

## Impact of Ensemble Member Size on Confidence-based Selection in Bankruptcy Prediction

Na-Ra Kim  
School of Business,  
Ewha Womans University  
(*knara@ewhain.net*)

Kyung-Shik Shin  
School of Business,  
Ewha Womans University  
(*ksshin@ewha.ac.kr*)

Hyunchul Ahn  
School of Management Information  
Systems, Kookmin University  
(*hcahn@kookmin.ac.kr*)

.....

The prediction model is the main factor affecting the performance of a knowledge-based system for bankruptcy prediction. Earlier studies on prediction modeling have focused on the building of a single best model using statistical and artificial intelligence techniques. However, since the mid-1980s, integration of multiple techniques (hybrid techniques) and, by extension, combinations of the outputs of several models (ensemble techniques) have, according to the experimental results, generally outperformed individual models. An ensemble is a technique that constructs a set of multiple models, combines their outputs, and produces one final prediction. The way in which the outputs of ensemble members are combined is one of the important issues affecting prediction accuracy. A variety of combination schemes have been proposed in order to improve prediction performance in ensembles. Each combination scheme has advantages and limitations, and can be influenced by domain and circumstance. Accordingly, decisions on the most appropriate combination scheme in a given domain and contingency are very difficult. This paper proposes a confidence-based selection approach as part of an ensemble bankruptcy-prediction scheme that can measure unified confidence, even if ensemble members produce different types of continuous-valued outputs. The present experimental results show that when varying the number of models to combine, according to the creation type of ensemble members, the proposed combination method offers the best performance in the ensemble having the largest number of models, even when compared with the methods most often employed in bankruptcy prediction.

.....

Received : May 22, 2013      Accepted : June 19, 2013

Type of Submission : Excellent Paper of Conference      Corresponding author : Kyung-Shik Shin

---

\* This work was supported by the National Research Foundation of Korea Grant funded by the Korean government (NRF-2010-330-H00003).

## 1. Introduction

Over the past few decades, changes in the business environment and advancements in computer and communication technologies have spurred the development of a knowledge-based system for business decision support. The knowledge-based system for bankruptcy prediction has also been constantly evolving according to its importance.

Bankruptcy prediction is a representative classification problem and a very important issue in the business field, in that it influences lending decisions, financing for companies, profitability, and even survival for lending institutions such as banks. Furthermore, corporate failure also affects stockholders, managers, employees, clients, affiliated companies, subcontractors, auditors, policy makers, and others. Perhaps not surprisingly, a great amount of research on issues relating to bankruptcy prediction has been conducted over the past half century (Beaver, 1966; Altman, 1968; Ohlson, 1980; Odom and Sharda, 1990; Leshno and Spector, 1996; Jo and Han, 1996; Back et al., 1996; Jo et al., 1997; Shin and Han, 1998; Atiya, 2001; Mar-Molinero and Serrano-Cinca, 2001; Hong and Shin, 2003; Neophytou and Mar-Molinero, 2004; Kim, 2004; Abdelwahed et al., 2005; Chan et al., 2006; Alfaro et al., 2008; Chen and Du, 2009; Kim and Kang, 2010; Tai and Shin, 2010).

The prediction model, which is a component of a knowledge-based system, is the main factor affecting its performance. Earlier studies on prediction modeling have focused on the building of a single best model using statistical and artificial

intelligence techniques. However, since the mid-1980s, integration of multiple techniques (hybrid techniques) and, by extension, combinations of the outputs of several models (ensemble techniques) have, according to the experimental results, generally outperformed individual models (Clemen, 1989; Hansen and Salamon, 1990; Leshno and Spector, 1996; Lin and McClean, 2001; Jo and Han, 1996; Olmeda and Fernandez, 1997; Shin and Han, 1998; Lin and McClean, 2001; West et al., 2005; Chan et al., 2006; Ravi et al., 2008; Sun and Li, 2008; Verikas et al., 2010). Integration has been an integral part of the effort to reinforce the ultimate performance of knowledge-based systems by improving the performance of the prediction model.

An ensemble is a technique that combines outputs of multiple models and produces one final prediction. The ensemble can overcome the limitations of individual models and produce better performance than the best individual model by combining outputs of multiple models.

The combination schemes to combine outputs of individual models composing the ensemble is one of the important issues affecting the prediction accuracy of an ensemble (Verikas et al., 1999; Kuncheva, 2004). A variety of combination schemes have been proposed for improvement of prediction accuracy: majority voting, weighted majority voting, the highest rank, Borda count, averaging, weighted averaging, confidence-based integration, and others (Ho et al., 1994; Cho and Kim, 1995; Hashem, 1997; Heskes, 1997; Taniguchi and Tresp, 1997; Shin and Han, 1998; Ruta and Gabrys, 2000; Lin and McClean, 2001; Hayashi and Setiono, 2002;

Kuncheva, 2004; West et al., 2005; Chan et al., 2006; Alfaro et al., 2008; Ravi et al., 2008; Verikas et al., 2010; Finlay, 2011). Those most often applied to bankruptcy prediction are majority voting, averaging, and weighted averaging (Verikas et al., 2010).

