Development of Personal-Credit Evaluation System Using Real-Time Neural Learning Mechanism

Jong U. Choi*
Hong Y. Choi**
Yoon Chung***

〈Abstract〉

Many research results conducted by neural network researchers have claimed that the classification accuracy of neural networks is superior to, or at least equal to that of conventional methods. However, in series of neural network classifications, it was found that the classification accuracy strongly depends on the characteristics of training data set. Even though there are many research reports that the classification accuracy of neural networks can be different, depending on the composition and architecture of the networks, training algorithm, and test data set, very few research addressed the problem of classification accuracy when the basic assumption of data monotonicity is violated.

In this research, development project of automated credit evaluation system is described. The finding was that arrangement of training data is critical to successful implementation of neural training to maintain monotonicity of the data set, for enhancing classification accuracy of neural networks.

+ 본 논문은 1994년 한국학술진흥재단의 공모과제 연구비에 의하여 수행되었음.
* Sangmyung University **LG EDS *** Hankuk University of Foreign Studies

1. Introduction

Neural-based training has been one of the hottest topics in AI research community. The technique has been successfully applied to various fields of classification: character recognition, signal processing, acoustic sound recognition, automobile licence plate recognition, business classification, and so on. For example, Kang et al. [1994a, 1994] applied neural network training to Korean character recognition in automobile licence plate and found that the accuracy rate of the system is much higher than that of an expert system. Even in business classification problem requiring experience of many years, the neural network systems have shown that the classification accuracies of the neural system are better than or at least equal to those of conventional systems or expert systems.

However, in a series of our research on neural network application to credit evaluation system, it was found that the prediction(classification) accuracy strongly depends on the characteristics of the data included in training and testing. In other words, neural network approach to classification(prediction) may not be applicable to every pattern classification problem. In fact, it is quite likely that neural-based training system, though many research results have claimed much better classification accuracy, can be worse than other classification techniques [Weigend and Gershenfeld, 1994].

Originally, this research was initiated to develop an automated credit evaluation system based on neural network training technique. The first experiment with a small set of training data was successful, showing almost 80 percent classification accuracy. However, as the size of training data and testing data increases, the system showed very low classification accuracy. To investigate the relationship between data characteristics and classification accuracy of the neural systems, the generalized learning vector quantization(GLVQ) algorithm[Pal, Bezdek, and Tsao, 1992] was applied to two sets of credit evaluation data. GLVQ algorithm is a modified version of Kohonen's learning vector quantization(LVQ) algorithm[Kohonen, 1992] which provides unbiased classification capability. In the experiments, it was found that the original data sets include too much inconsistent credit data that the network's learning capability, basically relies on monotonicity of the training data, is very low. The much lower classification accuracy is explained in that inconsistency in sample data set works in the reverse direction to monotonicity, which leads to misclassifications, as the network classifies data based on interpolated value.

The successful results of previous

researches might have been obtained only because the network was trained with sound data set, having little number of noisy data or being large enough to control noisy data. Or, the higher performance results might be obtained as "a consequence of having taken a chance" [Wang, 1995, p.556]. As a conclusion, the neural system might not be the best solution, or a panacea to classification problem.

In the following section, credit evaluation system and neural systems built on back propagation algorithm are introduced to help understanding domain problem. In section three, a general concept of GLVQ algorithm is introduced, and experimental results of LVQ algorithm and GLVQ algorithm are described in detail. Section four is reserved for discussion and future research.

