Expert Systems with Applications 29 (2005) 1–16

Expert Systems with Applications

www.elsevier.com/locate/eswa

A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms

Kidong Lee[a], David Booth[b,*], Pervaiz Alam[c]

[a]College of Business Administration, University of Incheon, Incheon 402-749, South Korea
[b]Department of Management and Information system, College of Business Administration, Kent State University, Kent, OH 44242, USA
[c]Department of Accounting, College of Business Administration, Kent State University, Kent, OH 44242, USA

Abstract

In this study, two learning paradigms of neural networks, supervised versus unsupervised, are compared using their representative types. The back-propagation (BP) network and the Kohonen self-organizing feature map, selected as the representative type for supervised and unsupervised neural networks, respectively, are compared in terms of prediction accuracy in the area of bankruptcy prediction. Discriminant analysis and logistic regression are also performed to provide performance benchmarks. The findings suggest that the BP network is a better choice when a target vector is available.

Keywords: Supervised network; Unsupervised network; Bankruptcy prediction; Kohonen self-organizing feature map

1. Introduction

Artificial neural networks (ANNs) as alternative classification technologies to statistical modeling have been frequently used in business largely due to improved prediction accuracy. In particular, the back-propagation (BP) network, one of the supervised networks, has been the most popular neural network model used for bankruptcy prediction during the last decade (O'Leary, 1998; Tam & Kiang, 1992).

However, two major concerns are associated with the supervised neural network approach, including BP, in bankruptcy prediction: (1) most supervised studies were performed in the retrospective manner in that they relied mostly on the analysis of historical data and (2) improved accuracy is an elusive measurement when an underlying business environment changes rapidly. In fact, the supervised networks need input vectors and target vector together to make training possible. Often this target vector is available only in the retrospective way, which is the major limitation of supervised training. For the same reason, the supervised approach may also be unable to provide a real-time response to a problem. Because of the fast changing nature of information technologies today, it is difficult to assume that the improved accuracy of a study is easily transferable and applicable to future studies.

Due to the inherent limitations of the supervised approach, some of the previous studies might not be as effective as has been proclaimed. In today's fast changing business environment, we need to develop a method that can detect the changing pattern of a firm in a more timely fashion, rather than retrospectively. Thus one needs timely intervention in the case of bankruptcy prediction in order to save more resources and prevent the deterioration of a firm.

Artificial neural networks are often classified into two distinctive training types, supervised or unsupervised. As mentioned earlier, supervised training requires training pairs, input vectors and corresponding target vectors. The BP network is a good example of the supervised training type, and is the most popular training method in the ANN literature. The reason for the successes of the multi-layer perceptron (MLP) and its most common learning algorithm, BP, is that the outputs of the BP network are the estimates of posterior probabilities that have a central role in statistical pattern classification theory (Booth, Khouja, & Hu, 1992). However, supervised networks such as

* Corresponding author. Tel.: +1 330 672 1143; fax: +1 330 672 2458.
E-mail address: dbooth@bsa3.kent.edu (D. Booth).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.01.004

the BP network must be provided with a target vector. In case of bankruptcy prediction tasks, the target vector is '*whether or not a firm has failed*' which must be embedded in the supervised training process. To launch a supervised analysis, researchers may have to wait until the target information of 'whether or not a firm has failed' is collected. Thus, this supervised approach is, indeed, performed in the retrospective mode.

In some situations, especially in today's fast-changing, real-time-based business environment that demands prompt responses, such extra information may not be readily available for training. In such circumstances, unsupervised neural networks might be more appropriate technologies to be use. Unlike supervised networks, unsupervised neural networks need only input vectors for training. Therefore, the extra information of whether or not a firm has gone bankrupt need not be embedded in the unsupervised training process. Developed by Kohonen (1982, 1997) and many others, the training algorithms for the unsupervised networks modify their weights to process input vectors into some similar output classes (or clusters).

As mentioned earlier, the BP supervised network has been the most widely used network for bankruptcy prediction. It is, however, desirable that these two different approaches, supervised and unsupervised, be investigated so that the feasibility and effectiveness of diverse neural network algorithms may be better understood. Strictly speaking, it is somewhat difficult to compare the supervised with unsupervised networks and across all dimensions because of their radically different orientations. In this study, we confine ourselves to the BP network and the Kohonen self-organizing feature map, selected as the representative type for the supervised and unsupervised neural networks, respectively. Discriminant analysis and logistic regression are also performed to provide performance benchmarks.

Since the BP network (supervised) utilizes one more critical variable, the target vector, in its training process, it might be expected that the BP network would be more accurate than the Kohonen self-organizing feature map (unsupervised). However, the focus of this study is to explore how much less accurate the Kohonen network could be compared with the BP network and the two other statistical techniques, in term of prediction accuracy in the context of bankruptcy prediction. Thus, we test the level of prediction accuracy of the Kohonen self-organizing feature map as compared with the BP network.

The rest of this paper is organized as follows. Section 2 discusses the prior research on the outlier detection paradigm, the BP network, the Kohonen self-organizing feature map, and bankruptcy prediction. Section 3 provides the research design and methodologies in terms of data, variables, cross-validation scheme, and classification technologies used for the study. Experimental results are shown and discussed in Section 4, where the results of each of the four individual classification techniques are first presented and their comparisons are then given. The limitations, the future research directions, and the conclusions for the study are given in Section 5.

2. Prior literature

In this section, the theoretical background for bankruptcy prediction studies from the perspective of an outlier detection paradigm is provided. Then, the BP network and the Kohonen self-organizing feature map, each selected as the representative type for supervised and unsupervised learning, are reviewed with respect to their training particular processes, followed by their applications to bankruptcy prediction.

2.1. Outlier detection paradigm

Outliers are 'out-of-control' observations that lie beyond a standard process control limit. Recognition of such abnormal entities is a very important activity in everyday life. In business settings, identification of outliers may be presented in a variety of forms: bankruptcy identification (Alam, Booth, Lee, & Thordarson, 2000; Booth, 1982; Booth, Alam, Ahkam, & Osyk, 1989; Lenard, Alam, & Booth, 2000), production process monitoring (Hamburg, Booth, & Weinroth, 1997), robot evaluation and selection (Booth et al., 1992), valuation of companies (Aupperle, Acar, & Booth, 1986) and even in the nuclear safeguard areas (Grznar, Booth, & Sebastian, 1997; Hamburg, Booth, & Weinroth, 1996).

Bankrupt firms are outliers from the perspective of a group of healthy firms. The fact that only about 2% of all firms go bankrupt in a normal economic period suggests (Lenard, Alam, & Madey, 1995; O'Leary, 1998) that bankrupt firms can indeed be treated as outliers. Within the outlier identification paradigm, this study attempts not only to extend methodological choices for bankruptcy prediction by comparing different neural network types, but also gives a chance to enhance the understanding of different ANN's behavior.

2.2. Back propagation (BP) network

The BP algorithm, a systematic training method for a Multi-Layer Perceptron (MLP), has been the most widely used in bankruptcy prediction tasks (O'Leary, 1998; Zhang, Hu, Patuwo, & Indro, 1999). Fig. 1 shows MLP with one hidden layer. This architecture shows i nodes in the input layer, h nodes in the hidden layer, and t nodes in the output layer.

The actual output value, $\mathbf{Y}$, in the output node of the BP network is computed as in Eq. (1),

$$\mathbf{Y} = f(\mathbf{X} * \mathbf{W}) \tag{1}$$

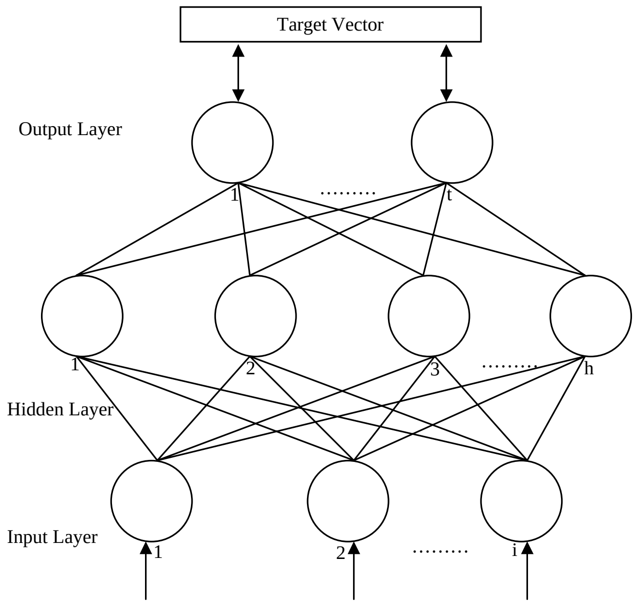

Fig. 1. MLP with one hidden layer.

where **Y** stands for the output vector, **X** the input (row) vector, **W** the weight vector (including bias), and $f(\cdot)$ denotes an activation function. The activation function, $f(\cdot)$, transforms the sum of input values into output values of the node. Typical choices of the activation function consist of the logistic, the tangent, the sign, and the linear. The logistic function used in this study is shown in Eq. (2),

$$\mathbf{Y} = f(\mathbf{X} * \mathbf{W}) = 1/(1 + \mathrm{e}^{-(\mathbf{X}*\mathbf{W})}) \tag{2}$$

Then, these actual output values, **Y**, are compared with the target (or desired) values as shown in Fig. 1, and the difference (or error) between the target values (**Y**-Target) and the actual network outputs by some error measure such as sum of squares, are computed. The main idea of the BP network is to reduce this error to some desirable level by adjusting the weight vector (O'Leary, 1998; Wasserman, 1989; Zhang et al., 1999).