This paper aims to improve classification performance by developing an effective combination method in ensembles for bankruptcy prediction. In the present study, we examined the impact of the size of ensemble members on the proposed confidence-based selection by varying, according to the creation type of ensemble members, the number of models to combine. On the basis of the results, we herein propose our confidence-based selection approach, which can be applied to heterogeneous ensembles consisting of different kinds of models as well as to homogeneous ensembles with several artificial neural networks.

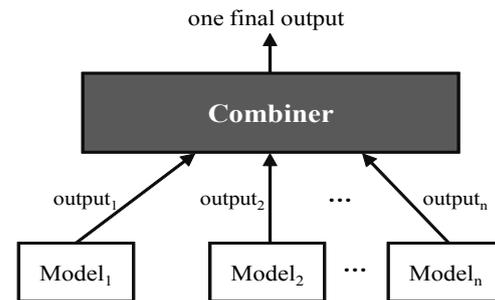
## 2. Review of Related Studies

### 2.1 Combination Schemes in Ensembles for Classification

Bankruptcy prediction that is the research field of this paper is a typical classification problem in the field of management. As mentioned earlier, studies on modeling have been progressed from the building of a single best model until ensemble techniques, which combine the outputs of several models.

In order to improve prediction performance in ensembles for classification problems, a variety of combination schemes, for instance majority

voting, weighted majority voting, the highest rank, Borda count, averaging, weighted averaging, confidence-based integration, fuzzy integral, stacked generalization, and others, have been proposed.



<Figure 1> Combination of Outputs of Multiple Models in an Ensemble

Among these, the major combination schemes are as follows. First, majority voting, which selects the class label that classifiers assign the most as a final decision for a given case, is one of the oldest and the most frequently used decision-making methods for consensus. It is also one of the most popular combination schemes, in that it is simple to use and can be applied to situations in which classifiers forming an ensemble produce outputs that cannot be interpreted in the same way (Kuncheva, 2004; Verikas et al., 2010).

In majority voting, the weights of all ensemble members are equal; weighted majority voting, contrastingly, by assigning weights for classifiers, allows the more capable among them to exercise more power in making the final decision, and selects the class label that has the largest value (Lin and McClean, 2001; West et al., 2005; Alfaro et al., 2008; Finlay, 2011).

The highest rank assigns the highest rank among the rankings given by the classifiers for each class, and then, as the final classification decision, selects the class label that has the highest (minimum) rank (Ho et al., 1994; Ruta and Gabrys, 2000).

Borda count sums the scores corresponding to the rankings given by the classifiers for each class, and selects, as the final classification decision, the class label having the largest sum total (Ho et al., 1994; Cho and Kim, 1995; Ruta and Gabrys, 2000).

Averaging calculates the average of continuous-valued outputs produced by the classifiers for each class, and selects, as the final decision for an ensemble, the class label having the largest average (Taniguchi and Tresp, 1997; Hayashi and Setiono, 2002).

Weighted averaging assigns weights for classifiers, and selects the class label that has the largest weighted average (Hashem, 1997; Heskes, 1997; Chan et al., 2006; Ravi et al., 2008).

Confidence-based integration measures confidence, which is the degree of trust in the class-label predictions of classifiers, using continuous-valued outputs of artificial neural network (ANN). It then selects, as the final decision, the class label predicted by the classifier with the highest confidence (Shin and Han, 1998; Shin and Lee, 2004).

## 2.2 Bankruptcy Prediction using Ensembles

Ensembles have been used in order to improve prediction performance in various classification problems such as bankruptcy prediction, medical diagnosis, word recognition, and face-based identity

verification.

The previous studies on ensembles relating to bankruptcy prediction are as follows. Jo and Han (1996) made five models using Case-Based Forecasting, Neural Network and Discriminant Analysis for bankruptcy prediction, and used weighted averaging as a model-integration method. The weights were obtained experimentally by trial and error. The integration approach outperformed the individual models.

Olmeda and Fernandez (1997) integrated multilayer perceptron (MLP), liner discriminant analysis (LDA), logistic regression (LR), Multivariate Adaptive Regression Splines (MARS), and the C4.5 decision tree (DT) into an ensemble for bankruptcy prediction, and used voting and weighted sum as combination schemes. They used the Evolutionary Programming (EP) algorithm to find the combination weights and showed that both methods were more accurate than any single model.

Shin and Han (1998) integrated three neural network models constructed using three feature groups for bankruptcy prediction and suggested a confidence-based integration that takes the decision of the highest-confidence neural network model as a combination scheme. The experimental results showed that significant improvement of prediction performance was achieved with this approach.

Lin and McClean (2001) combined MLP, LR, LDA, and C5.0 decision tree using two feature-selection methods, human judgment and the ANOVA statistical method, and used voting with different weights as the fusion method. The weights were

obtained from the prediction accuracy of the ensemble members in the training sample. Empirical tests showed that the combination of several models yielded better prediction accuracy than the individual models.

West et al. (2005) constructed ensembles with 100 MLPs made using bagging, cross validation, and AdaBoosting for credit scoring and bankruptcy prediction, and used majority vote and weighted vote. The weights were obtained on the basis of their accuracy. Their research results proved that the neural network ensembles were more accurate than the single best MLP model.