2. A Neural-Based Credit Evaluation System

2.1 Credit Evaluation

Credit evaluation is one of the most important and difficult tasks usually assigned to experienced officers in credit card companies, mortgage companies, banks, consumer goods companies and other financial institutes. Traditionally, credit scoring has been the most widely used method[Capon, 1982; Carlet and Catlett, 1987] in which applicant's credit is evaluated by picking up appropriate score corresponding to categories of evaluation value, than by summing up into total credit for thresholding. Recently, various method have been introduced to replace the credit scoring system and to provide more objective and convenient tools : statistical method [Majone, 1968; Apilado, Warner, and Dauten, 1974; Edelstein, 1975; Muchinsky, 1975; Beranek, Taylor, 1976; Borzar, 1978; Capon, 1982]. Induction trees(ID3, C3) [Carter and Catlett, 1987], expert system approach [Dungan, 1982; Dungan and Chandler, 1985; Kastner, Apte, Griesmer, Hong, Karnaugh, Mays, and Tozawa, 1986; Messier and Hansen, 1987; O'Leary, 1987; Pilote and Fillion, 1991; Still, Short, Williams, and Golibersuch, 1991; Talebzadeh, Mandutianu, and Winner, 1995; Sangster, 1995], and rough set approach[Slowinski, and Zopounidis, 1995].

2.2 Neural Network Approaches

Very recently, the neural network researchers have shown that prediction accuracies of the neural network system, that is, the degrees of generalization, are better than or at least equal to those of the statistical methods [Kim, 1992; Odem and Sharda, 1990; Schumann and Lohrbach, 1992]. For example, Kim[1992] conducted a comparative analysis of prediction accuracy with regression analysis, logit, discriminant analysis, and 3-layered neural network. He reported that the prediction accuracy of neural network was 84.5%, much superior to regression analysis(77.84%), logit(75.0%), and discriminant analysis(76.67%). Chung and Tam [1992] compared and reported the performance results of various approaches: analogical reasoning(48%), 2-layer neural network(73%), ID3 with threshold(79.5%), and 2-layer neural network(85.3%).

The experimental results have lead researchers to believe that credit evaluation systems of practical use can be developed using neural network technology and thus have devoted their research efforts to enhance degree of generalization. In determining the degree of generalization, involved are many internal and external factors : the network architecture(number of hidden nodes, input nodes, hidden layers, initial weights, learning rate, momentum, etc.), training algorithm(back propagation, self-organization map, quickprop algorithm, activation function, etc.) and composition of training data set and test data set. Researchers experimented with various architectures by modifying factor values and learning algorithm.

There are some experimental reports on the relationship between training data set and the degree of generalization, with recommendation of the ways to achieve higher generalization capability. Whitley and Karunanithi [1991] proposed a partitional learning strategy in which the training space is divided into a set of subspace according to the data characteristics and then each subspace is trained using a separate network. In the data selection step, emphasized is decision boundaries and the central tendencies of decision regions. In the test of 'two-spiral' problems, they achieved almost 100 percent correctness ratio, using the border patterns. Higher accuracy, with partitioning training data set into separate subsections, may be achievable, because partitioning can increase homogeneity of the training set. Very similar result was obtained in a research which was conducted on bankruptcy data in Korea[Lee, 1995]. The research categorizes credit data into multiple clusters based on learning vector quantization(LVQ) algorithm, and then teaches underlying mechanism using back propagation algorithm. The two-stage model showed a little higher classification accuracy than a simple neural training system. Fu and Chen [1993] investigated the sensitivity of input vectors on generalization capability, and found that the norm of Jacobian matrix measures the sensitivity of the network performance with respect to its vector and that good generalization must imply insensitivity to changes in the input vectors.

2.3 Automated Credit Evaluation System

This research was initiated to develop an integrated on-line credit evaluation system which would monitor system performance and enhance prediction accuracy through constant feedbacking customer's credit data. Especially, the neural network mechanism was adopted as a credit evaluating processor in this research. The neural network could predict the output values by nonlinear mapping through the hidden layer, even though it didn't know the direct relations between the input values and the output values.

E LTD. is one of leading companies in the Korea fashion business. A credit card system of this company is adopted to achieve 'Big Share' in fashion market. Recently, the number of card holders of this company has reached 180,000 and every month the number of overdue or delinquent credits reached 3,500 cases. Such delinquent customers inflict a serious loss to the company and thus the company had to devise a measure to solve the financial problem caused by continuously accumulated bad debts. One of the ideas popped up was to develop an automated credit evaluation system which continuously monitors evaluation system's performance and then can enhance the prediction accuracy through learning from customer's credit data.