Researchers must find an optimal architecture for the BP network with its associated parameters before they use it. This task involves determining a number of design variables for a BP network (i.e. activation function, error measure function, MLP architecture, and a particular training method). As shown, determining an appropriate architecture and parameters for a BP network is not simple (Refenes, 1995; Zhang, 1998). The detailed specification for this study is shown in Section 3.

2.3. Kohonen self-organizing neural network

Another popular neural network model is the Kohonen self-organizing feature map (Kohonen, 1982, 1997). The Kohonen self-organizing neural networks have appeared in many fields, for example, classification (Corridoni, Del Bimbo, & Landi, 1996; Deschenes & Noonan, 1995; Schonweiler, Kaese, Moller, Rinscheid, & Ptok, 1996), pattern recognition (Xin-Hua & Hopke, 1996), clustering (Kiviluoto, 1998; Martin-del-Brio & Serrano-Cinca, 1995), and forecasting (Der Voort, Dougherty, & Watson, 1996). In the Kohonen network, input vectors are presented without specifying the desired output (hence, unsupervised).

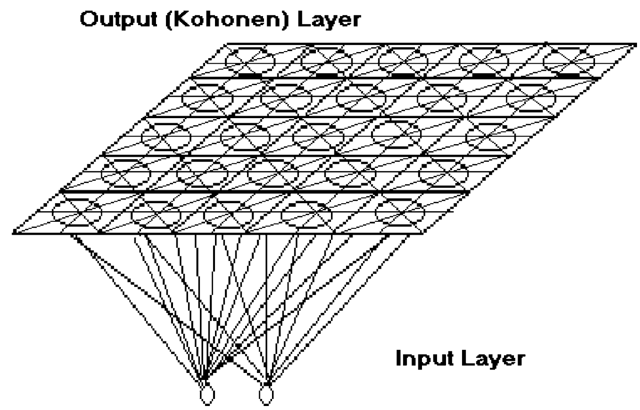

Fig. 2. Typical architecture of the Kohonen self-organizing feature map.

It appears that the unsupervised training type is more plausible in an ever-changing business environment today than is supervised training method (Hinton & Sejnowski, 1999; Kelly, 1994; Waldrop, 1992). Fig. 2 shows the typical architecture of the Kohonen self-organizing feature map with two input nodes and 5×5 output nodes. As is seen, the input vector space of the object is projected into the two dimensional Kohonen output space. It often ends up in the reduction of the dimensionality of the input space (not shown in Fig. 2).

The Kohonen training process succinctly summarized by Nour (1994) is shown below as:

1. Initialize $\mathbf{W}_i(t)$ to random values, and set $t=0$.
2. Present an input vector **X** to the network, and compute the distance (similarity) **D** using the Euclidean metric to find the closest matching unit c, to each input vector.

$$\mathbf{D}_i = ||\mathbf{X}(t) - \mathbf{W}_i(t)||, \forall i$$

$$N_c = \min\{\mathbf{D}_i\}, \forall_i$$

3. Update the weight vector according to the following rule.

$$\mathbf{W}_i(t) = \mathbf{W}_i(t) + \mathrm{lr}(t) * h(t,r) * \{\mathbf{X}(t) - \mathbf{W}_i(t)\} \quad i \in N_c$$

$$\mathbf{W}_i(t) = \mathbf{W}_i(t) \quad i \notin N_c$$

If $t > T$ stop, else increment t and go to Step 2.

In this algorithm, t is the iteration step, T is a number of iterations predefined, and $\mathrm{lr}(t)$ is a learning rate which ranges in (0,1). The term $h(t, r)$ is a neighborhood function, which decreases over the iteration step and the topological distance $r=||r_i-r_c||$ between unit i, and unit c (the winner), where r_i and r_c are the coordinates of units i and c, respectively.

The most basic competitive learning rule is a winner-take-all approach, whereby only the winning node adapts its weight (i.e. no neighborhood function). This winner-take-all approach is often called competitive learning. The Kohonen self-organizing learning method is a variation of the above competitive, winner-take-all approach in that not only the winner but also its neighbors can update their weights together. The neighbor nodes are nodes that are located around the winner.

In summary, the main characteristics of BP are: (1) both input and output are needed, (2) does not self-organize, and (3) computationally heavy. The primary features of the Kohonen network are: (1) only input is needed, (2) constructs an output, (3) is computationally light.

2.4. Bankruptcy prediction

Bankruptcy prediction, a binary decision-making problem, has long been a popular subject for business researchers. The improvement of this bankruptcy prediction area comes from the incessant effort of past researchers to develop ratio analyses, linear models, and nonlinear models including neural networks. Beaver (1966) was one of the first researchers to study bankruptcy prediction by testing the efficacy of several financial ratios in their classification and prediction capabilities. Altman (1968) introduced a class of models based on discriminant analysis in classifying bankruptcy prediction using the following five financial variables: working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, market value of equity/total debt, and sales/total assets. Ohlson (1980) used logistic regression to estimate the probabilities of a bankruptcy.

Neural networks were not used as a bankruptcy classification technology until the early 1990s. Odom and Sharda (1990) were the first researchers to investigate the feasibility of neural networks in firm failure prediction. They found that BP networks are at least as accurate as discriminant analysis. After this first neural experiment, a significant volume of neural network research followed (Alam et al., 2000; Jo, Han, & Lee, 1997; Kiviluoto, 1998; Martin-del-Brio & Serrano-Cinca, 1995; O'Leary, 1998; Salchenberger, Cinar, & Lash, 1992; Sharda & Wilson, 1996; Tam & Kiang, 1992; Tsukuda & Baba, 1994; Udo, 1993; Wilson & Sharda, 1994; Zhang et al., 1999).

Tam and Kiang (1992) compared a BP network experiment with a linear classifier, a logistic regression, *k*NN, and ID3. The 19 financial ratios were used in classification task for the 59 matched pairs of Texas banks that failed in the period of 1985–1987. They found that the BP network approach outperformed the other classification techniques. Salchenberger et al. (1992) initially selected 29 variables and then performed a stepwise regression to reduce the number of variables into the final five test variables. With a bank data set of the period January 1986–December 1987, their experiment was to test the possible performance difference of BP networks over a logistic regression. Udo's (1993) study compared the effectiveness of a BP network with a multiple regression in bankruptcy prediction. Udo's findings confirmed that the BP network is as accurate as or more accurate than a multiple regression model.

Tsukuda and Baba (1994) compared the effectiveness of a BP network versus discriminant analysis in bankruptcy prediction using financial data for 1 and 3 years prior to failure for two listed and unlisted company sets of Japanese corporations. The results showed that the BP network approach seems to work rather well with noisy data than its statistical counterparts. Wilson and Sharda (1994) and Sharda and Wilson (1996) used an experimental design of training and test sets to test BP's effectiveness compared to many statistical classification methods. With the use of Monte Carlo re-sampling techniques, they confirmed BP's prediction accuracy over more conventional statistical methods.

Martin-del-Brio and Serrano-Cinca (1995) applied the self-organizing neural network (Kohonen) to two financial data sets taken from the state of Spanish economy: the Spanish banking crisis of 1977–1985 and the financial state of Spanish companies in 1990–1991. Their results were interesting in that by applying the unsupervised training algorithm, the input feature spaces were projected into the natural clustering (or delimiting) of regions of interest on the output maps. Jo et al. (1997) compared three different techniques in bankruptcy prediction: discriminant analysis, case-based forecasting, and BP network. In classifying Korean firms during 1991–1993, matched by industry and average credit rating within industry, they found that the BP network was better than the two other techniques. Their experiments also showed that experiments with raw data produced a better result than with normalized data.

O'Leary (1998) launched a meta study comparing 15 prior neural network studies on bankruptcy prediction in terms of sampling, impact of different ratios of failed and non-failed firms, software used, input variables, the number of hidden layers and hidden nodes, performance measures, and misclassification costs. The findings of this comparative study confirmed that in general the BP approach outperforms other statistical classifiers, but often with a high cost of time and effort in experiments.

