Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 2639–2649
Expert Systems with Applications
www.elsevier.com/locate/eswa

Using neural network ensembles for bankruptcy prediction and credit scoring

Chih-Fong Tsai *, Jhen-Wei Wu

Department of Accounting and Information Technology, National Chung Cheng University, Taiwan

Abstract

Bankruptcy prediction and credit scoring have long been regarded as critical topics and have been studied extensively in the accounting and finance literature. Artificial intelligence and machine learning techniques have been used to solve these financial decision-making problems. The multilayer perceptron (MLP) network trained by the back-propagation learning algorithm is the mostly used technique for financial decision-making problems. In addition, it is usually superior to other traditional statistical models. Recent studies suggest combining multiple classifiers (or classifier ensembles) should be better than single classifiers. However, the performance of multiple classifiers in bankruptcy prediction and credit scoring is not fully understood. In this paper, we investigate the performance of a single classifier as the baseline classifier to compare with multiple classifiers and diversified multiple classifiers by using neural networks based on three datasets. By comparing with the single classifier as the benchmark in terms of average prediction accuracy, the multiple classifiers only perform better in one of the three datasets. The diversified multiple classifiers trained by not only different classifier parameters but also different sets of training data perform worse in all datasets. However, for the Type I and Type II errors, there is no exact winner. We suggest that it is better to consider these three classifier architectures to make the optimal financial decision.

Keywords: Bankruptcy prediction; Credit scoring; Neural networks; Classifier ensembles

1. Introduction

To predict business failure accurately is a very important issue in financial decision-making. Wrong decision-making in financial institutions can cause important consequences, e.g. financial crises or distress. Two well-known issues in financial decision-making are bankruptcy prediction and credit scoring.

Bankruptcy prediction and credit scoring have long been regarded as critical topics and have been studied extensively in the accounting and finance literature. The main impacts of such research are in lending decisions and profitability of financial institutions. Before extending a loan, banks need to predict the possibility of failure of the potential counterparty. Thus, predicting bankruptcy timely and correctly has become great importance for financial institutions (Atiya, 2001; Zhang, Hu, Patuwo, & Indro, 1999).

With the rapid growth in credit industry and the management of large loan portfolios, credit scoring models have been extensively used for the credit admission evaluation. The credit scoring models are developed to classify loan customers as either a good credit group (accepted) or a bad credit group (rejected) with their related characteristics such as age, income and martial status or based on the data of the previous accepted and rejected applicants (Chen & Huang, 2003). The benefits of using credit scoring include reducing the cost of credit analysis, enabling faster decision, insuring credit collections, and diminishing possible risk (West, 2000). A slight improvement in credit scoring accuracy might reduce large credit risk and translate into significant future saving.

Financial decision-making such as bankruptcy prediction and credit scoring described above, can be regarded as the binary classification problem of classifying an observation

* Corresponding author. Tel.: + 88 652720411x34519; fax: +88 652721197.
E-mail address: actcft@ccu.edu.tw (C.-F. Tsai).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.05.019

into one of the two pre-defined groups (in the bankruptcy prediction case, bankruptcy or non-bankruptcy). Artificial intelligence and machine learning techniques (e.g. artificial neural networks (ANN), decision trees (DT), support vector machines (SVM), etc.) have been used to solve the above financial decision-making problems (e.g. Atiya, 2001; Huang, Chen, Hsu, Chen, & Wu, 2004; Lee, Chiu, Chou, & Lu, 2006).

According to previous studies, they show that machine learning techniques are superior to that of traditional (statistical) methods in dealing with bankruptcy prediction and credit scoring problems, especially in nonlinear pattern classification (Huang et al., 2004; Ong, Huang, & Tzeng, 2005; Vellido, Lisboa, & Vaughan, 1999; Wong & Selvi, 1998). In particular, the neural network model trained by the back-propagation learning algorithm is the most popular tool used for financial decision-making problems, whose prediction accuracy outperforms than other models, such as logistic regression (LR), linear discriminant analysis (LDA), multiple discriminant analysis (MDA), k-nearest neighbor (k-NN), decision trees, etc. This indicates that choosing learning model/classifier is one major factor affecting the classification or prediction result. In this paper, we employ the multilayer perceptron neural network trained by the back-propagation learning algorithm as the baseline classifier to compare with multiple neural network classifiers.

Much related work focuses on identifying the single best model for a given financial decision-making problems. This reliance on a single model may be misguided. In West, Dellana, and Qian (2005) "multiple experts" (i.e. ensembles) of predictors have demonstrated the potential to reduce the generalization error of a single model from 5% to 70%. In order words, "multiple classifiers" may provide more accurate prediction results than "single classifiers". However, the performance of using multiple classifiers in the binary classification financial decision-making problems is not fully understood. Therefore, there are two research questions as the aim of this paper.

- Do multiple neural network classifiers outperform the single best neural network classifier in terms of predication accuracy based on a number of datasets?
- By considering the Type I and Type II errors, what kind of neural network classifiers provide the lowest prediction errors?

The organization of this paper is as follows. Section 2 describes the concept of pattern classification and application of multiple classifiers, with a particular attention given to artificial neural networks. Section 3 compares related work in bankruptcy prediction and credit scoring by using machine learning techniques. In Section 4, the experiments are based on comparing the performance of single and multiple classifiers in terms of average prediction accuracy and the type I and type II errors. Finally, the conclusion is made in Section 5.