S LTD. is an another leading credit card company which holds a big share in Korea market. The customer's behaviour of E LTD and S LTD is almost the same, in terms of credit standing. Even though the two companies employ different application variables, accordingly different application forms, the contents were almost the same.

The neural network training system, as usual in the back propagation systems, consists of three layers : input layer, hidden layer and output layer. From the customer's records, the eight variables which were believed to have a strong influence on customer's credit were derived as 'credit factors' : age, sex, marital status, occupation, organization, job position, residential condition, residential area. Selection of input variables for the system's training, that is, selection of critical factors significantly influencing on customer's creditability should be determined by consideration of customer's behaviour, social custom, and statistics. In this sense, the factors included in the current system reflect many features of Korean customers and social practices, and thus factors included in the current system might be much different from factors included in the system developed in other countries. For example, in the study of American loan application, occupation, length of employment, marital status, race and income level are important considerations[Capon, 1982], but work place dose not have a significant impact on credit evaluation. In contrast, work place might be the most important factor in determining an individual's credit status in Korea. Also, residential area might be very important factor.

According to the number of overdue payment, 'credit status' was divided into two status such as 'good' and 'bad'. When the customer's payment is not overdue or the number of overdue payment is less than 3 months, the customer's status was classified into 'good'. When the number of overdue payment exceeded 3 months, the customer was classified into 'bad' credit status. 2.4 Test Results In the beginning stage of this research, back propagation algorithm was employed. Many experiments with the back propagation algorithm showed that the system with more than 40 training data would not converge into an equilibrium state within a reasonable time, and thus another efficient algorithm was needed. Later, employed was the quickprop algorithm, an advanced form of back propagation, as a learning mechanism. 'Quickprop' algorithm suggested by Fahlman[1988] is well-known for speeding up convergence by jumping out directly the parabolic error space to the minimum point of the parabola. In this algorithm, the error defined as $\partial E/\partial w(t-1)$ is kept and then, for each weight, the weight change measured by the difference between current weight slope and previous weight slope is used for determining a parabola.

As predicted and assured by researchers [Fahlman, 1988], the quickprop algorithm effectively and quickly reduced total error, and thereby enabled the system to reach convergence in a reasonable time limit. Even though the epoch number to reach convergence state was quite different, there was no difference in classification accuracy between back propagation algorithm and quick-prop algorithm. As shown in Table-1, the classification accuracy of the back propagation was 65% and the accuracy of quick-prop was measured around 60%, a slightly lower than the back propagation algorithm.

When the number of training data set increases to 100 (50 bad creditors and another 50 good creditors), the back propagation system did not stop running. That is why the research, in back propagation learning system, could not extend testing the generalization capability beyond 80 cases. To the contrary, the quickprop algorithm with 100 training data easily reached the convergence state at the epoch of 267 and showed a little enhanced prediction capability, 67%. The same test results were obtained when the test of classification accuracy were conducted on S LTD data set. As shown in Table-2, the classification accuracy was less than 60%, which is too low for field application.

3. Generalized Learning Vector Quantization (GLVQ)

Disappointed with the test results, the research team attempted to classify credit data into two groups using unsupervised clustering algorithm, rather than using supervised back propagation algorithm. For the purpose, generalized learning vector quantization algorithm was employed.

3.1 Vector Quantization Clustering

Vector quantization(VQ) is defined as a technique which "searches for small but representative set of prototypes, which we can then use to match sample patterns with nearest neighbor techniques," [Kong and Kosko, 1992, p.1]. It is a technique developed for solving data encoding problem leading to minimization of reconstruction error in data compression and decompression [Ritter, Martinnetz, and Schulten, 1992]. In searching for an efficient data decompression skill, one seeks to describe as faithfully as possible "the distribution of data points in a high-dimensional space, using only a space of lower dimension" [Ritter, Martinnetz, and Schulten, 1992, p.238]. That is, the most efficient projection of original data onto lower dimensional planes can yield the smallest project error. The technique has been modified and successfully applied and to various fields : image classification [Cannon, Dave and Bezdek, 1986], phoneme signal processing[Kong and Kosko, 1992], and travelling sales person(TSP) problem [Rose, Gurewitz, and Geoffrey, 1993].

