A Review of Artificial Intelligence Models in Business Classification

In-goo Han*, Young-sig Kwon**, Hong-kyu Jo*

ABSTRACTS

Business researchers have traditionally used statistical techniques for classification. In late 1980's, inductive learning started to be used for business classification. Recently, neural network began to be applied for business classification. This study reviews the business classification studies, identifies a neural network approach as the most powerful classification tool, and discusses the problems and issues in neural network applications.

I. INTRODUCTION

Classification refers to separating distinct sets of objects or observations and allocating new objects or observations into previously defined groups. Classification needs an algorithm to separate and allocate objects or observations. This algorithm is called a classification technique. The ultimate goal of a classification method is to provide the relevant outcome or to replicate the expert's judgment. The relative performance of different classification techniques may depend on data conditions.

Areas of business classification research include bankruptcy prediction, accounting method choice, audit opinion decisions, credit rating prediction, and bank loan classification.

Business researchers have traditionally used statistical techniques for classification. Statistical methods frequently employed in business classification research include multivariate discriminant analysis (MDA), logit, and probit methods.

* Department of Management Information Systems, Korea Advanced Institute of Science and Technology, Seoul 130-012, Korea
** Department of Industrial Engineering, Dongguk University, Seoul 100-715, Korea

Recently, inductive learning, a subfield of artificial intelligence (AI), began to be applied to the classification research in business. Examples include stock market prediction (Braun and Chandler, 1987) and scholarship and fellowship grant cases (Garrison and Michaelsen, 1989). Inductive learning uses a data set of examples and determines a relationship between these examples via inductive inference. The induced rules can then be used to predict outcomes or to replicate judgments. Various inductive learning method have been developed, including ID3 which has been used most widely.

Quite recently, the artificial neural network (neural network for brevity) approach, another field of AI, started to be applied to the classification research in business. Neural network model is based on how human brain cells and their interactions are able to perform complex tasks. A neural network model consists of many processing elements (PE). These PEs are grouped into linear arrays called layers. A neural network model has an input layer and an output layer, and may or may not have hidden layers. Each PE calculates the linear combination of input signals and applies the transfer function to calculate the output value. Examples using the neural network approach include accounting inventory method choice (Liang, Chandler, Han, and Roan, 1992), bankruptcy prediction (Tam and Kiang, 1992 ; Jo, 1994), and bond rating (Surkan and Singleton, 1990).

Neural network can handle noisy data and adaptively adjust the model parameter when encountered new environments, decrementally retaining the importance of past data. Neural network also contains nonparametric advantages like ID3. While neural network lack explanatory capability which can explain the relative importance of input variable and how the output variable is attained, the advantage of neural network outweighs the drawback, in case of business classification where the prediction accuracy is more important than the explanation.

A review of literature indicates that neural network model is the most powerful classification tool. In this paper, we focus on the review of the business classification studies applying artificial intelligence classification techniques, followed by the design and implementation issues of neural network approach and the integration issue with other approach.

II. Performance of Classification Techniques

2-1. STATISTICAL CLASSIFICATION TECHNIQUES : MDA, probit/logit, and recursive partitioning

The statistical methods frequently employed in business classification research are MDA, logit, and probit. Discriminant analysis, evolved as a variant of univariate analysis of variance, is generally concerned with comparisons of the distribution of one or more variables across different groups or populations (Altman et al., 1981). Discriminant analysis can also be used to test for differences in variable mean vectors and/or covariance structures across groups.

Discriminant analysis assumes that the explanatory variables are distributed with a multivariate normal distribution. There are two types of MDA according to functional form : LDA (linear discriminant analysis) and QDA (quadratic discriminant analysis). LDA has been far more frequently used than QDA in business classification research.

Two years after Beaver's (1966) univariate approach, Altman (1968) introduced the LDA to bankruptcy prediction. As expected, the accuracy achieved by Altman's multivariate approach was higher than that by Beaver's univariate approach. Studies using LDA include Altman, Haldeman, and Narayanan (1977) and Mutchler (1985). LDA assumes that the variance-covariance matrices of groups are the same. If the data violate normality and independence assumptions of the classical linear regression or discriminant approaches, the usefulness of discriminant analysis will be reduced. Discrimant analysis may not also be appropriate for ordinal classification task such as bond rating, because classification by discriminant analysis doesn't specify ordinal relationship.

Probit and logit models evolved from the traditional regression model. Probit and logit (generally qualitative response (QR) models) postulate a line of causality running from exogenously determined independent variables and stochastic errors to a discrete dependent variable. QR models are used to estimate the conditional probability of an event given explanatory variables.

The probit model assumes that the conditional probability of category membership follows the standard normal distribution. The underlying assumption of the normal distribution for the probit or MDA model is sometimes justified by applying the central limit theorem and considering the large number of factors that may influence the probabilities. Studies using the probit method include Kaplan and Urwitz (1979), Zmijewski and Hagerman (1981), and Zmijewski (1984).

The logit model is identical to the probit model except that the conditional probability of category membership follows the standard logistic distribution. The distribution and density functions of the standard logistic are very similar in shape to the standard normal distribution. It is known that the use of probit or logit makes little difference except when data are heavily concentrated in the tails. Studies using the logit method include Ohlson (1980) and Gentry, Newbold, and Whitford (1985).