Kiviluoto (1998) applied the self-organizing map to 1137 Finnish industrial enterprises. In Kiviluoto's study, the self-organizing map is used for a way to indicate the '*bankruptcy zone*' in the output space. Using a five-fold cross-validation scheme, Zhang et al. (1999) provided a comprehensive review of a neural network approach on firm failure. They used six input variables (Altman's variables plus the current ratio) with a data set covering a 12-year period and confirmed that the BP network outperforms a logistic regression.

Alam et al. (2000) compared three different algorithms—a fuzzy clustering algorithm and two self-organizing neural network approaches—in a data set representing a real bankruptcy proportion in the real world. Their findings were interesting in that they identified some 'gray' areas between healthy and bankruptcy firms. Firms in the gray area could be possible bankrupt candidates that need to be watched closely.

In sum, the early ANN researchers heavily relied on the BP network but recently we see a pattern of neural network researchers that have attempted to experiment with the Kohonen self-organizing feature maps, rather than to test with the BP network alone. Once again, direct comparison between these two different learning styles is somewhat difficult, but we try to contrast supervised and unsupervised neural networks using their most frequently used models, hoping that the advantages and disadvantages of the Kohonen self-organizing feature map to the BP network becomes apparent. Next is the detailed research design for this study.

3. Research design and methodology

The data sets and variables used, the cross-validation scheme as well as the detailed specifications of neural networks are described below.

3.1. Data and variables

The data sample for this bankruptcy prediction study consists of Korean firms that filed for bankruptcy during 1995–1998. An initial search of bankrupt firms is made through the Security and Exchange Commission (SEC) filings stored in an on-line database (commercially available in an electronic format) of the Korea Investors Service (KIS), Inc. (www.kisrating.com), which is a strategic partner with Moody's Investors Service in the Korean security market.

Financial institutions such as commercial banks or investment bankers are excluded in this data set since in the Korean market, the fate of such financial intermediaries seem to be much more affected by government policies rather than by their financial position. Searching for failed firms resulted in 113 non-financial failed firms among the listed companies in the Korean stock market. Then, the availability of the financial ratios for the failed firms further reduced the final bankrupt sample size to 84 since some of them did not report their financial status on their bankruptcy filings.

Each failed firm is matched with a non-failed firm in terms of (1) asset size and (2) a two-digit Standard Industrial Classification (SIC) code as control measures. The asset size of a non-failed firm is matched with that of a failed firm using the 3-year period prior to bankruptcy filings. As a result, we have a matched sample of 168 firms, 84 failed firms and 84 non-failed firms. Two time-framed financial data sets, the 2- and the 3-years prior to bankruptcy filings, are prepared for this experiment in order to see if any of these classification tools can detect any discrepancy of the financial condition of a firm between this time-difference of the data. As noted, this selected period closely resembles the outbreak of the Asian financial crisis. Thus, we test the early warning capability of the classification tools using only 2 or 3 years ahead of bankruptcy filing, rather than using a longer time series data.

Each firm is described by Altman's five variables since the prediction capabilities of these ratios are well documented in the previous literature (Altman, 1968; Boritz & Kennedy, 1995; Odom & Sharda, 1990; Zhang et al., 1999):

1. WCTA = working capital/total assets as a measure of the net liquid assets of the firm to the total capitalization.
2. RETA = retained earnings/total assets as a measure of cumulative profitability.
3. EBITTA = earnings before interest and taxes/total assets as a measure of true productivity of the firm's assets.
4. MEDEBT = market value of equity/book value of total debt as a measure how much the firm's assets can decline in value before the liabilities exceed the assets and the firm becomes insolvent.
5. SALETA = sales/total assets as a measure of the sales generating ability of the firm assets.

This study does not use a data normalization method since researchers suggest that data normalization does not seem to improve the effectiveness of neural networks (Jo et al., 1997; Zhang et al., 1999).

3.2. Cross-validation scheme

Any bias due to changing data set composition could have a detrimental impact on determining neural network architecture and its parameters. A cross-validation technique is introduced to investigate the classification performance of neural networks in terms of sampling variation. The cross-validation technique enables us to use a whole data set so that any bias effect would be minimized (Tam & Kiang, 1992; Zhang et al., 1999).

In this study, a four-fold cross validation technique is used. The total data set that contains 84 matched firms (84 failed firms and 84 non-failed firms) is divided into quartiles of equal and mutually exclusive subsets, each of which contains 21 matched objects. Table 1 shows the details of this four-fold cross-validation scheme.

Training is conducted on any two of the four subsets while the remaining two sets are used for validation and testing purposes, respectively. The validation set is introduced as an early stopping technique, to improve generalization (MathWorks, 1997). Using this early stopping technique, the validation process is embedded

Table 1
Four-fold cross-validation technique scheme

	Set of matched and unmatched firms			
	First quartile	Second quartile	Third quartile	Fourth quartile
Subset 1	Training set (42 firms)		Validation set (21 firms)	Testing set (21 firms)
Subset 2	Testing set (21 firms)	Training set (42 firms)		Validation set (21 firms)
Subset 3	Validation set (21 firms)	Testing set (21 firms)	Training set (42 firms)	
Subset 4	Training set (21 firms)	Validation set (21 firms)	Testing set (21 firms)	Training set (21 firms)

into the training results, which would prevent a possible upward bias of the classification accuracy of a training set. Overall, the training set includes 168 firms, 42 testing set firms, and 42 validation set firms.

3.3. The BP network

For the BP experiment, the Matlab Neural Network package was used. A BP network with a three-layer model (also called a one-hidden layer model) is considered for this study. The previous studies proved that as long as there are sufficient numbers of hidden nodes provided, this architectural BP network is able to approximate any arbitrary function (Funahashi, 1989; Hornik, Stinchcombe, & White, 1989).

Five input nodes are used. In a cross-sectional study, the number of input nodes is usually the same as the number of independent variables used in a study (Zhang et al., 1999). For the output nodes, one output node is sufficient for a binary classification problem. A logistic transfer function is used for the hidden nodes and the output node. Therefore, the range of an output value from the BP network is (0, 1). If the output value exceeds 0.5, a firm is classified as bankrupt and non-bankrupt, otherwise. Also, all nodes except the input nodes have a bias term.

The number of hidden nodes is used as a major experimental factor: there is no definite rule to follow in this hidden node decision. Thus, one to ten hidden nodes are tested in this study. The Levenberg–Marquardt algorithm for training is used since it provides the fastest convergence and is specifically designed for the square error cost function. The Levenberg–Marquardt algorithm works in a way that when error size is large, the algorithm approximates gradient decent, whereas as error size gets smaller, the algorithm becomes the Gauss–Newton method which is faster and more efficient (MathWorks, 1997). The mean square error (MSE) function is used for the error function. It has been a popular choice in the past literature for theoretical considerations and provides a consistent error function (Berardi, 1988; Tam and Kiang, 1992).

3.4. The Kohonen self-organizing feature map

A two-dimensional Kohonen output layer is used to help provide a visual presentation. The Viscovery Som 3.0 system (www.eudaptics.com) software is used for this Kohonen experiment. Five input nodes are used corresponding to the five financial ratios. However, selecting the appropriate number of output nodes is quite difficult and this is usually experiment-dependent. There is no consensus among researchers about the subject. For example, Serrano-Cinca (1996) used a 14 by 14 grid for 74 training vectors. As a result, some output nodes may not win for the given input vector. Nour (1994) suggested that to obtain good mapping results, the number of output nodes in the Kohonen neural network should be at least 10–20% of training vectors (or objects). However, using too few output nodes may cause the congestion of input vectors over an output node, which may make it difficult to distinguish the characteristics of the output space. Thus, it seems that it is better to use a large number of output nodes and this study thus uses 200 output nodes.

The weight of each connection between an input node and an output node is initialized in a random value automatically generated by the Viscovery Som 3.0 system (www.eudaptics.com). The four subsets (generated by the cross-validation technique) are randomly sequenced to minimize any variation due to any possible input sequence pattern. So, with a combination of initial weight randomization and input sequence randomization, it is hoped that any bias effect can be minimized.

For the training cycle decision, there is no definitive stopping point. A heuristic is to use enough training cycles so that a network approaches a stable state. A preliminary experiment showed that the self-organizing feature map usually adjust its weights quickly to their inputs. Thus, it may not be necessary that the Kohonen self-organizing feature map needs too many training cycles. For a topological function and a distance function, this study uses the default functions of the Viscovery Som 3.0.

