Association for Information Systems
AIS Electronic Library (AISeL)

ICEB 2004 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-5-2004

A Risk Identification Method of Virtual Enterprise

Min Huang

Fei Xu

Xing-wei Wang

W.H. Ip

K.L. Yung

A Risk Identification Method of Virtual Enterprise

Min Huang[1], Fei Xu[1], Xing-wei Wang[1], W.H.Ip[2], K.L.Yung[2]

[1] Box 135, Faculty of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China, mhuang@mail.neu.edu.cn

[2] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Konghua.edu.cn

ABSTRACT

Virtual enterprise is the potential mode of future enterprise under global economic environment and the risk management for it is a popular research area recently. Risk identification is the key phrase of risk management. Considering the risk identification problem in the virtual enterprise, a new identification method based on the BP network is provided in this paper. Using this method and the data related to the problem, the possible to take place of the risks could be obtained by training the neural network, so that risks can be investigated in time and managed effectively. The simulation results prove its validity.

Keywords: virtual enterprise, risk management, risk identification, BP network learning algorithm

1. INTRODUCTION

Global economic environment is experiencing tremendous changes in recent decades as information technology and its application advance rapidly. Market demand is changing even rapidly and becoming more difficult to predict as the global market competition heats up. Under such background Virtual Enterprise (VE), a new organizational mode for enterprises, is gradually replacing traditional modes of enterprises [1] for it's more adaptive to new challenges. VE enables enterprises related to one product to combine together to form a temporary economic body so that they can grasp new opportunities and response to latest market demand as soon as possible. The new mode is considered profitable though it also brings in risks and uncertain factors, which might have harmful influences on enterprises involved. Traditional risk identification methods, depending on personal experiences and subjective judgments, are mostly qualitative [2,3]. They do not perform well in identifying new risks under the recent economic environment. Such shortage calls for a better risk identification method to ensure the secure operation of enterprises. Data mining (or knowledge discovering) technology, a process based on artificial intelligence and database technology in which some models or modes that are useful to data owner(s) are extracted from a large amount of data, is providing a new thought for resolving this problem. The present functions of the technology mainly include mode identification, association analysis, sorting, predicting and cluster analysis, etc.[4, 5]on the basis of algorithms and technologies of artificial intelligence. In the research risk identification problem is solved by data mining method based on BP network and with the help of Statistical Analysis System (SAS) software, the application is proved to be satisfactory.

2. PROBLEM DESCRIPTION

Successful risk management depends on scientific procedures. The process of risk management includes risk identification, risk estimation, risk evaluation, risk programming and effective risk control based on the previous steps. During risk identification, the primary step of risk management, it is necessary to determine the general objective of enterprise and the partial objectives needed for realizing the general one. Based on this, risks factors within the general objective and partial objectives are investigated and analyzed, which can be achieved through survey/ questionnaire, organization chart analysis, financial statement analysis, checking chart analysis, process analysis, input-output analysis, sensitivity analysis, accident-tree analysis, hierarchical holographic modeling analysis and expert investigation, etc [2,3]. These methods:

1) Require investigator or analyst's deep understanding of enterprises and long-term information accumulation;

2) Require long-term investigation and analysis;

3) Respond slowly to environmental changes.

However, the VE is formed dynamically, which means operators can only acquire information and data of pre-existing VE and original enterprises. Consequently, VE should decide based on the data whether the current risks are possible to take place. A new method towards such a problem is introduced in the research --- BP-network-based risk identification of VE. According to this method, every possible kind of internal or external risks is firstly analyzed and the information concerned defined; then the possibility of each kind of risk is got according to data of pre-existing VE, original enterprises and information of current VE.

3. NEURAL-NETWORK MODEL

BP-network is one of the most widely-used neural networks at present. Before putting BP-network into practice, the number of hidden hierarchies and that of units in each hierarchy have to be fixed so as to determine a network topology. Figure 1 shows a 3-hierarchy forward neutral network with only one output nodal point. From left to right aligns input, hidden and output hierarchy. Combination function between hierarchies is a lineal function. The net input of each unit j in hidden or output hierarchy is:

$$I_j = \sum_i w_{ij} o_i + q_j \tag{1}$$

Among the function, w_{ij} is the linking weight from previous unit i to unit j; o_i represents the output of unit i; q_j denotes the offset of unit j, which also presents a threshold here. Appropriate offset help to change the activity of the neuron.