For parametric statistical method discussed above mentioned, violation of the assumption may reduce the power of test, which could be a serious problem. Recursive partitioning, a recently developed Bayesian statistical technique, was applied to commercial bank loan classifications by Marais, Pattel, and Wolfson (1984) and financial distress prediction by Frydman, Altman, and Kao (1985). The recursive partitioning procedure, as an alternative to alleviate the statistical problems encountered as in DA, probit and logit, is a nonparametric method using recursive binary partitioning of the explanatory variables to classify observations. This method provides a treelike structure for classifying observations. Recursive partitioning selects and partitions the independent variable or linear combination of variables that most improves the homogeneity of class assignments based on misclassification costs and prior probabilities.

2-2. Artificial Intelligence Techniques : ID3, AQ, genetic algorithm, and neural network

Learning is defined as changes in a system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the population more efficiently and effectively the next time (Simon, 1983). AI researchers have devoted much effort to implanting learning capabilities in computer software. The computer modeling of learning processes constitutes the field of machine learning (Carbonell, Michalski, and Mitchell, 1983). The ability of human beings to make generalizations from scattered observations or to discover structures in collections of observations has been a long-standing issue of interest. The understanding of this process, called inductive inference, may be the key to an improvement of methods by which the computer can acquire knowledge. Inductive learning, a subfield of machine learning, is viewed as a heuristic search through a space of symbolic descriptions, generated by an application of various inference rules to the observational statements (Michalski, 1983).

Various inductive learning algorithms have been developed in AI domains, including analog concept learning (ACL) (Hunt et al.,1966), genetic algorithm (Holland, 1975), ID3 (Quinlan, 1979), AQ (Michalski et al., 1980), and Neural Network, etc.

ID3 has been the most frequently used in inductive learning applications. ID3 was developed for concepts which are qualitative in nature. Hence, it is expected that the ID3 method can deal with qualitative variables better than statistical methods. ID3 is a nonparametric method which may be preferred to statistical techniques based on specific assumptions when the data severely violate those assumptions. The ID3 method generates a decision tree, where each leaf node contains examples which are of the same class, based on Shannon's information theory.

ID3 have limitations in several ways. Messier and Hansen (1988) discussed three sources of limitations : the difficulty to apply to very large problem domains, the potential error introduced into a production system from omitting important instances or diagnostic attributes, and conflicting instances.

AQ is similar to ID3 in that it generates production rules or some form of text description using the symbolic concept acquisition approach. The language used by AQ is an extended version of Predicate Calculus, Annotated Predicate Calculus (APC). It extends Predicate Calculus by adding new forms and concepts to increase the expressiveness of the language. AQ is different from ID3 in such a way that in ID3, a decision rule can't involve any variables and ID3 is closer to statistical techniques in its adoption of the entropy notion while AQ is geared towards set operation (Chung & Tam, 1992).

Genetic algorithm (GA), developed by J. Holland (1975) who was inspired by biological evolution, is another possible method for inductively generating a set of production rules. Maintaining a population of candidate solutions to a given problem as chromosome represented in a form of bit strings, new solution is generated by biased reproduction, idealized recombination and mutation operators. GA is appropriate when problem requires good, but not optimal solution, acceptable performance measure and representation are available. Computational burden has been the major bottleneck to the practical usage of GA.

Scientists began to apply inductive learning to classification in early 1970's. Buchanan et al. (1976) developed a computer program called Meta-DENDRAL that assisted chemists in the discovery of rules from empirical data on mass-spectra. Meta-DENDRAL provided qualitative explanations of the characteristics of fragmentations and rearrangements among a set of molecules. Buchanan et al. showed that the inductive program was capable of rationalizing the mass-spectra data and suggested that it offered a powerful and useful complement to traditional methods for finding a structural relationship from spectral data.

It has been only a few years since business researchers began to apply the inductive learning approach to business domains. Braun and Chandler (1987) applied the ID3 to predict stock market movements. They suggested that inductive learning can be beneficial in developing a decision support system for market analysts or in developing their own decision processes.

Greene (1987) studied the classification performance between the genetic algorithm (GA), ID3 and logit models. His finding is that both the GA and the logit model produced clearly better results than that of ID3, but GA and logit models performed almost identically.

Garrison and Michaelsen (1989) applied ID3 to analyze Tax Court cases in determining scholarship or fellowship grant status. They chose fourteen attributes which are all nominal. The results using the holdout technique indicated that the rules from ID3 were more accurate than MDA or logit at predicting the outcome of Tax Court decisions. The authors suggested that ID3 may be more applicable to qualitative measurement situations such as tax studies than statistical techniques such as MDA or logit, which are based on interval parametric statistical theory.

Lee and Oh (1990) compared recursive partitioning (RP) with analog concept learning (ACL), a generalization of ID3, applying two methods to the bankruptcy prediction to see the conditions under which one of them works better than the other. Their results show that if prior probabilities are very biased, recursive partitioning performs worse than ACL which is independent of prior probabilities, suggesting recursive partitioning is recommendable where the estimates of prior probabilities are relatively accurate and ACL is useful where the estimation of prior probability is hard or the estimate is unreliable.

Cronan et al. (1991) compared the classification rates of recursive partitioning with those of ID3, MDA, Logit and Probit on mortgage, commercial and consumer lending problems. The empirical results show the recursive partitioning which utilized few variables provided notably higher classification accuracy than those of other methods, including ID3 which utilized many variables.

