ELSEVIER

Expert Systems with Applications 31 (2006) 241–247

Expert Systems with Applications

www.elsevier.com/locate/eswa

An evolutionary approach to the combination of multiple classifiers to predict a stock price index

Myoung-Jong Kim [a], Sung-Hwan Min [b,*], Ingoo Han [c]

[a] Kookin Bank, 36-3 Yeoido-dong, Yeongdeungpo-gu, Seoul 150-886, Korea
[b] Department of Business Administration, Hallym University, 39 Hallymdaehak-gil, Chuncheon Gangwon-do, 200-702, Korea
[c] Graduate School of Management, Korea Advanced Institute of Science and Technology, 207-43 Cheongryangri-Dong, Dongdaemoon-Gu, Seoul 130-012, Seoul, South Korea

Abstract

Multiple classifier combination is a technique that combines the decisions of different classifiers. Combination can reduce the variance of estimation errors and improve the overall classification accuracy. However, direct application of combination schemes developed for pattern recognition to solving business problems has some limitations, because business problems cannot be explained completely by the results provided by machine-learning-driven classifiers alone owing to their intrinsic complexity. To solve such problems, this paper proposes an approach that is capable of incorporating the subjective problem-solving knowledge of humans into the results of quantitative models. Genetic algorithms (GAs) are used to combine classifiers stemming from machine learning, experts, and users. We use our GA-based method to predict the Korea stock price index (KOSPI).

Keywords: Genetic algorithms; Machine-learning-driven classifier; Human-driven classifier

1. Introduction

Multiple classifier combination is a technique that combines the decisions of different classifiers that are trained to solve the same problem, but make different errors. Numerous studies of combination systems have been attempted in a wide range of pattern-recognition fields, including machine printed word/-character recognition (Ho, Hull, & Srihari, 1994), handwritten character recognition (Xu, Krzyak, &, Suen, 1992; Huang & Suen, 1995), speaker recognition (Radova & Psutka, 1997), face identification (Brunelli & Falavigna, 1995), text-to-phoneme translation (Wolpert, 1992), remote sensing (Benediktsson, Sveinsson, Ersoy, & Swain, 1997), military target recognition (Dasarathy, 1994), and biomedical signal processing (Hu, Parlreddy, & Tompkins, 1977). The findings show that combination can reduce the variance of estimation errors, thereby improving the overall classification accuracy owing to the reduced variance in the estimated decision boundary (Tumer & Ghosh, 1996).

Recently, combining classifiers to achieve greater accuracy has become an increasingly important research topic in business. Business problems are generally complicated and lack complete quantitative models. There have been significant efforts to combine the classification results provided by multiple classifiers in order to solve business classification problems (Blattberg & Hoch, 1990; Kim, Kim, & Lee, 2000). The appropriate combination of multiple classifiers can perform better than a single classifier in solving such problems.

However, direct application of the combination schemes developed for pattern recognition to solving business problems has limitations. Most of the combination studies examining pattern recognition have concentrated on combining the outputs of individual classifiers driven by various machine-learning techniques and the success of the combination depends largely on the individual performance of these machine-learning-driven classifiers. Business problems cannot be explained completely using the results provided by machine-learning-driven classifiers alone, owing to their intrinsic complexity. In fact, expert decision-makers do not depend on quantitative models entirely; they appear to work with their subjective knowledge framework to solve these problems, in addition to using quantitative models. The multiple classifiers used to solve such problems can be derived from machine learning, experts, and users. The classifiers obtained from these sources have different, but

* Corresponding author. Tel.: +82 33 248 1841; fax: +82 33 256 3424.
E-mail address: shmin@hallym.ac.kr (S.-H. Min).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.09.020

complementary, features. Machine-learning-driven classifiers are consistent because they are derived automatically from historical instances that possess regularities that are useful for interpreting some parts of phenomena. Conversely, human-driven classifiers obtained from experts and users are inconsistent because experts and users have different ways of thinking, although they have the advantage of being adaptable to the changing environment.