2. Artificial neural networks and multiple classifiers

2.1. Pattern classification

Pattern classification considers assigning a label to an input. In general, pattern classification is the problem to classify given patterns into several classes. After finding a set of classes, the input represented by a number of features is allocated to the correct class. The general model first determines the class, and then observations are obtained regarding the class. Finally, the model attempts to assign the correct class to the input based on the observations. When defining classes, one can state explicit rules. However, it is better to define through training examples.

There are two approaches regarding how pattern classification is done. One is "decision-theoretic approach." In this approach, the pattern is represented as a feature vector in a feature space and then a decision algorithm is used to decide which class the pattern belongs to. Another one is "structural approach." In this approach, the pattern is represented by its structure, e.g. a graph connecting the primary elements, etc. Subsequently, it uses parsing (grammatical) or graph matching to perform pattern classification (Witten & Frank, 2000).

2.2. Artificial neural networks

A neural network (or an artificial neural network) (Haykin, 1999) is an information processing paradigm that is inspired by the way of biological nervous systems, such as the brain to process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. Neural networks, like people, learn by examples. That is, neural networks learn by experience, generalize from previous experiences to new ones, and can make decisions.

The most common type of neural networks consists of three layers of units: input layers, hidden layers, and output layers. It is called multilayer perceptron (MLP). A layer of "input" units is connected to a layer of "hidden" units, which is connected to a layer of "output" units. The activity of the input layers represents the raw information that is fed into the network. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units. The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units. Fig. 1 shows an example of three-layer neural network including input, output, and one hidden layers.

In multilayer networks, the predicted outputs for each training example are calculated, and then it figures out the difference between each predicted output and the corresponding target output. The error is then adjusted so that the error rate will be reduced next time when the training example is presented to the network. Thus, the algorithm

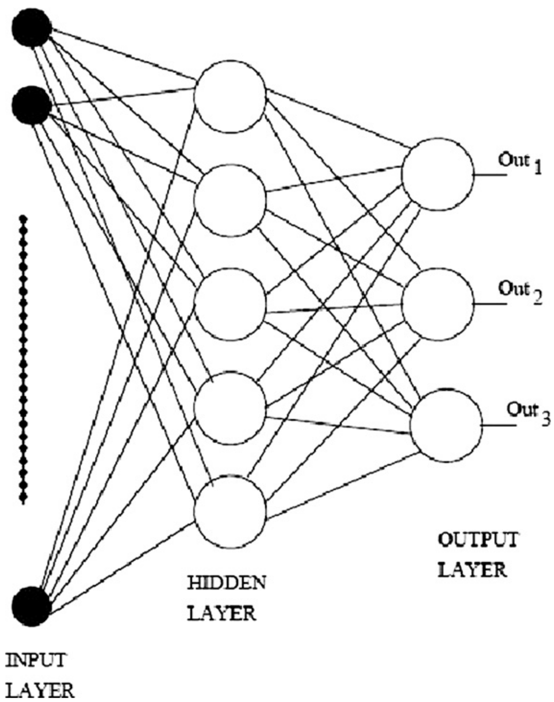

Fig. 1. The three-layer neural network.

finds out properties of the inputs that are most relevant to learning the target function.

2.3. Multiple classifiers

As an alternative to a single classifier approach, the combination of multiple classifiers has been considered in recent years. The key idea in multiple classifier systems is to combine a number of classifiers such that the resulting combined system achieves higher classification accuracy and efficiency than the original single classifiers. This is considered a property emerging from the combination of relatively simple pattern recognition devices with often limited individual performance profiles. The aim of multiple classifier systems is to design a composite system that outperforms any individual classifier by pooling together the decisions of all classifiers. The rationale is that it may be more difficult to optimize the design of a single complex classifier than to optimize the design of a combination of relatively simple classifiers (Frosyniotis, Stafylopatis, & Likas, 2003; Kang & Doermann, 2003; Roli, Kittler, & Windeatt, 2004).

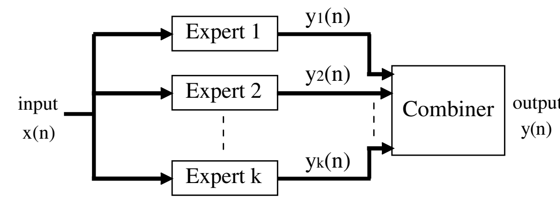

Fig. 2. Architecture of multiple classifiers (Haykin, 1999).

Combining classifiers is primarily to achieve higher classification accuracy and efficiency. It works better, depending that each learner be trained well, but different learners generalize in different ways, i.e., there is diversity in the ensemble (Ghosh, 2002). According to Kitter (Kittler, Hatef, Duin, & Matas, 1998), combining classifiers is particularly useful if they are different. This can be achieved by using many different ways, such as using different feature sets, or using different training sets, or randomly selected, or based on a cluster analysis.

The simplest method to combine classifiers is majority voting. The outputs of the several numbers of individual classifiers are pooled together. Then, the output which receives the largest number of votes is selected as the final classification decision (Kittler et al., 1998). In general, the final classification decision that reaches the majority (greater than half) of votes is taken. Fig. 2 shows the general architecture of a multiple classifier. The outputs of several different trained neural networks are combined to produce a final output.