Clustering through VQ is accomplished by partitioning the patterns $x \in R^n$ into k decision classes $\{D_i\} \in R^m$, the prototypes or reference vectors:

$$R^n = \bigcup D_j, \text{ and } D_i \cap D_j = \emptyset \text{ for } i \neq j.$$
$$x \in D_1 \quad \text{if } d(x, s_1) < d(x, s_2)$$
$$x \in D_2 \quad \text{if } d(x, s_1) > d(x, s_2).$$

where $d(x_i, s_i)$ is defined as the distance measure between the pattern xi and prototype s_i. S_1 and s_2 are prototypes belonging to the decision classes D1 and D2, respectively.

The VQ system attempts to find appropriate decision class ($D_1, D_2, \ldots, D_k$) and centroids ($\overline{S1}, \overline{S2}, .. \overline{Sk}$), from the patterns (X1, X2, ..., Xp), the pattern belongs to. In the view of data compression, the patterns x_i will not completely vary, but rather will be correlated to next patterns. Thus, the essential problem of VQ technique is to find mapping functions from the patterns to hidden variables $r_1, r_2, \ldots, r_m$, for $M < P$, with minimal variances. The variables r_i provide a more economical description of the observed phenomenon. In the linear discriminant functions [Kong and Kosko, 1992; Kosko, 1991], the function behaves as a separating hyperplane in the pattern space R^n, that is, setting up K-dimensional hyperplane lying in the N-dimensional data space. The variables ri can account for the total data variation. However, if the actual distribution of data points is deviated from the hyperplane, the description resulting from a projection on the principal axes of the distribution will be worse [Ritter, Martinetz, and Schulten, 1992; Cichocki and Unbehauen, 1993]. To overcome this problem, the linear principal axes or hyperplanes are replaced by curved surfaces, which may provide a better description of nonlinear data distributions. This can be interpreted geometrically as "a minimization of the mean-squared perpendicular distance $d(x, s_i)^2$ between the data points and the hyperplane" [Ritter, Martinetz, and Schulten, 1993, p.247].

Learning vector quantization (LVQ), suggested by Kohonen, is considered as an approximation procedure for the computation of principal curves, surfaces, or higher-dimensional principal manifolds [Ritter, Martinetz, and Schulten, 1993]. The LVQ system tries to discover cluster substructure hidden in unlabeled N-dimensional data and extract M-dimensional features. The prototypes $S=\{S_1, S_2, ..., S_k\}$ are a array of unknown cluster centers $S_i \in R^m$ for $1 \leq i \leq k$. In LVQ, learning refers to finding values for the $\{S_i\}$ [Pal, Bezdek, and Tsao, 1993]. When an input vector X_i is submitted to the system, the distance between the input vector and prototypes $d(X, S_i)$ is calculated and then the prototype with the shortest distance becomes a winner. The next step is to update the centroid of the prototype using update rules. The typical LVQ rule of finding the winner node and update is as following:

$$\| X_k - S_{i,t-i} \| = \min_{1 \leq i \leq k} \{ \| X_k - S_{j,t-1} \| \} \text{ for finding}$$

$$S_{i,t} = S_{j,t-1} + \alpha(X_k - S_{j,t-1}) \text{ for updating.}$$

Although the LVQ algorithm has some nice theoretical foundation, it suffers from a serious problem : initialization problem. As the initial position of centroid Si,0 have too strong influence on subsequential position updates, especially when they are outside the convex hull of the input data [Kang, Hwang and Yoo, 1994], it may not produce any meaningful clusters[Pal, Bezdek, and Tsao, 1993]. Also, as the winner node only update its position, the result of clustering might be biased by the gravitational force of winners. To overcome these problems, Pal, Bezdek and Tsao [1993] suggested the GLVQ algorithm which updates either all the centroids of prototypes or none, for each new input vector. When there is a perfect match to the winner node, no node is updated.