Quite recently, the neural network approach started to be applied to business classification. A neural network system consists of many simple interconnected PEs. Each PE calculates the weight of input value and applies the transfer function to generate the output value. The transfer function remains unchanged, but the weights for the linear combination can dynamically be adjusted to produce a desirable output. The PEs are grouped into linear arrays called layers. A neural network model has an input layer and an output layer, and may or may not have hidden layers.

In most business classification domain, back-propagation neural network has been mainly used. The application of back-propagation neural network involves two phases. In the forward stage, the input is propagated forward through the network to compute the output value and the error. In the backward stage, the recursive computation of the error term is performed in a backward direction to modify weights and correct errors. The back-propagation algorithm converges very slowly and may get stuck at local minimum.

The studies applying neural network models show that the neural network approach is very powerful tool for business applications. Dutta and Shekhar (1988) showed that predictive accuracy of the neural network outperformed that of regression in bond rating, although their neural network implementation might not be adequate.

Surkan and Singleton (1990) applied neural networks with single and multiple hidden layers for bond rating. The results show that neural networks models with single or multiple hidden layers outperform discriminant analysis. They also show that networks trained with two hidden layers outperform a network having only one hidden layer containing a comparable number of PEs during the testing phase, leading to the conjecture that redistributing an adequate number of single hidden layer elements to many optional pairings of hidden layers can produce a significant performance improvement without an unacceptable increase in training time.

Odom and Sharda (1990) also compared discriminant analysis and neural network, experimenting with different training sample proportion for bond rating and showed higher prediction accuracy of neural network model.

Liang, Chandler, Han, and Roan (1992) compares the performance of neural network, ID3, and probit models under various data conditions. The classification models include eight numeric variables (financial ratios and accounting numbers) and one nominal variable (industry classification). They shows that the neural network model mostly outperform the probit and ID3 methods in predicting accounting inventory method choice.

Tam and Kiang (1992), using bank failure data, compares a neural network approach with linear discriminant function, logit model, k nearest neighbor, and ID3. They proposed a backpropogation learning algorithm modified to include prior probabilities and misclassification cost. The empirical results shows that the neural net is a promising method of evaluating bank conditions in terms of predictive accuracy, adaptability, and robustness, especially under the conditions of multimodal distribution, adaptive model adjustment.

Chung and Silver (1992) compared linear models derived by logit model with rule-based systems produced by two induction algorithm, ID3 and the genetic algorithm (GA). The techniques performed comparably in modeling the experts at one task, graduate admission, but differed significantly at a second task, bidder selection, implying that categorical conclusions concerning the relative performance of the linear model and the induction algorithms are not appropriate. The other findings are that, for both induction algorithms, predictive performance depends on characteristics of the problems-solving task under consideration, and the linear models should not necessarily be abandoned as a useful paramorphic model of human expertise, which attempts to simulate expert decisions without regard to the cognitive processes through which those decisions were reached.

Kim (1992) evaluated three different tool families of statistical methods, ID3 and neural network to find a better approach for the bond rating, using Standard and Poor's compustat financial data. The experimental results show the neural network model is a better tool than regression, Logit, discriminant analysis and ID3 in light of prediction and classification.

Chung and Tam (1992) contrasted ID3, AQ (extended version of Predicate Calculus) and backpropagation algorithm of neural network, experimenting for their performance with two performance measure: the predictive accuracy and the representation capability. When applied to the construction project performance-assessment and bankruptcy-prediction tasks, it has the highest predictive accuracy in validation tests. The major limitation of backpropagation is in the representation capability, providing no explanation of the relationship between the input and output variables. In contrast, results generated by the symbolic approach such as ID3 and AQ are more self explanatory while both methods have less predictive accuracy than backpropagation, especially due to their wide variation in predictive performance.

Jo (1994) applied the MDA, neural network, and analogical reasoning which is included in AI techniques in bankruptcy prediction problem. The analogical reasoning method is the fundamental ability of humans to easily understand new situations by relating them to old ones and to solve problems based on previous experience of analogous problems. The method which had the highest classification ability among the three methods was neural network. He suggested the data refinement method to reduce the irregularity of source data. Two theorems were applied in building the architecture of neural network, those could determine the outline of architecture. The outline are related to the following two questions, how many hidden layer are needed and how many PEs are needed in hidden layer.

The performance the results of selected previous studies are summarized in table 1.

As discussed above, there exist various classification techniques available to researchers. Especially in most business classification task, where environments are dynamically changing, there are too many variable to be considered, including qualitative variables besides quantitative variables. Business data available usually contains noise. In such a situation, statistical approach and inductive method such as ID3 lack adaptability and generality to new environment.

In contrast, neural network can handle noisy data and adaptively adjust the model parameter when encountered new environments, changing the importance of past data. Neural network also contains nonparametric advantages like ID3. While neural network lacks explanatory capability which can explain the relative importance of input variable and how the output value is attained, the advantage of neural network outweighs the drawback, in case of business classification where the prediction accuracy is more important than the explanation.

The review of previous studies shows that the neural network model is the most powerful classification tool although there remain many unresolved issues on the design and implementation of neural network. In the next section, the design/implementation issue of neural network and data/task characteristics affecting the performance of neural network be discussed.

