Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Leveraging an Open Source VPN Technology to End User

Farok Azmat

Faculty of Information Technology and Quantitative Sciences

MARA University of Technology (UiTM)

40450 Shah Alam, Selangor, MALAYSIA

Tel: 60-3-5521-1148 E-mail: farok@tmsk.uitm.edu.my

Md Ikhwan Hamdan

Faculty of Information Technology and Quantitative Sciences

MARA University of Technology (UiTM)

40450 Shah Alam, Selangor, MALAYSIA

Tel: 60-3-5521-1148 E-mail: zegxro@gmail.com

Hj. Kamaruzaman Jusoff (Corresponding author)

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

Tel: 60-3-8946-7176 E-mail: kjusoff@yahoo.com

Abstract

Virtual Private Network (VPN) can be implemented by corporate and public user in a few methods. The methods included proprietary or open source solutions. It has shown great potential due to its encrypted traffic flow which gives it a secure environment. With a good reception of internet in Malaysia it is unfortunate that the major Internet Service Providers (ISPs) have imposed packet filtering. In effect they have throttled the fast broadband to become a slow medium. As a result it gives a disadvantage to active end user. To overcome an ISP packet filtering which slows down internet traffic, this paper intends to study and provide a sample of VPN solution service to end user to improve the current problem.

Keywords: BitTorrent, Broadband, File sharing, ISP throttling, Open Source, OpenVPN, P2P, VPN

1. Introduction

The Peer-to-Peer (P2P) network is evolving. (Normalkid, 2003) The first generation started by Napster where a client connect to a server to make a search and later downloading is done between peer and peer without server intervention.

The second generation is where peer and peer connect together directly without involving a central server. Therefore there is no particular center point to track this network. It is truly distributed network.

The third generation of P2P improves further in that any client no matter it is Kazaa, Grokster, Morpheus etc although they are different clients they access the exact same file throughout the internet. The enhancement made by this FastTrack protocol is that it includes supernode and spawning which improve searches and download speed.

The fourth generation is dominated by BitTorrent where big files are divided into small pieces of 64kB until 4mB where it does not implied much internet bandwidth to the seeder. Files are shared between peers. The pieces may not be in sequence.

2. Challenges of ISP

With the new phenomenon of file sharing where upstream and downstream traffic have increased ISP has to look into its broadband infrastructure. Most of the existing ISP is not optimized for “broadband aware” traffic. (Cisco system, undated) When this fact is not observed ultimately it will cause network congestion and performance drops. It is also observed that the common pitfall of ISP when handling the increase in traffic by file sharing is caused by:

(1) The lack of physical attributes of its cable infrastructure. To withstand the new environment to sustain the increase

in upstream network traffic it needs a high speed data quality cables such as fiber optic nodes.

(2) It also noted that file sharing traffic requires expensive network access point especially on the international link not so to the local network.

With the above facts, it also suggested that the ISP can no longer ignore the current traffic patent used by users. Consequently ISP should start looking at planning and infrastructure upgrade.

3. Bandwidth throttling

It is found that 35% of all internet traffic is used by BitTorrent while the rest are used to distribute all matters of files, distributions of new application and system backup. (Broadbandreports.com, 2004)

Due to the P2P activities most universities have quietly imposed packet filtering using solution such as Ellacoya switch which has captured about US$200 million market. Other ISP such as Shaw Network (Canada) has also used the same tactic to regain some of their "lost" bandwidth.

4. Bandwidth throttling in Malaysia

As in Malaysia the first encounter of packet filtering is felt when Maxis imposed in its *"Terms and Conditions"* a statement which states that an automatic disconnection to the customer's internet connection is imposed after a certain period of inactivity. (Asohan, 2007) The period may vary from 20 to 30 minute of inactivity. Another means of blocking by Maxis is to limit customer total usage per month NOT exceeding 3GB of data volume per customer.

It is also believed that another ISP, TM Net the provider of Streamyx broadband service has also imposed a similar tactic since February 2006 but it was greatly felt sometimes mid-September 2006. When customer complaint about the slowness of Streamyx TM Net conceded that 20% of its customer uses P2P services and causes the maximum utilization of its network capacity but it still hold back its upgrade to its network infrastructure to a later requirement. (Teoh, 2006)

5. Managing heavy traffic

According to the Cisco systems white paper, (Cisco system, undated) ISP should realize that there are factors to be taken into account when considering planning network bandwidth. It gives the following facts and suggestions as in Table I.

<<Table 1. Factor for planning network bandwidth (Source: Cisco systems white paper)>>

The paper also highlighted that the present major selling attraction to ISP is the broadband. It should also realize that the current technology has deviated from traditional "client-server" model to a multiple "client-server" model as shown by most sharing program protocols. They are: KaZaA, Gnutella, Winny, WinMX, eDonkey, BitTorrent, DirectConnect, Manolito, Kuro, Soulseek, Filetopia, iTunes, Napster, Waste, Mute, Share etc.

Moreover, the present internet technology has advanced from its traditional 'client-server' model which is also called Content Delivery Networks (CDN) and it has up-grown to a hybrid CDN which consumes more bandwidth. With this hybrid CDN as shown in the file sharing technology and high quality video delivering, ISP should take these into account when designing its infrastructure. (Klinker, 2008)

Thus based on the above factors the ISP can easily analyze, manage and implement its infrastructure to its benefits or it's Return on Investment (ROI) will not be achieved.

6. Practical method of bypassing ISP throttling

To address the traffic shaping by TM Net Streamyx, one possible solution is to implement OpenVPN technology. There are two components involved. One is the VPN server and the other is the client.

The main question here is where will the VPN server be located?

Basically Figure 1 illustrates the implemented location of VPN server

<<Figure 1. Location of The Implemented VPN Server>>

The Figure 1 above shows the implemented VPN server. It is located before the 10Gb International Line backbone to Malaysia (TM, 2007) by hiring a reasonable price at a site located in US. The VPN server is actually on the high speed backbone line. Bear in mind although the server is on the very high speed backbone the user will be expected a fully 1Mb/s speed. Thus the effect of using VPN technology:

(1) All applications used by the client will pass through the VPN server.

(2) Due to the unblocked port as granted by the ISP, the speed of transaction is available at a considerably higher speed as compared to the previous one. See Figure 2 and Figure 3 for the comparison between P2P downloading without and with VPN respectively.

<<Figure 2. Downloads Performance without VPN>>

<<Figure 3. Downloads Performance with VPN>>

The Figure 2 above shows the speed of normal downloading without VPN. It is 9.5kByte/s. This shows that TMNET is imposing the bandwidth throttling. While Figure 3 shows the downloading speed with VPN. It has improved to 108kByte/s. Therefore it shows that VPN is a valid method to circumvent the bandwidth throttling done by TMNET Streamyx.

A simple check is required here. (helpwithpcs.com, 2008) If the user bandwidth provided by TMNET Streamyx is 1Mbit/s it can be written as:

$$S = \frac{BW(bit/s)}{8} \tag{1}$$

Where:

$S =$ speed in Byte/s

$BW(bit/s) = bandwidth$

In this case applying Equation (1) to the Streamyx bandwidth of 1Mbit/s becomes:

$$S = \frac{1M(bit/s)}{8}$$

$$S = 125kByte/s\#$$

Comparing this ideal case to the speed obtained from the exercise which is 108kByte/s the speed is most appreciated.

7. Simple equations for user

The download time for any given file size can be determined by a general equation given as below:

$$T(\text{sec}) = \frac{F(MByte \times 8)}{BW\left(\frac{Mbit}{s}\right)} \tag{2}$$

Now consider an ideal condition as follows.

Example 1

Let the file size be

$$F = 1MByte$$

Let the broadband bandwidth be:

$$BW = \frac{1Mbit}{s}$$

Note that the bandwidth of the line is also normally known as the throughput of the line.

Then using Equation (2) the time to download a given file is:

$$T(\text{sec}) = \frac{F(MByte \times 8)}{BW\left(\frac{Mbit}{s}\right)}$$

$$T(\text{sec}) = \frac{1(MByte \times 8)}{1\left(\frac{Mbit}{s}\right)}$$

$$T(\text{sec}) = \frac{8Mbit}{\frac{1Mbit}{s}}$$

$$\underline{T(\text{sec}) = 8\ sec}\ \#$$

Therefore the time to download a 1MByte of data in a 1Mbit/s line is in 8 sec.

Example 2

Let's consider another example

Let the file size be:

$$F = 1GByte$$

Let the broadband bandwidth be:

$$BW = \frac{1Mbit}{s}$$

Then using Equation (2) again the time to download the given file is:

$$T(sec) = \frac{1(GByte \times 8)}{1(\frac{Mbit}{s})}$$

$$T(sec) = \frac{1(10^9 Byte \times 8)}{1(\frac{10^6 bit}{s})}$$

$$T(sec) = 8000sec$$

$$T(min) = \frac{8000}{60}$$

$$T(min) = 133.33sec$$

$$T(hr) = \frac{133.33}{60}$$

$$T(hr) = 2.2\ hour$$

$$\underline{T(hr) = 2\ hr\ 12\ min}\ \#$$

Therefore the time to download a 1GByte of data in a 1Mbit/s line is in 2 hour 12 min.

But the above formula is only valid for an ideal download through a 1Mbit/s broadband line. And in actual fact there exist a time interval between when a network station seeking access to a transmission channel and when the access is granted or received. This amount of time elapsed between receiving and retransmitting of the LAN packet is called latency. (Themanualpage.org, 2002)

Latency also depends on the speed of the transmission medium such as the copper wire used, optical fiber or radio wave. It also includes the delay of the transmission devices such as modem and router. Thus a low latency figure indicates high network efficiency.

Therefore in a real case there are a number of factors to be considered such as interference of data which caused delay, the quality of telephone line and distance of user from substation. These factors or latency will degrade the download traffic to 260ms or 26% of the actual speed. (Cheshire, 1996)

From the above argument the real download time in hours considering the worst case of 30%, the Equation (2) becomes:

$$T(\text{sec}) = \frac{F(MByte \times 8)}{BW(\frac{Mbit}{s})} \times (100 - 30)\% \quad (3)$$

Applying the Equation (3) to the result obtained in Example 2 the time of 2 hour 12 min becomes 2 hour 52 min. Thus in real life the download time is much longer then the ideal download time by about 30%.

8. Conclusion

This paper has clearly illustrated that VPN is able to circumvent the ISP throttling in Malaysia. The important point to note is that the VPN server should be located on the high bandwidth backbone.

The other advantages gained by end user are:

(1) Encrypted traffic – Nobody can sniff the traffic even the ISP.

(2) Able to surf anonymously – Not to reveal the actual IP address of origin. To be used as if the surfer is from the server site IP.

(3) Able to bypass site which has been blocked by country of origin.

References

Asohan, A (2007). Throttling broadband access for all, Available online: http://thestar.com.my/news/story.asp?file=/2007/2/8/focus/16811744&sec=focus, Downloaded: August 13, 2007

Broadbandreports.com (2004). Shaw Throttling Bit Torrent?, Available online: http://www.broadbandreports.com/shownews/56419, Downloaded: May 27 2007

Cisco systems (undated), Managing Peer-To-Peer Traffic With Cisco Service Control Technology, White paper, Available online: http://www.cisco.com/en/US/prod/collateral/ps7045/ps6129/ps6133/ps6150/prod_white_paper0900aecd8023500d_ps6135_Products_White_Paper.html, Downloaded: April 1, 2008

Klinker, E (2008). When should you choose P2P?, Available online: http://www.streamingmedia.com/article.asp?id=10068, Downloaded: April 15 2008

helpwithpcs.com (2008). Internet Connections explained, A guide to dial-up, ADSL and Cable connections, Available online:

http://www.helpwithpcs.com/internet/internet-connections.htm, Downloaded: 31 March 2008

Normalkid, L.L.C (2003). Peer to Peer (P2P) Introduction and History, Available online: http://www.mac-p2p.com/p2p-history/, Downloaded: May 11 2008

Teoh, S (2006). TM Net: P2P traffic clogging broadband, Available online: http://star-techcentral.com/tech/story.asp?file=/2006/3/1/technology/13540730&sec=technology, Downloaded: August 13, 2007

Cheshire, S (1996). Latency and the quest for Interactivity, Available online: http://www.stuartcheshire.org/papers/LatencyQuest.ps, Downloaded: 31 March 2008

Themanualpage.org (2002). Glossary – L, Available online: http://www2.themanualpage.org/glossary/glo_l.php3, Downloaded: 31 March 2008

TM News Release (April 2007). TM leads 17-member Consortium of Telcos to establish Asia-America Gateway (AAG) – The first submarine cable system linking South East Asia directly to USA, Available online: http://www.tm.com.my/about_tm/newsroom/2007/070427.htm, Downloaded: Feb 14, 2008

Table 1. Factor for planning network bandwidth (Source: Cisco systems white paper)

Parameter	Network planning
Upstream / Downstream traffic ratio	The amount of upstream and downstream traffic flow should be tailored according to the correlation to application requirement according to its users
Time of day and percentage of activity	Based on the users usage there must exist “rush hour” or “peak hour” during B2B or B2C activities. Most home user may peak at the evening till night, weekend and holidays
Traffic destination and peering point	Traffic within the local network backbone and international connection should be reviewed and considered
Estimated traffic volume	Network bandwidth is fixed for all users and a certain over subscriptions assumption should be planned accordingly to calculate current network capacity

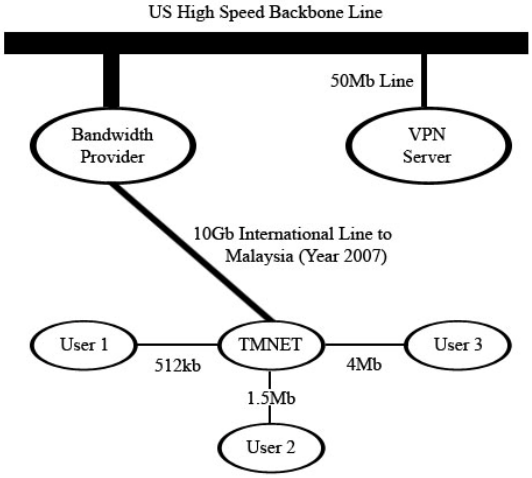


Figure 1. Location of the Implemented VPN Server

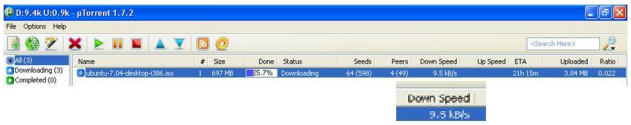


Figure 2. Downloads Performance without VPN

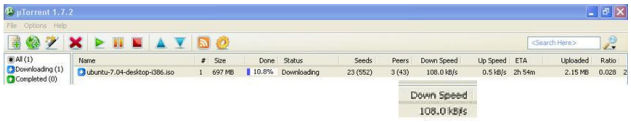


Figure 3. Downloads Performance with VPN

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Study on the TOPN Abnormal Detection Based on the NetFlow Data Set

Hongzhuo Zhang

School of Computer Science and Technology, University of Electronic Science and Technology of China

Chengdu 610054, China

Abstract

In recent years, with the increase of the scale and the complexity of the network, various abnormity flows begin to occur in the network. To know the running state of the network, the technology of NetFlow emerges as the times require. The NetFlow data are transmitted directly by the router which supports the function of NetFlow. Comparing with traditional data acquirement technology, the NetFlow technology needs not deployment in advance and acquires data conveniently, and it is gradually turning into the important data sources for the network management, maintenance, supervision and control. At present, there are still few abnormity detection methods based on the NetFlow data set. In this article, we introduced the principle and functions of the NetFlow, and put forward the TOPN flow abnormity detection method based on the NetFlow technology. This method can effectively detect the flow state in the network, reflect the network state and offer the information about abnormal network flows.

Keywords: NetFlow, TOPN, Abnormity detection

1. Introduction

In recent years, as Internet is using in the life, learning and work by more and more people, and the running of the whole society has been carved by the sign of Internet, and the Internet has been developed from single industry tool to the social and popular toll entering into various industries in China. The scale and complexity of Internet increase day by day, and multifarious network applications are correspondingly emerging. However, with the normal application flows of Internet, various abnormity flows successively occur, which have influenced the normal running of Internet and threatened the security and normal use of the host computer. In order to know the running and use state of network and find out the possible abnormal flows in the network, we need an effective flow detection method, so we put forward the TOPN abnormity detection method based on the technology of NetFlow. Most traditional detection methods need attacking the samples to detect the flow by describing the special mode of each attack. By analyzing the data packet in virtue of the net capture data software to compare the attach samples and complete the detection, this method will possesses higher precision rate. But its abilities of capturing data and analyzing speed especially in the high speed network are not ideal, and the invasion detection system based on characteristics is implemented only depending on the alarming rules established by human beforehand, so it lacks in the self-learning function when it faces changing network attacks continually. The method we put forward in this article is based on the NetFlow technology, and it can realize the data acquirement and analysis in high speed network, and the TOPN detection model we established has certain self-learning ability and ideal abnormality detection ability.

2. Brief introduction of the NetFlow technology

The technology of NetFlow is a set of network flow monitoring technology developed by Darren Kerr and Barry Bruins of Cisco Company in 1996, and it have obtained the supports from many mainstream manufacturers such as Juniper and Extreme. At present, it has been embedded in most Cisco routers, and is gradually becoming into the industrial standard. It is not only an exchange technology, but also a flow analysis technology, and it is one of mainstream charging technologies in the industry. It can solve following questions about the flow such as who interviews whom, where it interviews, which protocol is used and how much the concrete flow is. Because of the technology and the market occupancy ratio of the Cisco network product, the NetFlow technology has been one of mainstream flow analysis technologies at present. Its work principle includes storing the data into the cache by the format of flow record, and educing the data through the protocol of UDP until the data fulfill the conditions. The NetFlow data in this article are transmitted to the data capture host computer through the protocol of UDP by the router of Cisco, and because the data scale is huge, we write the data by the form of file to store in order to ensure the effective acquirement of the data. To be convenient for data analysis, we established a set of Oracle data sever which could translate the written data file into the

Oracle database and analyze the data by operating the Oracle database.

There are many data formats for NetFlow, we adopt the data with NetFlow Edition 5 to study, and the data fields used in this article mainly include unix_secs (the second amount from 0000 UTC 1970 to now), srcaddr (source IP address), dstaddr (destination IP address), dOctets (the total amount of byte on the third layer in the data packet of information flow), srcport (TCP/UDP source port number), dstport (TCP/UDP destination port number), and prot (IP protocol, such as 6=TCP and 17=UDP).

3. Abnormity detection based on TOPN analysis

The detection principle of TOPN includes the statistical analysis of TOPN and the model establishment based on the acquired NetFlow data. The TOPN analysis is to implement statistical ordering aiming at certain selected index, and select the top parameters which can fulfill the conditions. In the article, we mainly explain six TOPN statistical analysis items including host-computer initiating connection amount ratio TOPN analysis, host-computer emitting data amount ratio TOPN analysis, protocol use ratio TOPN analysis, destination port use ratio TOPN analysis, source IP and destination IP pair ratio TOPN analysis, and destination IP ratio TOPN analysis.

(1) The host-computer initiating connection amount ratio TOPN analysis is mainly used to find out the IP addresses, rankings and connection amount ratios of the top N host-computers which initiate the most connections in certain period.

(2) The host-computer emitting data amount ratio TOPN analysis is mainly used to find out the IP addresses, rankings and connection amount ratios of the top N host-computers which emit the most data in certain period.

(3) The protocol use ratio TOPN analysis is mainly used to find out the top N used protocols, rankings and the ratios relative to the total protocols in certain period.

(4) The destination port use ratio TOPN analysis is mainly used to find out the top N used destination port numbers, rankings and the ratios in the all ports in certain period.

(5) The source IP and destination IP pair ratio TOPN analysis is mainly used to find out the top N used source IP and destination IP pairs, rankings and the ratios in all source IP and destination IP pairs in certain period.

(6) The destination IP ratio TOPN analysis is mainly used to find out the top N used destination IP addresses, rankings and the ratios in all destination IP addresses.

3.1 TOPN ranking detection

TOPN ranking detection is a sort of usual TOPN detection method, and it is mainly used to confirm a general ranking sequence, and once the ranking by the statistics has certain transition with the consulted ranking, the result is treated as abnormity. Because we adopt the technology of NetFlow, we aim at the wide area network (WAN). There are numerous IP addresses in a huge WAN, and in the short term, the occurrences of top IP addresses of TOPN in different time periods will be dispersed, it is not ideal for ranking detection, so we can find out the destination ports and protocol types with few changes by the experiment to establish the model which would perform the abnormity detection.

3.2 TOPN ratio detection

There are numerous IP addresses in a huge WAN, and in the short term, the occurrences of top IP addresses of TOPN in different time periods will be dispersed, but the use ratios of top N IP addresses accord with the normal distribution, so we can establish the model according to the change range of ratio to implement the abnormity detection.

The mass data are adopted to establish the model first, and the mass data in one week can cover the abnormity. After that, the model should be renovated periodically, and the time of one weak probably is used to renovate the model, and the data to renovate the model are the normal data selected by the model.

The main method establishing the model is to use the average value to add the n times of standard deviation.

(1) The host-computer initiating connection amount ratio TOPN model mainly includes the top N TOPN and the corresponding connection amount ratios.

(2) The host-computer emitting data amount ratio TOPN model mainly includes the top N TOPN and the corresponding data amount ratios.

(3) The protocol use ratio TOPN model mainly includes the top N TOPN and the corresponding protocol use ratios.

(4) The destination port use ratio TOPN model mainly includes the top N TOPN and the corresponding destination port use ratios.

(5) The source IP and destination IP pair ratio TOPN model mainly includes the top N TOPN and the corresponding source IP and destination IP pair ratios.

(6) The destination IP ratio TOPN model mainly includes the top N TOPN and the corresponding destination IP ratios.

When the ratio of relative TOPN by the statistic is higher than the corresponding ratios of the model, the data in this period are regarded as abnormity, and the system will emit alarm and perform the abnormity detection aiming at the abnormity and data. By the experiment, we found that the alarm rate was a little high when the model used the single standard deviation, and the detection effect was good when the model used the double standard deviations, and the detection effect was bad when the over-double standard deviations were used.

4. Experiment and analysis

By setting up the router of the school network information center, we use it to transmit the NetFlow data to the data capture host-computer by UDP protocol, and store the data into the Oracle data server, and analyze the server to perform relative data analysis and detection. The experiment environment is seen in Figure 1.

4.1 TOPN ranking detection

In the experiment, we can find out the ranking distribution rules of protocol and port in the network. The main distribution sequence tendency of the protocol generally is UDP, TCP, ICMP and Reserved. Sometimes UDP and TCP will exchange their places, which are related to the flow of P2P, and the ICMP and Reserved will also exchange their places sometimes. The main distribution sequence tendency of the port generally is from 80 (HTTP), 8000 (Tencent QQ sever opens this port), 015000, 28000, 0 (Reserved), 443 (Https), 8080 (WWW agent opens this port), 10000 (network data management protocol), 26000 to 21 (FTP). According to the sequence, we establish the normal ranking model, and when the statistical port ranking and protocol ranking change and the abnormity data occur, the system will emit alarm and perform the abnormity detection.

4.2 TOPN ratio detection

Table 1 ~Table 4 are the ratio detection models in certain period.

Table 5 and Table 6 respectively listed the TOPN detections of the destination port and protocol in five minutes, and when the system detected the abnormity, the system would compare the abnormity data with the sample base of attack type to judge the attack type.

The ratio values of the second model and the eighth model exceed the corresponding values of the normal model, so the system emitted alarm and compared the data of abnormity with the sample base, and the detection results respectively were “Teardrop 68: refusing service attack” and “TCP SYN 61: refusing service attack”.

The ratio value of the first model exceeded the corresponding value of the normal model, the system emited alarm and compared the data of abnormity with the sample base, and the detection result was “Teardrop 68: refusing service attack”.

5. Conclusions

Through analyzing the principle and functions of NetFlow, we put forward a sort of TOPN abnormity detection method based on NetfFow, which could effectively detect the abnormity flows in the network by means of the traditional ranking transition detection method and the corresponding ranking use ratio value detection method, and be effectively applied in the wide area network with high speed. Nowadays, this method has been successfully applied in certain practical project of University of Electronic Science and Technology of China.

References

Lai, Jibao, Wang, Huiqiang & Jin, Shuang. (2007). Study of Network Security Situation Awareness System Based on NetFlow. *Application Research of Computer,* No.8.

Li, Xiangguo & Fei, Lingling. (2008). The Research of NetFlow-based Flow Analyzing Technology. *Control & Automation,* No.15.

Wang, Hua. (2004). Key Software Technique and Algorithm of Analyzing NetFlow Traffic Data. *Computer Engineering,* No.z1.

Yan, Jiahao, Ma, Rui & Wu, Yibo. (2006). Application and Design for Internet Traffic Monitoring Technology. *Designing Techniques of Posts and Telecommunications,* No.4.

Table 1. TOPN model of host-computer initiating connection amount

TOPN ranking N	Ratio
1	1.7237%
2	1.4924%
3	1.3819%
4	1.3179%
5	1.1980%
6	1.1417%
7	1.0807%
8	1.0516%
9	1.0327%
10	1.0071%

Table 2. TOPN model of destination IP ratio

TOPN ranking N	Ratio
1	1.4124%
2	1.2801%
3	1.1746%
4	1.1195%
5	1.0332%
6	0.9919%
7	0.9486%
8	0.9216%
9	0.9012%
10	0.8783%

Table 3. TOPN model of protocol use ratio

TOPN ranking N	Ratio
1	57.9436%
2	44.7775%
3	1.2337%
4	0.8144%
5	0.0197%
6	0.0017%

Table 4. TOPN model of destination port use ratio

TOPN ranking N	Ratio
1	12.9209%
2	3.0820%
3	2.7714%
4	2.3662%
5	1.8860%
6	1.3052%
7	1.0643%
8	1.0027%
9	0.9635%
10	0.9272%

Table 5. TOPN statistics of destination port

TOPN ranking N	No. of destination port	Ratio	Start time
1	80	5.29374271142224	2008-12-10 10:56:03
2	28000	3.35425422559238	2008-12-10 10:56:03
3	15000	1.92983437718383	2008-12-10 10:56:03
4	0	1.42244514436732	2008-12-10 10:56:03
5	53125	1.41290074150137	2008-12-10 10:56:03
6	22222	1.16968302708965	2008-12-10 10:56:03
7	8000	1.05471136957793	2008-12-10 10:56:03
8	33333	1.0075378841485	2008-12-10 10:56:03
9	30433	0.848354797269204	2008-12-10 10:56:03
10	4576	0.755763118891445	2008-12-10 10:56:03

Table 6. TOPN statistics of protocol

TOPN ranking N	No. of protocol	Ratio	Start time
1	17(UDP)	59.6278341116958	2008-12-10 10:56:03
2	6(TCP)	39.1158152949604	2008-12-10 10:56:03
3	1(ICMP)	1.24362472285577	2008-12-10 10:56:03
4	47	0.0127258704879384	2008-12-10 10:56:03

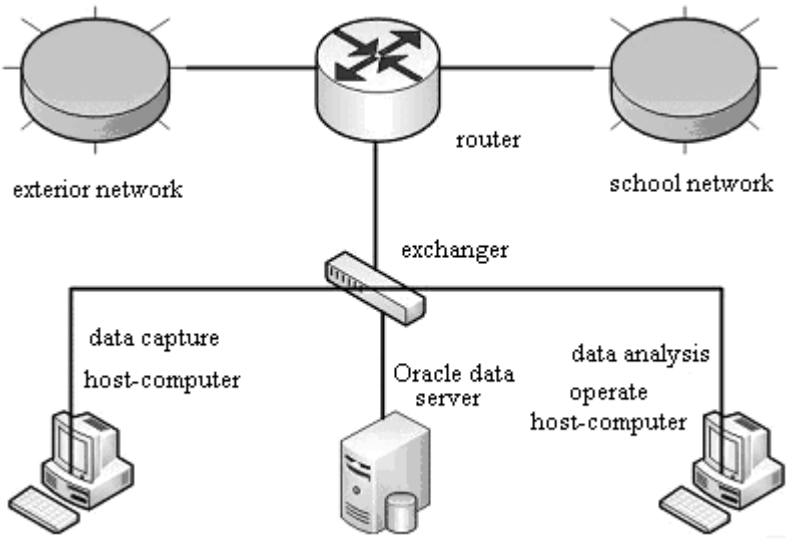


Figure 1. Experiment Environment

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Study on the Design Principles of Data Disaster Recovery System for Hospitals

Ping Xiao

Yunnan University of Finance and Economics, Kunming 650221, China

E-mail: Xiaoping@ynufe.edu.cn

Abstract

The medical information system has been applied in hospitals widely, which makes hospitals acquire notable management benefit and economic benefit. When hospitals enjoy the quick service decision and convenient management because of informationization, they are also in the danger of data loss. Data disaster recovery is the guarantee of the hospital informationization construction, and the base to ensure the normal medical service of hospital. In the article, the data disaster recovery mode in the medical information system is studied and the design principles of hospital data disaster recovery system are systematically analyzed.

Keywords: Hospital information system, Disaster recovery system, Design principles

1. Introduction

The medical information system has been applied in hospitals widely, which makes hospitals acquire notable management benefit and economic benefit. With the deeply using of the system, the data quantity in the server increases several-fold. At the same time, the downtime also occurs in the operation of the system, and though the system can be recovered depending on daily copy, and the loss is limited, but the normal work is influenced. So in the system, not various network equipments but data stored in the server are most precious. The urgent problem of the medical information system is to ensure the healthy running of the system and prevent the data loss because of default or disaster. If the data security of the hospital can not be guaranteed, large numbers of network investments should lose meanings. When hospitals enjoy the quick service decision and convenient management because of informationization, they are also in the danger of data loss. Because of virus attack, loss of data copy will seriously influence the sustainable works of hospitals.

2. Disaster recovery system

To prevent above possible disasters and reduce possible losses furthest, the disaster recovery system (DRS) is often established for pivotal operations. The establishment of DRS needs two parts, i.e. the data disaster recovery and the application disaster recovery. The data disaster recovery means to establish a distant data system which is a real-time copy of the local key application data. The application disaster recovery is to establish a set of complete copy application system corresponding with the local production system in the different place, and in the disaster, the remote system could quickly replace the local system. The data disaster recovery is the guarantee to fight disaster, and the application disaster recovery is the construction target of the disaster recovery system.

There are different definitions for the disaster recovery system in the world, but the total standards are same, i.e. first, the local data is copied, then the copy tape is stored at the remote place, and the remote copy system which can replace the local system in time is established, and the recovery time can be from several days to hour class, minute class, second class or zero data loss.

3. Measurement indexes of the disaster recovery system

Technically, there are two main indexes, RPO (Recovery Point Object) and RTO (Recovery Time Object) to measure the disaster recovery system, and the RPO presents the data quantity allowed to be lost when the disaster happens, and RTO presents the recovery time of the system.

To reduce data loss as much as possibly, a remote data storage system needs to be established, and it can perform the image copy of data with the production system. To reduce the recovery time of the system, a set of complete copy application system matching with the production system needs to be established in the disaster recovery center based on the data disaster recovery system. When the disaster happens, the disaster copy center could quickly replace the operation running, which can not only reduce the data loss and the system recovery time to the largest extent, but also

largely enhance the continual usability of the hospital operation system.

If the construction investments are available, the ideal construction of the disaster recovery system is to achieve RTO=0 and RPO=0. The realization of RTO=0 is mainly decided by the concrete construction mode of the application system, such as the concrete applications of the system disaster examination technology and the data migration technology in the system. And the realization of RPO=0 is mainly decided by the concrete applications of data storage and copy technology and the data fault-tolerant technology. The synchronous replication of system and data can be adopted, but how to realize the synchronous replication and ensure the integrality of data will directly influence the investment cost of the system.

When the disaster recovery system is established, RPO and RTO should be reasonably selected according to users' demands. When RPO and RTO are smaller, the usability of the system is higher and user's investments will be larger. Of course, the class of the disaster recovery system is also decided by the protection class and the significance of the operation application, and the construction should be based on the effective capital utilization and the existing system rebuilding. RPO and RTO must be confirmed by different operation demands after the risk analysis and operation influence analysis are performed. To different operations, the demands of RPO and RTO are also different.

4. Design principles of hospital data disaster recovery system

4.1 Principle of operation demand

First of all, according to the operation contents of the hospital information system, the problem which application systems need disaster recovery should be confirmed. The operation flow of each subsystem should be known in detail, and RTO and RPO are defined in advance, and the problem which application system needs which disaster recovery structure should be confirmed.

The work mode and the data acquirement contents of each subsystem in the hospital information system are analyzed first, and the requirements of each subsystem are established in Table 1.

4.2 Principle of physical demand

The differences between the disaster recovery center and the operation production center are more, the variety of disaster which can be resisted is more, and the difference is embodied by the physical distance to the large extent.

According to the demand of application, the local copy center and the remote disaster recovery center can be selected and established. The cost of the local copy center is less, and the technology is relatively simple. But once destructive disasters such as earthquake, flood, fire and electric deficiency happens in the local place, all local copy data will be lost and the operation cannot be recovered.

When the main computer, storage and network structure which can be started at any moment in the remote place, and the copy station is not only a separate copy system, but it should be in the running state, can provide the production application service and quick operation replace. Generally, both the data center system and the copy center system are in the running state, but the operation processing system only runs in the production center. Any amendments to the data by the operation system will be copied to the copy center synchronously. When the faults happen in some parts of the production center, the application system on the fault machine will be automatically replaced by other machines in the production center by the copy software. When the whole system of the production center breaks down because of accidents, the copy server port software will start the whole operation application system in the copy center according to the appointed rules. After the system is recovered, the application system will be replaced to the production center from the copy center, and the copy center system returns to the copying state.

4.3 Principle of software demand

Physical defaults will make the system break down, which can be found easily. But the logical default is hard to be found, but only if the original data exist, the data can be recovered, so it is necessary to copy the historical data for long term. Therefore, good "data protection system" and "data protection project" will reduce the disaster loss to the minimum degree.

The disaster recovery software can automatically discover the default, and give clues to help deciding whether to perform the switch, and automatically perform data interview, application online and reorientation of the network. The disaster recovery software can often nondestructively test the disaster recovery fame with low costs, fully embody existing IT frame, be one part of DR plan, automatically perform necessary approaches, test and claim the happening of disaster, decide the default switch point, test the loss quantity of data and decide the acceptable data loss quantity.

The disaster recovery center must support the operation replacement (i.e. all operation flows are implemented in the system of the disaster recovery center) and the switch (i.e. after the operation replacement completes and the original data center could support the operation, and the operation will be switched to the original data center from the disaster recovery center). According to the design requirements of the disaster recovery center, the operation switch should be

performed in the time of RTO value. By the support of the disaster recovery software, enterprise can effectively protect, interview, store and manage important information data at any time, which can help enterprise to enhance the information usability to large extent and relieve the fear of trouble in the rear of the enterprise.

4.4 Input and output analysis of the disaster recovery system

Generally, the disaster recovery system needs much investment, but the use probability is low, so the total cost of ownership (TCO) and the return on investment (ROT) should be seriously analyzed and computed. TCO and ROT are main indexes to measure the investment and return of the disaster recovery system, and the cost and benefit analysis (CBA) of the disaster recovery system emphasizes the analysis of the benefit of investment, and considers the rationality of the investment from the view of the development of the operation system.

First, the disaster recovery system which is prepared to be constructed and the continuality of the operation system which is running should be considered. To protect the former investments, the large-scale reconstruction of the original operation system should be avoided to the best. Second, the influence of the operation system extension to the disaster recovery system, especially the influences of the added storage capability and communication line load should be considered.

As viewed from the economy, the optimal disaster recovery solution is not certainly the disaster solution with the best performance, and TCO and ROI of the disaster recovery system are very important design indexes for many users. TCO includes the total investments of establishing system, maintaining system and extending system, and because of low starting probability of the disaster recovery system, the development of new technology and the enhancement of the cost performance of new product will certainly induce the depreciation of the disaster recovery equipments. Therefore, for the disaster recovery system, TCO is higher, ROI is lower.

According to the statistics of relative institutions, for the bank industry which requires the key operations highly, each downtime of the computer system will lose 10 million dollars averagely, and immaterial asset loss which can not be measured to the reputation of the company, but the cost of the disaster recovery project only needs one million dollars averagely. And the return ratios of investments will be relatively lower for the medical industry.