The updates rule of GLVQ is

$$S_{i,t} = S_{i,t-i} + \alpha(X_k - S_{i,t-i}) \frac{D^2 - D + \| X_k - S_{i,t-i} \|^2}{D^2}$$

$$S_{i,t} = S_{i,t-i} + \alpha(X_k - S_{i,t-i}) \frac{\| X_k - S_{i,t-i} \|^2}{D^2} \quad (r \neq i)$$

where i is the best matching node,

$D = \sum_{r=1}^{c} \| X_k - V_r \|^2 . k=1, 2,.. n; r=1, 2,$

...k and t is time

The simple modification of updating rule in GLVQ algorithm makes a significant difference in clustering. Most importantly, LVQ clustering is seriously affected by the initial data, while GLVQ algorithm provides more robustness to the location of initial data. The centroids of the clustering network in LVQ are at first located at a random initial places and subsequentially moves into the direction attracting input nodes through 1-nearest neighbor prototype principle. That is, the size of centroids movement is affected by a single input value. As a result, when the initial locations of clustering centroids are located far away from each other, it is very difficult to drag the centroids out of the convex hull into n-partitioned planes. Because of this, the initial Alpha value should be very large enough to move the clustering centroids into energy-minimum states, although the large number of network iterations has a slight impact on the clustering speed.

To the contrary, all the centroids of the GLVQ network is affected by input values, in proportion to the distance between the centroid and input value. Thus, the centroids moves into the direction of mass gravitational forces of all the input values. As a result, reasonable clustering can be obtained very quickly. The significant difference between LVQ algorithm and GLVQ algorithm come from the principle that a single input value does affect movement of all centroids in GLVQ, but a single nearest centroid in LVQ.

3.4 Experiments with GLVQ

In the experiments with GLVQ, two sets of credit evaluation data were employed for performance comparison. In the first experiment, as shown in the figures below, the system showed very low classification accuracy. As the system was developed based on GLVQ algorithm, the system was not sensitive to the modification of learning parameter, alpha and number of iterations. As shown in Figure-1 and Figure-2 of E LTD case, the system partitioned 120 data consisting of 60 'bad' customer and another 60 'good' customer data into two clusters: cluster-1 and cluster-2. The data numbered 0 to 59 should be group-1, while the data numbered 60 to 119 should be in group-2. In other words, the data numbered 0 to 59 and the data numbered 60 to 119 should be not be in the same group to be cohesive. But, the clustering result is that the cluster-1 has 74 units of data including 33 data units from one group and other 41 data units from another group.

This means that clustering the credit data is not so meaningful for real world application, indicating that the data included in clustering does not have any meaningful relationship with each other in the same group. In other words, neural classification cannot impose any meaningful decision rules on clustered data. As shown in Table-4 the same thing was observed in S LTD case. Figure-1 and Figure-2 show that inconsistencies between natural clustering of GLVQ algorithm and original credit clustering exist in case of S LTD data set. In Figure-1 and Figure-2, two different data set of the homogeneous category belong to the same cluster while the other two data sets of the same category belong to the other cluster. For example, the shape of square and circle should not be in the same cluster. Irrespectively of the data characteristics, the credit data are randomly scattered. It implies that the natural clustering through GLVQ algorithm may not produce any meaningful classification result.

