, 68: 335-342.

Liang, G. and Wang, A. (1992). Personnel Placement in a Fuzzy Environment, *Computers Operations Research*, 19(2): 107-121.

Mon, D.L., Cheng, C.H. and Lin, J.C. (1994). Evaluation Weapon System Using Fuzzy Analytical Hierarchy Process Based on Entropy Weight, *Fuzzy Sets and Systems*, 62: 27-134.

Othman, M., Ku-Mahamud, K.R. and Bakar, A.A. (2003). Quality Evaluation Using Two Phase Fuzzy Approach, *Proceeding DSI 7TH International Conference*, Shanghai, China.

Othman, M., Ku-Mahamud, K.R. and Bakar, A.A. (2004a). Fuzzy Rule Generation Using Membership Function, *Proceeding The 4th Information and Computer Engineering Postgraduate Workshop, Phuket Thailand*, p. 115 – 120.

Othman, M., Ku-Mahamud, K.R. and Bakar, A.A. (2004b). Fuzzy Similarity Function For Ranking River Basin Planning Alternatives, *Preprints Fifth International Workshop on Artificial Intelligence in Agricultural*, Cairo, Egpyt, p. 95 – 100.

Othman, M., Ku-Mahamud, K.R. and Bakar, A.A. (2004c). Ranking The River Basin Planning Alternatives Using Fuzzy Approximate Reasoning, Chiang Mai Journal of Science.

Othman, M., Ku-Mahamud, K.R. and Bakar, A.A. (2004d). Subjective Evaluation Using Fuzzy Rule Approach, Journal of Institute of Mathematics & Computer Sciences (Comp. Sc. Series), 15(1).

Pedrycz,W. and Gomide F. (1998). *An Introduction to Fuzzy Sets Analysis and Design*, England: The MIT Press.

Saaty, T.L. (1995). *The Analytic Hierarchy Process*, Pittsburgh: RWS Publications.

Sonja P. L. (2001). Personnel Selection Fuzzy Model, *International Transactions in Operational Research*, 8(1): 89-105.

Turban, E., Zhou, D. and Ma, J. (2000), A Methodology for Grades of Journals: A Fuzzy Set-Based Group Decision Support System, *Proceedings of the 33rd Hawaii International Conference on System Science*.

Turksen, I.B. (1992). Fuzzy Expert Systems for IE/OR/MS, *Fuzzy Sets and Systems*, 51:1-27.

Turksen, I.B. and Wilson, I.A. (1994). A Fuzzy Sets Preference Model for Consumer Choice, *Fuzzy Sets and Systems*, 68: 253-266.

Yamasita, T. (1997). On a Support System for Human Decision Making By the Combination of Fuzzy Reasoning and Fuzzy Structural Modeling, *Fuzzy sets and Systems*, 87: 257-263.

Table 1. Membership Set Score

TMS	C_1										
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Λ_1	0	0	0	0.5	1	0.5	0	0	0	0	0
Λ_2	0	0	0	0	0	0.5	1	0.5	0	0	0
Λ_3	0	0	1	0	0	0	0	0	0	0	0

Table 2. Grade Mid-Point and Mid-Interval Mark

Grade	**Mid-Point Mid-interval (*X*)**				
A	90.0	92.5	95.0	97.5	100.0
B	70.0	75.0	80.0	85.0	90.0
C	50.0	55.0	60.0	65.0	70.0
D	30.0	35.0	40.0	45.0	50.0
E	0.00	7.5	15.0	22.5	30.0

Table 3. Fuzzy Set Membership

TMS	C_1										
Λ_1	0/0	0/10	0/20	0.5/30	1/40	0.5/50	0/60	0/70	0/80	0/90	0/100
Λ_2	0/0	0/10	0/20	0/30	0/40	0.5/50	1/60	0.5/70	0/80	0/90	0/100
Λ_3	0/0	0/10	1/20	0/30	0/40	0/50	0/60	0/70	0/80	0/90	0/100

Table 4. Grade Fuzzy Set

Fuzzy Set											
Grade	0	10	20	30	40	50	60	70	80	90	100
E	1	0.67	0.33	0	0	0	0	0	0	0	0
D	0	0	0.5	1	0.5	0	0	0	0	0	0
C	0	0	0	0	0	0.5	1	0.5	0	0	0
B	0	0	0	0	0	0	0.5	1	0.5	0	0
A	0	0	0	0	0	0	0	0	0.33	0.67	1

Table 5. Similarity Value

TMS	C_1				
	E	**D**	**C**	**B**	**A**
Λ_1	0	0.17	0.67	0.17	0
Λ_2	0	0	0.67	1	0.11
Λ_3	0	0.67	0	0	0

Table 6. Maximum Similarity Value

TMS	**Factor**		**Max Similarity Value**	**Grade**	**Fuzzy Mark**
Λ_1	C_1		0.67	C	60
	C_2		0.67	D	40
	C_3		1.00	B	80
	C_4		1.00	D	40
	C_5		0.67	D	40
Λ_2	C_1		1.00	B	80
	C_2		0.67	C	60
	C_3		0.67	C	60
	C_4		0.67	C	60
	C_5		1.00	B	80

Λ_3	C_1		0.47	D	35
	C_2		1.00	D	40
	C_3		0.67	D	40
	C_4		0.67	D	40
	C_5		0.67	C	60

Table 7. Normalised Synthetic Score Value

TMS	**Factor**				
	C_1	C_2	C_3	C_4	C_5
Λ_1	0.6000	0. 4000	0.8000	0.4000	0.4000
Λ_2	0. 8000	0.6000	0.6000	0.6000	0.8000
Λ_3	0.4000	0.3500	0.4000	0.6000	0.6000

Table 8. Multi-criteria Rules Combination

Decision Criteria	**Factor Rule**	**Linguistic Variable**	**Description**	**Appraisal Set**
1	$C_2 \cap (C_3 \cup C_4)$	A_1	Satisfactory	v
2	$C_3 \cap (C_2 \cup C_4)$	A_1	Satisfactory	v
3	$C_4 \cap (C_2 \cup C_3)$	A_1	Satisfactory	v

Table 9. Factor Rule Value

	C_1	C_2	C_3
Λ_1	0.4000	0.4000	0.4000
Λ_2	0.6000	0.6000	0.6000
Λ_3	0.3500	0.4000	0.4000

Table 10. Appraisal Fuzzy Value for Decision Criteria DC_1

	Appraisal Set										
Λ_1	0.6000	0.7000	0.8000	0.9000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Λ_2	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000	1.0000	1.0000	1.0000	1.0000
Λ_3	0.6500	0.7500	0.8500	0.9500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 11. Appraisal Product Value

	Appraisal Set										
Λ_1	0.2160	0.3430	0.5120	0.7290	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Λ_2	0.0640	0.1250	0.2160	03430	0.5120	0.7290	1.0000	1.0000	1.0000	1.0000	1.0000
Λ_3	0.2340	0.3675	0.5440	0.7695	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 12. Calculated Range of α, $\Delta\alpha_l$, and $H_l(E_{m\alpha})$

l	Range α	$E_{m\alpha}$	$H_l(E_{m\alpha})$	$\Delta\alpha_l$
1.	0.0000 < α ≤ 0.2160	{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}	0.50	0.2160
2.	0.2160 < α ≤ 0.3430	{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}	0.55	0.1270
3.	0.3430 < α ≤ 0.5120	{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}	0.60	0.1690
4.	0.5120 < α ≤ 0.7290	{0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}	0.65	0.2170
5.	0.7290 < α ≤ 1.0000	{0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}	0.70	0.2710
6.	1.0000 < α ≤ 1.0000	{0.5, 0.6, 0.7, 0.8, 0.9, 1}	0.75	0.0000
7.	1.0000 < α ≤ 1.0000	{0.6, 0.7, 0.8, 0.9, 1}	0.80	0.0000
8.	1.0000 < α ≤ 1.0000	{0.7, 0.8, 0.9, 1}	0.85	0.0000
9.	1.0000 < α ≤ 1.0000	{0.8, 0.9, 1}	0.90	0.0000
10.	1.0000 < α ≤ 1.0000	{ 0.8, 1}	0.95	0.0000
11.	1.0000 < α ≤ 1.0000	{1}	1	0.0000

Table 13. Results of WS

Method	**Mon *et al*.**		**Subjective evaluation**	
	Performance	**Ranking**	**Performance**	**Ranking**
Λ_1	0.3368	2	0.6100	2
Λ_2	0.3392	1	0.7006	1
Λ_3	0.3241	3	0.6043	3
Acc %				100%

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Formulas of Retrieval Rate in Database Based on Normalization Design

Bin Yan

Library, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

E-mail: yanb@njupt.edu.cn

Abstract

A normalization database is considered as mathematics logical design to avoid redundancies and inconsistencies in the stored data which brought about null data and many relationship tables in the physical database. A retrieval rate is affected directly in running a practical system of the normalization design. We argue in the paper that rate formulas given out retrieval speed via testing the vast amount of data with decomposing fields and data in practical or ideal network environment. The rate formulas show that relationship tables increased must decrease retrieval speed and a minimum table must increase retrieval speed on the contrary, which compared with normal form according to the database theory and practical requirement. A speed is predictable to calculate the rate formulas. The retrieval speed is mainly concerned with server performance base on normal form too.

Keywords: Database design, Normal form, The vast amount of data, Server group, Retrieval speed, Rate formula of dynamics

1. Introduction

Traditionally, database design activities are partitioned into distinct phases in which a logical design phase precedes physical database design. The objective of the logical design step is to eliminate redundancies and updating anomalies using the notion of data dependencies, while leaving the physical design step to consider how the database schema may be restructured to provide more efficient access. Over the last decade, management system is more and more requisition for database design following functional dependencies of all sort of normal forms. The design process keeps a series of good properties which ensure their correct of functioning and limited redundancy (P. Cordero, A. Mora, I. P. de Guzman, M. Enciso, 2008, pp. 911-923). The theory of normalization has a special prominence as it proposes some properties that database must meet in order to avoid redundancies and inconsistencies in the stored data. Three new normal forms-the relaxed 3NF, replicated 3NF and relax-replicated 3NF, induced by the strong and weak functional dependencies, provide a theoretical framework for designing database schemas which are more efficient and practical, while not compromising the integrity of the underlying database. Relations in these new normal forms will not suffer from undesirable updating anomalies (Tok Wang Ling, Cheng Hian Goh, Mong Li Lee, 1996, pp.601-608).

Attribute value can be now accomplished in general database management system as non-decomposable minimum datum unit according to 1NF. A relation schema R had a set of attributes U and a set of functional dependencies F as R (U, F) of 2NF or 3NF could present more than two concrete relationship, then have many relationship tables and increase foreign keys set associate. A divided table to minimum block according to 5NF defined can limit redundancy and must observe other normal forms. Negation normal form has been emerging recently as a new direction of research for dealing with the query operation of general propositional reasoning (Tong Chun-Ling, Zuo Yi, Li Yu-chen, 2006, pp. 228-231), which decreased connecting operation, numbers of index and foreign keys. The basic notion of a relation was augmented by the concepts of functional dependencies and normal forms in an attempt to provide integrity by reducing undesirable updating anomalies (Codd, E. F, 1997, pp.33-64). A particularly undesirable form of redundancy is the presence in a relation of an attribute whose value can always be derived from other attributes and whose value is not needed to derive other attributes' values. The relation schema may attain to 3NF but not to BCNF (Li Yuchen, Zhang Linjian, Shi Bing, 1998, pp. 907-910) if it is decomposable both integrity and function dependence. Normalization and Ab-normalization are only a logic concept, which practical thing is not always to comply with. Some perfect mathematic relation schema and database behavior are decided physical system by hardware, database size and physical design, data storage and access method, optimization level and access amount of DBMS. In current time a main frequency of CPU has achieved GHz degree and transmission ability does Gbps degree but the retrieval ability does not attain it. When retrieval function is used into an information management system high frequently, design DBMS will be main effect under software and hardware. A software interfusing hardware show, that is, directly to reflect on retrieval results and not theory analysis. Many database designers had not obeyed all normal forms and only obeyed 2NF, little obeyed 3NF because of frequency query data need (Ye Zhong-jie, 2002, pp. 34-38). Sometimes purposedly reserving

partial redundancy in design may be convenient data query (Xi Yan, Pan Yu, Xu Guanghui, 2003, pp. 107-108). Sometimes software engineer lacked business knowledge to bring about redundancy in design system. We know that each time of data relation is calculated a time by a computer and the more relation tables are, the more a computer consumes resource.

2. Test methods and data collection

There is not previous paper to discuss on testing computer by the vast amount of data. Concurrent testing and computer simulation (J. Dennis Balenovich, David M. Gillott, John W. Motto Jr, 1973, pp. 227-235) was a testing method for circuit and was to verify capability of thyristor high-frequency. The moment equation (dm_k/dt) that is a state vector in the phase space and the evolution of time-dependent is nonlinearity in dynamical system (K. Sobczyk, 2001, pp. 475-498). The test purpose in our paper is to determine bottleneck caused retrieval speed down in a server with the vast amount of data which stored were derived from the library management system attained to 3NF. We can not test the vast amount of data for hardware and is only to see running hardware phenomena. Server used by us is Fujitsu GP7000F Model 400A (SPARC64-GP300MHz/8MB double CPU, 2GB), 18.2GB and 36.2GB 10000RPM disks. There is 15% CPU load and 30% physical memory holding in normal condition, but it took more than 30 seconds to retrieve one time in thousands or million of daily record data.

2.1 Testing data from database system

We chose one data table from book current system in library to migrate to another server in order to analyze the structure of data table and connection between tables. The BIBLI table we used was data of the bibliography information. There were 39 fields (columns) in the table included rec_ctrl_id, title, addition_title, volume,stock_num, publisher,publish_date, physical_descrip, unit_num, ISBN_1, ISBN_2, first_author, second_author, hird_author, other_respon_1, other_respon_2, subject_term_1, subject_term_2, subject_term_3, subject_term_4, subject_term_5, series_title, ISSN, abbre_spell_title, abbre_spell_author_1, abbre_spell_author_2, abbre_spell_author_3, abbre_other_respon_1, abbre_other_respon_2, book_search_no, call_no_1, call_no_2, item_type, orig_edit_mark, language_code, libary_code, cater_id, ISBN_X, check_control. The most of fields in the table had 164,906 rows of records.

The other test method was to analyze processing ability of a server. There were 13 fields in Check_in_out_log table included department_id, cur_date,book_barcode, reader_barcode, readertype_id, operate_code,user_id, call_no, language_code, workplace_id, rec_ctrl_id, binding_no, note and 2,969,373 rows of records transferred out from current system. The records were divided into 10% (325,575), 20% (600,900), 30% (904,380), 40% (1,201,500), 50% (1,507,894), 60% (1,801,500), 70% (2,101,100), 80% (2,401,445), 90% (2,726,965), 100% (2,969,373) and 110% (3,294,948) that 100% added 10% and 120% (3,570,273) that 100% added 20%. These parts were copied to different servers. Thus it is a more simple method to analyze ability of each server.

2.2 Equipment and its parameters

(1) Stopwatch: JD-3A model digital stopwatch, 1ms precision.

(2) UPS: USA ACP SmartCellXR 3KV.

(3) Computers: COMPAQ PROLIANT 800 (Pentium Pro, 200MHz, 256K Cathe) /64MB (two sets), SCSI 2G, 4G and 9G disks, 10/100 adapter; LONGSHAO 6060 (Pentium 3,930MHz) /128 MB, IDE 40G disk, 10/100 adapter; Fujistu Primergy MS 610 (double CPU PIII Xeon 700) /2G, IDE 18.2G, 10/100, adds ESCORT DA 3500 array disks (SCSI 73G×8, 128K Cathe, RIAD 0+1).

(4) Network: Cisco 1548, IBM 8271 and 999 9S3024B 10/100 Fast Ethernet switch.

(5) Software: WINDOWS NT 4.0, WINDOWS 2000, PowerBuilder 6.5 and MS SQL Server 7.0.

2.3 Testing method and environment

(1) Choice of Time Segment. We selected three time segments according to rush hour and low hour in network, such as AM 8:15~9:30, noon 12:00~1:15 and PM 2:15~3:30. The test kept two weeks that the first week was on the idea network condition and the second week on the practice one.

(2) Choice of Method. There were six items to test circularly and each item tested three times sequentially. The methods of different enter were used by different reader hobby to find differential connection involving refresh and reenter.

(3) Choice of Network. The network of library is sub-network in campus network and connects 105 sets of computers. The practice network can connect to Internet and the idea one can not do in order to judge out-network affection. The idea network is composed of one computer and two servers connected with hub.

(4) Choice of Retrieval Word. The retrieval word must be retrieved out 20 rows more or less in those fields in which there was not same keyword. We chose keyword “the principle of operate system” as retrieval word in title field and could retrieve out 20 rows with fuzzy retrieval method.

We chose keyword "753586" as retrieval word in book_barcode field to exist in each data part for the test of different data quantity. They were 10% (8 rows), 20% (16 rows), 30% (26 rows), 40% (36 rows), 50% (42 rows), 60% (46 rows), 70% (51 rows), 80% (55 rows), 90% (55 rows), 100% (55 rows), 110% (63 rows) and 120% (71 rows).

(5) Data Processed. We counted the average of three times for each item and data of same method every day and reserved 4-bit significance digits.

(6) Divided Data. The single table preserved primary all fields and data as a referencer. The new tables were made of the table of existed data and the table of null data in fields related with rec_ctrl_id field so called the table of divided fields. The divided data table was one divided into two from primary data with rec_ctrl_id field related. Each table was placed in different servers in order to analyze effect of server group.

(7) Test Network. The system of network was used as a client/server mode.

3. Results: analysis

3.1 Comparing to divide fields in a table

Dividing fields in a table means those fields of no data in an original table would be deleted and created new table with rec_ctrl_id field relation. There were 14 fields that had data in the original table such as rec_ctrl_id, title, publisher, publish_date, physical_descrip, first_author, second_author, series_title, book_search_no, call_no_1, item_type, orig_edit_mark, language_code, cater_id except 25 fields of no data. The test data for original table were as the standard. See Table 1:

The retrieval speed of the dividing tables was slower than the original table. But when directly retrieve for the table of data, the speed was quicker than the original table after deleting the no data fields. See Table 2:

The result was that the speed of a narrow table was quicker than a wide table. But the retrieval speed was slow if a prescribed method was to divide data into two tables related each other. The method used the narrow table and a few related tables could raise retrieval speed.

$$t_1=(1+am-bn)\cdot t_0 \quad (1)$$

In Eq. (1) it is shown retrieval time t_1 for affect factors of database can be completed successfully provided *am* and *bn,* where *a* is divided table coefficient, we set value 0.12, and *m* is number of tables related, only one data table here, *b* is deleting no data fields coefficient, we set value 0.015, *n* is number of deleting no data fields, about 25 fields deleted. The symbol t_0 is initial retrieval time, without loss no data fields and setting new tables related.

3.2 Dividing data into two tables

The data divided tables were that 164,906 rows of records in original table were divided two parts. One part had 81,693 rows of records in a table and the other part had 83,213 rows of records in other table. Two tables were related with rec_ctrl_id field. One of them was set in the same server or another server. Server group were composed of two more than servers and could use to conjugate retrieval. Test results see the Table 3.

We can see a result in Table 3 that data-divided-tables are obviously quicker than one table but data divided tables in server group are quicker than in one server. Although the processing possibility increased a time with two computers than only one computer if the computers were same, test speed was only to raise a little more than 20% under 50% data quantity. It is the best method for retrieving the vast amount of data. The retrieval time *t* should amount to various terms, which we have it practically in Eq. (2):

$$t=(t_1+t_2)/2 \quad (2)$$

$$t_2=(1-cp-dq)\cdot t_0 \quad (3)$$

In Eq. (3) it is a express formula of the retrieval time t_2 in detail, where t_2 denotes the retrieval time of dividing data in data table, therein *c* is dividing data coefficient, we obtain result 0.125 via calculating data, and *p* is number of data divided times, and *d* is server group coefficient, there is 0.029, and *q* is server number of composed of server group.

3.3 Comparing in a whole day

From test data in a whole day the table and field divided tables had a bit fluctuation but data divided tables had big fluctuation no matter one server or server group. See Table 4:

This fluctuation was not regular and was a bit improvement to steady for space after starting server.

3.4 Comparing the vast amount of data divided tables

We created new database log_ana in LONGSHAO model 6060 computer and data of log_ana were set 2,969,373 rows from Check_in_out_log data table in the current system. The first time test datum was discarded because connecting back end database caused retrieval time protraction. Test records show in Table 5 at 21°C after test data was stable with no other server connection. Counting out retrieval numbers per second sees Table 6 and average is 180,436 rows per

second.

Test reports show in Table 7 in the same condition for COMPAQ PROLIANT 800 server with additional 9G SCSI disk as slave. System construction adopted client/server pattern that LANGCHAO model 6060 was viewed as client and COMPAQ PROLIANT 800 regarded as server. Table 8 shows rows per second counted and their average was 78,726 rows per second.

The ideal network condition for test could be composed of Fujistu Primergy MS 610 server with ESCORT DA 3500 array disks and LANGCHAO model 6060 computer through connecting Cisco 8541 exchanger. The Table 9 is the test report at 21°C and the Table 10 is data to count row number per second retrieval, averages 609,840 row/s.

This purpose was to know the ability of processing the vast amount of data by computer and found inflection point of test value in order to improve most processing ability. Those data could be moved out after data of the inflection point, called the vast amount of data split. But the phenomenon did not happen from beginning to end during the test work. The results see as follow every table.

(1) Determining the effect of the vast amount of data

The synthetically processing ability of computer shows better blend technique of hardware and software. The result of any collocation will be a reflection of output equipment, showing on the screen of a monitor. There was not obvious change at retrieval speed to test different data quantity. We counted out row number per second of retrieval and again out average row number per second. There were maximum, minimum and average values for each server and the absolute value between max or min and average and their percent of average value could be seen in the Table 11.

Max and min values were lower than 10% and a server was normally processing ability. But the values were over 10% and appeared unstable of retrieval speed. It was that a computer caused effect for data processing ability and, further speaking, the computer emerged the effect of the vast amount of data. We propose that concept of the vast amount of data is to determine retrieval ability of a computer and decrease depositing data capacity in a computer for developing most ability of effectively processing data.

(2) Data quantity direct proportion retrieval speed

It can be seen in table 5, table 7 and table 9 that retrieval time was prolonged for different servers according to increase data quantity. The average number was approximate to the same value of processing data quantity by computer if different percent data quantity divided by every average time of test item made numbers per second. The numbers per second were in swing in the average numbers, and they, processing quantity per second, were linear that data quantity increased and time was consumed long. So we assume that data quantity and retrieval time are in direct proportion (see Eq. (4))

$$r=ka\cdot(1+f) \tag{4}$$

where r denotes the dynamics rate (data amount/time), therein k is rate constant associated with computer integrated process capability, i.e. Hz, disk(r/m), a is normal processing rate, f is vibrating coefficient and its value ranges from 0.15 to 0.30 that is over 10% with appeared unstable of retrieval speed.

$$k = \frac{nt_0}{\sum_{i=1}^{n} t_i} \tag{5}$$

In Eq. (5) rate constant k is placed a status in great data amount when $\sum t_i \approx t_0$, $k=1$.

(3) Server performance main effect factor for retrieval

The whole performance of server determined processing ability under client/server network pattern because data were deposited in database in server. The client computer was changed and server only changed each from COMPAQ PROLIANT 800 to LANGCHAO model 6060 to Fujistu Primergy MS 610. The test results in table 5, 7 and 9 shows the performance from better to good to best and client performance was unrelated.

(4) CPU as main performance

We noted that 30% and 70% in table 7 shows lower retrieval speed than 20% and both sides. We farther found the phenomenon was CUP occupancy resource from 30% to 60% during retrieval high percent data.

4. Conclusion

Above individual event retrieval speed tested we obtained some influence coefficients. All sorts of influence factors could become much more intricacy during computer processing data and transmission in network. The final rate we discussed should be a function of data processed and transduction $r=r(x)$and x is relevant to time. We give a rate formula:

$$r(x)=k_x a(1+f) \quad (6)$$

In Eq. (6) we know computer processed data amount unit time a is invariable and the rate coefficient k_x affects by more factors. The problem of increasing rate becomes to resolve the problem of rate coefficient k in fact.

We can give the evaluated dynamic rate formula through above analysis for test data. The fastest retrieval speed is narrow data table according to 5NF after deleting no data fields. And faster speed that is data divided tables according to 3NF in server group, but unstable phenomena could occur. The longest retrieval time is the first logging database because of needing time to connect it and retrieval time is stable once connection in success. Major factors of the effected retrieval speed are not room temperature and bandwidth of network but server performance. Server performance can be changed if some hardware such as disk is renewed, otherwise some strange phenomena could happen such as retrieving out double same rows. Data quantity is main factor to affect server performance and available divided data may relieve processing limit of server. It is an effective pathway of protective equipment investment that performance of different servers used reasonably has got server group of management systems.

These formulas are different from the results of information dynamics. One of formulas was proofed a=0.125, b=0.018 within formula (1) by my student undergraduate thesis. Our test results highlight that the server group has more processing ability to compare with a server under same data quantity. The processing ability is how to change along data quantity increase in same condition of network, equipments and temperature with the different ability and different data quantity of each server. The testing result may judge how many data quantity moved into other server in order to relieve burden of the server.

Acknowledgments

I thank Ms. Yang Jinjing and Mr. Wang Haifeng and Mr. Chen Kai, my students, of Nanjing University of Posts and Telecommunications for testing work during their undergraduate theses. The project was supported by the Scientific Research Foundation of Nanjing University of Posts and Telecommunications (NY206024).

References

Codd, E. F. (1997). Further normalization of the data base relational model. In Data Base System, R. Rustin, Ed., *Courant Computer Science Symposium*, vol. 6, Prentice-Hall, Englewood Cliffs, N. J., pp. 33-64.

J. Dennis Balenovich, David M. Gillott and John W. Motto Jr. (1973). Thyristor high-frequency ratings by concurrent testing and computer simulation, *IEEE Transactions on Industry Application*, Vol. IA.9, No. 2, pp.227-235 MARCH/APRIL.

K. Sobczyk. (2001). Information Dynamics: Premises, Challenges and Results, *Mechanical Systems and Signal Processing*, 15(3), pp.475-498.

Li, Yuchen, Zhang, Linjian and Shi, Bing. (1998). A consistency algorithm for realizing the reconfiguration of partitioned database in DDBS. *Journal of Computer Research and Development*, ISSN: 1000-1239, 35(10):907-910.

P. Cordero, A. Mora, I. P. de Guzman and M. Enciso. (2008). Non-deterministic ideal operators: An adequate tool for formalization in Data Bases. *Discrete Applied Mathematics*, ISSN: 0166-218X,156 pp. 911-923.

Tok Wang Ling, Cheng Hian Goh and Mong Li Lee. (1996). Extending classical functional dependencies for physical database design. *Information and Software Technology*, ISSN: 0950-5849, 38, pp.601-608.

Tong, Chun-Ling, Zuo, Yi and Li, Yu-chen. (2006). Normalization and Abnormalization of Relation Schema Design. *Journal of Jinan University (Sci.&Tech.)*, ISSN: 1671-3559, 20(3):228-231.

Xi, Yan, Pan, Yu and Xu, Guanghui. (2003) Design Formula of Database. *Modern Electronic Technique*, ISSN: 1004-373X, (24) pp.107-108.

Ye, Zhong-jie. (2002). Database normal formula and application. *Journal of Zhejiang vocational and Technical Institute of Transportation*, Vol. 3 No. 4, Dec.pp. 34-38.