When using unsupervised learning neural networks such as the Kohonen, researchers are faced with two crucial decisions: (1) the problem of how many clusters to be included in a particular context, and (2) the identification of each cluster's characteristics in the output space. First, for the decision about cluster numbers, there is no definite rule about how many clusters should be included for a problem. Thus, researchers often rely on their own judgment for each particular application at hand. Since bankruptcy prediction is a binary decision process, it would be desirable if there exists only two distinctive clusters, one containing mostly bankrupt firms and the other containing mostly

Table 2
Effects of hidden nodes on BP experiment (2-years of data used)

Number of hidden nodes	Type	Subset 1		Subset 2		Subset 3		Subset 4	
		#	%	#	%	#	%	#	%
Panel A for training sets									
1	Correct classification	**66**	**78.57**	**64**	**76.19**	59	70.24	**69**	**82.14**
	Type I error	**4**	**4.76**	**6**	**7.14**	15	17.86	**3**	**3.57**
	Type II error	**14**	**16.67**	**14**	**16.67**	10	11.90	**12**	**14.29**
2	Correct classification	65	77.38	63	75.00	**59**	**70.24**	68	80.95
	Type I error	4	4.76	6	7.14	**15**	**17.86**	4	4.76
	Type II error	15	17.86	15	17.86	**10**	**11.90**	12	14.29
3	Correct classification	63	75.00	65	77.38	56	66.67	65	77.38
	Type I error	3	3.57	5	5.95	11	13.10	6	7.14
	Type II error	18	21.43	14	16.67	17	20.24	13	15.48
4	Correct classification	63	75.00	68	80.95	58	69.05	67	79.76
	Type I error	9	10.71	7	8.33	10	11.90	5	5.95
	Type II error	12	14.29	9	10.71	16	19.05	12	14.29
5	Correct classification	64	76.19	68	80.95	62	73.81	63	75.00
	Type I error	10	11.90	8	9.52	15	17.86	8	9.52
	Type II error	10	11.90	8	9.52	7	8.33	13	15.48
6	Correct classification	59	70.24	63	75.00	59	70.24	65	77.38
	Type I error	3	3.57	6	7.14	10	11.90	10	11.90
	Type II error	21	25.00	15	17.86	15	17.86	9	10.71
7	Correct classification	61	72.62	64	76.19	57	67.86	64	76.19
	Type I error	3	3.57	6	7.14	12	14.29	6	7.14
	Type II error	19	22.62	14	16.67	15	17.86	14	16.67
8	Correct classification	64	76.19	57	67.86	58	69.05	62	73.81
	Type I error	5	5.95	3	3.57	6	7.14	15	17.86
	Type II error	15	17.86	24	28.57	20	23.81	7	8.33
9	Correct classification	66	78.57	69	82.14	50	59.52	70	83.33
	Type I error	4	4.76	6	7.14	25	29.76	6	7.14
	Type II error	14	16.67	9	10.71	9	10.71	8	9.52
10	Correct classification	64	76.19	57	67.86	64	76.19	66	78.57
	Type I error	9	10.71	3	3.57	9	10.71	6	7.14
	Type II error	11	13.10	24	28.57	11	13.10	12	14.29
Panel B for test sets									
1	Correct prediction	**31**	**73.81**	**30**	**71.43**	31	73.81	**25**	**59.52**
	Type I error	**8**	**19.05**	**4**	**9.52**	6	14.29	**5**	**11.90**
	Type II error	**3**	**7.14**	**8**	**19.05**	5	11.90	**12**	**28.57**
2	Correct prediction	30	71.43	25	59.52	**33**	**78.57**	25	59.52
	Type I error	6	14.29	5	11.90	**2**	**4.76**	6	14.29
	Type II error	6	14.29	12	28.57	**7**	**16.67**	11	26.19
3	Correct prediction	31	73.81	30	71.43	30	71.43	25	59.52
	Type I error	5	11.90	3	7.14	4	9.52	6	14.29
	Type II error	6	14.29	9	21.43	8	19.05	11	26.19
4	Correct prediction	27	64.29	25	59.52	32	76.19	25	59.52
	Type I error	11	26.19	4	9.52	5	11.90	6	14.29
	Type II error	4	9.52	13	30.95	5	11.90	11	26.19
5	Correct prediction	31	73.81	28	66.67	33	78.57	25	59.52
	Type I error	7	16.67	3	7.14	6	14.29	7	16.67
	Type II error	4	9.52	11	26.19	3	7.14	10	23.81
6	Correct prediction	29	69.05	28	66.67	32	76.19	25	59.52
	Type I error	5	11.90	1	2.38	4	9.52	7	16.67
	Type II error	8	19.05	13	30.95	6	14.29	10	23.81
7	Correct prediction	29	69.05	28	66.67	33	78.57	24	57.14
	Type I error	8	19.05	1	2.38	3	7.14	8	19.05
	Type II error	5	11.90	13	30.95	6	14.29	10	23.81
8	Correct prediction	28	66.67	28	66.67	32	76.19	24	57.14
	Type I error	12	28.57	0	0.00	4	9.52	11	26.19
	Type II error	2	4.76	14	33.33	6	14.29	7	16.67
9	Correct prediction	29	69.05	27	64.29	32	76.19	24	57.14
	Type I error	11	26.19	4	9.52	4	9.52	8	19.05
	Type II error	2	4.76	11	26.19	6	14.29	10	23.81

(*continued on next page*)

Table 2 (*continued*)

Number of hidden nodes	Type	Subset 1		Subset 2		Subset 3		Subset 4	
		#	%	#	%	#	%	#	%
10	Correct prediction	29	69.05	26	61.90	33	78.57	23	54.76
	Type I error	10	23.81	2	4.76	4	9.52	5	11.90
	Type II error	3	7.14	14	33.33	5	11.90	14	33.33

Note that # denotes the number of objects classified into each category and that % shows the percent of objects classified into each category. Type I error occurs when a bankrupt firm is classified as a non-bankrupt firm while Type II error occurs when a non-bankrupt firm is classified as a bankrupt firm.

non-bankrupt firms, in the output map. That is not the case in reality. Thus, we use two to four clusters in this binary decision making situation. In fact, any number of clusters less than 84, the number of objects used in this study, could be used for the number of clusters decision. One of the most difficult questions in applying unsupervised types of classification techniques is to decide how many clusters should be included for a data set at hand (Alam et al., 2000).

Then, the resulting clusters need to be characterized (or labeled) so that a particular cluster can be classified into a bankrupt group or a non-bankrupt group. Labeling a cluster (e.g. naming a cluster as a bankrupt group) is based upon a majority rule in this study. That is, if a cluster contains more bankrupt firms than non-bankrupt ones, the cluster is named as a bankrupt cluster, and a non-bankrupt, otherwise. Inserting object labels into the output map after training has been completed does this. Unlike the BP network, the Kohonen procedure does not use a target vector during the training process. The injection method was used by Martin-del-Brio and Serrano-Cinca (1995) and Serrano-Cinca (1996). The performance results of the Kohonen self-organizing map are also shown in a tabular form so that they can be easily compared with the outcomes of the BP and other statistical classifiers.

3.5. Statistical models

Two widely used statistical techniques, discriminant analysis and logistic regression, are conducted to provide performance benchmarks. For discriminant analysis, a quadratic discriminant analysis (QDA) is used rather than a linear analysis because the covariance matrices are different (Hair, Anderson, Tatham, & Black, 1995). The range of the expected output value of logistic regression is (0, 1), so it is usually interpreted as the probability of class belonging. Further, it has been suggested in practice that a logistic regression approach is often preferred, especially when the basic OLS assumption of normality of the variables is not met (Hair et al., 1995).

In sum, the four classification technologies compared are: (1) the BP network, (2) the Kohonen self-organizing feature map, (3) logistic regression analysis, and (4) discriminant analysis. For statistical analysis, the SAS program is used. For neural network experiments, the Matlab Neural Network Toolbox 5.3 is used for the BP network while the Viscovery Som 3.0 is used for the Kohonen self-organizing feature map. Neural network experiments were done on three Pentium III personal computers. Note that for the BP experiment, five different random seeds of initial weights were generated and then, five different runs were done for each factor. For the Kohonen experiment, the Viscovery Som 3.0 provides a random initial value to each arc of the network.

The prediction accuracy of the test sets obtained by each run was ranked by total misclassifications, and the median run was taken as the reported result. Performance outcomes of each classifier are measured in terms of the number and percentage of correct classification/prediction, and the number and percentage of misclassification: Type I error (bankrupt firms being classified as non-bankrupt) versus Type II error (non-bankrupt firms being classified as bankrupt).

4. Experimental results

4.1. The BP network

Tables 2 and 3 show the hidden node effects on the BP neural network performance when applied to the four-fold cross-validation scheme to the two data sets, 2-year (Table 2) or 3-year (Table 3) prior to a bankruptcy filing. Four subsets of the whole data were iteratively tested with one to ten hidden nodes of the architectural design of the BP networks. The four-fold cross-validation scheme with the ten different hidden-node combinations resulted in 40 cells explored in each table.

In both tables, each of the hidden node experiments shows the number (and percentage) of the correct *classification* for the four training sets (Panel A) as well as of the correct *prediction* for the four test sets (Panel B). Note that the validation process using the early stopping is embedded in the training period so that this embedded training would minimize an upward bias. Tables 2 and 3 also list the Type I error and Type II errors. Since we did not incorporate a misclassification cost difference between Type I and II errors within the training procedure itself, the only thing that matters in this study is the total misclassification number, rather than each separate misclassification number or cost.