4. Discussion and Future Research

Unexpectedly low classification accuracy may be explained by the inconsistency of the sample data. Because neural network training basically is nothing but a monotonic mapping between input values and out values, inconsistency of the sample data may destroy the interpolation capability. Assumption of the monotonicity is simply stated as

$$[\,[\ x_1 > x_2] \Leftrightarrow [\ \Phi(X \| x_j = x_1)) > \Phi(X \| x_j =:$$

for a monotonically increasing function. For a two-class classification problem in which the total score of hidden nodes multiplied by synaptic weight values is bi-sected into two hyperplanes, if the larger the value the input variable, the larger the classification score output value becomes, then the output variable is monotonic in the values of input variable[Archer and Wang, 1993]. For example, higher income level supports higher credit-worthiness in a non-linear fashion, even though there is no direct and linear mapping between income level and credit score.

However, the assumption of monotonicity between credit-worthinessand input values is violated, because of the inconsistencies in the sample data, in other words many conflicting cases were identified. For example, a department head of a business company usually earns better salary than other employees in the department and thus, the head is supposed to be much better in credit standing than others in the department. However, in the review of raw data it was found that the credit standing of employees is not correlated with income level of the employee. Even CEOs of business companies, though CEOs of large group companies are exceptional, were as bad as young undergraduates with less than one year's job experience.

When the assumption of monotonicity is violated, the classification capability can be lowered by the interpolation of the network. For example, if the first sample case is

$$[[x_1 > x_2] \Leftrightarrow [\Phi(X \| x_j = x_1)) > \Phi(X \| x_j = :$$

and the next sample case is

$$[[x_1 < x_2] \Leftrightarrow [\Phi(X \| x_j = x_1)) > \Phi(X \| x_j =$$

then, the network, if other things being equal, converges into the final state of

$$\Phi(X \| x_j = x_1) \cong \Phi(X \| X_j = x_2).$$

That is, the boundary of the two clusters will be blurred, which lowers the classification capability of the network.

The conflict of credit data collected in the research can be indirectly explained by company's policy and social customs. First of all, the first priority of Korean companies is not in achieving higher profit, but in taking bigger market share. For the reason, the companies are very aggressive in expanding their own market in issuing credit cards: no review of the credit standing of the applicants. As a result, the credit card holders' information is not accurate and bad credit cases are rapidly increased. In other words, with the rapid increase of sales volume and credit market in Korea, many business companies have not imposed any restriction on credit card applicants. This is because, different from American companies with hundreds of years of experienced in financial market, Korean Companies pursue the goal of market penetration and market expansion through granting credit cards to any applicant without any scanning efforts. Therefore, the training data set obtained from credit companies does not have strong consistency, or statistical trends. That is, why any other prediction method cannot earn much better results than the results achieved by this research.

Neural network can be an efficient method for achieving prediction accuracy based on training of collected past data set. Generally, as claimed by neural network researchers, the network training algorithms is better than, or at least equal to conventional prediction methods such as regression methods, expert system approach, or machine learning mechanism.

However, we believe that the better performance of the network training systems can be achieved because the data set included in training and testing is noisy-low set. In other words, even though the prediction power of neural networks is excellent, it strongly depends on the data characteristics. In a series of experiments with noisy data set for developing automated credit evaluation systems, it was found that the non-linear projection of training data set into hyperplanes cannot completely eliminate uncertainty included in white noise. As a conclusion, very high accuracy cannot be achieved with noisy data, and thus the neural network learning systems should be carefully designed before system development.

〈References〉

Apilado, V. P., D. C. Warner and J. J. Dauten, "Evaluative Techniques in Computer Finance - Experimental Results and Policy Implications for Financial Institutions", *Journal of Financial and Quantitative Analysis*, March, 1974

Archer, N.P., and S.Wang, "Application of the Back Propagation Neural Network Algorithm with Monotonicity Constraints for Two-Group Classification Problems," *Decision Sciences*, Vol.24, No.1, Jan. 1993, pp.60-75,.

Beranek, W. and W. Taylor, "Credit - Scoring Models and the Cut-off Point A Simplification", *Decision Sciences*, vol. 7, 1976.