Table 1. One server and one table comparing with fields dividing tables (second)

First day average		Second day average		Third day average		Fourth day average	
table	tables	table	tables	table	tables	table	tables
28.75	32.23	28.78	32.21	28.70	32.16	28.84	32.57

Table 2. One server and one table comparing with table of all data fields (second)

First day average		Second day average		Third day average		Fourth day average	
table	data table	table	data table	table	data table	table	data table
28.75	21.48	28.78	21.47	28.70	21.45	28.84	21.62

Table 3. One table comparing with tables in one server(1) and servers(2) (second)

First day average			Second day average			Third day average			Fourth day average		
table	tables 1	tables 2	table	tables 1	tables 2	table	tables 1	tables 2	table	tables 1	tables 2
28.75	25.25	23.23	28.78	24.80	23.40	28.70	24.94	23.15	28.84	25.50	23.97

Table 4. Data divided tables in server or servers (second)

	First day average		Second day average		Third day average		Fourth day average	
No.	server	servers	server	servers	server	servers	server	servers
1	25.92	25.84	25.66	25.86	25.59	25.93	25.80	25.73
2	25.71	25.53	25.64	26.09	25.75	25.20	25.98	25.02
3	25.71	25.50	25.69	26.00	25.60	25.43	25.94	25.06
4	25.54	25.65	25.66	25.97	25.54	25.22	25.84	25.13
5	25.68	25.20	25.62	26.07	25.65	25.79	25.72	25.01
6	25.61	25.27	25.56	26.08	25.56	25.15	25.73	25.28
7	26.33	22.16	24.40	22.11	24.41	21.98	26.02	25.03
8	26.10	22.03	24.38	22.10	24.41	22.00	25.94	24.17
9	26.32	22.06	24.57	21.95	24.49	22.05	25.65	24.28
10	26.01	22.12	24.10	22.04	24.44	21.68	25.66	24.42
11	24.53	21.90	24.29	21.96	24.56	21.82	25.70	24.33
12	24.70	22.09	24.41	22.05	24.94	21.96	25.74	24.43
13	24.33	22.11	24.38	22.16	25.01	22.06	25.71	24.03
14	24.31	22.16	24.42	22.24	24.70	22.09	25.80	24.26
15	24.32	22.17	24.44	22.30	24.58	22.05	25.67	24.24
16	24.34	22.02	24.34	22.11	24.65	22.07	25.61	24.36
17	24.42	22.28	24.48	22.07	24.58	22.07	25.70	23.99
18	24.55	22.12	24.36	22.08	24.43	22.07	25.91	24.23
average	25.25	23.23	24.80	23.40	24.94	23.15	25.78	24.61

Table 5. LANGCHAO model 6060 server test report (second)

	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
first	1.89	3.36	4.87	6.71	8.10	9.83	11.56	13.37	14.74	15.63
second	2.08	3.98	4.96	6.89	8.03	9.66	11.60	13.17	14.99	15.46
third	1.93	3.37	4.87	6.74	8.16	9.51	11.80	13.23	14.83	15.43
averag e	1.97	3.57	4.90	6.78	8.10	9.67	11.65	13.26	14.85	15.51

Table 6. LANGCHAO model 6060 server (rows per second)

data	10%	20%	30%	40%	50%
row/s	165,266	168,319	184,567	177,212	186,160
data	60%	70%	80%	90%	100%
row/s	186,298	180,352	181,104	183,634	191,449

Table 7. COMPAQ PROLIANT 800 server(9G)test report (second)

	10 %	20 %	30 %	40%	50%	60%	70%	80%	90%	100%	110%	120%
first	4.14	9.74	8.64	14.73	19.86	26.03	30.45	37.43	43.43	39.73	32.14	36.06
second	4.14	9.76	8.54	14.86	19.86	25.95	30.39	37.24	43.43	39.49	32.26	35.99
third	4.14	9.79	8.53	14.84	19.81	26.01	30.37	37.40	43.43	39.48	32.29	36.04
average	4.14	9.76	8.57	14.81	19.84	26.00	30.40	37.36	43.43	39.57	32.23	36.03

Table 8. COMPAQ PROLIANT 800 (9G) server (rows per second)

data	10%	20%	30%	40%	50%	60%
row/s	78,641	61,568	105,529	81,128	76,003	69,288
data	70%	80%	90%	100%	110%	120%
row/s	69,115	64,279	62,790	75,041	102,232	99,092

Table 9. Fujistu Primergy MS 610 server test report (second)

	10%	20%	30%	40%	50 %	60%	70%	80%	90%	100%
first	0.58	1.08	1.60	2.04	2.46	2.85	3.28	3.78	4.12	4.54
second	0.66	1.13	1.55	2.05	2.43	2.89	3.24	3.76	4.07	4.53
third	0.53	1.10	1.59	1.96	2.49	2.89	3.35	3.74	4.07	4.51
average	0.59	1.11	1.58	2.02	2.46	2.88	3.29	3.76	4.09	4.53

Table 10. Fujistu Primergy MS 610 server (rows per second)

data	10%	20%	30%	40%	50%
row/s	551,822	541,351	572,392	594,802	612,965
data	60%	70%	80%	90%	100%
row/s	625,521	638,632	638,682	666,740	655,491

Table 11. Average row number per second to retrieve and fluctuation ratio

server	LANGCHAO	COMPAQ	Fujistu
average	180,436	78,726	609,840
max	6%	34%	9%
min	8%	22%	11%

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Effectiveness of Corporate Intranet in Selected Malaysian Companies

Mohamad Noorman Masrek

Faculty of Information Management, MARA University of Technology

Puncak Perdana Campus, Section U10, 40150 Shah Alam Selangor, Malaysia

Tel: 60-3-7962-2002 E-mail: mnoorman@salam.uitm.edu.my

Kamaruzaman Jusoff (Corresponding author)

Faculty of Forestry, Universiti Putra Malaysia 43400 UPM Serdang Selangor, Malaysia

Tel: 60-3-8946-7176 E-mail: kamaruz@forr.upm.edu.my

Nor Shariza Abdul Karim

Faculty of Information and Communication Technology, International Islamic University

Jalan Gombak, 53100 Kuala Lumpur, Malaysia

Tel: 60-3-6196-5681 E-mail: shariza@iiu.edu.my

Ramlah Hussein

Faculty of Information and Communication Technology, International Islamic University

Jalan Gombak 53100 Kuala Lumpur, Malaysia

Tel: 60-3-6196-5627 E-mail: ramlah@iiu.edu.my

The research is financed by Universiti Teknologi MARA

Abstract

This study was undertaken with the purpose of measuring intranet effectiveness in selected Malaysian companies. Defining effectiveness as composite constructs of intranet characteristics, intranet usage and individual performance, the study employed a cross-sectional survey method. A total of 359 users participated in the study. The intranet characteristics namely, task fit, information quality and systems quality are found to be significantly related to intranet usage. Likewise the three intranet variables, namely, usage mode, usage for decision support and usage for knowledge sharing are also found to be significantly related to individual performance measured in terms of task innovation, task productivity and personal sense of accomplishment. Hence, the study had indeed confirmed that intranets users in Malaysian companies have perceived their intranet to be effective

Keywords: Intranet, Corporate, Usage, Characteristics, Individual, Performance

1. Introduction

Since its first inception a decade ago, the intranet has achieved major advancement and sophistication. At present, intranet technologies have significantly mature and they exist in all sizes, shapes, and forms. In fact, more sophisticated terms like intranet portal, enterprise portal, enterprise information portal or EIP (Shilakes & Tylman, 1998) have been coined to reflect the advancement and complexity of the technology. Concurrently, EIP vendors also started flourishing in the market. Nevertheless, as intranets are being equipped in most Malaysian corporate workplace environment, studies measuring their effectiveness from the user-level perspective are still very limited. Against this background, this study was undertaken with the purpose gauging intranet effectiveness. Adapting the work from Delone & Mclean (1992, 2003), intranet effectiveness is defined as composite constructs of intranet characteristics, usage behavior and individual performance. Within the domain of IS, various theories and models have been developed to explain technology

adoption and measuring its effectiveness. Accordingly, the model of this study is conceptualized based on the work of Trice & Treacy (1988), Delone & Mclean (1992), Goodhue & Thompson (1995), Ali & Money (2005) and Jeyaraj *et al*.(2006). Most of these models and frameworks were developed based Theory of Reasoned Actions or TRA (Fishbein & Ajzen, 1975) and Theory of Planned Behavior or TPB (Ajzen, 1991). In essence, both TRA and TPB posited that individual behavior is determined by his or her belief and attitudes towards a particular object. Models such as Information Systems Success Model or ISSM (Delone & Mclean, 1992) and Task Technology Fit or TTF (Goodhue & Thompson, 1995) posited that technology usage is a predictor of individual impact or performance. Various studies adopting these two models have demonstrated that individual performances can be greatly enhanced through technology usage (*e.g.* D'Ambra & Wilson, 2004; Iivari, 2005). Hence, the conceptualized framework, defined as the intranet effectiveness model from the user perspective was used in this study.

2. Methods

Several corporate enterprises were contacted to participate in the study. However, in order to be eligible, the intranet being used should be of high maturity i.e. being integrated with organizational information systems. Consequently, four corporate enterprises were found to fulfill the criteria and therefore chosen to participate. After lengthy discussion with representatives of these enterprises, it was decided that the respondents of the study should be the executives in the headquarters only. The rationale being that majority of the staffs ranging from the executives to top management were located here and furthermore, staffs working here directly involved in the core-business and core-functioning such as human resource, marketing, finance and service operation the organization. Hence, in the course of such activities the use of ICT such as intranet would be very rampant. A stratified random sampling technique was adopted so as to insure that respondents were well represented by various departments. Accordingly, 700 questionnaires were distributed to these participating companies, upon which 423 were returned but 359 were usable.

Other than demographic information, the questionnaire contained open-ended questions with seven point Likert scale. To measure task-fit, four-item scales adapted from Ali & Money (2005) were used. To gauge information quality and systems quality, six-item scales and ten-item scales adapted from Ahn, Ryu & Han (2004) and Roca *et. al*.(2006) were used. To measure decision support, four-item scales adapted from Doll & Torkzadeh (1998) was developed. To measure knowledge sharing, an instrument comprising of six-item scales adapted from De Vries *et. al.* (2006) was developed. To measure individual performance which consists of task innovation, task productivity and personal sense of accomplishment, ten-item scales adapted from Torkzadeh & Doll (1999) and Staples *et. al*. (2002) was used. Prior to actual data collection, the developed questionnaire was pilot tested involving 40 usable responses and reliability analysis conducted on the research instruments unveiled that it was soundly reliable.

3. Results and discussion

Based on 359 usable responses, data were analyzed using SPSS version 14.0. Non-response biases were checked by comparing early responders and late responders using independent sample t-test. Apparently the results showed that the responses were free from non-response biases. ANOVA test were also performed across companies to check whether significance differences could be observed on all the research variables. Evidently, no significant difference could be observed and following the results, responses from the four organizations were treated as one sample in subsequent analysis. Considering that this study was exploratory, factor analysis involving principal axis factoring with Varimax rotation was then executed on multi-items measures. In interpreting which factor loadings are worth considering, this study adopted loadings of 0.3, as suggested by Hair *et al.* (1998). All the measures were entered into principle axis factoring with Varimax rotation.

The exercise unveiled that despite few items that experienced cross-loadings, most items measuring intranet characteristics and individual performance variables were loaded into conceptualized variables. Due to the huge difference and the relatively higher factor loadings between factors, the cross-loaded items were retained on the factors representing the conceptualized variables. However, one item measuring knowledge sharing had to be eliminated as it had cross-loaded with other factor and the difference of the factor loading is very small. Hair *et al.* (1998) asserted that cross-loading reflects the high shared variance between variables and cautioned that when correlation among factors becomes higher; it becomes more cumbersome to differentiate which item load uniquely on the factor. In addition, two items measuring usage modes i.e. publishing and recording did not meet the cut off mark and thus removed from further analysis. Possible explanation for the low loadings of these two items could be attributed to the confounding effect of other two usage modes i.e. interacting and transacting. The practices of information publishing, especially memos are usually embedded together within interacting mode. In other words, users publish these types of information and consequently disseminate them via e-mail. Likewise, the practice of recording organizational knowledge could be perceived as similar to transacting mode. Recording of knowledge usually takes place via some form of computerized systems such as Knowledge Management Systems (KMS).Following the outcome of the factor analysis, reliability analyses were conducted on the variables employing multi-items measures. Accordingly, items measuring usage behavior had to be reduced into three as two items were found not to meet the cut-off value. Upon further investigation

on reliability analysis discovered that Cronbach's alpha for information quality, systems quality, usage mode, decision support, knowledge sharing, and individual performance were 0.903, 0.942, 0.899, 0.858, 0.910, and 0.952 respectively. Clearly, these values suggest that the instrument used in the study were highly reliable.

3.1 Demographics

Table 1 presents the demographic profiles of the research samples. Between male and female, the former seemed too outnumbered the later with 54.9% as opposed to 45.1%. Age group between 31 and 35 was most dominant and contributed to 29.5% of the sample. In terms of qualifications, 284 respondents indicated to have gotten first degree while 23 indicated to have obtained Masters. 306 respondents indicated as holding executives posts while 53 were holding middle management post. The average length of service was 7.62 while intranet experience recorded a mean of 6.92.

<<Table 1. Demographic Profiles>>

3.2 Relationship between variables

In order to test the formulated hypotheses, correlation and regression analyses were performed. Correlation entails the provision of a yardstick whereby the intensity of strengths of a relationship can be measured (Bryman & Cramer, 2001). However, correlation analysis gauges only the degree to which two variables are related or tend to move together but there is no assumption that one is causing or affecting the other (Alreck & Settle, 1995). Hence, to measure the degree and direction of influence of the independent variable on the dependent variable, the regression analysis was also applied in this study. The result of exercise is shown at Table 2. Evidently, resulting from the correlation analysis, the values of the Pearson r between all independent and dependent variables are well above 0.5, suggesting moderate correlation but significant at 0.05 level. In addition, the results of the regression analysis between independent variables and dependent variables unveiled that the values of R^2 are well above 0.30 indicating that the former are influential in explaining 30% or more variation in the latter. To this effect, it can be safely concluded that all of the formulated hypothesis are fully supported. Hence, the findings of this study further support previous studies by Phelps & Mok (1999), Tang (2000), Young (2000), Miller (2004), Wilkie (2005), Deltour (2005) and Baptista *et al.* (2006). The findings also confirmed TRA, TPB, ISSM and TPCM in the context of intranet usage.

<<Table 2>> Results of the correlation and regression analysis

3.3 The Relationship between Intranet characteristics and intranet usage

Theories such as TRA posited that object-based beliefs do have an effect on usage behavior. Model such as ISSM also posited similarly. In the context of this study, the three intranet characteristics namely, task-fit, information quality and systems quality were equivalent to object-based beliefs in TRA and ISSM. In congruent to these theories and models, all the three intranet characteristics were found to be strong determinants of intranet usage variables. In ensuring successful intranet usage among users, the intranet itself should inherit those characteristics that match users' job requirements. This situation which is also known as task fit is very critical as failing to accommodate users' need and requirements would further burden or cause more difficulties to users. Intranet functionalities' need and requirement varies across users. To respond to this diversity of needs and requirements, some intranet portals are equipped with the customization or personalization tool which enables users to personally customize their intranet functionalities according to their own preferences. The closer the intranet to fulfilling users' needs and expectations, the more they would be utilized by users. The crux of implementing information systems is to effectively manage information. In this information age, information is no longer treated as a byproducts of business transactions, rather, it is now considered as critical capital to organization well being. Even at the user level perspective, information usage especially among white collar jobs such as in this study, is so critical that without them almost no job can be carried out. Respondents of this study had evidently showed the importance of information quality in determining usage behavior. When the intranet matched users' information quality requirements, it would results in users utilizing the intranets.

In relation to information quality, systems quality also receives equal attention in terms of its importance from the users' perspective. The analogy between information quality and systems quality would be like the term 'effectiveness' and 'efficiency'. If the intranet could produce quality information as expected by users, then it deserves to be called effective. However, if the time taken to produce quality information is very slow because of its poor systems quality, then it is likely to be called inefficient. Apparently, both information quality and systems quality goes hand-in-hand. Therefore, addressing the systems quality of the intranet would guarantee usage among intranet users as evident by the findings of this study. The implication of the abovementioned finding is that it further strengthens TRA, TPB, ISSM and TPCM and suggests their applicability and relevancy in the context of intranet settings.

3.4 The Relationship between Intranet Usage and Individual Performance

Models such as ISSM and TPCM postulate that IS usage should lead to individual performance or effectiveness. Evidently, the findings fully support this and that intranet usage measured in terms of usage modes, usage for decision

support and usage for knowledge sharing were found to be predictors of individual performance. As users engaged in the intranet in varying usage modes such as transacting, interacting and searching, various outcomes could be expected. As a result of transacting with various corporate information systems via the intranet, various corporate information could be retrieved in the form of report summaries, etc. This information would certainly be very useful input to the generation of new ideas or innovations. Transacting with automated information system would also lead to time saving, and hence resulting in better task productivity. Also, as work tasks are automated or computerized via the intranet, jobs become less laborious and would eventually lead towards better job satisfaction, higher work commitment and greater personal sense of accomplishment.

In terms of interacting modes, users could engage in diverse and intense communication among corporate members. By means of interacting, users could query or participate in various online discussions and consultations that would eventually develop new ideas and innovations. Similarly, by means of interacting, the information needs of users, either for job-related or personal usage, would be better served. Essentially, by means of interacting, users can always seek the help of others in the company. As such, help can be acquired either for job purposes or personal matter and would eventually translate towards better work productivity and enhanced personal sense of accomplishment. The searching mode of intranet usage enables users to access and retrieve diverse corporate information and knowledge. As a result of understanding corporate processes and procedures through the help of information and knowledge via means of searching from the internet, users could devise or develop new innovative ideas that are beneficial to the corporate well-being. Also, when corporate information searching and access becomes unproblematic but efficient and effective through the intranet, work tasks can be easily completed, hence boosting productivity. By means of the searching mode, the individuals could easily fulfill their diverse informational needs that would further relate to personal sense of accomplishment and self-satisfaction. In the course of completing work tasks, users would certainly deal with crucial decision making or problem solving. In certain circumstances, decision making or problem solving could not he handled by oneself and hence suggestions or inputs from others are seen critical. By means of utilizing the intranet, users could either search for critical information required for the decision making or problem solving, or alternatively they could seek the help from corporate members via means of the intranet. In the process, various ideas could be germinated and once the problems are solved and decisions have been made, new knowledge (i.e. the solution or decision) are created. In due course, users would definitely feel relieved and satisfied as their mission of problem solving have been accomplished.

Individuals vary in knowledge, skills and competencies. Via means of knowledge sharing among peers and colleagues or even among subordinates and superiors, one's skills and knowledge could be enriched and deepened. In an corporate context, the intranet has always been considered an effective tool in materializing knowledge sharing. By sharing knowledge with others, new knowledge would be created. Also, by sharing knowledge with those who are in dire need of that knowledge, this may even help them in improving work productivity, especially when the knowledge relates task or job performance. Similarly, when the required knowledge is obtained, for instance by means of the intranet, that individual would be surely pleased as his or her needs are accomplished. In sum, the findings have clearly shown that in an intranet computing environment, intranet usages do have significant relationship with individual performance.

4. Conclusion

Findings of the study have indeed ascertained that intranet users have perceived their intranet to effective evident from the supported hypotheses. Nonetheless, the major contribution of this study can be assessed from two perspectives i.e. theoretical and practical. Intranet, despite its widespread adoption among public and private organizations, has captured little interest of the IS researchers to investigate the nature of usage and the corresponding determinants and impacts especially at the user level. Consequently, the empirical based framework that depicts the determinants and impacts of intranet usage at the user-level perspective is unavailable until the establishment of the framework used in this study. The framework, which mainly derived from the fundamental Input-Process-Output model and other supporting theories such as TRA, TPB, ISSM and TPCM has been successfully validated in the context of this study. Indirectly, it also implies that the findings of this study have strengthened the theories, frameworks and models upon which the model has been built upon.

Future research should consider adopting every intranet user irrespective of their job level as respondents. As such, differences and comparisons can be made between job-level and status of intranet usage. Secondly, the perceptual measures adopted in the survey instrument are subject to individual interpretation and understanding. Hence, instead of using self-reported measures for gauging individual intranet usage, a more accurate approach would be to install software-tracking systems onto the intranet that would both monitor and record individual usage. However, such approached would be quite difficult unless permission and access are granted to the organization's intranet. It would be more interesting if the measures could capture the type of knowledge being shared via the usage of intranet.

References

Ahn, T., Ryu, S. & Han, I. (2004). The impact of the online and offline features on the user acceptance of internet

shopping malls. *Electronic Commerce Research and Applications,* 3, 405–420.

Ajzen, I. (1991). The theory of planned behavior. *Organization Behavior and Human Decision Processes*, 50(2), 179–211.

Ali, A.S.B., & Money, W.H. (2005). A study of project management systems acceptance. *Proceedings of the 38th Hawaii International Conference of Systems Science (HICSS2005), Hawaii, USA.*

Alreck, P.L., & Settle, R.B. (1995). *The survey research handbook*. Second Edition. Chicago: Irwin.

Bryman, A., & Cramer, D. (2001). *Quantitative data analysis with SPSS release 10 for Windows (also suitable for SPSS release 11): a guide for social scientist*. New York: Routledge.

D'Ambra, J., & Wilson, C. S. (2004). Explaining perceived performance of world wide web: uncertainty and the task technology-fit model. *Internet Research,* 14(4), 294–310.

De Vries, R. E., Den Hooff, B. V. & De Ridder, J. A. (2006). Explaining knowledge sharing: the role of team communication styles, job satisfactions and communication beliefs. *Communications Research*, 33 (2), 115 – 135.

Delone, W.H. & Mclean, E.R. (2002). The Delone and Mclean model of information systems success: a ten-year review. *Journal of Management Information Systems*, 19(4), 9–30.

Delone, W.H., & Mclean, E.R. (1992). Information systems success: the quest for the dependent variable. *Information Systems Research*, 3(1), 60–95.

Deltour, F. (2005). Satisfaction, acceptance, impacts: linking three modalities to evaluate intranets at individual level. *Proceedings of the 12th European Conference on Information Technology Evaluation (ECITE2005), Turku, Finland.*

Fishbein, N., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: an introduction to theory and research*. Reading, MA: Addison-Wesley.

Goodhue, D.L., & Thompson, R.L. (1995). Task-technology fit and individual performance. *MIS Quarterly,* 19(2), 12 –236.

Hair, J.F., Anderson, R.E., Tatham, R.L. & Black, W.C. (1998). *Multivariate data analysis.* New Jersey: Prentice-Hall.

Iivari, J. (2005). An empirical test of the Delone-Mclean model of information systems success. *The DATABASE for Advances in Information Systems*, 36(2), 8–27.

Jeyaraj, A., Rottman, J.W. & Lacity, M.C. (2006). A review of predictors, linkages and biases in IT innovation adoption research, *Journal of Information Technology*, 21, 1–23.

Miller, R. (2004). Measuring intranet effectiveness. *Prescient Digital Media*. Retrieved August 18, 2005 From http://www.prescientdigital.com/Prescient_Research/Articles/Intranet_Articles/Measuring_Intranet_Effectiveness.htm

Phelps, R., & Mok, M. (1999). Managing the risks of intranet implementation: an empirical study of user satisfaction. *Journal of Information Technology*, 14, 39–52.

Roca, J.C., Chiu, C.M. & Martinez, F.J. (2006). Understanding e-learning continuance intention: an extension of the Technology Acceptance Model. *International Journal of Human Computer Studies*, 64(8), 683–696.

Shilakes, C., & Tylman, J. (1998). *Enterprise information portals*. New York: Merrill Lynch, Inc.

Staples, D.S., Wong, I., & Seddon, P.B. (2002). Having expectations of information systems benefits that matched received benefits: does it really matter? *Information and Management*, 40(2), 115–131.

Tang, S.M. (2000). An impact factor model of intranet adoption: an exploratory and empirical research. *The Journal of Systems and Software*, 51, 157–173.

Torkzadeh, G., & Doll, W.J. (1999). The development of a tool for measuring the perceived impact of information technology on work. *Omega*, 27 (7), 327–339.

Trice, A.W. & Treacy, M.E. (1988). Usage as dependent variable in MIS research. *Database*, 33–41.

Wilkie, K.J. (2005). An investigation on how intranets supports user task. *Unpublished masters dissertation. The University of North Carolina at Chapel Hill, USA.*

Young, L.Y. (2000). Factors affecting user satisfaction on Intranet. *Unpublished Master's thesis, National Sun Yat-Sen University*, *Taiwan.*

Table 1. Demographic Profiles

	Gender	Freq	%
Gender	Male	197	54.9
	Female	162	45.1
Age	20 – 25	33	9.2
	26 – 30	91	25.3
	31 – 35	106	29.5
	36 – 40	85	23.7
	41 – 45	33	9.2
	46 – 50	6	1.7
	> 50	5	1.4
Qualifications	Master	23	6.4
	Degree	284	79.1
	Diploma	28	7.8
	Others	24	6.7
Job Level	Executive	306	85.2
	Mid. Mgt.	53	14.8

Table 2. Results of the correlation and regression analysis

Independent variable	Dependent variable	r/β	R Square	Adjusted R^2	F (1, 357)	t	p value	Hypothesis Results
Task fit	Usage mode	0.575*	0.331	0.329	176.537	13.287	0.000	H1a: *Supported*
Information quality	Usage mode	0.567*	0.321	0.319	168.762	12.991	0.000	H2a: *Supported*
System quality	Usage mode	0.579*	0.336	0.334	180.374	13.420	0.000	H3a: *Supported*
Task fit	Decision support	0.571*	0.326	0.324	172.310	13.127	0.000	H1b: *Supported*
Information quality	Decision support	0.576*	0.332	0.330	177.159	13.310	0.000	H2b: *Supported*
System quality	Decision support	0.536*	0.288	0.286	144.268	12.101	0.000	H3b: *Supported*
Task fit	Knowledge sharing	0.599*	0.358	0.357	199.398	14.211	0.000	H1c: *Supported*
Information quality	Knowledge sharing	0.620*	0.385	0.383	223.408	14.947	0.000	H2c: *Supported*
System quality	Knowledge sharing	0.569*	0.324	0.322	171.151	13.082	0.000	H3c: *Supported*
Usage mode	Individual performance	0.638*	0.407	0.405	244.709	15.643	0.000	H4: *Supported*
Decision support	Individual performance	0.571*	0.326	0.324	172.341	13.128	0.000	H5: *Supported*
Knowledge sharing	Individual performance	0.635*	0.403	0.402	241.482	15.540	0.000	H6: *Supported*

** Correlation is significant at 0.01 level (2-tailed)*

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Analysis of Network Security and Risks Prevention Strategies of Hongqiao Airport West Terminal

Chongjun Fan (Corresponding author)
Business School, University of Shanghai for Science and Technology
Shanghai 200093, China
E-mail: cjfan@sh163.net

Honglin Xiong
Business School, University of Shanghai for Science and Technology
Shanghai 200093, China
E-mail: dianyi0467@yahoo.com.cn

Haiying Zhang
Shanghai Airport Construction Headquarters
Shanghai 201202, China

Wujun Liu
Shanghai Airport Construction Headquarters
Shanghai 201202, China

Liqiang Wang
Shanghai Airport Construction Headquarters
Shanghai 201202, China

Xiaohui Song
Shanghai Airport Construction Headquarters
Shanghai 201202, China

The research is financed by the research project of Science and Technology Commission of Shanghai Municipality of China. No. 06211212051 (Sponsoring information)

Abstract

The sharing and openness of network are convenient to the airport operations and passenger information inquiry, but also result in information security issues. In the design of Hongqiao Airport West Terminal network system, information security issues have been taken as important topics, and preventive strategies which aim at the specific circumstances have been formulated by comprehensive analyzing the causes of potential security hazard of airport network system security. This project can provide reference and help to related Chinese airport's planning and construction projects.

Keywords: Airport information system, Network security, Risk prevention

1. Introduction

With the accelerated pace of life, convenient flights information is a prerequisite for people on easy travel, which requires the airport network system to access internet stably and to provide open, timely, relevant and reliable information. Airport network system is the communications infrastructure for the whole airport information system and it supports the business operations on internet by connecting a wide range of interfaces to the various related system.

Based on the principle of "unified planning, phased construction", Hongqiao Airport West Terminal network system has used leading mature and reliable technology to built Hongqiao wet terminal main network, west terminal wireless network, West Terminal POS network, the West Terminal security network system, west terminal internal communication system, which provide the 24-hour continuously reliable and safty data and media operating platform for information system. This platform is required to support the operation of Hongqiao Airport and to supply timely, accurate, systematic and comprehensive flight information services to the passengers, airlines, Hongqiao integrated transport public information platform and the business management.

In order to optimize manufacturing operations and equipment running, the platform is also demanded to help each airport departments to conduct multi-airport, multi-terminal scheduling management under the command of AOC/TOC (Airport Operation Center, Terminal Operation Center) agreement.

Hongqiao Airport is an important aspect in the Hongqiao comprehensive transportation hub projects. And these projects extend to outer ring in the east, Huaxiang Road in the west, Beizai Road in the north, Luqingping Road in the south. The completed airport will be 5.8 times as big as present scale. Such a large project can not be separated from the complex network system and the network system is the priority among priorities, because a security and reliable network is the core link of the whole projects.

2. The Planning and Designing Of Hongqiao West Terminal Network

2.1 the main planning and designing of West terminal network

2.1.1 the main planning of West terminal network

As main network, the joint room of West terminal network AOC, TOC is the core node for the entire airport's network the MDF-1 and MDF-2 are the main aggregation nodes, the MDF-3 and MDF-4 are node cluster when the West Gallery is enlarged in the long term. Core nodes and aggregation node are connected by using stars model. Set aside the network interface with the East Terminal. In the future West Terminal AOC will connect with the Pudong AOC remotly in order to make the two airports share information and resources for effective allocation and scheduling. It is shown in Figure 1.