Table 3
Effects of hidden nodes on BP experiment (3-years of data used)

Number of hidden nodes	Type	Subset 1		Subset 2		Subset 3		Subset 4	
		#	%	#	%	#	%	#	%
Panel A for training sets									
1	Correct classification	61	72.62	**59**	**70.24**	**56**	**66.67**	61	72.62
	Type I error	10	11.90	**11**	**13.10**	**16**	**19.05**	11	13.10
	Type II error	13	15.48	**14**	**16.67**	**12**	**14.29**	12	14.29
2	Correct classification	**61**	**72.62**	65	77.38	56	66.67	59	70.24
	Type I error	**9**	**10.71**	6	7.14	17	20.24	18	21.43
	Type II error	**14**	**16.67**	13	15.48	11	13.10	7	8.33
3	Correct classification	61	72.62	65	77.38	59	70.24	58	69.05
	Type I error	13	15.48	6	7.14	12	14.29	14	16.67
	Type II error	10	11.90	13	15.48	13	15.48	12	14.29
4	Correct classification	59	70.24	59	70.24	58	69.05	57	67.86
	Type I error	9	10.71	18	21.43	9	10.71	19	22.62
	Type II error	16	19.05	7	8.33	17	20.24	8	9.52
5	Correct classification	62	73.81	66	78.57	55	65.48	65	77.38
	Type I error	6	7.14	8	9.52	8	9.52	6	7.14
	Type II error	16	19.05	10	11.90	21	25.00	13	15.48
6	Correct classification	60	71.43	73	86.90	62	73.81	64	76.19
	Type I error	18	21.43	5	5.95	14	16.67	7	8.33
	Type II error	6	7.14	6	7.14	8	9.52	13	15.48
7	Correct classification	62	73.81	64	76.19	64	76.19	70	83.33
	Type I error	9	10.71	6	7.14	9	10.71	2	2.38
	Type II error	13	15.48	14	16.67	21	25.00	12	14.29
8	Correct classification	66	78.57	71	84.52	65	77.38	**62**	**73.81**
	Type I error	4	4.76	5	5.95	10	11.90	**8**	**9.52**
	Type II error	14	16.67	8	9.52	9	10.71	**10**	**11.90**
9	Correct classification	68	80.95	64	76.19	62	73.81	68	80.95
	Type I error	4	4.76	6	7.14	13	15.48	5	5.95
	Type II error	12	14.29	14	16.67	9	10.71	11	13.10
10	Correct classification	70	83.33	61	72.62	62	73.81	65	77.38
	Type I error	7	8.33	10	11.90	7	8.33	9	10.71
	Type II error	7	8.33	13	15.48	15	17.86	10	11.90
Panel B for test sets									
1	Correct prediction	29	69.05	**30**	**71.43**	**30**	**71.43**	21	50.00
	Type I error	10	23.81	**2**	**4.76**	**7**	**16.67**	12	28.57
	Type II error	3	7.14	**10**	**23.81**	**5**	**11.90**	9	21.43
2	Correct prediction	**30**	**71.43**	26	61.90	24	57.14	25	59.52
	Type I error	**7**	**16.67**	2	4.76	11	26.19	7	16.67
	Type II error	**5**	**11.90**	14	33.33	7	16.67	10	23.81
3	Correct prediction	30	71.43	26	61.90	30	71.43	25	59.52
	Type I error	10	23.81	2	4.76	5	11.90	7	16.67
	Type II error	2	4.76	14	33.33	7	16.67	10	23.81
4	Correct prediction	29	69.05	27	64.29	24	57.14	24	57.14
	Type I error	7	16.67	2	4.76	9	21.43	7	16.67
	Type II error	6	14.29	13	30.95	9	21.43	11	26.19
5	Correct prediction	30	71.43	27	64.29	28	66.67	25	59.52
	Type I error	8	19.05	4	9.52	6	14.29	7	16.67
	Type II error	4	9.52	11	26.19	8	19.05	10	23.81
6	Correct prediction	28	66.67	26	61.90	30	71.43	27	64.29
	Type I error	13	30.95	3	7.14	8	19.05	7	16.67
	Type II error	1	2.38	13	30.95	4	9.52	8	19.05
7	Correct prediction	29	69.05	28	66.67	30	71.43	25	59.52
	Type I error	9	21.43	4	9.52	3	7.14	6	14.29
	Type II error	4	9.52	10	23.81	9	21.43	11	26.19
8	Correct prediction	29	69.05	25	59.52	29	69.05	**28**	**66.67**
	Type I error	8	19.05	6	14.29	8	19.05	**7**	**16.67**
	Type II error	5	11.90	11	26.19	5	11.90	**9**	**21.43**
9	Correct prediction	30	71.43	28	66.67	29	69.05	25	59.52
	Type I error	9	21.43	2	4.76	7	16.67	6	14.29
	Type II error	3	7.14	12	28.57	6	14.29	11	26.19

(*continued on next page*)

Table 3 (*continued*)

Number of hidden nodes	Type	Subset 1		Subset 2		Subset 3		Subset 4	
		#	%	#	%	#	%	#	%
10	Correct prediction	29	69.05	26	61.90	28	66.67	28	66.67
	Type I error	10	23.81	7	16.67	4	9.52	7	16.67
	Type II error	3	7.14	9	21.43	10	23.81	7	16.67

Note that # denotes the number of objects classified into each category and that % shows the percent of objects classified into each category. Type I error occurs when a bankrupt firm is classified as a non-bankrupt firm while Type II error occurs when a non-bankrupt firm is classified as a bankrupt firm.

The following procedure identifies the best model (boldfaced) for a subset by exercising two criteria: (1) the highest *prediction* rate in terms of the test set results and (2) if there was a tie among the model results, the parsimonious model, the one with the least number of hidden nodes, was chosen as the best model. In Table 2, for example, where the 2-year data sets are used, the best model for Subset 1 is the one-hidden node model. In fact, the correct prediction rate for Subset 1 is highest at 73.81% in the one-, the three-, or the five-hidden node models (when was the highest prediction rate criterion applied). But with the use of the parsimony criterion, the one-hidden node model is one that shows not only the highest correct prediction rate but is also the most economic model.

It is interesting to see that in Tables 2 and 3, the best model for each subset is either the one- or the two-hidden node model except the eight-hidden node model for Subset 4 in Table 3. The neural network studies focusing on the issue of tradeoff between sample size and model selection (Booth et al., 1992) suggest that as the size of data increases, the algorithm tends to use more complex models such as adding more hidden nodes. In other words, for a small sample size, the simpler models are usually preferred (Booth et al., 1992). Though research findings in this area may not be conclusive at the present time, the findings of this study also support the above argument that the model selection is constrained by the sample size being used. Training objects of this study are less than 90 and thus, simpler models such as one- or two-hidden node models are selected as the best models as shown above.

Prediction rates vary considerably across the four subsets. In Table 2, the correct prediction rates for the test sets are 73.81% (Subset 1), 71.43% (Subset 2), 78.57% (Subset3) and 59.52% (Subset 4). In Table 3, the correct prediction rates for the test set are 71.43% (Subset 1 through Subset 3) and 66.67% (Subset 4). The prediction accuracy for the 2-year data set is greater for two subsets, equal for one subset, and lower for the remaining subset. From these results, we cannot conclusively state that the BP network identifies and differentiates the 2-year data sets better than the 3-year data sets.

4.2. The Kohonen self-organizing feature map

Figs. 3–6 give present selected portions of the performance results of the Kohonen self-organizing feature map. Note first that each of these Kohonen self-organizing maps is shown in two-dimensional space. Two to four output clusters are actually applied for this study.

The 2-year data set is used to draw Fig. 3 (classification result) and Fig. 4 (prediction results). The 3-year data set is

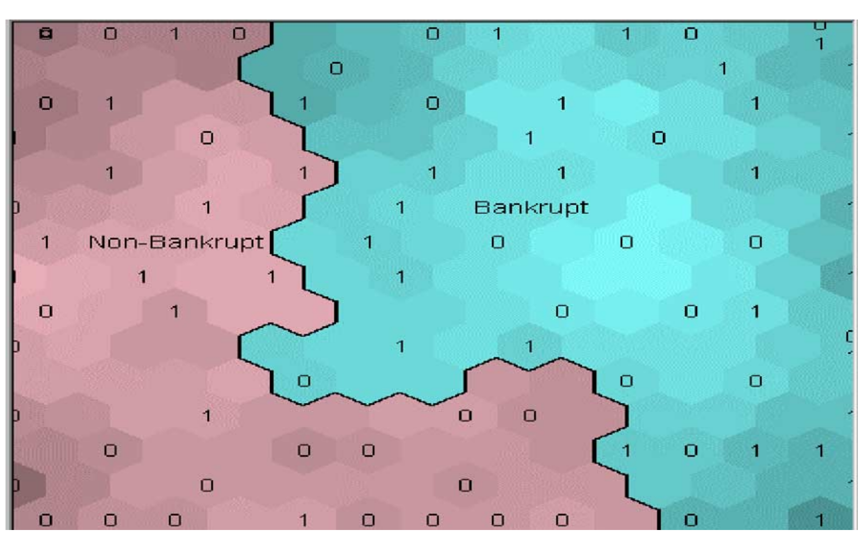

Fig. 3. Performance results of the Kohonen network for the Subset 2 (training set) in 2-year.