3. Related work

This section compares recent related work using MLP for bankruptcy prediction and credit scoring as shown in Table 1.

Regarding the above comparison, some issues as the limitations of literature, are listed below.

1. Most studies only use one chosen dataset for system validation. However, only one chosen dataset may not be

Table 1
Comparisons of related work

Work	Classifier	Datasets	Evaluation methods	
			Average accuracy	Type I/II error
Fan and Palaniswami (2000)	MLP	Australian credit	Yes	No
West (2000)	MLP	Australian/German credit	Yes	No
Atiya (2001)	MLP	US credit	Yes	No
Lee et al. (2002)	MLP + LDA	Taiwan credit	Yes	Yes
Huang et al. (2004)	MLP	Taiwan/US credit	Yes	No
Min and Lee (2005)	MLP	Korea credit	Yes	No
Shin et al. (2005)	MLP	Korea credit	Yes	No
West et al. (2005)	MLP ensembles	Australian/German credit/Bankruptcy dataset	Yes	No
Lee et al. (2006)	MLP	Taiwan credit	Yes	Yes

reliable to make a conclusion. It is necessary to consider a certain number of different datasets for system validation.

2. Most studies only examine average prediction performance of their models without considering the Type I and Type II errors except Lee, Chiu, Lu, and Chen (2002) and Lee et al. (2006).
3. Very few studies consider multiple classifiers on both credit scoring and bankruptcy prediction problem domains. Only one work, i.e., West et al. (2005), uses MLP ensemble strategies. However, they only investigate average prediction accuracy.

As a result, examining the performance of multiple classifiers including average accuracy and Type I and Type II errors by using a number of datasets is the aim of this paper.

Table 2
The three datasets

	Total cases	Good/bad cases	No. of attributes
Australian credit	690	307/383	14
German credit	1000	700/300	20
Japanese credit	690	307/383	15

Table 3
Prediction accuracy

Learning epoch	Hidden nodes	Australian credit	German credit	Japanese credit
50	8	0.8867	0.7338	0.8531
	12	0.8639	0.7426	0.8632
	16	0.8643	0.7825	0.8502
	24	0.9024	0.7707	0.8378
	32	0.8462	0.7546	0.8448
100	8	0.8462	0.7735	0.8538
	12	0.8667	0.7782	0.8763
	16	0.8473	0.7738	0.8277
	24	0.9732	0.7735	0.8756
	32	0.8502	0.7733	0.8621
200	8	0.8612	0.7415	0.855
	12	0.902	0.76	0.8565
	16	0.8679	0.75	0.8794
	24	0.8702	0.7604	0.879
	32	0.8744	0.7897	0.8639
300	8	0.8604	0.7535	0.8571
	12	0.8785	0.7872	0.8656
	16	0.8626	0.7649	0.8426
	24	0.8483	0.7703	0.8586
	32	0.8776	0.786	0.879
Avg.		*0.8725*	*0.766*	*0.8591*

Table 4
Prediction accuracy based on the Australian credit dataset

Learning epoch	Hidden nodes	Single classifiers	Numbers of multiple classifiers							
			3	3	5	5	7	9	11	13	15
50	8	0.8867	V		V	V	V	V	V	V	V
	12	0.8639			V					V	V
	16	0.8643	V		V				V	V	V
	24	0.9024	V	V	V	V	V	V	V	V	V
	32	0.8462			V						
	Avg.		*0.8723*		*0.8752*						
100	8	0.8462			V						
	12	0.8667	V		V				V	V	V
	16	0.8473			V						
	24	0.9732	V	V	V	V	V	V	V	V	V
	32	0.8502	V		V						
	Avg.		*0.8766*		*0.8737*						
200	8	0.8612			V						V
	12	0.902	V	V	V	V	V	V	V	V	V
	16	0.8679			V			V	V	V	V
	24	0.8702	V		V			V	V	V	V
	32	0.8744	V		V		V	V	V	V	V
	Avg.		*0.8839*		*0.8795*						
300	8	0.8604			V						V
	12	0.8785	V		V	V	V	V	V	V	V
	16	0.8626	V		V					V	V
	24	0.8483			V						
	32	0.8776	V		V		V	V	V	V	V
	Avg.		*0.8723*		*0.8679*						
Avg.		*0.8725*		*0.8766*		*0.8723*	*0.8752*	*0.8752*	*0.8752*	*0.8752*	*0.8737*

Table 5
Prediction accuracy based on the German credit dataset

Learning epoch	Hidden nodes	Single classifiers	Numbers of multiple classifiers								
			3	3	5	5	7	9	11	13	15
50	8	0.7338			V						
	12	0.7426			V						
	16	0.7825	V		V	V	V	V	V	V	V
	24	0.7707	V		V				V	V	V
	32	0.7546	V		V						V
	Avg.		*0.7968*		*0.8018*						
100	8	0.7735			V		V	V	V	V	V
	12	0.7782	V		V	V	V	V	V	V	V
	16	0.7738	V		V		V	V	V	V	V
	24	0.7735	V		V			V	V	V	V
	32	0.7733			V			V	V	V	V
	Avg.		*0.8008*		*0.7968*						
200	8	0.7415			V						
	12	0.76	V		V						V
	16	0.75			V						
	24	0.7604	V		V					V	V
	32	0.7897	V	V	V	V	V	V	V	V	V
	Avg.		*0.8338*		*0.8258*						
300	8	0.7535			V						
	12	0.7872	V	V	V	V	V	V	V	V	V
	16	0.7649			V					V	V
	24	0.7703	V		V				V	V	V
	32	0.786	V	V	V	V	V	V	V	V	V
	Avg.		*0.7938*		*0.8038*						
Avg.		*0.766*		*0.7998*		*0.8038*	*0.7968*	*0.7988*	*0.7998*	*0.8118*	*0.8098*