Table 1. Performance results of previous studies

Source	Statistical method	AI method	Domain
Braun and Chandler (1987)	53.8 (DA)	63.8 (ID3)	Stock market
Green (1987)	79.9 (Logit)	71 (ID3), 80.7 (GA)	Simulated data
Dutta et al. (1988)	64.7 (Reg.)	88.3 (2-layer NN) 82.4 (3-layer NN)	Bond rating : result using ten variables
Surkan and Singleton (1990)	39.0 (DA)	65.0 (3-layer NN) 88.0 (4-layer NN)	Bond rating
Odom and Sharda (1990)	59.26 (DA)	81.48 (3-layer NN)	Bankruptcy prediction: results of 50/50 training sample portion
Chung and Silver (1992)	79.0 (Logit) 83.8 (Logit)	77.8 (ID3), 80.0 (GA) 50.5 (ID3), 88.6 (GA)	Graduate admission task Bidder selection task
Kim (1992)	77.84 (Reg.) 75.0 (Logit) 76.67(DA)	76.67 (ID3) 84.5 (3-layer NN)	Bond rating
Cronan et al. (1992)	85.9 (RP) 62.5 (RP) 89.0 (RP)	78.4 (ID3) 50.0 (ID3) 80.0 (ID3)	Mortgage loan Commercial loan Consumer credit
Tam and Kiang (1992)	84 (DA) 81.8 (Logit) 77.2 (NN) 77.2 (NN)	79.5 (ID3) 81.8 (2-layer NN) 85.2 (3-layer NN)	Bank bankruptcy prediction : One year period hold-out sample case
Chung and Tam (1992)		38.5 (ID3 (threshold)) 59.6 (ID3 (chi-square)) 48 (AQ) 73 (2-layer NN) 79.5 (ID3 (threshold)) 79.5 (ID3 (chi-square)) 77.5 (AQ) 85.3 (2-layer NN)	Construction project performance assessment : test set Bankruptcy prediction : One-year period, test set
Jo (1994)	88.75 (DA)	90.91 (3-layer NN) 87.68 (AR)	Bankruptcy prediction

Note : Reg.=Regression, DA=Discriminant Analysis, GA=Genetic Algorithm, RP=Reculsive Partitioning, 1 NN=1 Nearest Neighbor, 2 NN=2 Nearest Neighbor, n-layer NN=n-layer Neural Network, AR=Analogical Resoning.

III. Design and implementation issues on NN

A neural network, based on the structure of the brain, consists of a number of simple, highly interconnected processing elements (neuron) which processes information by its dynamic state response to external inputs. Processing elements (PEs) are usually grouped into layers. Each processing elements (PE) do little more than receives inputs from other PEs, calculate a weighted sum of all its inputs, and produce an output applied by a transfer function. The output is propagated to other PEs through the topology of the neural network. This processing continues until a certain conditions is met, according to the network paradigm in which there are many variations.

A neural network can be configured as a directed graph with nodes representing PEs and links representing connection among PEs. In a neural network, knowledge is represented by connection weights among its PEs as numerical values which are assigned to each links. The connection weights, adaptively adjusted by learning mechanism, represents how the neural network reacts to external inputs. To get an appropriate connection weights, there are several methods proposed.

There has yet been a formal method to design an appropriate neural network for a certain classification task, including the number of PEs and hidden layers, learning methods adopted, etc., so that exploratory experiments are usually attempted to find a good architecture. This is one of the difficulties encountered in designing the neural network.

Therefore, several issues arise in designing the neural network for classification which we will describe below. Designing a neural network include neural network paradigm selection for a specific application, initialization and implementation for the selected paradigm.

3-1. Selection of the neural network paradigm

Since a neural network can be configured as a directed graph with nodes representing PEs and links representing connection among PEs, a lot of neural network paradigm are proposed according to the type of layer connections of PEs, transfer function, learning type, learning algorithm, etc.

Each paradigm is specifically developed to be suited for solving a particular type of application domain. Just to name a few, backpropagation (initiated by P. Werbos, 1974, Rumelhart et al. 1986) has multi-layer, feed-forward, hierarchical structure, using a generalized learning rule. Counterpropagation (developed by R. Hecht-Nielsen, 1987) similar to backpropagation, is nonhierarchical, feed-forward, bi-directional, using Hebbian type learning rule. ART1 (Grossberg's adaptive resonance theory, 1987) consist of two layers. The Inner layer is fully connected, feed-forwarded to the outer layer which is on-center/off-surround (competitive learning) type intra-layer connection.

Besides these paradigm, there are Hopfield network, Bi-directional Associative Memory, Fuzzy Cognitive Map, Brain-State-in-a-Box, ADALINE, and MADALINE, and many other paradigm exists.

Therefore, appropriate selection of the network paradigm for classification task could be made in light of the comparison between the classification task requirements and the paradigm capabilities and characteristics. The network paradigms are specified by the type of connectivity (topology), transfer function, type of learning and associative memory, etc., which is described below.

1) The type of connectivity

The type of connectivity concerns about how PEs are linked together. The layer connections of PEs are categorized into inter-layer type and intra-layer type. In inter-layer type connections, PEs are fully connected, partially connected, feed-forward connected, feed-back connected, bi-directionally connected or hierarchically connected. In intra-layer type connections, PEs' connections are recurrent or on-center/off-surround.

Hopfield network is a two-layer, fully connected network with output layer PEs having recurrent intra-layer connections. Kohonen's self organizing network is the type of intra-layer connection with on-center/off-surround connections (called competitive learning structure). In ART1 and recurrent back-propagation, PEs are feed-back connected.

Most of the neural network application to classification tasks adopted the fully connected, feed-forward connected network type, including Jhee (1989), Surkan & Singleton (1990), Tam & Kiang (1992), etc.