This paper proposes a genetic algorithm (GA)-based method for combining measurement-level classifiers derived from machine learning, experts, and users. The proposed method was developed in order to solve unstructured business problems. We apply our GA-based method to predict the Korea stock price index (KOSPI), which is categorized into four groups {Bear, Edged-Down, Edged-Up, Bull} according to the weekly return. The machine-learning-driven classifier is derived automatically by applying neural networks to technical indicators of the stock market, while the two human-driven classifiers are derived from the evaluations of an expert and a random user using qualitative information associated with the stock market. This study uses GAs to evolve a set of weights indicating the degree of importance of each classifier for the four stock market levels.

The performance of seven combination schemes was compared with the GA method, including the majority vote, weighted majority vote, Borda count, weighted Borda count, Bayesian formalism, BKS, and Dempster–Shafer theories of evidence. The experimental results show that the importance of each classifier differs for the four stock market classifications. Both human-driven classifiers can provide more accurate classifications than the machine-learning-driven classifier in the turbulent decision-making situations Bear and Bull, while the machine-learning-driven classifier performs better than both human-driven classifiers in typical situations: Edged-Down and Edged-Up. Moreover, the predictive accuracy of the combined results of the GA-based method is significantly better than that of individual classifiers, and the GA-based method has better predictive accuracy than other combination methods.

The remainder of this paper is organized as follows. Section 2 reviews the combination algorithms used in this paper. Section 3 explains the GA-based combination method. In Section 4, we discuss the experimental design. Section 5 describes the results of the experiment. Concluding remarks and further research issues are described in Section 6.

2. The combination methods

This section provides a review of the combination algorithms. The relatively well-known combining methods include the majority vote, Borda count, Bayesian method, Behavior–knowledge space (BKS), and Dempster–Shafer theories of evidence. Recently, there have been some works using neural networks and fuzzy algorithms (Cho & Kim, 1995; Huang, Liu, & Suen, 1994; Yamaoka, Lu, Shaout, & Shridhar, 1994).

2.1. The majority vote

The majority vote method of combining classifier decisions selects the class that is supported by the majority of the classifiers. Each classifier has one vote that can be cast for any one class in the majority voting. The majority voting method selects the class, which gets votes more than half of the classifiers. When there is no agreement among more than half of the classifier, the input is rejected (Xu et al., 1992). This method is very simple and needs no extra memory, but it treats classifiers equally without considering their differences in performance. The weighted majority voting and Borda count can be seen as extension of the majority voting method. The weighted majority voting method assigns each classifier different weight according to its performance (Suen, Nadal, Mai, Legault, & Lam, 1992).

2.2. Borda count

The Borda count is a simple yet effective method of combining rankings. The Borda count method for ranked lists combination can be considered as a generalized form of the majority voting method (Ho et al., 1994). Let B_k^i be the ranking of the ith class provided by the kth classifier. The Borda count for class i is calculated as:

$$B(i) = \sum_{k=1}^{K} B_k^i \tag{1}$$

where K is the number of classifiers used. The final decision is given by selecting the class yielding the largest Borda count.

2.3. Bayesian method

Let S be a pattern space which consists of M mutually exclusive sets $S = C_1 \cup \cdots \cup C_M$ be a set of class labels, each of C_i, $i \in \Lambda = \{1, \ldots, M\}$ representing a set of specified patterns called a class. e_k means classifier where $k = \{1,\ldots,K\}$ and x denotes an input pattern. For a given input pattern x, $e_k(x) = \{m_k^i(x) | \forall\, i(1 \le i \le M), \forall k(1 \le k \le K)\}$ means that classifier k assigns the input pattern x to each class i with a measurement value $m_k^i(x)$ and the class label of $e_k(x)$ is determined by assigning class label with the largest measurement value. Bayesian method assumes that the classifiers are mutually independent (Franke & Mandler, 1992). Let $e_1(x),\ldots,e_k(x)$ be the class labels assigned to x by classifier. The independence assumption leads to the belief function for class i represented in the sum of conditional probabilities as follows:

$$\mathrm{BEL}(i) = \eta \prod_{k=1}^{K} P(x \in C_i | e_k(x) = j), \quad \text{for } i = 1, \ldots, M \tag{2}$$

where η is a normalization coefficient that satisfies $\sum_{i=1}^{M} \mathrm{BEL}(i) = 1$. The class with highest combined belief measure is chosen as a final classification decision. Alternatively selection of any class may be rejected if the combined

belief is smaller than a specified threshold value (Xu et al., 1992).