4. Experiments

4.1. Study 1: Single classifiers vs. multiple classifiers

4.1.1. Experimental setup

Three financial datasets are chosen for the experiments, which are Australian credit[1], German credit[2], and Japanese credit[3]. Table 2 shows the content of these datasets.

Each dataset was divided into training and testing data randomly, in which there are 70–30% training and testing sets per dataset.

For the ANN model construction, we used the three-layer back-propagation network to train ANN. The number of nodes in the hidden layers ranges from 8 to 32 and each of the ANN classifier is constructed by four different learning epochs (50, 100, 200, and 300) as the stopping criteria for training.

For each dataset, a single classifier as the baseline is constructed based on the above parameters to compare with the multiple classifiers. The most used method to construct multiple classifiers is the voting strategy (West et al., 2005). We selected an odd number ($n = 3$, 5, 7, 9, 11, 13, and 15) as the numbers of multiple classifiers and decide the result which is counted more than half votes ($>n/2$) as the final classification output for each set of the multiple classifiers.

4.1.2. Results

Table 3 presents the results of the single ANN classifier based on different numbers of hidden nodes and learning epochs under the three datasets, respectively. For the three datasets, the best ANN classifiers provide 97.32%, 78.97%, and 87.94% accuracy, respectively.

Tables 4–6 present the performance of different numbers of multiple classifiers over the three datasets, respectively. The best prediction results are underlined under different numbers of multiple classifiers. When $n = 3$ and 5, there are two strategies to combine three and five multiple classifiers. For example, when the learning epoch is 50, the best three classifiers (hidden nodes = 8, 16, and 24) were chosen for the comparison. The second strategy of combining three multiple classifiers is based on the three best classifiers over the four different learning epochs. Therefore, combining 7, 9, 11, 13, and 15 multiple classifiers is based on the second case.

Following the performance of single and multiple classifiers presented above, Figs. 3–5 compare the best multiple

[1] <http://www.liacc.up.pt/ML/statlog/datasets/australian/australian.doc.html>
[2] <http://www.liacc.up.pt/ML/statlog/datasets/german/german.doc.html>
[3] <http://www.ics.uci.edu/~mlearn/MLRepository.html>

Table 6
Prediction accuracy based on the Japanese credit dataset

Learning epoch	Hidden nodes	Single classifiers	Numbers of multiple classifiers								
			3	3	5	5	7	9	11	13	15
50	8	0.8531	V		V						V
	12	0.8632	V		V			V	V	V	V
	16	0.8502	V		V						
	24	0.8378			V						
	32	0.8448			V						
	Avg.		*0.8591*		*0.8698*						
100	8	0.8538			V						V
	12	0.8763	V		V	V	V	V	V	V	V
	16	0.8277			V						
	24	0.8756	V		V	V	V	V	V	V	V
	32	0.8621	V		V			V	V	V	V
	Avg.		*0.8714*		*0.8729*						
200	8	0.855			V					V	V
	12	0.8565			V					V	V
	16	0.8794	V	V	V	V	V	V	V	V	V
	24	0.879	V	V	V	V	V	V	V	V	V
	32	0.8639	V		V		V	V	V	V	V
	Avg.		*0.8760*		*0.8744*						
300	8	0.8571			V				V	V	V
	12	0.8656	V		V		V	V	V	V	V
	16	0.8426			V						
	24	0.8586	V		V				V	V	V
	32	0.879	V	V	V	V	V	V	V	V	V
	Avg.		*0.8729*		*0.8729*						
Avg.		*0.8591*		*0.8729*		*0.8714*	*0.8760*	*0.8760*	*0.8760*	*0.8714*	*0.8714*

classifiers ($n = 3$–15 based on Tables 4–6) with the single best classifiers (based on Table 3) in term of prediction accuracy. Note that both three and five multiple classifiers have five different prediction results. We selected the best one out of the five results for further comparisons.

These results indicate that multiple classifiers perform better than the single best classifiers only when the German credit dataset is used. This comparative result indicates that on average the single best classifier outperform multiple classifiers over the three datasets.