Chan et al. (2006) made bagged data sets performing bagging, and trained each Radial Basis Function Neural Network (RBFNN) on a separate bagged data set selecting features separately with mutual information feature grouping. The trained RBFNNs, combined using weighted sum, outperformed the others. The multiple classifier system built using the weighted sum as a fusion method performed better than any of the others, including majority voting and average.

Ravi et al. (2008) combined, by majority voting and weighted averaging, nine different models built using various neural networks and statistical techniques. Their experiment demonstrated that the ensemble generated lower Type I and Type II errors compared with its constituent models.

Sun and Li (2008) built multivariate discriminant analysis (MDA), LR, ANN, DT, support vector machine (SVM), and case-based reasoning (CBR) using stepwise discriminant analysis as a features selection method. They combined these

six models by weighted majority voting, the voting weights of the ensemble members having been determined by the prediction accuracies. An empirical experiment confirmed that the proposed approach can improve average prediction accuracy and stability.

Tsai and Wu (2008) combined multiple neural networks that were trained based on different parameters through adjustments of learning epoch and hidden nodes for bankruptcy prediction and credit scoring. The classifier ensembles performed better in only one of the three datasets.

Finlay (2011), for the purposes of credit-risk assessment, constructed three ensemble systems using LR, LDA, CART, NN and KNN models as well as sampling techniques such as bagging and boosting. The empirical results showed that some of the ensemble systems produced significantly better accuracy than the single best model.

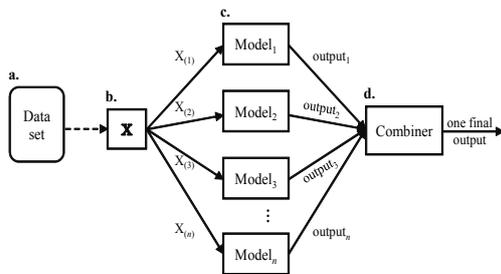
### 3. Research Methodology

#### 3.1 Introduction

An ensemble is a technique that constructs a set of multiple models, combines their outputs, and produces one final prediction. To obtain a better prediction performance than any single model, highly various studies have been conducted.

<Figure 2> illustrates approaches to building ensembles: a. Data approach: Using different training data subsets, b. Feature approach: Using different input variables, c. Model approach: Using different base models, and d. Combination approach: Using different combination schemes (Pal and Pal,

2001; Kuncheva, 2004).



<Figure 2> Approaches to Ensemble Design  
(adapted from Kuncheva, 2004)

The data, feature, and model approaches are methods for the creation of multiple models, and the combination approach is a method for the fusion of model outputs.

### 3.2 Approaches to Creation of Ensemble Members

#### 3.2.1 Data Approach

The first approach for ensemble design is the data sampling approach that trains ensemble members using different training data sets. It has been proven that a change in a training set can improve classification accuracy in an ensemble. And among the various methods using different data subsets, it is known that bagging and boosting are successful (Kim and Kang, 2010; Sewell, 2011).

Bagging, which is the abbreviation for Bootstrap AGGREGatING, generates multiple versions of a training data set using random sampling with replacement, on which bootstrap replicates respective models are trained. The outputs of the ensemble members are combined by plurality vote

for classification and by the average for prediction of a continuous-valued output (Breiman, 1996; Kuncheva, 2004; Sewell, 2011).

Boosting is a method that gradually constructs the ensemble, adding one ensemble member at a time while manipulating the training data sets. In boosting, the outcome of a classifier previously learned provides information necessary for learning of the next classifier, and learning progresses in order to compensate for the fault of the previous classifier (Freund and Schapire, 1996; Quoted in Quinlan, 1996; Alfaro et al., 2008).

#### 3.2.2 Feature Approach

The second feature approach is to apply different groups of input variables to ensemble members. Because a major cause of a given problem can be left out from a group of input variables applied to a classifier, it can be useful to vary the selection of feature subsets for ensemble members in the way that predictions from different perspectives are performed (Shin and Han, 1998).

The groups of input variables are generally extracted by statistical techniques, for example t-test, ANOVA, and stepwise selection method, artificial intelligences techniques such as genetic algorithm, and opinions from human experts.

#### 3.2.3 Model Approach

The third model approach entails the composition of base models forming the ensemble. The ensemble consists of multiple models that respectively produce prediction outputs. Therefore,

it can be diversely built, depending on how ensemble members are composed.

Ensembles are largely divided between homogeneous and heterogeneous types, according to whether the ensemble members are identical or dissimilar models. Even if the ensemble is constructed of identical base models, various ensemble members can be made using different input variables, different training data sets and different architectures (Shin and Han, 1998; West et al., 2005; Chan et al., 2006; Tsai and Wu, 2008).

### 3.3 Combination Approach to Classifier Outputs

Finally, there is the combination approach for combining the outputs of ensemble members. A variety of combination schemes have been proposed for improvement of classification accuracy in ensembles: majority voting, weighted majority voting, the highest rank, Borda count, averaging, weighted averaging, confidence-based integration, and others (Xu et al., 1992; Ho et al., 1994; Cho and Kim, 1995; Taniguchi and Tresp, 1997; Shin and Han, 1998; Ruta and Gabrys, 2000; Lin and McClean, 2001; Hayashi and Setiono, 2002; West et al., 2005; Chan et al., 2006; Alfaro et al., 2008; Ravi et al., 2008; Verikas et al., 2010).