5. Conclusions

The data disaster recovery is the guarantee of the information-based construction for hospitals, and the base to ensure normal medical service of the hospital. Good works and bases will help to future construction and development. In this aspect, long-term thinking and investment will exert important functions for future online running of more application management systems.

By the disaster recovery system, the medical information system can achieve high usability, high security, high efficiency, high expansibility and high management property. As viewed from the operation and application layers, the disaster recovery processing and high usability of the data center can enhance the efficiency of service and increase users' satisfactions and competitive forces through ensuring continual 24-hours key operations.

References

Chen, Qiang, Ma, Liya & Zhao, Wei. (2006). Digitized Hospital Standard System Construction and Question. *Medical Information*. No.19(1). P.27-29.

IBM. (2005). White Book of Disaster Recovery. [Online] Available: http://www.ibm.com.

Informationization Work Lead Group Office of Chinese Ministry of Public Health. (2001). *Basic Function Standard of Hospital Information System.* March of 2001.

Xie, Changsheng, Han, Desheng, Li, Huaiyang & Cao, Qiang. (2004). Grade and Technology of Data Disaster Recovery and Copy. [Online] Available: www.csip.cn/new/st/al/2004/0730/342.htm. (July 30, 2007).

Zhang, Feng. (2004). Expert Talking of Storage Technology: Disaster Recovery Builds a Port. [Online] Available: www.dostor.com/info/netstor/2004-10-29/0001921045. (Oct 29, 2004).

Table 1. Operation analysis of hospital information system

Application system	Significance	Real time	Recovery time
Outpatient doctor work station subsystem	Pivotal	High	Less than 10 minutes
Hospital doctor work station subsystem	Pivotal	High	Less than 2 hours
Nurse work station subsystem	Pivotal	High	Less than 2 hours
Clinical examination subsystem	Pivotal	High	Less than 2 hours
Blood transfusion management system	Important	Middle	Less than 6 hours
Medical image system	Pivotal	High	Less than 2 hours
Surgery anesthesia subsystem	Important	Middle	Less than 6 hours
Drug management subsystem	Important	Middle	Less than 6 hours
Clinical and emergency register subsystem	Important	Middle	Less than 6 hours
Clinical and emergency pricing and charge subsystem	Pivotal	High	Less than 10 minutes
Inpatient in, out and transfer management subsystem	Important	Middle	Less than 6 hours
Hospitalization charge subsystem	Pivotal	High	Less than 2 hours
Material management subsystem	Important	Middle	Less than 6 hours
Equipment management subsystem	Important	Middle	Less than 6 hours
Financial management subsystem and economic accounting management subsystem	Pivotal	High	Less than 2 hours
Disease case management subsystem	Pivotal	High	Less than 2 hours
Medical statistical subsystem	Secondary	Low	In one day
Hospital director comprehensive inquiry and analysis subsystem	Pivotal	High	Less than 2 hours
Patient consultation service subsystem	Secondary	Low	In one day
Medical insurance system interface	Important	Middle	Less than 6 hours
Community medical system interface	Important	Middle	Less than 6 hours
Remote medical consultation service interface	Important	Middle	Less than 6 hours

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Pattern Extraction and Rule Generation of Forest Fire Using Sliding Window Technique

Ku Ruhana Ku-Mahamud

Graduate Department of Computer Science

Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Tel: 60-4-928-4717 E-mail: ruhana@uum.edu.my

Khor Jia Yun

Intel Technology Sdn. Bhd.

Kulim Hi-Tech Park, 09000 Kulim, Kedah, Malaysia

Tel: 60-4-433-6939 E-mail: khorjy@hotmail.com

The research is financed by MOHE, FRGS No. 11705

Abstract

Patterns can be extracted from historical data and used to make predictions of future events. Such predictions are useful to support decision makers in various areas. In this study the sliding window technique is used to reveal forest fire patterns that relate four meteorological conditions (temperature, relative humidity, wind speed and rainfall) with burnt area size. Extracted patterns are then being grouped based on the size of burnt area. Rules are then generated resulting in eight distinct patterns of meteorological conditions that could predict the size of forest fire. Experimental results showed that extracted patterns produced good prediction accuracy.

Keywords: Pattern extraction, Sliding window, Forest fire

1. Introduction

Forest fire plays an important role in shaping forest ecosystems all over the world. For example, the Mountain Ash forests in Australia depend on fire for the regeneration of the ecosystems. There is increasing evidence to show that the link between climate change and El Nino phenomenon is causing an escalation in number and size of forest fire. There is new evidence from Amazon showing that tropical forests that have experienced burning before would be more susceptible to future burning. Thus, there is an increased possibility that wildfire episodes will occur more frequently, and in the magnitude not endurable by the tropical forest ecosystem. Scientists believed that the entire Amazon would be threatened, and this will affect the biodiversity and climate change globally (Rowell & Moore, 2000). Forest fire occurrence can caused massive destruction. A severe drought throughout Siberian Peninsular in 2005 caused massive fire in northern and central Portugal, destroying more than 150 000 ha land and killing 15 people (Voigt et al., 2007).

Fire, being a chemical reaction, needs heat, oxygen, and fuel for ignition and spread. Forest fire happens when an uncontrollable fire starts and spreads itself to natural vegetation, thus the ignition cause and environmental conditions affects the probability of occurrence. Influential factors of forest fire are like fuel's distribution and quality, weather, topography and human factors (Orozco, 2008). In addition, fires are easily ignited during draught seasons.

Umamaheshwaran et al. (2007) applied image mining techniques that used satellite images to produce forest fire prediction model. Laneve and Cadau (2007) assessed the quality of fire hazard prediction model based on Fire Potential Index and using SEVIRI/MSG satellite images. Meanwhile, it is well known that weather conditions and fire risks are related, and weather conditions are a crucial factor which determines whether wildfire would occur, and how far it will spread (Wells et al., 2007). A recent statistical survey indicated that weather and forest Fire Weather Index (FWI) components are highly influential in forest fires in Portugal (Carvalho et al., 2008).

Predicting natural hazards using satellites has begun in 2000, and it has changed the way natural disasters are being assessed (Gillespie et al. 2007). The use of satellite images in fire prediction and detection could be seen in the literature

from Umamaheshwaran et al. (2007), and Laneve and Cadau (2007). While using satellite images provides almost real-time data for analyst to predict natural hazards, atmospheric interference such as clouds, smoke and haze cause distortions in the images retrieved. Apart from that, spatial satellite images have low resolution. In addition, using satellites require high equipment and maintenance costs (Cortez & Morais, 2007).

Earlier studies have shown that weather and forest Fire Weather Index (FWI) components have significant influence on forest fire (Cortez & Morais, 2007; Carvalho et al., 2008). Cortez and Morais (2007) have concluded in their studies that further research is needed to confirm that direct weather conditions (i.e. temperature, rain, relative humidity and wind speed) are preferable over accumulated values in predicting forest fire behaviour.

Previous study by Cortez and Morais (2007) showed that predicting forest fire burnt area using Support Vector Machine (SVM) fed with meteorological data (i.e. rain, wind, temperature and humidity) gives the best performance compared with other four data mining approaches (Multiple Regression, Decision Tree, Random Forest, Neural Network). In their study, the five data mining approaches were being used with data consisting of spatial, temporal, meteorological and Fire Weather Index (FWI) data from Montesinho Natural Park in northern region of Portugal. The proposed solution, however, achieved a lower predictive accuracy for large fires. Thus, further exploratory research is required for predicting burnt forest area by using only meteorological data.

Data mining can be used to discover interesting salinity and temperature patterns and the patterns can be used to predict future events (Huang et al., 2008). By using association rules mining spatial-temporal patterns revealing the salinity and temperature variations can be discovered. Furthermore to evaluate the rules generated that have antecedent and consequent, an important measurement is proposed. The proposed process of mining association rules consists of several steps: transforming quantitative attributes, discovering frequent inter-transaction itemsets, generating association rules, and identifying insight association rules. During the frequent inter-transaction itemset discovery, sliding window technique is used so that during analysis only rules within the window would be considered and thus minimizing effort in mining uninteresting rules. The proposed method could generate rules about salinity and temperature patterns but the research did not evaluate the performance of the proposed important measurement.

This research has investigated patterns of weather in relation to the size of forest fire. The patterns are then classified and rules that can be used for decision making are formulated. Section 2 presents the research approach while Section 3 describes the extraction of the forest fire patterns. The rule generating activity is explained in Section 4 and concluding remarks are presented in Section 5.

2. Research Approach

Figure 1 depicts the approach that has been used in conducting the research. The sliding window technique used by Huang et al. (2008) in their study to discover rules of ocean salinity and temperature variations has been adopted and adapted in this study.

Forest fire data which consist of forest fire and meteorological information, have been collected from UCI Machine Learning Repository and from the study by Cortez and Morais (2007). In the data preparation stage, the attributes are described, records with missing values were removed or missing values were replaced with estimated values, or ignore the missing values during analysis. After the cleaning stage, transforming data by performing discretization to change the data type from continuous to categorical is a significant task in data mining process. A suitable discretization technique used can potentially improve the performance of the data mining technique significantly. Pattern extraction and rule generation using the sliding window technique are performed after the data preparation stage.

The data contain forest fire occurrence and forest fire weather index (FWI) components in Montesinho natural park, located in northern Portugal region, from 2000 until 2003. The data was integrated with weather observations (wind speed, temperature, relative humidity and rainfall) obtained from Braganca Polytechnic Institute. The occurrences of forest fires were within the Montesinho Natural Park. The park is being divided into eight distinct X and Y location by placing a 9x9 grid on the map. There is a total of 81 combinations of X and Y used in this study. The forest fire data consist of 517 where 270 records or 52% of the data consists of forest fire occurrences with burnt area more than or equals to 0.01ha and the other 247 instances or 48% of the data contains forest fire occurrences with burnt area less than 0.01ha. There are 5 attributes namely temperature, relative humidity (RH), wind speed, rainfall and the burnt size that have been included in this study.

3. Forest Fire Pattern Extraction

The actual forest fire data were in continuous value. Data transformation has been performed for these data whereby the continuous values have been changed to categorical form. This is a significant task in data mining process. A suitable discretization technique used can significantly improve the performance of the data mining technique. In pattern extraction study, data need to be in the categorical form and thus, discretization must be performed before the analysis could begin. The continuous values of the attributes temperature, relative humidity, wind speed and rain are being transformed into categorical form by using the ranges applied by practitioners (Kottlowski, 2006; Pearce, 2008).

The values for temperature are transformed into six categories as shown in Table 1 while the values for relative humidity are transformed into three categories as presented in Table 2. Temperatures are measured in Celcius and classified from 'very cold' to 'extremely hot' while relative humidity is from 'low' to 'very high'. The codes are assigned to facilitate the analysis of patterns with sliding window technique process.

Tables 3 and 4 display the categories used for wind speed and rainfall measurement. There are thirteen categories for wind speed which starts from 'calm' to 'hurricane' and it is measured in km/hr. Six (6) categories have been used for rainfall where it is measured in mm/hr. The rainfall is categories from 'very light rain' to 'extreme rain'. Again, codes are assigned for the purpose of patterns analysis.

The data was divided into the training and testing sets after the transformation process is complete. 80% of the data or 414 records were allocated for the training and the remaining 20% (103 records) for validation. The sliding window technique is used to extract patterns from the data. Each window slice will capture a set of patterns which consist of temperature, relative humidity, wind speed and rainfall with the associated burnt area. The sliding window technique captures data pattern as it moves down the data set. The captured patterns are then recorded. Figure 2 depicts the process of extracting patterns using the sliding window technique.

A total of 32 forest fire patterns have been obtained in this stage (refer Table 5). There are patterns that occurred many times such as patterns number 22 and 31 which occurs 97 times and 88 times respectively. However, there are patterns which occurred only once (refer patterns number 1, 5, 10, 14, 16, 17, 19, 24 and 26).

All the 32 patterns obtained from the training data were validated using the validation dataset. If a pattern correctly predicts the fire size, then the pattern is given a count for "*True Positive*". Otherwise, the pattern is given a count for "*True Negative*". The percentage of accuracy for each pattern is calculated by taking the total number of "*True Positive*" to be divided by the total number of validation against a pattern. A total of 18 patterns (56%) have been validated while the remaining 14 patterns (44%) could not be validated as the patterns were not found in the validation data. However, this does not affect the results as the total occurrences of each of these 14 patterns are small. Most of them (8 patterns) have only 1 occurrence, while 4 patterns have 2 occurrences, 1 pattern has 4 occurrences and 1 has 5 occurrences (see Table 5). The 18 patterns that have been validated have high occurrences. All the patterns that have been validated are summarized in Table 6. Each pattern is given an ID to simplify the process of eliminating patterns, classifying patterns and generating rules. Table 6 shows that out of the 18 patterns, three patterns have less than 50% accuracy (pattern ID 1, 7 and 33). Only 15 patterns were used in the classification and rule generation processes. In the pattern classification stage, the patterns were grouped according to size of burnt area (refer Table 7). There are 14 patterns under the target class "M" representing medium sized fire (1-500ha), and 1 pattern under target class "L" representing large sized fire (>500ha). There is no pattern associated with target class "S" representing small sized fire (<1ha).

4. Rule Generation

In the rule generation stage, the patterns that are being extracted and classified are translated into rules that can be used for predicting fire behaviour according to burnt area size. First, the categories for the four attributes are converted into their interval representations (refer Table 8).

From the interval representation, it can be seen that there are some intervals from two different patterns that could be merged into one. Patterns with ID 3 and 4 have similar values for the attributes except for the wind speed. It is apparent that these two records can be merged together. Similar merging can be performed to patterns with ID 5 and 6. Thus, patterns with ID 3, 4, 5 and 6 are merged to form one rule. Merging patterns with ID 9, 10, 11 and 12 will produce another rule. As a result of merging, 8 rules are generated as depicted in Table 9.

The first rule obtained from pattern ID 17 can be interpreted as, "*when the temperature is between 21 and 28 degrees Celsius, relative humidity between 0 and 50 percent, wind speed between 1 and 5 km/hr and rainfall of less than 0.25 mm/hr, the predicted forest fire size will be large, i.e. more than 500 hectare*". From the obtained rules, it is observed that whenever there is forest fire, the rainfall must be minimal or none i.e below 0.25 mm/hour. The first rule generated from the pattern with ID 17 has very high occurrence (refer Table 5, record 31) and the validation shows that this pattern is highly related to forest fire behaviour, and could cause large fire. Another pattern highly related to forest fire behaviour (refer Table 5, record 22) is the pattern with ID 11 which is later being used to form the 5th rule. It can also be concluded that high temperature is significant to forest fire because out of the eight rules, six of them shows that the temperature is in the range of 14 to 35 degrees Celsius.

5. Conclusion

Patterns that relate the size of the forest fire and meteorological attributes namely temperature, relative humidity, wind speed and rainfall were obtained using the sliding window technique. These patterns showed that different combinations of the four meteorological conditions will affect the size of forest fires. The rules generated from the obtained patterns can be used to identify the size of forest fire burnt area. By knowing the behaviour of a potential fire, the decision

maker can plan for the management of the fire more effectively. Effective prediction of forest fire could increase the efficiency in fire management, thus saving lives and natural resources. Future study may include the use of fuzzy technique with the sliding window approach in extracting the forest fire patterns as classification of meteorological attributes is very subjective and depends on experts.

References

Carvalho, A., Flannigan, M. D., Logan, K., Miranda, A. I., & Borrego, C. (2008). Fire activity in Portugal and its relationship to weather and the Canadian Fire Weather Index System. *International Journal of Wildland Fire,* 17, 328-338.

Cortez, P. & Morais, A. (2007). A data mining approach to predict forest fires using meteorological ata. *Proceedings of the 13th Portugese Conference on Artificial Intelligence*, 512-523.

Gillespie, T., Chu, J., Frankenberg, E. & Thomas, D. (2007). Assessment and prediction of natural hazards from satellite imagery. *Progress in Physical Geography*, *31*(5), 459-470.

Huang, Y. P., Kao, L. J. & Sandnes F. E. (2008). Identifying insightful salinity and temperature variations in ocean data. *Proceedings of the 8th IEEE International Conference on Computer and Information Technology*, 191-196.

Kottlowski, D. (2006). *Humidity*. Retrieved August 30, 2008, from AllExperts, Accuweather Web site: http://en.allexperts.com/q/Accuweather-2546/Humidity-2.htm

Laneve, G. & Cadau, E. G. (2007). Quality assessment of the fire hazard forecast based on a fire potential index for the Mediterranean area by using a MSG/SEVIRI based fire detection system. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 2447-2450.

Li, H. F., Ho, C. C., Kuo, F. F. & Lee S. Y. (2006). A new algorithm for maintaining closed frequent itemsets in data streams by incremental updates. *Proceedings of the Sixth IEEE International Conference on Data Mining Workshops*, 672-676.

Orozco, S. J. (2008). Forest fire risk model for Michoacan, Mexico. Unpublished M.Sc. thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

Pearce, V. (2008). *Windows to the Universe: Classification of Wind Speeds*. Retrieved August 25, 2008, from the University Corporation for Atmospheric Research (UCAR) Website: http://www.windows.ucar.edu/tour/link=/earth/Atmosphere/wind_speeds.html&edu=high

Rowell, A. & Moore, D. P. F. (2000). Global review of forest fires. International Union for Conservation of Nature and Natural Resources.

Umamaheshwaran, R., Bijker, W., & Stein, A. (2007). Image mining for modeling of forest fires from meteosat images, *IEEE Transactions on Geoscience and Remote Sensing,* 45(1), 256-253.

Voight, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., & Mehl, H. (2007). Satellite image analysis for disaster and crisis-management support. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6), 1520-1528.

Wells, G., Rohling, K., Kapus, J., & Swedberg, T. (2007). The fire-climate connection. *Fire Science Digest:* Research Supporting Sound Decisions.

Table 1. Categorization of temperature

Temp (T) (Celcius)	**Description**	**Code**
$0 \leq T \leq 7$	Very cold	C1
$7 < T \leq 14$	Cool	C2
$14 < T \leq 21$	Mild	M
$21 < T \leq 28$	Warm	W
$28 < T \leq 35$	Hot	H
$T > 35$	Extremely hot	E

Table 2. Relative humidity categorization

Relative Humidity (RH) (%)	**Description**	**Code**
RH<50	Low	L
50≤RH≤80	High	H
RH>80	Very high	V

Table 3. Wind speed categorization

Wind speed (WS) (km/hr)	**Description**	**Code**
WS<1	Calm	L
1≤WS<6	Light air	LA
6≤WS<12	Light breeze	LB
12≤WS<20	Gentle breeze	GB
20≤WS<29	Moderate breeze	MB
29≤WS<39	Fresh breeze	FB
30≤WS<50	Strong breeze	SB
50≤WS<62	Moderate gale	MG
62≤WS<75	Fresh gale	FG
75≤WS<89	Strong gale	SG
89≤WS<103	Whole gale	WG
103≤WS≤117	Storm	S
WS>117	Hurricane	H

Table 4. Rainfall categorization

Rainfall (R) (mm/hr)	**Description**	**Code**
R<0.25	Very light	VL
0.25≤R<1.0	Light	L
1.0≤R<4.0	Moderate	M
4.0≤R<16.0	Heavy	H
16.0≤R≤50.0	Very heavy	VH
R>50.0	Extreme	E

Table 5. Patterns obtained from pattern generation

No	Temp	RH	Wind	Rain	Occurrence
1	C1	H	L	VL	1
2	C1	H	LA	VL	6
3	C1	H	LB	VL	5
4	C1	L	LB	VL	5
5	C1	V	L	VL	1
6	C1	V	LB	VL	2
7	C2	H	L	VL	2
8	C2	H	LA	VL	16
9	C2	H	LB	VL	10
10	C2	L	L	VL	1
11	C2	L	LA	VL	7
12	C2	L	LB	VL	12
13	C2	V	LA	VL	7
14	C2	V	LB	VL	1
15	H	L	LA	VL	19
16	H	L	LB	VL	1
17	M	H	L	VL	1
18	M	H	LA	VL	51
19	M	H	LA	L	1
20	M	H	LB	VL	16
21	M	L	L	VL	4
22	M	L	LA	VL	97
23	M	L	LB	VL	25
24	M	V	LA	VL	1
25	W	H	LA	VL	5
26	W	H	LA	H	1
27	W	H	LA	M	2
28	W	H	LB	VL	4
29	W	H	LB	M	2
30	W	L	L	VL	3
31	W	L	LA	VL	88
32	W	L	LB	VL	17

Table 6. Validated patterns

ID	Pattern	No. of true positive	No. of true negative	Accuracy (%)
1.	C1/H/LA/VL	0	1	0
2.	C1/H/LB/VL	1	0	100
3.	C2/H/LA/VL	2	1	67
4.	C2/H/LB/VL	2	1	67
5.	C2/L/LA/VL	4	0	100
6.	C2/L/LB/VL	4	0	100
7.	C2/V/LA/VL	0	1	0
8.	H/L/LA/VL	5	0	100
9.	M/H/LA/VL	7	0	100
10.	M/H/LB/VL	3	2	60
11.	M/L/LA/VL	18	0	100
12.	M/L/LB/VL	6	1	86
13.	M/V/LA/VL	1	0	100
14.	W/H/LA/VL	1	2	33
15.	W/H/LB/VL	1	0	100
16.	W/L/L/VL	1	0	100
17.	W/L/LA/VL	31	0	100
18.	W/L/LB/VL	5	1	83

Table 7. Classification of patterns

ID	Pattern	Burnt size (ha)	Class
17.	W/L/LA/VL	<= 1090.84	L
2.	C1/H/LB/VL	<= 11.19	M
3.	C2/H/LA/VL	<= 37.02	M
4.	C2/H/LB/VL	<= 11.24	M
5.	C2/L/LA/VL	<= 8.31	M
6.	C2/L/LB/VL	<= 61.13	M
8.	H/L/LA/VL	<= 64.1	M
9.	M/H/LA/VL	<= 82.75	M
10.	M/H/LB/VL	<= 7.31	M
11.	M/L/LA/VL	<= 212.88	M
12.	M/L/LB/VL	<= 34.36	M
13.	M/V/LA/VL	<2.21	M
15.	W/H/LB/VL	<= 54.29	M
16.	W/L/L/VL	<= 11.06	M
18.	W/L/LB/VL	<= 15.45	M

Table 8. Classified Patterns

ID	Temp	RH	Wind Speed	Rainfall	Class
17.	21-28	<50%	1-5	<0.25	L
2.	0-7	50-80%	6-11	<0.25	M
3.	7-14	50-80%	1-5	<0.25	M
4.	7-14	50-80%	6-11	<0.25	M
5.	7-14	<50%	1-5	<0.25	M
6.	7-14	<50%	6-11	<0.25	M
8.	28-35	<50%	1-5	<0.25	M
9.	14-21	50-80%	1-5	<0.25	M
10.	14-21	50-80%	6-11	<0.25	M
11.	14-21	<50%	1-5	<0.25	M
12.	14-21	<50%	6-11	<0.25	M
13.	14-21	>80%	1-5	<0.25	M
15.	21-28	50-80%	6-11	<0.25	M
16.	21-28	<50%	<1	<0.25	M
18.	21-28	<50%	6-11	<0.25	M

Table 9. Generated rules

ID	Temp	RH (%)	Wind Speed	Rainfall	Class
17.	21-28	0-50	1-5	<0.25	L
2.	0-7	50-80	6-11	<0.25	M
3,4,5,6	7-14	0-80	1-11	<0.25	M
8.	28-35	0-50	1-5	<0.25	M
9,10,11,12	14-21	0-80	1-11	<0.25	M
13.	14-21	>80	1-5	<0.25	M
15,18	21-28	0-80	6-11	<0.25	M
16.	21-28	0-50	<1	<0.25	M

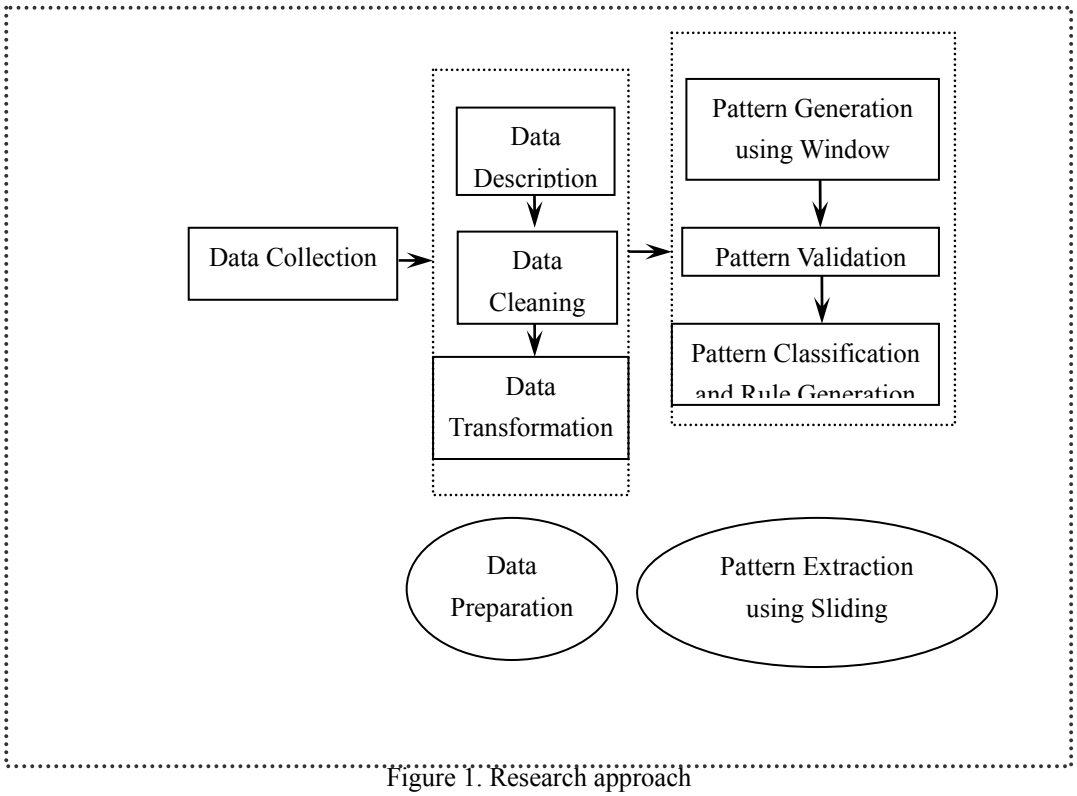


Figure 1. Research approach

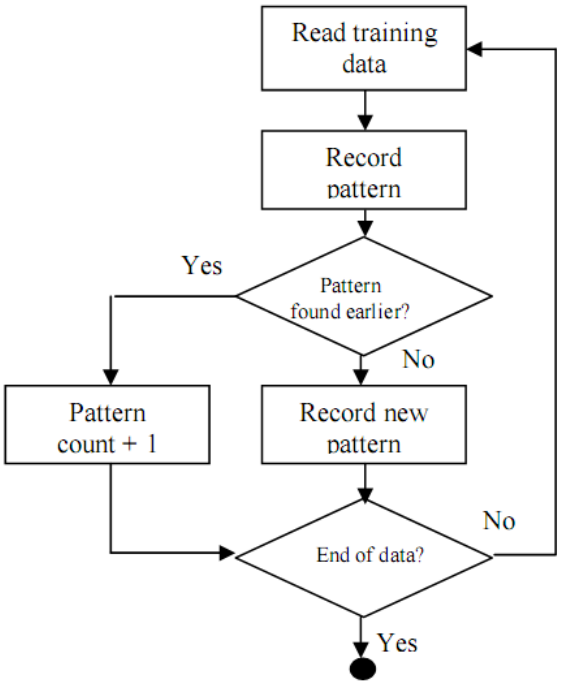


Figure 2. Process flow for extracting patterns

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

Study on the Fault Diagnosis Technique of Aerial Camera Exposure Board Based on Petri Net

Chengjun Tian
School of Electronics and Information Engineering
Changchun University of Science and Technology
Changchun 130022, China
Tel: 86-431-8558-3328 E-mail: tcjciom@126.com

Zhiyong An, Lijuan Li, Chengyu Yang, Jie Duan & Kang Wang
School of Photoelectric Engineering
Changchun University of Science and Technology, Changchun 130022, China
Tel: 86-431-8558-3328 E-mail: zhiyong_an@126.com

Abstract

Because of the mechanical situation, aerial carrier condition, environment condition and usage maintenance of the aerial camera, faults always occur. So it is necessary to check and maintain the aerial camera on the ground, and the interior structure of the aerial camera is very complex, and the aerial camera is composed by many circuit boards, and the checking must be performed from the whole camera to each component. Aiming at the demands of aerial camera fault diagnosis, the board level fault diagnosis system of aerial camera was developed in the article. The failure mode of the exposure board in the aerial camera electric control system was completely analyzed, and the exposure board fault board model based on Petri net was established, and the dynamical transition of the system status was solved by the relational matrix method, and the TPS program of the exposure board was compiled, and the fault diagnosis of the exposure board was primarily actualized in the article.

Keywords: Fault diagnosis, Petri net, Dynamical transition

1. Introduction

Aerial camera is the platform of aerial reconnaissance. The accurate imaging components are installed in the camera and the ground imaging workstation and the intelligence picking personnel all depend on the images shot by the aerial camera to pick up intelligences. The work of aerial camera directly influences the generation quality and generation cycle of intelligences.

The interior structure of aerial camera is very complex, and the electric control system is composed by multiple circuit boards. The fault diagnosis must be performed from the whole camera to each component. To test and diagnose the electric control system of aerial camera, the aerial camera fault model based on Petri net was established in the article, and aiming at the checking demands of the army, the board level fault diagnosis system of aerial camera was developed, which is the ground guarantee equipment of the panoramic aerial camera and is used to check the work of the camera, test and diagnose the various functional circuit boards of the fault camera, and orientate the fault in the minimum replaceable unit.

2. Total scheme design of the aerial camera board level fault diagnosis system

According to the work environment and setup function, the board level fault diagnosis system adopts the ATE design principle and many functional modules such as PC104 embedded industrial control computer structure, additive signal simulation function module, communication data bus interface module, checking signal acquirement conditioning function module and other suited function modules to complete main functions of the system (Tian, 2008). For the hardware design, various additive encouragement signal sources, nonstandard signal conditioning and testing control

function module, and the data interfaces of various checking sub-channels correspond with the diagnosed circuit board interfaces to actualize the fault diagnosis of various function circuit boards. The total design structure of the system is seen in Figure 1.

3. Principle of Petri net

3.1 Brief introduction of Petri net

Petri Net is a sort of mathematical model proposed by German Scientist Petri, and it is also a sort of figure method whish is used to denote and analyze the dynamical behaviors of the system, and it adopts some basic figure symbols to describe the relation between conditions and events (Yuan, 2005). The dynamical transition process of the system status can be studied through the shifts of data, resources and conditional decisions. The place node and the transition node can be used to implement static structured analysis to the system, and the token on the node can be used to implement the dynamical behavior analysis.

3.2 Petri net and the description of event logic relationship

Petri net is a sort of directional figure composed by position, transition, directional arc and token, and it includes following definitions (Jiang, 2003).

Definition 1. Tetrad PN= (P, T, A, M) is the Petri net, and it fulfils $P \cup T \neq \Phi$, $P \cap T = \Phi$ and $A \subseteq (P \times T) \cup (T \times P)$.

Pi= {Pi | Pi is the place, $1 \leq i \leq I$ (I is a integer)}

T= {Ti | Ti is the transition, $1 \leq i \leq I$}

A= {Ai |Ai is the directional arc from the place to the transition or from the transition to the place}

M= {Mi |Mi is the amount of certain place with sign which denotes a sort of status whether the condition or event of the place is tenable}

Definition 2. The transition in the Petri net is enabled, i.e. if the each input place of certain transition in Petri net contains a sign at least, the transition in Petri net is enabled.

Definition 3. Transition fire: if the transition is enabled, and remove one sign in all input places of the transition and add one sign in all output places of the transition, so the transition is fired.

Based on above definitions, the logic relation described by the Petri net can be obtained as seen in Figure 2.

4. Establishment of the exposure board fault model based on Petri net

The camera electric control system is composed by multiple circuit boards including the exposure board. Through analyzing the fault events of the exposure board, the Petri net fault model (seen in Figure 3) can be obtained. In Figure 3, P1 denotes component SN54HC04J is in failure, P2 denotes component SN74HC02 is in failure, P3 denotes component S26LS32M is in failure, P4 denotes component ADC0808 is in failure, P5 denotes component DAC1230LCJ is in failure, P6 denotes component ADOP071 is in failure, P7 denotes component TIL117 is in failure, P8 denotes component LF156J is in failure, P9 denotes component SG1524BJJ is in failure, P10 denotes component ADC0808 is in failure, P11 denotes component ADOP072 is in failure, P12 denotes the communication between the main control board and the exposure board is abnormal, P13 denotes the exposure board is not reactive to the scale change of the exposure, P14 denotes the light-stick motor control circuit is abnormal, P15 denotes the step motor drive circuit is abnormal and P16 denotes the exposure board is in failure.

From Figure 3, the events when the exposure board works abnormally includes the communication between the main control board and the exposure board is abnormal, the exposure board is not reactive to the scale change of the exposure, the light-stick motor control circuit is abnormal and the step motor drive circuit is abnormal.

5. Solving the dynamical transition of the system status by the relational matrix method

The solutions of the status transition of the diagnosis model system based on Petri net mainly include the relational matrix method and the ladder diagram solution, and the relational matrix method (Hu, 2001) is adopted to solve the dynamical transition of the system.

The column vector M is used to denote the sign of the Petri net diagnosis model, and the sign amount of the place Pi is denoted by the elements in i'th row, and the status transition of the system denotes the dynamical evolvement process of the sign. The next status of the system is decided by the present status and the logic structure of the system. To compute the status transition of the system, following definitions should be made.

(1) M_K denotes the k'th evolvement status of the system, and M_0 is the initial status of the system.

(2) A^T is the relational matrix of the system, the rows of the matrix denote places which represents transition, and when the i'th place is the output place of the j'th transition, the element $(A^T)_{ij}$ in the i'th row and the j'th column of A^T is 1, and when the i'th place is the input place of the j'th transition, $(A^T)_{ij}$ is -1, and when there is not the direct

relation, $(A^T)_{ij}$ is 0.

(3) S^k are column vectors, and when the element in the j'th row is 1, the transition T^j is fired.

Based on above definitions, the status transition expression of the system is

$$M_{k+1}=M_k+A^T S_k \quad (1)$$

From the recursion relation of the formula (1), the following formula can be obtained.

$$M_n=M_0+A^T\sum_{k=0}^{n-1} S_k \quad (2)$$

From the formula (2), the final status of the system is accumulated from the initial status to each transition. According to the formula (1), the k+1'th status of the system M_{k+1} is decided by the k'th status M_K, and the logic structure of the system decides A^T, and the trial and error method (Xu, 2005) is adopted to confirm the system status S_k.

The concrete approaches of the trial and error method include that supposing n_T denotes the transition amount in the exposure board fault diagnosis model, S_k^i (i=1, 2…, n_T) denote the status of the i'th transition in the k+1'th evolvement, and if S_k^i=1, the transition is fired, and if S_k^i=0, the transition is not fired. First, order S_k^i=0 (i=1, 2…, n_T), and order S_k^i=1 in turn, and compute $A_T S_k$, and if all elements exceed 0, so the i'th transition in the k+1'th evolvement fulfills the fire condition, S_k^i=1 comes into existence, or else, the fire condition of the i'th transition in the k+1'th evolvement is not fulfilled, order S_k^i=0 again. When S_k^i=0(i=1, 2…, n_T) all are confirmed, S_k is confirmed, and formula (1) can be used to solve S_{k+1}. Supposed that the initial sign of the exposure board Petri net diagnosis model $M_0=(1000010000000000)^T$, and according to the definition, the relational matrix of the exposure board fault model can be confirmed first.