(1) Pudong Airport connection. Between Hongqiao Airport AOC and Pudong Airport AOC a dedicated network interface is established which is optical fiber-based links and the telecommunications backup line is reserved. The network interface can provide two integrated systems, flight tracking system and within system which were interoperability required.

(2) Interface with Hongqiao integrated transport hub public information platform. Through the airport security isolation equipment the network share information platform interface with hub, and use the public flight information server to provide dynamic information-sharing platform for flight information.

2.1.2 The design of the core Main-net of West terminal

The core Main-net of Western Terminal is shown in Figure 2. The Western Terminal's AOC/TOC joint machine room act as the data center and main-net center, the Main Distribution Frame MDF-1, MDF-2 in gallery Area A and Area B is the Grade Ⅰ network node of CL in this building; the BHS machine room is the Grade Ⅱ network node of CL in this building; the Eastern Traffic Hub MDF, the outfield managing center (OMC)MDF, the aviation business administration building MDF is the Grade Ⅱ network node of CL outside this building; while the correlative IDF is the Access Layer. All the core layer, the Grade Ⅰ network node and part of the Grade Ⅱ network node need bi-core allocation(multilink-bind among exchangers). Among core network node exchangers and from network node of CL exchangers to core network node exchangers, bi-link design is used.

The Main-net links to Hongqiao integrated transport public information platform externally, also links to Pudong Airport and Eastern Terminal internally.

Four security-outer-connection is built among four areas: the Western Terminal's AOC/TOC joint machine room, the Main Distribution Frame of gallery Area B MDF-1, the Main Distribution Frame of gallery Area C MDF-2 and the Main Distribution Frame of Eastern Traffic Hub MDF-TIC, to safeguard a reliable out connection locally.

2.2 Planning and design other sub-network

Commercial POS system can run on the platform. The public commercial areas of west terminal can be covered with network. Inside, it connects with wireless LAN interfaces; outside, it connects with bank, customs and Internet

interface.

2.2.1 The security network of the west terminal

It mainly includes hand-baggage and checked -baggage screening systems. The security network of west terminal is designed with two-tier architecture. The core layer is located in security room and the access layer is scattered in the office of votes and the IDF which near the security zone.

2.2.2 The Commercial POS system of the west terminal

It is designed with three-tier architecture, the core layer of the network is located in AOC/TOC joint equipment room, the aggregation layer of the network is located in MDF-1, MDF-2 and MDF-TIC and the access layer of the network dispersed between the relevant IDF.

2.2.3 The wireless LAN of the west terminal

In view of the needs of a variety of wireless applications, the wireless LAN is designed in the corresponding floor of the West Terminal, and in according with the needs of BRS and under the permission of air traffic control nearly 45 seats are covered directional. The visitors in Terminal area and VIP lounge can enjoy wireless Internet and information Inquiry Service by using wireless terminal.

Computer Room A: the computer room of CAAC and China Civil Aviation Information Network Co., Ltd

Computer Room B: the computer room for security inspection information system

Computer Room C: the computer room for baggage management information system

3. Analysis of network security and risks prevention

3.1 Analysis the causes of airport network system security risks

First of all, standing the point of the Hongqiao Airport network infrastructure, we have found that many core components of network infrastructure (such as the main network and server subsystems, etc) ,related technology and application software (for example, large-scale database applications)depend on foreign manufacturers to provide; at same time, some of the airport network security engineers and managers can hardly understand the potential risks, lack necessary technical facilities and related processing experience, so facing increasingly serious network security, they often don't know how to do and what to do. It is just because of technical limitations, many people only know how to guard against viruses but lack overall awareness of network security,.

Secondly, when the airport is under constructed, its network hardware facilities and application software in a certain period of time is a relatively safe. However, a variety of hardware and application software when they have been made out can't be tested out all the flaws or loopholes. With the rapidly development of network technology and application software, some drawbacks and vulnerabilities of hardware and software will be exposed gradually. Because hardware and software lag behind when technology is making progress, the network security risks are inevitable.

Thirdly, some man-made factors also threaten with the airport. Man-made network intrusion makes the maintenance of the airport network security to be very difficult, How to deal with all kinds of cyber crime is a good example. Some hackers are taking network system to paralyze as their goal; some lawless are developing those malicious program (such as deleting or changing, even theft the information which is related to airport), these forces can not be disdained.

Finally, there are unconventional threats with network security, such as natural disasters or terrorist attacks, these are unexpected threatens.

Therefore, we need a comprehensive understanding of the airport network threats, taking effective measures to prevent and combat these threats in time. The rapid development of aviation industry asks for more stringent, conscientious and careful airport network security than before.

3.2 Different perspectives of the network safety precautions

3.2.1 Based on TCP / IP protocol security

TCP / IP protocols are a four-tier architecture, which includes the application layer, transport layer, Internet layer and network interface layer. But essentially, TCP / IP has only three layers, namely, application layer, transport layer, Internet layer, because the bottom layer of the network interface has no specific content (Xie Xiren, 2003). However, in according with the specific network layout of the airport, according to the characteristics of the construction of the airport network and TCP / IP architecture, airport network can be divided into the corresponding network application layer, system layer, network layer.

(1). Application layer security. As directly providing service to Application process, application-layer has many protocols in the Internet, such as the World Wide Web's HTTP protocol , supporting for SMTP e-mail protocol, the FTP file transfer to support the agreement, ect (Xie Xiren, 2003). In accordance with these agreements' function, we can

prevent the possible loopholes by use of software methods (see **3.2.3**) to prevent the virus from the erosion and illegal users' visiting; In addition, establishing data backup systems, which make the network can rapidly recover when network has failed to work, and minimize data loss.

(2). System Layer security. Frequent exchange of visits exists in Airport network (including the exchange of visits emerges between Hongqiao Airport and Pudong Airport), a wide range of Resource sharing make the virus more easily transmitted. To protect security of Layer system, first of all , the security of servers need to be protected. Each server has its dedicated software and a dedicated service, in order to prevent the server from attacking, the best way is to close unnecessary services and ports which leave no opportunity to hackers (Qin Ying, 2007). Secondly, checking the backup server's system log regularly, recording the users' operation, analysing log files, reforming the hidden danger of system in time and nipping in the bud. In addition, study show that 80% of network attacks and worm attacks come from internal LAN, the viruses and Trojans have the characteristics of self-replication, so those infected computers will often affect other LAN users. Also now there are a lot of rebound Trojans, which can take the initiative to connect the client and publish harmful information. So in the implementation of security software, firewall software and precaution Trojan horse software should use coordinately. The practice shows that this method will protect your computer from viruses effectively, Trojan horses and other malicious code threats, as well as the invasion of illegal users.

(3). Network layer security. The main function of Network layer is to provide communications for different hosts, and its detailed functions will not be repeated here. During sending data packet, segment network layer composition is packaged into groups or packages by network layer, so is user data. In general, the network layer security can be considered from the following aspects: One approach is to use IP packaging technology, the nature of the package is that plain text is encrypted (a kind of security in 3.2.3 Software-Based Security), encapsulated in the outer layer of the IP report, when it used to encrypt the packets routing on the Internet to reach the other side, the outer IP header has been opened, that is to say, the messages be decrypted (William Stallings, 2002); Another method is to make reasonable use of VLAN division, the VLAN is unrelated to position, making the airport network divide into virtual VLAN network segments, which can suppress broadcast storm on the network effectively; The third way is IP address binding, IP address management is an important part of the network construct in LAN network of the airport. The final paragraph into the address corresponds to the VLAN, at the same time , the IP address of network nodes, and MAC address of equipment are bound. The IP address management methods have greatly improved the safety rules and the availability of traffic control rules, reducing the LAN IP attacks possibility, and improving network security and reliability.

IPSec security measures is also very important to protect the network security, who offers a Clear structure, considering its security, identity validated, integrality, secure secret keys changing, and protection measures (So-Hee Park, Jae-Hoon Nah, and Kyo-Il Chung, 2005). IPSec is widely used in organizations or enterprises to build VPN(Virtual Private Network). IPSec in the IP layer achieves its security services and protects directly all data packets in order to achieve security purposes (Deng Lei, Ai Jisong, 2009).

3.2.2 hardware-Based Security

From the aspect of hardware to address the airport network security, in the design of the airport network system we conduct from the following aspects.

(1). The location selection of the AOC/TOC core computer room need to be relative covert and secure (Of course, geographical location should also be reasonable, in order to network calling), which is basic preconditions of the airport network system's safety. This requires the location selection of the AOC/TOC core computer room to have a certain security and construction safety. On the one hand, it can prevent the interference or destruction of the lawless elements. On the other hand, it can withstand some natural disasters. In addition, we should choose stable capability and reliable quality brands in purchasing the AOC/TOC core network to promote the safety factor of the airport network system.

(2). In order to ensure all the main airport networks and subnets to run 24 hours all day, the airport network servers (especially the AOT / TOC core network server) implement a Hot-Standby hot backup mode (Wu Nian zu, 1999) that is, in the normal situation of the server, a backup server is in a waiting state that once the server on work breaks down, the backup server wakes up immediately, automatically replaces broken-down server to work, and recovers the normal operation of the whole network system, in order to achieve continuous communications of the whole network, then it can meet the airport's requirement of 24-hour continuous work..

(3). Taking hardware firewall to protect the airport network system effectively. It can provide a physical and logical isolation for the external and internal network; what's more, it can be multi-layered defense. In addition, a hardware firewall (especially high-end firewall) can also provide better remote management capabilities except the defense of the internal network security threats. Hongqiao Airport and Pudong Airport will eventually implement the sharing of the information and the unified configuration, which requires interconnection in the network system. Constructing in this remote LAN structure, hardware firewall is essential.

(4). For the high level data security, physical isolation is a good choice (all types of hardware firewall is not absolutely

safe. For example, a firewall can not detect encrypted, and the capability of prevention Web applications is insufficient).That is to say, the relative and connected LAN with high security level should not be connected to the Internet, in order to prevent lawless from using the loopholes to access to these data.

3.2.3 Software-Based Security

Carrying out hardware security precautions, software security should be done as well. Combining current technology, the solution is as follows:

(1). Installing anti-virus software. Anti-virus software is an essential part of network security procedures. Before Hongqiao Airport AOC and Pudong Airport AOC connect with each other, it's inevitable that there are software vulnerabilities that haven't been checked out during software testing(Because many systems that are used don't came from the same software company, when information is delivered between different systems , "seamless" is emerged, what's worse, it is a step in the virus), so a safety hazard have broke out. Correct configuration and enforcement anti-virus software can reduce the harmful of networks malicious programs. The administrator should examine virus alarm log regularly, and learn the whole situation of network virus timely, accurately, and remote scanning the threat of computer. In this way, we can prevent the spread of the virus in the local area network effectively.

(2). Network information encryption. Taking into account the safety of network information, encryption has been implemented for data transmission and information storage. The purpose of encryption is to protect network data, documents, passwords and control information, and online transmission of data. After the encryption of systems information, it will protect the information that can not be malicious acquisition and be made use of during transmission and storage.Information encryption coordinate with firewall technology which is used to enhance information systems and information security and confidentiality is one of the key technologies. When the Hongqiao Airport project have been complete, the data of Hongqiao Airport AOC and Pudong Airport AOC will have backed up each other. Switches is the core of two LAN, VTP protocol attack is a threat to the security of switch. VTP domain has been encrypted, so someone haven't know the password or password is incorrect, switch will not be able to learn the news of VLAN. This can effectively prevent VTP protocol attacker entering each of the other users, who are in the same VLAN.

(3). Access Control. Access control is to identify legitimate users own the authority of the system resources, in order to prevent the illegal invasion and legitimate users using non-permission resources, which is the core of application-layer security architecture (Zhang Yun, Kong Fang, 2005). At present, there are many solutions of access control, such as MAC address filtering, VLAN isolation, access control list which bases on the IP address, etc. Each computer system which is belong to the airport network has user access control, as soon as file access permission rights are set. File access control restrictions will effectively carry out legitimate user access to acts within its competence, which maximize the sharing of resources. Implementation of access control information can safeguard the integrity of the information; reduce the HIV infection the opportunity; slow down the speed of the spread of infection; protect the confidentiality of important information and so on.

(4). Intrusion detection. Check whether there is breached strategy behavior or attacked signs by collecting and analyzing computer network or some important information of computer system. The combination of detection software and hardware of intrusion is called Intrusion Detection System(IDS).IDS plays a warning role in network security. Hongqiao Airport West Terminal will provid wireless Internet area, which makes passengers surf the Internet conveniently. However, such opened network also make network security system be in risk, At this moment, it's time for IDS show its function. Whether visitors have illegal intention can be analyzed, thus the corresponding information are feed backed to system control center which judge the visitors are legitimate users or not, ultimate, determine to give access rights if the visitors are legitimate, otherwise refuse being visited.

In a word, to ensure network system security no matter from the TCP/IP protocol architecture perspective or from analysis of hardware and software perspective, they are carried out interlaced when they defense network security. Therefore, all aspects of technical factors are the inevitable choice of implementation countermeasure (as illustrated in Figure 3).

Some special attention should be paid that man factor is the premise of all factors. We have advanced technology and equipment to maintain airport network security hardly enough, reasonable management system, administrator's operation habit, and the safety awareness of network security are the indispensable supplement.

4. Conclusions

Airport network is a quite complex system engineering, with the improvement of technology, ensure airport network security is a dynamic and perfection process. In order to solve this problem, first of all, we must think about from different angles and multiple levels, adopt diverse security measures; secondly, both the openness and security of network should be considered, understanding the potential threats in the network, analyzing network structure deeply in any times, making out suitable and reasonable management programs on the base of actual situations, carrying on protection network from the whole aspect, that's to say, we should defense airport network security with the means

which will have coordinated all aspects of the technological superiority, only in this way can make network in the maximum safety and ensure airport network system security, stable and efficient operation.

The construction of Hongqiao Airport West Terminal network system have been first-class in the domestic at present, studying and summing up this experiments have the role of inspiration and setting a model to domestic airport construction in the future.

5. Acknowledgements

This research was supported by the research project of Science and Technology Commission of Shanghai Municipality of China (No. 06211212051). Expresses the thanks.

References

Deng, Lei & Ai, Jisong. (2009). A Study of Automatic Generation of IPSec Security Policies. *Computer Security*, 2009(1): 4-8.

Qin, Ying. (2007). Airport Research on Design and Application Technology of Airport Network Security. *Jiangsu Aviation*, 3(2): 32-33.

So-Hee Park, Jae-Hoon Nah, and Kyo-Il Chung. (2005). The Implementation of IPsec-Based Internet Security System in IPv4/IPv6 Network. LNAI 3397: 109-116.

William Stallings. (2002). Foundation Course of Network security. Beijing: Tsinghua Press.

Wu, Nian zu. (1999). The Construction of Pudong International Airport—Information System. Shanghai: Publishing House of Shanghai Science and Technology.

Xie, Xiren. (2003). Computer Network(The Furth Edition). Beijing: Publish House Of Electronics Industry.

Zhang, Yun, Kong, Fang. (2005). Design and Research of Application Layer Security Architecture. *Microcomputer Development*, 15(3): 16-17.

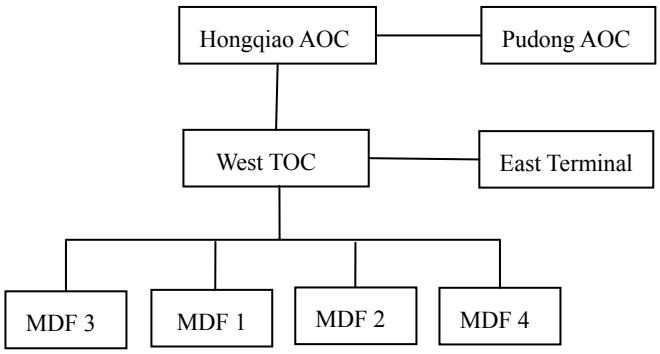


Figure 1. main network plans

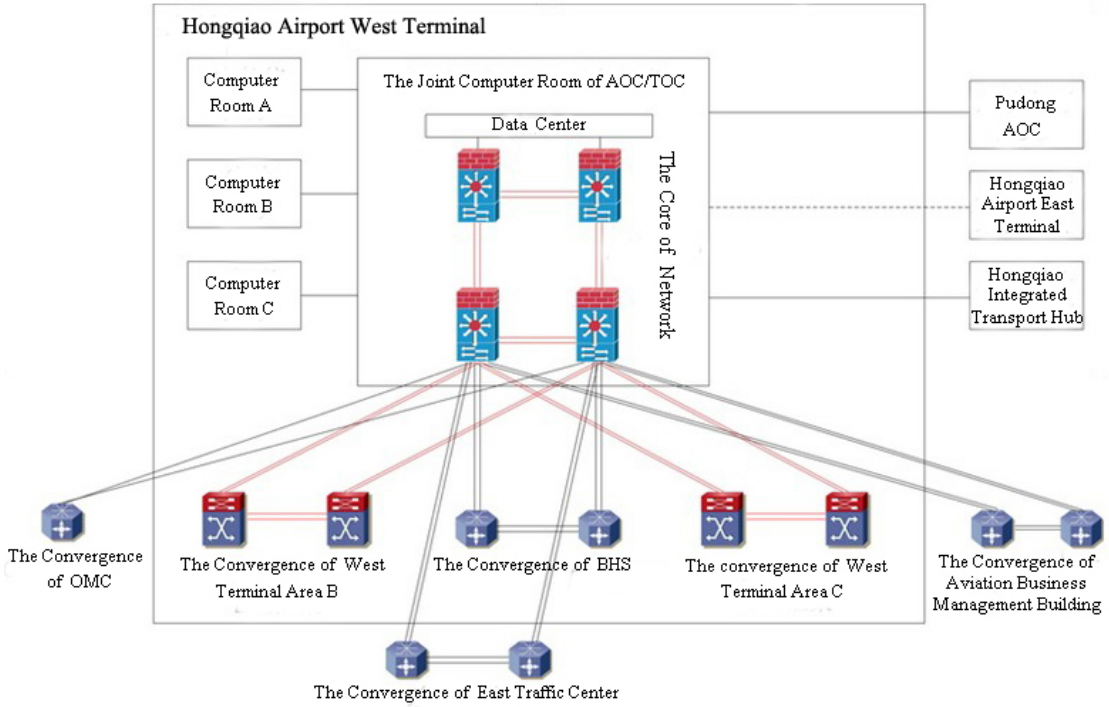


Figure 2. Hongqiao west terminal core network of main network

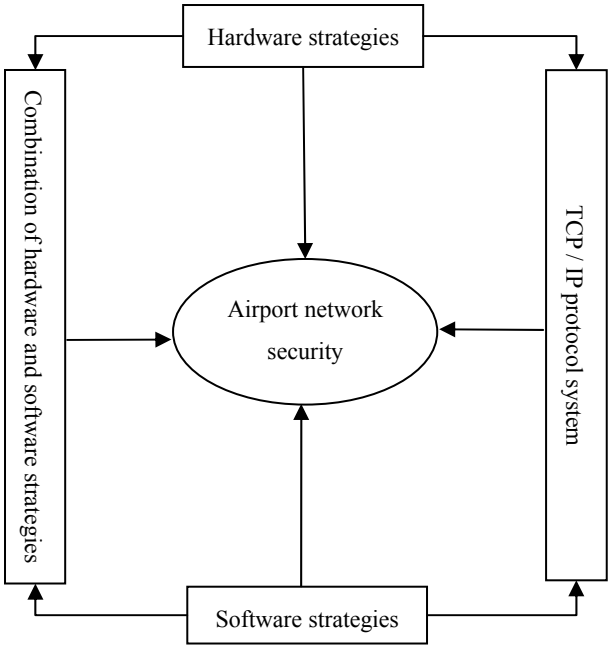


Figure 3. Coordination of the strategy to maintain the airport network security

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 2, No. 3
August 2009

Paper Prototyping as a Rapid Participatory Design Technique

Aznoora Osman
Faculty of Computer Science and Mathematics, Universiti Teknologi MARA Perlis
02600 Arau, Malaysia
Tel: 60-4-988-2811 E-mail: aznoora@gmail.com

Hanif Baharin
Faculty of Computer Science and Mathematics
Universiti Teknologi MARA Perlis, 02600 Arau, Malaysia
E-mail: ahmadhanif@perlis.uitm.edu.my

Mohammad Hafiz Ismail
Faculty of Computer Science and Mathematics
Universiti Teknologi MARA Perlis, 02600 Arau, Malaysia
Tel: 60-4-988-2811 E-mail: mohammadhafiz@perlis.uitm.edu.my

Kamaruzaman Jusoff (Corresponding author)
Faculty of Forestry, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
Tel: 60-3-8946-7176 E-mail: kamaruz@aeroscan.com.my

Abstract
This paper describes participatory activities with university lecturers to design an online community. The objective of this study is to engage the users of an online community to collaboratively design their online community. We speculated that by involving them in the design team, we can identify their specific requirements, and they will accept and use the system. However, lecturers have heavy workload and tight schedule. For that reason, we thought that paper prototyping is the most suitable tool to be used because it is fast and easy to create. Therefore, paper prototyping technique has been adapted in a two-day participatory design session. We found that paper prototyping is indeed the most suitable technique to elicit requirements from the end users under a time constraint. Moreover, participants came out with unexpected requirements and novel interface.
Keywords: Participatory design, Paper prototyping, Requirements elicitation, User centered design

1. Introduction

University lecturers need to communicate with each other in their very own online community. Online community comprises of many standardized features such as forums, announcement, advertisement column and poll. However, for a group of lecturers, they need specific features that might not be needed by other user groups. Thus, online community for lecturers should be designed to fulfill their requirements. They must participate in the design process because it is possible to get a deep knowledge of their work context and needs (Dix, Finlay, Abowd & Beale, 2004, p.198). This can be achieved through a participatory design session. Nevertheless, the lecturers are busy with their teaching schedule. So, there should be a method to gather them and engage them within a short time frame. That is why we chose paper prototyping technique in the participatory design session because it is cheaper, faster and easier to elicit feedback from the users (Snyder, 2003).

1.1 Participatory design

Participatory design allows end users to become part of a design team as well as to test the usability of a system. Therefore, when they are actively involved in the design process, the system requirements and system design can be

refined iteratively. In this session, the end users will sit together with the developers and designers to discuss their requirements. Dix et al. (2004, p.466) stated that participatory design aims to improve the work environment and task by the introduction of the design. It also enforces collaboration between the end users and the developers. The approach is iterative in which the design is subject to evaluation and revision until an optimal design has been produced. There are many methods to conduct participatory design session such as brainstorming, storyboarding and paper prototyping. Butler & Fitzgerald (1997) discussed that participatory design activities have indeed contributed to the development of systems that adequately captured user requirements and hence satisfied user informational needs. A study conducted by Denef, Ramirez, Dyrks, Schwartz and Al-Akkad (2008) on the other hand shown that participatory design can be used to evaluate multimodal application by inviting participants to join participatory design workshops.

1.2 Paper prototyping

Paper prototyping is a method to brainstorm, design, create, test and communicate user interfaces (Snyder, 2003). It uses only papers, pencils, sticky notes, scissors and highlighters to create a mock-up of the system design. This technique requires less budget and time to develop as compared to computer prototyping. Since it is easy and fast to create, developers are more willing to accept changes to the design because there has been no effort in writing the program code yet. Paper prototyping technique can be employed to design the interface of any software, websites and devices. A study conducted by Bailey et al. (2008) revealed that paper prototyping can be used to develop effective interfaces for Multi Display Environment (MDEs). However, they claimed that designers must employ methods that allow them to rapidly generate and test alternative designs early in the design process. The technique also needs to be adapted to effectively simulate the use of multiple displays and allow testing with groups of users. In another context, paper prototyping has successfully helped in structuring the feature of the interface for children's collaborative handheld application (Black et. al, 2004).

1.3 Usability testing

Usability testing is a process of evaluating the degree to which a product or system meets specific usability criteria. The evaluation should be made by the participants who are representative of the target population (Rubin, 1994). In usability testing, the participants work on typical tasks using the system (or the prototype), and the evaluators use the results to see how the user interface supports the users to do their tasks (Zaphris & Kurniawan, 2007). A study on conducting usability testing of a website using paper prototyping technique has proven that it required low cost and time, generated critical feedback from the users and lead to improved usability (Grady, 2000). Sousa & Furtado (2005) discussed user participation throughout requirements elicitation, requirements validation and prototyping process of an interactive application. The process contributes to user satisfaction and the design of usable prototypes. Participatory design has also been used successfully to design and evaluate mobile phones as described by Massimi (2007). In order to conduct a usability test using paper prototype, the team members need to take a role-play for four different roles. The first role is the user. The person who takes this role must know the needs of the real users. The user will interact with the prototype to accomplish a set of tasks. Another person becomes the "Computer" who will simulate the system behavior. A team member with more experience in usability should become the facilitator to conduct the session. The rest of the team should become note-taking observers (Snyder, 2003).

The objective of this study is to engage users in a participatory design session to design an online community using paper prototyping technique. The study is carried out to elicit requirements from the end users by inviting them to be part of the design team. By involving the end users in the design process, the end-product will be accepted and used.

2. Methodology

Four participatory design sessions were conducted in two days. Eight participants were selected among Computer Science lecturers. They were divided into two groups whereby the researchers acted as the facilitator for each group. In each group, a participant became the user, another participant became the "Computer" and the other two participants became the observers. Table 1 summarizes the participatory design sessions.

2.1 Session 1

Participants were briefed about the overall objectives of the sessions and the goals to be accomplished at the end of the first day. Then they were introduced to the paper prototyping technique. Five aspects were emphasized during the introductory briefing:

The history of paper prototyping – The participants were brief about the history of paper prototyping to show them the importance and the relevance of the prototyping technique in the industry and how it relates to participatory design. The participants were also reminded that they were a part of the design team, and not merely end-users that answer questions and give feedbacks.

The workshop is a hands-on process - The participants were told that they were learning paper prototyping techniques as they develop their first paper prototype, and the researches acts as facilitators to guide them through the process.

There is no right or wrong way to develop a paper prototype and the participants were free to explore their creative side with the guidance from the facilitators.

Roles in the workshop – The two groups worked separately in developing the prototypes, which mean that at the end of the sessions, two different prototypes were made. All the members in each group collaboratively developed the prototypes. But in doing the usability test, each team members was assigned a role, either as Computer, Observer or Facilitator. Two members from a group acted as the users for the other group and vice versa.
Materials used in paper prototyping and paper prototypes sample – The participants were briefed about the stationeries and materials used to develop paper prototypes. This was followed by showing the participants samples of paper prototypes. By examining the samples the participants quickly learned how to manipulate the materials to create interface widgets such as drop-down menu, radio buttons, or how to make windows and internet browsers.
The benefits of paper prototyping techniques - The benefits and the positive aspects of paper prototyping were highlighted throughout the briefing, to convince skeptics and to encourage the participants to give their full commitments.
In the session, participants were reminded that they were part of the design team, and not merely end-users that answer questions and give feedback. The task design is an adaptation of Snyder's. In the first step, the concept of user goals was explained to the participants. They were told that they were the users and that they were developing an online community for Computer Science lecturers. Thus, as the users, they know their specific needs. They were asked to think about the things that they do frequently and the things that were important. Next, they were asked to list a set of questions regarding the functionality, navigation and terminologies to be used in the prototype. In the last step, the participants wrote their tasks in the template provided. They were shown several examples of tasks. The facilitators were involved with the participants in the task design process. All user goals and questions came from the participants and they worked hand-in-hand with the facilitator to translate the user goals and questions into tasks. The facilitators made sure that the instruction for the tasks did not reveal the way to navigate the website. Session 1 took about four hours with half an hour break between introductory briefing and task design.

2.2 Session 2

Once the tasks have been designed and written in the task sheets, the groups started developing the prototype with the assistance of the facilitators. This is the moment where they could explore their creative side to design an online community. Before ending the session, a walkthrough was done. Walkthrough is like a rehearsal of usability test, in which any problems of the prototype can be detected and corrections can be made. During the walkthrough, the facilitators acted as the user in a usability test. As a result, several mistakes were found in the prototype such as incomplete interface and missing links. Corrections were made and the participants practiced and get ready for the usability test on the second day. The session wrapped up in two hours.

2.3 Session 3

This session took three hours to complete. Two members from each group acted as users for the other group in a usability test. The facilitators observed the participants and helped them to identify usability issues encountered during the test. The results of the usability test are discussed in the results and discussion section. After the usability test, the participants rejoined their group and made refinement to their prototypes according to the findings of the usability test.

2.4 Session 4

Once the participants finished refining the prototypes, a second walkthrough was done. The aim was to document the prototype navigation. One of the team members read the instruction from the task sheet and another member used the prototype to complete the task. Using a video camera, the facilitator recorded the walkthrough session, focusing on the prototype interface. The video recording will be an important reference for the researchers in using the prototype for the next iteration of the design process.