None of the combination schemes always outperforms the others under any circumstances. Each combination scheme has advantages and limitations, and can be influenced by domain and circumstance. Accordingly, decisions on the most appropriate combination scheme in a given domain and contingency are very difficult.

## 4. Model Development

### 4.1 Introduction

The ensemble, which is one of the techniques used to improve prediction accuracy, combines the outputs of multiple models and produces one final prediction. The way in which the outputs of ensemble members are combined is one of the important issues affecting prediction accuracy (Verikas et al., 1999; Kuncheva, 2004).

A variety of combination schemes have been proposed. Among them, the most popular combination schemes in ensembles for bankruptcy prediction are majority voting, averaging, and weighted averaging. However, when majority of ensemble members predict wrongly in the ensemble that adopts these combination methods, the final prediction result is wrong or it is difficult to give the correct prediction result. In other words, opinions of a small number of models that predict correctly are ignored or can be weakly reflected in the final decision-making.

In comparison, confidence-based integration, which measures confidence using continuous-valued outputs of artificial neural networks for bankruptcy prediction and takes the decision of model having the highest confidence, can produce correct prediction result, even when majority of models predict wrongly, if one model having the highest confidence predicts correctly. However, it cannot be applied to combining continuous-valued outputs of different level produced by different kinds of models, because it was developed to combine homogeneous models of three artificial

neural networks.

Thus, this paper proposes a confidence-based selection method that can overcome the limitations of existing combination schemes in ensembles for bankruptcy prediction by measuring the confidence that can use different types of numerical outputs.

#### 4.2 Confidence-based Selection Approach

The confidence-based combination schemes that measure confidence, which signifies the degree of trust in the class-label predictions of classifiers, and take the decision of model having the highest confidence as the final classification decision on a case by case basis completely exclude the equivalence of votes in the multiple experts (models) system. And an ensemble built using this combination method can predict correctly corresponding case, even when majority of models predict wrongly, provided that only the one ensemble member having the highest confidence

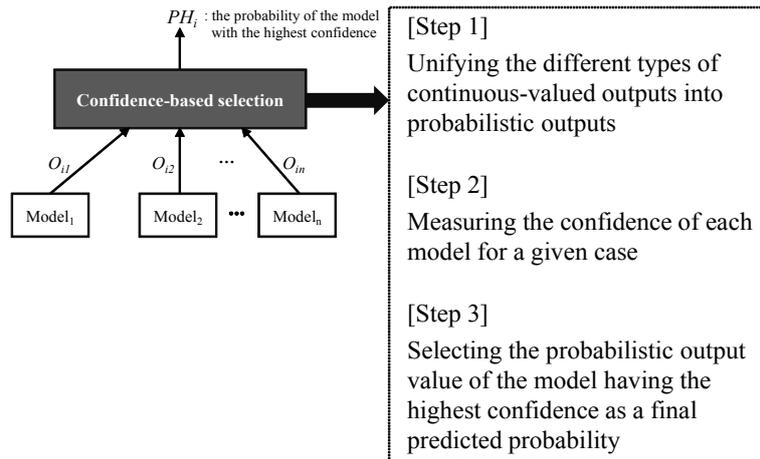
predicts correctly given case. Therefore, these characteristics can be expected to improve prediction accuracy according to domain and contingency because the greater the confidence, the larger the possibility that the prediction is right. Accordingly, this paper proposes a confidence-based selection method that can measure confidence adjusted into the same level and thus can be applied to heterogeneous ensembles as well as homogeneous ensembles.

The suggested method unifies the continuous-valued outputs of different types into probabilistic outputs using a logistic link function. For a two-class ( $Y = 0$  or  $1$ ) problem, it is as follows (Platt, 1999):

$$P(\text{class} \mid \text{input}) = P(y = 1 \mid x) \quad (1)$$

$$= p(x) = 1 / (1 + \exp^{(-\text{output})})$$

The logistic link function was proposed as one method for producing probabilistic outputs



<Figure 3> Confidence-based Selection Procedure

(Wahba, 1992; Wahba, 1999; Platt, 1999). This method is used to obtain the probability of the predicted value for classification from discriminant analysis, artificial neural network, and others in IBM SPSS Statistics, IBM SPSS Modeler, and XLMiner, which are statistics and data-mining software programs (Duda et al., 2001; Kuncheva, 2004; Shmueli et al., 2010; IBM, 2012).

In the proposed method, the confidence for the homogeneous and heterogeneous ensembles is defined as the closeness between the probability of belonging to class 1 of the classifier and the actual target probabilities. And the confidence is measured in the form of the distance between the probability of the classifier and 0.5, which is the median probability as a cutoff value.

$$\text{Confidence} = |0.5 - P_{ij}|, \quad (2)$$

where  $P_{ij}$  is the probabilistic output value for case  $i$  produced by model  $j$ , that is, the probability of belonging to class 1 for case  $i$  produced by model  $j$ .

For the two-class problem, the following expression is established:

$$\begin{aligned} &\text{The probability of belonging to class 0} \quad (3) \\ &= 1 - \text{The probability of belonging to} \\ &\quad \text{class 1} \end{aligned}$$

If the probability value of belonging to class 1 for a certain observation produced by a model is closer to 1, the probability of belonging to class 1 ( $Y = 1$ ) is higher; conversely, if it is closer to 0, the probability of belonging to class 0 ( $Y$

$= 0$ ) is higher. In other words, the larger the difference between the probability of the classifier and 0.5 (i.e. the median probability), the greater the confidence.