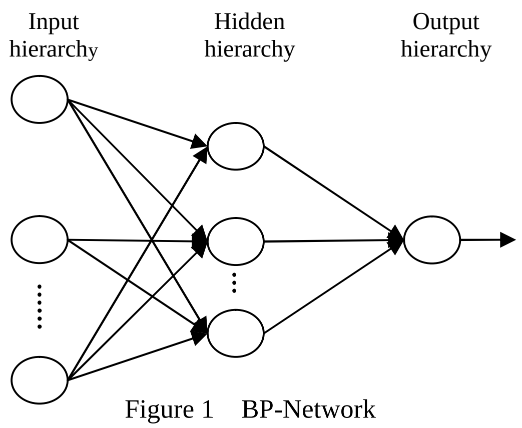


Figure 1 BP-Network

After estimating the net input in each unit, an activating function has to be chosen to estimate the output value. Simoid activating function, the most common type, is chosen here:

$$o_j = \frac{1}{1+e^{-I_j}} \tag{2}$$

In BP-network training, classical BP algorithm is applied, showing a forward transmission of input and a backward transmission of error. Linking weight and offset are modified by calculating the error between the neural network predicted and the value of sample. If linking weight and offset end up in convergence after a number of iterations, the learning process stops; otherwise re-training or adjustment of the network structure is needed.

4. APPLICATION AND KEY PROBLEMS

SAS, one of world's most famous statistical software, is developed by SAS Institute of the US and used to help realize the process of risk identification in this research. As the technology of data mining thrives these years, SAS institute has developed for this field a tool called SAS/EM（Enterprise Miner）, which then became a best data mining software because of its integrated and powerful statistical function. The outcome of data mining also shows a progress since the software has been worked on as a product and improved. Its user-friendly interface also brings in convenience to the data mining process [6-8].

Before data mining, should be processed for preparation so as to ensure validity and high accuracy:

1) Data Set Division

Over-fitting is considered a very significant problem in neural network training. It can be defined as a process in which a neural-network model fits the current data set too well(this might even mean a transitional fitting) to adapt to a new set flexibly. A common way to avoid over-fitting is to divide the data set into a training set and a valid set so that the training process may be applied to two sets at the same time. Training set is for BP-network training and valid set for evaluation to the model. Taking the average error as a performance measurement of training process, it shows a continuous decrease of average error in training set as the training process goes on while the average error in valid set decreases first and then increases. This can be explained as a consequence of over-fitting of neural-network to training set. In this research, as a principle, the weight value, when the average error of valid set is the lowest, is selected as the final solution.

2）Data Standardization

Usually statistics within a data set is not acquired under a same commensurability but by individual method and unit of measurement, which indicates that different input variables of BP-network may have unbalanced influences on target variables. The unbalancing situation among variables caused by disunified measurement is obviously irrational and it might waken some variables of great significance. In order to avoid any mistake caused, the following method of standard deviation (each variable is divided by the standard deviation of sample) is applied. The standard deviation of variable number k that is denoted as variable X_k is estimated in the following formula:

$$\hat{s}_k = \left(\frac{1}{n} \sum_{i=1}^{n} \left(x_k(i) - m_k \right)^2 \right)^{\frac{1}{2}} \tag{3}$$

m_k, denoting the average of X_k, is estimated through sample average

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^{n} x_k(i) \tag{4}$$

Ultimately the inaccuracy brought by disunified measurement is eliminated by calculating

$$x'_k = \frac{x_k}{\hat{s}_k}. \tag{5}$$

After the two steps above, BP-network training begins.

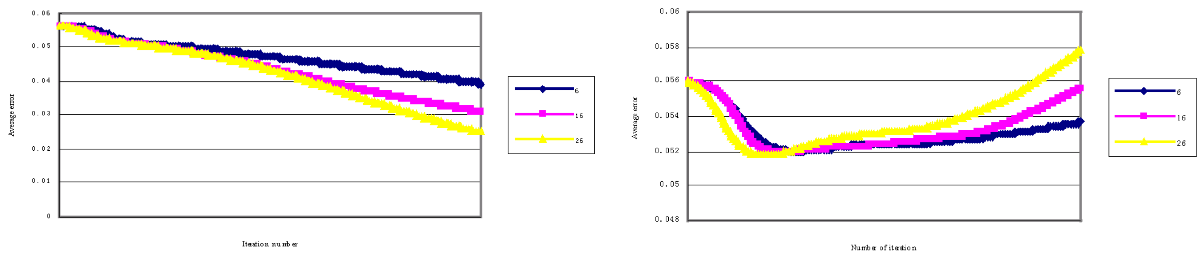
The training can be carried out by the Neural Network Function in SAS/EM.