3. Results and discussion

The participatory design session and usability test were able to elicit user requirements that are unique for a group of lecturers. It has helped in determining the best navigation technique for the online community as well as proper terminologies to be used. Nevertheless, it was observed that doing participatory design with busy people has its weaknesses as well.

3.1 Requirements elicitation

The most significant outcome was when a participant revealed unexpected requirements, which was equipment booking. The equipment that can be booked are printers, digital cameras and laptops. The idea came from the head of the Computer Science department who has been handling the equipment booking process manually when any lecturers come to her office. The process takes away some of her precious administrative time since she has to record every item that has been booked, taken and returned. The record will keep track of the availability of equipment. So, by having equipment booking facility in the online community, she hopes that it will ease her burden and that she can use her time for more productive tasks.
Furthermore, a participant suggested having a function to update research information. At the moment, research information can be updated in the departmental website. However, it is done solely by the webmaster. So, whenever a

lecturer wants to update or add research information, he or she has to submit it to the webmaster. This is a waste of time for both parties. Therefore, the participant wishes that the online community will provide a channel for every lecturer of the department to update research information instantly. The necessary information includes research title, team leader, team member, grant awarded and its amount, duration of research and research synopsis.

There were few other requirements disclosed by the participants. The first group suggested a function for files uploading and downloading. They believed that it is important to share any type of files among the lecturers. Another function was links to useful websites. They were mainly links to electronic journals, electronic library, electronic newspapers and university websites. Besides that, the second group suggested a function to post advertisement and announcement.

3.2 Novel interface

The participants came out with a novel interface for the equipment booking function. They suggested creating the booking form in a table where information of a particular item can be viewed in a single row. The information includes availability, the person who is currently using it and the date it will become available. Hence, by looking at the table, the user will not only know the availability of the item, but he or she can also book the item directly without navigating to a different booking page.

3.3 Terminologies

The terminologies used should be more descriptive to reflect its function. For example, Resource Sharing did not represent uploading and downloading of files. As a result, it was changed to Files Uploading and Downloading. The term Advertisement was inappropriate because there were two options under it which were Classified and Announcement. The users were confused when given a task to advertise used car. Therefore, the keyword was changed to e-notice board and the two options under it were removed.

3.4 Navigation

The navigational issue also came into concern. The first group did not employ the 'breadcrumb trails' in their interface. It is a navigation technique that helps users to keep track of their current location within a program or website (Dix et al., 2004, p.764). For a website, it allows users to follow back to the starting point or home. It was discovered that users prefer to navigate using breadcrumb trails. Thus, it was added onto every interface.

3.5 Weaknesses of the study

There are few issues that need to be addressed when conducting a participatory design session with busy people. Since the end users were divided into two groups, we assumed that there might be some pressures to compete with each other. We observed that there was a tendency to build a perfect or nicely drawn prototype. This took them a longer time, hence making them feel bored in the end. Towards the end of the session, there was lack of enthusiasm among few participants even though they enjoyed using papers and pencils in the beginning. Therefore, we believe that in order to conduct an effective participatory design session using paper prototyping, the participants must be assured that it is an alternative method to computer prototyping. On top of that, they must also be convinced that paper prototyping is able to refine design ideas as effective as computer prototyping. This will help to minimize or even eliminate skepticism among them.

4. Conclusion

The participatory design using paper prototyping technique has helped in eliciting requirements from the end users of an online community for Computer Science lecturers. It is indeed the most effective technique to reveal user requirements within a short time frame. The users who acted as part of the design team were more willing to make changes to their design because paper prototyping is easy and fast to create. Since it requires no skill in computer programming, it is possible to conduct the same session to develop an online community for lecturers from non-technical departments as well.

References

Bailey, B.P., Biehl, J.T., Cook, D.J. & Metcalf, H.E. (2008). Adapting Paper Prototpying for Designing User Interfaces for Multiple Display Environments. *Personal and Ubiquitous Computing*, 12, 269-277.

Black, J.T., Hawkes, L.W., Jean-Pierre, K., Johnson, I., II & Lee, M. (2004, September 13-16). A Paper Prototype Study of the Interface for a Children's Collaborative Handheld Learning Application. *6th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 305-309). Glasgow, Scotland.

Butler, T. & Fitzgerald, B. (1997, December 14-17). A Case Study of User Participation in the Information Systems Development Process. *Proceedings of the Eighteenth International Conference on Information Systems* (pp. 411-426). Atlanta, Georgia.

Denef, S., Ramirez, L., Dyrks, T. & Al-Akkad, A. (2008, October 20-22). Participatory Design Workshops To Evaluate Multimodal Applications. *Proceedings of the NordiCHI2008* (pp.459-462). Lund, Sweden.

Dix, A., Finlay, J., Abowd, G.D., & Beale, R. (2004). *Human Computer Interaction* . Essex: Pearson Education Limited.

Grady, H.M. (2000, September 24-27). Website Design: A Case Study in Usability Testing Using Paper Prototypes. *Paper presented at the 18th Annual ACM International Conference on Computer Documentation: Technology & Teamwork* (pp. 39-45). Massachusets, USA.

Massimi, M., Baecker, R.M. & Wu, M. (2007, October 15-17). Using Participatory Activities with Seniors to Critique, Build and Evaluate Mobile Phones. *Paper presented at the 9th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2007* (pp. 155-162). Arizona, USA.

Rubin, J. (1994). *Handbook of Usability Testing: How To Plan, Design and Conduct Effective Tests.* John Wiley & Sons. Retrieved February 01, 2008, from Books24x7 database.

Snyder, C. (2003). *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces.* San Francisco: Morgan Kauffman Publishers. Retrieved January 01, 2008, from Books24x7 database.

Sousa, K. & Furtado, E. (2005, September 26-27). From Usability Tasks To Usable User Interfaces. *Proceedings of the TAMODIA 05* (pp. 103-110). Gdansk, Poland.

Zaphiris, P. & Kurniawan, S. (2007). *Human Computer Interaction Research in Web Design and Evaluation.* Pennsylvania: IGI Publishing. Retrieved January 25, 2008, from Books24x7 database.

Table 1. Sessions in the Participatory Design

	Session 1	Session 2
First Day	Introduction to paper prototype Introduction to task design Designing tasks	Paper prototype development Walkthrough 1
	Session 3	Session 4
Second Day	Usability test Prototype refinement	Walkthrough 2 (Video documentation)

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

The Design and Implementation of the Distributed Computing Platform for Bioinformatics

Juan Huang

School of Information Science and Engineering, Central South University, Changsha 410075, China

Tel: 86-731-265-5236 E-mail: dealhelen@yahoo.com.cn

Yaoping Fei

Network Center, Central South University, Changsha 410075, China

Abstract

The processing of the huge amounts of information in the bioinformatics has been the bottleneck to restrict its development, and in this article, we used the distributed computation to solve this problem, and we described the structure, the design, the implementation of task decomposition, the distributed application program and the database management of the distributed computing platform. The distributed computing platform first decomposes one problem into many subtasks, and then the client sends the request of task computation for the server end, and the server end responses the request and takes out the information of the minimum subtask, and distribute the information to the client end. When the client end acquires the information of the subtask, it will transfer the operation module to compute, and when the task is completed, it will upload the result to the server end, and the server end repeats above process until all subtasks are completed, and according to the computation results of all subtasks, we can obtain the solution of the problem.

Keywords: Bioinformatics, Distributed computation, Task decomposition, RMI

1. Introduction

Along with the development of the human genome project in the world, huge amounts of genome information and the data about the structure of nucleic acid and protein issued present exponential growth every year, and the biological experiment data enter into the situation of “big bang”. In the data processing and analysis process about the huge amounts of information for bioinformatics, many processes such as the split joint of gene sequence, the sequence homology comparison, the establishment of the molecule cladogram, the structured information analysis and the forecast of the gene function all require the computation environment with higher performance (Shi, 2001, P.161-165). The distributed computation researches how to divide the problem which can only be solved by the super computer into many small parts, and distributing these parts to many computers, and integrating these computation results to obtain the result of the problem finally. The advantages that we adopt the distributed computation include that it can acquire same even better computation ability comparing with the super computer, and its expenses are much smaller than the expenses of the super computer.

2. Total structure and work flow of the distributed computing platform

The distributed computing platform needs to decompose the problem into many small tasks, and distribute these small tasks to the free computers in the network for computation, and manage these subtasks at the same time.

The bioinformatics distributed computing platform can realize the computation of the problem which needs the computer with high performance in the bioinformatics by the distributed mode in the Internet, and the solution of the problem about the bioinformatics is decomposed into the solutions of many subtasks. The platform is composed by three parts including the database, the server and the client end. The database stores relative information about the task, users and operation. The server realizes the task decomposition, the task management and distribution, the client management and the result processing. The client end realizes the task data download, the task computation and the task upload. The system structure of the platform is seen in Figure 1.

The personnel who participate in the distributed computation include the project mangers and clients, and the flows that the personnel participate in the distributed computation are different.

(1) The flow that the project mangers participate in the distributed computation.

First, describe the biological problem by the task tree, decompose the task into many subtasks and write the task information into the database.

Second, start the service of the server end, login the remote method for the users' transfer, and manage the data in the database.

(2) The flow that the clients participate in the distributed computation.

First, user interviews the website, enter into the login unit, input the information needed by the login, and complete the login procedure.

Second, download and install the application program in the download unit of the website.

Third, fill in user information in the client end, click "save", and the save user information at home, and the client end will automatically complete many operations such as the requesting the server, downloading task data, transferring the operation module and updating the task result.

3. Implementation of the task decomposition

The problem of the bioinformatics: give the set of DNA, $S = \{s_1, s_2, s_3, s_4 \ldots s_n\}$, and the length of each string is m, and find the string closing to other strings most (Ma, 2000, P.99-107).

If we adopt the distributed computation to solve this problem, we can decompose the set S into many small subsets. Every string has two states which are respectively denoted as 0 and 1, and the binary tree is used to store all combinations. In the original set S, the first layer is divided into the left child and the right child which are used to denote s_1, and the left child denotes the set S_10 has no the string s_1, and the right child denotes the set S_11 has the string s_1, and the second layer is used to denote s_2. In this way, the task tree with n layers can be generated, and the sub-trees which don't accord with the conditions will be deleted at the same time. If the amount of the string in the minimum subtask from the task decomposition is c, and the amount of the string in the decomposed set achieves c or the amount of the string which is removed in the set achieves n-c, we think this sub-tree doesn't according with the conditions and it will be deleted. For example, the amount of the factor in the set S, n is 6, and the amount of the factor in the minimum subtask set c is 3, and the established binary tree is seen in Figure 2.

According to the characters of the task tree, we code the task trees, and define the root node to denote the character "1", and the left branch to denote the character "0", and the right branch to denote the character "1", so we can take the character string composed by the branch characters from the root node to the leaf node as the code of the node.

From the coding rule, the code of the left child of the node is the code of the node adding "0", and the code of the right child of the node is the code of the node adding "1", so to interview the left node is to interview the node with the code adding "0", and to interview the right node is to interview the node with the code adding "1", and to interview the father node is to interview the node eliminating the last bit. The condition to achieve the minimum subtask is that the amount of "1" has achieved the maximum or the amount of "0" has been less than the amount of "the total – the amount of the maximum "1"".

Based on the recursion definition of the binary tree, the process of the task tree traversal includes four approaches.

(1) Judge whether the minimum subtask is achieved, if it is, turn to (4).

(2) Traverse the left sub-tree post-orderly.

(3) Traverse the right sub-tree post-orderly.

(4) Interview the root node, and write the task information into the database.

4. Implementation of the task management

The project managers need to grasp the task distribution and the development of the project, and manage the tasks in the database, and inquire about the task information table, the client information table and the operation information table. And the information of the inquiry possesses the comparability, and they are in same module, so the database management can be divided into two models, i.e. the task management and the information inquiry.

4.1 Management of the task tree

There are two sorts of representations for the task tree, and one sort is to list all minimum subtasks, which is similar with the list box, and all minimum subtasks can be inquired in the database, and they are taken as the roots of the tree structure to be inserted into the box. The other sort is to represent the task information by the tree form, and it's the hierarchy can be presented very intuitively. Because the records in the task information table are numerous, the time to generate the tree once needs long time, so we adopt the mode of classification generation, i.e. when the user clicks, its next level task tree can be generated. To process the event of "constructor" and the event of "itempopulate" in the

control of "TreeeView", and in the event of "constructor", if the task tree is presented by the tree form, so we only generate the tree roots, and judge whether the root has children to set up the attribute of the root node. In the event of "itempopulate", add the child which clicks the node in the tree, and judge when its child has children to set up the attribute of the child.

4.2 Information inquiry

Because of the comparability among the inquiry of task information, the inquiry of client information, the inquiry of project information and the inquiry of operation, the system can use the general interface to offer the user selection table information, and generate the list option of the inquiry table which user wants to inquire according to the selected table, and when the user select the list selection item, the comparison conditions will be generated according to the data types (including the numerical value, the character and the time). According to user's selection and the query of input, the information needed by the project manager will be picked up to the data window. To achieve this effect, two more tables storing table information and list information should be established, the information can be obtained from the table. Different information will be showed in same data window, and we can use different data windows to set the proper data into the data window according to different table information.

5. Design and implementation of the distributed application program

The distributed application programs include project processing, the communication between the server and the clients, and the task management. The multithreading mechanism transferring RMI and Java by the remote method based on Java language offers a sort of application oriented and simple solution to realize the software implementation of the distributed computation. RMI encapsulates the distributed processing details on the bottom, eliminates the complexity brought by the network distributing and the diversity of the agreements, and offers transparent remote object method transfer mechanism, and the multithreading character of Java make the parallel computation and control be realized easily in the distributed computation environment (Zhu, 2003, P.60-62). The connection of the server and client based on RMI is seen in Figure 3.

The multilayer model distributed application based on RMI generally includes the remote interface, the remote object, the service program and the client program.

5.1 Implementation of the remote interface

The remote interface regulates the interactive interface between the client program and the service program, and to create a RMI program, we first should create the interface to extend the interface. In this interface, we can add the method which can be transferred in the remote computer. With this interface, we can transfer the object in the remote computer like in the local computer. The task information about the user distribution and the method uploading task result should be included in the interface (Zhang, 2002, P.415-451).

5.2 Implementation of the remote object

The remote object can offer real implementation for every method regulated by the remote interface. The functions of the remote method include the identification of user' ID and the judgment of user's operation type. If the user asks for participating in the task computation, modify the state of the minimum subtask which has not returned the result for a long time, inquire the minimum subtask which task state is un-computed, modify the task information table, the operation information table and the user information table, and return the task name of the minimum subtask which was computed. For the computation that the user has participated in the subtask, obtain the result of the task, upload the task result, write the task result into the task information table, and correspondingly modify the data in the operation information table and the user information table. Compile the written remote method class by the RMIC compiler, and generate the peg and framework.

5.3 Implementation of the service program

The server charges for establishing the service example and logging in the remote method logging place which includes the position of the logging place, the selective end number and the server name. The static state method can be referenced by the remote object placed in the logging place which charge for the index through the class of "java.rmi.Naming". After logging in the service, client can interview the service. To make the client go to special service, the banding process establishes a logging item in the logging place. And if the service stops, it will cancel the banding by itself. If the banding is correct, the remote method transfer can be used (David, 2003, P.19-37).

5.4 Implementation of the client end

The client program first acquires the user name and password from the local computer, reads the information in the task configuration files, and judge the task state of the computer, and if the computer has no the task information or the task result has been uploaded, it will send the request to the server program end (i.e. transferring the remote method). The client program first acquires the citation of the remote method in the register of service program, and once the citation is acquired, the client program will use the remote transfer method to obtain the task information. When the client obtains

the task information, the client will take out the data needed by the task computation in the database, write the data into the local file which takes the task name as the file name, and transfer the operation model of the client end, and circularly test whether the result file has been generated. When the result file is generated, transfer the remote method once again, upload the task result to the server, and write the result into the database. The key technologies that the client end needs to realize include inquiring and transferring the remote method, read-out and write-in the text, writing the local last user information into the log file convenient for user operation, writing the task data in the database into the file of the local computer which is used for the data file for the operation of the module computation, and testing the result (utilizing the thread to test whether the result file is generated in every period of time).

The distributed application is the core of the whole distributed computation platform, and it connects the server with hundred thousands of client ends and harmonizes the task assignment and manages many aspects, and its design will directly influence the performance of the distributed computation. It is especially important to adopt the distributed computation technology based on RMI for the development of large-sized system, because it will make the idea which distributes the resources and processes loads on multiple computers possible (Zhang, 2002, P.415-451).

6. Example and conclusions

For the connectivity of the Mesh network, the connectivity is denoted by the form of probability. If the amount of total node and the amount of the dead node in the Mesh network are known, the probability of the connection can be denoted by the following formula.

The connection probability of Mesh network = the connection times/ the total times

Here, the connection times equals to the connection times of Mesh network in each error state.

By the data obtained by the computation, we can analyze the connectivity of Mesh network. Though the problem is simple, but because the amount of the node in the Mesh network is always very numerous, so the computation amount of the connection probability is very large, and it can achieve hundred thousands even more.

Take the connectivity of the Mesh network as the running example, this platform can not only solve this complex problem, but can test the running status of the platform. The process using the platform to solve the connectivity of the Mesh network includes following approaches.

(1) Establish the connectivity problem of Mesh network in the database, i.e. GTT.

(2) Add the program code about the Mesh network connectivity in the client end program.

(3) Start the server end program which connects the background database, and start the tcp monitoring thread and the http monitoring thread.

(4) Install and implement the client end program in various client computers, various client end programs ask for data to the server end program, and return the connection times of the subtask after processing.

(5) The server end program processes the return data from the client end, accumulate the connection times of various subtasks, and correspondingly modify the background database.

(6) When the task is finished, the server end program gives the finally total connection times and the connection probability, and the work stops.

For the connectivity problem of Mesh network tested by us, the total amount of the network node is 120, and the total amount of the dead node is 5. The time to solve this problem in the single computer averagely is 72.554900 hours. The platform used 12 computers to solve it, and the solving time reduced to 6.627607 hours. The test result is seen in Table 1, and the result indicated the platform works normally. Because the input spaces of many biological problems are finite state sequence, the platform can be further applied to solve the problem of biological computation.

The test indicated that for the bioinformatics problems with huge amounts of computation and better distributed computation characters, we could utilize the idea of distributed computation to divide the original problem into many sub-questions, and concurrently exert the subprograms solving the sub-questions by the form of remote method on multiple computers in the network by virtue of Java RMI mechanism and its multithreading mechanism, which could effectively reduce the computation time, and this distributed computation mode required little for the hardware and software environment, and it was very simple and quite effective.

References

David Reilly & Michael Reilly, interpreted by Shenfeng. (2003). *Java Network Programming and Distributed Computation.* Beijing: China Machine Press. P.19-37.

Hu, Kai, Hu, Jianping & Wang, Qiang. (2000). Research of Distributed and Parallel Computing Network Architecture. *Mini-Micro Systems.* No.21(2). P. 113-125.

Ma, Bin. (2000). A polynomial time approximation scheme for the closest substring problem. *11th Annual Symposium*

on Combinatorial Pattern Matching. P. 99-107.

Shi, Xiaoqiu & Kong, Fansheng. (2001). Application of Computer Science in Bioinformatics. *Journal of Zhejiang University of Technology,* No. 29(2). P.161-165.

Wang, Jianxin, Huangmin & Li, Shaohua. (2006). Design and Implementation of a Distributed Computing Platform Based on Task Tree. *Mini-Micro Systems.* No. 27(5). P. 940-944.

Zhang, Yao, Zhang, Qing & Guo, Lishan. (2002). *Tutorial of Java Program Design.* Beijing: Metallurgical Industry Press. P. 415-451.

Zhu, Jianzhong. (2003). Implementation of Distributed Computing Based on RMI. *Computer Engineering.* No. 29(10). P. 60-62.

Table 1. The running time to solve the connectivity of the Mesh network with 120 crunodes and 5 dead dots

No.	Running time of single computer	Running time of double computers	Running time of four computers	Running time of twelve computers
1	72.566700	37.089070	18.995645	6.740065
2	72.670900	37.485003	19.243402	6.788951
3	72.589133	37.198142	18.989061	6.718954
4	72.600190	37.896455	19.358326	6.852362
5	72.576899	37.574122	19.177826	6.770057
6	72.498750	36.599087	18.779556	6.593502
7	72.709003	37.990450	19.495235	6.865860
8	72.598706	37.434912	19.137332	6.745364
9	72.602004	37.377904	19.150420	6.739288
10	72.554900	36.608720	18.758371	6.627607
Average	72.596719	37.325387	19.108517	6.744201

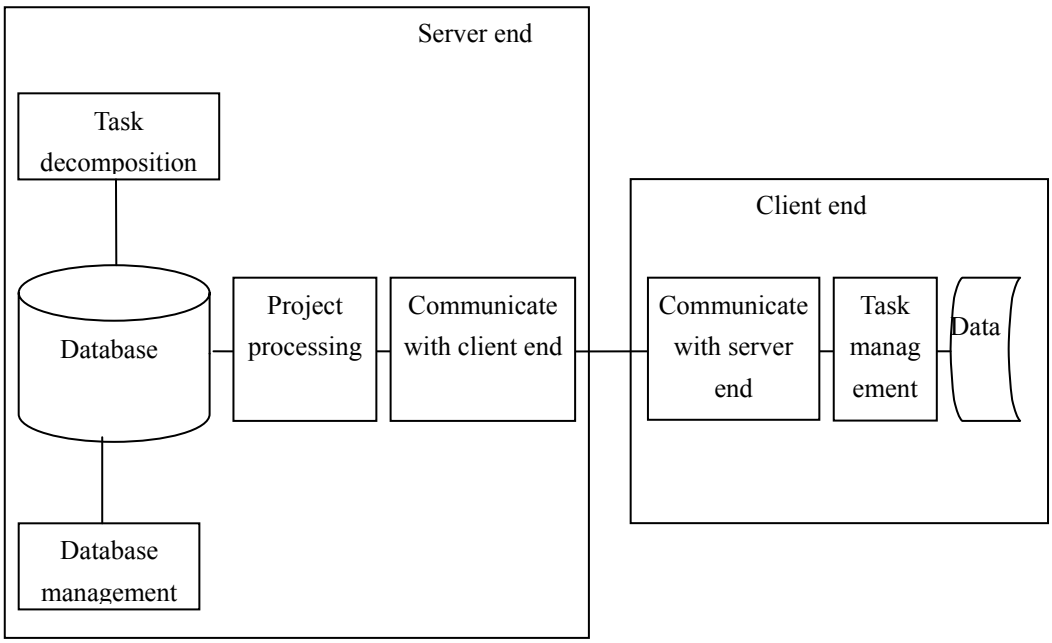


Figure 1. Platform System Structure

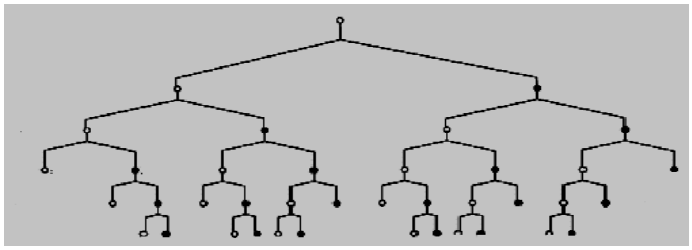


Figure 2. The Task Tree (n=6, c=3)

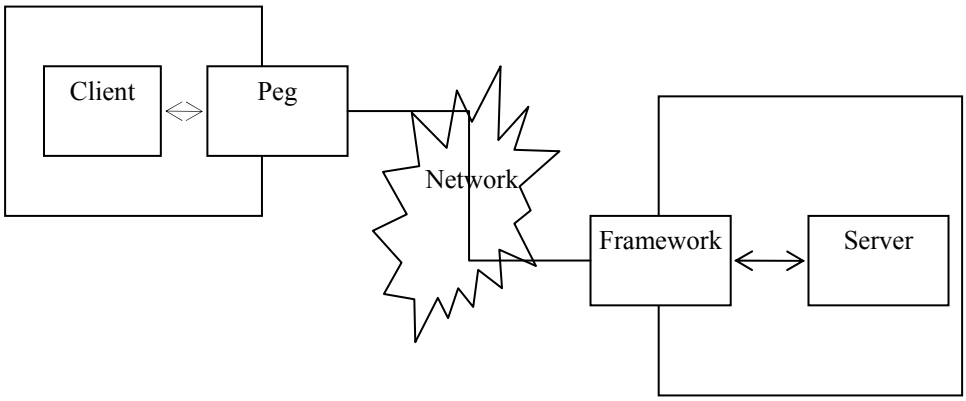


Figure 3. The Client End and Server End Connected by Peg and Framework

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Positive Affects Inducer on Software Quality

Sharifah Lailee Syed-Abdullah

Faculty of Computer Science and Mathematics, Universiti Teknologi MARA

02600, Arau, Perlis. Malaysia

Tel: 60-4-987-5076 E-mail: shlailee@perlis.uitm.edu.my

Mazni Omar

Faculty of Computer Science and Mathematics, Universiti Teknologi MARA

02600, Arau, Perlis. Malaysia

Tel : 60-4-928-4735 E-mail: mazni@isiswa.uitm.edu.my

Mohd Nasir Abdul Hamid

Centre for Islamic Thought and Understanding (CITU), Universiti Teknologi MARA

02600, Arau, Perlis. Malaysia

Tel: 60-4-987-4702 E-mail: mdnasir@perlis.uitm.edu.my

Che Latifah bt Ismail

Centre for Islamic Thought and Understanding (CITU), Universiti Teknologi MARA

Perlis 02600, Arau, Perlis. Malaysia

Tel: 60-4-987-5050 E-mail: chelatifah@perlis.uitm.edu.my

Kamaruzaman Jusoff (Corresponding author)

Faculty of Forestry, Universiti Putra Malaysia

43400, Serdang, Selangor. Malaysia

Tel: 60-3-8946-7176 E-mail: kamaruz@putra.upm.edu.my

Abstract

This paper presents an early empirical study on an agile methodology (Extreme Programming) using Positive Affect metric. The question of interest is whether an agile methodology has any distinct outcome on the positive affectivity of the software developers. And whether these affects will contribute to the quality of software produced. Quantitative methods were utilized, including participative observation and simple statistical tests such as Spearman Correlation and Mann-Whitney test. The results showed that Extreme Programming has positive affectivity which leads to the increase in software quality. This study suggests that when people experience joy and mild contentment, they are more likely to be more creative over wider range of problems, become more resilient over time and are more likely to develop long-term plans and goals.

Keywords: Agile methodology, Empirical study, XP, Positive affect, Software quality

1. Introduction

The traditional methodologies imposed a disciplined process upon software development, with the aim of making software development more efficient in order to produce better quality systems. The detailed process places a strong emphasis on planning and was inspired by other engineering disciplines. The most frequent criticism of these methodologies is that they are bureaucratic thus slowing the development process. The second problem with these methodologies is that the requirements specifications are not flexible. In reality, it is difficult to get the software customer to identify their requirements. Even if the requirements can be identified, the business world is forever changing. As a reaction to these problems, agile methodologies evolved. Agile methodologies welcome change and unpredictability because they are more adaptive than predictive, and more people-oriented than process-oriented. Adaptive approaches are better when the requirements are uncertain or dynamic as in the new type of software application being developed nowadays. When faced with unpredictable user requirements and changes that must be

accommodated during software-in-progress development, developers often experienced stressful emotions such as anxiety and depression (Sharifah Lailee, Holcombe, & Gheorge, 2006a, 2006b).

1.1 Agile Methodology

Due to rapid demand of technological changes, agile methodologies have emerged to alleviate the uncertainty of business requirements. The need to deliver quality software at economical cost is the main issue in software industry. Therefore, the Agile Alliance (2001) has expressed the values in Agile Manifesto as:

Individuals and interactions over processes and tools

Working software over comprehensive documentation

Customer collaboration over contract negotiation

Responding to change over following a plan

The major agile methodologies consist of Scrum, Dynamic Systems Development Method (DSDM), Crystal Methods, Feature Driven Development (FDD), Lean Development and Extreme Programming (XP). Based on the agile manifesto, people and communication are the key ingredients towards producing quality software. Agile employs a lightweight process, which communication plays an important role over comprehensive documentation. This method focuses more on people-oriented approach, which relies on tacit or interpersonal knowledge whilst developing software. Study on agility level of these methodologies (Qumer & Henderson-Sellers, 2008) discovered that Crystal methodology has the highest degree of agility followed by XP, Scrum, and DSDM. Agility characterization were based on flexibility, speed, leanness, learning and responsive features that were calculated quantitatively.