$A^T =$

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
P1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P2	0	−1	0	0	0	0	0	0	0	0	0	0	0	0	0
P3	0	0	−1	0	0	0	0	0	0	0	0	0	0	0	0
P4	0	0	0	−1	0	0	0	0	0	0	0	0	0	0	0
P5	0	0	0	0	−1	0	0	0	0	0	0	0	0	0	0
P6	0	0	0	0	0	−1	0	0	0	0	0	0	0	0	0
P7	0	0	0	0	0	0	−1	0	0	0	0	0	0	0	0
P8	0	0	0	0	0	0	0	−1	0	0	0	0	0	0	0
P9	0	0	0	0	0	0	0	0	−1	0	0	0	0	0	0
P10	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	0
P11	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0
P12	1	1	1	0	0	0	0	0	0	0	0	−1	0	0	0
P13	0	0	0	1	1	1	0	0	0	0	0	0	−1	0	0
P14	0	0	0	0	0	0	1	1	1	0	0	0	0	−1	0
P15	0	0	0	0	0	0	0	0	0	1	1	0	0	0	−1
P16	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

By the trail and error method, $S_0=(1000010000000000)^T$ and $S_1=(0000000000001100)^T$, so $M_2=M_0+A^T\sum_{i=0}^{1} S_i=(000000000000002)^T$, i.e. the peak place P16 is signed, so the following diagnosis conclusions can be obtained. The events of "component SN54HC04J is in failure" and "component ADOP071 is in failure" represented by the bottom places P1 and P6 induce the occurrence of peak event, i.e. induce the fault of the exposure board of the aerial camera.

6. Conclusions

Aiming at the ground checking demands of the aerial camera of the army, the fault diagnosis system of aerial camera is studied in the article, and the fault diagnosis method based on Petri net is adopted to diagnosis the faults of the exposure board of the aerial camera, and the systematic fault model is established. The method is simple and convenient for modeling, and it can conveniently perform knowledge denotation, diagnosis reasoning and logic relation expression, and it is very effective to systematically analyze the complex net topological structure, and it is a fault diagnosis method based on model with very wide applied foreground (Chen, 2000). At present, the panoramic aerial camera has been

equipped in the scout plane, and the aerial camera fault diagnosis system has been approved. The fault diagnosis system is stable in reliability and servicing, and its fault diagnosis cover rate of the camera electric control system can achieve 100%. The system can provide powerful guarantee for the development and usage of the panoramic aerial camera and the accomplishment of the scout tasks.

References

Chen, Yubao. (2000). Study on the Application of Petri Nets to Fault Diagnosis. *China Mechanical Engineering.* No.12.

Hu, Changhua. (2001). *Analysis and Design of Fault Diagnosis and Fault Tolerant Control in the Control System.* Beijing: National Defense Industry Press. Feb, 2001.

Jiang, Changjun. (2003). *Behavior Theory and Application of Petri Net.* Beijing: Higher Education Press. Jan, 2003.

Tian, Chengjun. (2008). Design of the Infield Test Equipment in the Aerial Camera. *Ship & Ocean Engineering.* No.2.

Xu, Yongjun, Wang, Yi & Li, Guanxiong. (2005). A Study of a Fault Diagnosis Method Based on the Petri Net. *Industrial Instrumentation & Automation.* No.6.

Yuan, Chongyi. (2005). *Principle and Application of Petri Net.* Beijing: Electronic Industry Press. March, 2005.

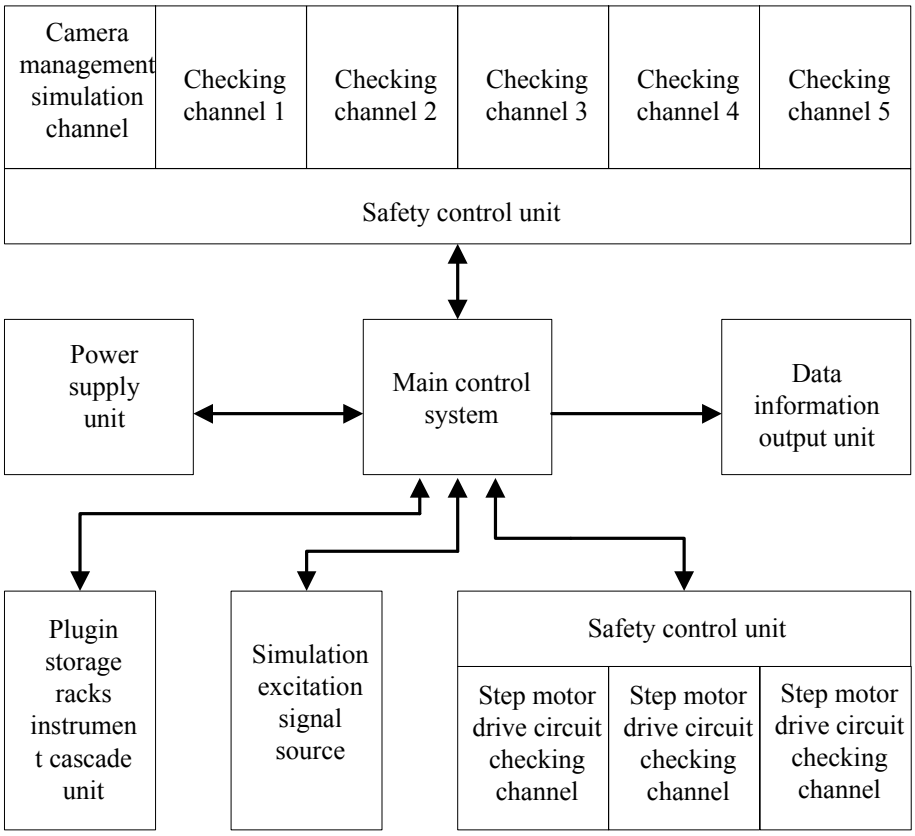


Figure 1. Total Structure of the Fault Diagnosis System

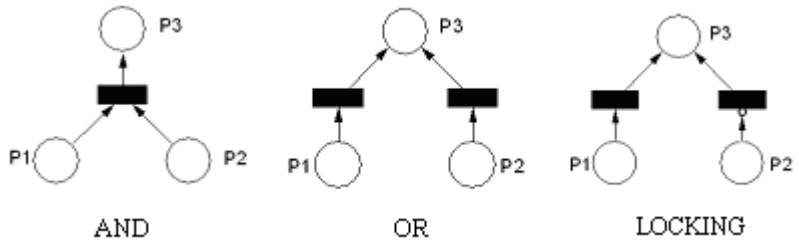


Figure 2. Petri Net Model of Logic Operation

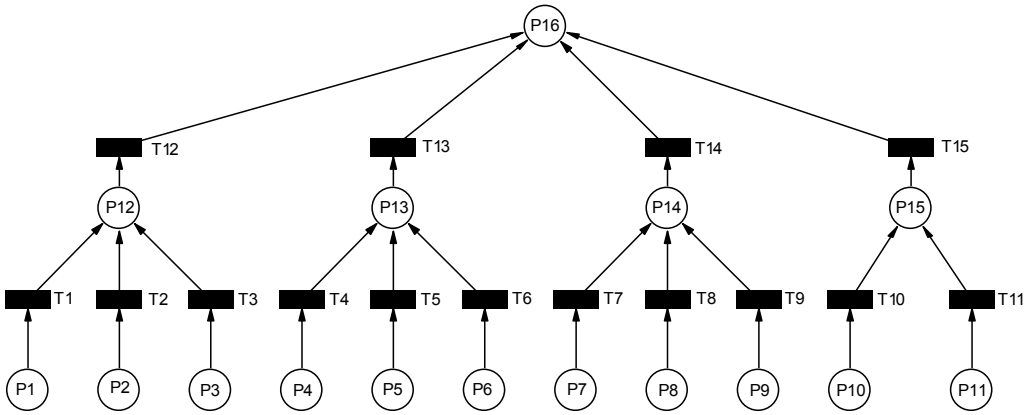


Figure 3. Exposure Board Petri Net Fault Model

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Estimation of Saturation Percentage of Soil Using Multiple Regression, ANN, and ANFIS Techniques

Khaled Ahmad Aali (Corresponding author)

Department of Irrigation and Reclamation, Faculty of Water and Soil Engineering

University of Tehran, Karaj, Iran

Tel: 98-914-185-3937 E-mail: khaled.ahmadauli@gmail.com

Masoud Parsinejad

Department of Irrigation and Reclamation, Faculty of Water and Soil Engineering

University of Tehran, Karaj, Iran

Bizhan Rahmani

Soil Science Department, College of Agriculture

Azad Islamic University of Karaj, Karaj, Iran

Abstract

The saturation percentage (SP) of soils is an important index in hydrological studies. In this paper, artificial neural networks (ANNs), multiple regression (MR), and adaptive neural-based fuzzy inference system (ANFIS) were used for estimation of saturation percentage of soils collected from Boukan region in the northwestern part of Iran. Percent clay, silt, sand and organic carbon (OC) were used to develop the applied methods. In additions contributions of each input variable were assessed on estimation of SP index. Two performance functions, namely root mean square errors (RMSE) and determination coefficient (R^2), were used to evaluate the adequacy of the models. ANFIS method was found to be superior over the other methods. It is, then, proposed that ANFIS model can be used for reasonable estimation of SP values of soils.

Keywords: Saturation percentage, Soils, MR, ANN, ANFIS, Boukan

1. Introduction

Saturation percentage (SP) is related to the mechanical constituents of soils and can, therefore, be regarded as a quantitative measure of soil texture, water-holding capacity, and cation exchange capacity. Soil profiles may be described in terms of SP, and soil maps may be developed to represent quantitative changes in soil texture within a region. Furthermore, measurement of soil water content is important in simulation of all aspects of hydrological cycle, for estimation of plant water use, and for characterizing most soil physical, chemical, and biological processes. Chemically, water serves as transport agent for dissolved inorganic chemicals and suspended biological components, involved in the processes of soil development and degradation.

The saturation percentage is defined as the ratio of the amount of water added to saturate dry soil samples, to total mass of the fully dried soil. Direct measurement of saturation percentage is time consuming and relatively expensive. In the conventional procedure, initially dried soil samples are saturated with deionized water and then oven dried at 105 C° for a period 24 hrs.

Indirect methods are, then, used as an alternative solution. Numerous attempts have been made to correlate base-saturation percentage with pH in saline suspensions. (Keeney and Corey, 1963; Shaw, 1952) Such efforts have been, at best, only partially successful, particularly if one attempts to estimate precise levels of base saturation from pH measurements.

In view of the importance of accurate estimation of saturation percentage (SP), using basic and readily available soil information, adoption of modern techniques such as artificial neural networks (ANNs) and fuzzy inference system (FIS) can be a viable alternative. Because of the non-linear structure in ANNs models and ambiguity in variables in FIS models, (Piotrowski et al. 1996; Mukhopadhyay 1999), researchers are, recently attracted in using hybrid models such

as Adaptive Neural-based Fuzzy Inference System (ANFIS) to further analyze the variables, which are spatially distributed. (Lee, 2000)

In this study, efficiency of Adaptive Neural-based Fuzzy Inference System (ANFIS), artificial neural networks (ANN) and multiple regression (MR) models were examined in estimation of saturation percentage (SP) using measured data of clay, silt sand and organic carbon (OC), in Boukan plain in the West Azerbaijan Province, Iran.

2. Material and methods

2.1 Description of the study area

The study area is Boukan region which is located in southern part of West Azerbaijan Province, Iran. Boukan covers an area of 47300 hectares (Figure.1), with latitude of 36º 32' and longitude of 46º 13'. Average elevation in this region is 1330 m above sea level. The area falls under the semiarid climate with an average rainfall of 517 mm/year. 600 measured values of clay, silt, sand, organic carbon (OC %) and saturation percentage (SP %) which was previously collected by the Iranian Rojhalat Soil Lab were used in this study. The Walky black method was used to determine OC content in the samples and soil texture (percent sand-silt-clay) were determined using Hydrometer method (Schumacher 2002).

A summary of obtained results and their basic statistics is presented in Table1. The SP values ranged between 28 and 66 (%) with an average value of 48.18%. The respective average values of effective organic carbon, percent of clay, percent of silt and percent of sand were determined as 0.72, 24.43, 51.34 and 24.23 %.

Simple regression analysis was performed to initially establish the predictive relationship between measured parameters. The relations between SP and other measured parameters were analyzed using linear, power, logarithmic, and exponential functions. Models with statistically significant and strong correlations were then selected for further analysis (Table 2). Regression equations were also established among index parameters with SP (Table 3). All obtained relationships were found to be statistically significant according to the Student's t-test at 99% level of confidence.

2.2Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are based on current understanding of biological nervous systems, though much of the biological details are neglected. ANNs are massively parallel systems composed of many processing elements connected by links of variable weights (Lippman, 1987).

In Figure 2 a three-layered neural network consisting of i, j and k layers with the interconnection weights W_{ij} and W_{jk} between layers of neurons is illustrated (Hagan and Menhaj 1994; Kisi and Uncuoglu 2005). The weights are computed through an iterative process based on back propagation algorithm in such a way that the difference between computed and given output (or any error criterion such as mean square error) is sufficiently small. The hidden layer node numbers of each model were determined after trying various network structures, since there is no theory yet available to tell how many hidden units are needed to approximate a given function. Cross validation mode (checking mode) monitors the error to find the optimal termination point for training and also avoid overtraining. Testing mode is used to determine how accurately the network can simulate input-output relationships.

All collected data were divided into three sets, namely; training (3/5 of all data), test (1/5 of all data), and verification (remaining 1/5 of all data). In this study MatLab 7.4 software was used in neural network analysis having a three-layer feed-forward network that consisted of an input layer, one hidden layer, and one output layer. Logsigmoid (transfer) functions for both hidden and output layers were used for analysis network activation.

2.3 Adaptive Neural-based Fuzzy Inference System (ANFIS)

Adaptive Neural-based Fuzzy Inference System (ANFIS) is capable of approximating any real continuous function on a compact set to any degree of accuracy (Jang et al., 1997). Specifically, ANFIS system of interest here is functionally equivalent to the Sugenofirst-order fuzzy model (Jang et al., 1997; Drake, 2000). The hybrid learning algorithm, introduced as follows, combines gradient descent and the least-squares method. As a simple example a fuzzy inference system is assumed with two inputs x and y and one output z. The first-order Sugeno fuzzy model, a typical rule set with two fuzzy If-Then rules, can be expressed as:

$$\text{Rule 1: If x is } A_1 \text{ and y is } B_1, \quad \text{then } f_1 = p_1 x + q_1 y + r_1 \tag{1}$$

$$\text{Rule 2: If x is } A_2 \text{ and y is } B_2, \quad \text{then } f_2 = p_2 x + q_2 y + r_2 \tag{2}$$

The resulting Sugeno fuzzy reasoning system is presented in Figure 3, where the output *z* is the weighted average of the individual rule outputs and is itself a crisp value. The corresponding equivalent ANFIS architecture is presented in Figure 4. Nodes at the same layer have similar functions. The node function is described next. The output of the *i*th node in layer *l* is denoted as O_{li}.

Layer 1: Every node *i* in this layer is an adaptive node with node function

$$O_{1,i} = 礎_i(x), \qquad \text{for } i = 1,2, \text{ or}$$
$$O_{1,i} = 碟_{i-2}(y), \qquad \text{for } i = 3,4$$

where x (or y) is the input to the *i*th node, and A_i (or B_{i-2}) is a linguistic label (such as ''low'' or ''high'') associated with this node. In other words, O_{1i} is the membership grade of a fuzzy set A (= A_1, A_2, B_1,or B_2) and it specifies the degree to which the given input x (or y) satisfies the quantifier A .

$$\mu A_i(x) = \frac{1}{1+[(x-c_i)/a_i]^{2bi}} \qquad (3)$$

where $\{a_i, b_i, c_i\}$ is the parameter set. As the values of these parameters change, the bell-shaped function varies accordingly, thus exhibiting various forms of membership functions on linguistic label A_i. In fact, any continuous and piecewise differentiable functions, such as commonly used in triangular shaped membership functions, are also qualified as candidates for node functions in this layer (Jang, 1993). Parameters in this layer are referred to as premise parameters. The outputs of this layer are the membership values of the premise part.

Layer 2: This layer consists of nodes labeled Π, which multiply incoming signals and sending the product out. For instance,

$$O_{2,i} = W_i = 礎_i(x)\,碟_i(y), \quad i = 1,2. \qquad (4)$$

Where, each node output represents the firing strength of a rule.

Layer 3: In this layer, the nodes labeled *N* calculates the ratio of the *i*th rule's firing strength to the sum of all rules' firing strengths

$$O_{3,i} = \overline{W}_i = \frac{W_i}{W_1 + W_2} \qquad i = 1,2. \qquad (5)$$

the outputs of this layer are called normalized firing strengths.

Layer 4: This layer's nodes are adaptive with node functions

$$O_{4,i} = \overline{W}_i f_i = \overline{W}_i (p_i x + q_i y + r_i) \qquad (6)$$

where $\overline{W}_i$ is the output of layer 3, and $\{p_i, q_i, r_i\}$ is the parameter set. Parameters of this layer are referred to as consequent parameters.

Layer 5: This layer's single fixed node labeled Σ computes the final output as the summation of all incoming signals

$$O_{5,i} = W_i f_i = \sum_{i=1} \overline{W}_i f_i = \frac{\sum_i W_i f_i}{\sum_i W_i} \qquad (7)$$

In the present study, the triangular and Gaussian membership functions were used. In each application, different numbers of membership functions were tested and the best one, with minimum root mean square error (RMSE) and the maximum R^2, was selected.

A hybrid intelligent system called ANFIS (the adaptive neuro-fuzzy inference system) for predicting SP was also applied. ANFIS was trained with the help of Matlab (version 7.4) and SPSS (15.0 package), and two top models, namely ANFIS11 and ANFIS12 were selected based on RMSE and R^2. ANFIS parameter types for the two models and their values are presented in Table 4.

2.4 Multiple Regression Models

Multiple Regression (MR) is a statistical technique that allows us to predict someone's score on one variable on the basis of their scores on several other variables. MR was used in order to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable. The general form of these models is $y = b_1x_1 + b_2x_2 + ... + b_nx_n + c,$ where $\{b_1, b_2, ..., b_n\}$ the regression coefficients are, and y is now written as a function of n independent variables; $x_1, x_2, x_3, ...x_n$. C is the y-intercept (Milton et al., 1997; McClave et al., 1997). Eight MR analysis were carried out to correlate the measured SP to various combinations of measured parameters; namely clay, silt, sand and OC content

3. Application and results

in this study various combinations of measured parameters (clay, silt, sand, organic carbon percentage, and saturation percentage (SP %)) were examined as inputs to ANN models for evaluation of effect of each variable on SP (%). Several NF models were established with different variable added into the input combination at one time. Thus, the input combinations evaluated here were: (i) clay; (ii) silt; (iii) sand; (iv) OC; (v) OC and clay; (vi) silt and sand; (vii) OC and silt; (ix) OC and sand; (x) clay and silt; (xi) clay and sand; (xii) clay, silt and sand (xiii) OC, clay, silt and sand.

In this study, estimated and observed values were statistically compared and analyzed using root mean square error (R-MSE) and determination coefficient (R^2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (8)$$

Where n is the number of observations, O_i and P_i are the measured and predicted values, respectively. The calculated indices are presented in Table 5. The obtained results shows that ANN models with OC, clay, silt and sand inputs (combination (xiii)) had the smallest RMSE and the highest R^2. This observation emphasizes that all of these parameters have a weight on prediction of SP (%). The best two ANN models among all applied models are presented in the Table 5.

The coefficient of correlation between measured and predicted values is considered as a valuable indicator of the predictability of models. These relationships, presented in Figure 5, were found to be highly correlated. Cross-correlation between predicted and observed values (Figure 5) indicated that the constructed ANN model is acceptable for prediction of SP.

Multiple regression models were developed to predict SP with different combinations of inputs and the two best models were selected based on R^2 and RMSE indices (Table 6). Cross-correlation between predicted and observed values (Figure 6) indicated that the constructed MR model is acceptable for prediction of SP.

According to the RMSE and R^2 values (Table 7) and cross-correlation between predicted and observed values (Figure 7), ANFIS model constructed has a high prediction performance for prediction of SP.

Analysis of ANFIS models with various combinations of inputs and different type of membership functions with different numbers showed that the model having OC, clay, silt, and sand as inputs with triangular membership function type has the best performance, followed by, the model having clay, silt, and sand combination as inputs with Gaussian membership function.

4. Conclusion

In this study Adaptive Neural-based Fuzzy Inference System (ANFIS), artificial neural network (ANN) and multiple regression (MR) techniques were used for prediction of Saturation Percent (SP) using characteristics of soils; namely, percent sand, silt, clay, and organic carbon (OC). Appropriate models were developed by scrutinizing their performance degrees and the model with minimum RMSE and maximum R^2 was selected as best model. The results showed that constructed ANFIS and ANN models were effectively able to predict SP. The comparison of developed models showed that ANFIS and ANN models are superior as compared with MR functions in estimating SP indices. In this study, option of estimating SP using proposed empirical relationship and models is acknowledged.

References

Drake, J.T. (2000). Communications phase synchronization using the adaptive network fuzzy inference system. Ph.D. Thesis, New Mexico State University, Las Cruces, New Mexico, USA.

Gardner, W.H. (1986). Water content. In: A. Klute Ed. Methods of Soil Analysis, Part 1—Physical and Mineralogical Methods, 2nd ed. American Society of Agronomy, Soil Science Society of America, Madison, WI, pp. 493–544.

Hagan M.T., and M.B. Menhaj. (1994). Training feed forward networks with the Marquard algorithm. IEEE Trans. Neural Networks 6, 861– 867.

Jang, J. S.R. (1993). ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Trans. Syst. Manag. Cyber. 23 (3), 665–685.

Jang, J. S.R. and C. T, Mizutani, E. (1997). Neuro-Fuzzy and soft computing: A computational approach to learning and machine intelligence. Prentice-Hall, Upper Saddle River, NJ.

Keeney, D. R. and R. B. Corey. (1963). Factors affecting the lime requirements of Wisconsin soils. Soil Sci. Soc. Amer. Proc. 27: 277-280.

Kisi, O. and E. Uncuoglu, (2005). Comparison of three backpropagation training algorithms for two case studies, Indian J. of Eng. and Materials Sci., Vol. 12.

Kisi, O. (2007). Development of Stream flow-Suspended Sediment Rating Curve Using a Range Dependent Neural Network. International Journal of Science and Technology Vol. 2, No 1, 49-61.

Lee E.S. (2000). Neuro-Fuzzy Estimation in Spatial Statistics. Journal of Mathematical Analysis and Applications, 249, 221-231.

Lippman, R. (1987). An introduction to computing with neural nets. IEEE ASSP Mag. 4, 4–22.

Matlab 7.4.0. (2007). Software for technical computing and model-based design. The MathWorks Inc.

Mcbratney, A.B. and I.O.A. Odeh, (1997). Application of Fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma, 77, p.85-113.

McClave, J. T., Dietrich, F. H., and Sinicich, T. (1997). Statistics. Prentice-Hall. Inc.

Milton, J. S., McTeer, P. M., and corbet, J. J. (1997). Introduction to statistics. McGraw-Hill.662 PP.

Mukopadhyay A.. (1999). Spatial Estimation of Transmissivity Using Artificial Neural Network. Ground Water, 37(3), 458-464.

Piotrowski J.A., F. Bartels, A. Salski, and Schmidt G. (1996). Geostatistical Regionalization of Glacial Aquitard Thickness in Northwestern Germany, Based on Fuzzy Krigging. Mathematical Geology, 28(4), 437-448.

Shaw W. M. (1952). Report on exchangeable hydrogen in soils. Interrelationships between calcium sorption, exchangeable hydrogen, and pH values of certain soils and subsoils. J. Assoc. Offic. Agr. Chemists 35: 597-621.

SPSS 15.0.0. (2006). Statistical analysis software (Standard Version). SPSS Inc.

Brian A. Schumacher. (2002). Methods for the determination of total organic carbon (TOC) in soils and sediments. NCEA-C- 1282 EMASC-001.

Zimmerman, H. J. (1996). Fuzzy Set Theory-and its applications. Boston, Kluwer Academic Publishers, p.435.

Table 1. Statistical analysis of the measured soil parameters

	SP (%)	OC (%)	Clay (%)	Silt (%)	Sand (%)
Minimum	28.00	0.08	6.00	16.00	3.00
Maximum	66.00	2.16	52.00	71.00	72.00
Range	38.00	2.08	46.00	55.00	69.00
Skew ness	-0.45	0.83	0.54	-1.58	1.33
Kurtosis	0.25	2.18	0.37	4.62	3.11
Mean	48.18	0.72	24.43	51.34	24.23
Std. dev	6.28	0.26	7.01	7.90	10.66

SP: Saturation Percentage OC: Organic Carbon

Table 2. Correlation coefficients (R^2) obtained from simple regression between SP and other measured parameters

model	Dependent	OC	clay	silt	sand
linear	SP	0.089	0.335	0.264	0.586
logarithmic	SP	0.083	0.333	0.267	0.528
power	SP	0.089	0.325	0.313	0.527
exponential	SP	0.09	0.318	0.304	0.616

Table 3. Predictive relationship for assessing SP, using available measured values.

Input combination	Predictive model	R^2
SP- clay	SP=35.517+0.519clay	0.335
SP- clay	SP=9.252+12.344Ln(clay)	0.333
SP- clay	$SP=20.571+ clay^{0.267}$	0.325
SP- clay	$SP=36.430e^{0.011clay}$	0.318
SP- sand	SP=59.103-0.451clay	0.586
SP- sand	SP=80.093-10.312Ln(clay)	0.528
SP- sand	$SP=96.031clay^{-0.226}$	0.527
SP- sand	$SP=61.025e^{-0.010clay}$	0.616

Table 4. Type and values of parameter used for training ANFIS

ANFIS parameter type	Value	
	ANFIS11	ANFIS12
Number of input	3	4
MF type	Trigonal function	Gussian functin
Number of MFs	2 2 2	2 2 2 2
Number of nodes	34	55
Number of linear parameters	8	16
Number of nonlinear parameters	18	16
Total number of parameters	26	32
Number of training data pairs	360	360
Number of checking data pairs	120	120
Number of fuzzy rules	8	16

Table 5. Statistical analysis for the best two ANN models in train and test period

Input combinations	Neuron number	Train		Test	
		R^2	RMSE	R^2	RMSE
clay, silt and sand	23	0.642	3.946	0.5878	3.6458
OC, clay, silt and sand	5	0.654	3.919	0.5886	3.6796

Table 6. Performance indices (R^2 and RMSE) for two preferred MR models for prediction of SP

Input combinations	R^2	RMSE
Clay, silt sand	0.564	3.793
OC, clay, sand and silt	0.577	3.7507

Table 7. Preferred ANFIS's parameter type and their performance indices

Input combinations	Membership functions type	Number of membership functions	train		test	
			R^2	RMSE	R^2	RMSE
clay, silt and sand	Triangular membership function (trimf)	2 2 2	0.649	3.886	0.607	3.585
OC, clay, silt and sand	gussian membership function (gussmf)	2 2 2 2	0.686	3.676	0.593	3.647

Figure 1. The study area

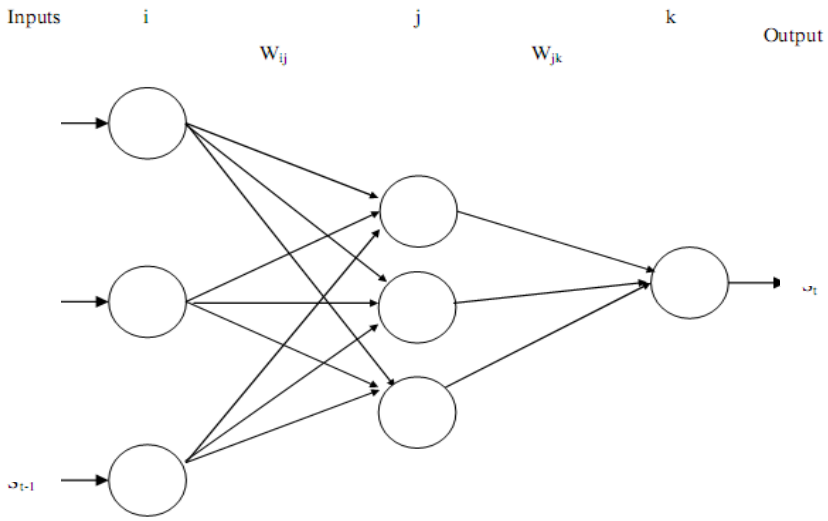


Figure 2. Architecture of three-layer feedforward network (Kisi 2007)

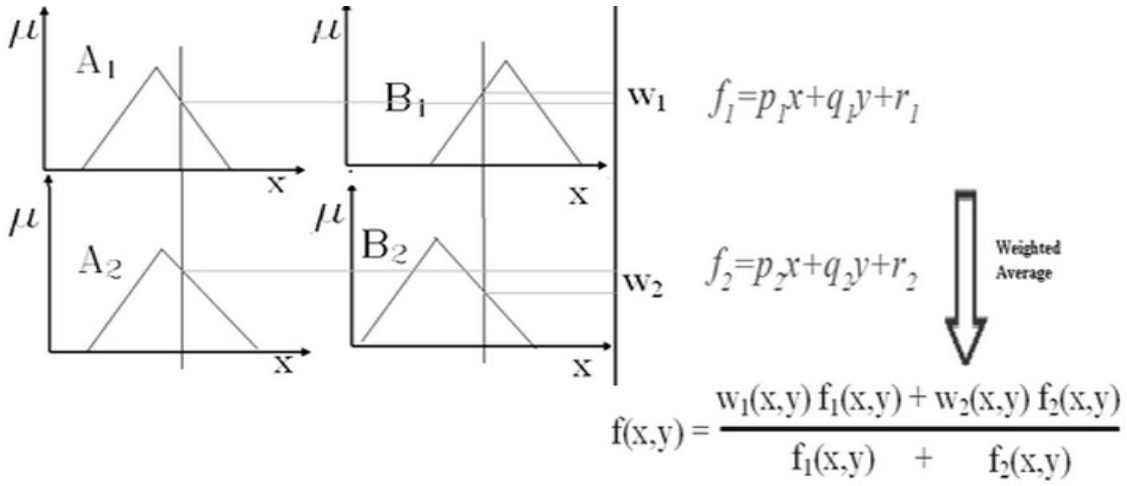


Figure 3. First-order Sugeno fuzzy model with two rules

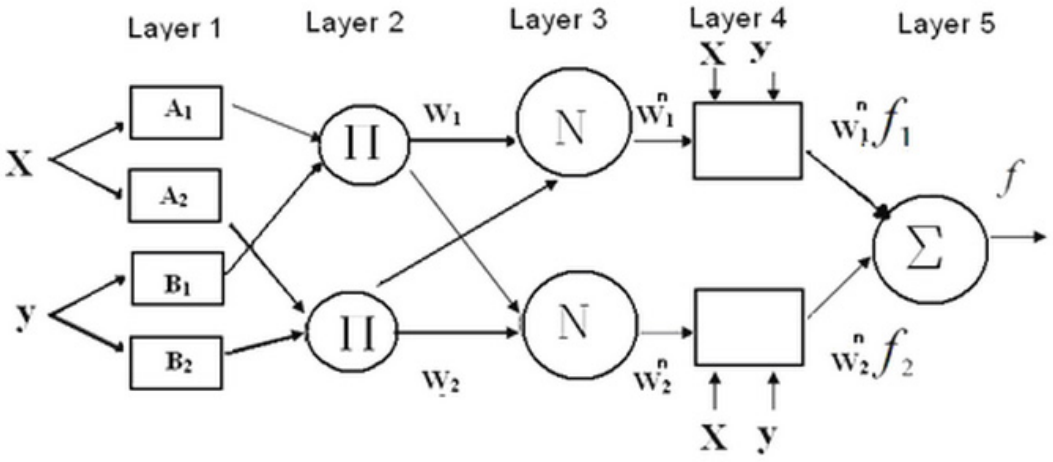


Figure 4. Equivalent ANFIS architecture.

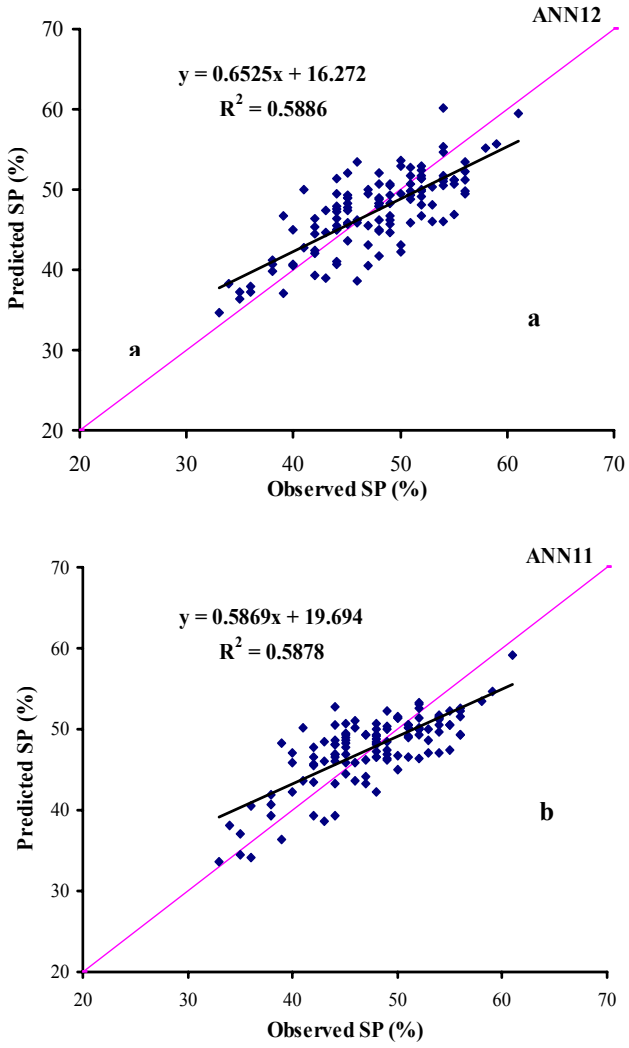


Figure 5. Observed and estimated SP (%) in test period in ANN models with 2 sets of inputs: **a)** OC, clay, silt and sand b**)** clay, silt and sand

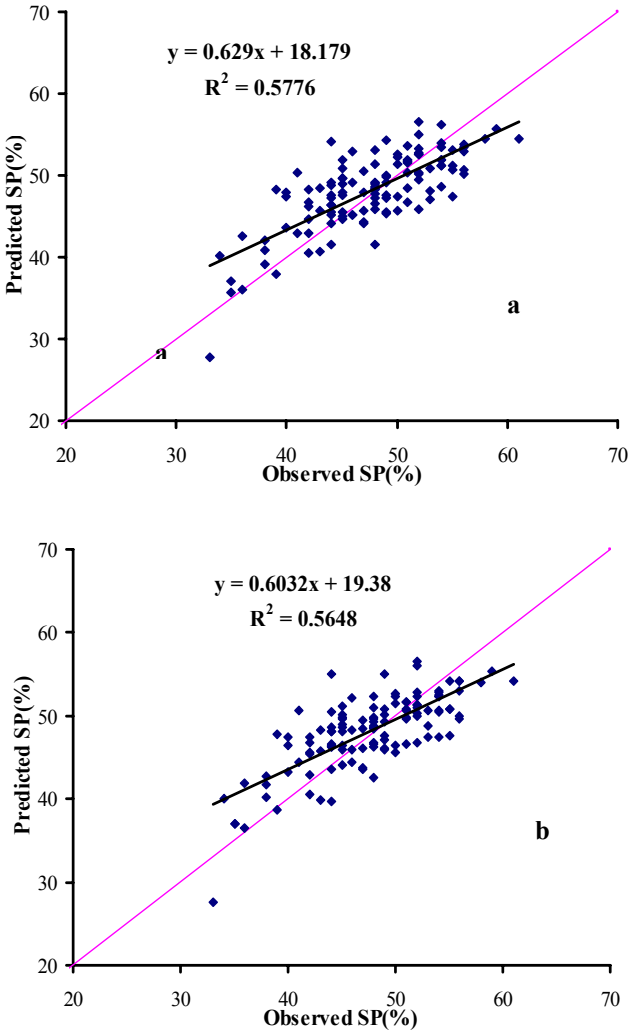


Figure 6. Observed and estimated SP (%) in test period in multiple regression models with 2 sets of inputs: **a)** OC, clay, silt and sand **b)** clay, silt and sand

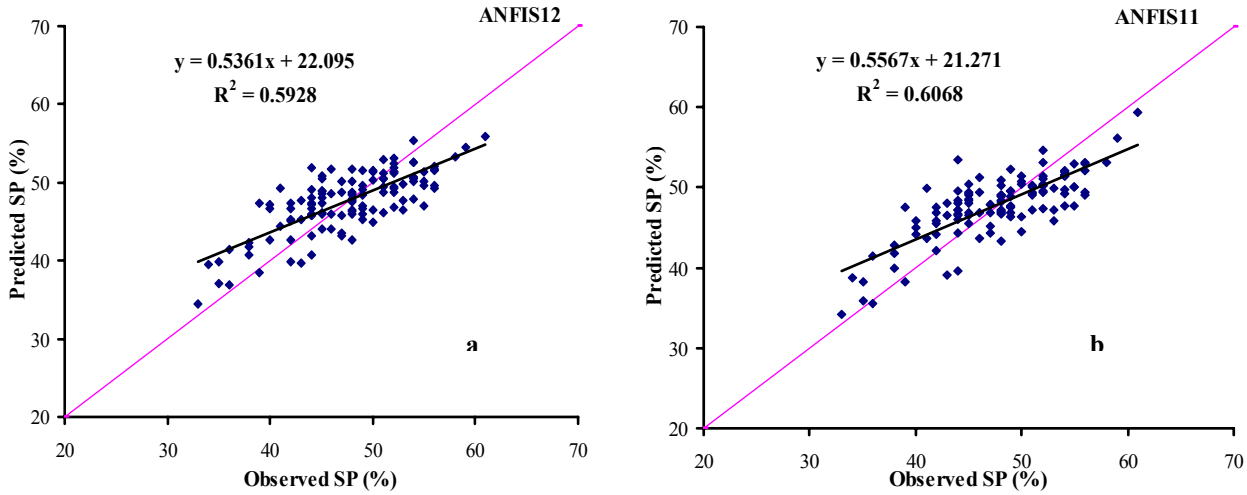


Figure 7. Observed and estimated SP (%) in test period in ANFIS models with 2 sets of inputs: **a)** OC, clay, silt and sand **b)** clay, silt and sand

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

An Optimised Method-Based an Improved Neural Network Classifier

Mengxin Li

Faculty of Information and Control Engineering, Shenyang Jianzhu University

Shenyang 110168, China

E-mail: limengxinf1972@yahoo.com.cn

Chengdong Wu

Northeastern University, Shenyang 110000, China

E-mail: wuchengdong@ise.neu.edu.cn

Hui Lin

Faculty of Information and Control Engineering, Shenyang Jianzhu University

Shenyang 110168, China

Abstract

A hybrid method is presented to accelerate network training for traditional BP networks and to improve the classification accuracy of features for automatic visual inspection of wood veneers. In order to achieve an optimal network structure, the uniform design method is employed to optimise the parameters taking advantage of typical experimental data and good data representation, and the optimal combination is confirmed using a nonlinear quadratic programming (NLPQL) from a response surface model., and the 'best' level-combination is obtained to further improve the performance of the hybrid classifier. By comparison, the classifier using the optimal factors shows more powerful performance with a classification accuracy of 98.99% and a fast speed, which means greater potential for practical applications.