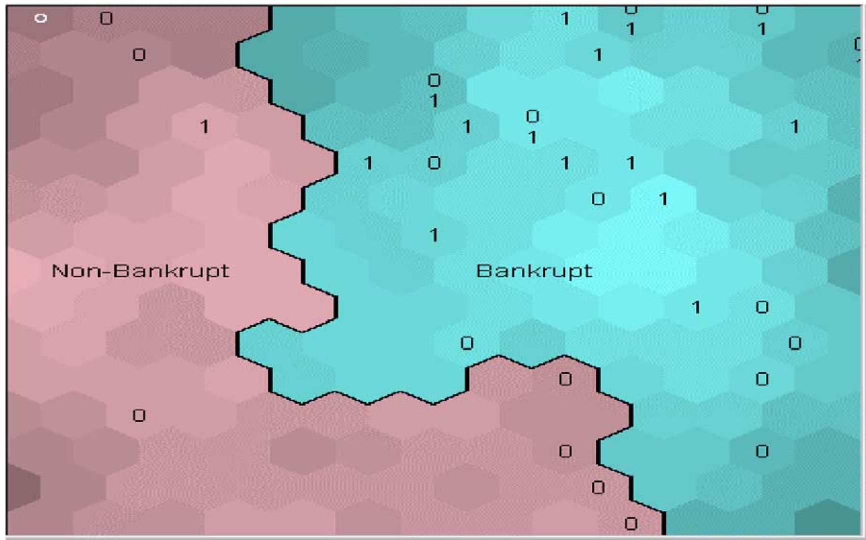

Fig. 4. Performance results of the Kohonen network for the Subset 2 (test set) in 2-year.

used for Fig. 5 (classification results) and Fig. 6 (prediction results). In Figs. 3–6, the label 1 denotes a bankrupt firm and the label 0 a non-bankrupt firm. The naming of a cluster is as follows. A cluster is named 'bankrupt' if it contains more bankrupt firms than non-bankrupt ones. Otherwise it is named a non-bankrupt group.

Table 4 portrays the experimental results of the Kohonen self-organizing feature maps in tabular form. When the 2-year data sets are used, the correct classification rates for the training sets range from 64.29 to 73.81% while the correct predication rates for the test sets run from 54.76 to 66.67%. When the 3-year data sets are used, the correct classification rates of the Kohonen self-organizing feature maps for the training sets are in between 66.67 and 70.24% while the correct prediction rates for the test sets are in the range of 52.38–76.19%.

In a bankruptcy prediction study, early 'detection' is very important. That is why we test the 2-year versus the 3-year data sets using neural networks or statistical classifiers, to see if either methodology can detect any early distress call. In this Kohonen experiment, it is difficult to conclude that the Kohonen network differentiates the 2-year data set from the 3-year.

4.3. Statistical models

First, discriminant analysis (DA) is performed. Since the assumption of the equal covariance matrices is not met in this paired bankruptcy study, quadratic DA is conducted and the performance results are presented in Table 5. Discriminant analysis is implemented using SAS procedure DISCRIM. In all cases, the prior probability proportional to group size option was used. Where the 2-year data sets are used, the correct classification rates for the training sets ranged from 64.29 to 70.24%. The correct prediction rates for the test sets are between 54.76 and 66.67%.

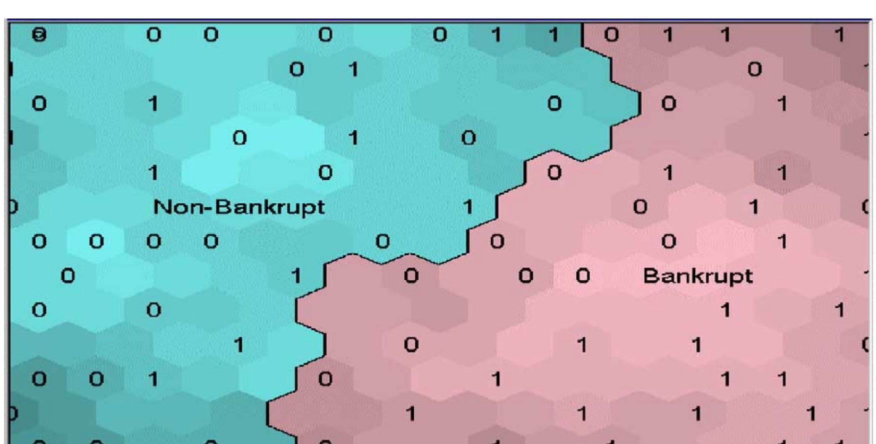

Fig. 5. Performance results of the Kohonen network for Subset 3 (training set) in 3-year.

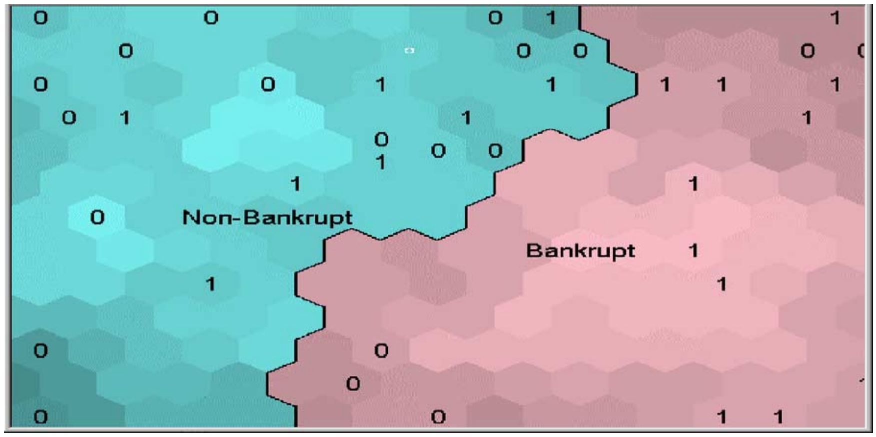

Fig. 6. Performance results of the Kohonen network for Subset 3 (test set) in 3-year.

Where the 3-year data sets are used, the correct classification rates for the training sets go from 60.71 to 73.81% and the correct prediction rates for the test sets are in between 61.90 and 64.29%.

Second, the performance results of logistic regression are shown in Table 6. Logistic regression is implemented using SAS procedure LOGISTIC. For the 2-year data sets, the correct classification rates for the training sets range from

Table 4
Performance results of the Kohonen self-organizing feature map

Years of data used	Type		Subset 1		Subset 2		Subset 3		Subset 4	
			#	%	#	%	#	%	#	%
2	Training	Correct classification	58	69.05	54	64.29	56	66.67	62	73.81
		Type I error	12	14.29	17	20.24	22	26.19	13	15.48
		Type II error	14	16.67	13	15.48	6	7.14	9	10.71
	Test	Correct prediction	28	66.67	27	64.29	28	66.67	23	54.76
		Type I error	1	2.38	14	33.33	13	30.95	7	16.67
		Type II error	13	30.95	1	2.38	0	0.00	12	28.57
3	Training	Correct classification	59	70.24	58	69.05	56	66.67	59	70.24
		Type I error	12	14.29	14	16.67	15	17.86	14	16.67
		Type II error	13	15.48	12	14.29	13	15.48	11	13.10
	Test	Correct prediction	32	76.19	25	59.52	28	66.67	22	52.38
		Type I error	4	9.52	15	35.71	5	11.90	9	21.43
		Type II error	6	14.29	2	4.76	9	21.43	11	26.19

Note that # denotes the number of objects classified into each category and that % shows the percent of objects classified into each category.

Table 5
Performance results of the quadratic discriminant analysis

Years of data used	Type		Subset 1		Subset 2		Subset 3		Subset 4	
			#	%	#	%	#	%	#	%
2	Training	Correct classification	59	70.24	56	66.67	54	64.29	56	66.67
		Type I error	16	19.05	19	22.62	17	20.24	20	23.81
		Type II error	9	10.71	9	10.71	13	15.48	8	9.52
	Test	Correct prediction	25	59.52	28	66.67	28	66.67	23	54.76
		Type I error	8	19.05	12	28.57	10	23.81	12	28.57
		Type II error	9	21.43	2	4.76	4	9.52	7	16.67
3	Training	Correct classification	62	73.81	51	60.71	51	60.71	54	64.29
		Type I error	15	17.86	24	28.57	26	30.95	25	29.76
		Type II error	7	8.33	9	10.71	7	8.33	5	5.95
	Test	Correct prediction	26	61.90	27	64.29	27	64.29	26	61.90
		Type I error	10	23.81	14	33.33	13	30.95	14	33.33
		Type II error	6	14.29	1	2.38	2	4.76	2	4.76

Note that # denotes the number of objects classified into each category and that % shows the percent of objects classified into each category.