The creators of Agile Alliance agreed that detailed project tactic should be continuously innovated in order to discover a larger set of agile software practices (Cockburn, 2007). Therefore, it is not surprising to discover different set of practices that is similar and complementary to each other. In this paper, four agile methodologies, namely XP, Scrum, Crystal and DSDM are compared according to its principles, methods and tools.

According to Strode (2005), all of the agile methodologies highlight iterative, incremental, and rapid software delivery to deal with flexible requirements in addition to active participation between the team and stakeholders. Except for Scrum and DSDM methodologies that that cater for large projects, the other methodologies are more suitable for small and medium scale project.

Agile modeling (AM) is a practice-based methodology for effective software modeling. AM is not a prescriptive process but focuses more on a portion of an overall software process needed by other agile methodologies. Most agile methods implicitly defined its model to be used, thereby indicating that working software is imperative compared to documentation. However, it is also possible to use object-oriented modeling such as Unified Modeling Language (UML) and software tools to speed up development process. These tools are needed to support the main activities such as requirements management, coding, configuration and testing activities. Each agile methodology serves different purposes and has its own principles and methods, even though they share common characteristics of agile values. Table 1 show the comparison made according to principles, methods, and tools.

For the purpose of this study, a comparison between an agile methodology (XP) and formal methodology (RUP) was conducted. The study focus on the agility of each software methodology to induce positive affect on each member of SE teams and how these affects contribute to the level of software quality.

1.2 Extreme Programming

Beck (2000) introduces XP as a solution to the problems encountered by the formal methods. The XP methodology was created to address requirement changes and project risk. XP begins with 4 values; Communication, Simplicity, Feedback and Courage. It then builds up to practices that XP projects should follow. In XP, features that provide the most business value to the customer must be developed first because the real goal of this approach is to deliver the software that is needed when it is needed. Requirements are written as user stories, which are chunks of functionality that are valuable to the customers. Chunking is a technique in cognitive learning strategy that allows information to be broken down into smaller and meaningful collection of knowledge. Through the use of story cards, it is easier for developers to group different stories according to main functions. Chunking assist developers because human has limited memory capacity and often have difficulty to memorize a large amount of functions or information (Mazni, Sharifah Lailee, & Holcombe, 2009). XP encourages communication by having developers collectively own all codes and work in pairs. Collective code ownership considers code as belonging to the team and not to the individual developers, thereby encourages every developer to contribute new ideas to all segments of the project. Pair programming requires two developers to sit side by side in front of a computer. One person types and thinks tactically about the methods being created, while the other thinks strategically about how the methods fit into the class. Pair programming changes the environment from criticism and competition to learning and cooperation thus improving group cohesiveness and communication. Complex requirements must be simplified to enhance understanding between

team members. XP simple design evolves through constant refactoring, which is guided by suitable metaphor and implemented in accordance to common coding standards. Fowler, Beck, Brant, Opdyke, & Roberts (1999) defined refactoring as the removal of redundant or unused functionality and the restructuring of obsolete designs in order to improve smelly codes. Tong (2004) considered too many comments as useless and further suggested refactoring of these comments into codes. System metaphor is a narrative that everyone (customer, programmer and managers) can associate with when discussing new functionality. The reason for using a metaphor is to achieve a common vision and shared vocabulary. On-site customer is a practice which requires the customer to sit with the development team on a full-time basis. Holcombe (2002) is more realistic by suggesting regular visits and meetings at both the development site and the business site. The humanistic aspect of the communication and the simplicity aspect promote good teamwork because it is an important towards developing quality software.

1.3 Rational Unified Process

Rational Unified Process (RUP) is a software development methodology originated by Rational Software, IBM (Dennis, Wixom, & Tegarden, 2005; Runeson & Greberg, 2004). RUP methodology claimed to be a heavyweight methodology due to its demands to produce extensive documentation. This approach is a use case driven and requires tool to support the software development activities. Unlike agile (XP), RUP emphasizes on specific roles that tailored for a particular project. Examples of these roles are team leader, programmer, user interface designer, software tester and quality assurance. RUP has four distinct phases, which are inception, elaboration, construction and transition phase (Bennet, McRobb, & Farmer, 2006). Although RUP has formal phases, it allows software development teams to design and develop software in iterative and incremental manner.

1.4 Positive affect

Past research has shown that a positive affect induction leads to a greater cognitive flexibility and facilitates creative problem solving across a broad range of settings. Research works by Carnevale & Isen (1986), Aspinwall (1998), Ashby, Isen, & Turken (1999) and Isen (2001), suggest that positive affect increases a person's ability to organize ideas in multiple ways, to access alternative perspectives and also to improve performance in several tasks that are typically used as indicators of creativity or innovative problem solving. In a study of the role of affect on human life, Norman and colleagues (Norman, Ortony, & Russell, 2003) show that affect makes humans smart because affect is always passing judgments and presenting them with immediate information about the world. The affective signals work through neurochemicals, bathing the relevant brain centres and changing the way humans perceive, decide, and react. These neurochemicals change the parameters of thought, adjusting such things as whether reason is primarily 'depth first' (focused, not easily distracted) or 'breadth first' (creative, out of the box thinking, but easily distractible).

It is the intention of this paper to discuss Extreme Programming (XP) as a positive affect inducer and to discuss the findings of the possible impact of the selected XP practices on the software quality. To achieve this, a comparison study was conducted on the software engineering teams consisted of third year students at University Utara Malaysia (UUM). Findings revealed that the XP methodology does have an impact on the positive affectivity and level of software quality.

2. Method

A replicated study (Sharifah Lailee et al., 2006b) were carried out in UUM, to determine empirically, whether teams using XP methodology would experience higher positive affectivity than the teams using the Rational Unified Process (RUP) approach. To measure the developers' state of the positive affect, the positive affect scale of the Positive and Negative Affect Schedule (PANAS) was used. Positive affect was induced by introducing and requiring the XP methodology to be used by half of the development teams. The studies do not include the negative affect because previous research has shown that positive affect can operated as a single construct, indicating that the fluctuation of the positive affectivity, has no effect on the negative affectivity of a person (Anderson & Thompson, 2004).

The validity and reliability of PANAS scale has been demonstrated by other studies (Watson & Clerk, 1997; Watson, Pennebaker, & Folger, 1987; Watson & Tellegen, 1985).The Positive Affect scale showed a satisfactory internal consistency coefficient, Cronbach alpha = 0.78 during the first reading (Week 2), Cronbach alpha = 0.89 during the second reading (Week 6) and Cronbach alpha = 0.87 during the third reading (Week 15). At the beginning of the study, Independent sample t-test was used to compare the total mean score for Positive Affect variables and the result showed no significant difference between Formal teams [N=28 , Mean Score (M) =34.12, Standard Deviation (SD)=4.77] and Agile (XP) teams [N=30, Mean Score (M) =34.53, Standard Deviation (SD)=4.41].

3. Results and discussions

Statistical test Mixed between-within ANOVA was conducted. The analysis indicated that there is no significant difference between the three intervals; Reading 1 (Week 2), Reading 2 (Week 6) and Reading 3 (Week 15) for both methodologies; Formal (N=28, M_1= 34.12 SD_1 = 4.77; N=28 M_2 =33.61, SD_2 = 5.07; N=28 M_3 =33.86, SD_3 = 4.57) and Agile (N=30, M_1=34.53, SD_1 = 4.41; N=30 M_2 = 34.17, SD_2 =5.90; N=30 M_3 =35.37, SD_3 = 6.37) (see Figure 1and

Table 2). This may due to the small effect size (eta squared =0.010). Besides, the results may be moderated by others factors such as partial adoption of Agile (XP) practices during this study. This finding supported earlier finding on the positive effect of Extreme Programming on SE teams (Sharifah Lailee, Holcombe, Karn, Cowling, & Gheorge, 2005).

At the end of Week 15, each information system projects were graded by teams of evaluators which consisted of project client and a lecturer. The mean scores awarded by both evaluators were assessed. The following graph shows grades achieved by both teams according to the projects. The Mann-Whitney non parametric statistical test was used to compare the mean scores and the results showed significant differences in the mean scores for the Formal teams [M=21.09, SD=2.91] and the Agile (XP) teams [M=23.96, SD=1.31]. The graph indicates that teams using Agile (XP) approach were awarded higher score than Formal (RUP) teams (see Figure 2).

In this study, XP methodology was chosen as a positive inducer because of the existence of several XP practices that warrant feedback to the developers. Positive feedback about one's performance has been known as a positive affect inducer (Estrada, Isen, & Young, 1997; Sharifah Lailee et al., 2006a). The XP practices associated with feedback seeking are simple design, pair programming, continuous testing, continuous integration and frequent review (release).

Studies by Aspinwall (1998) and Muraven, Tice, & Baumeister (1998) noted that people must have a surplus of resources such as time, energy and attention to engage in a proactive behaviour. Using XP approach, the developers experienced a surplus of time during coding because less time was engaged in the designing phase. Using simplified design XP developers are actually releasing the stressful task of creation, thus liberating the mind to be more creative and innovative. By reducing the technical aspect of design, the mind was able to approach the problem solving task through a breadth first approach. Design is only an early manifestation of ideas, whereas the coding process allows the developers to realize their idea in a more concrete way. This approach is considered as a positive affect inducer because it allows feedback on the design through the programming code. The ability to see the results and identify the flaws in the design allows the developers to be more creative in the next part of the system. This is the reason why simple design can accommodate flexible requirements because the process of creating part of the system in this manner allows developers to be more innovative in the problem solving process. It was observed that the practice of pair programming started with initial socializing amongst the pair thus creating a positive mood before any formal programming commenced. The ability to discuss the advantages and disadvantages of certain coding ideas enables the pair to seek improvements and avoid specific weaknesses. Studies on pair programming have provided the evidence about the benefits of pair programming (Coman, Sillitti, & Succi, 2008; Succi, Marchesi, Pedrycz, & Williams, 2002; Williams, 2002). With pair programming practice, positive affect is induced through early socializing, more attention and immediate feedback amongst the pair.

Continuous testing allows feedback on the developed code. In the normal software testing domain, testing is usually left at the end of the development cycle thus leaving a very mood which is experienced and the attention of the two developers allow the pair to engage in a more short time for complete testing. In this situation, often the developers were faced with products that have too many defects, as the bugs were discovered too late. The benefit of testing as the software is developed is that the developers are always certain that the software developed is always test compliant. Continuous testing is a practice that is structured so that different levels of testing can be conducted as the solution is being built. In the study by Trope & Pomerantz (1998), participants in whom positive affect has been induced showed greater interest in the part of the test they had failed than did neutral mood participants. The emphasis of the continuous testing enables the developers to feel more confident about the correctness of the code and therefore bolster their confidence and self-esteem.

Continuous integration is another feedback seeking practice, which allows the developers to address performance problems earlier in the development process. The more frequently the developers were able to test the integrated system, the more often they were able to check the functional integrity of the application as some problems do not manifest themselves until they are in the integration environment, such as when a database application is finally tested in a genuine load. The ability to address the performance problems early and to continuously improve the system allows the developers to enjoy a level of self regard or positive affect. Developers using this practice had the advantage of improving their self-esteem continuously, as they worked to perfect the functionality of the integrated system. Frequent release (review) is another practice that commands feedback. Feedback from the client be it positive or negative, is also a positive affect inducer. Accumulating evidences suggest that positive affect can create an increased interest in information about one's liabilities. A study by Trope & Neter (1994) has shown that prior positive experience subsequently increased the interest in feedback of high rather than low self-relevance, even when the feedback was expected to diagnose weaknesses rather than strengths.

4. Conclusion

In this study, even though there was no significant difference in positive affect between the methodologies, it is interesting to observe the impact of the XP on the positive affectivity of the developers which result in higher score being awarded to the agile teams. The teams using a more flexible approach, such as the XP methodology, were able to

incorporate the constant changes made by the clients and thus able increase their positive mood. When a person experiences a positive affect, they show a greater preference for a larger variety of actions and are able to see and think of more possibilities and options to solve whatever problem is faced. People with a positive affect are more likely to take action because they are proactive. This study suggests that when people experience joy and mild contentment, they are more likely to think of a wider range of actions, become more resilient over time and are more likely to develop long-term plans and goals.

Acknowledgement

The authors would like to express our appreciation to Software Engineering Project 1 course lecturers, from Universiti Utara Malaysia, Dr Azman Yasin and Dr Haslina Mohamad who have given their support towards the completion of this research. In addition, we would like to thanks all clients, supervisors and students in this study.

References

Alliance, A. (2001). *Manifesto for Agile Software Development*. Retrieved 31 March, 2009, from http://www.agilemanifesto.org

Anderson, C., & Thompson, L. L. (2004). Affect from top down: How powerful individuals' positive affect shapes negotiation. *Oranizational Behavior and Human Decision Processes, 95*, 125-139.

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529-550.

Aspinwall, L. G. (1998). Rethinking the Role of Positive Affect in Self-Regulation. *Motivation and Emotion, 22*(1), 1-32.

Beck, K. (2000). *Extreme Programming Explained: Embrace Change*. USA: Addison-Wesley.

Bennet, S., McRobb, S., & Farmer, R. (2006). *Object-oriented systems analysis and design using UML*. Berkshire: McGraw-Hill.

Carnevale, P. J. D., & Isen, A. M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational Behaviour and Human Decision Processes, 37*, 1-13.

Cockburn, A. (2007). *Agile Software Development: Cooperative Game (2nd Ed.)*. USA: Addison-Wesley.

Coman, I. D., Sillitti, A., & Succi, G. (2008). Investigating the Usefulness of Pair-Programming in a Mature Agile Team. In P. Abrahamsson, R. Baskerville, K. Conboy, B. Fitzgerald, L. Morgan & X. Wang (Eds.), *Lecture Notes in Business Information Processing*. Berlin-Heidelberg: Springer-Verlag.

Dennis, A., Wixom, B. H., & Tegarden, D. (2005). *Systems analysis and design with UML version 2.0: an object -oriented approach with UML (2nd Ed.)*. Hoboken, NJ: John Wiley and Sons.

Estrada, C. A., Isen, A. M., & Young, M. J. (1997). Positive Affects Facilitates Integration of Information and Decreases Anchoring in Reasoning among Physicians. *Organizational Behaviour and Human Decision Processes, 72*(1), 117-135.

Fowler, M., Beck, K., Brant, J., Opdyke, W., & Roberts, D. (1999). *Refactoring: Improving the Design of Existing Code*. USA: Addison-Wesley.

Holcombe, M. (2002). Extreme Programming for Real: A disciplined, agile approach to software engineering. Sheffield: University of Sheffield.

Isen, A. M. (2001). An influence of positive affect on decision making in complex situations: Theoritical Issues with practical implications. *Journal of Consumer Psychology, 11*(2), 75-85.

Mazni, O., Sharifah Lailee, S.-A., & Holcombe, M. (2009, 3-4 February 2009). *Being Agile in Classroom: An Improvement to Learning Programming.* Paper presented at the National ICT in Education Seminar, Ipoh, Malaysia.

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology, 74*, 774-789.

Norman, D. A., Ortony, A., & Russell, D. M. (2003). Affect and machine design: Lessons for the development of autonomous machines. *IBM Systems Journal, 42*(1), 38-43.

Qumer, A., & Henderson-Sellers, B. (2008). An evaluation of the degree of agility in six agile methods and its applicability for method engineering. *Journal of Information and Software Technology, 50*, 280-295.

Runeson, P., & Greberg, P. (2004). *Extreme Programming and Rational Unified Process-Contrast or Synonyms?* Paper presented at the Software Engineering Research and Practice (SERPS'04), Sweden.

Sharifah Lailee, S.-A., Holcombe, M., & Gheorge, M. (2006a). *Extreme Programming and its ability to improve the creativity and innovative of SE teams.* Paper presented at the National ICT Conference, Perlis, Malaysia.

Sharifah Lailee, S.-A., Holcombe, M., & Gheorge, M. (2006b). The Impact of an Agile Methodology on the Well Being of Development Teams. *Empirical Software Engineering, 11*(1), 143-167.

Sharifah Lailee, S.-A., Holcombe, M., Karn, J., Cowling, T., & Gheorge, M. (2005). The Positive Affect of the XP methodology. In *Lecture Notes in Computer Science* (Vol. LNCS 3556, pp. 218-221). Berlin-Heidelberg: Springer-Verlag.

Strode, D. E. (2005). *The agile methods : an analytical comparison of five agile methods and an investigation of their target environment.* Massey University, Palmerston North, New Zealand.

Succi, G., Marchesi, M., Pedrycz, W., & Williams, L. (2002, May, 26-30). *Preliminary Analysis of the Effect of Pair Programming on Job Satisfaction.* Paper presented at the 3rd International Conference on Extreme Programming and Flexible Processes in Software Engineering (XP2002), Sardinia, Italy.

Tong, K. L. (2004). Essentials Skills for Agile Development: Macau Productivity & Tech.

Trope, Y., & Neter, E. (1994). Reconciling competing motives in self-evaluation: The role of self-control in feedback seeking. *Journal of Personality and Social Psychology, 66*, 646-657.

Trope, Y., & Pomerantz, E. M. (1998). Resolving Conflicts Among Self-Evaluative Motives: Positive Experiences as a Resource for Overcoming Defensiveness. *Motivation and Emotion, 22*(1), 53-72.

Watson, D., & Clerk, L. A. (1997). Extraversion and its positive emotional core. In R. Hogan & J. A. Johnson (Eds.), *Handbook of Personality Psychology* (pp. 767-793). San Diego, Ca: Academic Press.

Watson, D., Pennebaker, J. W., & Folger, R. (1987). Beyond negative affectivity: Measuring stress and satisfaction in the workforce. *Journal of Organizational Behavior Management, 8*(2), 141-157.

Watson, D., & Tellegen, A. (1985). Towards a consensual structure of mood. *Psychological Bulletin, 98*, 219-235.

Williams, L. (2002). Pair Programming: Why Have Two Do the Work of One? In M. Marchesi, G. Succi, D. Wells & L. Willaims (Eds.), *Extreme Programming Perspectives* (pp. 23-33). Boston: Addison-Wesley.

Table 1. Comparison of four agile methodologies according to its principles, methods and tools

Criteria	**Extreme Programming (XP)**	**Scrum**	**Crystal**	**DSDM**
Principles	Programmer-oriented Based on five values: communication, simplicity, feedback, courage, and respect Small and medium scale project	Project-management oriented Facilitates tracking activities Small and large scale project	Problem-solving for well-defined problem Allow adjusting to project size and criticality Small and medium scale project	Business value is imperative Project that are requirements are flexible Small and large scale project
Methods	Iterative and incremental development planning game, pair programming, refactoring, simple design, continuous integration, test-first programming, collective ownership, coding standards, short releases, metaphor, sustainable pace, on-site customer	Incremental phases adjustment Product backlog, sprint, sprint goal, sprint backlog, daily meeting, sprint review meeting	Iterative and incremental development Project interview, team workshop, reflection workshop	Iterative and incremental development Timeboxing, prioritise requirements, prototyping, regular meeting
Tools	Automated testing and configuration tool	Project management tool	Automated regression testing tool	Requirements analysis, system prototyping, testing and configuration management

Table 2. Descriptive statistics of positive affectivity of both teams (UUM 2008)

Method	**N**	**M_1**	**SD_1**	**M_2**	**SD_2**	**M_3**	**SD_3**
Formal (RUP)	28	34.12	4.77	33.61	5.07	33.86	4.57
Agile(XP)	30	34.53	4.41	34.17	5.90	35.37	6.37

Note: $p < 0.05$

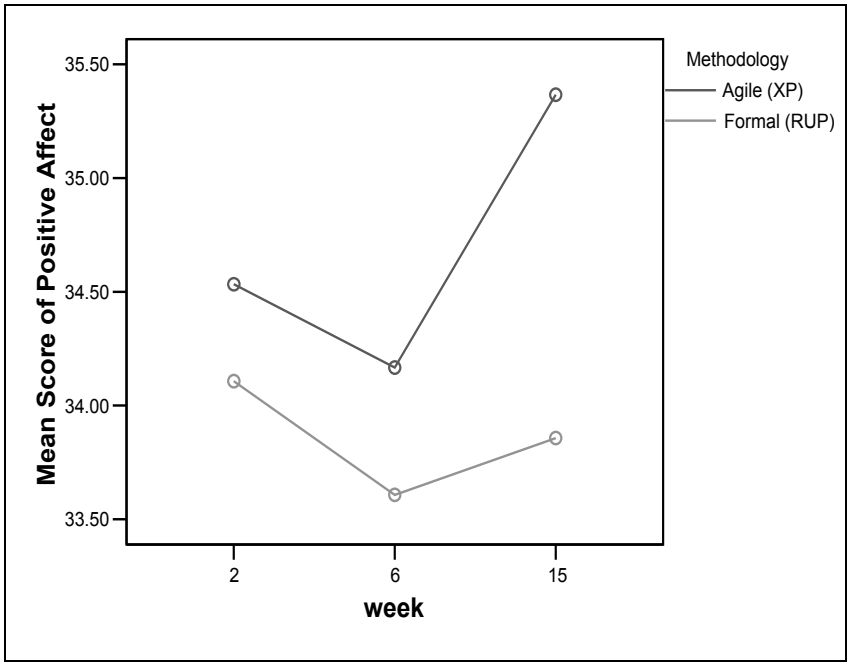


Figure 1. Line graph of the positive affect of the two teams before treatment (week 2) and after treatments (week 6 and week15)

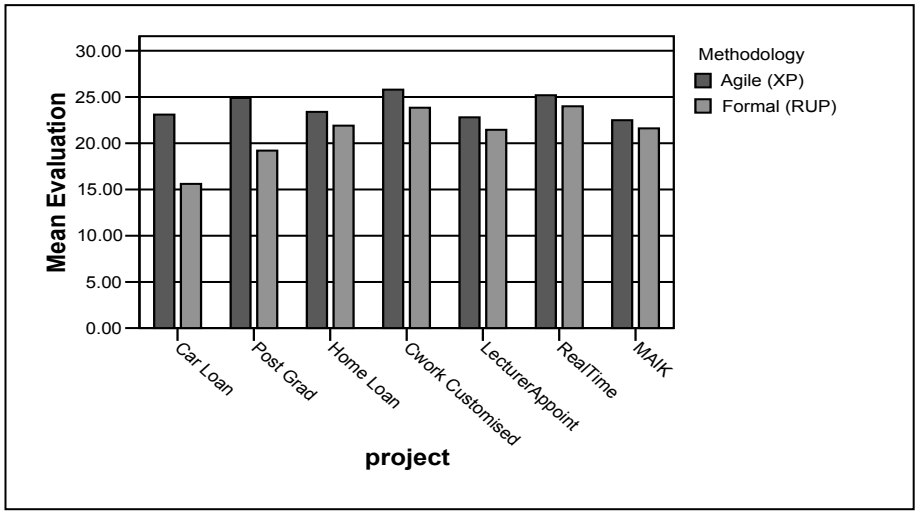


Figure 2. Bar graph showing teams performance according to project

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Text Emotion Computing under Cognition Vision

Jingzhong Wang & Lu Zhang

Information Engineering College, North China University of Technology

Beijing 100144, China

E-mail: jingzhongwang@163.com, elaine_ncut@163.com

Abstract

The emotion computation is an artificial intelligence popular research area. This article in view of text emotion computation research present situation, introduces in the emotion recognition the cognition linguistics and the psychology knowledge, attempts solves the existing problem from a new angle. Learned from experiment's result, the text emotion cognitive model's experiment effect is good.

Keywords: Emotion, Computing, Cognition vision

1. Introduction

With the development of Internet,the information which appears by text form are more and more frequent. It becomes us one kind of the most easily to gain and the richest interactive resources. Presently, however, the text emotion analysis aspect's research is little.Nowadays,with the uncausing development of natural language technology,people could extract emotion information from text through analyzing grammer strcture, Semantic information and emotion glossary methods etc. From the massive texts withdraws the emotion information which is contained in them has the broad application prospect in many aspects.For example, automated analysis the received mail emotion information,and give relevant emotion computing result before user reads. Through emotion analysis to the information in homepage, it may examine the homepage which contain violent emotion,realize homepage filtration and ensure the net's information security.

Along with natural language processing technology unceasing development, the text emotion computation has a richer method and reliable theory basis.

At present, the text emotion computation has the widespread application prospect in many domains. For example,speech synthesis, information security, intelligent robot, pattern recognition, personalized text, and analysis article emotion structure aspects and so on.

2. Related research

Regardless of the method of key word frequency statistics or the method based on semantic characteristic,the emotion glossary in the sentence is an important factor of emotion recognition. In the reality many sentences do not contain the emotion glossary, also has some sentences contain the complex grammar structure and the meaning, only join the corresponding grammar and the syntax analysis can it understand. Whereas the present majority Chinese syntaxes and the grammar analyzer's accuracy rate is low to the complex sentence analysis, introduce the glossary syntactic relation possibly to bring more noises. In view of the above situation, this article take the cognition linguistics and the mood psychology as a foundation, attempts from multiple perspectives and multi-dimension to solve text emotion computation's question, further enhances the accuracy rate of recognition.

2.1 Mood, emotion and sentiment in psychology

Meng Zhaolan in book"Mood Psychology" has made certain discrimination to mood,emotion and sentiment etc expressions. In this article, the emotion includes basic emotion which the humanity and the animal altogether have, and also includes the human some unique high-level emotions, it used American MIT laboratory professor R.Pcard's definition in her monograph "Emotion Computation" in 1997.

In the psychology research, many psychologists have elaborated the relations of mood and cognition. Liang Ning jian pointed out that person's psychological phenomenon divides into two parts which are both mutually connected and distinguished: Mechanism and personality. And the mechanism divides into the cognition process, mood emotion process and will process

2.2 Related knowledge of Text emotion recognition

2.2.1 The text emotion recognition particularity

Text emotion recognition has similarity in emotion production mechanism with appearance ,sound and so on emotion recognition, but actually has the very big difference in the emotion manifestation and the concrete recognition methods. The text emotion recognition mainly concentrates in the semantic analysis to the context and the glossary. Because the environment where emotion occurs and the stimulating factor and so on are indicated through the language, rather than real stimulation through hears, saw and touches, therefore the language understanding veracity is the key of text emotion recognition, is also the most important part. To well understand the language meaning, from multi-dimensional, the three-dimensional angle explore syntax and the semantic structure, this article has introduced the cognition linguistics, the cognition pragmatics and the transference theory and so on related knowledge,enriches text emotion recognition method, expands its research view.

2.2.2 Cognitive context

The cognition pragmatics's foundation is connection theory,and that connection is a concept which depends on linguistic environment. Therefore, the connection theory's linguistic environment view has then become the cognition pragmatics research focal point, this linguistic environment is different with traditional linguistic environment concept,it is one kind of cognitive context view. The traditional context concept is jumbles together the latent context factor and the real context, had not explained how the context does affect the words understanding in the spoken language process, thus is unable to make the explanation for some questions.

The cognitive context is a psychological construction body, He Ziran thought that the cognitive context is composed by three kind of information: the logic information, encyclopedic information and words information. Xiong Xueliang thought that the cognitive context involves to the scene knowledge (concrete situation) the language context knowledge (work memory) and the background knowledge (knowledge structure), also includes the group consciousness which the social group altogether has, namely social token.

3. Text emotion cognitive model

This article uses the above mood theory, cognition psychology and cognition pragmatics and so on related knowledge foundation, synthesize text emotion recognition's characteristic, has given the emotion cognitive model. This model describe current sentence emotion through formula3.1

$$E=(S,P,CC,TR) \quad 3.1$$

And E expressed the current sentence emotion, S represented the last sentence emotion, namely the current emotion foundation, P expressed the current emotion main body's charactor characteristic, CC represented the current sentence the cognitive context, TR represented the emotion migration rule, here is mainly refers to some emotion migration rule which counts from the corpus. The cognitive context is the essential part to define text emotion, therefore this article uses the cognition pragmatics related knowledge, describe cognitive context through formula 3.2

$$CC=(W,K,L) \quad 3.2$$

W expressed the current sentence glossary information, it is the generalized scope glossary, including emotion word, negative term, mark language and so on specific glossary, also includes syntax and grammar and so on information. K is cyclopedic knowledge, is refers to each kind of scheme emotion related in the text emotion recognition, L expresses the logic information, it is the ability which relates the glossary information and cyclopedic knowledge, the logic information decided how this concept contact and relates with what concept.

The text emotion cognitive model, from the emotion cognition's angle, synthesized each kind of possibility influence emotion experience factor, took each kind of factor into account, promoted text emotion recognition. For example, the glossary information supplied the sentence basic meaning, the glossary information’s rich and accurate can activate the correct topographic model, thus inferred the appropriate emotion.