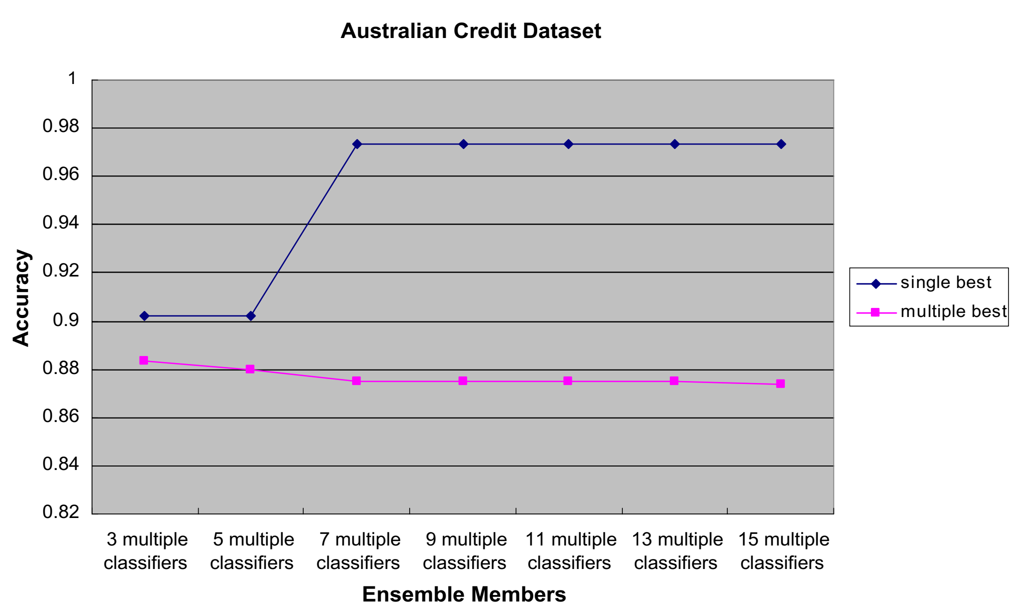

Fig. 3. Prediction accuracy based on the Australian credit dataset.

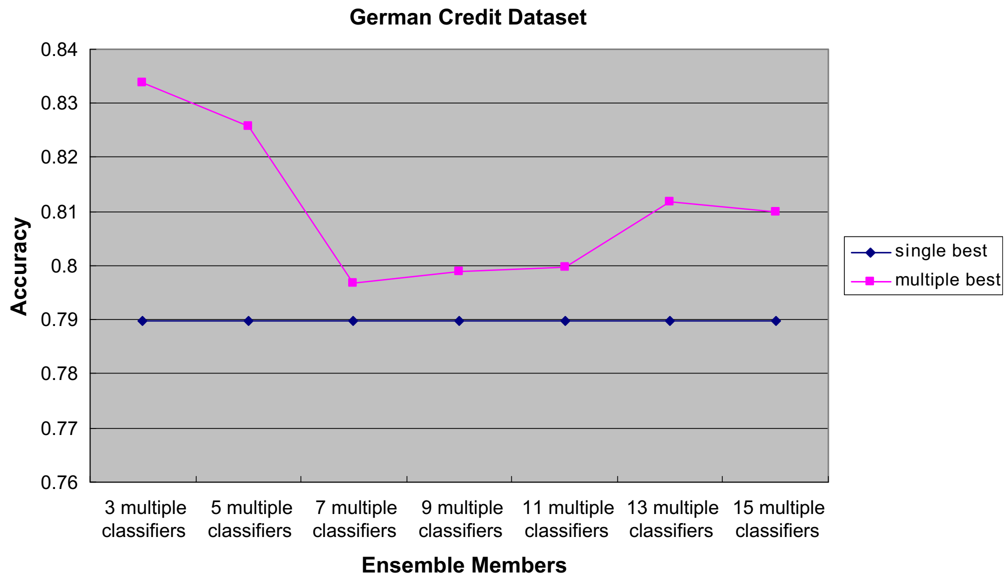

Fig. 4. Prediction accuracy based on the German credit dataset.

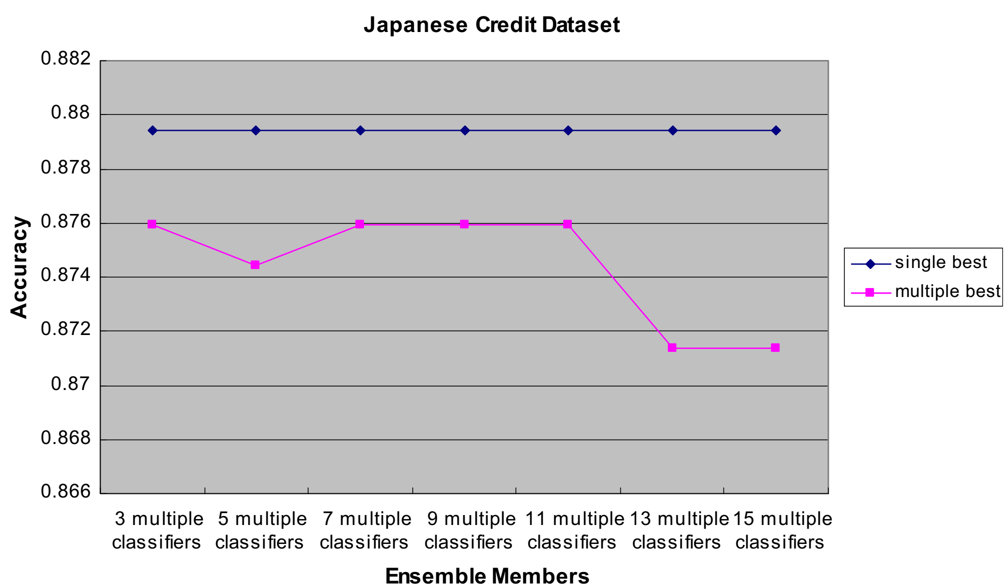

Fig. 5. Prediction accuracy based on the Japanese credit dataset.