2.4. The behavior–knowledge space

Most combination methods assume independence of the decision made by individual classifiers. This is in fact not necessarily true and behavior–knowledge space (BKS) does not require this condition. It derives its weights from the knowledge space, which concurrently records the decision of all the classifiers on each learned sample during the training process. The BKS method embodies multidimensional space where each dimension corresponds to the decisions of one classifier. The intersection of the decisions of individual classifiers called focal unit constitutes one of the BKS space. In general, each unit in K-dimensional space can be denoted as BKS($e_1,\ldots,e_k$) where classifier 1 gives its decision as e_1 and classifier K gives its decision e_k. Each BKS unit contains three types of data: the total number of incoming samples: $T_{e_1,\ldots,e_k}$ the best representative class: $R_{e_1,\ldots,e_k}$, and the total number of incoming samples for each class: $n_{e_1,\ldots,e_k}(m)$.

To combine the decisions by each classifier, BKS method is implemented in two stages: knowledge modeling and decision-making. During the knowledge modeling stage, it makes K-dimensional BKS for K classifiers. In decision-making stage, it decides the final result using the following rules:

$$F(x) = \begin{cases} R_{e_1,\ldots,e_k}, & \text{if } T_{e_1,\ldots,e_k} > 0 \text{ and } \dfrac{n_{e_1,\ldots,e_k}(R_{e_1,\ldots,e_k})}{T_{e_1,\ldots,e_k}} \geq \lambda \\ \text{reject}, & \text{Otherwise} \end{cases} \tag{3}$$

where λ is a threshold controlling the reliability of the final decisions (Huang & Suen, 1995).

2.5. Dempster–Shafer theories of evidence

The Dempster–Shafer theory, also known as the theory of belief functions, is a generalization of the Bayesian theory of subjective probability. Whereas the Bayesian theory requires probabilities for each question of interest, belief functions allow us to base degrees of belief for one question on probabilities for a related question. These degrees of belief may or may not have the mathematical properties of probabilities; how much they differ from probabilities will depend on how closely the two questions are related. For details on using Dempster–Shafer theory, readers may refer to (Le Hegarat-Mascle, Bloch, & Vidal-Madjar, 1997; Klein, 1999).

3. The GA-based combination method

Genetic algorithms (GAs) are heuristic search techniques that are based on the theory of natural selection and evolution (Holland, 1975). They are effective and robust in searching very large spaces in a wide range of applications. GAs are particularly suitable for multi-parameter optimization problems in which an object function is subject to numerous hard and soft constraints. This paper proposes a GA-based method to combine measurement-level classifiers stemming from machine learning, experts, and users. The proposed method was developed in order to solve unstructured business problems. The GA-based method for combining multiple classifiers including machine learning-, expert-, and user-driven classifiers is as follows.

Let S be a pattern space that consists of M mutually exclusive sets $S = C_1 \cup \cdots \cup C_M$, each of C_i, $i \in \Lambda\{1,\ldots,M\}$ representing a set of specified patterns called a class. $e_k(x)$ means a classifier where $k = \{1,\ldots,K\}$ and x denotes an input pattern. For a given input pattern x, $e_k(x) = \{m_k^i(x) | \forall\, i(1 \leq i \leq M), \forall k(1 \leq k \leq K)\}$ means that classifier k assigns the input pattern x to each class i with a measurement value $m_k^i(x)$. Let $W = \{w_k^i | \forall\, i(1 \leq i \leq M), \forall k(1 \leq k \leq K)\}$ be the set of weights, where w_k^i means the degree of importance of the kth classifier for class i and has a positive value between 0 and 1.

For an input pattern x, the final output $E_i(x)$ for class i is calculated as the weighted sum of measured values $m_k^i(x)$ and the corresponding weight values w_k^i and is expressed as:

$$E_i(x) = \sum_{k=1}^{K} w_k^i m_k^i(x) \tag{4}$$

The final decision is given by selecting the class label with the highest output value $E_i(x)$.

For a GA, each weight vector should be encoded into a string called a chromosome. We use a real value string to represent a chromosome. The initial population consists of a set of weights distributed randomly from the systems. Once the initial population is generated, the GA evaluates each individual according to the fitness function. The role of the fitness function is to encode the performance of each individual numerically. In this study, the objective of the GA method is to find the set of weights capable of generating the optimized combination results. The fitness function for the lth weight set is defined as follows:

$$Z' \text{ Fitness}(W_l) = \left[\frac{\sum_{i=1}^{N} \text{HR}(W_l)}{N} \right] \tag{5}$$

where N is the total amount of training data.