4. Emotion computation model appraisal

4.1 Emotion computation model appraisal

The text emotion appraisal did not have a quite authoritative evaluation method at present, however a good evaluation method may expose the main question, neglects these non-essence question .It is possible to guide the emotion analysis develop to a better direction, enhance labelling precise rate quickly. Regarding the sentence emotion appraisal, this article uses the following two appraisal method:

Appraisal method one: The sentence appraisal method takes the sentence as the division unit, in an article the labeled correct sentence number divides entire chapter article sentence number. As formula 4.1

$$P_i = \frac{1}{m}\sum_{k=1}^{m} f(k) \quad 4.1$$

If the kth sentence labelling is correct, the function value is 1, otherwise the value is 0. m is the language total sentence number, P_i is the language i emotion analysis accuracy rate. The result total accuracy weighs through the macro average and the micro average two methods. The macro average P_{ma} namely is to compute each emotion category rate of accuracy's mean value, micro P_{mi} is similar with the formula 4.1, with all labelling correct sentence dividing total sentence

$$P_{ma} = \frac{1}{n}\sum_{i}^{n} E_i \quad 4.2$$

The formula 4.2 gave the computational method of macroscopic accuracy rate, P_{ma} expressed the macroscopic rate of accuracy, P_{mi} expressed that the ith kind of emotion's rate of accuracy, n expressed the emotion category number.

This evaluation method evaluated each sentence isolatedly , has shut off the sentence emotion connection relations, can only consider accuracy from the single sentence, but cannot effectively appraisal entire chapter article emotion labelling quality and the emotion continuity. For example possibly has following labelling result:

Correct labeling:tranquil, joyful, joyful, joyful, affection, tranquil, joyful, likes

Machine labeling1: tranquilly, joyful, joyful, joyful, affection, affection, affection, tranquil

Machine labelling2: tranquil, affection, joyful, affection, affection, affection, joyful, tranquil

We can see, although labels 1 and labels 2 has 5 labelling correct sentences, labeling result is bad in show emotion's continuous aspect. Therefore besides method one, this article has also used the appraisal method two, to make up and strengthen the emotion continuous appraisal.

Appraisal method 2: (union appraisal law): consult machine translation BLEU evaluation method, altogether the present glossary represents the target language by dual above the fluent degree. This article uses altogether the present multi-dimensional emotion the rate of accuracy which matches in standard labelling to weigh emotion labelling the continuity

$$C_d = \frac{\sum_{i}\sum_{k=1}^{m_i-d+1} f(k,d)}{\sum_{i} m_i} \quad 4.3$$

Formula 4.3 C_d expressed that d continual sentence's emotion labelling rate of accuracy, the function f (k,d) expressed starting in the documents from k to take d labelling the emotion sentence and the correct emotion labelling match, if entire equal, is 1, otherwise is 0. i expressed that the ith documents, m expresses the ith documents total sentence .

4.2 Experimental result and analysis

This article takes periodical as the experiment materials, total 26757 sentences, nearly 800,000 characters. And training language materials 17457, the testing language materials 9300. Based on the above language materials, the experiment compared three method experimental results separately, the appraisal method has used the sentence appraisal and union appraisal method.

The method 1, the key word term frequency statistic

The method 2, emotion recognition based on the semantic characteristic

The method 3, emotion recognition based on cognition appraisal theory

Sentence appraisal's experimental result as shown in table 1, union appraisal method experimental result like table 2

May see from the table experimental result, the method 3 uses the text emotion cognitive model, the micro average rate of accuracy is 45%. method two uses the condition stochastic territory the method labelling rate of accuracy is 43%. Generally speaking, the text emotion cognitive model's experiment effect is good.

5. Conclusions and future works

Recognition methods based on the cognition introduced the cognition linguistics and the psychology knowledge into the emotion recognition, attempts to solve the existing problem from a new angle.

Future prime task: One, continues to expand the main body scale of the emotion glossary, enhance the automatic study the rate of accuracy. Second, further consummates the emotion corpus construction work, increases language materials labelling quantity, and will label the tool and the artificial revision automatically unifies, reduces the manual participation, speeds up the labelling speed. Third, consummates presently in the labelling speed and the rate of accuracy two aspects the some emotion labelling method, continues from the cognition linguistics and the mood

psychology various exploration emotion recognition method, the rich emotion's topographic model, joins the individuality characteristic factor to the emotion the influence and so on.

References

Benjamin K Y, Kwong Y, Wong W L. (2005). Sentiment and content analysis of Chinese news co-verage. *International Journal of Computer Processing of Oriental Languages*, 2005, 18(2):171-183.

Lafferty J, McCallum A, Pereira F. (2001). Conditional random fields:probabilistic models for segmenting and labeling sequence data. *International Conference on Machine Learning,Morgan Kaufmann,* San Francisco, 2001:282-289.

Liu H, Lieberman H, Selker T. (2003). A model of textual affect sensing using real-world know-ledge. *international Conference on Intelligent User Interfaces*, Miami,Florida,2003:125-132.

Liu H, Selker T, Lieberman H. (2003). Visualizing the affective structure of a text document. Conference on Human Factors in Computing Systems, Lauderdale, Florida, 2003:740-741.

Ma C L, Pendinger H, Ishizuka M. (2005). Emotion estimation and reasoning based on affective textual interaction. *International Conference of Affective Computing and Intelligent Interaction*, Beijing, 2005:622-628.

Rosalind W P. (1997). Affective Computing. Cambridge: MIT Press,1997.

Su W P. (2007). Personality and emotion-based high-level control of affective story chara-cters. IEEE Transactions on Visualization and Computer Graphics, 2007, 13(2):281-293.

Wang G J, Wang Z L, Li L X. (2005). An affective model of artificial psychology. IEEE International Conference on Communications, Circuits and Systems, Hong Kong, 2005:1061-1064.

Wang H,Prendinger H,Tgarashi T. (2004). Communicating emotions in online chat using physiolo-gical sensors and animated text. Conference on Human Factors in Computing Systems, Vienna, Austria, 2004:1171-1174.

Zhang Y, Li Z M, Ren F J et al. (2005). Semi-automatic emotion recognition from textual input based on the constructed emotion thesaurus. International Conference on Natural Language Processing and Knowledge Engineering, Wuhan, 2005:571-576.

Table 4.1 Three method experimental results

appraisal type	method1	method2	method3
macro mean	0.2846	0.3168	0.4057
micro mean	0.4130	0.4314	0.4462
happy,	0.5056	0.3534	0.4858
anger	0.1108	0.4744	0.8949
sorrow	0.1686	0.1202	0.1665
fears	0.1625	0.0092	0.2761
wicked	0.2409	0.4838	0.2941
startled,	0.1022	0.0895	0.1150
tranquil	0.6284	0.6872	0.5997

Table 4.2 Results of union sentence evaluation

accuracy rate	method1	method2	method3
C2	0.1958	0.2423	0.2174
C3	0.0967	0.1409	0.1128

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Back Propagation Wavelet Neural Network Based Prediction of Drill Wear from Thrust Force and Cutting Torque Signals

Xu Yang (Corresponding author)

Department of Mechanical System Engineering, Faculty of Engineering, Gunma University

1-5-1 Tenjin-cho, Kiryu 376-8515, Japan

Department of Mechanical Engineering, Dalian Polytechnic University, Dalian 116034, China

E-mail: yangxv1994@163.com

Hiroyuki. Kumehara

Department of Mechanical System Engineering, Faculty of Engineering, Gunma University, 1-5-1

E-mail: kumehara@me.gunma-u.ac.jp

Wei Zhang

Department of Mechanical Engineering, Dalian Polytechnic University, Dalian 116034, China

E-mail: dlzhangwei@yahoo.com

Abstract

The fast monitoring of tool wears by using various Cutting signals and the prediction models developed rapidly in recent years. Comparatively, various wear forecast models based on artificial neural networks (ANN) perform much better in accuracy and speediness than the conventional prediction models. Combining the prominent dynamic properties of back propagation neural network (BPNN) with the enhanced ability of a wavelet neural network (WNN) in mapping nonlinear functions, a Back propagation wavelet neural network (BPWNN) is newly established to perform prominent prediction of drill wear. In this work, a multilayer BPWNN with wavelet algorithm has been applied to predict the average wear of a K10 carbide drill bit for drilling on a high silicon aluminum work piece. Mean value of the thrust force, cutting torque, and drilling depth, spindle speed and feed-rate are inputs to the network, and drill wear is the output. Drilling experiments have been carried out over a wide range of cutting conditions and the effects of drill wear, cutting conditions (spindle speed, drilling depth and feed-rate) on the thrust force and cutting torque have been investigated. Performance of BPWNN has proved to be satisfactory by experimental result. The accuracy of the prediction of drill wear using BPWNN is found to be better than using BPNN, and that BPWNN can learn the pattern faster compared to BPNN and could be used advantageously in online drill wear monitoring and prediction.

Keywords: Back propagation wavelet neural network, Drilling, Thrust force, Cutting torque, Drill wear monitoring

1. Introduction

Manufacturing industries are trying to reduce the operation cost and better quality of product. In machining process, tool wear and breakage often has a fatal effect on the product. So, automation with online monitoring and prediction in metal cutting operations is a new approach toward improvement of the quality of the product as well as reduction of the cost of the product. Especially prediction of tools wear plays a very significant role to realize a automated manufacturing system. Drilling represents over 30% of all cutting operations performed in industries (Young Jun Choi, Min Soo Park, Chong Nam Chu, 2008). Monitoring and prediction of drill wear is an important issue since wear on drill affect the hole quality and tool life of the drill. Therefore, monitoring and prediction of drill wear is an important area of research. For improving the performance of decision-making in tool condition monitoring, many works have been reported in the broad field of different type of intelligent monitoring systems.

Tool wear detection methods can be direct or indirect. Indirect methods, in which various signals of sensors (thrust force and torque, current, acoustic emission, vibration) correlating to tool wear have been extensively applied to prediction of tool wear detection (E. Jantunen, 2002). Mauri and Matti (R. Mauri, S Matti, 1995) reported that the online measurements of cutting power, spindle motor current, torque and thrust force did not vary significantly in entire

tool life until the failure of the drill bit. Li and Wu (T.I. Liu, S.M. Wu, 1990) introduced thrust force and torque which were selected as the features relevant to the four drills wear states. Bhattacharyya and Sengupta (L. Xiaoli, S.K. Tso, 1999) established the combinations of signal processing techniques for real-time estimation of tool wear in face milling using cutting force signals are presented. Xiaoli and Tso (P. Bhattacharyya, D. Sengupta, S. Mukhopadhyay, 2007) established the correlation between drill wear and motor current at different cutting conditions. Xiaoli (L. Xiaoli, 1999) applied wavelet transform on motor current signals to detect the breakage of a small diameter drill. AE has been used to tool monitoring successfully during turning operations (E. Govekar, J. Gradisek, I. Grabe, 2000). Abu-Mahfouz (I. Abbu-Mahfouz, 2003) reported that the drilling wear detection and classification uses vibration signals and artificial neural network. Parameters for monitoring drill conditions, thrust force and cutting torque have been used because they are the most sensitive to the changes of drill conditions (G.S. Li, W.S. Lau, Y.Z. Zhang, 1992).

Artificial neural network can effectively map the non-linear relationship of the sensors signals and wear state. Multilayer neural network with back propagation learning, a simple, robust and non-linear modeling device, has been widely applied to develop systems for monitoring wear of turning, milling and drilling tools (D.E. Dimla, P.M. Lister, N.J. Leighton, 1997), Hong et al. (G.S. Hong, M. Rahaman, Q. Zhou, 1992) used cutting force signal input to back propagation neural network for tool wear monitoring. Lin and Ting (S.C. Lin, C.J. Ting, 1996) used back propagation neural network, and determined the best architecture for drill wear predicting. Panda and Singh (S.S. Panda, A.K. Singh, D. Chakraborty, S.K. Pal, 2006) introduced a method to chip thickness as an additional input parameters entered into the neural network to obtain a better training effect. Panda and Chakraborty (S.S. Panda, D. Chakraborty, S.K. Pal, 2008) used thrust force, torque and vibration the input to two different types of different back propagation neural network (BPNN) and radial basis function network (RBFN) for predicting flank wear in drills.

Wavelet neural network (WNN) is a feed forward neural network based on wavelet transform (WT) (H. Zhong, J. Zhang, M. Gao, J. Zheng, G. Li, L. Chen, 2001). This successful synthesis of theories has generated a new class of networks called Wavelet Neural Networks (WNN) (Q. Zhang, A. Benveniste, 1992). The distinct feature of these networks is that they use wavelet functions as the activation functions of hidden layer neurons. Using theoretical features of wavelet transform, network construction methods could be developed (Szu H, Telfer B, Kadambe S., 1992). The network has advantages of the wavelet transform (in denoising, background reduction and recovery of characteristic information) and also has neural network capacity of "universal" approximation (H. Zhong, J. Zhang, M. Gao, J. Zheng, G. Li, L. Chen, 2001).

The present paper, effects of cutting conditions (spindle speed, drilling depth and feed-rate) and drill wear on thrust force and cutting torque has been tested; and the drilling process was modeled by a neural network with a back wavelet algorithm and weights gradient descent method. Comparison has also been made between BPNN and BWFN in drill wear monitoring. The performance of BPWNN has been found to be satisfactory while validated with experimental result.

2. Back propagation wavelet neural network

2.1 Back propagation wavelet neural network architecture

Back propagation wavelet neural network (BPWNN) is an ANN that is integrated with wavelet techniques and has been used successfully in many fields. Instead of conventional nonlinear sigmoid transfer functions, the transfer function of the nodes in a wavelet neural network is wavelet bases. Because wavelet bases have the characteristics of time precision in high frequency domains and frequency precision in low frequency domains due to dilating and translating the mother wavelet, the ability of a WNN in mapping complicated nonlinear functions is enhanced considerably.

BPWNN has been used in the present work. Basic structure of BPWNN with a single hidden layer is shown in Figure 1. Functioning in three stages, namely learning or training or testing. The network consists of three layers: an input layer, an hidden layer, and an output layer. Input layer receives information from the external sources to the network for processing. Hidden layer receives information from the input layer, and processes the information, and output layer receives processed information from the network, and sends the results out to an external receptor. Where, W_{ji} is the weight between i th neuron of the input layer and j th neuron of the hidden layer, V_{kj} the weight between j th neuron of the hidden layer and k th neuron of the output layer, θ_j the bias at j th neuron of the hidden layer, τ_k the bias at k th neuron of the output layer, respectively.

Back propagation wavelet neural network algorithm has been used in the present work. And thrust force, torque, drilling depth, spindle speed and feed-rate are used as input parameters and corresponding wear are used as the output parameter as showing in Figure 1. The normalized data sets are used for training the network. The data sets are normalized in the range of 0.1–0.9 by using Eq. (1):

$$I_i = \frac{0.8}{d_{\max} - d_{\min}}(d_i - d_{\min}) + 0.1 \quad (1)$$

Where d_i is the actual value, $d_{\max}$ is the maximum value of d_i, $d_{\min}$ min is the minimum value of d_i, and I_i is the normalized value corresponding to d_i.

Outputs signal of input layer are modified by interconnection weight (W_{ji}) of i th node of input layer to j th node of Hidden layer. The sum of the inputs signals is then modified by a Morlet transfer function. (ψ). Similarly, outputs signal of hidden layer are modified by interconnection weight (V_{kj}) of j th node of hidden layer to k th node of output layer. The sum of the modified signal is then modified by sigmoid transfer function (f) and output is collected at output layer.

Let $I_l^P = (I_1^P, I_2^P, \cdots\cdots, I_i^P, \cdots\cdots, I_l^P)$ is the i th input data from the P th training pattern of the data set to the input layer. l is the total number of input parameter applied to the network.

Output from a neuron in the input layer by using Eq. (2):

$$O_i^P = I_i^P, \; i = 1,2,\cdots\cdots l \quad (2)$$

Output from a neuron in the hidden layer by using Eq. (3):

$$O_j^P = \psi\left(\sum_{i=0}^{l} \frac{W_{ji} O_i^P - b_j}{a_j} + \theta_j\right), \; j = 1,2,\cdots\cdots, m \quad (3)$$

Output from a neuron in the output layer by using Eq. (4):

$$O_k^P = f\left(\sum_{i=0}^{l} V_{kj} O_j^P + \tau_k\right), \; k = 1,2,\cdots\cdots, n \quad (4)$$

Suppose $\psi(x) \in L^2(R)$ is a mother wavelet, a series of daughter wavelets can be developed through dilating and translating function $\psi(x)$:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right) \quad (5)$$

Where a is the dilation factor, b is the translation factor.

Mother wavelet is suitable for this work to adopt the wavelet as the transfer function of the BPWNN in consideration of high resolution in both time and frequency domains. The Morlet wavelet is defined as follows (Szu H, Telfer B, Kadambe S, 1992):

$$\psi(x) = \cos(\omega \cdot x) \times \exp(-x^2/c)$$

In Hidden layer transfer function (ψ) the signal function does not undergo wavelet transformation and inverse transformation. Therefore, it is applicable in this work to use the type of Morlet wavelet as follows (Eq. (6)):

$$\psi(x) = \cos(1.75x) \times \exp(-x^2/2) \quad (6)$$

Output layer transfer function (f) of sigmoid transfer function as follows (Eq. (7)):

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (7)$$

2.2 Learning or training in Back propagation wavelet neural network

Batch mode type of supervised learning has been used in the present case, where, interconnection weights are adjusted using delta rule algorithm after sending the entire training sample to the network (H. Zhong, J. Zhang, M. Gao, J. Zheng, G. Li, L. Chen, 2001). During training, the predicted output is compared with the desired output, and the mean square error is calculated. If the mean square error is more than a prescribed limiting value, it is back propagated from output to input, and weights are further modified till the error or number of iterations is within a prescribed limit.

Mean square error (MSE), E_P for pattern P is defined as

$$E_P = \frac{1}{2}\sum_{k=1}^{n} (Y_k^P - O_k^P)^2 \quad (8)$$

Where Y_k^P is the target output, O_k^P is the computed output.

At any epoch η, the gradient descent algorithm is well known, the modified with the adaptive learning factor and the momentum factor, can be summarized as the adjustment of the weight change according to Eq. (9) and Eq. (10).

$$\Delta W(n+1) = \alpha \times \Delta W(n) - \eta \times \frac{\partial E_P}{\partial W} \tag{9}$$

$$\Delta V(n+1) = \alpha \times \Delta V(n) - \eta \times \frac{\partial E_P}{\partial V} \tag{10}$$

η is the learning rate, $0 < \eta < 1$. α is the momentum coefficient, $0 < \alpha < 1$.

In the present case, prediction has been considered for training of the wavelet neural network, which consists of the following steps:

1) Training samples are normalized and read.

2) Weights of hidden layer and output layer are initialized to randomly within [0,1] values, and bias of hidden layer and output layer are initialized to randomly within [0,1] values, and the dilation and translation factors a and b as well.

3) Determine other initial parameters as the following through trial and error. The momentum factor is fixed on 0.5-0.9. The initial learning rate is set at 0.3-0.9, respectively. The stop conditions are the limitation of the mean square error (MSE) of 0.00003 and the maximum training iteration epochs of 2000.

4) Output from hidden layer was calculated as per Eq. (3). Output from output layer was calculated as per Eq. (4).

5) MSE of training sample was calculated. If the MSE training not reach the goal specified then weight is updated based on gradient descent method. The weight was updated in the present case in batch mode.

6) The process was carried out for a designated number of iteration.

2.3 Testing of Back propagation wavelet neural network

An experimental data set is divided into the training set and the testing set. The error on the testing set is monitored during the training process. The testing error is decreasing during the training. When the testing error increases to specified number of iterations, the training will stop. The weights when getting the minimum value of the testing error are returned.

3. Experimental set-up

The schematic diagram of the experimental setup is shown in Figure 2 The experiments were performed on a 5-axis CNC vertical drilling machining center (MAKINO MC1010-5XA). The thrust force and cutting torque are recorded through a piezoelectric drilling dynamometer (YDZ-II02). Measuring range of thrust force varies from -5 to 5 kN with a sensitivity of 3.8 pC/N. Similarly range of torque varies from-100 to 100 Nm with a sensitivity of 1.68 pC/Ncm. Signals from the dynamometer were passed through low pass filter, amplified through charge amplifier (YE5850), and stored in the computer through a data acquisition system (NR-2000). The analog outputs from the sensors were converted to digital signals by A/D board of data acquisition system sends to personal computer (PC).

In the experiments, three types of drill flank(i.e. Multifacet Ⅰ flank, Multifacet Ⅱ flank and Standard flank) are considered. The Multifacet Ⅰ, Multifacet Ⅱ and Standard drills are shown in Figure 3. Scanning Electron Microscope has been used to measure the drill wear. Figure 4 shows the photographs of flank faces and corresponding wear of three drill's flanks (a) Multifacet Ⅰ, (b) Multifacet Ⅱ (c) Standard in the drilling experiments. Carbide drill bits with different flank have been used for drilling in high silicon aluminum alloy workpiece under different cutting conditions. A series of tests were carried out under different cutting conditions to confirm the effectiveness of the proposed method. The experiment conditions are shown in Table 1.

4. Experimental results

Drilling operations have been conducted over a wide range of cutting condition. Spindle speed has been varied in the range of 3000-6000 rpm in six steps. Feed rate has been varied from 0.1 to 0.2 mm/rev in three steps. Drilling depth has been varied from 10 to 20 mm in three steps. K10carbide drills of three different Flank figures (standard, Multifacet Ⅰ and Multifacet Ⅱ) have been used for drilling in high silicon aluminum alloy A390 plates. Different combinations of three design variables viz. spindle speed; feed rate and drilling depth have been used to perform 18 different drilling operations on A390 plate. For each of these conditions, thrust force and torque have been measured using dynamometer and the data are stored in the computer through the WAVE SHOT! 2000 software. Corresponding to each cutting condition, the wear has also been measured using scanning electron microscope. The results of the

experiments are tabulated in Table 2, which shows the thrust force, torque and the drill wear corresponding to 18 different cutting conditions.

4.1 Effect of Input parameters on thrust force

Figure 5–Figure 7 shows the effect of important cutting parameters on thrust force and torque during drilling operation.

Figure 5 shows variation of thrust force and torque with spindle speeds (3000 and 4000rpm) for different feed rate and different flank drill. It could also be observed that the decreases trend of thrust force and torque with Spindle speed is more pronounced at flank of Standard (feed rate 0.1mm/rev) and Multifacet Ⅱ (feed rate 0.2mm/rev). Similarly it could be observed that thrust force and torque decreases with increase in spindle speed and this is due to facts that increase in spindle speed increase the temperature generation during shearing action of cutting tool, and hence it softens the material of the work piece which results in the reduction of the thrust force. While on the contrary be observed that the increasing trend of thrust force and torque with Spindle speed is more pronounced at flank of Multifacet Ⅰ (feed rate 0.15mm/rev), Because the extend of the cutting edge, the cutting temperature of Multifacet Ⅰ drills is not increases with the increment of the rotating speed obviously.

Figure 6 shows variation of thrust force and torque with spindle speeds (4500 and 5500rpm) for different feed rate and different flank drill. It could also be observed that the decreases trend of thrust force and torque with Spindle speed is more pronounced at flank of Standard (feed rate 0.2mm/rev) and Multifacet Ⅱ (feed rate 0.15 mm/rev). While on the contrary be observed that the increasing trend of thrust force and torque with Spindle speed is more pronounced at flank of Multifacet Ⅰ (feed rate 0.1mm/rev), the condition in the range of 4500 and 5500rpm spindle speeds is same with 3000 and 4000rpm spindle speeds. Because the extending of the cutting edge, the cutting temperature of Multifacet Ⅰ drills does not increase with the increment of the rotating speed obviously.

Figure 7 shows variation of thrust force and torque with spindle speeds (5000 and 6000rpm) for different feed rate and different flank drill. It could also be observed that the increasing trend of thrust force and torque with Spindle speed is more pronounced at flank of all. Because increase of cutting temperatur was not obvious, it doesn't soften the material,so it leads the increment of the thrust force and torque.

The thrust force, torque and the drill wear which correspond to cutting conditions are shown in Table 2, it could be observed that the cutting parameters with the drill wear have been at nonlinear mapping relations. Combining the prominent dynamic properties of back propagation neural network with the enhanced ability of a wavelet neural network in mapping nonlinear functions, a back propagation wavelet neural network is newly established in this work to perform prominent prediction of drill wear.

4.2 Drill wears monitoring and prediction by neural network

In this case, the network has five input nodes and one output node. Neural network architectures, prepared using various combinations of input parameters (Table 2), such as spindle speed, feed rate, drilling depth, thrust force, torque, and with flank wear as the only output of the network. From the 18 data sets obtained in the experiment, 10 have been selected at front for training the network and remaining 8 are used for testing. Optimum network architecture include number of hidden layers, number of neurons in the hidden layers, learning rate (η) and momentum coefficient (α) has been obtained by error based on mean square error MSE in training, MSE in testing, and convergence rate (number of iterations).

4.2.1 Optimum neural network architecture of BPWNN and prediction performance

Several architectures of the BPWNN are trained for different learning rates and momentum coefficient values. Performances of some of the architectures for ten training and eight testing result are shown in Table 3. Best network architecture has been arrived at by trial and error of different combination of (η) and momentum coefficient (α) of hidden nodes. In the present case, based on these observations, based on these observations, the optimum network obtained is 5-15-1 with $\alpha = 0.9$ and $\eta = 0.8$.

Figure 8 shows the variation of MSE in training and testing with number of iterations for network of 5-15-1 with $\alpha = 0.9$ and $\eta = 0.8$. It could be observed that network could be trained till 38 iterations after which it starts over fitting and at this point MSE for training as 0.0000274 and MSE for as testing 0.0000051.

4.2.2 Optimum neural network architecture of BPNN and prediction performance

The BPNN apply architectures are similar to those of the BPWNN described in Table 3, it is BP networks of 5-input model for drill wear forecasts, the activation functions are all sigmoid without wavelet.

Monitoring amd prediction performances of BPNN of the architectures for ten training and eight testing result are shown in Table 4. In the present case, the optimum network obtained is 5-5-1 with $\alpha = 0.9$ and $\eta = 0.8$.

Figure 9 shows the variation of MSE in training and testing with number of iterations for network of 5-5-1 with $\alpha = 0.9$ and $\eta = 0.8$. It could be observed that network could be trained till 52 iterations after which it starts over

fitting and at this point MSE for training as 0.0000483 and MSE for testing as 0.0000064.

4.3 Comparison of monitoring amd prediction performances between BPNN and BPWNN

The BPWNN has been trained; it has been verified with testing sample. Figure 10 is optimum to choose the architecture of BPWNN. It shows percentage of error between actual value and the predicted value, it can be observed that the present 5-15-1 BPNN $\alpha = 0.9$ and $\eta = 0.8$ predicts the result within 9.38%. Similarly, Figure 11 is optimum to choose the architecture of BPNN. It expresses percentage of error between actual value and the predicted value and it can be observed that the present 5-5-1 BPNN $\alpha = 0.9$ and $\eta = 0.8$ predicts the result within 12.31%.

Figure 12 shows network testing set by performing a linear regression between each value of the network response and the corresponding target. It gives similar comparisons clearly.

5. Conclusions

When combination of the process parameters such as drilling depth, spindle speed, feed is supplemented with sensors signals such as thrust force and torque for training a Back propagation wavelet neural network, The enhanced basis function network is used to predict drill wear. It has been showed in the work that both BPNN and WNN can predict the drill wear reasonably well. From the present work, the following specific conclusions have been drawn:

Inclusion of thrust force, torque signals, drilling depth and Cutting conditions as input to train neural network results in a well-trained network, which can predict the wear with prominent accuracy.

This paper presents a new model for prediction drill wear based on the Back propagation wavelet neural network (BPWNN) and weights are adjusted using gradient descent method. Experiments prove that the model is capable of mapping drill wear that is usually highly nonlinear. This is because the BPWNN combines the advantages of both the BPNN and WNN.

Experiments show that the value of the prediction MSE by the trained BPWNN described in the paper is less than those by other typical models (BPNN). It proves that the effort for improving the prediction precision in the paper is fruitful, and BPWNN can monitor online drill wear faster than BPNN.

References

D.E. Dimla, P.M. Lister, N.J. Leighton, (1997). Neural network solutions to the tool condition monitoring problem in metal cutting—a critical review of methods, *International Journal of Machine Tools & Manufacture*. 37 (9) (1997). 1219–1241.

E. Govekar, J. Gradisek, I. Grabe, (2000). Analysis of acoustic emission signals and monitoring of machining processes, *Ultrasonic* .38 (2000). 598–603.

E. Jantunen, (2002). A summary of methods applied to tool condition monitoring in drilling, *International Journal of Machine Tools & Manufacture*. 42 (2002) 997–1010.

G.S. Hong, M. Rahaman, Q. Zhou, (1996). Using neural network for tool condition monitoring based on wavelet decomposition, *International Journal of Machine Tools & Manufacture*. 36 (5) (1996) 551–566.