After measuring the confidence of each model for a given case, confidence-based selection approach compares the confidence of all ensemble members, and selects the probabilistic output value of the model having the highest confidence as the final predicted probability of the ensemble, when the ensemble members have different predicted values for classification:

$$\begin{aligned} PH_i &= P_{ik}, \text{ where } |0.5 - P_{ik}| \\ &= \arg \max |0.5 - P_{ij}|, \\ &(j = 1, \dots, m) \end{aligned} \quad (4)$$

where  $m$  is the number of ensemble members. The  $PH_i$  of the final predicted probability in the ensemble is the probabilistic output of the model showing the largest difference between the probability and 0.5 for case  $i$ .

## 5. Experiments

### 5.1 Experimental Design

#### 5.1.1 Research Data

The research data consists of 2,944 non-audited medium and small-sized manufacturing firms for which a bankruptcy occurred during the period from 2003 to 2008, along with an equal number of non-bankrupt firms, all from South Korea. The data on the bankrupt firms amounts

to between one year and two years' worth of financial statements prior to the bankruptcy occurrence. The numbers of bankrupt and non-bankrupt firms are equal in each year.

The dataset is randomly divided into a training dataset and a validation dataset (80 : 20). The training dataset is used to build models, and the validation set is used to assess performance.

Bagging, employed as a sampling technique, produces three versions of a training dataset using the original training dataset.

### 5.1.2 Features Selection

In this paper, three groups of input variables are selected through two steps for each training dataset. At the first common step, the number of financial ratios is reduced from 106 to 31 by statistical analyses such as descriptive statistics, factor analysis, univariate t-test, factor analysis, and correlation analysis, and the opinions of experts. In the second step, three groups of input variables are selected using three different methods for each training dataset. First, group 1 is selected by the stepwise selection method through Multivariate Discriminant Analysis. Group 2 consists of financial ratios selected by experts, who are credit analysts of a credit rating agency and a commercial bank in Korea. Finally, the input variables of group 3 are selected by the genetic algorithm using software package Evolver™ 5.5 of the Palisade Corporation. At this time, the hit ratio is used as the fitness function, and the crossover and mutation rates are set at 0.5 and 0.1, respectively.

### 5.1.3 Composition of base Models

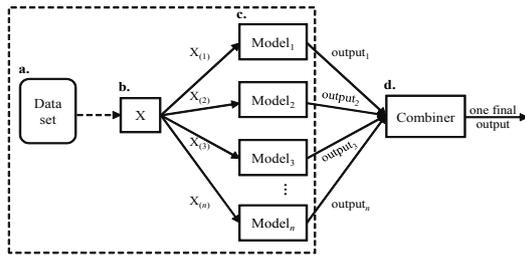
We construct homogeneous and heterogeneous ensembles. The homogeneous ensembles are constructed using artificial neural networks (ANN) with different training datasets or different feature groups. ANN has been one of the most popular methods used in bankruptcy prediction problems, and has also been utilized frequently as a base model for ensemble construction. The heterogeneous ensembles employ multivariate discriminant analysis (MDA), logistic regression (LR) and artificial neural network (ANN) as the ensemble members. MDA and LR have been widely employed in academic research and in applications for bankruptcy prediction. They have also been used often as a base model for ensemble construction.

### 5.1.4 Creation of Ensemble Members

In order to study the effectiveness of confidence-based selection in the various ensemble constructions, the creation of ensemble members was divided into seven types: Type 1 is generated using different training datasets, Type 2, using different input variables, Type 3, using different base models, Type 4, using different training datasets and input variables, Type 5, using different training datasets and base models, Type 6, using different input variables and base models, and Type 7, using different training datasets, input variables, and base models.

We vary the number of models to combine considering the total number of cases, according to the ensemble member creation type, in order

to reflect the effect of all combinations. Accordingly, Types 1, 2, and 3 each have three models forming ensembles, Types 4, 5, and 6 nine models, and Type 7, 27 models.



<Figure 4> Components of Ensemble Member Creation

### 5.2 Results and Analysis

This paper proposes a confidence-based selection approach as part of an ensemble bankruptcy-prediction scheme that can measure unified confidence, even if ensemble members produce different types of continuous-valued outputs.

To evaluate the proposed approach, four combination methods, which are the most popular combination schemes for bankruptcy-prediction ensembles, are used as benchmarks: majority voting,

weighted majority voting, averaging, and weighted averaging (the result of Borda count is equal to that of majority voting in a two-class problem).

Learning for bankruptcy prediction is performed with a training dataset, and the results are validated with the identical validation dataset (which is not used in learning) and expressed as the classification accuracy rate.

<Table 1> lists the classification accuracy rates of the seven ensemble types employing the five combination methods, and indicates the number of ensemble members for each type.