As shown in Eq. (5), the fitness function is measured using the sum of HR(W_l) for all input training data. For an input pattern x, HR(W_l) can be calculated as:

$$\text{HR}(W_l) = \begin{cases} 1 & \text{if correctly matched} \\ \left[\dfrac{E_j^{W_1}(x)}{\sum_{i=1}^{M} E_i^{W_1}(x)} \right] & \text{otherwise} \end{cases} \tag{6}$$

where j is the actual class for input x, W_l is the lth candidate

Table 1
The economic meaning of and formulas for the technical indicators

Technical indicator	Economical meaning and formula
MA	MA is a popular way of defining where the price trend of recent days lies $\left[\frac{\text{the sum of the closing price over 6 days}}{6}\right]$
RSI	RSI, which moves on a scale from 0 to 100, highlights overbought (70 and above) or oversold (30 or below) conditions $\left[\frac{\text{the sum of the positive closing price over 25 days}}{\text{the sum of the closing price over 25 days}}100\right]$
PSY	PSY measures the psychological stability of investors $\left[\frac{\text{the days when the price has close up over 12 days}}{12}100\right]$
MOM	MOM measures the speed of price change. It highlights periods when the market is overbought or oversold and consequently prone to reversal [the most recent closing price − the closing price 6 days ago]
STOD	STOD shows buy (30 and below) or sell (70 or above) signals $\left[\frac{\text{the most recent closing price–the low price over 6 days}}{\text{the high price over 6 days–the low price over price over 6 days}}\right]$
VR	VR measures the stability of the trend $\left[\frac{\text{the sum of the volume whenthe price has close up over 6 days–the sum of the volume when the price has close down over 6 days}}{\text{the sum of volume over 6 days}}100\right]$
OBV	OBV is a running cumulative total and it should confirm the price trend [the sum of volume when the price has close up over 6 days − the sum of volume when the price has close down over 6 days]
DIS	DIS shows the stability of the most recent closing price $\left[\frac{\text{the most recent closing price}}{\text{6 day MA}}100\right]$
ROC	ROC shows buy (130 and above) and sell (70 and below) signals $\left[\frac{\text{the most recent closing price}}{\text{the price of 6 days ago}}\right]$

weight vector, and $E_i^{W_1}(x) = \sum_{k=1}^{K} w_k^i m_k^i(x)$ is the final output for class *i*. If an individual classifies the input correctly, the score for the classifier is increased by one. Otherwise, it is increased by the ratio of the measurement for the actual class to the sum of all measurements to consider the potential hit of the individual.

Once each weight set is evaluated using the fitness function, GAs select the best solution to reproduce new offspring using genetic operators, such as crossover and mutation. The good offspring replace old candidates with a low fitness value. After replacement, the new population is evaluated using the fitness function. This process continues iteratively until a predefined stopping condition is reached.

4. Experiments

4.1. The experiment design of a machine-learning-driven classifier

The input variables for a machine-learning-driven classifier are technical indicators. We used the following technical indicators: the moving average (MA), relative strength index (RSI), psychology (PSY), momentum (MOM), stochastic %D (STOD), volume ratio (VR), On balance volume (OBV), disparity (DIS), and rate of change (ROC). The meanings and formulas for the technical indicators are shown in Table 1.

The output variables are four stock market levels: Bear, Edged-Down, Edged-Up, and Bull. Criteria used by stock market analysts (−3%, 0%, +3%) were used to determine the appropriate classification. If the return for the next week is greater than 3%, the corresponding stock market level is regarded as Bull. Similarly, Edged-Up, Edged-Down, and Bear are used when the return for the next week is between 3% and 0, between 0 and −3%, and less than −3%, respectively.

We applied a two-stage input selection process. In the first stage, one-way ANOVA between technical indicators and the stock market level was used to select the appropriate technical indicators. We selected seven that were significant at the 5% level: MA, STOD, RSI, ROC, DIS, MOM, and VR. In the second stage, we selected five technical indicators using a stepwise method: MA, STOD, RSI, ROC, and DIS. Table 2 shows the F-ratio and level of significance obtained from the one-way ANOVA and the coefficients with the significance level obtained from the stepwise MDA.