2) Transfer function

Each PE computes the weighted sum of all its input and applies transfer function to produce the output value to be routed inside the topology of the neural network. There are a number of functions used, including sigmoid functions, signum function, hard-limiting function, hyperbolic tangent function, linear function, etc. While sigmoid function is most widely used as in Surkan & Singleton (1990), Odom & Sharda(1990), Tam & Kiang (1992) etc., any continuous and differentiable function could be tried.

3) Type of learning

Learning is the process by which a neural network adjust their connection weights when presented with external input. There are two types of learning : supervised and unsupervised. In supervised learning, the actual output of a neural network is compared to the desired output, and the connection weights are adjusted to minimize the error between the actual output and desired output. Therefore, a correct answer should be presented to the network. The learning rules to adjust the weights include , delta rule, generalized delta rule, etc.

In unsupervised learning, called self-organizing learning, the network is given only the input without any information on what the desired output should be, and then network self organized by adjusting its weights according to a well-defined learning algorithm. The unsupervised type learning rule include Hebbian rule, Kohonen's rule, Grossberg learning rule.

For the business classification, the supervised learning type is ubiquitous in most application, although Garavaglia (1993) attempted the counter-propagation type unsupervised learning, of which design was not described in detail.

4) Associative memory characteristics

Association is a major feature of neural network, which is the capability of associating output for a given input. Association is categorized into two types: auto-associative and hetero-associative.

In auto-associative memory, input vector is the same as the output vector and this type of neural network is capable of recalling complete information when only a partial information is available. Auto-associative type neural network paradigm include the Hopfield network, brain-state-in-a-box network and ART, etc.

In hetero-associative memory, when a certain type of input is given, an output different from input type is produced. Bi-directional associative memory and Kohonen's' self organizing network are hetero-associative memory.

3-2. Implementation issues for the selected paradigm

After the application requirements is compared with the capabilities of the neural network paradigm, appropriate paradigm is selected, initial network parameters and condition should be specified.

1) The number of layers and the number of PEs in the layer

Some of the neural network paradigm such as ADALINE, MADALINE, Hopfield network and ART, have one or two layers. Other paradigms such as back-propagation allows a variable number of layers. Hidden layers act as layers of abstraction, pulling features from input. Increasing the number of hidden layers augments the processing power of the neural network but significantly complicates training so that tracing of error propagation become intractable.

The number of PEs in the layer also affect the performance of neural network. Too few nodes in the hidden layer prevents it from accurately training the network and too many nodes impedes generalization.

Deciding an appropriate number of layers and PEs is difficult, but there are two related theorems. Those are functional approximation capability of neural network theorem and Kolmogorov's mapping neural network existence theorem. Functional approximation capability of neural network theorem means that one hidden layer architecture could reduce the error as small as possible. Kolmogorov's mapping neural network existence theorem suggests that the maximum number of nodes in the hidden layer is restricted to 2n%1, where n is the number of input node.

Klimasauskas (1991) suggested the top down approach and the bottom up approach. In the top down approach, the network training starts with maximum number of hidden layers and PEs, and the optimal network is obtained by reducing the number. On the contrary, the bottom up approach starts with the smallest number of hidden layers, and then, adding layers and PEs, the optimal configuration is attempted.

Most researches tried different numbers of layers and PEs exploratory to map a classification task into a optimal configuration.

2) Learning parameter

When the learning rate is given high, the number of training iteration is smaller, and the system learns faster but the generalization capability of the network is deteriorated. On the contrary, with the low learning rate, the learning speed is slow, and the generality on the test data is increased. That is, if learning rate is high, the gradient descent for minimizing the error function oscillate widely, and if the learning rate is small, the gradient descent may be very small.

To avoid this problem and for the faster training, the addition of momentum term is often effectively and commonly used for back-propagation network (Plaut et at., 1986).

In practice, the default learning parameter was used as in Kim (1992) or incremental changes in parameters were tried as in Odom & Sharda (1990).

3) Input and output data

For the classification tasks, inputs to the neural network is assumed to be the attributes or features of the class and the outputs to be the class categories.

The input and output data type is dependent on the transfer function and learning type adopted ,so input and output pattern should be appropriately represented in alliance with the transfer function.

While Kohonen's' feature map requires a normalized data to perform well, most other network paradigms themselves do not usually need such a restriction on their data. But, for faster learning and convergence, input data should be normalized or scaled into the same range of the output of the transfer function. Normalization requires its training data to be scaled into the same length often by Euclidean length of data or logarithmic manipulation. Another method of normalization is to scale the training patterns into the same range of the transfer function used. If the TanH function is used, the input data is scaled between -1 and $+1$. If sigmoid is used, it is scaled between 0 and $+1$.

Table 2. Neural network design and implementation of previous studies

	Jhee (1989)	Surkan and Singleton (1990)	Odom and Sharda (1990)	Kim (1992)	Tam and Kiang (1992)	Chung and Silver (1992)	Jo(1994)
Application Domain	Bond rating	Bond rating	Bankruptcy prediction	Bond rating	Bankruptcy prediction	Construction project performance Bankruptcy prediction	Bankruptcy prediction
Paradigm of NN used	BPN	BPN	BPN	BPN	BPN with modified error function	BPN	BPN
Transfer function	Sigmoid			TanH/ Sigmoid	Sigmoid	Sigmoid	Sigmoid
Learning parameter at the time of convergency			Learning rate→0.x Momentum →.8	Learning rate→default Momentum →.4			
#of layers, #of PES	(130,60,30,4)	(7,10,5,2)	(5,5,x)	(8,7,6)	(19,10,x) (19, x)	2 layer 3 layer	1 hidden layer
Input data type	Normalized (0,x : binary)		Normalized (0∼x)		Different scaling factors for each input variables		Raw type and standardized type
Output type	0 or x						0 or 1
NN s/w used		Programmed in APL function GENNET	Neuroshell release x.x	Neural Ware professional plus II	Programmed in PASCAL	Programmed in PASCAL	Neural Ware professional plus II

The design and implementation features of neural network by previous studies are summarized in table 2.