4.2. Study 2: Single vs. multiple vs. diversified multiple classifiers

4.2.1. Experimental setup

To consider the diversity problem of designing multiple classifiers, we used different classifier parameters (numbers of hidden nodes, or different training epochs), and randomly assigned the input training examples in Study 1. However in this case, the input training data may be overlapped. In order to overcome this problem, we considered a more "*diversified*" case not only using different classifier parameters (described above) but also each of the multiple classifier was trained by different training data in the training set to make each of the combined classifiers as unique as possible. For example, when the "Australian credit" dataset (total cases = 690) is used to construct three multiple classifiers, we divided total 690 cases into three different training sets (201*3) to train the three classifiers respectively and one common testing set (87) for evaluation.

4.2.2. Results

Tables 7–9 present the average prediction accuracy of different numbers of the diversified multiple classifiers over the three datasets, respectively. The best prediction results are also underlined under different numbers of multiple classifiers.

Table 7
Prediction accuracy based on the Australian credit dataset

Learning epoch	Numbers of diversified multiple classifiers								
	3	3	5	5	7	9	11	13	15
50	*0.8708*		*0.8766*						
100	*0.8780*		*0.8664*						
200	*0.8795*		*0.8708*						
300	*0.8809*		*0.8694*						
		0.8737		*0.8679*	*0.8607*	*0.8665*	*0.8563*	*0.8665*	*0.8563*

Table 8
Prediction accuracy based on the German credit dataset

Learning epoch	Numbers of multiple classifiers/diversified multiple classifiers								
	3	3	5	5	7	9	11	13	15
50	*0.7528*		*0.7568*						
100	*0.7938*		*0.7598*						
200	*0.7448*		*0.7628*						
300	*0.7688*		*0.7628*						
		0.7538		*0.7578*	*0.7558*	*0.7988*	*0.7508*	*0.7157*	*0.7017*

Table 9
Prediction accuracy based on the Japanese credit dataset

Learning epoch	Numbers of multiple classifiers/diversified multiple classifiers								
	3	3	5	5	7	9	11	13	15
50	*0.8591*		*0.8622*						
100	*0.8621*		*0.8637*						
200	*0.8698*		*0.8530*						
300	*0.8683*		*0.8698*						
		0.8622		*0.8698*	*0.8545*	*0.8515*	*0.8254*	*0.8469*	*0.8392*

Figs. 6–8 compare the best diversified multiple classifiers (in Study 2) with the multiple classifiers and the single best classifiers (in Study 1) in term of prediction accuracy. The results show that most of diversified multiple classifiers perform worse than single best classifiers and multiple classifiers.

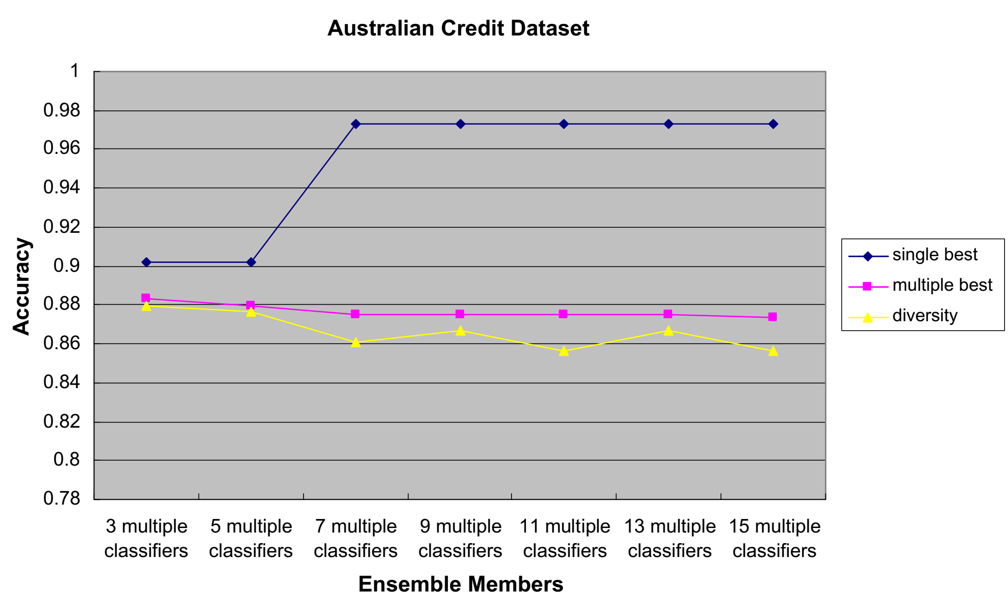

Fig. 6. Prediction accuracy based on the Australian credit dataset.