Keywords: Hybrid classifier, Uniform design, Parameter Optimisation, Defect Inspection

1. Introduction

Production rates in a plywood factory are very high, with the wood sheets being conveyed at a speed of 2-3m/s and an interval of only one or two seconds is allowed for human inspection (Pham and Alcock,1996,p.45-52). This makes the workers extremely stressed and a little disturbance or loss of attention will result in a misclassification. Huber, Mcmilin and Mckinney(1985,79-82) made a series of experiments and found an accuracy of 68% with human inspection of boards. It is necessary to develop an automatic visual inspection system to relieve the human inspector and improve the classification performance.

Neural networks are of great interest due to their proven adaptability, parallel and distributed architecture and ability to learn. Backpropagation (BP) neural networks are widely used in various applications. Currently, the BP architecture is considered the most popular, effective and easy-to-learn model for complex, multi-layered network. However, there is still a need for improvement of the BP algorithms to overcome its shortcomings and to achieve a better structure of the network.

This research proposes a hybrid method to accelerate network training and to improve the classification accuracy of features for automatic visual inspection of wood veneers. In order to achieve an optimal network structure, the uniform design method is employed to optimise the parameters, and the 'best' level-combination is obtained to further improve the performance of the neural network classifier. Section 2 describes a hybrid BP algorithm to tackle slow convergence of BP algorithms, and feature extraction from the defect images. It also discusses the normalisation of feature values and the encoding of the classifier output. Then the structure of the classifier is determined. Section 3 adopts uniform design for parameter optimisation of the neural network. A uniform design table is formed and a regression model designed between the response and the factors to find the best results among the all the responses and corresponding level-combinations of the factor values. Section 4 presents the results through a comparison between the improved BP

network and the traditional BP network. With the optimal parameter combination, the performance of the classifier is further improved. Finally, Section 5 draws conclusions of the research and recommends further work.

2. Improved BP Network

Typical problems of the backpropagation algorithms are the slow speed of convergence and the possibility of a local minimum of the error function. Thus the following three improved methods are used to overcome these weaknesses. The first two methods are applied separately or combinatively (Peng and Mo,1999,p.169-171). The third method is proposed as a new method to reach the desired error quickly. The hybrid method of this research makes use of their advantages adequately.

2.1 Improved BP Algorithm

Additional momentum

This method considers both the effects of the gradient direction and the influences of change tendency on the gradient direction. To some extent, it accelerates the adjustment process and avoids getting into a local minimum.

Self-adaptive learning rate

The self-adaptability learning rate is induced to solve issues such as unsteadiness caused by a very high learning rate and long training time caused by a very low learning rate. The self-adaptability learning rate can reach a reasonably high efficiency while stable training is maintained. Note that learning rate is a sensitive parameter, and it has to change in a small area in order to avoid the training failure.

Dynamic error segmenting

The training process actually is looking for the global least value on the error surface. If the initial weights are given with some less value randomly, the error gradient easily gets into a local minimum and results in error vibration at the beginning of training. Usually after several vibration periods, the adjustment direction may tend to reach the global minimum. Assuming that the neural network is trained according to the error accuracy desired at the beginning, there should be such a long time to meet the training needs and accordingly the generalisation of the network gets worse. To overcome this shortcoming, a dynamic error segmenting method is presented. First, the training process begins with a larger error to accelerate to get the global minimum. Then, the error is lowered gradually till the desired error is satisfied. The error is divided into 3 to 4 grades.

In this research, the initial error is set to 0.4. According to geometric proportion error training, the error is divided into 4 grades. First, an error amplification ratio is set: A=0.4/*ERROR*, and *ERROR*=0.03, where *ERROR* is the desired error accuracy. Then, geometric proportion is set: $B=1/A^{1/(n-1)}$, where n is a known grade number. At last, the segmented errors are determined as 0.4, 0.4*B, 0.4*B^2, …, *ERROR*.

2.2 Structure of the Improved BP Neural Network

2.2.1 Features extraction from defect images

The images of veneer sheets acquired consist of 512*512 picture elements (pixels), each with a grey level value between 0 (black) and 255 (white) inclusive. Once the defect area is found, a window of size 60 pixels in the X-direction and 85 pixels in the Y-direction is placed on the defect. The origin of the window is in the window of the defect. The size of this window corresponds to 3 square centimetres on the sheet and is large enough to cover any of the defects under consideration except certain large barks.

The grey level frequencies are recorded from the feature extraction window. The grey level histograms for samples belonging to the same defect have similar shapes. 17 typical features which represent the wood veneer defects are extracted from their image of every sample for training and testing the neural network. The features of wood veneers are shown in Table 1.

2.2.2 Normalisation of feature values

Because of different scales and ranges of the features for wood veneer defects, normalising the data is important to ensure that the distance measure accords equal weight to each variable. The features are scaled between -1 and 1 for use as the network input. To perform the normalisation, each image feature is converted to the standard distribution by the following transformation

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

where μ is the mean and σ the standard deviation of the original distribution, x is the original feature value and Z is a new transformed variable with a standard normal distribution (mean 0 and standard deviation=1). This ensures that 99.04% of the data will lie within the range ±3. The Z values are further divided by 3 to limit the input values between -1 and 1. This method of normalisation was used by Kjell, Woods and Freider (1995,p.1222-1226). The normalised

feature data is eventually fed to the neural network for training.

2.2.3 Encoding of the classifier output

The neural network has 13 output neurons, each of which corresponds to one defect type as indicated in Table 2. Because the output values of a neural network are usually real numbers, it is essential to convert them into a binary form suitable for the classification of defects. This can be realised with several methods. The maximum method is chosen here, which sets the highest output value to 1 and the others to 0. This means that the defect class chosen corresponds to the output neuron with the highest value.

The network has 17 input neurons, each corresponding to an extracted feature, while 13 output neurons correspond to the defect classes respectively. This research uses one hidden layer with a number of different neurons to determine the suitable network. Because networks with biases, a hidden layer and an output layer are capable of approximating any function with a finite number of discontinuities. The hidden layer uses the tansig activation function. Initial weights and biases are generated randomly. The output layer uses the purelin activation function, and output layer determines the class that the features in the input layer belong to. The number of neurons in the hidden layer is 50 determined by a series of experiments.

3. Uniform Design for Network Parameter Optimisation

Taking into account the difficulties of determining the neural network parameters, uniform design (UD) is introduced to solve parameter optimisation of the neural network. UD is an experimental design method proposed by Fang (1980,p.363-372). It has been recognised as an important space-filling design, which plays a key role in large systems engineering design. UD is equivalent to generating a set of design points that are uniformly scattered in the experiment domain, which reflects the main features of the system. It can solve optimisation problems by finding the maximal or minimal value for the fitness or an error function (Xie and Fang, 1997,p.101-111).

All the UD designs are based on U-type design such as UG-type and UL-type. U-type design gives a good structure. Suppose that there are s factors with q levels for each of the factors. There are q^s level-combinations.

DEFINITION 1. A U-type design for simplicity denoted by $U_n(n^s)$, is a matrix of n rows and s columns. A U-type design can be considered as a design with n levels and s factors.

UD is often expressed as a table, called a UD table, and a number of UD tables can be found on the UD web. Note that for a given set of (n, s), the corresponding UD is not unique. Two U-type designs are called equivalent if one can be obtained from another by permuting the rows and the columns.

3.1 Forming a UD Table

UD defines the minimum set of parameter level-combinations to be tested in an experiment in order to gain an estimate of the average effects of each parameter. In the structure of the neural network to be designed, two parameters are considered, learning rate and the number of hidden layer neurons that play an important role in the classifier performance. Therefore the UD table is built in Table 3.

The six levels marked by 1,2…,6, are transformed into the real levels of the factors and record the corresponding yield Y. Specifically the heading of (1, 2) represents the UD table for the two factors, i.e. the number of hidden neurons and learning rate. Such a table can be found from the UD-web. The heading (X_1 and X_2) represents the actual experimental values for the two factors. The last column Y gives the response of the experimental results from the misclassification accuracy.

3.2 Building the Response Surface Model (RSM)

RSM (Li and Wu, 2001,p.68-73) is an efficient tool for modelling on a few observations. In general, for n design parameters $x=(x_1, x_2, \ldots, x_n)$, the system response Y can be written as

$$Y=f(x)+\varepsilon \tag{2}$$

where ε is a random error component. An accurate model of the true system requires a model of degree two or higher to approximate the curvature in the actual surface. In most cases, the second order model is an adequate approximate. The second order (or quadratic) model is

$$Y = \beta_0 + \sum_{i=0}^{n} \beta_i X_i + \sum_{i=0}^{n} \sum_{j=0}^{n} \beta_{ij} X_i X_j + \varepsilon \tag{3}$$

For a system with two design parameters (n=2), it can be expressed by the following equation

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_{11} X_1^2+\beta_{22} X_2^2+\beta_{12} X_1 X_2+\varepsilon \tag{4}$$

In the equation, the linear and quadratic components and the first order interaction are included. A reasonable approximation of the true response of most systems can be fitted using the second order model. The model coefficients can be estimated through a response surface design.

To estimate the coefficients of a second order model, a design of experiment with at least three levels per parameter is required since two points can only decide a straight line.

The coefficients of the second order model can be estimated for a system with two parameters if none of the interactions

are included. The fitted model is described by substituting n=2 and ignoring the interactions in equation (4) to give equation (5).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 \tag{5}$$

where β_0 is the mean of all the observed response values. Many runs in an UD average the random error to zero which is why the ε term has been dropped. The two model coefficients are the linear components (β_1, β_2) and the quadratic components (β_{11}, β_{22}) of the two parameters. Ignoring the higher order components and interactions gives an advantage to the fitted model. This model avoids fitting a surface exactly through all the observed data which is affected by random noise.

3.3 Modelling Neural Network Performance

Using Table 1 and the method described, the model coefficients are obtained by the following formula.

$$B = Y^{-1} \times X$$

$$= [7.9995\ 0.0511\ 78.2435\ -0.0036\ -136.8136]'$$

A model with no interactions which describes the classifier's performance using the estimated values can be written as

$$Y = 7.9995 + 0.0511 \times x_1 + 76.2435 \times x_2 - 0.0036 \times x_1^2 - 136.8136 \times x_2^2 \tag{6}$$

3.4 Finding the Optimal Settings

Finding the process of optimal parameters is actually a problem of the sequential quadratic programming algorithm NLPQL (Nonliear quadratic programming) of Schittkowski [12]. The NLPQL algorithm is to solve nonlinear mathematical programming problems with equality and inequality constraints. Therefore in this paper, the following formulae need to be satisfied with NLPQL.

$$\begin{cases} \min(Y) \\ Y \geq 0 \\ 0 < x_1 \leq 60 \\ 0 < x_2 \leq 0.06 \end{cases} \tag{7}$$

Using the fitted model and the constraints, $0<x_1\leq 60$ and $0<x_2\leq 0.06$, an optimal setting for each design is found for the classifier: Learning rate = 0.0127. Number of neurons in the hidden layer = 57. For the learning rate, two decimal places are kept, e.g. 0.01(See Fig.1) . The results indicate a small learning rate and less than 60 neurons in the hidden layer give the best classification performance, which are consistent with above optimal results. It is worth noting that the learning rate cannot be 0, which is meaningless to the network learning.

4. Experimental Results

In the simulation experiments, 80% of the 232 samples are selected at random to form the training set and the remaining 20% for the test set. Experiments are carried out in 3 groups. 17 features are considered as input for training the improved neural network.

The classification accuracy is 98.99 %for the test set, and short running time of 9.67s with a larger amount of sample data (i.e. 232 samples). The running time is taken largely to train the sample data, while the testing time is only in milliseconds. It therefore has greater potential for practical applications with respect to both accuracy and real-time inspection. In comparison with the traditional BP neural network, the improved BP neural network presented is more accurate and has a faster convergence speed. Figure 2 shows the training process of the improved BP network.

5. Conclusions

The improved BP algorithm improves the accuracy and reduces the possibility of getting into a local minimum. It is also inspiring that the training time has been decreased from 92.69s to 9.67s. The improved BP classifier for wood veneer defects is more effective. Furthermore, the UD method has optimised the network parameters with effective algorithms adopted to further improve the accuracy to 98.99%. For further work, more parameters of the neural networks such as the epochs, running time or momentums could also be considered to further be optimised.

References

Fang K.T. (1980). The Uniform Design: Application of Number-Theoretic Methods in Experimental Design. *Acta Math. Appl. Sinica*, 3, 363-372.

Huber H.A., Mcmilin C.W. and Mckinney J.P. (1985). Lumber Defect detection Abilities of furniture Rough Mill Employees.*Forest Products Journal*, 35 (11/12), 79-82.

Kjell B., Woods W.A. and Freider O (1995). Information Retrieval Using Letter Tuples with Neural Network and Nearest Neighbor Classifiers.*IEEE International Conference on Systems, Man and Cybernetics*, Vancouver, Canada, 1222-1226.

Li M.X. and Wu C.D.: Parameter Optimization of a Neural Network Classifier Based on Response Surface Method. *The 20th Chinese Control Conference,* Dalian, China, (2001)68-73.

Peng M.S. and Mo Y.L (1999). Texture Classification Based on Mathematical Morphology and Neural Network. *Journal of Shanghai University*, 5(2), 169-171.

Pham D.T. and Alcock,R.J (1996). Automatic Detection of Defects on Birch Wood Boards. *Journal of Process Mechanical Engineering*, 45-52.

Xie M.Y. and Fang K.T. (1997). Admissibility and Minimaxity of the Uniform Design Measure in Nonparametric Regression Models. *International Symposium on Contemporary Multivariate Analysis and Its applications*, Hong Kong.101-111.

Table 1. Typical features for wood veneers

Code	Features
1	Mean grey level
2	Mode grey level
3	Median grey level
4	Standard deviation
5	Skewness
6	Kurtosis
7	Number of dark pixels
8	Number of bright pixels
9	Lower grey level
10	Higher grey level
11	Tail length on the dark side
12	Tail length on the bright side
13	Number of edge pixels (threshold=μ)
14	Number of pixels (threshold=μ-2δ)
15	Number of edge pixels for feature 14
16	Number of pixels (threshold=μ+2δ)
17	Number of edge pixels for feature 16

Table 2. Defect classes for wood veneer

Defect class	Desired outputs presentation
Holes	1 0 0 0 0 0 0 0 0 0 0 0 0 0
Pin knots	0 1 0 0 0 0 0 0 0 0 0 0 0 0
Rotten knots	0 0 1 0 0 0 0 0 0 0 0 0 0 0
Roughness	0 0 0 1 0 0 0 0 0 0 0 0 0 0
Splits	0 0 0 0 1 0 0 0 0 0 0 0 0 0
Streaks	0 0 0 0 0 1 0 0 0 0 0 0 0 0
Discoloration	0 0 0 0 0 0 1 0 0 0 0 0 0 0
Coloured streaks	0 0 0 0 0 0 0 1 0 0 0 0 0 0
Bark	0 0 0 0 0 0 0 0 1 0 0 0 0 0
Worm holes	0 0 0 0 0 0 0 0 0 1 0 0 0 0
Curly grain	0 0 0 0 0 0 0 0 0 0 1 0 0 0
Clear wood	0 0 0 0 0 0 0 0 0 0 0 1 0 0
Sound knots	0 0 0 0 0 0 0 0 0 0 0 0 1 0

Table 3. U_6 (6^2) and related design

No. of runs	1	2	X_1	X_2	Y
1	5	5	50	0.05	5.1223
2	4	1	40	0.01	5.0505
3	2	2	20	0.02	9.0909
4	3	6	30	0.06	6.0606
5	1	4	10	0.04	9.0909
6	6	3	60	0.03	3.0303

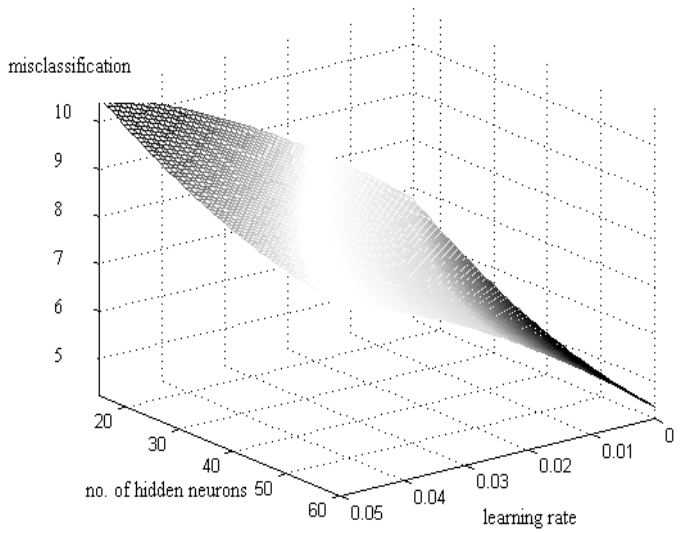


Figure 1. Response with the learning rate and number of hidden neurons

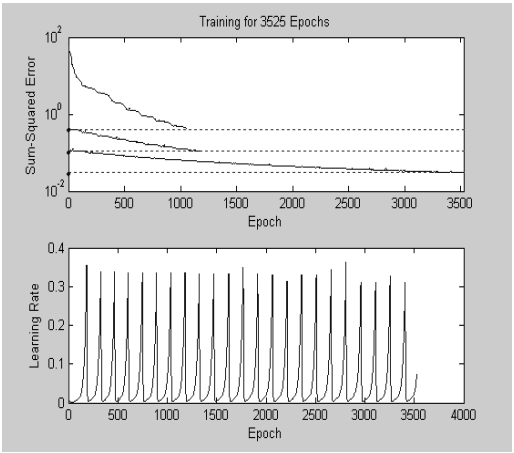


Figure 2. Training of Improved BP Network in Group 2

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Relative Truth Degree of Logic Formulas and Graded Fuzzy Logic

Li Fu

Department of Mathematics, Qinghai Nationality College, Xining 810007, China

Tel: 86-971-829-2658 E-mail: fl0971@163.com

The research is financed by The Education department chunhui program (Z2006-1-81001)

Abstract

The concept of logic proposition induced functions is proposed in the present paper, then the concept of relative truth degree of propositions with respect to a logic theory Γ is introduced by means of infinite product of evenly distributed probability spaces and integrated semantics respectively w.r.t. discrete and continuous situations, and a graded approximate reasoning theory is established. Next, theory of consistency degrees of finite logic theories is also proposed. Finally, the simple application of graded fuzzy logic in fuzzy inference is given by examples.

Keywords: Logic proposition induced functions, Truth degree of propositions, Approximate reasoning, Consistency degree of finite logic theories

1. Preliminaries

Let $S = \{p_1, p_2, \cdots\}$ be a countable set, and $\overline{0}$ be a special element not contained in S, and $F(S)$ be the free algebra of type $(\&, \vee, \wedge, \rightarrow)$ generated by $S \cup \{\overline{0}\}$, where $\&, \vee, \wedge, \rightarrow$ are binary operations respectively, i.e.,

(i) $S \subset F(S), \overline{0} \in F(S)$

(ii) If $A, B \in F(S)$, then $A \& B, A \vee B, A \wedge B,$ and $A \rightarrow B \in F(S)$

(iii) Elements of $F(S)$ can be obtained by (i)and (ii) in a finite step of calculus.

Members of S are called atomic propositions or atomic formulas (briefly, atoms), that of $F(S)$ are propositions or formulas or wffs. $\overline{0}$ is called contradiction. Moreover, in what follows $\neg A = A \rightarrow \overline{0}$. For classical two valued propositional logic system L, the following are axiom schemes.

(L1) $A \rightarrow (B \rightarrow A)$

(L2) $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$

(L3) $(\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$

The deduction rule is modus ponens (MP), i.e., from $A \rightarrow B$ and A, B follows. The following theorem holds in L (Ref A.G. Hamilton.(1979)):

Theorem 1 In classical propositional logic system L, suppose that $\Gamma \subset F(S), A, B \in F(S)$,and $\Gamma \cup \{A\} \vdash B,$ then $\Gamma \vdash A \rightarrow B,$ and vice versa. In L, we have $A \vee B = \neg A \rightarrow B$, $A \& B = A \wedge B$

$= \neg(\neg A \vee \neg B)$.

In the fuzzy prepositional logic system L^* the following are axiom schemes

(L^*1) $A \rightarrow (B \rightarrow A \wedge B)$

(L^*2) $(\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$

(L^*3) $(A \rightarrow (B \rightarrow C)) \rightarrow (B \rightarrow (A \rightarrow C))$

(L^*4) $(B \rightarrow C) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$

(L^*5) $A \rightarrow \neg\neg A$

(L^*6) $A \to A \vee B$

(L^*7) $A \vee B \to B \vee A$

(L^*8) $(A \to C) \wedge (B \to C) \to (A \vee B \to C)$

(L^*9) $(A \wedge B \to C) \to (A \to C) \vee (B \to C)$

(L^*10) $(A \to B) \vee ((A \to B) \to \neg A \vee B)$

where $A \vee B = \neg((((A \to (A \to B)) \to A) \to \neg((A \to B) \to B))$, $A \wedge B = \neg(\neg A \vee \neg B)$. The deduction rule in L^* is also MP, and we have the following

Theorem 2 Suppose that $\Gamma \subset F(S), A, B \in F(S),\quad \Gamma \cup \{A\} \vdash B,$ then $\Gamma \vdash A^2 \to B,$ and vice versa, where $A^2 = A \,\&\, A = \neg(A \to \neg A).$ In L^* we have $A \wedge B = \neg(\neg A \vee \neg B)$, $A \,\&\, B = \neg(A \to \neg B),$ and $A \vee B = \neg((((A \to (A \to B)) \to B) \to B) \to \neg((A \to B) \to B))$ (see Ref G.J.Wang(2003),S.M.Wang, etc(2003)).

2. Truth degrees of propositions in L and L^*

Let $A = A(p_1, \cdots, p_n)$ be a formula in $F(S)$, then A induced a function $\overline{A}(x_1, \cdots, x_n) : V^n \to V,$ where $V = \{0,1\}$ or $V = [0,1]^n$, as follows: the value of $\overline{A}(x_1, \cdots, x_n)$ can calculated in V by connecting $x_1, \cdots, x_n$ with the operators $\&, \vee, \wedge, \to$ and $\neg$ in the same way as A is obtained by connecting $p_1, \cdots, p_n$ with the logic connectives $\&, \vee, \wedge, \to$ and $\neg$ respectively. For example, if $A = (p_1 \to p_2) \to \neg p_3 \vee p_4$, then $\overline{A}(x_1, x_2, x_3, x_4) = (x_1 \to x_2) \to (1 - x_3) \vee x_4$. In case $v = \{0,1\}, \overline{A}(0,0,1,0) = (0 \to 0) \to (1-1) \vee 0 = 0 \to 0 = 1,$ and in case $V = [0,1],$ $\overline{A}(0.8, 0.3, 0.9, 0.2) = (0.8 \to 0.3) \to (1 - 0.9) \vee 0.2 = 0.3 \to 0.2 = 0.7,$ where $\to$ is the R_0 -implication operator, $R_0 = a \to b = \begin{cases} 1, a \le b \\ (1-a) \vee b, a > b \end{cases}$,(see Ref G.J.Wang.(2003,1998)).

In the following $\overline{A}(x_1, x_2, ...x_n) : V^n \to V$ will be called the **proposition A induced function.**

Since $F(S)$ is the free algebra generated by $S \cup \{\overline{0}\}$ and for every valuation $v : F(S) \to V, v(\overline{0}) = 0$ holds, hence v is determined by $v \mid S,$ i.e., v is determined by $(v(p_1), v(p_2), \cdots) \in V^\infty$. Let μ be the infinite product of evenly distributed probability measure on $V = \{0,1\}$,and Ω be the set consisting of all valuations of $F(S)$,then there is a one-one correspondence between Ω and $\{0,1\}^\infty$. In fact, assume that $v(p_k) = v_k \quad (k = 1,2,...)$,then $\vec{v} = (v_1, v_2, ...) \in [0,1]^\infty$.

Conversely, assume that $\vec{v} = (v_1, v_2, ...) \in [0,1]^\infty$, then there exists an unique $v \in \Omega$, such that $v(p_k) = v_k, (k = 1,2,...)$,hence $\varphi : \Omega \to \{0,1\}^\infty$ is a bijection, where $\varphi(v) = \vec{v}$.

Let $\varphi : \Omega \to \{0,1\}^\infty$ be the bijection, then $\forall A \in F(S),$ define $[A]$ and $\tau(A)$ as follows respectively: $[A] = \left\{ \vec{v} \in X \,\middle|\, v \in \Omega, v(A) = 1 \right\}, \tau(A) = \mu([A]), \tau(A)$ is called the truth degree of A . It is clear that $0 \le \tau(A) \le 1$

In case $V = [0,1], \forall A \in F(S),$ call $\tau(A) = \int_\Delta \overline{A}(x_1, \cdots, x_n) dw = \int_0^1 \cdots \int_0^1 \overline{A}(x_1, \cdots, x_n) dx_1 \cdots dx_n$

integrated truth degree of A, where $\Delta = [0,1]^n$ and $dw = dx_1 \cdots dx_n$ whenever $A = A(p_1, \cdots, p_n).$ Specifically, we can define integrated truth degree of A with respect to diverse implication operators. In reference P.H´ajek(1998), you can the following definition:

Definition 1 Suppose that $A(p_1, \cdots, p_n)$ is a formula in L^*,define

$\tau_R(A) = \int_0^1 \cdots \int_0^1 \overline{A}(x_1, \cdots, x_n) dx_1 \cdots dx_n$, then $\tau_R(A)$ is called **integrated truth degree of A with respect to implication operator R.**

Example 1 $\tau(p) = \int_0^1 x dx = \frac{1}{2}$

$\tau(p \vee q) = \int_0^1 \int_0^1 (x \vee y) dx dy = \int_0^1 \int_y^1 x dx dy + \int_0^1 \int_x^1 y dx dy = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$

$\tau(p \wedge q) = \int_0^1 \int_0^1 (x \wedge y) dxdy = \int_0^1 \int_y^1 y dxdy + \int_0^1 \int_x^1 x dxdy = \frac{1}{3}$

$\tau(p \wedge \neg p) = \int_0^1 (x \wedge (1-x) dx = \int_0^{\frac{1}{2}} (1-x) dx + \int_{\frac{1}{2}}^1 x dx = \frac{3}{4}$

$\tau(p_1 \wedge \cdots \wedge p_n) = \int_0^1 \int_{x_1}^1 x_1 dxdy + \cdots + \int_0^1 \int_{x_n}^1 x_n dxdy = \frac{1}{n+1}, \tau(p_1 \vee \cdots \vee p_n) = \frac{n}{n+1}$

$\tau(p \to p) = \int_0^1 \int_y^1 (1-x+y) dxdy + \int_0^1 \int_x^1 dxdy = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$, where $\to$ is Lukasiewicz implication operator $R_{Lu}(a,b) = (1-a+b) \wedge 1, a, b \in [0,1]$.

Example 2 Suppose that $A = p_1 \to (\neg p_2 \to \bar{0})$, let $\to$ be R_0 implication operator, then $\tau(A) = \int_{[0,1]^2} (x_1 \to x_2) dx_1 dx_2 = \int_{[0,1]^2} (x \to y) dxdy$. Divide $[0,1]^2$ into 3 parts as follows

$\Delta_1 = \{|(x,y) \in [0,1]^2 | x \le y\}, \Delta_2 = \{|(x,y) \in [0,1]^2 | y < x, 1-y \le x\}$ $\Delta_3 = \{|(x,y) \in [0,1]^2 | y < x < 1-y\}$.

Then $[0,1]^2 = \Delta_1 \cup \Delta_2 \cup \Delta_3$, and $\Delta_1, \Delta_2, \Delta_3$ are disjoint from each other, $\tau(A) = \int_{\Delta_1} 1 dxdy + \int_{\Delta_2} ((1-x) \vee y) dxdy + \int_{\Delta_3} ((1-x) \vee y) dxdy = \frac{1}{2} + \int_{\Delta_2} y dxdy + \int_{\Delta_3} (1-x) dxdy = \frac{1}{2} + \frac{1}{8} + \frac{1}{8} = \frac{3}{4}$.

3. Graded approximate reasoning theory

Suppose that $V = \{0,1\}, \Gamma$ is a finite theory in L, i.e., Γ is a finite subset of $F(S)$, say, $\Gamma = \{A_1, \cdots, A_n\}$, and $B \in F(S)$.

Definition 2 Define $\tau(\Gamma \vdash B) = \tau(A_1 \to (A_2 \to \cdots (A_n \to B) \cdots))$, call $\tau(\Gamma \vdash B)$ **truth degree of** B **relative to** Γ of the fact that Γ implies B in L, simply, **the truth degree** $\Gamma \vdash B$.

Theorem 3 $\Gamma \vdash B$ holds if and only if $\tau(\Gamma \vdash B) = 1$.

Proof Assume that $\Gamma \vdash B$ holds, then it follows from the deduction theorem of L that $\tau(A_1 \to (A_2 \to \cdots (A_n \to B) \cdots))$ is a theorem, hence it follows from Definition 1 that $\tau(\Gamma \vdash B) = 1$. Conversely, assume that $\tau(\Gamma \vdash B) = 1$, we can prove $\Gamma \vdash B$ similarly.

Theorem 3 told us from Γ can imply B, then Γ implies B is a tautology, i.e., its truth degree equals 1 in L.

Example 3 If $\Gamma = \{p \to q, \neg q \to r\}$, and $B = \neg p$, then $\tau(\Gamma \vdash B) = (\tau(p \to q) \to ((\neg q \to r) \to \neg p)) = \frac{3}{4}$, in fact, there are 8 possible (p,q,r) in $\{0,1\}^3$ and only 2 of them, i.e., (1,1,0)and (1,1,1) imply that $(p \to q) \to ((\neg p \to r) \to \neg p)$ equals 0.

Suppose that $V = [0,1]$ and $\Gamma = \{A_1, \cdots, A_n\}$, and B are finite theory and a formula in $Ł^*$ respectively, and $Ł^*$ is complete (Ref D.W.Pei,etc(2002)).

Definition 3 Define $\tau(\Gamma \vdash B) = \tau(A_1^2 \to (A_2^2 \to \cdots (A_n^2 \to B) \cdots))$, and call $\tau(\Gamma \vdash B)$ **truth degree** of the fact that Γ implies B in $Ł^*$.

It is similar with theorem 3, for integrated truth degrees of formulas, we have:

Theorem 4 If $\Gamma \vdash B$ holds, then $\tau(\Gamma \vdash B) = 1$. Conversely, if $\tau(\Gamma \vdash B) = 1$, then $E = A_1^2 \to (A_2^2 \to \cdots (A_n^2 \to B) \cdots)$ is an almost tautology, i.e., the induced function $\bar{E}$ of E equals 1 almost every where in $[0,1]^m$, assume E contains m atomic propositions.

Proof Assume that $\Gamma \vdash B$ holds, then it follows from Theorem 2 that $E = A_1^2 \to (A_2^2 \to \cdots (A_n^2 \to B) \cdots)$ is a theorem of $Ł^*$, hence it follows from Definition 2 that $\tau(\Gamma \vdash B) = 1$. Conversely, assume that $\tau(\Gamma \vdash B) = 1$, then it follows from the definition of integrated truth degree of E that the integral of $\bar{E}$ equals 1, hence the function $\bar{E}$ equals 1 almost everywhere in $[0,1]^m$.

Theorem 4 means the theory of truth degrees of propositions is uniform with deduction theorem of finite theories in $Ł^*$.

This fact is significant, it can be used to discuss deduction theorem by truth theory, and to propose a theory of consistency degrees of finite theories in L^* in section 5, except that the concept of truth degrees of formulas in two valued logic has to be substituted by the concept of integrated truth degrees of formulas.

Example 4 If $\Gamma = \{p\}, B = \neg q,$ then it follows from the basic properties of R_0 -implication operator that, $\tau(\Gamma \vdash B) = \int_0^1 \int_0^1 (x \otimes x \to (1-y))dxdy = \frac{19}{24}$,where $\otimes : [0,1]^2 \to [0,1]$ is a t -norm such that $(\otimes, \to)$ is a residuated pair,(see Ref G.J.Wang,etc(2003)),that is, $a \otimes b = \begin{cases} a \wedge b, a + b > 1 \\ 0, a + b \le 1 \end{cases}$, and hence $a \otimes a = \begin{cases} a, a > \frac{1}{2} \\ 0, a \le \frac{1}{2} \end{cases}$.

In [9] a pseudo-metric was defined on F(S) in L as follows:

$\rho(A,B) = \int_0^1 \cdots \int_0^1 \left| \overline{A}(x_1, \cdots, x_n) - \overline{B}(x_1, \cdots, x_n) \right| dx_1 \cdots dx_n$, where $x_1, \cdots, x_n$ are the atoms simultaneously appeared in A and B , and $\rho(A,B)$ even A and B can contain different atoms , can use extensions A^* and B^* instead of A and B correspondingly , such that A^* and B^* contain one and the same group of atoms. e.g., if $A = p_1, B = p_2$ then let $A^* = (p_1 \vee p_2 \to p_1 \vee p_2) \to p_1, B^* = (p_1 \vee p_2 \to p_1 \vee p_2) \to p_2$, we see that $A^* \square A, B^* \square B$.