Ⅳ. Discussion and Concluding Remarks

Most current applications of neural network to business classification mainly adopts feed-forward, back-propagation network paradigm, including Jhee (1989), Liang et al. (1992), Chung and Tam (1992), and Jo (1994), etc. The main reason for adopting back-propagation is supposedly that back-propagation is easy to understand and use.

The successful use of back-propagation network contributed to locate the research niche in business classification. But the back-propagation network has some demerits. It may require excessive training time to converge and tends to get stuck at local minima. Backpropagation is a fixed system so that it has difficulties in learning from new data.

4-1. Potential use of clustering or unsupervised learning

Back-propagation network, which have been most widely used in business classification task so far, does not assume the intrinsic grouping within the data and it can't learn from an input pattern which does not have an associated output. Backpropagation fails to take advantage of the natural grouping which may exist within the data.

Such a drawback might be overcome via unsupervised learning mechanism, such as counter-propagation paradigm. In the first phase, using a clustering or an unsupervised learning, input patterns can be clustered into several groups, and in the second phase, a certain group can have an association with an output category, assigning a new input to a correct output. Such an approach could promisingly enhance the predictive capability of the neural network approach.

Collins et al. (1988) applied this concept to mortgage underwriting compared extremely favorably to expert judgments. Using this approach, they could find the data structure to simplify the problem. Reducing the dimensionality of the problem, in the supervised learning phase, classification task became much easier.

4-2. Potential use of optimization and/or simulation techniques in neural networks

Neural network requires a large number of data to accomplish satisfactory level of learning in a specific application domain. However, we also should admit the fact that such data might include various kinds of errors or noise such as measurement errors, gathering errors, etc., which can cause serious deterioration of neural network performance.

Liang et al (1992) applied a semi-Markov processes techniques to trim away the noise or errors hidden in training data, yielding drastic decrease of size of training data. With integrating optimization techniques with neural network applications, we can expect the following advantages :

1. The redundant and inconsistent data can be removed from the training data. This increases the reliability of the resulting model.

2. The size of training data can be reduced drastically.

Simulation techniques have been widely used in business applications. They are characterized by its time-dependent, algorithmic, and mathematical modeling property. As computer hardware skill develops, simulation techniques have also undergone dramatic change from conventional algorithm-oriented approaches to more advanced intelligent approaches including expert systems, neural network, etc. Especially applying simulation techniques to neural networks can yield a number of unique features. For example, neural networks operation can be organized in the simulation mechanism. Since neural network operation is similar to the simulation, those two models can be naturally integrated into a fusion framework. If such integration is accomplished, then neural network experiments become easier and analytical. Most of all, due to the intrinsic nature of simulation, any kind of parametric combination is possible during the neural network experiments.

4-3. Integration of neural network with expert systems

Neural network proves a good alternative to avoid the bottleneck in knowledge acquisition for the expert systems. On the contrary, neural network lack explanatory capability. As a preprocessor, neural network hands in processed data to expert system for explanation. The integration aspects have been also advocated by other disciplines. Liang, Chandler, and Han (1990) discussed the integration issues between statistical and artificial intelligent approaches.

Lee (1992) applied integrated method of neural network and fuzzy system in strategic planning area. He suggested that neural network be used to strengthen the inference ability of conventional knowledge base. Fuzzy theory was used to make self-evolving mechanism, because it could handle the uncertainty.

Jhee et al. (1992) suggested neural network-based paradigm for automating the controversial identification stage of the Box-Jenkins method, in which a time series is classified into an autoregressive moving average (ARMA) model. It used neural network to select the representative parameter sets from ARMA model.

Kwon et al. (1993) proposed the conceptual framework of simulation-assisted neural network. It is consisted of five components : problem analyzer, neural network modeler, neural network simulator, evaluator, and flexibility identification system. They were interested in the characteristics of original problem. To analyze the a given problem, some statistical model and operation research models were used. The basic nature of the problem is analyzed, then simulation models must be sought which fit the nature of the problem.

Lee et al. (1994) proposed the new algorithm called to cooperative inference engine which integrate inductive learning and neural network to solve the bankruptcy prediction problem. That research was based on relation of error types of neural network. When the system error rate is lower, the type I error (predicting sound firm as bankrupt firm) is reduced and type II error (predicting bankrupt firm as sound firm) is increased. From above relation, they selected the result of prediction method. For example, when system error rate is low, they used neural network to predict bankrupt firm.

Kim (1994) applied neural network to estimate the man-hours requirements for shipbuilding assemblies. He used neural network as a main process, stepwise regression is used for variable selection and relation base. To reflect the proximity of the new case to existing ones, four measures (constant weight, Euclidean distance, MLE, and cardinal value conversion) were proposed and accommodated in the backpropogation algorithm.