H. Zhong, J. Zhang, M. Gao, J. Zheng, G. Li, L. Chen, (2001). Chemom. Intell. Lab. Syst.59 (2001) 67.

I. Abbu-Mahfouz, (2003). Drilling wear detection and classification using vibration signals and artificial neural network, *International Journal of Machine Tools & Manufacture*. 43(2003) 707–720.

L. Xiaoli, (1999). On-line detection of the breakage of small diameter drills using current signature wavelet transform, *International Journal of Machine Tools & Manufacture*. 39 (1999) 157–164.

L. Xiaoli, S.K. Tso, (1999). Drill wear monitoring based on current signals, *Wear* 231 (1999) 172–178.

P. Bhattacharyya, D. Sengupta, S. Mukhopadhyay, (2007). Cutting force-based real-time estimation of tool wear in face milling using a combination of signal processing techniques, *Mechanical Systems and Signal Processing* 21 (2007) 2665–2683.

Q. Zhang, A. Benveniste, *(1992). IEEE Trans. Neural Network*.3 (1992) 889.

R. Mauri, S Matti, (1995). Tool wear and failure in the drilling of stainless steel, *Journal of Materials Processing Technology*. 52 (1995) 35–43.

S. Li, W.S. Lau, Y.Z. Zhang, (1992). In-process drill wear and breakage monitoring for a machining center based on cutting force parameters, *International Journal of Machine Tools and Manufacture*. 32 (1992) 855–867.

S.C. Lin, C.J. Ting, (1996). Drill wear monitoring using neural network, *International Journal of Machine Tools & Manufacture*. 36 (4) (1996) 465–475.

S.S. Panda, A.K. Singh, D. Chakraborty, S.K. Pal, (2006). Drill wear monitoring using back propagation neural network,

Journal of Materials Processing Technology 172 (2006) 283–290.

S.S. Panda, D. Chakraborty, S.K. Pal, (2008). Flank wear prediction in drilling using back propagation neural network and radial basis function network, *Applied Soft Computing* 8 (2008) 858–871.

Szu H, Telfer B, Kadambe S, (1992). Neural network adaptive wavelets for signal representation and classification. Opt Eng 1992; 36(9): 1907–1916.

T.I. Liu, S.M. Wu, (1990). On-line detection of drill wear, *ASME Journal of Engineering for Industry*. 112 (1990) 299–302.

Young Jun Choi, Min Soo Park, Chong Nam Chu, (2008). Prediction of drill failure using features extraction in time and frequency domains of feed motor current, *International Journal of Machine Tools & Manufacture*.48 (2008) 29-39.

Table 1. Cutting conditions

Tool	Drill type: Standard flank, Multifork Ⅰ flank and Multifork Ⅱ flank
	Diameter: 8mm
	Tool material: K10carbide
	TiAlN coated
Cutting conditions	Spindle speed: 3000, 4000, 4500, 5000,5500 and 6000 rpm
	Feed-rate: 0.1, 0.15and 0.2 mm/rev
	Exterior and interior lubrication
Work-piece	High silicon aluminum alloy A390

Table 2. Experimental input/output data for neural network

Serial number	Drill type	Speed (rpm)	Feed (mm/rev)	Deepness(mm)	Thrust force(N)	Torque(Nm)	Wear(mm)
1	Standard	3000	0.1	100	2148.04	10.34	0.178
2	Multifork Ⅰ	3000	0.15	150	4155.97	15.68	0.093
3	Multifork Ⅱ	3000	0.2	200	3032.05	15.35	0.176
4	Standard	4000	0.1	150	1983.18	9.32	0.157
5	Multifork Ⅰ	4000	0.15	200	4607.97	17.6	0.084
6	Multifork Ⅱ	4000	0.2	100	2876.34	14.63	0.061
7	Multifork Ⅰ	4500	0.1	100	2972.67	11.81	0.097
8	Multifork Ⅱ	4500	0.15	150	2090.95	10.04	0.113
9	Standard	4500	0.2	200	3125.68	11.97	0.25
10	Multifork Ⅱ	5000	0.1	200	1740.67	7.19	0.137
11	Standard	5000	0.15	100	2318.48	8.84	0.136
12	Multifork Ⅰ	5000	0.2	150	3813.8	15.74	0. 067
13	Multifork Ⅰ	5500	0.1	200	3335.29	12.66	0.065
14	Multifork Ⅱ	5500	0.15	100	1987.52	8.96	0.063
15	Standard	5500	0.2	150	2799.22	10.54	0.178
16	Multifork Ⅱ	6000	0.1	200	1818.54	7.9	0.101
17	Standard	6000	0.15	150	2411.55	10.54	0.133
18	Multifork Ⅰ	6000	0.2	100	3971.78	16.92	0. 102

Table 3. Convergence performances for different back propagation wavelet neural network architectures

Neural network Architecture	Momentum Coefficient (α)	Learning rate (η)	MSE Training	MSE Testing	Number of Iterations	Maximum predicted Error (%)	Minimum predicted Error (%)
5-5-1	0.9	0.8	0.0000297	0.0000088	98	9.38	0.000774
5-5-1	0.8	0.6	0.0000291	0.0000068	143	9.14	0.0044
5-5-1	0.6	0.4	0.0000297	0.0000069	280	9.43	0.00582
5-5-1	0.5	0.3	0.0000298	0.0000131	316	9.42	0.00583
5-10-1	0.9	0.8	0.0000291	0.0000118	55	9.12	0.000527
5-10-1	0.8	0.6	0.0000294	0.0000169	82	9.27	0.00146
5-10-1	0.6	0.4	0.0000294	0.0000089	140	9.26	0.00226
5-10-1	0.5	0.3	0.0000295	0.0000023	227	9.58	0.00667
5-15-1	0.9	0.8	0.0000274	0.0000051	38	9.18	0.000733
5-15-1	0.8	0.6	0.0000278	0.0000227	68	9.39	0.000636
5-15-1	0.6	0.4	0.0000290	0.0000028	132	9.33	0.00902
5-15-1	0.5	0.3	0.0000294	0.0000067	168	9.15	0.00658

Table 4. Convergence performance for different back propagation neural network architectures

Neural network Architecture	Momentum Coefficient (α)	Learning rate (η)	MSE Training	MSE Testing	Number of Iterations	Maximum predicted Error (%)	Minimum predicted Error (%)
5-5-1	0.9	0.8	0.0000483	0.0000064	52	12.31	0.00233
5-5-1	0.8	0.6	0.0000481	0.0000523	75	11.42	0.0148
5-5-1	0.6	0.4	0.0000490	0.0000029	126	11.39	0.0112
5-5-1	0.5	0.3	0.0000487	0.0000396	183	10.19	0.0172
5-8-1	0.9	0.8	0.0000482	0.0000095	64	10.16	0.00188
5-8-1	0.8	0.6	0.0000478	0.0000285	74	11.29	0.009360
5-8-1	0.6	0.4	0.0000492	0.0000289	128	10.55	0.00202
5-8-1	0.5	0.3	0.0000483	0.0000053	243	14.47	0.00147
5-12-1	0.9	0.8	0.0000422	0.0000046	201	14.88	0.000247
5-12-1	0.8	0.6	0.0000423	0.0000039	358	15.08	0.000430
5-12-1	0.6	0.4	0.0000496	0.0000045	375	7.26	0.000508
5-12-1	0.5	0.3	0.0000456	0.0000049	563	8.48	0.000443

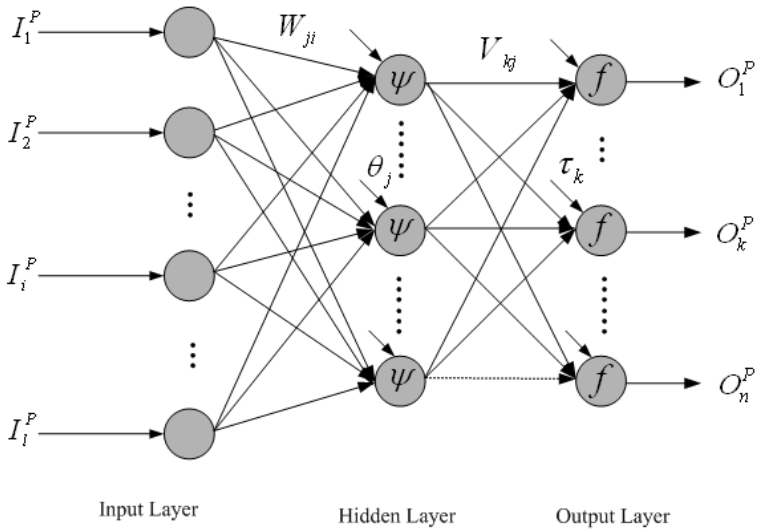


Figure 1. Schematic diagram of Neural network with nine input nodes and one output node

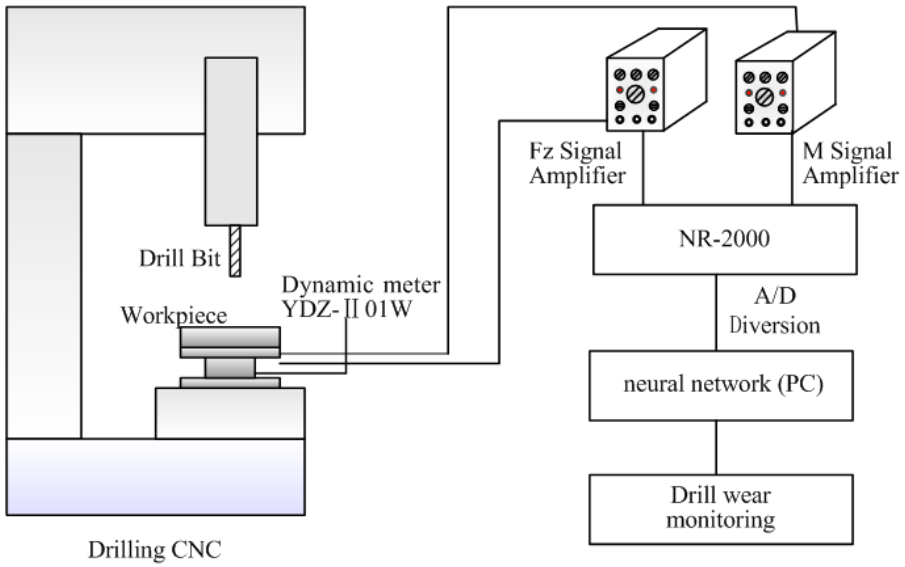
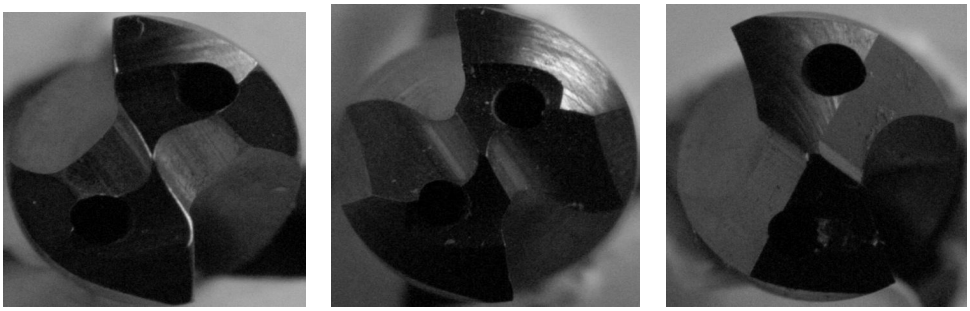
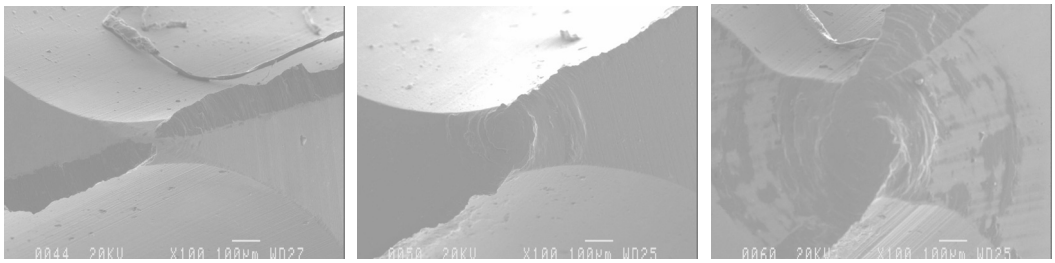


Figure 2. Schematic diagram of the experimental set-up



(a) (b) (c)

Figure 3. Flank of standard and Multifacet drill



(a). (b) (c)

Figure 4. (a) Drill wear of Multifacet Ⅰ at spindle speed 4500 rpm and feed rate 0.1 mm/rev.
(b) Drill wear of Multifacet Ⅱ at spindle speed 4500 rpm and feed rate 0.15mm/rev.
(c) Drill wear of standard at spindle speed 4500 rpm and feed rate 0.2 mm/rev.

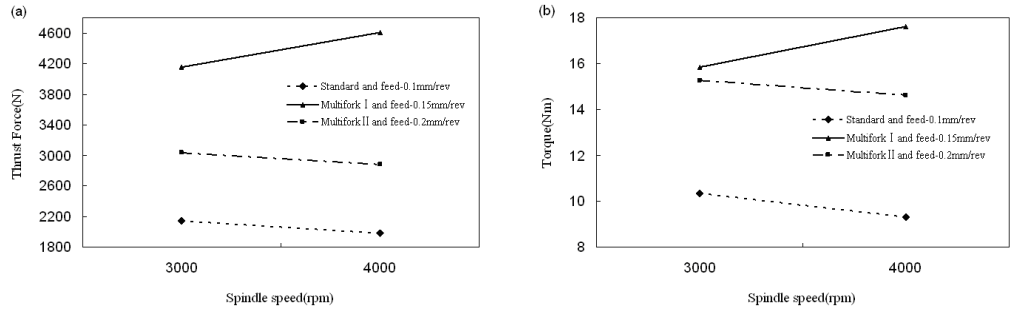


Figure 5. (a) Variation of average thrust force with spindle speed (3000 and 4000rpm) for different feed rates and drills flank. (b) Variation of average torque with spindle speed (3000 and 4000rpm) for different feed rates and drills flank.

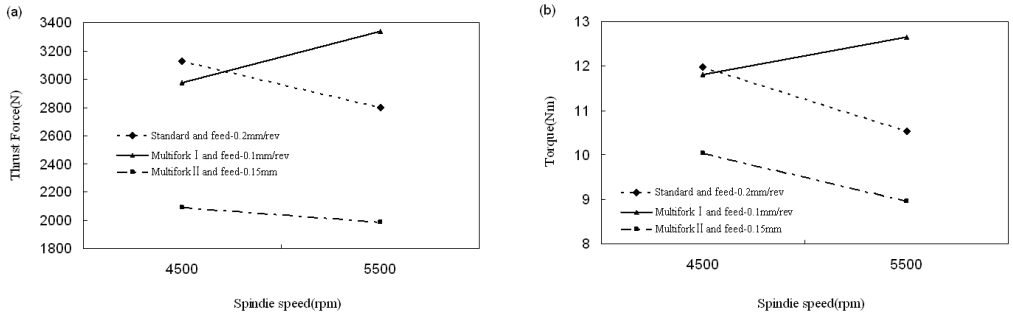


Figure 6. (a) Variation of average thrust force with spindle speed (4500 and 5500rpm) for different feed rates and drills flank. (b) Variation of average torque with spindle speed (4500 and 5500rpm) for different feed rates and drills flank.

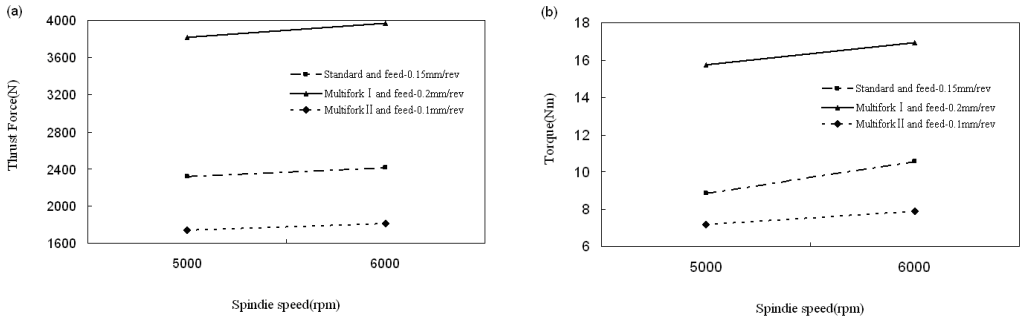


Figure 7. (a) Variation of average thrust force with spindle speed (5000 and 6000rpm) for different feed rates and drills flank. (b) Variation of average torque with spindle speed (5000 and 6000rpm) for different feed rates and drills flank.

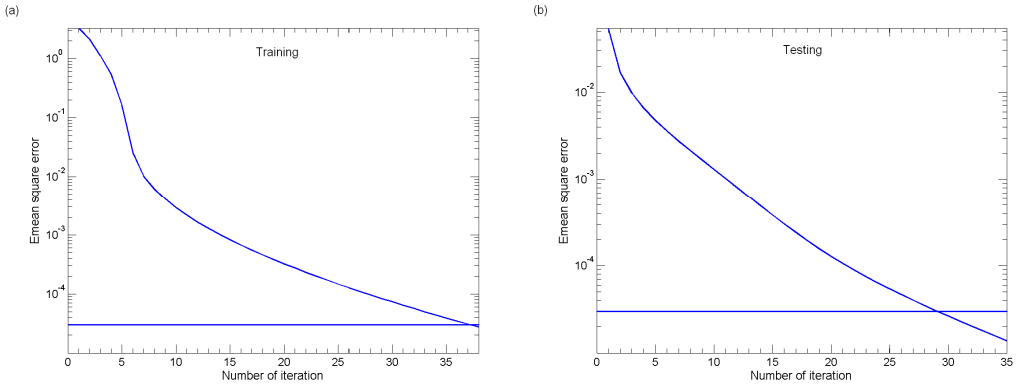


Figure 8 (a). Variation of MSE with number of iteration for the for 5-15-1 neural network with $\alpha = 0.9$ and $\eta = 0.8$ (BPWNN). (b) Variation of MSE with number of iteration for the for 5-15-1 neural network with $\alpha = 0.9$ and $\eta = 0.8$ (BPWNN)

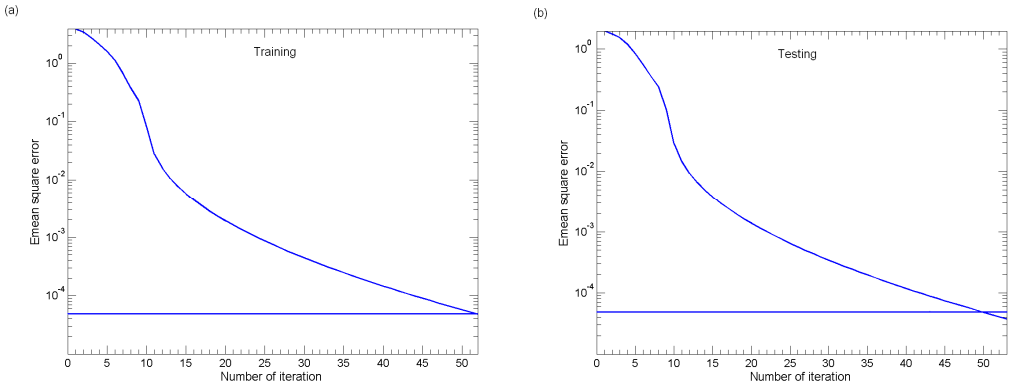


Figure 9(a). Variation of MSE with number of iteration for the for 5-5-1 neural network with $\alpha = 0.9$ and $\eta = 0.8$ (BPNN). (b) Variation of MSE with number of iteration for the for 5-5-1 neural network with $\alpha = 0.9$ and $\eta = 0.8$ (BPNN)

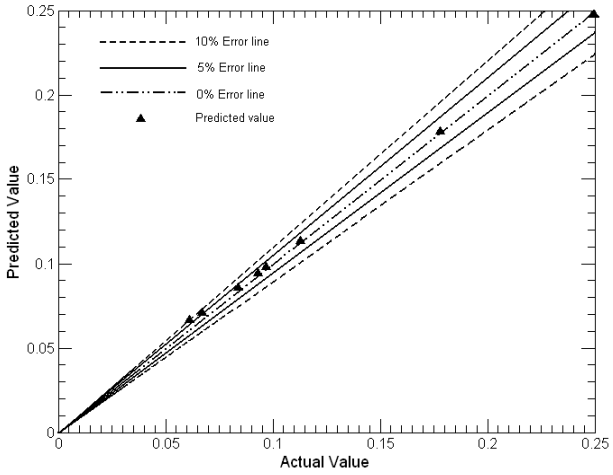


Figure 10. Comparison of predicted values with actual values for 5-15-1 Back propagation wavelet neural network with $\alpha = 0.9$ and $\eta = 0.8$.

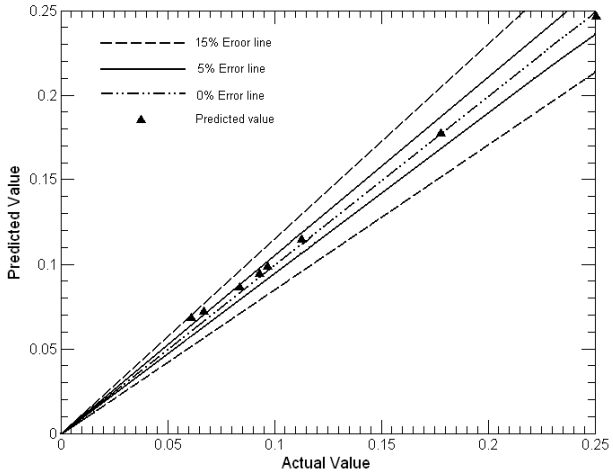
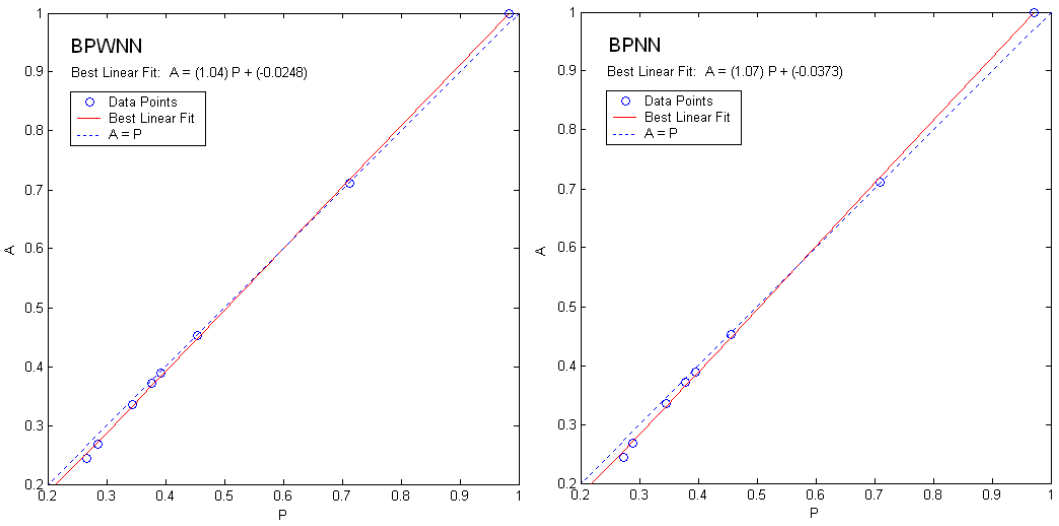


Figure 11. Comparison of predicted values with actuals value for 5-5-1 Back propagation neural network with $\alpha = 0.9$ and $\eta = 0.8$.



(a) (b)

Figure 12. Comparison between the predictions of drills wear by two models, i.e. (a) BPWNN and (b) BPNN

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Study on the NN Decoupling Control System of Air-cushioned Headbox

Zhongjun Xiao

College of Electric and Information Engineering, Shaanxi University of Science & Technology, Xi'an 710021, China

Tel: 86-29-8616-8631 E-mail: xiaodongjun@sust.edu.cn

Mengxiao Wang

Xi Wei Measurement and Control Corporation, Xianyang 712081, China

Tel: 86-29-3357-7920 E-mail: qywjs@163.com

Abstract

The headbox is the key hinge to link the pulp supply system with the sheet forming in the papermaking process. The primary parameters include the total pressure and the stock level which couple each other in the box, and they decide the distribution of the web cross-directional basis weight and influence the paper forming quality. Here taking widely used air-cushion type headbox as the plant, a new neural network (NN) decoupling control system is proposed to overcome the coupling relation between the total pressure and the stock level, decrease the adjusting time as well as to achieve robustness, fault-tolerance and the self-study ability in different environments. The practice has proved that the headbox control system based on NN decoupling algorithm could satisfy the industrial control requirements in the papermaking process well.

Keywords: Papermaking process, Air-cushioned headbox, Neural network (NN), Decoupling control

1. Introduction

The regularity distribution and the sheet quality of the finished paper are decided by the dispersion degree of the fiber in the headbox and the regularity of the jetting pulp in the stock inlet of the headbox, so the design and operation of the headbox become into the key issues in the papermaking process (Tang, 2005, P.129-134). In the work process of the air-cushioned headbox, the main control parameters include the total pressure and the stock level. The intention to control the total pressure is to acquire the even pulp flow velocity and flux jetting from the headbox to the net. The intention to control the stock level is to acquire proper pulp flow area to reduce the changes of crosscurrent and concentration, and produce the controllable overfall to limit the agglutination of the fibers. But the total pressure and the stock level couples each other, at present, to overcome the coupling between the total pressure and the stock level, many control systems such as the fuzzy control, NN decoupling control and self-adaptive control to quickly and exactly control the total pressure and the stock level (Tang, 2006, P.107-114).

According to the characteristics of the headbox and the papermaking process, the NN control algorithm is adopted to exactly control the total pressure and the stock level of the headbox and ensure the web cross-directional basis weight and the quality of the moisture measurement.

NN is the complex network linked by large numbers of nerve cells each other. It could simulate the distributed work characters and the self-organized function of human brain cell, and it possesses very strong abilities such as parallel processing, self-adaptability, self-study and nonlinear mapping, and it can be adapted for the highly uncertain process control system and the nonlinear process control system.

2. Control flow of air-cushioned type headbox

Figure 1 can be used to describe the fourdrinier wet flow of the headbox automatic adjustment system. The stocks from the flow system enter into the headbox through the control valve, and the headbox jets the stocks to the net according to acquired flux and velocity, and the wet sheets are formed through the net dehydration (Wang, 2003).

Figure 1 is the sketch map of the air-cushioned headbox system, and in the figure, *FW* is the water jetting capacity of the headbox, *FAI* is the air-cushioned air inflow of the headbox, *FV* is the spillway discharge of the headbox, *CO* is the

stock outflow concentration of the headbox, and *FR* is the stock inflow capacity of the headbox. In the air-cushioned headbox, the following formula exists between the stock jetting velocity *V* and the total pressure of the headbox *P*.

$$V = \mu\sqrt{2g \cdot P}$$

In the formula, μ is the coefficient about the material property and the form of the head box, and the total pressure *P* includes the air-cushioned pressure (P_{air}) and the static pressure of the stock level (H_{stock}). Therefore, the stock level or the air-cushioned pressure or the total pressure all can be adjusted to adjust the stock jetting velocity. In the total pressure and the stock level, the key is to stabilize the total pressure in order to stabilize the stock velocity. The intention to control the stock level is only to keep the needed flow character in the transmission process of the head box, and under the premise that the total pressure doesn't change, the fluctuation of the stock level in small range is allowed.

So the control application project seen in Figure 2 is proposed. The control of the total pressure is to change the stock adjustment valve by the total pressure controller (PC), i.e. adjusting the stock inflow capacity. The control of the stock level is actualized by the stock level controller (LC) which adjusts the air-cushioned outflow capacity, and the cascade control system composed by the ratio between the stock velocity and the net velocity and the total pressure is added. This control method can not only control the total pressure and the stock level, but can control the ratio between the stock velocity and the net velocity. The emphasis of the control is the adjusted parameter, i.e. the total pressure, and the adjustment capacity is the stock inflow capacity. Therefore, when the adjustment system controls the total pressure well, the flux of the headbox can be stabilized, and the stock inflow capacity must balance with the outflow capacity.

The intention of the headbox control is to acquire stable and even flux and stock jetting velocity of the headbox, and both are always decided by the total pressure when the inlet degree is certain, so the total pressure is the most important and strict parameter in the headbox control. The proper stock level change will not obviously influence the stability and the regularity of the flux of the headbox, i.e. the stock level is the secondary control parameter in the headbox control. Because of the coupling function of the total pressure and the stock level, the general adjustment system can hardly control them in the optimal state simultaneously, so to achieve the control objective of the headbox and the control the total pressure in the optimal state, the stock level should be allowed to change little. As viewed form the stock balance, only the inflow stock capacity and the outflow stock capacity of the headbox achieve balanced, the flux of the headbox can return to the stable state. In other words, the key of the headbox automatic adjustment is to exactly adjust the stock capacity flowing into the headbox and make the capacity equal to the appointed flux of the headbox.