Table 6
Performance results of the logistic regression

Years of data used	Type		Subset 1 #	Subset 1 %	Subset 2 #	Subset 2 %	Subset 3 #	Subset 3 %	Subset 4 #	Subset 4 %
2	Training	Correct classification	64	76.19	56	66.67	56	66.67	62	73.81
		Type I error	8	9.52	12	14.29	13	15.48	9	10.71
		Type II error	12	14.29	16	19.05	15	17.86	15	17.86
	Test	Correct prediction	24	57.14	29	69.05	33	78.57	27	64.29
		Type I error	8	19.05	10	23.81	8	19.05	6	14.29
		Type II error	10	23.81	3	7.14	1	2.38	9	21.43
3	Training	Correct classification	59	70.24	52	61.90	52	61.90	56	66.67
		Type I error	12	14.29	15	17.86	17	20.24	10	11.90
		Type II error	13	15.48	17	20.24	15	17.86	18	21.43
	Test	Correct prediction	26	61.90	26	61.90	28	66.67	26	61.90
		Type I error	5	11.90	10	23.81	9	21.43	8	19.05
		Type II error	11	26.19	6	14.29	5	11.90	8	19.05

Note that # denotes the number of objects classified into each category and that % shows the percent of objects classified into each category.

66.67 to 76.19% while the correct prediction rates for the test sets are in between 57.14 and 78.57%. For the 3-year data sets being used, the range of the correct classification rates for the training sets give 61.90–70.24% while the correct prediction rates for the testing sets go between 61.90 and 66.67%.

4.4. Performance comparisons among classification techniques

Table 7 provides a summary of performance comparison among the four methodological choices used in this bankruptcy prediction study: the two neural networks, the BP networks and the Kohonen self-organizing feature maps, and of the two statistical classifiers, QDA and logistic regression. In Table 7, the best *classification* model (training group) for each subset, is identified by *underlining*. As expected, the BP networks show the highest classification accuracy. Among the total eight sets, seven from the BP network, one from the Kohonen network, and another from QDA are selected as the best classification models. Note that there is a tie in the classification result between the BP and the Kohonen in Subset 3 in the 3-year data set being used.

Again, the best *prediction* model (test group) for each subset is identified in the bold-faced character. The BP networks show the best prediction capabilities across the sample variability. Among the total eight sets, six from the BP network, two from the logistic regression, and one from the Kohonen network are recognized as the best prediction

Table 7
Summary of performance results for training and test firms organized by subsets and by the number of years of data used before bankruptcy filing

Years of data used	Subset	Type	BP #	BP %	Kohonen #	Kohonen %	QDA #	QDA %	Logistic #	Logistic %
2	Subset 1	Training	66	78.57	58	69.05	59	70.24	64	76.19
		Test	**31**	**73.81**	28	66.67	25	59.52	24	57.14
	Subset 2	Training	64	76.19	54	64.29	56	66.67	56	66.67
		Test	**30**	**71.43**	27	64.29	28	66.67	29	69.05
	Subset 3	Training	59	70.24	56	66.67	54	64.29	56	66.67
		Test	**33**	**78.57**	28	66.67	28	66.67	**33**	**78.57**
	Subset 4	Training	69	82.14	62	73.81	56	66.67	62	73.81
		Test	25	59.52	23	54.76	23	54.87	**27**	**64.29**
3	Subset 1	Training	61	72.62	59	70.24	62	73.81	59	70.24
		Test	30	71.43	**32**	**76.19**	26	61.90	26	61.90
	Subset 2	Training	59	70.24	58	69.05	51	60.71	52	61.90
		Test	**30**	**71.43**	25	59.52	27	64.29	26	61.90
	Subset 3	Training	56	66.67	56	66.67	51	60.71	52	61.90
		Test	**30**	**71.43**	28	66.67	27	64.29	28	66.67
	Subset 4	Training	62	73.81	59	70.24	54	64.29	56	66.67
		Test	**28**	**66.67**	22	52.38	26	61.90	26	61.90

The Year column indicates the number of years of data used prior to bankruptcy filing in developing estimates. The best classification model (for training) for each subset is identified in the underlined character while the best prediction model (for test) is shown in the bold-faced character. Note that # column shows the correct number of object classified and % column shows the percentage of the correct number of object classified. BP refers to the back-propagation network and QDA represents quadratic discriminant analysis.

Table 8
Effect of data set year on performance accuracy

Type	2-year Mean	3-year Mean	F-value	P-value
Training	59.44	56.69	3.45	0.0729
Test	27.63	27.31	0.10	0.7544

models. Note again that there is a tie between the BP network and the logistic model in Subset 3 when the 2-year data set is used.

The effect of data set year is tested in Table 8. It gives the SAS ANOVA results for both training and test sets. This test is performed to see whether there is a performance difference in the two different data sets, namely, the 2-year or the 3-year data sets prior to bankruptcy filing. In the training sets, there is a performance difference at the 10% significance level. However, in test sets, ANOVA results indicate that there is no significant difference using either of these two different timed data sets. In other words, the year-effect does not seem to be a critical factor influencing the prediction capabilities of the classification techniques compared here. One speculation for no year effect presented is that for most bankrupt firms in these particular data sets, financial deterioration had already begun well before the 3 fiscal years, and thus, no financial strength difference is detected between the 2- or 3-year data sets being used.

Effectiveness of each individual technique tested is shown in Table 9. The table gives ANOVA results with the Duncan Option for both training and test sets. Letters in the Duncan grouping columns show the group to which the mean of each group belongs. Different letters in the Duncan Option indicate that groups are significantly different at a 5% level. In other words, the same letters in the Duncan grouping mean that they can be grouped statistically together. Table 9 shows that the BP network seems to be the most accurate model. The runner-up is the logistic regression as shown in the test section of the Duncan grouping. The performance accuracy of the Kohonen network and DA shows that they are not as accurate as the BP network for both training and test. But their prediction rates are comparable to logistic regression.

Table 9
Results of Duncan's multi-group test for training and test firms

Set	Technique used	Means	Duncan grouping
Training	BP	62.000	C
	Logistic regression	57.125	D
	Kohonen	57.750	D
	Discriminant analysis	55.375	D
Test	BP	29.625	C
	Logistic regression	27.375	C, D
	Kohonen	26.625	D
	Discriminant analysis	26.250	D

Similar letters in the Duncan grouping column indicates that the groups are statistically similar at the 5% level.

Back in Table 7, discriminant analysis does not seem to show a comparable performance to the remaining three other techniques, and thus its presence might obscure some statistical tests. For this reason, we drop the discriminant analysis method in the following ANOVA test, which focused on the prediction (test) performance of the remaining three individual classification techniques. A one-way ANOVA with one repeated-measure design is used to compare the possible performance (test) rate differences among the three classification techniques.

The null hypothesis is that there is no difference in the prediction rates of the three classification techniques. Since the P-value is 0.0434 ($F_{2,14}=3.96$) of the one-way ANOVA with one repeated-measure design, our test result rejects the null hypothesis of no differences in the mean levels of prediction rates among the three classification techniques. Thus, we performed the paired comparison, between methods: (1) the BP versus the Kohonen, (2) the Kohonen versus the logistic regression, and finally (3) the logistic regression versus the BP networks.

Our results show that prediction rate difference between the BP network and the Kohonen network at about 1% level (F-value$=11.45$). Also, the performance rates between the BP network and the logistic regression differ nearly at a 5% significant level (F-value$=5.30$). It means that the BP networks show the best performance results among the three classification tools. Finally, there is no significant difference between Kohonen network and the logistic regression. When the null hypothesis of no differences in the mean levels of the prediction accuracy among the three classification techniques is tested, the P-value is 0.0434 (F-value$=3.96$). The results of this study confirm the findings of the previous literature that the BP networks provide a better mapping function for bankruptcy indication (Berardi, 1988; Tam and Kiang, 1992; Zhang et al., 1999).

5. Conclusions

The main purpose of this study is to investigate two different training (or learning) types of neural networks using their representative networks—the BP network (supervised) versus the Kohonen self-organizing feature map (unsupervised)—in terms of their performance accuracy in the area of bankruptcy prediction. Discriminant analysis and logistic regression have also been introduced to give some performance benchmarks for the neural network classifiers.

The test sample for this study is from the Korean listed companies. The findings of this study can be summarized as follows: the impact of data set size in neural network experiment is particularly important. It should be noted that training data sets (84 paired objects) used in this study is indeed a small one. Usually, BP networks provide a good posterior probability when they have enough objects to be learned. It is because the neural network paradigm is, in essence, a data driven non-parametric approach. However, we show that even with the small sample size, the BP network consistently outperforms the logistic regression as well as other classification techniques.