5. SIMULATION ANALYSIS

Simulation data includes 5822 observation results gained per hour by per-existing enterprises, each of which contains 83 input variables and 1 target variable. 83 input variables are enterprise-related data, including the number of partner enterprises, communicative efficiency within partners, information-safety rate and product-purchasing rate, etc. The target variable denotes the possibility of the risk of product quality. The data set is divided randomly into a training set, which contains 70%(4075 pieces) of observation results and a valid set, which contains 30% (1747 pieces) of observation results. There are many methods for evaluating the model, such as error-classifying rate, average error, profits/loss, etc, but in regards to the nature of risks, wrong prediction of the possibility of a risk, and its deviant receive more attention. Consequently error-classifying rate and average error are chosen to be the evaluating measurement to BP network model.

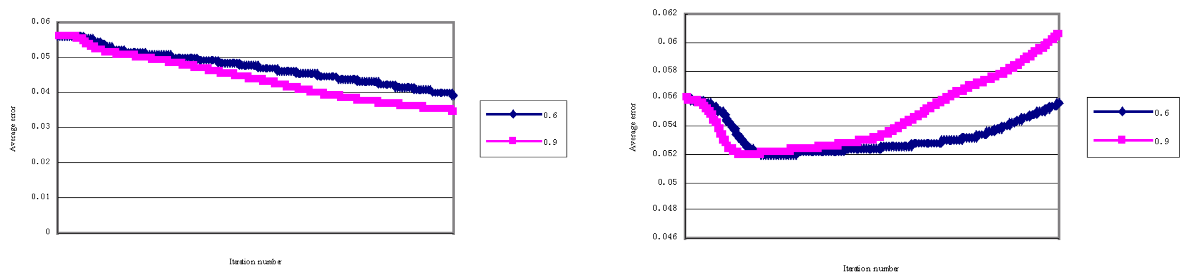
BP network training is a repeating and continuously adjusting process, which needs several parameters to be setting first: the initial value of weight, the number of units in hidden hierarchy, learning rate and numbers of iteration. In these parameters, the initial value of weight almost have no effect on the training result. The numbers of iteration should be proper; too small the training will be not convergence, too big the training will last long time. However, the numbers of units in hidden hierarchy and learning rate effect the accuracy and process of training largely. Figure 2 and 3 show respectively the influence of them on average error in both training and valid set.

The numbers of units in hidden hierarchy is crucial to BP network training because it directly affects the quality of prediction. Insufficient units may result in divergence or dissatisfactory accuracy of prediction while too many units sharply increase training time and arouse more difficulties in learning process. As the training goes further, it is shown in Figure 2(a) that under the same parameters, the increasing number of units in hidden hierarchy results in a smaller model-predicted average error but too many units takes a much longer time for training. Also, the Figure 2(b) shows the numbers of units in hidden hierarchy should be proper. So, in general, 16 is chose as the numbers of units in hidden hierarchy when Figure 2 (a) and (b) is considered synthetically.



(a) Training set (b) Valid set

Figure 2 The influence of the numbers of units in hidden hierarchy on average error



(a) Training set (b) Valid set

Figure 3 The influence of learning rate on average error

Figure 3 shows the influence of learning rate on the training process. Figure 3(a) shows that lower learning rate results in slower training but an excessive learning rate (bigger than 8) may make the training process swing among unsuitable solutions or even results in divergence. Figure 3(b) shows although the average error of the learning rate 0.6 is higher in training set, it is lower in valid set. Hence, 0.6 is chose for learning rate.

Figure 2(b) and 3(b) also show that the average error in valid set decrease first and then increase, which is the “Over-fitting” process. To avoid this, the weight of the smallest average error point in valid set is used as the final training result.
Figure 4 shows the training curve and valid curve illustrating the changing average error in neural network training. The vertical line shows the smallest average error point in valid set, to which the weight corresponds is used as the ultimate result.

Based on the analysis above, the structure of BP-network model and parameters for training are determined as shown in Table 1.