Definition 5 Let $A, B \in F(S), \xi_R(A,B) = \int_\Delta R(\overline{A}, \overline{B}) \wedge R(\overline{B}, \overline{A}) dw$ is called R **- integral similar degree** between A and B , where $dw = dx_1 \cdots dx_n$ and $\Delta = [0,1]^n$, R is implication operator. If $\xi_R(A,B)$ = 1, then A and B is R **-integral similar** . Such, we have $\rho(A,B) = 1 - \int_\Delta (\overline{A} \to \overline{B}) \wedge (\overline{B} \to \overline{A}) dw = 1 - \xi_R(A,B)$. Furthermore, by the concept of truth degrees of formulae of $F(S)$, in L^* , **a pseudo-metric** can be defined as follows: $\rho(A,B) = 1 - \tau(A \to B) \wedge \tau(B \to A), A, B \in F(S)$.

4. Consistency degrees of finite theories

It is well known that a theory Γ (i.e., a set of wffs) is consistent if the contradiction $\overline{0}$ is not a conclusion of Γ , i.e., $\Gamma \vdash \overline{0}$ does not hold, otherwise Γ is inconsistent. Hence we see that

if $\Gamma \vdash \overline{0}$ is (fully) true, then Γ is (fully) inconsistent.

Suppose that we can in certain way measure the truth degree, say, α ,of "$\Gamma \vdash \overline{0}$", then it is natural to call α the inconsistency degree of Γ , and the consistency degree of Γ can then be defined to be $1 - \alpha$. In this sub-section we propose the concept of consistency degrees of finite theories in accordance with this idea.

Definition 6 Suppose that $\Gamma = \{A_1, \cdots, A_n\}$ is a finite theory in L , let $\Gamma \vdash \overline{0} = A_1 \to (A_2 \to (\cdots (A_n \to \overline{0}) \cdots))$, and define $\xi(\Gamma) = 1 - \tau(\Gamma \to \overline{0}) = 1 - \tau(A_1 \to (A_2 \to (\cdots (A_n \to \overline{0}) \cdots)))$,

then $\xi(\Gamma)$ is called the **consistency degre**e of Γ .

Example 5 (i) Suppose that p is an atomic formula, then $\xi(p) = 1 - \tau(p \to \overline{0}) = 1 - \tau(\neg p) = 1 - \frac{1}{2} = \frac{1}{2}, \xi(\neg p) = 1 - \tau(\neg p \to \overline{0}) = 1 - \tau(p) = 1 - \frac{1}{2} = \frac{1}{2},$

(ii) Since $\tau((p \to q) \to q) = \frac{3}{4}, \tau((p \to q) \to (q \to \overline{0})) = \frac{1}{2}$, hence it follows that

$\xi(\{p \to q, q \to \overline{0}\}) = 1 - \tau((p \to q) \to q) = \frac{1}{4}$, and $\xi(\{p \to q, q\}) = 1 - \tau((p \to q) \to (q \to \overline{0})) = \frac{1}{2}$. The consistency degree of $\Gamma = \{p, q\}$ in L is

$\xi(\Gamma) = 1 - \tau(\Gamma \to \overline{0}) = 1 - \tau(p \to (q \to \overline{0})) = \int_0^1 \int_0^1 (x \to (y \to 0)) dxdy = 1 - \frac{3}{4} = \frac{1}{4}$, where $\to$ is

R_0 implication.

Based on theorem 2, we can propose a theory of consistency degrees of finite theories in L^* by truth degrees of formulas.

Definition 7 Suppose that $\Gamma = \{A_1, \cdots, A_n\}$ be a finite theory in L^* , let

$\Gamma \vdash \overline{0} = A_1^2 \to (A_2^2 \to \cdots (A_n^2 \to \overline{0}) \cdots))$, and define

$\xi(\Gamma) = 1 - \tau(\Gamma \vdash \overline{0}) = 1 - \tau(A_1^2 \to (A_2^2 \to \cdots (A_n^2 \to \overline{0})\cdots))$, then $\xi(\Gamma)$ is called the **consistency degree of** Γ.

Example 6 The consistency degree of $\Gamma = \{p, q\}$ in L^* is $1 - \tau(p^2 - (q^2 \to \overline{0}))$, Since $\tau(p^2 \to (q^2 \to \overline{0})) = \int_0^1 \int_0^1 (x^2 \to (y^2 \to 0))dxdy = \frac{3}{4}$, hence $\xi(\Gamma) = \frac{1}{4}$.

5. The simple application of graded fuzzy logic

We can use the ideas of graded fuzzy logic to discuss the simple application of the triple I method in graded fuzzy logic. Suppose that A, B, A^* are different propositions of $F(S)$, the following deduction rule is called generalized modus ponens (briefly, GMP),(Ref P.H´ajek(1998)):

$$\frac{\begin{array}{c} A \quad \to \quad B \\ A^* \end{array}}{B^*} \tag{1}$$

Zadeh first investigated GMP-like problem where A, A^* and B, B^* were supposed to be fuzzy subsets $A(x), A^*(x)$ and $B(y), B^*(y)$ on universes X and Y respectively, and the Compositional Rule of Inference (briefly, CRI) was proposed and the conclusion B^* can be calculated by the following formula(Ref G.J.Wang. (2003,2005))

$$B^*(y) = \sup\{A^*(x) \wedge R_Z(A(x), B(y)) \big| x \in X\}, y \in Y \tag{2}$$

where R_Z is Zadeh's implication operator such that

$$R_Z(a,b) = (1-a) \vee (a \wedge b), a, b \in [0,1] \tag{3}$$

Professor Wang proposed Triple I method and it's generated form $\alpha -$ triple I method to improve the CRI method , where conclusion B^* is the smallest fuzzy subset of Y satisfying the condition that

$$(A(x) \to B(y)) \to (A^*(x) \to B^*(y)) = 1, x \in X, y \in Y \tag{4}$$

where $\to$ is any implication operator having an adjoint t-norm (i.e., $a \otimes b \le c$ if and only if $a \le b \to c$), and B^* can then be computed by the formula :

$$B^*(y) = \sup\{A^*(x) \otimes (A(x) \to B(y)) \big| x \in X\}, y \in Y \tag{5}$$

and for considering α restriction

$$(A(x) \to B(y)) \to (A^*(x) \to B^*(y)) \ge \alpha, x \in X, y \in Y \tag{6}$$

then B^* can then be computed by the formula

$$B^*(y) = \sup\{A^*(x) \otimes (A(x) \to B(y)) \big| x \in X\} \wedge \alpha, y \in Y \tag{7}$$

(4) and (6) also mean the truth of formula $(A(x) \to B(y)) \to (A^*(x) \to B^*(y))$ equals 1 and over than α, (5)and (7) are R_0-triple I and $R_0 - \alpha$- triple I algorithm respectively.

Example 5 Let $X = Y = [0, 1], A, A^* \in F(X) B, B^* \in F(Y)$, $A(x) = \frac{x+1}{4}, B(y) = 1 - y$,

$A^*(x) = 1 - x, \alpha = \frac{5}{9}$, let $\to$ be R_0 implication, using Eq.(5), compute $B^*(y) = \frac{3}{4}$, and using Eq.(7), compute $B^*(y) = \begin{cases} \frac{5}{9}, y \le \frac{3}{4} \\ \frac{3}{4}, y > \frac{3}{4} \end{cases}$.

Conversely, for it's dual forms FMT,that is

$$\frac{\begin{array}{c} A \quad \to \quad B \\ B^* \end{array}}{A^*} \tag{8}$$

A^* can then be computed by the formula

$$A^*(y) = \inf\{B^*(x) \vee (A(x) \to B(y)) \big| x \in X\}, y \in Y, \tag{9}$$

and for considering α restriction , then A^* can then be computed by the formula

$$A^*(y) = \inf\{B^*(x) \vee (A(x) \to B(y)) \big| x \in X\} \vee \alpha', y \in Y, \tag{10}$$

$A \to B$ (9) and (10) are R_0 -triple I and $R_0 - \alpha$ - triple I MT algorithm respectively.

Example 6 Let $X = Y = [0, 1], A, A^* \in F(X) B, B^* \in F(Y), A(x) = \frac{x+1}{4}, B(y) = 1 - y$

$B^*(y) = \begin{cases} 1, y \le \frac{3}{4} \\ \frac{3}{4}, y > \frac{3}{4} \end{cases}, \alpha = \frac{5}{9}$, let $\to$be R_0 implication, using Eq.(9), compute $A^*(x) = \inf \varnothing = 1$, and using Eq.(10), compute

$A^*(x) = \inf \varnothing \vee \alpha' = 1 \vee \frac{4}{9} = 1 \cdot$

Remark From example 5,we can see if, $A(x) = \frac{x+1}{4}, B(y) = 1 - y$ and $A^*(x) = 1 - x$, the minimum $B^*(y)$ is given by example 5 such that Eq.(5) holds, moreover, form example 6, if A, B, B^* do not vary, let $A^* = 1$,then by Eq.(5), B^* is same as before. It means by Eq.(5),correspondence $(A, B, A^*) \to B^*$ is more to one.

References

A.G. Hamilton. (1978). Logic for Mathematicians. London: Cambridge University Press, 1978.

D.W.Pei and G.J.Wang, (2002). The completeness of the formal system L_, *Science in China (E) (in Chinese)*, 32(1) (2002): 56-64.

G.J.Wang, and L.Fu. (2005). unified form of triple I method, *Computers and Mathematics with Applications*, 49(2005): 923-932.

G.J.Wang, L.Fu and J.S.Song. (2002). Theory of truth degrees of propositions in two valued logic, *Science in China (A)*, 45(9) (2002):1106-1116.

G.J.Wang, Yee Leung. (2003). Integrated semantics and logic metric spaces. *Fuzzy Sets and Systems*, 36 (1)(2003): 71-91.

G.J.Wang. (2003). Introduction of mathematics Logics and resolution principle (in Chinese). Beijing: Science in China Press, 2003.

G.J.Wang. (2003). Non-classical Logics and Approximate Reasoning (in Chinese). Beijing: Science in China Press, 2003

G.J.Wang. (1999). On the logic foundation of fuzzy reasoning. *Information Sciences,* 117(1)(1999): 47-88.

G.J.Wang. (1998). Theory of tautologies in revised Kleene systems. *Science in China,* 41(1998): 188-195.

P.H´ajek, (1998). Metamathematics of Fuzzy Logic, Dordrecht: Kluwer Academic Press. 1998.

S.M.Wang, B.S.Wang and G.J.Wang, (2003). A triangular-norm-based propositional fuzzy logic, *Fuzzy Sets and Systems,* 136(1)(2003):55-70.

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Study on the UWB Rader Synchronization Technology

Guilin Lu

Guangxi University of Technology, Liuzhou 545006, China

E-mail: lifishspirit@126.com

Shaohong Wan

Ari Force No.95275, Liuzhou 545005, China

E-mail: wshhg@163.com

Kaiqing Xu

Ari Force No.95275, Liuzhou 545005, China

Abstract

In the article, the characteristics, synchronization method, approaches and process were described in detail, and the hardware platform of the system was designed. Furthermore, the article further pointed out that the research of the synchronization technology should be considered with the algorithm of channel estimation together, and the method of orthogonal wavelet could effectively solve the problem of channel estimation error.

Keywords: Synchronization, Channel estimation, Orthogonal wavelet

1. Introduction

The synchronization in the UWB radar system means making the code signals arriving at the receiver and the local reference signals consistent for the code design position and code clock rate on time, and it includes the capturing process and the following process. Though the uncertainties exist in the code phase and carrier frequency, but the receiver should decode signals normally, i.e. the resolving power of the carrier center frequency must make the dissociated signals in the region of correlative filter frequency, and make the local carrier frequency always aim at the carrier frequency of input signals to ensure the synchronization of code clock frequency. There are many methods such as the code synchronization, frequency synchronization, pulse synchronization and frame synchronization, and the capturing and following need not only the synchronization algorithms with good robustness, but the combination of the MB (multi-band) project which uses the frequency hopping and orthogonal frequency division multiplexing (OFDM) on the chip logic structure with the DS-UWB project which stores original UWB pulse attributes.

2. Brief introduction of UWB synchronization technology

The ultra wideband signals are transmitted to the receiver after modulating, and under the diffusion time delay extension and extent decline, these signals have been seriously distorted, and the pulse width of UWB carrier free is very narrow, and its performance is very sensitive with the time jitter. The demodulator is different with the receiver with narrowband wave, and it must use the channel estimation or the UWB receiver including the self-correlative receiver with equal performance when receiving or detecting, so the key technologies in the UWB radar technology are the synchronization technology and the channel estimation technology.

The synchronization processing includes following three steps (B. Pemson, 2001).

Step 1. Coarse synchronizing. The code phase receiving the pseudo-random code must be adjusted to the local pseudo-random code sequence.

Step 2. Code following and fine synchronizing. The phase separation of the pseudo-random code and the local pseudo-random code should be reduced to zero.

Step 3. Carrier return and carrier following. The picking and following of carrier phase can be implemented with code capturing and code following at the same time.

The hardware platform of synchronization technology in the UWB radar system is seen in Figure 1.

The fist task that the synchronization technology solves is to exactly know the frequency uncertainty induced by the diffusion distance and transmitting time, but the synchronization technology in the traditional IR-UWB interference system, the spread frequency synchronization technology in the DS-UWB and the synchronization technology in

TR-UWB all have large loss of receiving signal to noise ratio, which would influence the precision, and because of the limitation of the hardware resources, the existing algorithm designs all need large amounts of data which are acquired by the statistical method, and the constringency speed is slow. Therefore, considering the influences of the synchronization deviation parameters on the synchronization performance, we should first confirm the coarse time delay, frequency deviation synchronization and channel estimation under the special realization demand of random channel. Large numbers of literatures adopted the iterative ML algorithm to optimize the maximum likelihood estimation algorithm, and up to now, the synchronization technology theory based on the wavelet theory has been applied broad.

2.1 Research of capturing process algorithm

For any measurement based on the receiving code byte sequence, if the result could show obvious difference in the synchronization state and the non-synchronization state, we can use this measurement to identify the synchronization and the non-synchronization. According to the characteristics of several modules in the synchronization technology, consider the influences of timing delay deviation parameters on the performances of synchronization, obtain the pulse position by the dichotomy searching, and capture the signal frame by the baker slide circuit.

Because of the carrier free characteristic of UWB, in the capturing aiming at the spread frequency, when N frequencies, $f_1, f_2, \ldots f_N$, enter into the capturing circuit based on the matching filter, the outputs from various branches are adjusted on time after different time delays, and enter into the comparator, and the input of the comparator is the frequency amount from the channel summitor, and if the correlative peak output producing the sum exceeds the adaptive threshold, so the system will capture the frequency hopping design, and obtain the initial synchronization signals, or else, the system should continue to capture the signals.

The principle figure of capturing circuit is seen in Figure 2.

2.2 Characteristics of synchronization modules

The modules in the synchronization process include frame synchronization module, symbol synchronization module, frequency synchronization module and pulse synchronization module, and various modules supplement each other.

The mathematical model of synchronization clock is based on the time processing process observed by the clock output port, and the clock equipment is composed by the oscillator and the counter.

The input UWB pulse waveform is

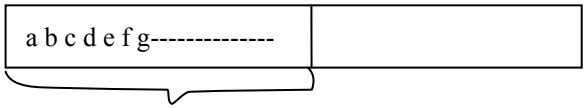
$$s_i(t) = [1 - 4\pi[\frac{t}{t_w}]^2]\exp(-2\pi[\frac{t}{t_w}]^2) \tag{1}$$

When the wave passes the synchronization module, the initial synchronization index signals can be denoted by

$$s_i(t) = A_i c_i(t - iT_i)\sin[w_i(t - iT_i) + \theta_i] + n_i(t - iT_i). \tag{2}$$

Where, $n_i(t)$ is the Gauss white noise, $\theta_i(t)$ is the carrier initial phase of the i'th frequency hopping signal and the phase between the synchronization module and the receiving signal, $w_i(t)$ is the intermediate frequency after frequency conversion, and $c_i(t)$ is the M sequence signal.

The format of one frame in the frame synchronization is



Frame header

The frame header of every frame is the pseudo-random sequence, and if the correlative function value between the transmitting sequence and the local sequence is detected, whether both phases are aligned can be judged, i.e. whether the capturing is successful can be judged, so the correct position of the frame header can be computed to directly demodulate the timing estimation and the frequency deviation estimation acquired before.

The synchronization characteristic of the code word and the code element is that if the spectrum character of the integral and differential signal and the spectrum character of the lead byte are consistent, and the power exceeds certain threshold, the synchronizations of both the code word and the code elements can be judged at the same time, and the information such as the frequency, phase and timing deviation needed by the digital transmission synchronization can be solved by taking the byte as the lead byte, which can ensure the coherence of the code phase and carrier frequency for receiver and transmitter.

We introduce the OFDM system into the UWB system, and the pilot frequency symbol can insert some known symbols and sequences on the frequency axis and the time axis simultaneously, and pick up the pilot frequency signal estimation channel transmission in the receiving port.

The local sample signals are composed by the multi-path fading, time delay and other estimations, and the delay lock function and the implementation process are to calculate the local signals and receiving signals correlatively. Then the capturing enters into the second state, i.e. the coincidence detection state, and if the correlative function values of the local signal sequence and the transmission sequence are aligned and the threshold test is passed, the system enters into the following state, or else, the system will continue to process the pseudo-random code phase, and repeat the above process until capturing correct code element phase.

2.3 United design theory of synchronization and channel estimation

As the system enters into the following state from the capturing state, it can prevent the false alarms induced by the noise and the cheating signals, but the errors in the frequency deviation timing will influence the frequency synchronization technology, and the improved method is to comprehensively consider the synchronization and the channel estimation, and the channel adopts the cluster structure.

The improved method of the local sample signal is to adopt the channel estimation algorithm with high efficiency and multi-path extent channel parameters and the proper double-orthogonal wavelet mathematical model, which can constitute the local sample signals composed by the channel multi-path fading and time delay, and the slide correlative operation can realize the inner product of the signal and wavelet on the same scale, and avoid getting in the local minimum point process, and simplify the estimator circuit, and make the receiver capture the energy maximum of the signal, and effectively estimate many channel parameters such as the multi-path fading extent and time delay.

Because the false alarm probability and the capturing probability are a pair of ambivalent parameters, the estimation algorithm based on the maximum likelihood rules in the past was adopted to enhance the capturing performance of the receiver and realize the united design of the synchronization of the demand radar system and the channel estimation.

In the beginning of 1990s, the second order statistics was adopted to complete the blind equilibrium and blind identifying, but because the blind equilibrium of the second order matrix is sensitive to the estimation of the channel order number, the order number of the channel can not be exactly estimated in practice, so we introduce the higher-order-statistics (HOS) algorithm to solve this problem.

HOS contains the phase characteristics and the extent characteristics of the system or the signal, so it can be applied broad in the identification of the system, but its intolerant deficiency is the low constringency speed, and to acquire the estimation value, we need huge amounts of data samples, which makes the nonlinearity of the channel become very serious and lose the function.

To solve the problem of the bad channel receiving estimation errors aiming at ill-conditioned channel, the orthogonal wavelet channel estimation method was proposed in many literatures and acquired anticipated effects and objective.

2.4 Application of wavelet theory in the synchronization technology

To enhance the synchronization precision to the level of subnanosecond, UWB system adopts the orthogonal scaling function and the wavelet function waveform, and utilizes the m sequence theory to decompose the wavelet.

Schmidt orthogonal method is the method which solves a standard orthogonal basis from one basis a1, a2 … in the n dimensional Euclidean space.

$$b_{k+1} = a_{k+1} - (a_{k+1}, e_1)e_1 - \cdots\cdots - (a_{k+1}, a_k)e_k \tag{3}$$

For the non-stationary signal such as UWB, the basic wavelet function in the article is required to be oscillated and converged quickly. Based on the character that the wavelet or the wavelet network can approach any function or pulse response, and the orthogonal scaling transform can weaken the noise in the detected data, the whole synchronization circuit can work on Mcps.

The wavelet transform theory can be extensively applied in many domains such as signal processing, quantized filed theory and radar. The wavelet analysis overcomes the deficiencies in the traditional Fourier analysis, and it possesses good local characteristics in the time domain and frequency domain. The wavelet function is defined by the function cluster $\{\psi_{ab}\}$ from the base function $\psi(t)$ through flexing and levelly shifting.

$$\psi_{ab}(t) = \frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a})$$

$$a, b \in R, a \neq 0$$

Because the wavelet function adopts the gradually refined time domain step length for the high frequency, so it can focalize any detail of the analyzed signal, and the important characteristics of the wavelet transform is that the wavelet transform possesses the gradual localization character, i.e. “varifocal”.

The expression of the continuous wavelet is the inner product of the signal and the wavelet function, and the discretized wavelet generally adopts the binary discrete system, which can fulfill the orthogonal requirement, and the discretized

continuous wavelet transform is generally used in the mallat algorithm of the filter group which can realize the reconstruction of the signal and largely reduce the computation. Figure 3 is the one dimensional signal reconstruction figure, and the discretized continuous wavelet transform adopts the mallat algorithm to recur the original waveform.

The multi-scale analysis to the signals in the article is to decompose the treating signals of the UWB radar system to different scales, and the former class signal is called as the smooth signal, and the latter disappearing signal is called as the detail signal, and the relations among signals on different scales are mainly completed by the wavelet transform.

The UWB radar signal input mode in this article is to evenly distribute the needed wavelet functions and scale functions designed by the UWB pulse waveform to multiple users (Zhu Lei Yu, 2004), utilize the fast algorithm of the orthogonal wavelet transform, i.e. the mallat algorithm, to quickly decompose and reconstruct the localization characteristics of the signals, realize the progressively taking process of the one-dimensional discreted wavelet transform of signal, separate the coarse weight and detail weight after discreted wavelet transform, transform the signals to the transmitting port of the UWB, implement wavelet transform to part of signals of the system and eliminate the detail information and transmit, reconstruct the original signals in the receiving port, implement discreted feedback to the difference between the original signals and the signals in the receiving port, and finally make the system achieve synchronization.

3. Conclusions

In this article, we researched the UWB radar synchronization technology in the implementation of all digital spread frequency, put forward the united design and research of the synchronization technology and the channel estimation by summarizing various characteristics of many UWB types such as DS-UWB, gave the hardware platform of the synchronization technology, the format of the frame, and the process of UWB from capturing to following, adopted the orthogonal wavelet channel estimation method to reduce the frequency deviation error. Base on the important character of the wavelet transform, the synchronization process of the UWB radar system is to process the wavelet coefficient decomposed by the threshold form, and reconstruct the signals to make the system to achieve the synchronization.

References

B. Pemson, D. E. DOd & R. J. Bohon. (2001). A segmented matched filter for CDMA code synchronization in systems with Doppler frequency offset. *Proceedings IEEE Glob~*. P_Abm, San Antonio, Texas. November 2001.

Zhu Lei Yu. (2004). Waveform Division Multiple Accesses Based on the Orthogonal Wavelet for Ultra-Wideband Impulse Radio Multiple-Access Communication. *Journal of Beijing University of Posts and Telecommunications.* No.27(4).

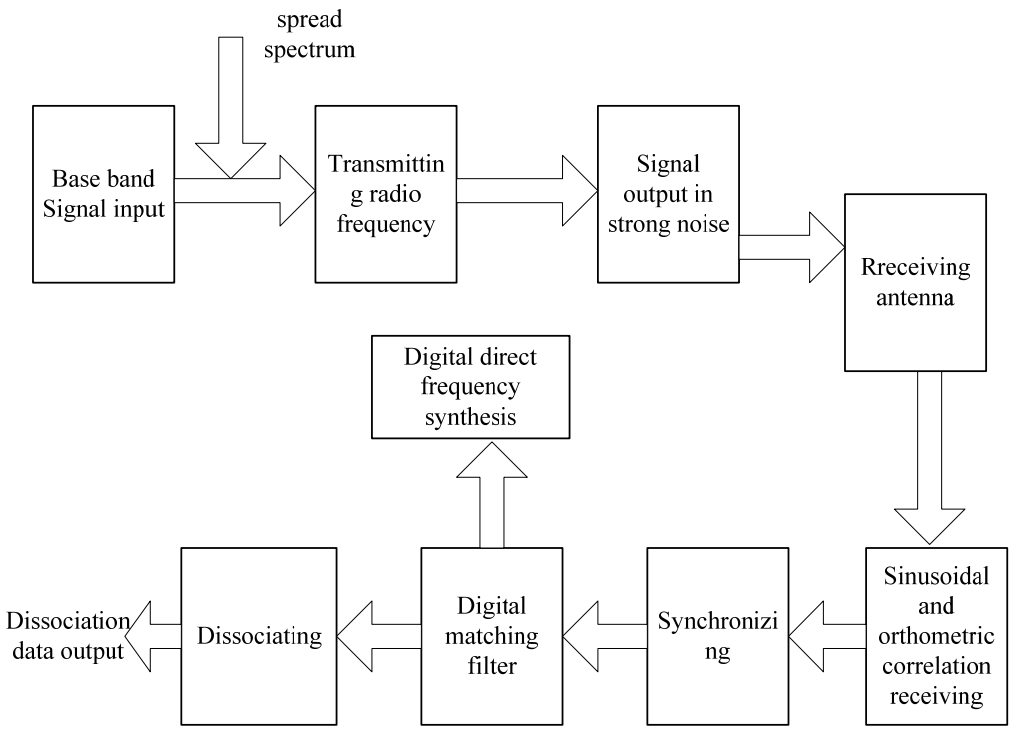


Figure 1. Hardware Platform of Synchronization Technology

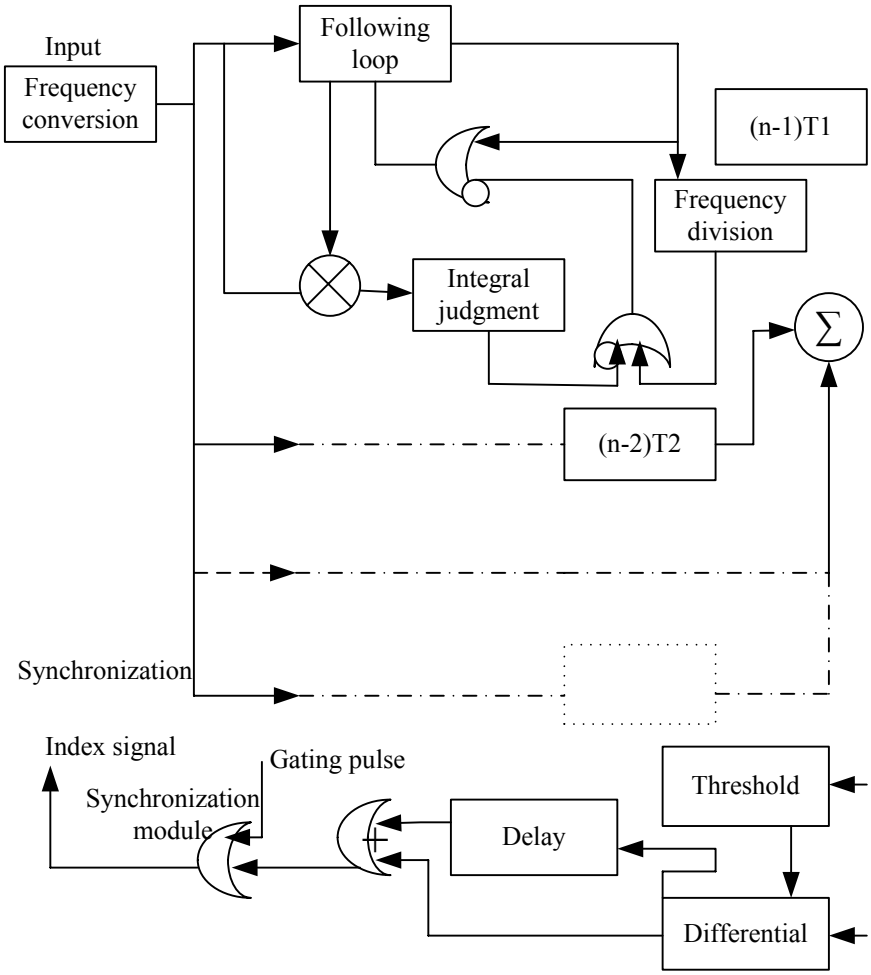


Figure 2. Capturing Circuit

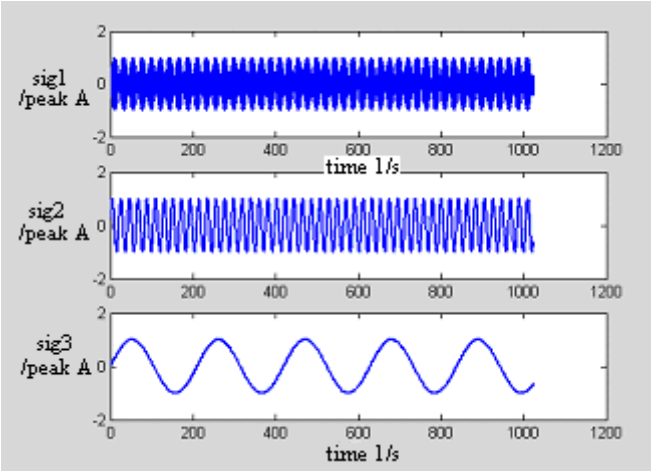


Figure 3. One-dimensional Wavelet Reconstruction

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

A Species-Coexistence Model Defending against Credit Cheating in E-Commerce

Zheng Li

School of Information and Technology, Zhejiang Vocational College of Commerce

470 Bin Wen Street Bin Jiang District, Hangzhou 310053, China

E-mail: nbl_zheng@126.com

Abstract

Based on a simplified evolutionary game model, this article analyzes the internal mechanism of credit cheating behaviors in E-Commerce business. With regard to potentially technological or rule defaults in credit system as well as traders' selfish rationality, presence of credit cheating is to some extent unavoidable. We then suggest a new kind of credit system, which is independently coexistent with traders and trading, should be set up to control credit cheating. Simulation experiments support the ideas aforementioned.

Keywords: E-Commerce, Credit system, Credit cheating, Evolutionary game theory, Model of species coexistence

1. Introduction

E-Commerce was introduced into China in late of the 1990s and has begun to boom recently. However, the problem of trust has been a bottleneck constraining the development of E-Commerce. (Kiku Jones and Lori N. K. Leonard, 2008) Due to the existence of endogenous information asymmetry in E-Commerce transactions, if there is no unimpaired credit system, it is difficult for trust to diffuse quickly among traders. The problem increases transaction costs and even hinders the transaction from taking place, thereby inhibiting the growth of online trade. (Zheng Li, 2008) As a result, many well-known E-Commerce transactions website (such as Taobao, eBay, Yahoo, PaiPai and others) have set up their own credit evaluation system and dealers (about 75%) have begun to concern about the credit evaluation in the trading activities.

However, in spite of the establishment of credit evaluation system, distrust events in various forms still occur at high frequency in China. (WU Liang and WEN Jing, 2007) Because there are rule or technical loopholes in the existing credit system, a new kind of distrust behavior appears--credit cheating which is hidden and infectious. Credit cheating behaviors have a lot of forms. If these behaviors can not timely be prevented and controlled, they would seriously harm the fairness of transactions in E-Commerce market. The academic research on E-Commerce credit cheating has nevertheless not yet kept up with the real situation. Through constructing a species-coexistence model based on evolutionary game theory and carrying out simulation experiments, we analyze the internal mechanism of credit cheating behaviors in E-Commerce business.

2. Analysis on the Internal Mechanism of Credit Cheating: Based on Evolutionary Game Theory

In the 1950s, Nash and other scholars' seminal works set off a wave of game theory research. Game theory is a discipline studying individual decision-making mechanism in an environment people's decisions are influenced by each other, and calculating the equilibrium of decision-making. However, there are many defects in the traditional game theory, such as it is limited in studying the dynamic game. As a result, starting in 1970s, Smith and other scholars developed the traditional game theory and put forward the evolutionary game theory based on the theory of biological evolution. In recent years, Chinese researchers have begun to use the evolutionary game theory to analyze the trust problem in E-Commerce. (LIU Feng-ming and DING Yong-sheng, 2007)

Assuming that an E-Commerce website has set up a credit system which has potential loopholes. By taking advantage of these loopholes, traders can make themselves be untruly evaluated. The website traders can be divided into two categories--ordinary individuals and mutative individuals. Ordinary individuals are traders that choose trust behavior, and we use t to indicate them. Mutative individuals are traders that choose credit cheating behavior, and we use c to indicate them. Let $1-\varepsilon$ and ε are the proportion for t and c, respectively. The reason why we name traders who choose credit cheating acts as mutative individuals is that the credit cheating acts are often derived from some accidental factors in the course of transactions. The traders can only choose trust behavior x or distrust behavior y.

Meanwhile, the game between traders is neither perfectly information symmetric, nor information insulation, but is a game with noisy signals. At the same time, all traders are homogenous, that is which type a trader belongs to can not be judged by his or her action. According to the assumptions aforementioned, the payoff matrix between traders is showed by Figure 1.

Now we suppose: (1) The loopholes in the credit system are potential but not well-known. As a result, the trader can take free ride , that is the trader can choose distrust behavior but not be found easily, so the payoff $R < T$. (2) The traders are rational, so the payoff $P > S$.

We can then get the conclusions as follows:

Conclusion One If there are potential loopholes in the E-Commerce credit system, the traders can take free ride, that is $R < T$, thus it is impossible to realize the equilibrium in which the traders will trust each other.

Proof: The expected utility of ordinary individuals that choose trust behavior x is $\mu_t(x) = R + (S - R)\varepsilon$. The expected utility of mutative individuals that choose trust behavior is $\mu_c(x) = T + (P - T)\varepsilon$. The condition for realizing the pure strategy evolutionary stable equilibrium in which the traders choose trust behavior is:

$$\mu_t(x) - \mu_c(x) = R - T + (S - P - R + T) \cdot \varepsilon > 0 \tag{1}$$

In Equation (1), the necessary and sufficient condition for the state under which traders trust each other to become the equilibrium is $\varepsilon < \varepsilon_1 = \frac{R - T}{P + R - T - S}$. However, since $R < T$, the parameter ε that satisfies $\varepsilon < \varepsilon_1$ and $\varepsilon > 0$ does not exist, thus it is impossible to realize the equilibrium of trust.

Conclusion Two If the E-Commerce traders are self-concern and rational, that is $P > S$, the equilibrium in which traders make credit cheating between each other is very likely to occur.

Proof: The expected utility of ordinary individuals that choose distrust behavior y is $\mu_t(y) = P + (T - P)\varepsilon$ .The expected utility of mutative individuals that choose distrust behavior is $\mu_c(y) = S + (R - S)\varepsilon$ .The condition for realizing the pure strategy evolutionary stable equilibrium of credit cheating is:

$$\mu_t(y) - \mu_c(y) = P - S + (T - P - R + S) \cdot \varepsilon > 0 \tag{2}$$

In Equation (2), the necessary and sufficient condition for the state under which traders cheat each other to become the equilibrium is $\varepsilon < \varepsilon_2 = \frac{P - S}{P + R - T - S}$. Because of $P > S$, parameter ε that satisfies $\varepsilon < \varepsilon_2$ and $\varepsilon > 0$ can be found easily(a ε just less than a certain value), so it is easy to realize the equilibrium of credit cheating.

As a result, as long as there are potential loopholes in the credit system of E-Commerce website and traders are rational, the presence of credit cheating would be inevitable.

3. Defending against Credit Cheating: Based on A Species-coexistence Model

Because the credit evaluation depends excessively on the traders, the traders can cheat to get untrue credit evaluation results by using the loopholes in the credit system. Therefore, we must construct some new kind of credit system. Through comparing the three cases mentioned below, we can make clear what kind of credit system is adequate.

Case One: credit evaluation excessively depends on traders.

In this case, traders can exist independently and depend only to some extent on credit evaluation. On the contrary, credit evaluation is not independent but completely depends on the behaviors of traders. According to the Logistic Law of species evolution, the evolution equation for the quantity of traders is:

$$\dot{x}_1(t) = r_1 x_1 (1 - \frac{x_1}{N_1} + \sigma_1 \frac{x_2}{N_2}) \tag{3}$$

In Equation (3), $x_1(t)$ refers to the quantity of traders at the time point t , r_1 refers to the inherent increasing rate of the quantity, $\dot{x}_1(t)$ refers to the actual increasing rate of the quantity, x_1 and x_2 refer to the values of the quantity and the correct rate of credit evaluation in the last period, respectively. N_1 and N_2 refer to upper limits for the quantity of traders and the correct rate of credit evaluation, σ_1 refers to traders' dependency on the credit evaluation.