In recent, Tsujino and Nishida (1994) proposed sophisticated knowledge acquisition system integrating decision tree and neural network. They defined the advantages and disadvantages of decision tree and neural network. Decision tree is one of the popular and powerful techniques for knowledge base construction, but appropriate pre- or post-processors have to be prepared when we need to achieve continuous input-output mapping. On the other hand, neural network provides enough ability of representing and learning the task having continuous input-output mapping, but structure of the network must be carefully determined beforehand to realize reliable and robust knowledge.

The neural network with explanation capability will surely be the important goal of future research.

4-4. Concluding Remarks

This paper have reviewed the statistical methods and artificial intelligence methods for business classification and discussed the design and implementation issues on neural network approach which appears to be the most powerful tool for business classification. This study provides some discussion on neural network approach including future directions.

REFERENCES

1) Altman, E. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." Journal of Finance, September 1968, pp.589~609.

2) Altman, E., R. Haldeman, and P, Narayanan. "ZETA Analysis : A New Model to Identify Bankruptcy Risk of Corporations." Journal of Banking and Finance, 1977, pp.29~54.

3) Bailey, D. and D. Thomson, "How to Develop Neural Network." AI EXPERT, June 1990, pp.38~47.

4) Bailey,D. and D. Thomson," Developing Neural Network Applications." AI EXPERT, Sept. 1990, 34~41.

5) Beaver, W., "Financial Ratios as Predictors of Failure." Journal of Accounting Research, supplement, 1966, pp.71~111.

6) Braun, H. and J. Chandler, "Predicting Stock Market Behavior Through Rule Induction : An Application of the Learning-From-Example Approach." Decision Science, 1987, pp.415~429.

7) Buchanan, B., D. Smith, W. White, R. Gritter, E. Feigenbaum, J. Lederberg, and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference. 22. Automatic Rule Formation in Mass Spectrometry by means of the Meta-DENDRAL program." Journal of American Chemical Society, September 1976, pp.6168~6178.

8) Carbonell, J. G., R. S. Michalski, and T. M. Mitchell, "An Overview of Machine Learning." In R. S. Michalski, J. G., Carbonell, and T. M. Mitchell (Ed.), Machine Learning : An Artificial Intelligence Approach. Tioga, Palo Alto, Calif., 1983.

9) Caudill, Maureen, "Neural Network Training Tips and Techniques." AI EXPERT, Jan. 1991, pp.56~61.

10) Caudill, Maureen, "Neural Network Primer." AI EXPERT, Dec. 1987-Aug. 1989

11) Collins, E., S. Ghosh and C. Scofield, "An Application of A Multiple Neural Network Learning System to Emulation of Mortgage Underwriting Judgments." Proceedings of the IEEE International Conference on Neural Networks, 1988, pp.459~466.

12) Cronan, T. P., L. W. Glorfeld and L. G. Perry, "Production System Development for Expert Systems Using a Recursive Partitioning Induction Approach : An Application to Mortgage, Commercial, and Consumer Lending." Decision Science, Vol. 22, 1991, pp.812~845.

13) Chung, H. M. and K. Tam, "A Comparative Analysis of Inductive-Learning Algorithm." Intelligent Systems in Accounting, Finance and Management, Vol. 2, 1992, pp.3~18.

14) Chung, H. M. and M. S. Silver, "Rule-Based Expert Systems and Linear Models: An Empirical comparison of Learning-By-Examples Methods." Decision Sciences, Vol. 23, 1992, pp.687~707

15) Dutta, S. and S. Shekhar, "Bond Rating : A Non-Conservative Application of Neural Networks." Proceedings of the IEEE International Conference on Neural Networks, San Diego, CA., 1988, pp.443~450.

16) Frydman, H., E. I. Altman, and D. Kao, "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress." Journal of Finance, March 1985, pp.269~291.

17) Garrison, L. and R. Michaelson, "Symbolic Concept Acquisition: A New Approach to Determining Underlying Tax Law Constructs." Journal of American Taxation Association, Fall 1989, pp.77~91

18) Gentry, J., P. Newbold, and D. Whitford, "Classifying Bankrupt Firms with Fund Flow Components." Journal of Accounting Research, Spring 1985, pp.146~160.

19) Greene, D. P., "Automated Knowledge acquisition : Overcoming the Expert System bottleneck." Proceedings of the 8th International conference on Information Systems, N.Y. : ACM Press, 1987, pp. 107-117.

20) Hecht-Nielsen, R., Neurocomputing, Addison-Wesley Publishing Company Press, 1990.

21) Holland, J., Adaptation in Natural and Artificial Systems. Ann Arbor, M. I. : University of Michigan Press, 1975.

22) Hunt, E., Marvin, J., and Stone, P. Experiments in Induction, N. Y. : Academic Press, 1966.

23) Jhee, W., "Learning Agency Judgment for CP Rating : A Comparative Study between Statistical Method and Connec-

tionist Network," Hong-Ik Review, 1989, pp.601~613.

24) Jhee, W. and J. Lee, "Performance of Neural Networks in Managerial Forecasting." Intelligent Systems in Accounting, Finance, and Management, 1993, pp.55~71.

25) Jo. H., Bankruptcy Prediction using Multivariate Discriminant Analysis, Analogical Reasoning, and Neural Network, Master's Thesis, KAIST, Korea, 1994.

26) Jones, W. and J. Hopkins, "Back-propagation : A Generalized Delta Learning Rule." BYTE Magazine, October 1987.