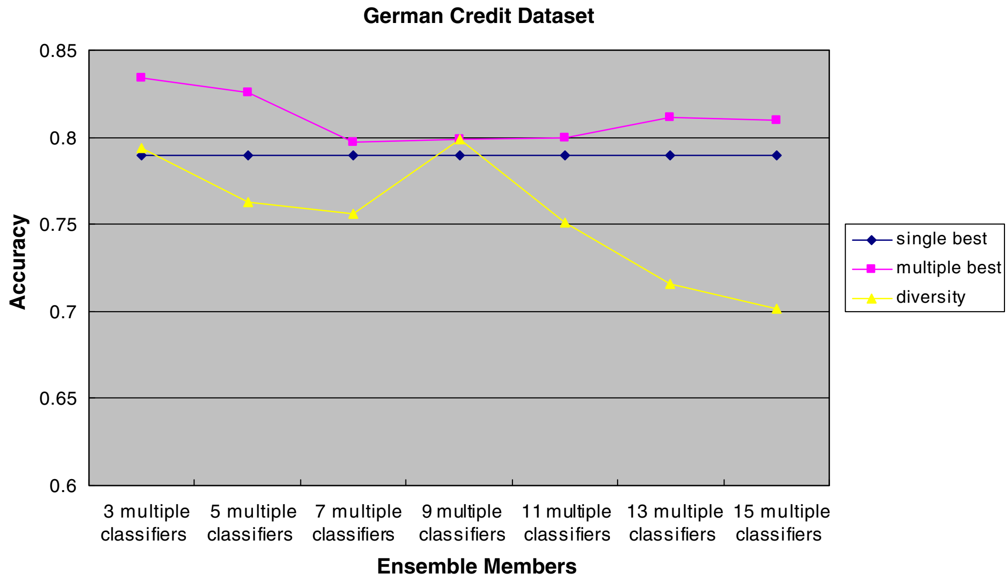

Fig. 7. Prediction accuracy based on the German credit dataset.

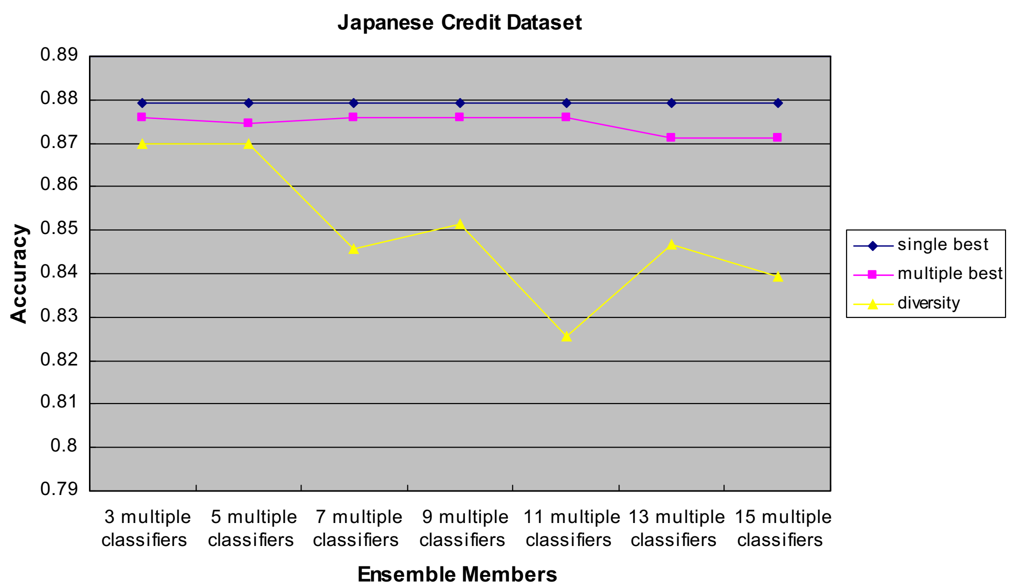

Fig. 8. Prediction accuracy based on the Japanese credit dataset.

4.3. Study 3: Type I and Type II errors

In addition to examining average prediction accuracy of the single and multiple classifiers, we compare these classifiers to examine their Type I and Type II errors, which are based on a confusion matrix shown in Table 10.

Table 10
The confusion matrix

		Predicted	
		Good credit group	Bad credit group
Actual	Good credit group		Type II error
	Bad credit group	Type I error	

- Type I error: This shows the rate of prediction errors of a model, which is to incorrectly classify the bad credit group into the good credit group.
- Type II error: Opposed to Type I error, this presents the rate of prediction errors of a model to incorrectly classify the good credit group into the bad credit group.

Table 11 presents the average error rate (%) of Type I error and Type II error over the single classifiers (S), multiple classifiers (M), and diversified multiple classifiers (D) under the three datasets. In addition, Tables 12–14 show *t*-test result of the Type I and II errors of using single, multiple, and diversified multiple classifiers over the three datasets, respectively.

Table 11
Average error rate of Type I and Type II errors

	Australian credit			German credit			Japanese credit		
	S	M	D	S	M	D	S	M	D
Type I error	*12.16*	12.85	14.28	*44.27*	45.25	59.82	15.02	15.07	*14.42*
Type II error	12.97	12.14	*11.55*	9.48	*8.46*	8.67	10.79	*10.00*	14.06

Table 12
Paired *t*-test results of type I and type II error over Australian credit dataset

Type I error/Type II error	Multiple classifiers *t*-value (significance)	Diversified multiple classifiers *t*-value (significant)
Single classifiers	−2.028 (0.062)/**2.112** (**0.053**)	−**3.173** (**0.007**)/**2.604** (**0.021**)
Multiple classifiers		−**2.265** (**0.04**)/1.090 (0.294)

Table 13
Paired *t*-test results of type I and type II error over German credit dataset