The evolution equation for the correct rate of credit evaluation is:

$$\dot{x}_2(t) = r_2 x_2 (-1 - \frac{x_2}{N_2} + \sigma_2 \frac{x_1}{N_1}) \tag{4}$$

In Equation (4), $x_2(t)$ refers to the correct rate of credit evaluation at the time point t, r_2 refers to the inherent increasing rate of correct rate of credit evaluation with the development of technology, $\dot{x}_2(t)$ refers to the actual increasing rate of correct rate of credit evaluation, the meaning of x_1, x_2, N_1 and N_2 are the same as mentioned in Equation (3), σ_2 refers to credit evaluation's dependency on the traders' behaviors.

Let

$$\begin{cases} \dot{x}_1(t) = r_1 x_1(1 - \frac{x_1}{N_1} + \sigma_1 \frac{x_2}{N_2}) = 0 \\ \dot{x}_2(t) = r_2 x_2(-1 - \frac{x_2}{N_2} + \sigma_2 \frac{x_1}{N_1}) = 0 \end{cases},$$

there are three balance points in the above equation group, that is $P_1(N_1,0), P_2(\frac{N_1(1-\sigma_1)}{1-\sigma_1\sigma_2}, \frac{N_2(\sigma_2-1)}{1-\sigma_1\sigma_2}), P_3(0,0)$. According to the stability theory of differential equation [6], the stability conditions of the three solutions are:

$$\begin{cases} P_1(N_1,0): \sigma_2 < 1, \sigma_1\sigma_2 < 1 \\ P_2(\frac{N_1(1-\sigma_1)}{1-\sigma_1\sigma_2}, \frac{N_2(\sigma_2-1)}{1-\sigma_1\sigma_2}): \sigma_1 < 1, \sigma_2 > 1, \sigma_1\sigma_2 < 1 \\ P_3(0,0): not\ stable \end{cases}$$

That is, when credit evaluation excessively depends on traders, there are two evolutionary stable equilibria: (1) when $\sigma_2 < 1, \sigma_1\sigma_2 < 1$, i.e. the credit evaluation's dependency on traders is less than 1 and the product of the two's interdependent degrees is also less than 1, the quantity of traders reaches to the maximal value N_1 and credit evaluation fails absolutely(the correct rate of credit evaluation decreases to zero), resulting in the prevalence of credit cheating; (2) when $\sigma_1 < 1, \sigma_2 > 1, \sigma_1\sigma_2 < 1$, i.e. the traders' dependency on credit evaluation is less than 1 and the credit evaluation's dependency on traders is more than 1, the quantity of traders and the correct rate of credit evaluation are $\frac{N_1(1-\sigma_1)}{1-\sigma_1\sigma_2}$ and $\frac{N_2(\sigma_2-1)}{1-\sigma_1\sigma_2}$, respectively.

Case Two: binding trade behavior with credit evaluation.

In this case, neither the traders or credit evaluation can exist independently. For example, if traders do not take part in credit evaluating or their credit evaluated result is less than a certain value, their qualification for taking part in transaction will be abolished. At the same time, the result of credit evaluation wholly depends on traders' behaviors. Therefore, the evolution equation for the quantity of traders is:

$$\dot{x}_1(t) = r_1 x_1(-1 - \frac{x_1}{N_1} + \sigma_1 \frac{x_2}{N_2}) \tag{5}$$

The evolution equation for the correct rate of credit evaluation is:

$$\dot{x}_2(t) = r_2 x_2(-1 - \frac{x_2}{N_2} + \sigma_2 \frac{x_1}{N_1}) \tag{6}$$

Let

$$\begin{cases} \dot{x}_1(t) = r_1 x_1(-1 - \frac{x_1}{N_1} + \sigma_1 \frac{x_2}{N_2}) = 0 \\ \dot{x}_2(t) = r_2 x_2(-1 - \frac{x_2}{N_2} + \sigma_2 \frac{x_1}{N_1}) = 0 \end{cases},$$

there are two balance points, that is $P_1(\frac{-N_1(1+\sigma_1)}{1-\sigma_1\sigma_2}, \frac{-N_2(1+\sigma_2)}{1-\sigma_1\sigma_2}), P_2(0,0)$. According to the stability theory of differential equation, the stability conditions of the two solutions are:

$$\begin{cases} P_1(\frac{-N_1(1+\sigma_1)}{1-\sigma_1\sigma_2}, \frac{-N_2(1+\sigma_2)}{1-\sigma_1\sigma_2}) : not\ stable \\ P_2(0,0) : always\ stable \end{cases}$$

We can then get a conclusion: when trade behavior and credit evaluation are wholly bound with each other, there is only one evolutionary stable equilibrium in which traders' quantity decreases to zero and credit cheating is prevalent. As a result, it is hard to maintain E-Commerce transaction in Case Two.

Case Three: credit evaluation independently coexists with trade behavior.

In this case, traders and credit evaluation can exist independently and promote each other. For example, the transaction qualification required for the traders depends only to a limited extent on the credit evaluation of their behaviors, and the result of credit evaluation also depends only partially on traders' behaviors. Therefore, the evolution equations for the quantity of traders and the correct rate of credit evaluation are:

$$\dot{x}_1(t) = r_1 x_1 (1 - \frac{x_1}{N_1} + \sigma_1 \frac{x_2}{N_2}) \tag{7}$$

$$\dot{x}_2(t) = r_2 x_2 (1 - \frac{x_2}{N_2} + \sigma_2 \frac{x_1}{N_1}) \tag{8}$$

Let

$$\begin{cases} \dot{x}_1(t) = r_1 x_1 (1 - \frac{x_1}{N_1} + \sigma_1 \frac{x_2}{N_2}) = 0 \\ \dot{x}_2(t) = r_2 x_2 (1 - \frac{x_2}{N_2} + \sigma_2 \frac{x_1}{N_1}) = 0 \end{cases},$$

there are four balance points, that is $P_1(N_1,0), P_2(0,N_2), P_3(\frac{N_1(1+\sigma_1)}{1-\sigma_1\sigma_2}, \frac{N_2(1+\sigma_2)}{1-\sigma_1\sigma_2}), P_4(0,0)$. According to the stability theory of differential equation, the stability conditions of the four solutions are:

$$\begin{cases} P_1(N_1,0) : not\ stable \\ P_2(0,N_2) : not\ stable \\ P_3(\frac{N_1(1+\sigma_1)}{1-\sigma_1\sigma_2}, \frac{N_2(1+\sigma_2)}{1-\sigma_1\sigma_2}) : \sigma_1\sigma_2 < 1 \\ P_4(0,0) : not\ stable \end{cases}$$

In result, when trade behavior and credit evaluation are independently coexistent, there is only one evolutionary stable equilibrium, i.e. $P_3(\frac{N_1(1+\sigma_1)}{1-\sigma_1\sigma_2}, \frac{N_2(1+\sigma_2)}{1-\sigma_1\sigma_2})$, the condition for which is $\sigma_1\sigma_2 < 1$.

Comparing the above three cases, we find that if $\sigma_1\sigma_2 < 1$, a credit system that is independently coexistent with traders is most favorable for the development of E-Commerce business. As both of them can promote each other and can also stand alone, the quantity of traders and credit evaluation system can achieve the optimal balance of coexistence. At the time, both the quantity of traders and the correct rate of credit evaluation will reach the optimal levels ($\frac{N_1(1+\sigma_1)}{1-\sigma_1\sigma_2} > \frac{N_1(1-\sigma_1)}{1-\sigma_1\sigma_2} > 0$, $\frac{N_1(1+\sigma_1)}{1-\sigma_1\sigma_2} > N_1 > 0$, $\frac{N_2(1+\sigma_2)}{1-\sigma_1\sigma_2} > \frac{N_2(\sigma_2-1)}{1-\sigma_1\sigma_2} > 0$). This can not only help the size of E-Commerce transactions expand constantly, but also effectively prevent and curb credit cheating.

4. Data simulation and analysis

The paper proves that a credit system independently coexistent with traders can effectively defend against credit cheating and promote the development of E-Commerce. Next we use Matlab7.1 to do a series of data simulation experiments to test the above conclusions.

Experiment One: Let $N_1 = 10000$ $N_2 = 0.2$ $\sigma_1 \in (0,1)$ $\sigma_2 \in (0,1)$. The values of σ_1 and σ_2 must meet the conditions of $\sigma_1\sigma_2 < 1$ and $\sigma_2 < 1$. Calculate the three cases in Section 3 for one hundred times repeatedly by stochastically choosing the values of σ_1 and σ_2. Then we get the evolutionary equilibria for the quantity of traders and the correct rate of credit evaluation. Figure 2 and figure 3 show the results.

In Figure 2 and Figure 3, the curve IC means the credit system that independently coexists with traders as described in Case Three, the curve OD means the credit system that excessively depends on traders as in Case One, the curve CT means the credit system wholly bound with traders as in Case Two. We find that as long as under certain conditions, that is $\sigma_1\sigma_2 < 1$ and $\sigma_2 < 1$, a credit system independently coexisting with traders is always superior to the other two kinds of credit systems: both the quantity of traders and the correct rate of credit evaluation are at highest levels. This result is consistent with the theoretical analysis in Section 3.

Experiment Two: Let $N_1 = 10000$ $N_2 = 0.2$ $\sigma_1 \in (0, 0.5)$ $\sigma_2 \in (1, 2)$. The values of σ_1 and σ_2 must meet the conditions of $\sigma_1\sigma_2 < 1$ and $\sigma_2 < 1$. Calculate the three cases in Section 3 for one hundred times repeatedly by stochastically choosing the values of σ_1 and σ_2. Then we get the evolutionary equilibria for the quantity of traders and the correct rate of credit evaluation. Figure 4 and figure 5 show the results.

As Figure 4 and Figure 5 indicate, the result from Experiment One has no change when the dependency of credit evaluation on the traders' behaviors be beyond one, which means the result that a credit system independently coexisting with traders is always superior to the credit systems either excessively depending on traders or wholly bound with traders does not depend on the parameter σ_2. In sum, the results of the data simulation experiments aforementioned support the theoretical analysis about the adequate relationship between the credit system and the traders.

5. Conclusions

Based on the evolutionary game theory, the paper divides E-Commerce traders into ordinary individuals and mutative individuals. By using a simplified evolution game model, we analyze the internal mechanism of credit cheating in E-Commerce business. We then construct a species-coexistent model, identifying what kind of relation credit evaluation should maintain with traders in order to effectively defend against credit cheating. The results of the model are supported by the simulation experiments. Our finding can be concluded as follow: although the emergence of credit cheating is unavoidable, credit cheating can be defended effectively by constructing a new kind of credit system independently coexistent with the traders. In the further research, we will concentrate on the topics like constructing the warning model for credit cheating in E-Commerce, and how to improve credit evaluation systems with various forms of credit cheating behaviors.

References

Alchian. A. (1995). Uncertainty, evolution, and the economic theory, *The Journal of Political Economy*, 58(3), 211-221.

Elinor Ostrom. (2000). Collective Action and the Evolution of Social Norms, *Journal of Economics,* 14(3), 145.

Jiang Qi-yuan. (1993). Mathematical model(the second version), *Higher Education Press*, 201.

Jin Xuejun, Mao Jie and Yuan Jia. (2004). Analysis on the Efficiency of Commercial Banks'Customer Service System, *Nanjing Business Review*, 3, 98-121.

Kiku Jones, and Lori N. K. Leonard. (2008). Trust in consumer-to-consumer electronic commerce, *Information and Management*, 45(2), 88-95.

LIU Feng-ming, DING Yong-sheng. (2007), Evolutionary model of trust game based on ecological network in P2P networks, *Computer Engineering and Applications*, 43(23), 24-27.

WU Liang, WEN Jing. (2007). Research on trust problems in C2C E-commerce, *Market Modernization*, 4, 139.

Zheng Li. (2008). Trust model in E-commerce based on ESS equilibrium, *Journal of Computer Applications*, 28(8), 2173-2176.

		(Trader j)	
		Trust behavior x	Distrust behavior y
(Trader i)	x	R , R	S , T
	y	T , S	P , P

Figure 1. Payoff Matrix for Game between Traders in E-Commerce

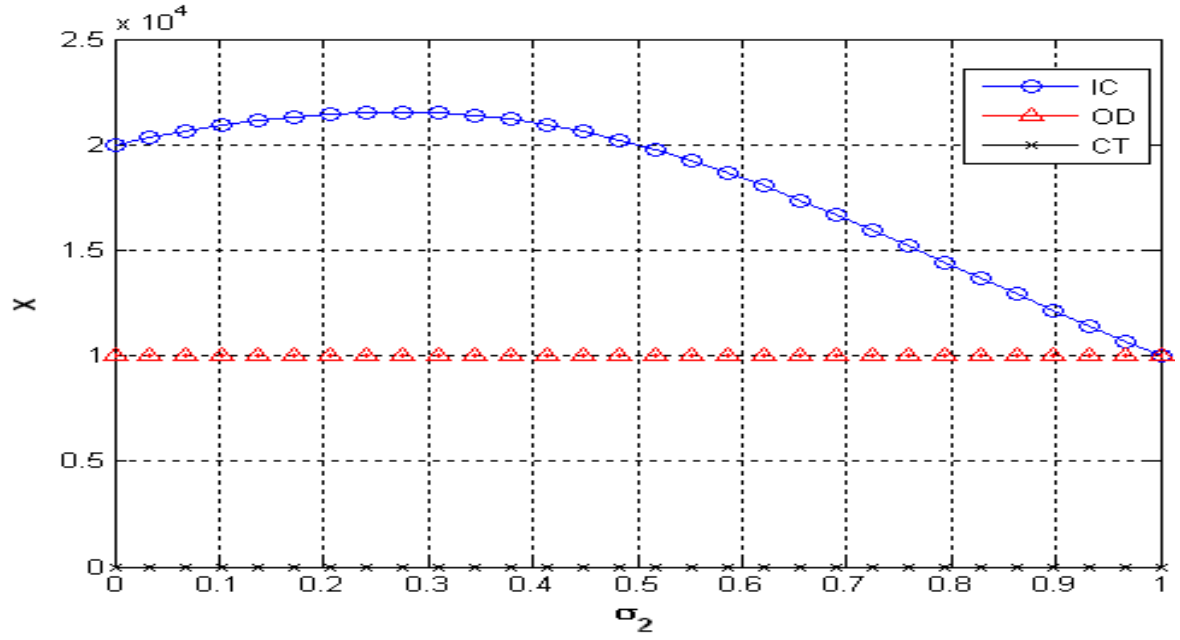


Figure 2. Quantity of Traders in Different Credit System

($\sigma_1\sigma_2 < 1$, $\sigma_2 < 1$)

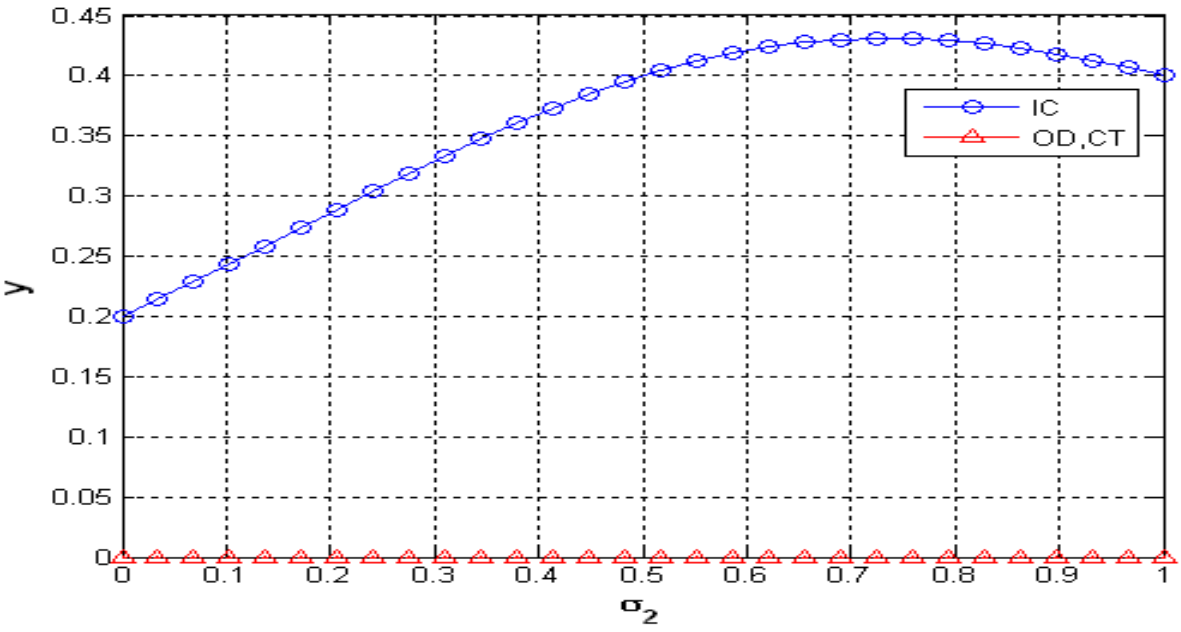


Figure 3. Correct Rate of Credit Evaluation in Different Credit System

($\sigma_1\sigma_2 < 1$, $\sigma_2 < 1$)

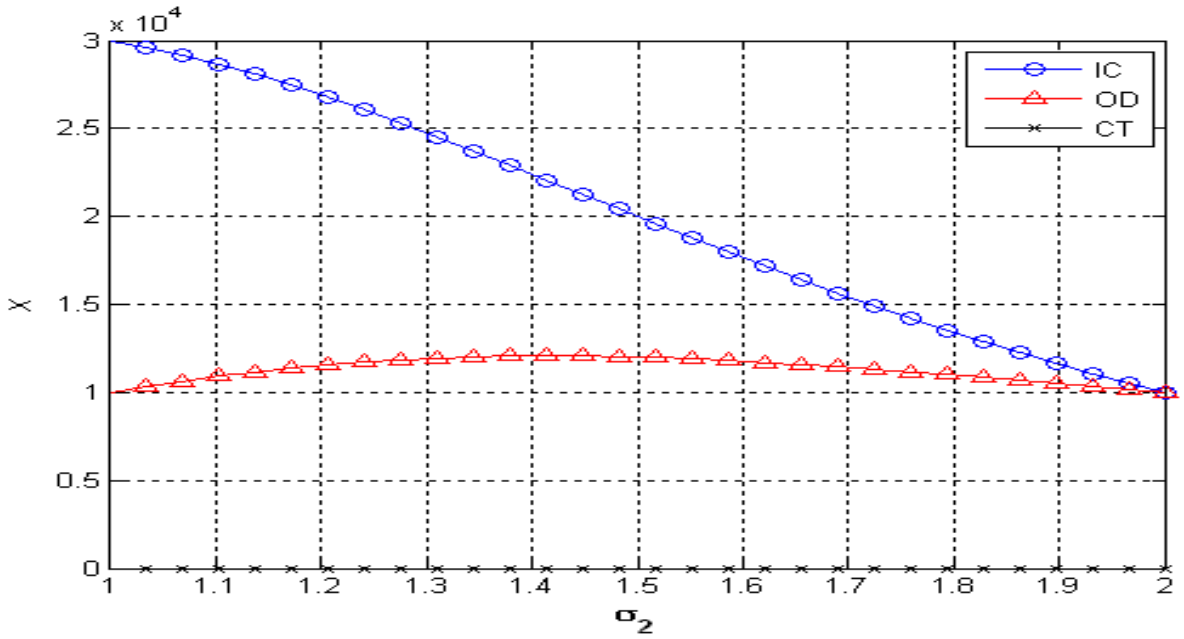


Figure 4. Quantity of Traders in Different Credit System

($\sigma_1\sigma_2 < 1$ and $\sigma_1 < 1$, $\sigma_2 > 1$)

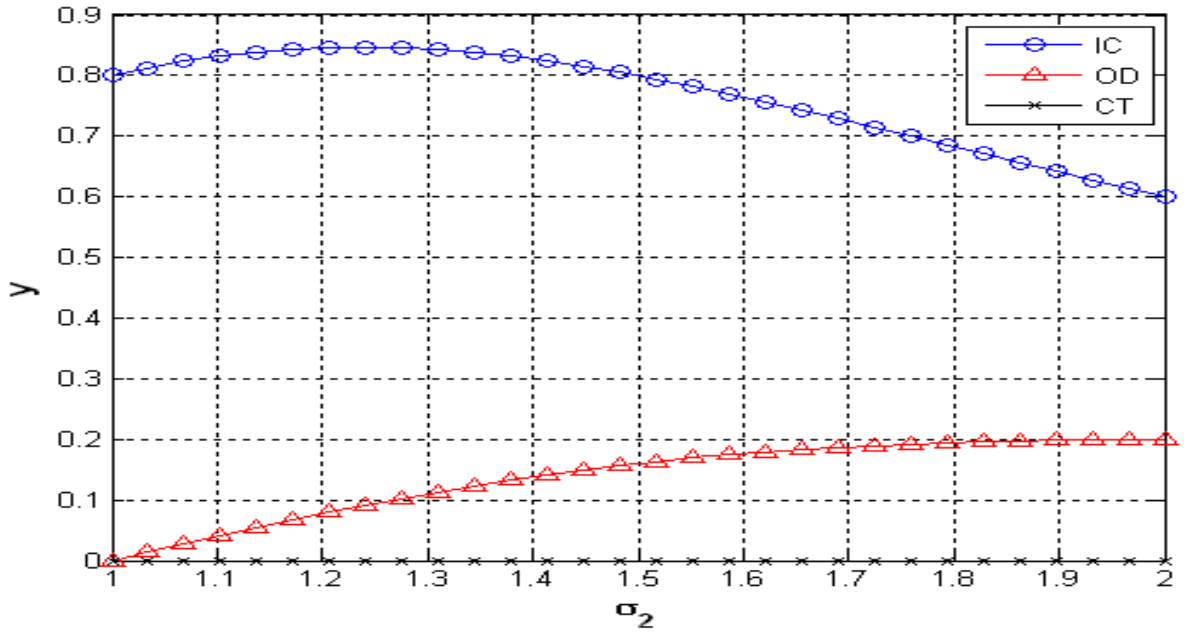


Figure 5. Correct Rate of Credit Evaluation in Different Credit System

($\sigma_1\sigma_2 < 1$ and $\sigma_1 < 1$, $\sigma_2 > 1$)

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Study on the Multiple-Key-Pair Generation Scheme Based on Grey Half-generation

Peng Li, Yulian Shang, Jifeng Wang & Yufei Zhang

College of Information Engineering, Taishan Medical University, Tai'an 271016, China

Abstract

The generation of the multi-key-pair in the RSA iterative encryption system was studied, and from the generation algorithm of key, a sort of multi-key-pair generation scheme was proposed in the article. In the scheme, the encryption key set and the decryption key set used in the encryption system were managed by the improved half-generation algorithm, and the management scheme that one secrete key was used to generate multiple key pairs could solve the difficult management problem for multiple key sets. Finally, the key generation algorithm was simulated by VB6.0 in the article.

Keywords: Iterative encryption, Key Management, Grey half-generation, Key generation

1. Introduction

The key safety is the key factor to ensure the safety of the encryption system, and the key is the core of the whole cipher system, so the safety of the key must be guaranteed. The key generation is the base of the key management, and it is one factor to directly influence the safety of the key, and one important part in the encryption system. Yulian Shang et al studied the RSA encryption system and gave a sort of iterative encryption scheme based on RSA which acquired better encryption effect in their article (Shang?, 2008, P.120-124). However, in that iterative encryption system, large numbers of encryption key sets and encryption key sets would be used, and their article didn't describe how to manage these key sets. In addition, in the RSA key generation, the effective method to generate any big random number has not been found. Based on the grey system theory (J.L. Deng, 1982, P.285-294 & J.L. Deng, 1989, P.1-24), the superset theory and the data generation technology, literatures (J.L. Deng, 1989, P.1-24 & K.Q. Shi, 2000, P.331-340 & T.S. Chen, 2001, P.57-64 & K.Q. Shi, 2000, P.215-224 & T.S. Chen, 2000, P.255-262) offered the concept of grey encryption, and the grey half-generation algorithm was proposed in this encryption system. Based on the grey half-generation algorithm, a sort of new parameter set (big prime number) generation scheme was proposed, which could improve the generation of the key set and enhance the safety of the key. The grey half-generation algorithm was proposed by Professor K.Q. Shi in the grey encryption system theory, and in the article, above algorithm would be improved. In the former grey differential equation, $z^{(1)}$ and $x^{(0)}$ are the neighboring average generation sequence $z^{(1)}(k)=0.5x^{(1)}(k)+0.5x^{(1)}(k-1)$ ($k=2,3,\cdots,n$), and in the article, the general form was adopted, i.e. $z_k^{(1)}=(1-\alpha)x_k^{(1)}+\alpha x_{k-1}^{(1)}$ ($k=2,3,\cdots,n$). And because of $\alpha=\frac{1}{a}-\frac{1}{(e^a-1)}$, $z_k^{(1)}$ changes according to the values of α, and the iterative method is adopted to confirm the grey half-generation set. The research result indicated that the algorithm could enhance the safety of the key.

2. Grey half-generation algorithm (J.L. Deng, 1989, P.1-24 & K.Q. Shi, 2000, P.331-340 & T.S. Chen, 2001, P.57-64 & K.Q. Shi, 2000, P.215-224 & T.S. Chen, 2000, P.255-262)

The correlative concept about the grey half-generation will be defined as follows, and it is the base of the grey encryption theory.

Definition 1: Supposes both sides of the encryption communication select one initial sequence *x*, *x* is the sequence composed by positive integers.

$$x=\{x_1,x_2,\cdots,x_n\} \tag{1}$$

$|x|\geq 4$, $\forall x_i\in N^+, i=1,2,\cdots,n$, $n\geq 4$, so the sequence *x* is called as the key seed sequence, *x* is secret.

Definition 2: Call $x^{(1)} = \{x_1^{(1)}, x_2^{(1)}, \cdots, x_n^{(1)}\}$ is the 1-AGO generation data sequence of *x*, and if

$$x_{(k)}{}^{(1)} = \sum_{j=1, x_j \in x}^{k} x_j; \quad k = 1,2,\cdots n \tag{2}$$

Suppose $x^{(0)}$ is the non-negative sequence, $x^{(1)}$ is the 1-AGO sequence of $x^{(0)}$, $z^{(1)}$ is the neighboring average generation sequence of $x^{(0)}$, $(a,b)^T = (B^T B)^{-1} B^T Y$, the definitions of parameters are seen in reference articles (J.L. Deng, 1989, P.1-24 & K.Q. Shi, 2000, P.331-340 & T.S. Chen, 2001, P.57-64 & K.Q. Shi, 2000, P.215-224 & T.S. Chen, 2000, P.255-262).

Definition 3: The complete solution of the grey differential equation $x^{(0)}(k) + az^{(1)}(k) = b$ is

$$\hat{x}_{k+1}^{(1)} = (x_1^{(1)} - \frac{b}{a})e^{-ak} + \frac{b}{a} \tag{3}$$

Definition 4: The half-solution of the grey differential equation $x^{(0)}(k) + az^{(1)}(k) = b$ is

$$\bar{x}_{k+1}^{(1)} = (x_1^{(1)} - \frac{b}{a})e^{-ak} = \beta e^{-ak}, \quad k = 1,2,\cdots,n \tag{4}$$

Obviously, if $k = 1,2,\cdots,r$, from (4), the data sequence $\bar{X}^{(1)}$ can be obtained.

$$\bar{X}^{(1)} = \{\bar{x}_1^{(1)}, \bar{x}_2^{(1)}, \cdots, \bar{x}_r^{(1)}\} \tag{5}$$

Because above definitions are based on the grey model *GM (1,1)*, so they are respectively called as grey complete solution generation and grey half solution generation.

Definition 5: Suppose x'' is the half-solution key set of the grey encryption system, and

$$x'' = \{x_1'', x_2'', \cdots, x_r''\} \tag{6}$$

Where, $x_i'' = x_i' \bmod p$, p is one big prime number selected randomly, $\forall \bar{x}_i^{(1)} \in \bar{X}^{(1)}$, $x_i' = INT(\bar{x}_i^{(1)})$, $i = 1,2,\cdots r$, and *INT* is the operator of Int.

3. Multi-key-pair generation scheme based on grey half-generation

3.1 Key generation scheme

Iterative encryption algorithm established by Yulian Shang et al (Shang, 2008, P.120-124) based on RSA enhanced the safety of the key, and the key generation algorithm would be studied in the article. When generating the key in the RSA system, the selections about correlative parameters such as the big prime number *p* and *q* are the necessary condition to generate the key set. Based on the grey half-solution generation algorithm, a new parameter set (big prime numbers) generation scheme would be proposed as follows, which could improve the generation of the key set and enhance the safety of the key. The concrete generation algorithm of the parameter set (selecting prime parameters) includes following steps.

Convention: Both sides of the encryption communication are A and B, and A is the sender and B is the receiver. Both A and B should first generate their respective key set before the communication.

3.1.1 Selecting data column

Both A and B respectively select one data sequence *x* as the key seed sequence, i.e. $x = \{x_1, x_2, \cdots, x_n\}$, $|x| \geq 4$, and the sequence is secret. The data sequence *x* can be composed by the birth date, telephone number, or correlative biological characters (such as fingerprint or eye iris) about the identity signs.

3.1.2 Confirming 1-AGO data column

Solve the 1-AGO generation data sequence $x^{(1)}$ of x according to the formula (2), where, $x^{(1)} = \{x_1^{(1)}, x_2^{(1)}, \cdots, x_n^{(1)}\}$ and

$$x^{(1)} = \sum_{j=1, x_j \in x}^{k} x_j, \quad k = 1,2,\cdots n.$$

3.1.3 Confirming grey half-generation set

From above theory introduction, according to the data sequence x, $x^{(1)}$ and definition 1~definition 5, the half-solution set $\overline{x}_{k+1}^{(1)} = (x_1^{(1)} - \frac{b}{a})e^{-ak} = \beta e^{-ak}$ ($k = 1,2,\cdots,n$) of the grey differential equation $x^{(0)}(k) + az^{(1)}(k) = b$ can be confirmed.

The least-square estimation parameter sequence $\hat{a} = (a,b)^T$ of the grey differential equation $x_k + az_k^{(1)} = b$ could fulfill

$$\hat{a} = (B^T B)^{-1} B^T Y$$

$$Y = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}, \quad B = \begin{bmatrix} -z_2^{(1)} & 1 \\ -z_3^{(1)} & 1 \\ \vdots & \vdots \\ -z_n^{(1)} & 1 \end{bmatrix}.$$

Here, based on the modeling mechanism and the application condition of *GM* (1,1), the scheme will improve the above algorithm. In the grey differential equation, $z^{(1)}$ is the neighboring average generation sequence $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1)$ ($k = 2,3,\cdots,n$) of $x^{(0)}$. In the article, the general form $z_k^{(1)} = (1-\alpha)x_k^{(1)} + \alpha x_{k-1}^{(1)}$ ($k = 2,3,\cdots,n$) is adopted. Because $\alpha = \frac{1}{a} - \frac{1}{(e^a - 1)}$, $z_k^{(1)}$ changes with the value of α, and the iterative method is adopted to confirm the grey half-solution set.

(1) From $z_k^{(1)} = (1-\alpha)x_k^{(1)} + \alpha x_{k-1}^{(1)}$, first take $\alpha = 0.5$, and replace it into the formula, i.e. $z_k^{(1)} = 0.5x_k^{(1)} + 0.5x_{k-1}^{(1)}$, and solve $z_k^{(1)}$;

(2) According to the method in formulas (4)~(7), solve the parameter sequence $\hat{a} = (a,b)^T$;

(3) Replace the parameter *a* into the expression $\alpha_{m-1} = \frac{1}{a} - \frac{1}{(e^a - 1)}$ to confirm the new value of α, and note it as α_{m-1};

(4) Replace the value of α into the expression $z_k^{(1)} = (1-\alpha)x_k^{(1)} + \alpha x_{k-1}^{(1)}$ to confirm the value of new $z_k^{(1)}$, and solve the value of the new parameter sequence $\hat{a} = (a,b)^T$. Repeat (1), (2) and (3) in turn.

(5) According to the iterative times set in advance, the solved parameter sequence $\hat{a} = (a,b)^T$ can be used to confirm the grey half-solution set $\overline{x}^{(1)} = \{\overline{x}_1^{(1)}, \overline{x}_2^{(1)}, \cdots, \overline{x}_r^{(1)}\}$.

3.1.4 Confirming parameter set and selecting big prime numbers *p* and *q*

Perform the Int operation to the grey half-solution set $\overline{x}^{(1)} = \{\overline{x}_1^{(1)}, \overline{x}_2^{(1)}, \cdots, \overline{x}_r^{(1)}\}$, and obtain the data sequence x', where, $\forall \overline{x}_i^{(1)} \in \overline{x}^{(1)}, x_i' = INT\left(\overline{x}_i^{(1)}\right)$ ($i = 1,2,\cdots r$), and *INT* is the operator of the Int operation. According to the formula (6), take an enough big prime number *p*, and solve the sequence $x'' = \{x_1'', x_2'', \cdots, x_r''\}$, where, $x_i'' = x_i' \bmod p$, i.e. the parameter set.

A and B respectively select two big prime numbers according to the conditions randomly from the parameter set x'', *p* and *q*, finally, generate the multi-key-set with iterative encryption by the key generation method based on the RSA cipher system.

3.2 Example of key generation

To explain the concrete implementation method and approach of the key pair generation algorithm, the following simple example is offered. The generation algorithm of the big prime number parameter set and the key set generation algorithm have been implemented by Matlab6.0 and VB6.0, and the key generation interface by the VB6.0 is seen in Figure 1.

The concrete implementation process includes following steps.

Step 1. The key seed sequence selected by any side of the encryption communication is *x*, $x = \{8,3,9,4,2,9,9\}$ (secret), and convenient for the computation, randomly select a seven bits data sequence as *x*;

Step 2. Seek the 1-AGO generation data sequence $x^{(1)}$ of the key seed sequence *x*, and obtain $x^{(1)} = \{8,11,20,24,26,35,44\}$;

Step 3. Replace data sequence x and $x^{(1)}$ into the following expressions,

$$Y = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}, \quad B = \begin{bmatrix} -z_2^{(1)} & 1 \\ -z_3^{(1)} & 1 \\ \vdots & \vdots \\ -z_n^{(1)} & 1 \end{bmatrix}$$

Take $\alpha = 0.5$ to seek the parameter sequence, $\hat{a} = (a,b)^T$;

Compute and obtain a= -0.1475, b= 2.5080;

Replace a into the formula $\alpha_1 = \frac{1}{a} - \frac{1}{(e^a - 1)}$,

And obtain $\alpha_1 = 0.5123$ (without loss of generality, the iterative times is two in the example)

Step 4. According to the value of α_1, compute and obtain $a_1 = -0.1482$, and $b_1 = 2.4814$;

Replace values of a_1 and b_1, and $x_1^{(1)} = 8$ into $\bar{x}_{k+1}^{(1)} = (x_1^{(1)} - \frac{b}{a})e^{-ak} = \beta e^{-ak}$, takes values of k are $k = 1,2,\cdots,n$ (n is random) in turn, and perform the Int operation, and obtain the grey half-solution generation set x';

$$x' = \{29,33,39,45,52,60,70,81,94,109,126,147\cdots\}$$

Take any big prime number, *p*=347, and obtain the parameter data sequence $x'' = x' \bmod p$;

$$x'' = \{29,33,39,45,52,60,70,81,94,109,126,147\cdots\}$$

Step 5. In the parameter set x'', select any big prime number, *p*=29, *q*=94 (in actual application, the prime number test is needed);

Obtain one key pair:

Public key e=37, private key d=2041.

End.

4. Conclusions

Based on the grey half-generation algorithm, a sort of multi-key-pair generation scheme based on the RSA iterative encryption system is proposed in the article, and the scheme possesses following characteristics.

(1) The algorithm thought of the scheme can be described as follows. In each key exchange and secret communication, both sides of the encryption communication select one secret key (secret key seed sequence) to randomly generate t encryption key sets and decryption key sets used in the encryption communication. Each encryption can replace key seed sequence and multi-key-pair, and from certain meaning, this scheme is closed to one-time-one-encryption, and its encryption performance and safety are higher than general cipher system. The key is established only when it is needed, and it is not necessary to be stored, or else, which will bring the exposed danger. That is obviously advanced for the key generation mode.