The experimental results show that when varying the number of models to combine, according to the creation type of ensemble members, the proposed combination method offers the best performance in the ensemble having the largest number of models, even when compared with the methods most often employed in bankruptcy prediction. In other words, the ensemble built using confidence-based selection in Type 7, where ensemble members were created using different training datasets, different groups of input variables, and different types of base models, has the highest

<Table 1> Classification Accuracy Rates

Type	Number of ensemble members	Majority voting	Weighted majority voting	Averaging	Weighted averaging	Confidence-based selection
Type1	3	66.02%	66.02%	65.68%	65.68%	65.59%
Type2	3	64.75%	64.75%	64.83%	64.92%	65.34%
Type3	3	67.03%	67.03%	66.86%	66.86%	66.36%
Type4	9	65.00%	65.00%	64.75%	65.08%	65.76%
Type5	9	66.36%	66.36%	65.93%	65.93%	66.61%
Type6	9	66.61%	66.61%	66.86%	66.95%	66.27%
Type7	27	65.17%	65.17%	65.85%	66.02%	67.54%

performance (67.54%). And the larger the number of models to combine, the greater the frequency of cases in which the performance of confidence-based selection is highest in the ensemble type with the same number of members.

This study utilized McNemar's test to verify the statistical significance between the classification accuracy ratio of the proposed method and that of each of the other methods used as benchmarks. <Table 2> shows that the classification accuracy ratio of the proposed method is significantly higher than those of the other benchmark methods at the 5% or 1% significance level.

<Table 2> McNemar's Test Results

	Weighted majority voting	Averaging	Weighted averaging	Confidence-based selection
Majority voting	1.000	0.200	0.110	0.004***
Weighted majority voting	-	0.200	0.110	0.004***
Averaging	-	-	0.500	0.019**
Weighted averaging	-	-	-	0.033**

\*\* Significance Level 5%, \*\*\* Significance Level 1%.

## 6. Conclusions

Prediction-modeling for bankruptcy prediction has been progressed from the building of a single best model to ensemble techniques. An ensemble is used to improve prediction accuracy for classification and estimation. It constructs a set of multiple models, combines their outputs, and produces one final prediction. How to combine the outputs of ensemble members is one of the

important issues affecting prediction accuracy.

A variety of combination schemes have been proposed in order to improve prediction performance in ensembles. Among them, the existing combination scheme using confidence can achieve better performance than the combination methods most often used in ensembles for bankruptcy prediction, depending on the situation, though it cannot be used in the case of ensembles consisting of models that produce different types of numerical outputs, because it was developed to combine homogeneous models of three artificial neural networks.

The proposed method unifies different types of continuous-valued outputs into probabilistic outputs using a logistic link function, and measures the confidence that can be applied to heterogeneous ensembles as well as homogeneous ensembles.

Thus, the present study's special contribution is the development of a combination scheme that relaxes the limitations of existing methods by measuring the confidence that can utilize different types of numerical outputs. Secondly, this study examined the impact of the size of ensemble members on the proposed method by varying the number of models to combine, according to the creation type of ensemble members.

Several research directions are suggested by the limitations of this study. First, in the experiments performed, the size of ensemble members was limited according to the creation type. In future research, changes in size must be considered at regular intervals. Second, the heterogeneous ensembles were constituted of only three kinds of

base model (ANN, MDA, and LR). Thus, there is a need to diversify the heterogeneous base model types. Third, only four combination methods, the most popular schemes for bankruptcy prediction, were used as benchmarks. New research will need to verify the proposed method by means of other combination schemes.

## References

- Abdelwahed, T. and E. M. Amir, "New Evolutionary Bankruptcy Forecasting Model based on Genetic Algorithms and Neural Networks," *Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence*, (2005).
- Alfaro, E., N. Garcia, M. Gamez, and D. Elizondo, "Bankruptcy forecasting : an empirical comparison of AdaBoost and neural networks," *Decision Support Systems*, Vol.45(2008), 110~122.
- Altman, E. I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The Journal of Finance*, Vol.23, No.4(1968), 589~609.
- Atiya, A. F., "Bankruptcy Prediction for Credit Risk Using Neural Networks : A Survey and New Results," *IEEE transactions on neural networks*, Vol.12, No.4(July, 2001), 929~935.
- Back, B., T. Laitinen, and K. Sere, "Neural networks and genetic algorithms for bankruptcy predictions," *Expert Systems with Applications*, Vol.11(1996), 407~413.
- Beaver, W., "Financial Ratios as Predictors of Failure," *Journal of Accounting Research*, Vol.4(1966), 71~111.
- Breiman, L., "Bagging predictors," *Machine Learning*, Vol.24, No.2(1996), 497~501.
- Chan, A. P. F., W. Y. Ng, D. S. Yeung, E. C. C. Tsang, and M. Firth, "Bankruptcy prediction using multiple classifier system with mutual information feature grouping," *IEEE international conference on systems, man, and cybernetics*, (2006), 845~850.
- Chen, W. S. and Y. K. Du, "Using Neural Networks and Data Mining Techniques for the Financial Distress Prediction Model," *Expert Systems with Applications*, Vol.36(2009), 4075~4086.
- Cho, S. B. and J. H. Kim, "Combining multiple neural networks by fuzzy integral for robust classification," *IEEE Transactions on System, Man, and Cybernetics*, Vol.25, No.2(1995), 380~384.
- Clemen, R. T., "Combining forecasts : A review and annotated bibliography," *International Journal of Forecasting*, Vol.5(1989), 559~583.
- Duda, R. O., P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, NY, 2001.
- Finlay, S., "Multiple classifier architectures and their application to credit risk assessment," *European Journal of Operational Research*, Vol.210(2011), 368~378.
- Freund, Y. and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Unpublished manuscript. Quoted in Quinlan, J.R., "Bagging, boosting, and C4. 5," *Proceedings of the National Conference on Artificial Intelligence*, (1996).
- Hansen, L. K. and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, No.10(1990), 993~1001.
- Hashem, S., "Optimal linear combinations of neural