27) Kaplan, R. S. and G. Urwitz, "Statistical Models of Bond Ratings : A Methodological Inquiry." Journal of Business, April 1979, pp.231~262.

28) Kim, H. D., Man-hours Requirement Estimation for Shipbuilding Assemblies using Neural Networks, MS Thesis, KAIST, Korea, 1994.

29) Kim, J. W., A Comparative Analysis of Rule Based, Neural Networks, and Statistical Classification Systems for the Bond Rating Problem, Unpublished doctoral dissertation, Virginia Commonwealth University, Richmond, Virginia, 1992.

30) Klimasauskas, C., "Applying Neural Networks." PC AI, May/June, 1991, pp.20~34.

31) Kwon, Y., K. C. Lee and I. Han, "Neural Network Applications for Business Classification : State of the Art," Korean OR/MS Society Conference Proceeding, 1993.

32) Lee, K. C., I. Han and Y. Kwon, "A Comparative Study of the Bankruptcy Prediction Accuracy of Statistical and Artificial Intelligent ModelL MDA, ACLS and Neural Network," Proceedings of 1994 Japan/Korea Joint Conference on Expert Systems, Tokyo, Japan, 1994, pp.277~282.

33) Lee, K. C., H. Kim and M. J. Kim, "An Inductive Learning-Assisted Neural Network Approach to Bankruptcy Prediction: Comparison with MDA, Inductive Learning, and Neural Network Models," Korean Management Study, 1994.

34) Lee, S. B. and S. Oh, "A Comparative Study of Recursive Partitioning Algorithm and Analog Concept Learning System." Expert Systems With Applications, Vol. 1, 1990, pp.403~416.

35) Liang, T., J. Chandler, I. Han, and J. Roan, "An Empirical Investigation of Some Data Effects on the Classification Accuracy of Probit, ID3, and Neural Networks." Contemporary Accounting Research, Fall 1992, pp.306~328.

36) Liang, T., H. Moskowitz, and Y. Yih, "Integrating Neural Networks and semi-Markov Approaches for Automated Knowledge Acquisition : An Application to real-time Scheduling." Decision Sciences, Vol. 23, 1992, pp.1297~1314.

37) Marais, M., J. Patell, and M. Wolfson, "The Experimental Design of Classification Models : An Application of Recursive Partitioning and Bootstrapping to Commercial Bank Loan Classifications." Journal of Accounting Research, supplement, 1984, pp.87~114.

38) Messier, W., Jr., and J. Hansen, "Inducing Rules for Expert System Development : An Example Using Default and Bankruptcy Data." Management Science, December 1988, pp.1403~1415.

39) Michalski, R. and Chilauski, R. "Knowledge Acquisition by Encoding Expert Rules versus Computer Induction from Examples : A Case Study involving Soybean Pathology." International Journal of Man-Machine Studies, Vol.

12, No. 1, 1980, pp.63~87.

40) Michalski, R. S., "A Theory and Methodology of Inductive Learning." in R. Michalski, J. Carbonell, T. Mitchell, (Ed.), Machine Learning : An Artificial Intelligence Approach. Tioga, Palo Alto, CA., 1983.

41) Mutchler, J. "A Multivariate Analysis of the Auditor's Going-Concern Opinion Decision." Journal of Accounting Research, Fall 1985, pp.668~682.

42) Odom, M. and R. Sharda, "A Neural Network Model For Bankruptcy Prediction," Proceedings of the IEEE International Conference on Neural Network, San Diego, CA., 1990, pp.163~168.

43) Ohlson, J. "Financial Ratios and the Probabilistic Prediction of Bankruptcy." Journal of Accounting Research, Spring 1980, pp.109~131.

44) Plaut, D., S. Nowlan, and G. Hinton, Experiments on Learning by Back-propagation, Technical Report CMU-CS-86-126, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA., 1986.

45) Quinlan, J., "Discovering rules by Induction from Large collection of Examples, In D. Michie (ed.)," Expert Systems in the Micro Electronic Age. Edinburgh, Scotland : Edinburgh University Press, 1979.

46) Shaw, M. J., J. A. Gentry, "Using an Expert System with Inductive Learning to Evaluate Business Loans." Financial Management, Autumn 1988, pp.45~56.

47) Simon, H. A., "Why Should Machine Learn." in R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Ed.), Machine Learning : An Artificial Intelligence Approach. 1983.

48) Standard and Poor's, Credit Overview : Industrial Rating, New York, 1986.

49) Surkan, A. and J. Singleton, "Neural Networks For Bond Rating Improved by Multiple Hidden Layers." Proceedings of the IEEE International Conference on Neural Network, San Diego, CA., 1990, pp.157~162.

50) Tam, K. and M. Kiang, "Managerial Applications of Neural Networks : The Case of Bank Failure Predictions." Management Science, July 1992, pp.926~947.

51) Tsujino, K. and S. Nishida, "Integration of Decision Trees and Neural Networks for Hybrid Knowledge Acquisition," Proceedings of 1994 Japan/Korea Conference on Expert Systems, Tokyo, Japan, 1994, pp.17~22.

52) Zmijewski, M., "Methodological Issues related to the estimation of financial distress prediction Models." Supplement to Journal of Accounting Research, 1984, pp.59~82.

53) Zmijewski, M. and R. Hagerman, "An Income Strategy Approach to the Positive Theory of Accounting Standard Setting/Choice." Journal of Accounting and Economics, August 1981, pp.129~149.