(2) When the algorithm is used to generate the key, one random key seed sequence (private key) needs to be stored or selected temporarily, and the large-sized key set is not necessary to be managed, or the safety of KDC is not necessary to be worried, and this method makes the key more easily to be managed.

(3) The safety of the method that that the grey half-solution generation theory is used to generate the key parameter set is based on "the problem to solve the discrete logarithm", so the generation scheme of the multi-key-pair and the iterative encryption system is safe for the algorithm, and the key is safe.

(4) The algorithm of the scheme is safe, and it is difficult to get all t keys for the attackers of the cipher system, because the key seed sequence is secrete, and even if the grey half-solution algorithm is known, the key can not be obtained. If the grey half-solution set is leaked by certain reason, it is still impossible to deduce key seed sequence from the half-solution key set, because its difficulty equals to "solve the problem of discrete logarithm", and the elements in the key set is infinite ($n \in N$), and p is the enough big prime number which can be selected randomly, and the possibility that the cipher attackers want to find the big prime number used in the generation of the key parameter set. Therefore, it is safe to adopt this algorithm to generate the key, and the safety of the key can be guaranteed.

References

J.L. Deng. (1982). Control Problems of Gray System. *Systems and Control Letters.* No.1(5). P.285-294.

J.L. Deng. (1989). Introduction to Gray System Theory. The Journal of Grey System. No.1(1). P.1-24.

K.Q. Shi & T.S. Chen. (2000). A Grey General Lock and Central Public Cryptosystem (I). *The Journal of Grey System.* No.12(4). P.331-340.

K.Q. Shi & T.S. Chen. (2000). On the Grey Encryption Problems of Information Security (I). *The Journal of Grey System.* No.12(3). P.215-224.

Yulian Shang, Wuyuan Jia & Peng Li etc. (2008). Study on the Scheme for RSA Iterative Encryption System. *Computer and Information Science.* No.1(3). P.120-124.

T.S. Chen & K.Q. Shi. (2001). A Grey General Lock and Central Public Cryptosystem (II). *The Journal of Grey System.* No.13(1). P.57-64.

T.S. Chen & K.Q. Shi. (2000). On the Grey Encryption Problems of Information Security (II). *The Journal of Grey System.* No.12(3). P.255-262.

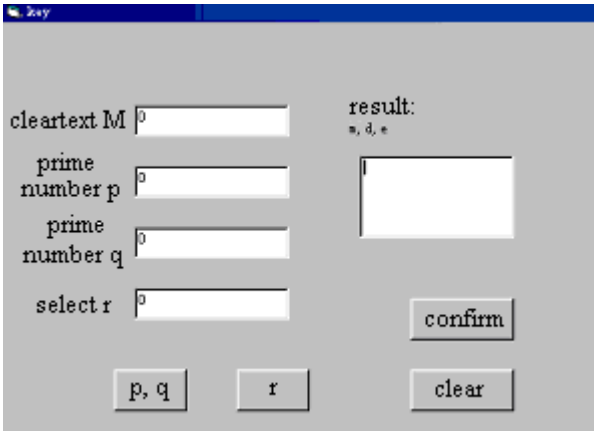


Figure 1. Key Generation and Parameter Selection

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Knowledge Management and Usability Model for Knowledge Management System

Nafaizatulnaniah Hashim

Department of Information & Communications Technology

Centre for Foundation Studies (Nilai Campus)

International Islamic University Malaysia

E-mail: nafaizatul@yahoo.com

Abu Bakar Md Sultan

Department of Information System

Faculty of computer Science and Information Technology

43400 UPM Serdang

E-mail: abakar@fsktm.upm.edu.my

Abstract

Many studies and works have been done to produce a Knowledge Management System (KMS) in which employees of any organization can access the organization's sources of information and solutions. However, there is still no standard knowledge measurement and usability model that can assist KMS user to select or evaluate the appropriate KMS. The aim of this paper is to analyze how the ISO Consolidated Usability Model suggested by Abran, Khelifi, Suryn and Seffah can be used in measuring knowledge and evaluating usability for any Knowledge Management System. The methodology used is a user-satisfaction questionnaire developed based on the ISO Consolidated Usability Model.

Keywords: Knowledge Management, ISO Model, Usability

1. Introduction

Knowledge Management System (KMS) is becoming a trend nowadays enabling employees of any organization to access the organization's sources of information and solutions. For example, using a KMS, a programmer of an IT company could know the existing libraries that he/she could use in his/her current project. Sharing this information organization wide can lead to more effective software design and it could also lead to ideas for new or improved software features.

However, the lack of a usability measurement framework for KMS may drive people away from using it. It also impedes a systematic comparison among KMS providing a similar functionality. Nevertheless, this situation can be overcome if appropriate usability models are in place. These models will assist KMS administrators in determining which provider best fits the organization's needs. Furthermore, it is most important to notice that, usable system is a must to ensure satisfied and returning users.

In determining usability of a software system, one must have specific knowledge about the end-user of the software systems. In ISO 9126-1, the first part of ISO 9126, usability is defined as "set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by stated or implied set of users" (Wikipedia, 2009). Usability here comprises of:

- Learnability – learning effort for different users, i.e. novice, expert, casual etc.
- Understandability – how system functions can be understood, relates to user
- Operability - ability of the software to be easily operated by a given user in a given environment.

The usability of product depends on the nature of the user, the task and the environment. For the end-user, software usability means how fast and efficient he/she can complete the expected task. For managers, software usability is a

criterion that must be considered in selecting a product. For software developers, software usability means issues related to design and documentation.

From KMS perspective and basing on ISO Consolidated Usability Model, usability refers to the capability of the KMS to be effective, efficient, satisfactory, learnable and secured under specified conditions. In particular, the usability model proposed in ISO Consolidated Usability model is enhanced. This model is validated using user satisfaction questionnaire given to Universiti Putra Malaysia KMS' users. We do not claim the model to be exhaustive, rather a starting point where new attributes and measures can be introduced as further experience is developed.

This paper is structured as follows: In Section 2 we give a brief overview of Knowledge measurement and usability, while in Section 3 the some usability models are reviewed. In section 4, we apply the ISO Consolidated Usability Model to KMS. Section 5 show the methodology used while in section 6 the results are discussed. The last section summarizes this paper and proposes future work

2. Knowledge Measurement

What is knowledge? According to (Trans4mind, 2009) , knowledge means:

1) Knowing that (facts and information)

2) Knowing how (the ability to do something)

Thus, in an organization knowledge can refer to the information that an employee may have know and the ability of an employee to perform assigned tasks. It can also mean facts and information that an organization has. In (Lethbridge, 1994), knowledge is defined as "any form of information that one might be able to manipulate in one's brain … involving the categorization, definition and characterization of things and their relationships".

Now, how to measure knowledge? In Computer Science and other fields, knowledge can be measured using metrics. In (Wikipedia(b), 2009) metric is defined as a "standard unit of measure, such as meter or gram, or more generally, part of a system of parameters, or systems of measurement, or a set of ways of quantitatively and periodically measuring, assessing, controlling or selecting a person, process, event, or institution, along with the procedures to carry out measurements and the procedures for the interpretation of the assessment in the light of previous or comparable assessments". There are two types of metrics; open-ended and closed-ended metrics (Lethbridge, 1994). A closed-ended metric refers to measurements that fall within a particular range and it can range from zero to one. An open-ended metric refers to measurements which one of the ends of its range is not absolutely fixed.

In Software Engineering, there are many metrics such as LOC (Lines of Code), Function points and COCOMO (Constructive Costs Modeling). In this paper, we will use metrics that are applicable to usability only. In ISO 9126 (international standard for the evaluation of software), usability is defined as the how easy and effective a software is. Usability here comprise of learnability, understandability and operability.

The next sections will describe in detail on usability and usability metrics.

3. Usability

In (Galin, 2004) usability is defined as "requirements deal with the scope of staff resources needed to train a new employee and to operate the software system". Boehm et. al. (1978) defines software usability as "the extent to which the product is convenient and practical to use". The measures of usability include measures of internal attributes and external attributes (Fenton, et. al., 1998). The internal attributes are:

1) well-structured manuals

2) good use of menus and graphics

3) informative error messages

4) help functions

5) consistent interfaces

External attributes of usability:

1) entry level: in terms of experience with similar classes of applications

2) learnability: speed of learning, hours of training required before independent use

3) handling ability: speed of working when trained, errors made when working at normal speed

The usability of the system plays a big role, not only in customer satisfaction but also in terms of additional functionality and life-cycle costs. There are many works that have been done to produce good usability models. We will further describe these works in section 3 consolidated usability model.

Usability metrics are the measurement used in measuring the usability of software, in other words, quantitative methods of measuring usability. Metrics can help managers in tracking design process or in deciding to purchase products. Examples of metrics; Percentage of Task accomplished, Time to achieve one task, Time to learn, Time spent on errors and Percentage of Task achieved per unit of time.

4. ISO Consolidated Usability Model

From our study, we found out that there is little attention in the usability during the development of any type of software. Software developers mainly stress more on the features of software and getting it done while neglecting the important aspect of a software which is usability. Therefore, in this paper, we aim to analyze how the ISO Consolidated Usability Model suggested in (Abran et. al., 2003) can be used for guiding any organization or institution in developing usability measurement for their KMS.

4.1 Consolidating the ISO Usability Models

According to Abran et. al., 2003, software usability can be characterized based on the target audiences since each audience has a specific expectation of the software usability. For the end user, good usability software will be software that allows him to achieve the expected task more efficiently. For the manager, usability will help him in deciding whether to select a product or not.

The authors used ISO 9241 and ISO 9126 as the basis of their studies. They argued that a clearly defined model must be established so that it can clarify usability definition and objectives of software through specified measurements. The paper argued that ISO 9126 is unclear in detailing the measures for each level and also lack of guidance in assessing results of measurement. The paper also stated that ISO 9241 is still not adequate as it does not tackle the learnability characteristics. Hence, it suggested a consolidated model, which was based from ISO 9241-11 but added with two additional characteristics: learnability and security.

The consolidated model is shown in figure 1 below.

Our paper will be based on this ISO Consolidated Usability Model. In order to ensure a usable KMS one must study design of user interfaces for the system and also the security part. We evaluate this model using Universiti Putra Malaysia KMS, the SSM. The SSM is the KMS system to evaluates the performance of the staffs. This is detailed out in methodology section.

We believe this model is also a good candidate for future work of KMS usability model.

4.2 Applying the ISO Consolidated Usability Model to KMS

In this paper we have selected ISO Consolidated Usability model as the basis for KMS usability model. In Table 1, the definition of the different attributes along with their relevance is shown.

Each of the sub characteristics above will determine whether a given KMS is usable or not. Thus, attributes need to be defined for each sub characteristics so that measurement can be done.

Each of the sub characteristics above will determine whether a given KMS is usable or not. Thus, attributes need to be defined for each sub characteristics so that measurement can be done.

Based from ISO Consolidated Usability Model, effectiveness is measured by:

1) Percentage of tasks accomplished

2) Ratio on failure of handling

3) Percentage of tasks achieved per unit of time

The following table 2 shows what we can measure to determine the effectiveness of a KMS.

Efficiency was measured by:

1) Repetitions' number of failed command

2) Documentation or help's use frequency

3) Errors' percentage

4) Time spent on error

5) Time to achieve one task

6) Number of good and bad characteristics recalled by users

7) Number of available commands not called upon

The following table 3 shows what we can measure to determine the efficiency of a KMS.

Satisfaction was measured by:

1) Percentage of users' favorable and unfavorable comments

2) Number of times that user expresses his frustration

3) Rating scale for users' satisfaction with functions and characteristics

The following table 4 shows what we can measure to determine the user satisfaction of a KMS.

Security was measured by:

1) Access audibility

2) Access controllability

3) Data corruption prevention

4) Data encryption

Security was more on whether participants were asked for password during the first time using the system and also when the system fails. The following table 5 shows what we can measure to determine the security of a KMS.

For learnability sub characteristic, we consider the capability of the software to enable the user to learn how the software achieves its aim. Learnability was measured by time to learn as shown in table 6.

4.3 The enhanced Usability Model

The following figure 2 shows the ISO Consolidated Usability Model together with the attributes that we suggested in previous section.

Based on the measurement given for each sub characteristics, usability of a KMS can be calculated accordingly. However, in this paper, we only validated the ISO Consolidated Usability Model using a simplified questionnaire as stated in the Methodology and Results and Discussion sections.

5. Methodology

In order to evaluate the suggested model in section 4, we developed a questionnaire based on the measurements laid out in the model and also sample taken from Software Usability Measurement Inventory (SUMI).

The usability questionnaire supports user subjective satisfaction with the KMS effectiveness, efficiency, security and learnability as well as with the attitude the system induces in users during its usage. Participants indicate level of their agreement with a questionnaire statement on a three-point Likert scale.

6. Results and Discussions

For this study, we have chosen a group of Universiti Putra Malaysia KMS's users, mainly lecturers from Fakulti Sains Komputer & Teknologi Maklumat, Universiti Putra Malaysia.

These participants have used the KMS at least once. Therefore, we assumed that they are quite familiar with the flow of the KMS.

6.1 Procedures

We use quantitative data in order to meet our objective by giving out structured questionnaires. In preparing the questionnaires, we employed the following steps:

- Categorized the questions by employing the characteristics in the ISO Consolidated Usability Model which are Effectiveness, Efficiency, Satisfaction, Security and Learnability.
- Chose type of data to be collected. Each question should have 3 likert scale or at least a Yes and No answer for a quick and easy data analysis, as well as open question to obtain detailed information.

The questionnaires consist of both closed questions and open-ended questions. For the closed questions, Likert Scale was used in which participant was asked to indicate his or her degree of agreement with the statement. For open-ended questions, participant was asked on the main reason of using the system.

What we have done during the questionnaire sessions are:

- Explain to the participants the purpose of the sessions
- Give participants around 10 to 15 minutes to fill up the questionnaires
- Ask for any additional suggestion from the participants

6.2 Measures

The validity of the model was determined from participants' answers to the questionnaire; all questions used a three-point Likert scale. The five aspects that are taken into considerations: effectiveness, efficiency, satisfaction, security and learnability.

In this study, for effectiveness, we only look at in terms of participants' ability to complete given task. For efficiency we look at participants' satisfaction with the system efficiency in terms of: repeating certain processes, error recovery and also usefulness of online help. For satisfaction, we look at whether the participants agree that system has all the expected functionalities. Security was more on whether participants were asked for password during the first time using the system and also when the system fails. For learnability, we look at the number of participants who agreed that the system is easy to learn.

6.3 Accomplished results

We received 14 respondents for this study. Table 6 shows the results of this study.

Here we can see 71.4% participants aware that they must use the system to fill up their activities for assessment of their work performances.

From table 7, we can see that most participants agree that the system is not effective.

Based on the responds, we can see from table 8 that most participants agree that the system is not efficient as:

- They need to repeat previous tasks if system stops suddenly
- The online information is not very helpful
- It takes a long time for the system to recover from error

From table 9, most of the participants do not satisfy with the system. Based on our short interviews with them, we gather that:

- System requires many steps
- Users need to scan their documents before entering data into the system. If not the system will not allow the users to save the transactions

From table 10 more than 60% respondents agree that the system is not easy to learn.

From table 11, the system is good in terms of security.

Looking at the accomplished results we can see that the Universiti Putra Malaysia KMS still need more enhancements in order to ensure its usability. We believe using this enhanced Usability Model, any KMS usability also can be determined.

7. Conclusions and Future Work

In this paper, we apply the characteristics suggested in ISO Consolidated Usability Model to see how this model can be used in KMS. This model is validated using user satisfaction questionnaire given to Universiti Putra Malaysia KMS' users. From the results, we can see that this model can be used to develop a usability measurement framework for any KMS. We do not claim the model to be exhaustive, rather a starting point where new attributes and measures can be introduced as further experience is developed.

References

Abran, A., Suryn, W., Khelifi, A., Rilling, J., Seffah, A.and Robert, F. (2003). Consolidating the ISO Usability Model. *In submitted to the 11th International Software Quality Conference and the 8th Annual INSPIRE Conference.*

Boehm, B.W., Brown, J.R., Kaspar, J.R., Lipow, M., MacLeod, G.J. and Merit, M.J. (1978). Characteristics of Software Quality. *Amsterdam, North-Holland.*

Fenton, N.E.and Pfleeger, S.L. (1998). Software Metrics: A Rigorous & Pratical Approach. Second Edition. *Course Technology.*

Galin, D. (2004). Software Quality Assurance: From theory to implementation. *Addison Wesley.*

Lethbridge, T.C. (1994). *Practical Techniques for Organizing and Measuring Knowledge*, Ph.D Thesis. University of Ottawa.

Trans4mind. (2009). http://www.trans4mind.com/personal_development/Philos/WhatIsKnowledge.htm

Wikipedia(b) (2009). http://en.wikipedia.org/wiki/Metrics

Wikipedia. (2009). http://en.wikipedia.org/wiki/ISO_9126

Table 1. Attributes Definition

Characteristic	Sub Characteristic	Definition	Relevance
Usability	Effectiveness	How well do the users achieve their goals using the system?	The KMS users must be able to complete their task at a given time using the system
	Efficiency	What resources are consumed in order to achieve their goals?	The KMS must have less errors
	Satisfaction	How do the users feel about their use of the system?	The KMS users must satisfy with the system functionality and capability
	Security	Ability of the system to prevent unauthorized access	The KMS must allow only authorized users
	Learnability	Time taken for the user to learn the system	The KMS must be easy to learn

Table 2. Measurement of Effectiveness KMS

Sub characteristic	Attribute	Measure
Effectiveness	Task	Number of task to perform
		Percentage of tasks accomplished
		Percentage of tasks achieved per unit of time
	Time	Time taken to complete a task
	Failure	Ratio on failure of handling

Table 3. Measurement of Efficiency KMS

Sub characteristic	Attribute	Measure
Efficiency	Failed Command	Number of failed command
	Documentation	Documentation or help use frequency
	Error	Number of error per screen
	Time	Time spent on error
		Time to achieve one task

Table 4. Measurement of Satisfaction

Sub characteristic	Attribute	Measure
Satisfaction	Comment	Number of comments per screen
	Rating	User rating scale

Table 5. Security of KMS

Sub characteristic	Attribute	Measure
Security	Password	The software asks for password

Table 6. Learnability measurement

Sub characteristic	Attribute	Measure
Learnability	Help	The software provide help
	Documentation	The software provide documentation
	Screen	Number of screens to achieve one functionality
	Time	Time to learn one screen

Table 7. The main reasons for using KMS

Purpose	%
Instruction by faculty to fill in activities for assessment of performance	71.4
To update directory or looking for forms	7.1
To store information	14.3
Not sure	7.1

Table 8. Questions on Effectiveness

Questions	Agree	Undecided	Disagree
I am able to complete my task at given time using this software	23.1%	38.5%	38.5%
There is too much steps to take in completing a task	92.3%	0.0%	7.7%
Tasks can be performed in a straightforward manner using this software	15.4%	15.4%	69.2%

Table 9. Questions on Efficiency

Questions	Agree	Undecided	Disagree
If this software stops for any reason, I do not have to repeat the process all over again	23.1%	23.1%	53.8%
If this software stops for any reason, it is easy to restart it	23.1%	23.1%	53.8%
The information (such as online help, on-screen messages, and other documentation) provided with this system is clear	7.7%	15.4%	76.9%
I like to use the online help whenever I am stuck	46.2%	7.7%	46.2%
The software does not take a long time to recover from error.	7.7%	38.5%	53.8%
I rarely face an error when using this software	23.1%	7.7%	69.2%
If the software stops during data entry process, I need to enter a new data again	76.9%	0.0%	23.1%

Table 10. Questions on Satisfaction

Questions	Agree	Undecided	Disagree
Overall, I am satisfied with how easy it is to use this software	15.4%	0.0%	84.6%
This software has all the functions and capabilities I expect it to have	0.0%	30.8%	69.2%

Table 11. Questions on Learnability

Questions	Agree	Undecided	Disagree
It was easy to learn using this software	23.1%	7.7%	69.2%
If there is any data entry error, the alert message able to describe the next action for me	15.4%	23.1%	61.5%

Table 12. Question on Security

Questions	Agree	Undecided	Disagree
The software requires me to enter a password before entering the main screen	100.0%	0.0%	0.0%
If the software suddenly quit, it will ask me to enter password again	76.9%	15.4%	7.7%

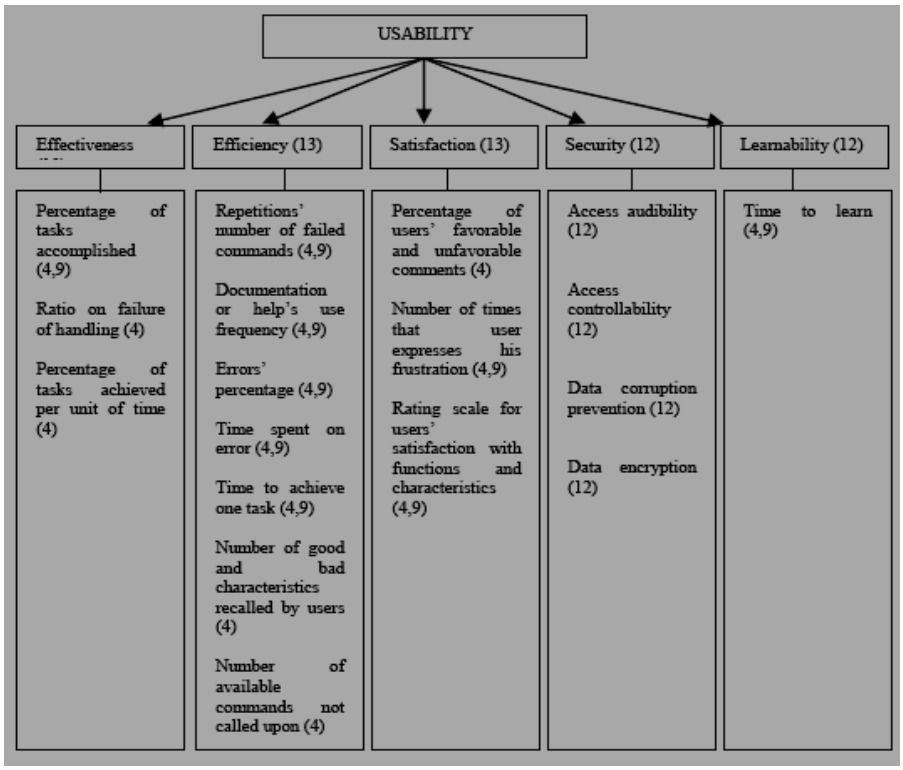


Figure 1. ISO Consolidated Usability Model

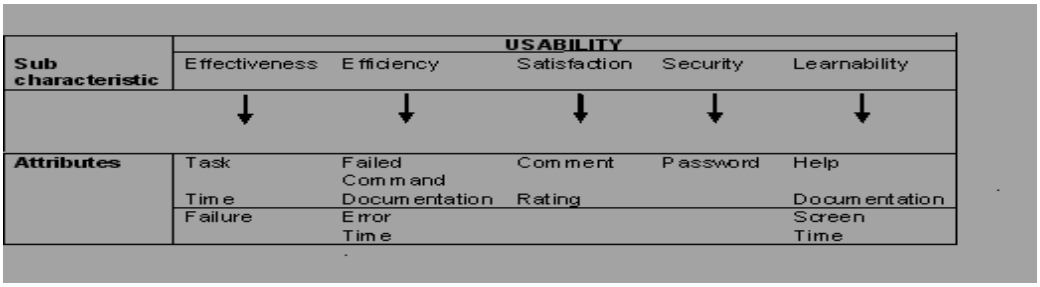


Figure 2. ISO Consolidated Usability Model

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

The Design of RS232 and CAN Protocol Converter Based on PIC MCU

Xianjun Wang & Wencheng Guo

School of Computer Technology and Automation, Tianjin Polytechnic University

63 Chenglin Road, Hedong District, Tianjin 300160, China

E-mail: wangxianjun1985@163.com

Abstract

CAN bus will be increasingly used in wide range of applications for its superiority, but it couldn't communicate with computer directly. The article describes the design of RS232 and CAN bus protocol converter depending on PIC Microcontroller, which solves the problem that CAN networks can not directly communicate with PC. At present, single-chip microcomputer of 51 series with CAN controller SJA1000 are widely used in domestic. Considering cost and converter size, the paper will talk about PIC18F2580 with integrated CAN microcontroller designed for RS232 and CAN protocol converters to facilitate the direct communication between computers and CAN bus.

Keywords: RS232, CAN, PIC MCU

1. Introduction

CAN bus are being used more and more widely in the areas of automotive, machinery, CNC machine tools, medical devices, smart sensors for its high-performance, high reliability and flexible design features. However, data on CAN bus could not directly communicate with computer.

At present there are more and more MCU with internal CAN controller integration. In this paper, CAN bus and RS232 protocol converter is designed based on Microchip's PIC18F2580 MCU, so that computer can read data on CAN bus directly and also send control commands to CAN bus through the RS232 interface.

2. The introduction of CAN Bus

CAN (Controller Area Network) is one of the most widely used serial communication bus in the world. CAN was first used in the automobile industry at the end of 1980s, which is proposed by the German Bosch company. CAN-bus specification has been developed as international standards the "ISO11898" (high-speed applications) and "ISO11519" (low-speed applications) by ISO. CAN bus optimizes the seven layers of communications architecture of Open System Interconnection model (ISO / OSI model) provided by ISO. And only the physical layer, data link layer and application layer can guarantee the data transmission speed and effectiveness. The basic design specification of CAN protocol not only demands a high bit rate, anti-electromagnetic interference, but also detects any possible errors.

CAN bus have the following main features: (1) Good real-time. CAN bus, which uses CSMA / CD media access method, is a multi-master field bus. This approach is similar with Peer to Peer approach of large-scale network, thus coincides with the open structure. Regardless of master-slave, information can be sent to free bus by any node at any time, which is a flexible means of communication, and this feature can also be used to constitute a (fault-tolerant) multi-machine backup system easily. Each node of the CAN receive all the information in the bus. Packet filtering and shielding mechanism enable each node quickly decide whether current packet enter into the receive buffer or not so that to reduce the information processing time which has nothing to do with the node. (2) Fault isolation is good. CAN uses non-destructive arbitration, when two or more nodes simultaneously send data to the network, the identifier of data frame determines the priority of the data frame, that small identifier means high-priority, large identifier equals to low priority. Nodes which send high-priority information continue to send data, and nodes which send low-priority information take the initiative to exit the bus. For a serious fault node, it will automatically turn off the bus function, which does not affect the work of other nodes but avoids network conflicts or a decrease of arbitration time. In the case of multi-transmission, lost information can be made up for through data fusion technology, which is caused by transmission or equipment failures (3) Long distance communication. The largest rate of CAN communication can be achieved 1Mbps/40m, the maximum transmission distance can reach lOKm/50Kbps. (4) Good anti-mistake transmission design. CAN with short-frame structure (8 effective bytes for each frame) allows short data transmission

time, low probability of interference and short re-sending time. And each frame of information has CRC error checking and other measures to ensure the low error rate of data transmission. (5) Theoretically CAN bus can connect 2000 nodes. In fact less than 110 is appropriate; it can satisfy the needs of majority users. CAN bus can transmit and receive data in several ways such as point-to-point, point-to-multipoint and overall broadcast. CAN bus transmission uses twisted pair and has no special requirement for the transmission medium.

3. PIC18F2580 introduction and hardware design

3.1 PIC18F2580CAN module introduction

PIC MCU has advanced RISC-like (Reduced Instruction Set Computer) structure which is reflected in each of efficient and powerful command. Among the command systems, all commands are single-cycle, single-word commands except procedures branch command which is two-cycle, single-word command. There is no cross-functional phase of instructions among these commands, so that all commands are with simplicity thus improve the efficiency of the software coding and reduce the required procedure storage unit, allowing the system with the highest efficiency and outstanding performance, which sets a new performance standard for 8-bit microcontroller market. Advantages of PIC microcontroller are also reflected in following ways: the function of each command is powerful, with high implementation efficiency. Commands digit for Low-grade, mid-grade, high-grade series are 12, 14, 16 respectively, and general command digit are only 33, 35, 58 while upward compatible. However, the CISC (Complex Instruction Set Computer) architecture microcontrollers usually have up to 50-110 multi-byte and multi-cycle commands. Therefore, PIC microcontroller series are easier to study, whose main features are as follows: (1) line structures of command can complete the implementation of a directive and the Fetch operation of next directive in a cycle simultaneously, which improves the efficiency of each internal clock cycle. (2) High-speed instruction execution time achieving 200ns at 20MHz clock or 160ns at 25MHz clock. It can carry out any operation of I / O port bit in a single cycle. Only operations on branch procedure CALL, GOTO and PC do require two-cycle execution time. (3) The system uses byte wide instructions, so that procedure volume is significantly reduced in the same circumstances, and the formation of the code is generally half of the microcontroller.

High-end product PIC18F2580 Single chip is introduced by Microchip of the United States, which applies 16 byte RISC instruction set with short instruction cycle, strong processing ability and high computing power. Data communication can be completed and communication protocols requirements can be satisfied without expanded memory. PIC18F2580 chip integrated single-chip A / D converter, the internal EEPROM memory, output compare, input capture, PWM output, I2C and SPI interfaces, asynchronous serial communication (USART) interface circuit, CAN Bus interface circuit, FLASH memory read / write and etc. Besides, it has powerful chips, strong I / O port drive capability, and can be connected directly with the LCD interface and the design of the circuit is simple and reliable. Addressable Universal Synchronous / Asynchronous Receiver Transmitter (USART) module and the CAN control module can easily achieve data communication in industrial field which has a good prospect.

Its main features include: (1) 1536 bytes RAM; (2) 32Kb FLASH, among which 1Kb is dedicated to USB buffer; (3) 256 bytes EEPROM data memory; (4) supporting RS232, RS485 and EUSART module of LIN serial interface; (5) supporting I2C and master synchronous serial port of SPI communications; (6) 10-bit A / D converter, precision up to ± 1LSB, which is equipped with up to 13 input channels; (7) two analog comparators; (8) with 16-bit data capture and the resolution capture / compare / PWM module; (9) enhanced capture / compare / PWM module, there is dead-zone control and fault protection input; (10) 4 timers (three 16 bytes timers and one 8 bytes timers); (11) programmable under-voltage reset and low voltage detection circuit; (12) enhanced online debugging features, add up to three hardware breakpoints.

3.2 PIC18F2580CAN module introduction

The Controller Area Network (CAN) module is a serial interface which is useful for communicating with other peripherals or microcontroller devices. This interface, or protocol, was designed to allow communications within noisy environments.

The CAN module in 18F2580 is a communication controller, which is the topic involved in this paper. It implementing the CAN 2.0A or B protocol as defined in the BOSCH specification. The module will support CAN 1.2, CAN 2.0A, CAN 2.0B Passive and CAN 2.0B Active versions of the protocol. The module features are as follows:(1) consistent with the ISO model; (2) Implementation of the CAN protocol CAN 1.2, CAN 2.0A and CAN 2.0B; (3) The CAN module supports the following frame types: Standard Data Frame; Extended Data Frame; Remote Frame; Error Frame; Overload Frame and so on (4) with two priority receive buffer and three priority send buffer; (5) 6 to accept filter: 2 high priority receive buffers, and the remaining 4 low priority receive buffers; (6) with 2 shielding filters, corresponding to two different receiver buffer; (7) with six kinds of operating mode settings: Request Configuration mode; Request Listen Only mode; Request Loop mode; Request Disable mode; Request Normal mode; (8)supporting short-frame structure, Standard and extended data frames, 0-8 bytes data length; (9) Programmable wake-up functionality with

integrated low-pass filter; (10)Signaling via interrupt capabilities for all CAN receiver and transmitter error states; (11)Programmable Loop mode supports self-test operation; (12)Programmable clock source; (13) Programmable link to timer module for time-stamping and network synchronization;(14)Low-power Sleep mode.

The CAN bus module consists of a protocol engine and message buffering and control. The CAN protocol engine automatically handles all functions for receiving and transmitting messages on the CAN bus. Messages are transmitted by first loading the appropriate data registers. Status and errors can be checked by reading the appropriate registers. Any message detected on the CAN bus is checked for errors and then matched against filters to see if it should be received and stored in one of the two receive registers. The CAN module uses the RB2/CANTX and RB3/CANRX pins to interface with the CAN bus. In normal mode, the CAN module automatically overrides TRISB<2>. The user must ensure that TRISB<3> is set.

CAN module of PIC18F2580 is configured with three send buffers. If needed, TXBnCON (n is 0, 1, 2) TXPR1, TXPR0 bit can be modified to set 4 different priorities. If the two send buffers have the same priority, then the send buffer with a bigger number has the higher priority.

Messages will be sent in order according to the level of priority. You can start sending messages by sending data to TXBnDm (n is 0, 1, 2, m is 0-8) and then set TXREG bit of TXBnCON. You can determine whether the messages are sent successfully through TXB0IF bit of PIR3.

CAN module of PIC18F2580 is configured of the 2 receive buffers, 6 receive filters and 2 receiver shielding filter. RXB0 corresponds to the receiving filter RXF0, RXF1; shielding filters RXM0, RXB1 correspond to the receiving filter RXF2, RXF3, RXF4, RXF5, shielding filters RXM1. In addition to these two receive buffers; PIC18F2580 is also equipped with a MAB (Memory Allocation Block) which receive all messages from the bus. When the node detects the messages on bus, the messages will be transmitted to MAB and will take its own arbitration and receive filters for comparison. If met, messages will be transmitted to the corresponding receive buffer. Shielding filter will decide which receive filter bit is effective.

3.3 Universal Synchronous Asynchronous Receiver Transmitter(USART)

The Universal Synchronous Asynchronous Receiver Transmitter (USART) module is one of the three serial I/O modules. (USART is also known as a Serial Communications Interface or SCI.) The USART can be configured as a full-duplex asynchronous system that can communicate with peripheral devices, such as CRT terminals and personal computers. It can also be configured as a half-duplex synchronous system that can communicate with peripheral devices, such as A/D or D/A integrated circuits, serial EEPROMs and so on.

3.4 Circuit hardware design

The system mainly includes two parts, the CAN bus driver circuit and the RS232 driver circuit. CAN bus transceiver chip and MCU are isolated by high-speed opt coupler 6N137 to avoid the affecting of CAN bus fluctuations on the normal operation of MCU and to improve the system anti-interference ability. Power and Ground wire on both sides of 6N137 must be isolated too. PCA82C250, which is interface between CAN bus protocol controller and the physical bus, is used as a CAN bus transceiver chip. It provides different send performances for bus and different reception performances for the CAN bus controller and it is fully compatible with the ISO11898 standard. The purpose of the use of PCA82C250 are increasing communication distance, improving the transient system anti-interference ability, protecting the bus, reducing radio frequency interference (RFI) and implementing thermal protection. It should be noted: resistor RS between PCA82C250 No. 8 pin and floor is called slope of resistance and its value determines whether the system is in high-speed working condition or the slope control mode. No. 8 pin should be connected directly with the floor so that the system will be in a high-speed working condition. In this manner, in order to avoid radio frequency interference, it is recommended to use shielded cable for the bus. While at a lower baud rate and short bus, slope control mode will be generally used and slope of the increase and decrease depends on the RS resistance value. CAN bus interface part is as shown in Figure 1.

RS232 has been widely applied as a standard computer serial communication interface.. However, the transmission distance is short with a maximum 15M. Transfer rate is also relatively low with a maximum of 20Kb/S. Imbalance are used to transmit data, that is, single-ended communication. RS232 protocol uses negative logic instead of TTL-level interface standard. Negative logic: logic "1":-3V ~-15V, logic "0": +3 V ~ +15 V. MAX232 chip can convert 5V TTL level of single-chip to ± 15V level which is needed by RS232. The system required only a single +5 V power. The serial ports TX and RX of PIC18F2580 are connected with MAX232 which can carry out full-duplex asynchronous communication with other RS232 interfaces. RS232 interface part is as shown in Figure 2.