- networks,” *Neural Networks*, Vol.10, No.4 (1997), 599~614.
- Hayashia, Y., and R. Setionob, “Combining neural network predictions for medical diagnosis,” *Computers in Biology and Medicine*, Vol.32 (2002), 237~246.
- Heskes, T., *Balancing between bagging and bumping*, In : Mozer, M. C., Jordan, M. I., Petsche, T. (Eds.), *Advances in Neural Information Processing Systems 9*. Cambridge, MIT Press, (1997), 466~472.
- Ho, T. K., J. J. Hull, and S. N. Srihari, “Decision combination in multiple classifier systems,” *Pattern Analysis and Machine Intelligence*, Vol.16, No.1(1994), 66~75.
- Hong, S. H. and K. S. Shin, “Using GA based Input Selection Method for Artificial Neural Network Modeling Application to Bankruptcy Prediction,” *Journal of Intelligence and Information Systems*, Vol.9, No.1(2003), 227~249.
- IBM SPSS Modeler Help, Modeling Nodes, Neural Networks, embedded in IBM SPSS Modeler, (2012).
- Jo, H. K. and I. G. Han, “Integration of Case-Based Forecasting, Neural Network, and Discriminant Analysis for Bankruptcy Prediction,” *Expert Systems With Applications*, Vol.2, No.4(1996), 415~422.
- Jo, H. K., I. G. Han, and H. Lee, “Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis,” *Expert Systems With Applications*, Vol.13(1997), 97~108.
- Kim, K. J, “Data Mining using Instance Selection in Artificial Neural Networks for Bankruptcy Prediction,” *Journal of Intelligence and Information Systems*, Vol.10, No.1(2004), 109~123.
- Kim, M. J. and D. K. Kang, “Ensemble with neural networks for bankruptcy prediction,” *Expert Systems with Applications*, Vol.37(2010), 3373~3379.
- Kuncheva, L. I., *Combining pattern classifiers : Methods and algorithms*, Hoboken, New Jersey, John Wiley & Sons, 2004.
- Leshno, M. and Y. Spector, “Neural network prediction analysis : The bankruptcy case,” *Neurocomputing*, Vol.10(1996), 125~147.
- Lin, F. Y. and S. McClean, “A data mining approach to the prediction of corporate failure,” *Knowledge-Based Systems*, Vol.14(2001), 189~195.
- Mar-Molinero, C. and C. Serrano-Cinca, “Bank failure : a multidimensional scaling approach,” *The European Journal of Finance*, Vol.7 (2001), 165~183.
- Neophytou, E. and C. Mar-Molinero, “Predicting Corporate Failure in the UK : A Multidimensional Scaling Approach,” *Journal of Business Finance and Accounting*, Vol.31 (2004), 667~710.
- Ruta, D. and B. Gabrys, “An Overview of Classifier Fusion Methods,” *Computing and Information Systems*, Vol.7(2000), 1~10.
- Odom, M. and R. Sharda, “A neural network model for bankruptcy prediction,” *Proc. Int. Joint Conf. Neural Networks*, San Diego, CA, (1990).
- Ohlson, J., “Financial ratios and the probabilistic prediction of bankruptcy,” *Journal of Accounting Research*, Vol.18, No.1(1980), 109~131.
- Olmeda, I. and E. Fernandez, “Hybrid classifiers for financial multicriteria decision making the case of bankruptcy prediction,” *Computational Economics*, Vol.10(1997), 317~335.
- Pal, S. K. and A. Pal, *Pattern Recognition : From Classical to Modern Approaches*, World Scientific, 2001.