Borzar, G. E., "Competition between Banks and Finance Companies : A Cross Section Study of Personal Loan Debtors", *The Journal of Finance*, vol. XXXII, No. 1, March, 1978

Cannon, R.L., J.V. Dave, and J.E. Bezdek, "Efficient Implementation of the Fuzzy c-Means Clustering Algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence* (Vol. 8, No.2), March, 1986.

Capon, N., "Credit Scoring Systems : A Critical Analysis", *Journal of Marketing*, (Vol.46), Spring 1982, pp.82-91.

Carter, C., and J. Catlett, "Assessing Credit Card Applications using Machine Learning", *IEEE EXPERT*, Fall 1987, pp.71-79.

Chung, H.M., and K.Tam, "A Comparative Analysis of Inductive-Learning Algorithm," *Intelligent Systems in Accounting, Finance, and Management*, Vol.2, 1991, pp.812-845.

Dungan, C. W., A Model of an Audit Judgment in the Form of an Expert System, Ph.d, Dissertation, Dept. of Accounting, University of Illinois at Urbana-Champaign (December, 1982).

Dungan, C. W. and J. S Chandlers, "Auditor : A microcomputer-based expert system to support auditors in the field", *Expert System* (October 1985), pp.210-224.

Edelstein, R. H., "Improving The Selection of Credit Risks : An Analysis of a Commercial Bank Minority Lending Program", *The Journal of Finance*, vol. XXX, No. 1, March, 1975

Fahlman, S. E., "Faster-Learning Variations on Back-Propagation: An Empirical Study," in Touietzky D., T. Sejnowski, and G.Hinton (eds.) *Proceedings of the '88 Connectionist Models Summer School*, CMU, 1988.

Fu, L., and T. Chen, "Sensitivity Analysis for Input Vector in Multilayer Feedforward Neural Networks," *Proceedings of IJCNN*, 1993, pp.215-218.

Hathaway, R.J., and J.C. Bezdek, "Switching Regression Models and Fuzzy Clustering, "*IEEE Trans. on Fuzzy Systems* (Vol.1, No.3), August 1993, pp.195-204.

Hathaway, R.J., and J.C. Bezdek, "Local Convergence of the Fuzzy C-means Algorithms", *Pattern Recognition* (Vol.19, No.6), 1986, pp.477-480.

Kang, B.H., J.H. Yoo, and S.J. Kang, "Speeding up Generalized Learning Vector Quantization," will be in *the Proceedings of ICONIP(International Conference on Neural Information Processing)*, Seoul, October, 1994a.

Kang, B.H., D.S. Hwang, and J.H. Yoo, "Square Error Clustering Scheme and Clustering Networks," in *the Proceedings of Izuka-94 International Conference on Neural Networks, Fuzzy and Soft Computing*, Izuka, Japan, 1994b.

Kastner, J., C. Apte, J. Griesmer, S. J. Hong, M. Karnaugh, E. Mays, and O. Tozawa, "A Knowledge-Based Consultant for Financial Marketing," *The AI Magazine*, Winter 1986, pp.71-79.

Kim, J.W., A comparative analysis of rule based, neural networks and statistical classification for the bond rating problem, Ph.d, Dissertation, Dept. of Information Systems, Virginia Commonwealth University Richmond, Virginia (April, 1992).

Kohonen, T., "The Self-Organizing Map," in Lau, C(eds), *IEEE Neural Networks*, IEEE (New York), 1992, pp.74-90.

Kong, S., and B. Kosko, "Differential Competitive Learning for Phoneme Recognition," in Kosk, B. (eds) *Neural Networks for Signal Processing*, Prentice-Hall (New Jersey), pp. 36-, 1992.

Majone, G., "Classification and Discrimination in The Analysis of Credit Risks : II", Management Sciences Research Report No. 128, Carnegie-Mellon University, March, 1968

Messier, and J. V. Hansen, "Expert Systems in Auditing : The State of the Art", *Auditing : A Journal of Practice & Theory*, Vol.7, No.1, Fall 1987, pp.94-105.