4. Software design

In order to improve the operating efficiency of firmware, the main program initializes the system, interrupts it then and makes some analysis and necessary treatment in interrupt service routine and set the corresponding variable signs and

data buffer. The main program queries the variable signs circularly and let appropriate subroutine for processing. This program structure allows the main program be able to deal with various data transfer tasks in foreground, and at the same time can treat bus cases in the background timely through disruption. Initialization of main program includes initializations of I / O port, timers, serial port, CAN controller, interrupt and user signs data. This process gives configuration definition to basic resource of PIC18F2580 and configures reused I/O resource CAN interface and serial port. After setting up a 450ms timer interrupt, defining the serial port and operating parameters of CAN interface, opening the corresponding interrupt source, putting the user signs data the initial value, initialization process ends. MCU enters into the variables circulatory state.

CAN initialization mainly includes configuration of the CAN module, setting send mail, receiving mail identifier and initialization data, setting baud rate and CAN work mode, initializing receive filter and receiving shielding and CAN communication module initialization procedure:

```
Void CAN init (void)
{TRISB = 0x08;            // Set CAN input and output
CANCON = 0X80;      // request to enter CAN configuration mode REQOP = 100
While (CANSTAT & 0X80 == 0) {;} //Wait to enter CAN configuration mode OPMODE = 100
  BRGCON1 = 0X01; / / set SJW and BRP, SJW = 1TQ, BRP = 01H
BRGCON2 = 0X90; / / set Phase_Seg1 = 3TQ and Prog_Seg = 1TQ
BRGCON3 = 0X42; / / set Phase_Seg2 = 3TQ / / Set send mail 0 symbols and data
TXB0CON = 0X03; / / Sending priority is the highest priority, TXPRI = 11
TXB0SIDH = 0XFF; / / Set 0 Standard Identifier for Send Buffer, the procedures use Standard Identifier
TXB0SIDL = 0XE0;
TXB0DLC = 0X08; / / Set the data length as 8 bytes
/ / Set receive mail 0 identifier and initialization data
RXB0SIDH = 0XFF; / / set the receive buffer identifier 0
RXB0SIDL = 0XE0;
RXB0CON = 0X20; / / just to receive valid information of standard identifier, FILHIT0 = 0 means RXB0 uses filter0
RXB0DLC = 0X08; / / set the length of data area for receive buffer 0
RXB0D0 = 0X00; / / initialize data from receive buffer 0 data zone
RXB0D1 = 0X00;
RXB0D2 = 0X00;
RXB0D3 = 0X00;
RXB0D4 = 0X00;
RXB0D5 = 0X00;
RXB0D6 = 0X00;
RXB0D7 = 0X00;
/ / Initialize the receiving filter 0 and receive shielding,
RXF0SIDH = 0XFF;
RXF0SIDL = 0XE0;
RXM0SIDH = 0X00;
RXM0SIDL = 0X00;
/ / Initialize I / O control register of CAN Module
CIOCON = 0X00;
CANCON = 0X00; / / enable CAN enter normal operating mode
While (CANSTAT & 0XE0! = 0) {;}
/ / Initialize CAN interruption
PIR3 = 0X00; / / clear all Interrupt Flag
```

PIE3 = 0X01; / / interrupt receivers which can receive buffer 0

IPR3 = 0X01; / / receiving interruption for receive buffer 0 is at highest priority

}

USART module initialization procedure:

Void USART init (void)

(TRISC = 0x80; / / set the serial input and output

SPBRG = 0X4d; / / select the baud rate for transmission 9600bps

TXSTA = 0X04; / / select asynchronous high-speed 8-bit data transfer

RCSTA = 0X80; / / permit synchronous serial port work

TXSTAbits.TXEN = 1; / / send permit

RCSTAbits.CREN = 1; / / accept the data permit

PIE1bits.RCIE = 1; / / Receive Interruption Enable

IPR1bits.RCIP = 0; / / receive interruption is at low priority

}

After the completion of all I/O Configuration and CAN initialization subroutine, the main program waits for the connection between MCU and PC. After communications between PC, procedures are ready for data exchange: If the system receive valid CAN data, then the data is sent to the host computer; if the system receive the effective data and then will send it to the CAN bus. The main program flow chart is as shown in figure 3.

5. Results of experiment

This system use serial port of personal computer to receive and transfer data on RS232, and connected to the CAN network through the PIC MCU and the CAN communication module in the MCU. Converters can be two-way communication: on the one hand it can receive data from the CAN bus and transfer it to the computer in standard RS232 format; On the other hand it can be put the data flow that received in RS-232 format into the data flow that compliance with CAN protocol and sent it to the CAN bus. Give full play to the CAN bus communication long distance, node capacity, high reliability, as well as through the host computer to monitor the operation of bus and can be directly sent instructions to CAN bus. PIC18F2580 chip in SOIC package also has greatly narrowed the converter size which makes the industrial applications convenient.

References

Feng Wu&Xinjian Chen. (2005). *PIC Single Chip C language Entry.* Beiijing: Beihang University Press.

Guoxian Chen. (2004). *PIC Single-chip Microcomputer Principle and Interface Technology*. Electronics Industry Press.

Microchip Technology Inc. *PIC18F2x8 data sheet*. [Online] Available:www.microchip.com

Xianhui Yang. (2000). *Field Bus Technology and Applications*. Beijing: Tsinghua University Press.

Yu Wang. (2006). *PIC Single Chip Entry and Improve*. Beijing: Machinery Industry Press.

Yuntao Rao, Jijun Zhou&Yongyun Zheng. (2003). *CAN Bus Principle and Application Technology*. Beijing: Beihang University Press.

Zexiang Liu. (2005). *Field Bus Technology*. Beijing: Machinery Industry Press.

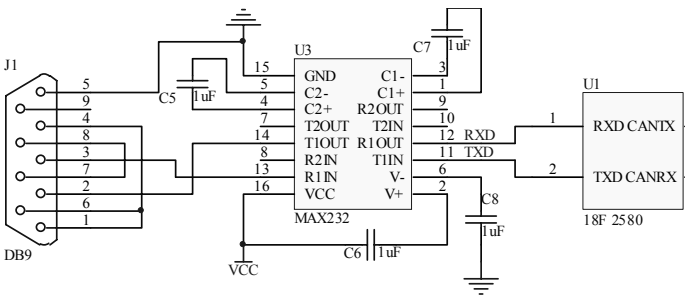


Figure 1. CAN Bus Interface Circuit

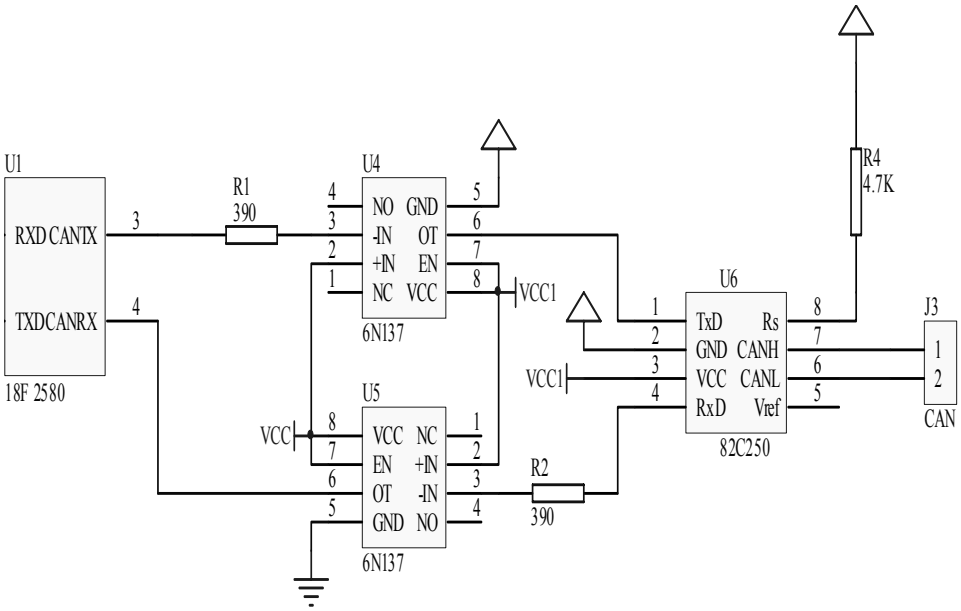


Figure 2. RS232 interface circuit

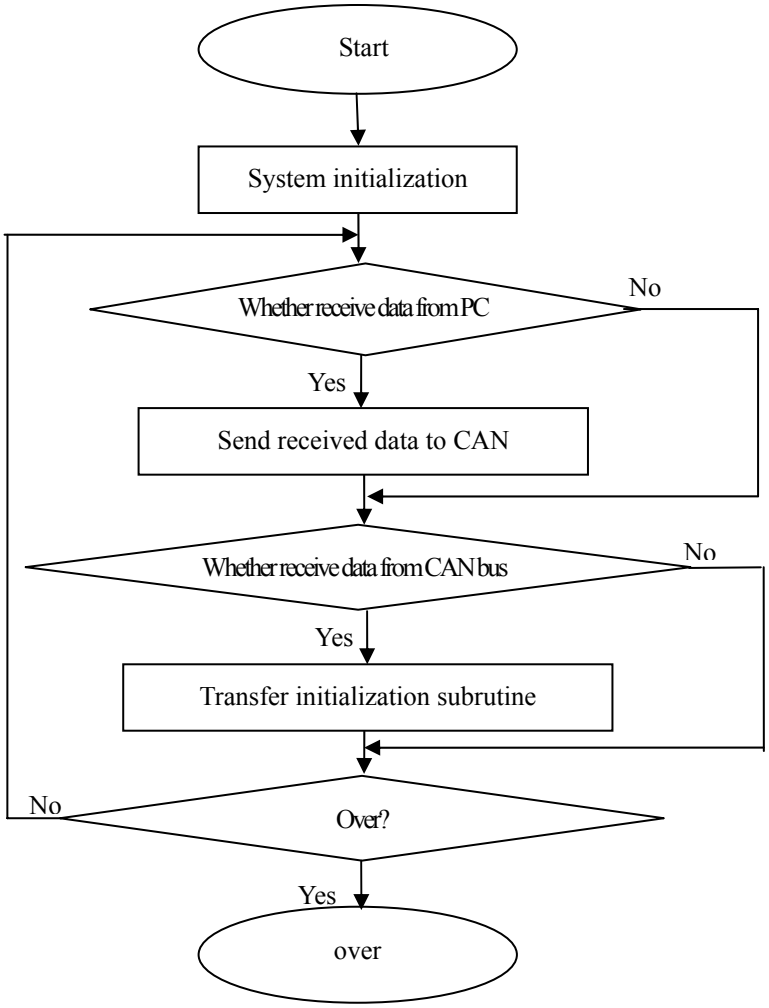


Figure 3. main program flow char

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 2, No. 3

August 2009

Development of Individual Learners:

Perspective on the Uncertain Future Contribution of E-Learning

Zaki Ahmad Dahlan

Department of Surveying Science and Geomatics, Faculty of Architecture Planning and Surveying

Universiti Teknologi MARA Perlis, 02600 Arau, Perlis, Malaysia

Tel: 60-4-988-2858 E-mail: zaki@perlis.uitm.edu.my

Nor Arzami Othman

Department of Computer Science, Faculty of Computer Science and Mathematics

Universiti Teknologi MARA Perlis, 02600 Arau, Perlis, Malaysia

Tel: 60-4-988-2811 E-mail: arzami@perlis.uitm.edu.my

Mohd Fazril Izhar Mohd Idris

Department of Mathematical Science and Statistic, Faculty of Computer Science and Mathematics

Universiti Teknologi MARA Perlis, 02600 Arau, Perlis, Malaysia

Tel: 60-4-988-2679 E-mail: fazrilizhar@perlis.uitm.edu.my

Hj. Kamaruzaman Jusoff (corresponding author)

Tropical Forest Airborne Observatory (TropAIR), Faculty of Forestry

Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Tel: 60-3-8946-7176 E-mail: kjusoff@yahoo.com

Abstract

E-learning in the educational world has grown and changed rapidly in recent years. Both private and public sector organizations have embraced the practice of reaching their students at a distance via new technologies afforded them by Information Communication Technology (ICT) infrastructure. E-learning is grounded on technology, and without it, its practice would be difficult, if not impossible. We can see that the use of the internet and ICT are becoming an important part of learning and teaching strategies in many educational institutions. Knowing that education has always been an important engine for economic development, the Malaysian government has taken initiative steps to implement public awareness on the ICT issues. An important part of e-learning has been to contribute to the development of individual learners whatever their life circumstances. E-learning as a field of educational endeavor is at a crucial juncture in its historical development. The notion of learning at a distance has gained wide acceptance across the developed world. Instructors, physically and temporally separated from learners using newly emerging information and communication technologies, are widespread. The potential of latest technology has adopted in creating new learning environments. The rational behind this endeavor is the expectation that unique features of the Information Communication technology. It also can include a range of powerful media forms and its interactive capability that support sophisticated range and interaction in teaching. As a result, these approaches will provide a rich environment for teaching.

Keywords: E-learning, Information and communication technology, Individual learners, Internet, Multimedia

1. Introduction

E-learning is not a recent phenomenon to Malaysia. Most of the people work harder for higher qualification, especially those in employment who aspires to be successful in their career development. E-learning was conducted by supplying

to registered students with reading materials, reference and assignments by accessing from the web. At the end of session students will have to sit for the scheduled examinations at different recognized centers in the country. In person teaching was not common, but educational entrepreneurs normally took the opportunity to provide some form of tuition to those students who were preparing for their examinations. In certain cases, there was no interaction at all between students and learning institution.

Information and Communication Technology (ICT) pushes our society from one of the industrial production to a dynamic, knowledge-based service and information society. Social processes change rapidly as is shown by individualization and globalization. Demographically, the most important issue is that more people grow old and fewer babies are born. Such trends lead to the need for 'life long learning'.

The need to learn is definitely there, but motivation is an important issue. All of the learners are required to participate fruitfully in our modern society. In addition, e-learning environments that engage learners are required learning processes. There are some issues that need to be looked into:

- What is the impact of social and technological changes on learners and their learning processes?
- What are the required learning processes in order to participate productively in our (future) society?
- How can we facilitate such learning processes?

E-learning describes the way new information and communications technologies (ICT) are set to re-invent education and learning in a digital world. It means that internet enabled learning by an exciting range of opportunities for educators and learners alike to use new skills and tools to prosper in an information society. E-learning means the delivery of learning with the assistance of interactive and electronic technology. The delivery and administration of learning opportunities are done via computer networked and web-based technology, to help individual performance and development. It is more than just training on a computer as it encompasses dissemination of information, performance support, and knowledge management. It involves not only access to training materials but also offers the management of learning by providing both content and administration.

Approaches such as Computer Based Training (CBT), Web Based Training (WBT), electronic performance support systems and video conferencing have to been integrated. Some people prefer to use the new approaches such as knowledge media, technology-assisted learning and *technology*-based training. Sometimes the term Technology Based Training (TBT) has been used to refer to computer-mediated learning. There is a better term to cover the entire alternative approaches. This e-word is a sibling of e-business where e-learning exploits the technology of the World Wide Web. E-learning is not, however, restricted to the Web. It also includes training delivered via stand-alone computers like CBT. Nowadays, we are moving toward Mobile Based Training.

2. Method

Learning processes should be reconstructed into a framework of competencies where the people can develop for a 'life long learning'. The categories are prerequisite literacy competencies (e.g. English, ICT), learning competencies, social competencies, professional competencies and career development competencies. Concerning the socio-constructivist learning theory, people can develop such competencies by being actively involved in constructive learning activities, within the setting of a 'rich' learning environment. Such an environment resembles real life, is motivating and provides learners with the opportunity to transfer learning experiences to daily practice. Along this lines, problem-based learning, case-based learning and methods such as learning by discovery are strongly preferred (Veerman, Riemersma & Veldhuis, 2004).

2.1 The role of e-learning

With the growing number of on-line courses, the increasing accessibility of computers, the rapid growth of internet and intranets, and the increasing number of computer users, students of all ages are taking advantages of distance learning/distance education or are using computers to enhance traditional classroom experience. E-learning, which describes the use of Web or Internet technologies to enhance teaching and learning experience, is therefore, a suitable solution to encourage education to anyone, anywhere, at anytime. In addition, the overall operating costs for the national education system can be reduced once the e-learning systems have been implemented.

Conventional classroom-based and distance education has collided in the realm of e-learning (Muirhead, 2005). The challenge now is to support university teachers in the new blended environment characterized by elements of both conventional classroom and distance learning. Our paper is a critically reflective case study of one strategic project to deliver such support. E-learning in university teaching is still generally considered an educational innovation rather than the use of teaching and learning strategies embedded in university culture and practice. An institutional strategic plan is essential but only the first step. Uptake of an educational innovation is about personal and, often, institutional change, whether desired or required.

While there is a growing base of literature on descriptive case studies of the implementation of e-learning strategies (Steeples & Jones, 2002), they observe that only a fraction use an explicit theoretical framework to examine the experience. This is not surprising as a recent wider study of published educational research concluded that more explicit theoretical engagement is needed (Tight, 2004).

Buch & Sena, (2001) posed important questions when looking at the benefits of using the Internet in education. The questions they asked were:

- Do individuals learn differently with on-line instruction than traditional classroom situations?
- Can the same teaching strategies be used with on-line as traditional methods and will they be as effective?
- Do individuals' react differently to on-line instruction and how can individual differences in learners be accommodated?

2.2 Independent Learning

Independent learning focuses on creation of the opportunities and experiences necessary for students to become capable, self-reliant, self-motivated and life-long learners. The desired outcome would be students who value learning as an empowering activity of great personal and social worth. Independent learning is that learning in which the learner, in conjunction with relevant others, can make the decisions necessary to meet the learner's own learning needs. In this method a process independent learners develop the values, attitudes, knowledge and skills needed to make responsible decisions and take actions dealing with their own learning. Independent learning is fostered by creating the opportunities and experiences which encourage student motivation, curiosity, self-confidence, self-reliance and positive self-concept; it is based on student understanding of their own interests and a valuing of learning for its own sake.

This method is part of an ongoing, lifelong process of education that stimulates greater thoughtfulness and reflection and promotes the continuing growth of students' capabilities and powers. More than the rote learning of facts and skills, this approach to learning encourages students to make meaning for themselves, based on their understanding of why and how new knowledge is related to their own experiences, interests and needs.

It is a direction for the process of education, not an absolute standard; this process takes different forms for different students and it varies according to subject matter and students' interests and abilities in the subjects. The process involves the teacher and learner in an interactive process that encourages students' intellectual development and their capacity for independent and reflective judgment.

Independent learning is fostered by a school environment which is sensitive, flexible, democratic and responsive to the needs of students. This encourages a strong sense of purpose and motivation on the part of students. Independent learning makes full use of the resources of the school and the community and fosters the development of independent learners in every grade and in every subject.

2.3 Technology Literacy

The goal of incorporating technological literacy is to develop individuals who understand how technology and society influence one another and who are able to use this knowledge in their every day decision making. What is desired are students who are able to analyze the technological influences on their lives and see themselves as having roles and responsibilities in shaping public policy related to technological changes?

Technological literacy can be described as the intellectual processes, abilities and dispositions needed for students to understand the link between technology, themselves and society in general. Technological literacy is concerned with developing students' awareness of how technology is related to the broader social system, and how technological systems cannot be fully separated from the political, cultural and economic frameworks which shape them. In order to achieve an informed, balanced and comprehensive analysis of the technological influences on their lives and then be able to act on the basis of their analysis, students require certain levels of knowledge, skills and abilities.

A technologically literate person is someone who critically examines and questions technological progress and innovation. To be critical about technology means to have the intellectual skills to analyze the pros and cons of any technological development, to examine its potential benefits, its potential costs, and to perceive the underlying political and social forces driving the development. Decisions about the creation and use of new technology involve human, social and environmental issues which place constraints and limits on the solutions. Values also influence intellectual processes, since anything that involves choice also involves consideration of whose values are shaping a particular technological development. The capacity to make critical judgments involving technology increases the ability of students (as citizens) to use such knowledge to shape and influence their environment.

2.3.1 Technology and society

An understanding of how technology shapes and is shaped by society is based on two principles. First, technology is an agent of social change. Technological systems have produced great increases in the speed of communication, made

mass production possible, and reduced physical labor for human beings. Technology has also provided an abundance of products and services; this requires individuals to make wise consumer decisions. Second, societies influence the course of technological development. Social, historical and cultural factors determine if, and how, a technology is used. More than ever before, students are immersed in an environment shaped by human technology. Citizens' responsibilities in a technological age are defined as the exercising of democratic rights on issues that affect the direction to be taken by society and its technological developments. Becoming technologically literate means coming to grips with the problems of living in, and exerting influence upon, the constructed world.

2.4 Challenges in E-learning

Since the emergence of the Internet and the World Wide Web (WWW) in providing instruction in the mid-1990s, there have been numerous studies about the problems of designing web-based instruction. Most of these studies have had "common shortcomings" in that they have failed to develop a theoretical or conceptual framework of web-based, or online instruction (Jung, 2001). Indeed, the process of designing online instruction can be so complex and difficult that educators often end up adopting curriculum to fit the technology rather than selecting the proper technology to support the curriculum.

Many educators now believe that the unique environment of online learning necessitates a reexamination of the learning process, in many instances a paradigm shift in pedagogical practice. For instance, changing a traditional face-to-face course to an online course does not mean posting lectures online in a text-based format. Rather, it involves a transformation of both teaching and learning, a process that requires training and possibly a change in an instructor's style and expectations.

Time is another challenge. Faculty must work with time constraints and communicate and follow-through with email, grading, discussion boards, and online chats. They must be able to support and nurture a community of learners, motivate and inspire, gain their attention, and get them to learn. At the same time, faculty must also be cognizant of available and evolving technologies and how to use them to effectively support and enhance student learning. As a result, educators need to constantly reflect upon, improve, and update their practice, understanding how to best design instruction to support student learning. These can be difficult, if not impossible, goals, given the time that most instructors of higher education must spend on teaching, research, and service (Turley, 2005).

Faculty may also need to learn new skills to create and implement rich online learning experiences. Those who want to augment their instruction with online components need to learn how to use those tools, such as synchronous meetings, tutorials, simulations, multimedia lessons, instant messaging, blogs, course management systems, and other interactive multimedia formats. Additionally, instructors need to understand human learning processes. According to Clark & Mayer (2003), when the limits of human cognitive processes are ignored, instruction that employs all of the technological capabilities to deliver text, audio, and video can actually reduce or hinder learning. An understanding of educational psychology, instructional design, multimedia production, graphics, and interface design are necessary to translate these principles into effective online instruction. Although new technologies ease the burden of knowing a programming language, it still takes from ten to twenty times more labor and skill to produce good courseware for online learning than for traditional classrooms (Clark & Mayer, 2003).

Another challenge of online learning environments is the shortage of technical staff to help faculty, students, and staff. This shortage can put a strain on developing web-based programs and delay worthwhile projects. Many issues continue to confront institutions of higher education in the realm of online learning. The following are the goals for development of distance learning programs.

- Reducing per-student costs
- Making educational opportunities more affordable for students
- Increasing institution enrollments
- Increasing student access by reducing time constraints for course taking
- Increasing student access by making courses available at convenient locations
- Increasing institutional access to new audiences
- Improving the quality of course offerings
- Meeting the needs of local employers

The good news is that instructors now have access to rich multimedia tools to enhance instruction. The bad news is that multimedia software is often used in instructionally-deficient ways. For instance, PowerPoint is multimedia software that is easy to use, but can be detrimental to learning if used in the wrong ways. Faculty in higher education may need to receive training on how to effectively integrate multimedia in instruction. This is indicated by the availability of training courses offered by various universities.

3. Result and discussion

3.1 Opportunities of E-learning environment

Western industrialized countries are undergoing a transformation from industrial societies to information and communications societies or knowledge-based societies. Almost all spheres of life are affected by these radical changes, which, however, also offer new opportunities. The education sector is both affected and challenged by this change. New content-related and structural requirements have to be met by initial and continuing training but at the same time the new media offer new opportunities for processing and presenting knowledge and for organizing the teaching process. Self-directed and assisted learning will be thoroughly transformed by the new media. Digital processing of knowledge becomes increasingly important, and new forms of teaching are emerging. Fresh impetus is given to the vision of self-directed learning at any place. Modern information and communications technologies are opening up new training and continuing training options which enable participants to independently organize their learning, thus adapting to quickly changing qualification requirements. Web-based learning breaks up traditional structures of learning and combines initial and continuing training more than in the past in terms of content and organization.

Broad application of information and communications technologies in the education sector and a multimedia-based teaching approach offer an opportunity to better prepare young people already at school for the requirements of the knowledge-based society in their private and professional lives and familiarize them early on with multimedia based learning, develop further the tried and tested, internationally recognized dual system of vocational education and training, support structural change in higher education and enhance the international competitiveness academic teaching, better prepare employees for changes at the workplace and on the labor market and include them in the organization of corporate development processes, facilitate self-directed site-independent learning, which benefits mainly those men and women who, due to family duties, illness, old age or other circumstances, want to learn at home, introduce new, cooperative forms of teaching and learning.

3.2 Technologies convergence shaping the future of E-learning

As we have now moved into the digital and internet information age, where e-learning courses can be accessed from anywhere, we must ask the question, how will new technologies affect the future of e-learning? First, what are the new technologies? Some of these new and future technologies consist of video conferencing, TV/PC bridging, broadband access for wire technology, virtual reality programs, and multi sensory technologies. All of these technologies and more could revolutionize the way digital learning is implemented in the future.

Video conferencing has been the dream of the future, for quite some time now, however, only businesses seem to use it. In order to set up videoconference, you must get an IP address, however your IP address changes every time you sign off and on. Meaning, if you want to video conference, you have to communicate with the person before, in order to get their IP address to reconnect with them, which is both inconvenient, and unnecessary. The technology is there, it just needs better implementation, and spread more universally before it can be used for online learning.

Video conferencing will actually allow people to take e-learning courses together, which will improve the integration process, as well as other processes, because interaction with other people will actually bring it closer to real life situations, in which the learner can relate and apply the knowledge to everyday life. The collaborative potential of Computer Mediated Communications (CMC) for learning was realized utilizing computer conferencing technologies.

Today's wireless network is fast enough for digital music and photos, but when you try to download a streaming video over a network, the quality is low, and it needs to be compressed, or it will stall and jump, which is not pleasant to watch. So finally, there is a new form of network connection on the horizon, and although the best thing now is the Ethernet cord, some new devices will be for sale by the end of the year. Broadband bandwidth is an important aspect for the future of e-learning, as it will allow more courses to be taken online without a CD, allowing more streaming audio, video, animation, and other graphics, plus more interactive two-dimension and three-dimension environments. These technologies will allow designers to create advanced e-learning courses that will engage the learner at an all-new level (Hammond, 2005).

Virtual reality and multi sensory technologies are still infants in terms of technology. However, with further development, these technologies could expand e-learning beyond the standard of what we're used to. E-learning will no longer be limited to the mouse, keyboard, and monitor. There will be gloves that will allow you to feel what somebody else is feeling. As they mold clay, you too will know what it feels like to sculpt. Learning will go beyond emotional and mental capacity, it will actually tap into your sensory motor skills and with virtual reality, and you could see something and feel something through advanced courseware that is purely based on new technologies linked with e-learning. Think about flight simulation programs that have been used for several years now, in training pilots, astronauts, and military personnel. With the technologies that already exist, and the potential of these other technologies, the possibilities of e-learning in the future are endless.

Overcoming the hurdle of limited wire-technology, improving the speed of networking, increasing the use of video conferencing, more online interactivity, and the introduction and application of virtual reality and multi sensory activated technologies, will allow digital learning to become a higher form of education. These are merely introductory products into the technological marketplace, and although they cannot fully be utilized in the marketplace of e-learning, because not enough people have them yet, there is the great potential for these products to be used in the future.

E-learning in terms of corporate objectives is far different than anything that universities are concerned with. The courses are based on money, as all things are with business. When corporations with thousands of employees are required by law to teach their employees about something, they can save time and money by hiring a company to make an e-learning course for them, as opposed to making them all take classes on the subject. For the corporation, the most important thing about the e-learning course is the learning outcome. They will set a learning outcome goal of 80%-90%, and once the employee has achieved this goal, they are done with the course, and the corporation is legally covered. Whether these courses are successful or not, is up for debate, however, these corporate e-learning situations make up the majority of all e-learning courses in existence.

The future of e-learning looks promising, as new technologies arise, and new and improved course designs and modules are developed, making some sort of prediction about what is to come with e-learning may not be to far out of reach, as the beginnings of several innovations are actually being utilized. E-learning will incorporate individual's learning styles and an emotional profile before the course is even taken, finding the best way to teach each individual learner. New interactive technologies will be integrated with the program, making the course exploration enjoyable and more interesting to the user. Learners will not be rushing through the course just to get it over with. In addition, with new technologies, digital learning may not be limited to just a keyboard, a mouse, and a computer screen. It is very likely that e-learning will become part of everyday learning, just as computers today are a part of everyday communication. Overall, digital learning has and will continue to affect the way people in society teaching (Hammond, 2005).

It is thought that E-learning is here to replace the Instructor Lead Training (ILT) with Computer Based Training (CBT). This is probably the most common reason why many people think that e-learning is a bad idea, can a machine replace the human mentor? But in fact, the true meaning of e-learning is to enhance the ILT with CBT. The best possible e-learning is the correct portion of Instructor Lead Training and Computer Based Training. Taking the best of both worlds creates a setting of more powerful, personalized and efficient learning.

Everybody is different, which means, everybody learns in his/her own personalized way. So what can online training be like? Synchronous training means that a certain lesson or task is taught within a time frame and everyone works within that timeframe. Asynchronous training lets the student study a lesson in his/her own time and pace. The quick learners work fast and slow learners take more time. The real advantage of Computer Based Training is that it's easy to get the training material.

One of e-learning's advantages is that resources are at hand. Traditionally teachers use books as the learning material, but we are moving into an era where teacher's material is all in a digital form. This ensures that the material is up-to-date, easy to distribute and maintain. When the course material is available on the Internet or Intranet, the student has access to it anytime and anywhere. Although having material digital requires that the teacher knows how to create the material and use computers.

E-learning gives a student an opportunity to practice practical exercises that would occur in real life, this is called real life interactivity. Without technology the student would only have a pen, paper and books. With the advantage of technology, the student can for example drive a virtual drive test with a car, which will point out all the errors the student makes and simulates real life traffic. This is just one example of what real life interactivity brings to learning. The following points must be considered prior to designing any e-learning program (Wall, 2004).

There is plenty of evidence to suggest that pure e-learning programmes do not work whereas there is significant evidence to suggest that blended learning programmes are much more likely to be successful.

- Information is not instruction.
- Instruction must be based on sound pedagogical principles.
- Individuals have different learning style preferences, different cognitive processes and different past experiences.
- Different learning situations require different learning strategies.
- Learners require direction and focus.
- A one-size fit all approach to e-learning has been shown to be ineffective.

Revolutions are exciting but technological revolution is no exception. They generate energy as they gather strength and bring more and more individuals and functions into their sphere of influence. The changes come so fast that is often

little opportunity to reflect on the implications, little opportunity to access, plan for, and cope with the effects of these changes (Wibunsin, 2002).

There are a number of benefits to tertiary learning online that are unique to the medium (Capper, 2001). (1) Any time. The Students can access the learning program at any time that is convenient for them. Not just during the specific numbers of hour that is set for a conventional course. It can be quick snatches at odd times or long late-night sessions. It is more flexible for them to learn. (2) Any where. The students do not have to meet face-to-face, meaning that this learning process can be anywhere. International sharing is feasible. Individuals can log on and access reading material, assignment either at work, home, the library, in a community learning centre or from their hotel when traveling. (3) Asynchronous interaction. By using an e-mail, the students will become more convenience to come out with questions and ideas. Electronic mail does not require participants to respond immediately. As a result, interactions can be more succinct and to-the-point, discussion can stay more on-track, and people can get a chance to craft their responses. This can lead to more thoughtful and creative conversations. (4) Group collaboration. Electronic messaging creates new opportunities for groups to work together, creating shared electronic conversations that can be thoughtful and more permanent than voice conversations. Sometimes aided by on-line moderators, these net seminars can be powerful for learning and problem solving. (5) New approaches. Many new options and learning strategies become economically feasible through online courses. For instance, the technology makes it feasible to utilize faculty anywhere in the world and to put together faculty teams that include master teachers, researchers, scientists, and experienced professional developers. Online courses also can provide unique opportunities for teachers to share innovations in their own work with the immediate support of electronic groups and expert faculty. (6) Computers Integration. The online learner has access to a computer, so computer applications can be used without excluding some participants. This means, for instance, that a mathematical model implemented in a spreadsheet can easily be incorporated into a lesson and downloaded so all participants can run, explore, and refine the model and then share their findings and improvements.

4. Conclusion

It can be conclude that the revolution of technology has been one of the most dramatic changes in modern history. Although not yet universal, this technology has been global in scope of human life. Learning is a process of active engagement with experience. It is what people do when they want to make sense of the world. It may involve an increase in skills, knowledge or understanding, a deepening of values or the capacity to reflect. Effective learning will lead to change, development and desire to learn more. By using or implementing e-learning, we can reduce training cost, training time and we also can increase learning effectiveness. I believe that in one day we can reach a peak at the future because e-learning gradually growth and promise for the developing world. A growing number of organizations are now delivering training and education over the Internet, including colleges and university, corporations, military institutions, and even secondary schools. Emerging of new technology and ICT infrastructure can help our nation to achieve their vision such as information society which it will culture ICT based economy. Knowledge society such as data worker and knowledge worker can produce knowledge products and services and lastly competitive society in which to face the challenge competitive knowledge economy.

References

Corre, B. K., & Sena, C. (2001). Accommodating diverse learning styles in the design and delivery of on-line learning experiences. *International Journal of Engineering Education, 17*(1), 93-98

Clark, R. C., & Mayer, R. E. (2003). E-Learning and the science of instruction. *San Francisco: Pfeiffer*.

Doris, W. (2002). Online Learning: The promise and the Reality *Retrieved April 10, 2005 from http://www.fareastern.ac.th/conference/docs/DorisPaper2.doc*

Hammond, P. (2005). E-Learning: Future Technologies, *Retrieved Jan. 25, 2006 from http://www.du.edu/~phammond/elearning6.html*

Joanne, C. (2001); E-Learning Growth and Promise for the Developing World. *Retrieved April 4, 2005, from http://www.techknowlogia.org/TKL_active_pages2/currentArticles/main.asp*

John, W. (2004). The Role Of E-Learning – A Review. *Retrieved March 14, 2005, from http://www.multimedia-instruction.com/Docs/*

Jung, I. (2001). Building a theoretical framework of web-based instruction in the context of distance education. *British Journal of Education Technology,* 32(5), 525-534.

Muirhead, B. (2005). A Canadian perspective on the uncertain future of distance education. *Distance Education,* 26 (2), 239-254.

Steeples, C. & Jones, C. (Eds.). (2002). Networked learning: Perspectives and issues. *London: Springer.*

Tight, M. (2004). Research into higher education: An a-theoretical community, *Higher Education Research & Development,* 23 (4), 395-411.

Turley, S. (2005). Professional lives of teacher educators in an era of mandated reform. Teacher Education Quarterly, 1. *Retrieved May 4, 2006 From http://www.findarticles.com/p/articles/mi_qa3960/is_200510/ai_n15743272.*

Veerman, A., Riemersma, J. & Veldhuis, G,. (2004). E-learning and the fading boundaries of society. *Retrieved May 5, 2005, from http://www.elearninternational.co.uk/docs/abstracts/Accepted%20Abstracts/Veerman.doc*

A journal archived in Library and Archives Canada
A journal indexed in CANADIANA (The National Bibliography)
A journal indexed in AMICUS
A journal included in DOAJ (Directory of Open-Access Journal)
A journal included in Google Scholar
A journal included in LOCKSS
A journal included in PKP Open Archives Harvester
A journal listed in Journalseek
A journal listed in Ulrich's
A peer-reviewed journal in computer and information science

Computer and Information Science

Quarterly

Publisher Canadian Center of Science and Education
Address 4915 Bathurst St. Unit # 209-309, Toronto, ON. M2R 1X9
Telephone 1-416-208-4027
Fax 1-416-208-4028
E-mail CIS@ccsenet.org
Website www.ccsenet.org
Printer William Printing Inc.
Price CAD.$ 20.00

ISSN1913-8989