Table 2 lists out the result of model evaluation. It shows that the average error and error-classifying rate are small enough in both training and valid set, which proves the model’s excellent ability in risk identification

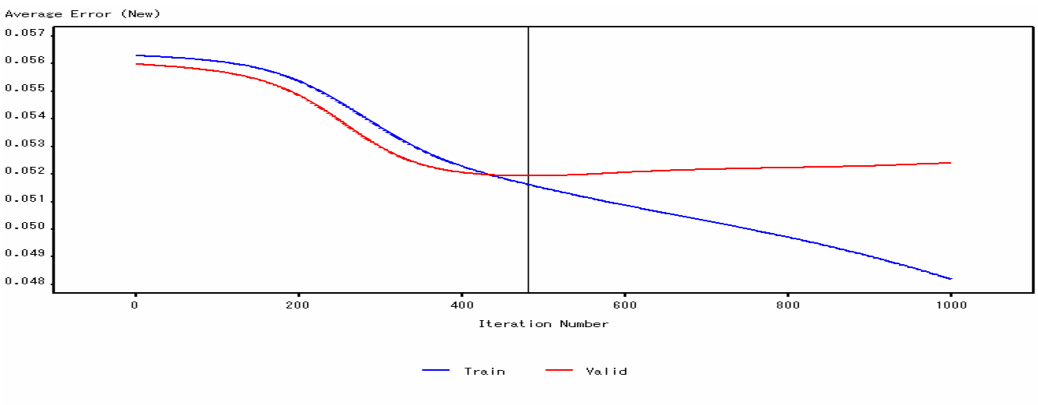


Figure 4 BP training process

Table 1. Structure of BP network model and training parameters

Number of input hierarchic units	83
Number of hidden hierarchic units	16
Number of output hierarchic units	1
Activating function of hidden and output hierarchy	$f(x)=\frac{1}{1+e^{-x}}$
Combination function between	$\sum_i w_{ij}o_i + q_j$
Learning rate	0.6
Iteration number	1000

Table 2 Results of BP identification

Measurement / Data set	Average error	Error-classifying rate
Training set	0.0516240742	0.0598773006
Valid set	0.0519343988	0.0595306239

6. CONCLUSIONS

It’s a brand-new attempt to solve risk identification problem by BP network. Simulation results have proved the method’s validity. Some excellent software for data mining also makes the application of this method possible. With an accurate identification of risks, enterprises are capable of taking more effective measures in their risk management.

ACKNOWLEDGEMENTS

The authors wish to thank the support of the National Natural Science Foundation of China (Project no. 70101006, 60473089, 60003006), Liaoning Province Natural Science Foundation of China (Project no. 20032019, 001015), the Scientific Research Foundation for the Returned Overseas Chinese Scholars by SEM of China, Modern Distance Education Engineering Project by MoE of China, and the Research Grants Council of Hong Kong, China (Project no. PolyU 5167/99E).

REFERENCES

[1] Huang, M., Yang H.M. and Wang X.W., "Genetic algorithm and fuzzy synthetic evaluation based risk programming for virtual enterprise", *Acta Automatica Sinica*, Vol. 30, No. 3, pp449-454, 2004.

[2] Haimes, Y. Y., *Risk Modeling, Assessment, and Management, 2nd ed*, New York: Wiley, 2004.

[3] Bier, V.M., Haimes, Y.Y., Lambert, J.H., Matalas, N.C. and Zimmerman, R., "Survey of approaches for assessing and managing the risk of extremes", *Risk Analysis*, Vol. 19, No. 1, pp 83-94, 1999.

[4] Han, J.W. and Kamber, M., *Data Mining: Concepts and Techniques*, Beijing: China Machine Press, 2001.

[5] Hand, D., Mannila, H. and Smyth, P., *Principles of Data Mining*, Cambridge, Mass., London: MIT Press, 2001.

[6] SAS Institute Inc., "Business intelligence systems and data mining", *SAS Institute White Paper*, Cary, NC: SAS Institute Inc., 1997.

[7] SAS Institute Inc., "Using data mining techniques for fraud detection", *SAS Institute Best Practice Paper in Conjunction with Federal Data Corporation*, Cary, NC: SAS Institute Inc., 1999.

[8] SAS Institute Inc., "Finding the solution to data mining: a map of the features and components of SAS enterprise miner TM software version 3", *SAS Institute White Paper*, Cary, NC: SAS Institute Inc., 2000.