One problem of the supervised approach is that within its retrospective mode, researchers often have to describe and explain experiments with past events. That is, their findings cannot be readily generalized to real-time or future mode, especially when the underlying business environment is radically changing. For this reason, we test how much tolerance, in terms of classification accuracy, we may get by using the unsupervised approach, such as the Kohonen self-organizing feature map, rather than the accuracy-based supervised neural network approach. Our findings confirm that the prediction accuracy of the Kohonen self-organizing feature map, as expected, is lower than the other supervised classification techniques. Though having observed the discrepancy the prediction accuracy of these classification tools, it is, indeed, up to practitioners to choose classification techniques for their problems at hand. But it is shown in this study that the Kohonen self-organizing feature map could be used as an alternative classification tool. The realization that such supervised classification techniques cannot provide on-line real time response and that a high accuracy with no-time-relevancy study, often the case of the supervised studies, may provide no additional contribution in reality, gives some weight to the usefulness of the Kohonen unsupervised neural network. We also learn that the Kohonen self-organizing feature map has some disadvantages: low accuracy, the decision about number of clusters to be included, and the identification of cluster characteristics when exposed to classification tasks.

5.1. Limitations and future research directions

There are a variety of neural networks available for pattern classification tasks. Each of these neural network classifiers has its own advantages and disadvantages based on its algorithms and architectures. In this study, only two neural network types, the BP and Kohonen networks, are compared in bankruptcy prediction. We must manage the speed of this fast changing track. The direct comparison between the BP network (supervised) and the pure Kohonen network (unsupervised) may not be a good experimental construct. However, in so doing, we have identified the limitations and feasibility of the Kohoen networks in classification settings.

Another way of using the Kohonen self-organizing neural network is tracking the financial condition of a firm over time. That is, each sample or population is plotted on the Kohonen map initially, and over a certain period of time, the position of a firm in the map is kept in track while the characteristic of each of the clusters is evaluated. The experimental prototype of this approach was given by Martin-del-Brio and Serrano-Cinca (1995). A major advantage of this approach is that we can examine the comparative financial condition of a firm in a real-time base. This method does not provide a high prediction rate, but it is more practical to detect a failing firm in a more timely fashion, the trait that is especially important in a fast changing business environment. Since the effectiveness of this kind of approach could be great when we have a dataset that contains a longer period of time than just 2–3 years ahead of bankruptcy filing, it is difficult to speculate on this aspect of the Kohonen network in this study. But our research team has already launched this real-time-based pattern research that will be shown later. This unsupervised type of neural network is a future-oriented approach. With this unsupervised network, monitoring a firm's track with its surroundings (which is represented by a cluster) can be evaluated at the same time. It implies that this clustering technique is more adaptable and possibly more useful in today's fast changing environment.

References

Alam, P., Booth, D., Lee, K., & Thordarson, T. (2000). The use of fuzzy clustering algorithm and self organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems With Applications*, *18*, 185–199.

Altman, E. L. (1968). Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. *Journal of Finance*, *23*(3), 589–609.

Aupperle, K. E., Acar, W., & Booth, D. (1986). An empirical criticism of in search of excellence: How excellent are the excellent companies. *Journal of Management*, *12*(4), 499–512.

Beaver, W. (1966). Financial ratios as predictors of failure, empirical research in accounting, selected studies 1966. *Journal of Accounting Research*, *4*, 71–111.

Berardi, V. (1988). *An Investigation of neural network ensemble methods for posterior probability estimation*. Unpublished PhD dissertation, Kent State University.

Booth, D. (1982). The analysis of outlying data points by robust regression: A multivariate problem bank identification model. *Decision Sciences*, *13*, 72–81.

Booth, D., Alam, P., Ahkam, S. N., & Osyk, B. (1989). A robust multivariate procedure for the identification of problem savings and loan institutions. *Decision Sciences*, *20*(2), 320–330.

Booth, D., Khouja, M., & Hu, M. (1992). A robust multivariate procedure for industrial robot evaluation and selection. *International Journal of Operations and Production Management*, *12*(2), 15–25.

Boritz, J. E., & Kennedy, D. B. (1995). Effectiveness of neural network types for Prediction of Business Failure. *Expert Systems With Applications*, *9*(4), 503–512.

Corridoni, J. M., Del Bimbo, A., & Landi, L. (1996). 3D object classification using multi-object Kohonen networks. *Pattern Recognition*, *29*(6), 919–935.

Der Voort, M., Dougherty, M., & Watson, S. (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, *4*(5), 307–318.

Deschenes, C. J., & Noonan, J. (1995). Fuzzy Kohonen network for the classification of transients using the wavelet transform for feature extraction. *Information Sciences*, *87*(4), 247–266.

Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, *2*, 189–192.

Grznar, J., Booth, D., & Sebastian, P. (1997). The use of robust smoothers in nuclear material safeguards. *Journal of Chemical Information and Computer Science*, *37*, 236–240.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis with reading* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Hamburg, J. H., Booth, D., & Weinroth, G. J. (1996). A neural network approach to the detection of nuclear material losses. *Journal of Chemical Information and Computer Sciences*, *36*, 544–553.

Hamburg, J. H., Booth, D., & Weinroth, G. J. (1997). A neural network approach to monitoring the quality of production processes. *International Journal of Applied Quality Management* .

Hinton, G., & Sejnowski, T. (Eds.). (1999). *Unsupervised learning: foundations of neural computation*. Cambridge, MA: The MIT Press.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multi-layer feedforward networks are universal approximators. *Neural Networks*, *2*, 359–366.

Jo, H. Y., Han, I. G., & Lee, H. Y. (1997). Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. *Expert Systems With Applications*, *13*(2), 97–108.

Kelly, K. (1994). *Out of control, The new biology of machines, social systems, and the economic world*. Reading, MA: Perseus Books.

Kiviluoto, K. (1998). Predicting bankruptcies with the self-organizing map. *Neurocomputing*, *21*, 191–201.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

Kohonen, T. (1997). *Self-organizing maps* (2nd ed.). Berlin: Springer.

Lenard, M. J., Alam, P., & Booth, D. (2000). An analysis of fuzzy clustering and a hybrid model for the auditor's going concern assessment. *Decision Sciences*, *31*(4), 861–884.

Lenard, M. J., Alam, P., & Madey, G. R. (1995). The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sciences*, *26*(2), 209–227.

Martin-del-Brio, B., & Serrano-Cinca, C. (1995). Self-organizing neural networks: the financial state of Spanish companies. In Refenes (Ed.), *Neural network in the capital markets*, 341–357.

MathWorks. (1997). *Matlab: Neural network toolbox, User's Guide*, The MathWorks, Inc., 24 Prime Park Way, Natick, MA 01760.

Nour, M.A. (1994). *Improved clustering and classification algorithms for the Kohonen self-organizing neural network*. Unpublished PhD dissertation, Kent State University.

Odom, M., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *Proceedings of the IEEE international conference on neural networks* , 163–168.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131.

O'Leary, D. E. (1998). Using neural networks to predict corporate failure. *International Journal of Intelligent Systems in Accounting, Finance and Management*, *7*, 187–197.

Refenes, A. P. (Ed.). (1995). *Neural networks in the capital market* (pp. 3–98). New York: Wiley. Chapters 1–6.

Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: a new tool for predicting thrift failures. *Decision Sciences*, *23*, 899–916.

Schonweiler, R., Kaese, S., Moller, S., Rinscheid, A., & Ptok, M. (1996). Classification of spectrographic voice patterns using self-organizing neuronal networks (Kohonen maps) in the evaluation of the infant cry with and without time-delayed feedback. *International Journal of Pediatric Otorhinolaryngology*, *38*(2).

Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems*, *17*, 227–238.

Sharda, R., & Wilson, R. L. (1996). Neural network experiments in business-failure forecasting: predictive performance measurement issues. *International Journal of Computational Intelligence and Organization*, *1*(2), 107–117.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management Science*, *38*(7), 926–947.

Tsukuda, J., & Baba, S. I. (1994). Predicting Japanese corporate bankruptcy in terms of financial date using neural network. *Computers and Industrial Engineering* , 445–448.

Udo, G. (1993). Neural network performance on the bankruptcy classification problem. *Computers and Industrial Engineering*, *25*, 377–380.

Waldrop, M.M. (1992). *Complexity*. Touchstone, Rochefeller Center, 1230 Avenue of the Americas, New York, NY 10020: Simon and Schuster, Inc.

Wasserman, P. D. (1989). *Neural computing: Theory and practice*. New York: Van Nostrand Reinhold.

Wilson, R. L., & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, *11*, 545–557.

Xin-Hua, S., & Hopke, P. K. (1996). Kohonen neural network as a pattern recognition method based on the weight interpretation. *Analytica Chimica Acta*, *336*, 57–66.

Zhang, G. (1998). Linear and nonlinear times-series forecasting with artificial neural networks, Unpublished PhD dissertation, Kent State University.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. *European Journal of Operational Research*, *116*, 16–32.