We use a three-layered backpropagation neural network as the machine-driven classifier. The total number of samples available consisted of data for the 572 weeks from January 1990 to December 2001. The dataset was split into three subsets according to the period used: neural net training (312 weeks from January 1990 to December 1995), test (156 weeks from January 1996 to December 1998), and validation (156 weeks from January 1999 to December 2001).

4.2. The experimental design of the human-driven classifier

The user was a stock market analyst with less than one year of experience, while the expert was a stock market analyst with more than 7 years experience. The expert- and user-driven

Table 2
The statistics of the one-way ANOVA and stepwise MDA

Variables	F-ratio (ANOVA)	Coefficient (stepwise MDA)
Moving average (MA)	5.777*	.979**
Stochastic %D (STOD)	9.600**	.971**
Relative strength index (RSI)	18.121**	.964**
Rate of change (ROC)	13.702**	.953**
Disparity (DIS)	22.978**	.881**
Psychology (PSY)	2.254	–
Momentum (MOM)	19.924**	–
Volume ratio (VR)	19.351**	–
On balance volume (OBV)	2.399	–

*Significant at the 5% level; **significant at the 1% level.

Table 3
The optimal set of weights obtained from the GA-based method

		BR	ED	EU	BL
W_K	MC	0.084	0.752	0.948	0.121
	EC	0.688	0.178	0.151	0.746
	UC	0.101	0.112	0.087	0.131

The rows represent the four levels of the stock market {Bear (BR), Edged-Down (ED), Edged-Up (EU), Bull (BL)} and the columns represent the classifiers {Machine learning (MC), Expert (EC), User (UC)}.

classifiers were their evaluations of qualitative information on the stock market, including economic prospects, the level of stock supply and demand, the amount of currency ready for buying stocks, and conditions favorable or unfavorable for the stock market trend. The evaluations of qualitative information by the expert and user are represented as possibility measures for the four levels of the stock market because it is difficult to represent their evaluations as probability measures. Therefore, these measures need to be transformed into possibility measures. We introduced the concept of Dubois and Prade's (Dubois & Prade, 1983) solution to convert possibility distributions into probability distributions, while maintaining consistency. Given a possibility measure, the possibility of classifying an object into class C_i can be defined as:

$$\pi(C_i) = \mu(C_i)/\max \mu(C_i) \quad \text{for } i = 1, \ldots, M \tag{7}$$

where $\mu(C_i)$ represents the possibility measure and $\pi(C_i)$ is a normalized possibility distribution for the class labels i. Let $\pi_1,\ldots,\pi_m$ be the possibility values in descending order, so that π_1 represents the highest value 1 and is π_m represents the lowest value. For any possibility value π_i the corresponding probability value p_i is given by

$$P(C_j) = \sum_{j=1}^{m} \frac{\pi_j - \pi_{j+1}}{j} \quad \text{with } \pi_{m+1} = 0 \tag{8}$$

Suppose that we have the following possibility measures for the four levels of the stock market: 0.8 for Bear, 0.6 for Edged-Down, 0.5 for Edged-Up, and 0.2 for Bull. Then, we have:

$$\pi(\text{Bear}) = 0.8/0.8 = 1$$

$$\pi(\text{Edged} - \text{Down}) = 0.6/0.8 = 0.75$$

$$\pi(\text{Edged} - \text{Up}) = 0.5/0.8 = 0.625$$

$$\pi(\text{Bull}) = 0.2/0.8 = 0.25$$

$$p(\text{Bear}) = (1-0.75) + (0.75-0.625)/2 + (0.625-0.25)/3 + (0.25-0.0)/4 = 0.5$$

$$p(\text{Edged} - \text{Down}) = (0.75-0.625)/2 + (0.625-0.25)/3 + (0.25-0.0)/4 = 0.25$$

$$p(\text{Edged} - \text{Up}) = (0.625-0.25)/3 + (0.25-0.0)/4 = 0.1875$$

$$p(\text{Bull}) = (0.25-0.0)/4 = 0.0625$$

If $\pi_1,\ldots,\pi_m$ are in descending order, then the probability values $p_1,\ldots,p_m$ will necessarily sum to one. We can now replace the original possibility distribution with a probability distribution that can be used in the combination.