Type I error/Type II error	Multiple classifiers *t*-value (significance)	Diversified multiple classifiers *t*-value (significant)
Single classifiers	−0.401 (0.694)/1.066 (0.304)	−**4.488** (**0.001**)/0.597 (0.560)
Multiple classifiers		−**5.949** (**0.000**)/−0.312 (0.759)

Table 14
Paired *t*-test results of type I and type II error over Japanese credit dataset

Type I error/Type II error	Multiple classifiers *t*-value (significance)	Diversified multiple classifiers *t*-value (significant)
Single classifiers	0.585 (0.568)/0.977 (0.712)	1.002 (0.333)/−**2.072** (**0.057**)
Multiple classifiers		0.747 (0.468)/−**2.337** (**0.035**)

On average, the single classifiers are the winner as the best model/classifier architecture for bankruptcy prediction and credit scoring. However, when we consider the Type I and II error rates, the single classifiers do not totally outperform multiple or diversified multiple classifiers, especially for the Type II error.

Overall, there is some level of significance between these three classifiers. That is, although single classifiers perform the best in the case of average prediction accuracy, multiple classifiers or diversified multiple classifiers should not be ignored in bankruptcy prediction and credit scoring. The results indicate that while considering the credit scoring problems, the decision maker should consider not only single classifiers but also multiple classifiers and diversified multiple classifiers.

5. Conclusion

In this paper, we compared the performance of the single neural network classifier with the (diversified) multiple neural network classifiers over three datasets for the bankruptcy prediction and credit scoring problems. Theoretically, multiple classifiers should perform better than single classifiers. However, regarding the experimental results of average prediction accuracy, multiple neural network classifiers do not outperform a single best neural network classifier in many cases. In particular, the results imply that the single best neural network classifier is more suitable than multiple or diversified multiple neural network classifiers for the bankruptcy prediction and credit scoring domains. On the other hand, by examining the Type I and Type II errors of these classifiers, there is no exact winner. In this case, the decision makers should consider the combination of multiple classifiers for bankruptcy prediction and credit scoring, besides a single classifier.

Regarding the experimental results, there are two issues to be discussed that multiple classifiers do not outperform single best classifiers. First, the divided training datasets may be too little to make the multiple classifiers and diversified multiple classifiers to perform worse. Second, in the binary classification domain problem as bankruptcy prediction and credit scoring, single classifiers may be a more stable model. In other words, the multiple classifiers and diversified multiple classifiers may not perform better in the binary classification problem.

Acknowledgement

This research was partially supported by National Science Council of Taiwan (NSC 94-2416-H-194-036).

References

Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: a survey and new results. *IEEE Transactions on Neural Networks, 12*(4), 929–935.

Chen, M.-C., & Huang, S.-H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications, 24*(4), 433–441.

Fan, A., & Palaniswami, M. (2000). Selecting bankruptcy predictors using a support vector machine approach. In: *Proceedings of the international joint conference on neural networks*, Vol. 6, pp. 354–359.

Frosyniotis, D., Stafylopatis, A., & Likas, A. (2003). A divide-and-conquer method for multi-net classifiers. *Journal of Pattern Analysis and Applications, 6*(1), 32–40.

Ghosh, J. (2002). Multiclassifier systems: back to the future. In: *Proceedings of the third international workshop on multiple classifier systems*, Cagliari, Italy, June 24–26, pp. 1–15.

Haykin, S. (1999). *Neural networks: a comprehensive foundation* (2nd ed.). New Jersey: Prentice Hall.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems, 37*, 543–558.

Kang, H.-J., & Doermann, D. (2003). Evaluation of the information-theoretic construction of multiple classifier systems. In: *Proceedings of the international conference on document analysis and recognition*, Edinburgh, Scotland, Aug. 3–6, pp. 789–793.

Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers. *IEEE transactions on pattern analysis and machine intelligence, 20*(3), 226–239.

Lee, T.-S., Chiu, C.-C., Chou, Y.-C., & Lu, C.-J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics and Data Analysis, 50*, 1113–1130.

Lee, T.-S., Chiu, C.-C., Lu, C.-J., & Chen, I.-F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications, 23*, 245–254.

Min, J. H., & Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications, 28*, 603–614.

Ong, C.-S., Huang, J.-J., & Tzeng, G.-H. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications, 29*, 41–47.

Roli, F., Kittler, J., & Windeatt, T. (2004). Multiple classifier systems. In: *Proceedings of the 5th international workshop on multiple classifier systems*, Cagliari, Italy.

Shin, K-S., Lee, T. S., & Kim, H.-J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications, 28*(1), 127–135.

Vellido, A., Lisboa, P. J. G., & Vaughan, J. (1999). Neural networks in business: a survey of applications (1992–1998). *Expert Systems with Applications, 17*, 51–70.

West, D. (2000). Neural network credit scoring models. *Computers and Operations Research, 27*(11/12), 1131–1152.

West, D., Dellana, S., & Qian, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers and Operations Research, 32*, 2543–2559.

Witten, I. H., & Frank, E. (2000). *Data mining: practical machine learning tools and techniques with Java implementations*. California: Morgan Kaufmann.

Wong, B. K., & Selvi, Y. (1998). Neural network applications in finance: a review and analysis of literature (1990-1996). *Information and Management, 34*, 129–139.

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.