- Platt, J. C., *Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods*, In *Advances in large margin classifiers*, Alexander J. S., P. Bartlett, B. Scholkopf, D. Schuurmans, eds., Cambridge, MIT Press, 1999.
- Ravi, V., H. Kurniawana, P. Thaia, and P. Kumarb, "Soft computing system for bank performance prediction," *Applied Soft Computing*, Vol.8 (2008), 305~315.
- Sewell, M., "Ensemble Learning," *Research Note*, RN/11/02, Dept. of Computer Science, UCL, 2011.
- Shin, K. S. and I. G. Han, "Bankruptcy Prediction Modeling Using Multiple Neural Network Models," *Proceedings of the Fall Conference on the Korean Operations Research and Management Science Society (KORMS)*, (1998), 145~148.
- Shin, K. S. and K. J. Lee, "Bankruptcy Prediction Modeling Using Multiple Neural Network Models," *Lecture Notes in Computer Science*, Vol.3214(2004), 668~674.
- Shmueli, G., N. R. Patel, and P. C. Bruce, *Data Mining for Business Intelligence : Concepts, Techniques, and Applications in Microsoft Office Excel With XMiner*, Hardcover, 2nd Edition, Wiley, 2010.
- Sun, J. and Li, h. "Listed companies' financial distress prediction based on weighted majority voting combination of multiple classifiers," *Expert Systems with Applications*, Vol.35 (2008), 818~827.
- Tai, Q. Y. and K. S. Shin, "GA-based Normalization Approach in Back-propagation Neural Network for Bankruptcy Prediction Modeling," *Journal of Intelligence and Information Systems*, Vol.16, No.3(2010), 1~14.
- Taniguchi, M. and V. Tresp, "Averaging regularized estimators," *Neural Computation*, Vol.9(1997), 1163~1178.
- Tsai, C. F. and J. W. Wu, "Using Neural Network Ensembles for Bankruptcy Prediction and Credit Scoring," *Expert Systems with Applications*, Vol.34(2008), 2639~2649.
- Verikas, A., A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers : A comparative study," *Pattern Recognition Letters*, Vol.20 (1999), 429~444.
- Verikas, A., Z. Kalsyte, M. Bacauskiene, and A. Gelzinis, "Hybrid and ensemble-based soft computing techniques in bankruptcy prediction : a survey," *Soft Computing*, Vol.14 (2010), 995~1010.
- Wahba, G., *Multivariate function and operator estimation, based on smoothing splines and reproducing kernels*, Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity, Proc. Vol.XII, Eds. M. Casdagli and S. Eubank, Addison-Wesley, 1992.
- Wahba, G., "Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV," *Technical Report*, NO.984rr, Department of Statistics, University of Wisconsin, 1998.
- West, D., S. Dellana, and J. Qian, "Neural network ensemble strategies for financial decision applications," *Computers and Operations Research*, Vol.32(2005), 2543~2559.
- Xu, L., A. Krzyzak, and C. Y. Suen, "Methods for combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.22(1992), 418~435.

Abstract

## 부도예측을 위한 확신 기반의 선택 접근법에서 앙상블 멤버 사이즈의 영향에 관한 연구

김나라\* · 신경식\*\* · 안현철\*\*\*

부도예측을 위한 지식기반시스템에서 모델은 실적에 영향을 끼치는 주요한 요인이다. 예측 모형의 개발에 있어 초기 연구들은 통계기법 및 인공지능기법들을 이용하여 최고 실적을 가지는 단일 모델을 만드는데 주력하였다. 1980년대 중반 이후에는 다수 기술의 통합(하이브리드), 더 나아가, 다수 모델의 결과의 결합(앙상블) 기법이 수많은 실험에서 개별 모델들보다 더 나은 결과를 보여왔다. 다수 모델들의 출력값들을 결합하여 한 개의 최종 예측값을 산출하는 앙상블 모델링에서 결합기법은 앙상블의 예측 정확도에 영향을 끼치는 중요한 이슈이다. 본 논문은 부도예측을 위한 앙상블 결합기법으로서 앙상블 멤버들이 다른 유형의 연속형 수치 출력값들을 산출하더라도 통일된 확신을 측정할 수 있는 확신 기반의 선택 접근법을 제안하고 이에 대한 앙상블 멤버 사이즈의 영향을 연구하였다. 실험 결과는 앙상블 멤버들의 생성 타입에 따라 결합하는 모델 개수를 변화시켰을 때 가장 많은 기본 모델들을 가지는 앙상블에서의 제안 결합기법이 부도예측에 가장 자주 사용되는 다른 방법들에 비해서도 가장 높은 실적을 가진다는 것을 보였다.

Keywords : 부도예측, 앙상블, 결합기법

\* 이화여자대학교 경영학과

\*\* 이화여자대학교 경영학과 교수, 교신저자

\*\*\* 국민대학교 경영대학 경영정부학부 교수

## 저자 소개



김나라

이화여자대학교에서 화학과 학사를 취득하고, 이화여자대학교 경영대학원에서 경영학 석사를 취득하였다. 현재 이화여자대학교 대학원 경영학과에서 경영정보시스템 전공으로 박사과정에 재학 중이다. 지식시스템 연구센터의 연구원으로 기업평가모형 개발 등의 다수 프로젝트에 참여하였으며, 주요 연구분야는 지능형 의사결정지원시스템, 인공지능과 데이터 마이닝 등이다.



신경식

이화여자대학교 경영대학 교수 겸 기획처장으로 재직 중이다. 연세대학교에서 경영학사, George Washington University에서 MBA, 한국과학기술원에서 경영정보학으로 박사학위를 취득하였다. 주요 연구분야는 빅 데이터 분석 및 활용, 지능형 의사결정지원시스템, 인공지능과 데이터 마이닝, 지식기반 시스템 등이며 이와 관련한 다수의 연구논문 및 산학연구를 수행하였다. 또한 가상화에 따른 인간, 조직 및 사회 변화연구, 사회관계망 분석 등에 관련된 연구를 수행하고 있다.



안현철

현재 국민대학교 경영대학 경영정보학부 조교수로 재직 중이다. KAIST에서 산업경영학사를 취득하고, KAIST 테크노경영대학원에서 경영정보시스템을 전공하여 공학석사와 박사를 취득하였다. 주요 관심분야는 금융, 고객관계관리 및 인터넷 보안 분야의 인공지능 응용, 정보시스템 수용과 관련한 행동 모형 등이다.