Qualitative information was collected from weekly KOSPI reports for the period from January 1996 to December 2001. The data were classified into two datasets used for learning (156 weeks from January 1996 to December 1998) and validation (156 weeks from January 1999 to December 2001).

4.3. The experiment design for combining multiple classifiers

Generally, the population size is based on the size of the problem. We used 100 strings in the population. Several parameters had to be defined for genetic operators because the values of these parameters can have a great influence on the algorithm. The crossover rate ranges from 0.5 to 0.7 and the mutation rate ranges from 0.06 to 0.12. The set of individuals was evolved for 3000 generations. In addition, the performance of the GA method was compared with seven combination schemes: the majority vote, weighted majority vote, Borda count, weighted Borda count, Bayesian formalism, BKS, and Dempster–Shafer theories of evidence.

The data for the combined methods consisted of 156 weeks from January 1996 to December 1998, while the data for validation consisted of 156 weeks from January 1999 to December 2001.

5. Discussion and results

The GA-based method generated the weight vector shown in Table 3.

Let $\text{IMP}_k(i)$ be the relative importance of the values of the kth classifier to the machine-learning-driven classifier for class

Table 4
The calculated IMPs

		BR	ED	EU	BL
IMP_K	MC	1.000	1.000	1.000	1.000
	EC	8.850	0.256	0.172	6.662
	UC	5.314	0.658	0.406	4.785

Table 5
The performance of the classifiers

Classifiers	BR (%)	ED (%)	EU (%)	BL (%)	Overall (%)
UC	56.0	48.2	49.0	62.5	51.9
EC	64.0	53.6	56.9	70.8	59.0
MC	48.0	71.4	76.5	50.0	66.0
Majority vote	60.0	75.0	78.4	66.7	72.4
Weighted majority vote	64.0	75.0	78.4	70.8	73.7
Bayesian	68.0	75.0	80.4	70.8	75.0
Borda count	68.0	75.0	78.4	70.8	74.4
Weighted Borda count	68.0	75.0	80.4	70.8	75.0
BKS	72.0	76.8	80.4	70.8	75.8
Dempster–Shafer	68.0	78.6	80.4	70.8	76.3
GA	76.0	76.8	86.3	75.0	79.5

i, which is defined as follows:

$$\mathrm{IMP}_k(i) = \frac{w_k^i / \sum_{i=1}^{M} w_k^i}{w_{\mathrm{MC}}^i / \sum_{i=1}^{M} w_{\mathrm{MC}}^i} \tag{9}$$

where MC is the machine-learning-driven classifier. Table 4 shows the results of the calculation of IMP.

The IMPs of the expert- and user-driven classifiers for the Bear and Bull levels were greater than one. In Bear and Bull, stock prices are changing rapidly and regularities are rarely found. In such situations, the subjective problem-solving knowledge of humans is useful in predicting stock market levels because human knowledge is more adaptable to the changing environment than machine-learning-driven classifiers. By contrast, the IMPs of the expert- and user-driven classifiers in Edged-Down and Edged-Up were less than one, which means that the machine-learning-driven classifier can better predict the regular situations Edged-Down and Edged-Up, because it can deal with regularities more effectively than human-driven classifiers.

Table 5 shows the performance of the different classifiers. Overall, the machine-learning-driven classifier outperformed the other individual classifiers, but both human-driven classifiers performed better than the machine-learning-driven classifier in Bear and Bull situations. Consequently, the machine-learning-driven classifier is useful in typical situations, while human-driven classifiers are suitable for predicting changing situations. The performance of the GA-based method was significantly higher than those of the individual classifiers and was also better than other combination algorithms.

6. Concluding remarks

This paper proposes a GA-based method for combining measurement-level classifiers stemming from machine learning, experts, and users. The GA-based combination scheme aims to solve unstructured business classification problems, such predicting the stock market. It introduces multiple classifiers stemming from machine learning-, expert-, and user-driven classifier to deal with business problems effectively. These classifiers have different, but complementary, characteristics. The results show that the machine-learning-driven classifier is suitable for explaining and predicting regular environments, while both human-driven classifiers are appropriate for predicting dynamically changing environments. This means that multiple classifier systems should be chosen carefully according to the decision-making environment. The GA-based method performs better than individual classifiers or any other combination scheme.