3. NN decoupling control system of air-cushioned headbox

3.1 Coupling characters of total pressure and stock level

In the control process of the total pressure and the stock level in the headbox, when any one of the values which control the stock inlet capacity and the air-cushioned flux acts, both the total pressure and the stock level will change. In one experiment, put the value which controls the air-cushioned air flux on the manual state, and change the stock inlet capacity, when the total pressure ascends from 15.2KPa to 15.96KPa, the stock level ascends from 27.0cm to 33.9cm, which indicates that the total pressure is associated and coupled with the stock level control loop. If the decoupling control is not implemented, the total pressure and the stock level can not achieve the optimal state simultaneously, which will make the character of the net sheet be influenced, and influence the papermaking quality of the paper. Therefore, to achieve optimal papermaking condition, it is necessary to implement decoupling control of the total pressure and the stock level in the headbox.

3.2 Characteristics of NN decoupler

There are strong coupling phenomena in many multi-variable control systems, and if certain decoupling measure is not adopted, the system can not acquire satisfactory control effect, even can not be controlled. In many situations, the decoupling equipments are very complex, and even they are not implemented in the physical practice. In addition, to implement decoupling, the mathematical model of the controlled process must be known in advance, and even if that can be done, the work of the system will not be ideal because of the nicety of the model, the nonlinearity and the time-varying of the system. NN possesses the self-study ability and the nonlinear mapping ability, so it can solve above problems well.

The decoupling of the system is divided into the static decoupling and the dynamic decoupling. In practice, the static decoupling is generally adopted, because the compensation designed by the dynamic decoupling can not be implemented, or can be implemented too complexly. When the system has time-varying character or obvious nonlinearity, it is impossible to realize the static decoupling, because the decoupling stage is required to possess the self-adaptive ability. As a new nonlinear system description method, NN can realize complex function by the composite of simple nonlinear functions. The trainability of NN and the commonality of the structure make NN possess self-adaptive ability to fulfill the requirement of decoupling. Therefore, if NN is used in locale decoupling, it can overcome the nonlinearity and the time-varying character of the system well, break past unchangeable decoupling mode

and enhance the decoupling precision of the system, so NN will the development direction in the future decoupling of control process as the decoupler.

3.3 Headbox NN decoupling control project

In the total pressure and stock level control system, according to the characteristics of NN, the feed forward NN is selected to develop and study as the decoupler.

Different connections between NN and controlled objective can compose different NN decoupling compensation projects. In the actual system, the pressure adjustment system stabilizes the total pressure by the stock inlet capacity adjusted by the stock pump, and the stock level adjustment system stabilizes the stock level by the electric conduction adjuster controlling the air-cushioned air flux. The decoupler is put in front of the controller or in the feedback loop. In the article, the decoupler is put in front of the controller. For the paired input and output system of the headbox, the control frame chart is seen in Figure 3.

4. NN decoupling controller

4.1 NN decoupling algorithm

The indexes adopted by the NN decoupling algorithm include the generalized objective Bristol first coefficient matrix containing the NN compensation, i.e. each sensitivity matrix of controlled variable about the deviation, and here, the paired input and output system is discussed as follows (Shen, 2004, P.151-156).

$$\mathbf{B}=\left[\frac{\partial Y}{\partial E}\right]=\frac{\partial(y_1,y_2)}{\partial(e_1,e_2)}=\begin{bmatrix}\frac{\partial y_1}{\partial e_1} & \frac{\partial y_1}{\partial e_2}\\ \frac{\partial y_2}{\partial e_1} & \frac{\partial y_2}{\partial e_2}\end{bmatrix} \tag{1}$$

If the generalized objective is decoupling, so the matrix B can be turned into the diagonal matrix through the line or row transformation.

The decoupling train objective is

$$\min_{W}\mathrm{E}=\frac{1}{2}\sum_{i=1}^{2}[B_i^{[s]}-\frac{\partial y_i(W)}{\partial \mathrm{E}_i}]^2+\frac{1}{2}\sum_{\substack{i=1\\ i\neq j}}^{2}[\frac{\partial y_i(W)}{\partial \mathrm{E}_j}]^2 \tag{2}$$

In above formula, $B_i^{[s]}$ is the generalized objective plus of the decoupling system including the plus of the adjuster, and W is the training parameter of NN.

The training algorithm of NN adopts the grads algorithm.

$$\Delta W=-\eta\frac{\partial E(W)}{\partial W} \tag{3}$$

η is the learning step length, and when η is small enough, the enough more reasonable interior units should be selected, and the function could make $E(W)$ is differentiable for W, and the grades algorithm is always convergent.

Note K_{ij}^{Q} is the static plus of the system stage Q between the i'th input and the j'th input, and Q can be any one of R (controller), P (controlled object) or N (NN). From (1), (2) and (3),

$$\frac{\partial E(W)}{\partial W}=\sum_{i=1}^{2}[B_i^{[s]}-\frac{\partial}{\partial e_i}y_i(W)]\frac{\partial^2 y_i(W)}{\partial W\partial e_i}+\sum_{\substack{i=1\\ i\neq j}}^{2}(\frac{\partial}{\partial e_j}y_i(W))\frac{\partial^2 y_i(W)}{\partial W\partial e_j} \tag{4}$$

$$\frac{\partial y_i(W)}{\partial e_i}=\frac{\partial y_i(W)}{\partial u_i(W)}\frac{\partial u_i(W)}{\partial L_i(W)}\frac{\partial L_i(W)}{\partial e_i}=K_{ii}^{P}K_{ii}^{R}K_{ii}^{N}(W) \tag{5}$$

$$\frac{\partial y_i(W)}{\partial e_j}=\frac{\partial y_i(W)}{\partial u_i(W)}\frac{\partial u_i(W)}{\partial L_i(W)}\frac{\partial L_i(W)}{\partial e_j}=K_{ii}^{P}K_{ii}^{R}K_{ij}^{N}(W) \tag{6}$$

So,

$$\frac{\partial^2 y_i(W)}{\partial W\partial e_i}=\frac{\partial}{\partial W}[K_{ii}^{P}K_{ii}^{R}K_{ii}^{N}(W)]=K_{ii}^{P}K_{ii}^{R}\frac{\partial K_{ii}^{N}(W)}{\partial W}=K_{ii}^{P}K_{ii}^{R}T_{ii}(W) \tag{7}$$

$$\frac{\partial^2 y_i(W)}{\partial W \partial e_j} = \frac{\partial}{\partial W}[K_{ii}^P K_{ii}^R K_{ij}^N(W)]$$
$$= K_{ii}^P K_{ii}^R \frac{\partial K_{ij}^N(W)}{\partial W} = K_{ii}^P K_{ii}^R T_{ij}(W) \tag{8}$$

Unite (7) and (8),

$$\frac{\partial E(W)}{\partial W} = \sum_{i=1}^{2}[B_i^{[s]} - K_{ii}^P K_{ii}^R K_{ii}^N(W)]K_{ii}^P K_{ii}^R T_{ii}(W)$$
$$+ \sum_{\substack{i=1 \\ i \neq j}}^{2}(K_{ii}^P K_{ii}^R K_{ij}^N(W))K_{ii}^P K_{ii}^R T_{ij}(W) \tag{9}$$

If the decoupler form is known, the above algorithm can be used to acquire the structure of NN decoupler.

4.2 NN decoupling compensator design

The decoupler adopts the multi-layer feed forward NN (Li, 2007, P.107-114). Because the control system has the paired input and out structure, so there are 2 input and output nodes in the network. The quantity of the hidden node should exceed the quantity of input node, and in the simulation, the system will find the proper network node quantity.

First, set up a few nodes to train the network, and test the approaching error of the network, and then gradually increase the quantity of node until the test error doesn't decrease obviously. If the node quantity is too more when above method is adopted, two hidden layers can be adopted, and the total quantity of the node in two hidden layers usually is less than the quantity of the node in one hidden layer.

In formula (9), K_{ii}^R and K_{ii}^P respectively are the static pluses between the i'th input and the j'th input of of *R* (controller) and *P* (controlled object). K_{ii}^R can be obtained directly from the controller, and K_{ii}^P can be measured by the experiment. In the feed forward three-layer NN, the input functions of various layers all adopt the logsig function (Zhong, 2007), and the network form is seen in Figure 4.

The weight matrixes of three layers respectively are $\mathbf{W}_1$ (2×2), $\mathbf{W}_2$ $(S\times 2)$ and $\mathbf{W}_3$ $(2\times S)$. Suppose the inputs of three layers respectively are $\mathbf{O}_1$, $\mathbf{O}_2$ and $\mathbf{L}$, the inputs of three layers respectively are $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$.

$$\mathbf{X}_1 = \begin{bmatrix} X_1(1) \\ X_1(2) \end{bmatrix}, \quad \mathbf{X_2} = \begin{bmatrix} X_2(1) \\ \vdots \\ X_2(S) \end{bmatrix}, \quad \mathbf{X}_3 = \begin{bmatrix} X(1) \\ X(2) \end{bmatrix}$$

$$\mathbf{O}_1 = \begin{bmatrix} O_1(1) \\ O_1(2) \end{bmatrix}, \quad \mathbf{O}_2 = \begin{bmatrix} O_2(1) \\ \vdots \\ O_2(S) \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} L(1) \\ L(2) \end{bmatrix}$$

From the analysis,

$$\mathbf{X}_1(j) = \sum_{i=1}^{2} e_i W_1(j,i) \qquad (j = 1,2) \tag{10}$$

$$\mathbf{X}_2(j) = \sum_{i=1}^{2} O_1(i) W_2(j,i) \qquad (j = 1,\cdots,S) \tag{11}$$

$$\mathbf{X}_3(j) = \sum_{i=1}^{S} O_2(i) W_3(j,i) \qquad (j = 1,2) \tag{12}$$

$$\mathbf{O}_1(j) = f(X_1(j)) \qquad (j = 1,2) \tag{13}$$

$$\mathbf{O}_2(j) = f(X_2(j)) \qquad (j = 1,\cdots,S) \tag{14}$$

$$\mathbf{L}(j) = f(X_3(j)) \qquad (j = 1,2) \tag{15}$$

The input and output function of various nodes is

$$y = f(x) = \frac{1}{1+e^{-x}} \tag{16}$$

So, $\frac{dy}{dx} = y(1-y)$, and $K_{ii}^{N}(W)$, $K_{ij}^{N}(W)$ and the one-order partial derivatives to $W_1(i,j), W_2(i,j), W_3(i,j)$ can be calculated.

Replace above formulas into (9) and (3), ΔW can be obtained, and the NN decoupler is

$$W = W + \Delta W \tag{17}$$

Computer the objective function (2), when $E < \varepsilon$ (very small positive number), end the iteration process.

5. Decoupling control simulation of total pressure and stock level

By the method of the mechanism modeling or the model distinguish (Gao, 2003), the one-order inertial model about the total pressure and the stock level of the headbox can be simplified.

$$\begin{bmatrix} P(s) \\ H(s) \end{bmatrix} = \begin{bmatrix} \frac{1.9}{36s+1} & \frac{0.7}{27s+1} \\ \frac{0.9}{36s+1} & -\frac{1.5}{27s+1} \end{bmatrix} \begin{bmatrix} u_F(s) \\ u_P(s) \end{bmatrix} \tag{18}$$

The parameters of both controllers respectively are

The control total pressure of the stock inlet capacity adjustment, $PI:\ K_p = 0.73, K_I = 0.03$.

The control stock level of the air-cushioned pressure adjustment, $PI:\ K_p = -0.8, K_I = -0.043$.

Use above algorithm to train the NN decoupler, and when the amount of the hidden node is 8, the learning step length is 0.001 and the training times are about 60000, the proper decoupler W can be obtained. Use W in the decoupling control, the simulation Figure 5 can be obtained.

From Figure 5 (a), the influences between total pressure and the stock level influence are large before decoupling, because the fluctuation would enter into the stable state after a certain period because of the coupling function of the system in the initial period. Figure 5 (b) indicates that the system control ability increases after decoupling, the coupling function between the total pressure and the stock level is almost overcomes, and the time entering into the stable state is reduced.

6. Conclusions

The analysis in the article indicates that the control ability of the system increases obviously after the headbox uses the NN decoupling control system. The coupling function between the total pressure and the stock level is almost overcomes, and the time entering into the stable state is reduced. So the decoupler composed by the feed forward NN has better decoupling ability.

In actual application, when the sheet type changes, the character of the pulp will change largely and the work point will move. NN possesses the trainability, and the train of NN can be implemented in the upper computer which can make NN re-modified with manual instruction and transmit the system parameters to lower computers, so NN can be applied in the system decoupling to exert the trainability and the adaptability of the network. But the training time of the network is too long and the training process needs human participation, so the real-time character of NN will be reduced. Therefore, the better training algorithm or decoupling algorithm should be sought, and to further study the self-adaptive decoupling of NN will more make for the use and push of NN as the decoupler.

References

Gao, Juan. (2003). *Artificial Neural Network Principle and Simulation Examples*. Beijing: China Machine Press.

Li, Minghui & Li, Yan. (2007). Decoupling and Time-delay Control Strategy of Basis Weight and Moisture Content Measurement in Papermaking. *Transactions of China Pulp and Paper*. No. 22(4). P. 107-114.

Shen, Guojiang, Hu, Dan & Sun, Youxian. (2004). Modeling and Control of Basis Weight and Moisture in Papermaking Process Using Artificial Neural Networks. *Transactions of China Pulp and Paper*. No. 19(1). P. 151-156.

Tang, Wei, Wang, Mengxiao & Li, Minghui. (2005). The Summarize about the Control of Headbox of Paper Machine. *Journal of Shaanxi University of Science and Technology (Natural Science Edition)*. No. 23(6). P. 129-134.

Tang, Wei, Wang, Mengxiao, Li, Minghui & Xue, Huijian. (2006). The Advanced Control Strategies and Decoupling Algorithms of Headbox. *Transactions of China Pulp and Paper*. No. 21(1). P. 107-114.

Wang, Mengxiao, Sun, Yu & Tang, Wei. (2003). *Measurement and Control System and Engineering for the Pulp and Papermaking process*. Beijing: Chemical Industry Press.

Zhong, Luo & Rao, Wenbi. (2007). *Artificial Neural Network and Its Integrated Application Technology*. Beijing: Science Press.

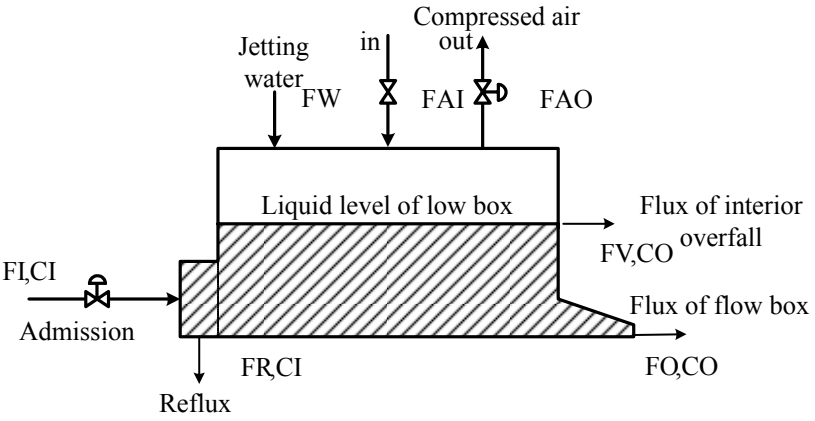


Figure 1. Sketch Map of Air-cushioned Type Headbox System

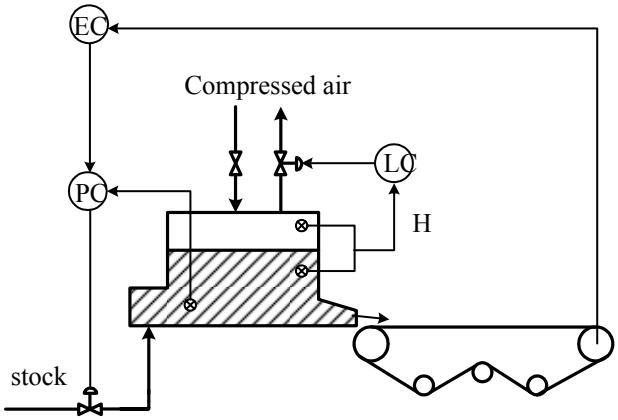


Figure 2. Adjustment Project of the Headbox

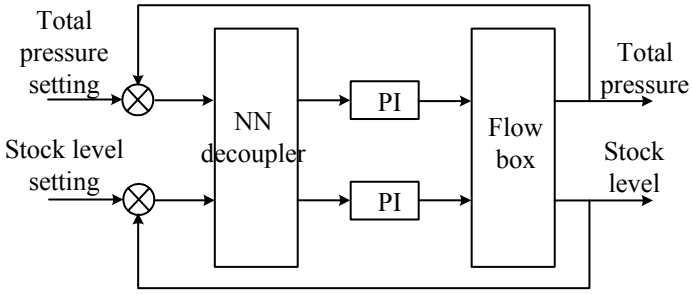


Figure 3. NN Decoupling Control Project

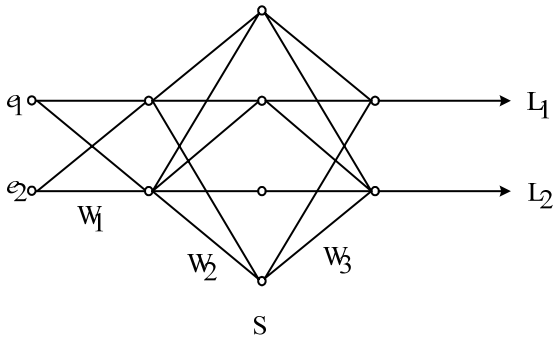
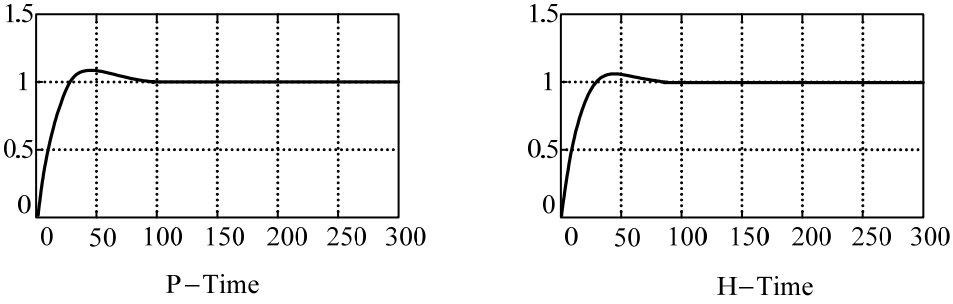
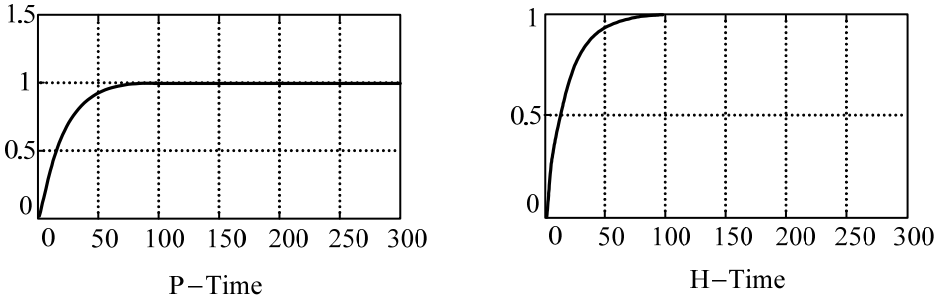


Figure 4. Three-layer NN Structure



(a) Simulation of total pressure (P) and Stock level (H) before decoupling



(b) Simulation of total pressure (P) and Stock level (H) after decoupling

Figure 5. Total Pressure and Stock Level Control System Simulation of the Headbox

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Security and Management of Local Computer Data

Ruicheng Li

Shandong Youth Administrative Cadre' College, Jinan 250103, China

Tel: 86-531-85857405 E-mail: LRC04@163.COM

Abstract

In the computer system, only the data storing in the computer are the real fortune, and the protection of information and data are crucial. In the day that data security becomes more and more important, the stable and self-contained backup system has been the key to ensure the normal running of the system. Data backup is not only the conservation of data, but includes more important content, i.e. the management, or else, the data will be lost even the whole hard disk will be dormant, and hell and gone losses will be induced.

Keywords: Computer system, Data security, Management, Backup

1. Introduction

Data security has been more and more important for the computer system, and there are two reasons which may induce data loss, i.e. the software fault and the hardware fault. Once the important data are destroyed or lost, it will induce huge influences even hell and gone losses. Large-sized data reconstruction needs long time and expensive costs, and the only effective method to reduce the loss minimum is to make the data backup.

The security backup of data has been developed for a long time, and when most domestic enterprises make the network layout and design, they always have not pay enough attention to the data storage and backup management which are proposed rarely even ignored completely sometimes in the design. When the network is established, because of the deficiency of reliable data protection, once the accident occurs, the reparation is too late (Fan, 2006).

Therefore, except for the consciousness, we should pay more attention to the storage and backup of data in practice, and regard the data storage and backup as the first important thing, and adopt advanced data storage backup equipment to ensure the security of the network data in the possible range. The facts have proved that only complete data storage and backup will offer perfectly safe data security and protection for people.

There are two meanings of the data security. The one is the logic security, for example, to prevent the damage from virus and hacking, and the other is the physical security, for example, man-made mistake or irresistible disaster. The former needs the safety protection of the system, and the latter needs the protection of data storage and backup or disaster tolerance.

2. Analysis of reasons (Yu, 2006)

2.1 Virus and other malicious programs

The most familiar malicious program is virus which is the unauthorized code inserting in the legal program. After the computer infects virus, some programs or data can not be used, and usually we can use then antivirus software to kill viruses for the computer, and all problems will be solved. But sometimes the data losses induced by the malicious programs are very serious, and relative to the damage of the data files, the prevention and killing of viruses are always lagged.

2.2 Other maliciously damages by people

In other maliciously damages by people such as the hacking, hacker can use many measures such as the filching and cracking to possess enough operation right, and any data protection of the system will be not safe any longer. Only depending on normal deleting, moving, formatting and other operations can also destroy the data. Some users only create necessary password for their own user names, but for the "Administrator" with the highest limitation, the password is blank, so the any illegal user connected with the network can implement any operations in the system.

2.3 Hardware fault

Hardware fault is one of biggest reasons of data loss. The hardware fault and the unstable pressure always will make the computer automatically restarted. The hardware fault is the most serious problem, and it includes the physical damage, being stolen and so on, and the possibility of the data recovery will be zero.

2.4 Incorrect manipulation

The incorrect manipulations mean some destroying behaviors induced by unfamiliar computer operations, misunderstandings of clue information from the system. For example, quit when the files are not saved, the files or the corrections of the files will be lost, or take out or replace the U disk when editing the files in the U disk, which will induce the editing files destroyed or the files in the new U disk destroyed, or incorrectly deleting, or incorrectly formatting and so on.

3. Hard disk structure and disk storage principle

To know the data structure of the hard disk and the storage principle of files can help use understand and grasp the data recovery technology.

3.1 Data structure of hard disk (Liu, 2007 & Dai, 2003)

The hard disk is generally divided into five parts, i.e. the master boot sector and the DOS boot record, the file allocation table, the directory area and the data area. The master boot sector shares 512 bytes, and it includes the main boot record (MBR) and disk partition table (DPT). The function of MBR is to check whether the DPT is correct and which sector is the boot sector, and transfer the start program of this sector into the memory and exert it when the program ends.

Dos boot record (DBR) is the first sector which the operation system can be directly interviewed, and it includes the boot program and the sector parameter block which is called as BPB (BIOS Parameter Block). The main task of the boot program is to judge whether the former two files in the root directory are the root files of the operation system, read the first file into the EMS memory, and give the control right to the file. BPB parameter records many important parameters such as the start sector, the end sector, the file storage format, the disk medium descriptor, the size of root directory, the amount of FAT, the size of allocation unit (cluster).

FAT (File Allocation Table) is used to indicate the allocation of various clusters and cluster chains occupied by one file, and mark the bad cluster and the usable cluster. There are two FATs in the disk, and the first one is the basic table, and the second FAT is the backup of the first FAT, and their sizes are decided by the size of the sector and the size of the file allocation unit. DIR (or FDT, i.e. File Directory Table) records the start unit and the file attribute of every file (directory). Data area is the real data storage area, and it occupies most space of the disk. Data in the Data area are explained by FDT and FAT, and if FDT and FAT describe the Data area as the “unused”, the corresponding Data area is the “unoccupied” free space which can be written by new data.

3.2 Data storage principle

When the operation system saves the files, it first find the free space in the FDT to write relative information including the file name, the file size and the created time, and find the free space in the Data area to save the files, and write the number of the first cluster of the flies in the Data area into FDT, or write the number of the last cluster in the Data area if the files end, and write the end mark in the last cluster of Data area. When reading the file, the operation system reads the file name, filename extension, file size, data modification, and the cluster number of the first cluster saved in the Data area and other information about the file in FDT, and read the data of corresponding cluster number in the Data area according to the cluster number in FDT, and find the corresponding unit of FAT, and if the content is the file end mark, the file ends, or else, save the cluster number of next cluster of data, and in this way, repeat the process until the file end mark occurs.

Through understanding the influences of the data storage structure and various operations to the disk data, we can understand why we can find the data back when the data are destroyed and there are no new data covers, which is the possibility of data recovery.

4. Usual data fault treatment methods

Data recovery can be divided into the pure software recovery and the recovery with the combination of software and hardware. There are many single data recovery software such as Easy Recovery, Disk Genius and so on. Different software programs have different advantages and disadvantages, and we can reasonably select and use them according to the practical situations.

4.1 The system can not find the hard disk

Most these faults occur in the connection cable or the port of IDE, and the fault rate of the disk is rare, but it will induce the system can not be started from the disk, and even can not enter into the disk C when starting from disk A, and the automatic test function of the CMOS can not find the disk too. We can find the fault by the replacement experiments such as re-plugging or re-pulling the disk cable or exchanging the IDE port and cable.

4.2 Recovery of incorrect data deletion

Start Easy Recovery, click the “data recovery”, and enter into the main menu of data recovery. Select “deleting the recovery”, and the deleted files can be recovered. Select the sector where the files are in, and if only recovering one or

two files, the quick scanning in the fault should be selected, and if recovering the whole directory containing subdirectory and files, the complete scanning should be selected. Click the "next", the system will scan the selected directory. When the scanning finishes, all recovered file information will be displayed on the screen. We can select the files what we want to recover like using "the resource manager", and click the "next". Because Easy Recovery needs not to reread the disk when recovering the data, the software only mirrors the FAT and the directory tables in the EMS memory, all recovered files are stored in the EMS memory, so we should select the place storing these files and write the recovered files in the memory on the disk.

4.3 Data recovery after formatting

If the files are deleted because of incorrect manipulation, we need to select "formatting recovery", and other operations are similar with the "deleting recovery". For the instance which the partition table is destroyed, Easy Recovery cannot recover the partition table information in disk, but scan the disk according to the cluster, and put the recovered files into different folders by the file type, which provides a sort of new data recovery method for us. When the partition table of disk is destroyed seriously, and we cannot use other recovery software to recover it, we can use this method. But for some big files, because they may be stored in multiple discontinuous clusters, and if the data are not recovered by the partition table, the recovered files may be incomplete. So we should first recover the disk partition table possibly and then recover the data.

5. Conclusions

Any one data recovery solution can not ensure to recover all data. To really protect the data, the most important work nips in the bud. The daily works include installing the anti-virus software and firewall and updating in time to prevent the invasion of viruses, maintaining the computer in time to prevent the system halted, avoiding the file loss induced by incorrect manipulation and other man-made factors, using disk defragment usually to make the data in the Data area continually stored possibly and largely enhance the success rate of data recovery, and especially paying attention to the effective backup of important files which is relatively simple and reliable.

References

Dai, Shijian & Chen, Yonghong. (2003). *Technology of Restoring DB*. Beijing: Publishing House of Electronics Industry. August of 2003.

Fan, Hong. (2006). Briefly Description of the Information Security Risk Evaluation Standards. *Netinfo Security,* No. 1.

Liu, Sanman. (2007). Analysis on Data Restore Technology of Computer. *Shanxi Electronic Technology,* No.1.

Yu, Jinyun & Luo, Yixin. (2006). Current Status of China's Information Security and Study on Countermeasures. *China Safety Science Journal,* No.1.