Muchinsky, P. M., "Consumer Installment Credit Risk : A Need for Criterion Refinement and Validation", *Journal of Applied Psychology*, vol. 60, No. 1, 1975.

Noel, C., "Credit Scoring Systems: A Critical Analysis", *Journal of Marketing*, (Vol. 46), Spring 1982, pp.82-91.

Odom, M.D., and R. Sharda, "A Neural Network Model for Bankruptcy Prediction," *Proceedings of IJCNN 1990*, Vol.II, 1990, pp.II-163-168.

O'Leary, D. E., "The Use of Artificial Intelligence in Accounting.", in Silverman B. G. (ed.) *Expert System for Business* (Addison-Wesley, 1987), pp.82-98.

Pal, N.R., J.C. Bezdek, E.C, Tsao, "Generalized Clustering Networks and Kohonen's Self-Organizing Scheme," *IEEE Trans. on Neural Networks* (Vol.4, No.4), July 1993, pp.549-557.

Pilote, M., and M. Fillion, "Automated Underwriting at Continental Canada: Results and Major Obstacles, in *the Proceedings of the First Int'l Conference on AI Applications on Wall Street*, New York, New York, Oct. 1991, pp.168-173.

Ritter, H., T. Martinetz, and K. Schulten, *Neural Computation and Self-Organizing Maps*, Addison-Wesley(Reading, Massachusetts), 1992.

Rose, K., E. Gurewitz, and G.C. Fox, "Constrained Clustering as an Optimization Method," *IEEE Trans. on Pattern Analysis and Machine Intelligence* (Vol.15, No.8), August 1993, pp.785-794.

Sangster, A. "The Bank of Scotland's Lending Adviser Expert System, COMPASS," *The Proceedings of the 11th Conference on Artificial Intelligence for Applications*, Feb. 1995, LA, CA, 1995, pp.24-30.

Slowinski, R., and C. Zopounidis, "Application of the Rough Set Approach to Evaluation of Bankruptcy Risk," *Intelligent Systems in Accounting, Finance, and Management*, Vol.4, 1995, pp.27-41.

Still, D.D., R.S. Short, D.B. Williams, and D.C. Golibersuch, "Guidelines: A Production Rule Based Implementation of Mortgage Insurance Underwriting Policy," *The Proceedings of the First Inte'l Conference on AI Applications on Wall Street*, New York, New York, Oct. 1991, pp.174-177.

Surkan, A. J., and J. C. Singleton, "Neural Networks for Bond Rating Improved by Multiple Hidden Layers," *Proceedings of IJCNN* 1990, Vol.II, pp.II-157-162.

Surkan, A. J., and X.Ying, "Bond Rating Formulas Derived Through Simplifying A Trained Neural Network," *Proceedings of IJCNN* 1991, 1991, Vol.II, pp.1566-1570.

Talebzadeh, H., S. Mandutianu, and C.F. Winner, "Countrywide Loan-Underwiring Expert System," *AI Magazine*, Spring 1995, pp.51-64.

Tam, K.Y., "Applying Conceptual Clustering to Knowledge-bases Construction, " *Decision Support Systems* (Vol.10), 1993, pp.173-198.

Tam, K.Y., and M.Y. Kiang, "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions," *Management Science*, Vol.38, No.7, July 1992, pp.926-947.

Wan, E. A., "Neural Networks Classification: A Bayesian Interpretation," *IEEE Transactions on Neural Networks* (Vol.1, No.4), Dec. 1990, pp.303-305.

Wang, S., "The Unpredictability of Standard Back Propagation Neural Networks in Classification Applications," *Management Science*, Vol.41, No.3, March 1995, pp.555-559.

Weigend, A. S., and N.A. Gershenfeld (eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison Wesley, 1994.

Whitley, D., and N. Karunanithi, "Generalization in Feed Forward Neural Networks," *Proceedings of IJCNN*, July 8-12, Seattle, Washington, 1991, pp. II-77 - II-82.