This study has useful implications for the combination of multiple classifiers. First, the structure of multiple classifier systems should be designed carefully according to the complexity of the problem to be solved. Multiple classifier systems for solving well-structured problems are less subjective and can be solved using a combination of quantitative methods, while ill-structured or semi-structured problems require the participation of classifiers obtained from subjective human knowledge. Second, the proposed GA-based combining method successfully predicted KOSPI more accurately than any other combination scheme. Therefore, the GA-based method can be an effective decision tool for unstructured business problems.

However, several limitations and issues for further research remain. First, this study used a small dataset due to the difficulty in collecting data associated with human-driven classifiers. Therefore, further work should examine a large, more general dataset to confirm the efficacy of the proposed GA-based combination method. In the future, we will apply our proposed method to additional classification problem domains and we will also consider further extension of our approach to integrate different level classifiers, such as rank level or abstract level.

Acknowledgements

This work was supported by the Research Grant from Hallym University, Korea.

References

Benediktsson, J. A., Sveinsson, J. R., Ersoy, O. K., & Swain, P. H. (1997). Parallel consensual neural networks. *IEEE Transactions on Neural Networks*, *8*(1), 540–564.

Blattberg, R. C., & Hoch, S. J. (1990). Database models and managerial intuition: 50% model +50% manager. *Management Science*, *36*(8), 887–899.

Brunelli, R., & Falavigna, D. (1995). Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(10), 955–966.

Cho, S. B., & Kim, J. H. (1995). Multiple network fusion using fuzzy logic. *IEEE Transactions on Neural Networks*, *6*(2), 497–501.

Dasarathy, B. V. (1994). *Decision fusion*. Los Alamitos, CA: IEEE Computer Society Press.

Dubois, D., & Prade, H. (1983). Unfair coins and necessity measures: Towards a possibilistic interpretation of histograms. *Fuzzy Sets and Systems*, *10*, 15–20.

Franke, J., & Mandler, E. (1992). A comparison of two approaches for combining the votes of cooperating classifiers. *Proceeding of the 11th international conference on pattern recognition* pp. 611–614.

Ho, T. K., Hull, J. J., & Srihari, S. N. (1994). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis, Machine Intelligence*, *16*(1), 66–75.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Harbor, MI: MIT Press.

Hu, Y. H., Parlreddy, S., & Tompkins, W. J. (1977). A patient-adaptable EDG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, *44*(9), 891–900.

Huang, Y. S., & Suen, C. Y. (1995). A method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *17*(1), 90–94.

Huang, Y. S., Liu, K., & Suen, C. Y. (1994). A neural network approach for multi-classifier recognition systems. *Proceeding of the fourth IWFHR* pp. 235–244.

Kim, E., Kim, W., & Lee, Y. (2000). Combination of multiple classifiers for the customer's purchase behavior prediction. *Decision Support Systems*, *34*, 167–175.

Klein, L. (1999). *Sensor and data fusion, concepts and applications*. Bellingham, WA: SPIE Optical Engineering Press.

Le Hegarat-Mascle, S., Bloch, I., & Vidal-Madjar, D. (1997). Application of Dempster-Shafer evidence theory to unsupervised classification in multi-source remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(4), 1018–1031.

Radova, V., & Psutka, J. (1997). An approach to speaker identification using multiple classifiers. *Proceedings of the IEEE international conference on acoustics, speech and signal processing, Munich, Germany, April 1997* pp. 1135–1138.

Suen, C. Y., Nadal, C., Mai, T. A., Legault, R., & Lam, L. (1992). Computer recognition of unconstrained handwritten numerals. *Proceedings of the IEEE*, *80*(7), 1162–1180.

Tumer, K., & Ghosh, J. (1996). Analysis of decision boundaries in linearly combined neural classifiers. *Pattern Recognition*, *29*(2), 341–348.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*, 241–259.

Xu, L., Krzyak, A., & Suen, C. Y. (1992). Method of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, *22*(3), 418–435.

Yamaoka, F., Lu, Y., Shaout, A., & Shridhar, M. (1994). Fuzzy integration of classification results in handwritten digit recognition system. *Proceeding of the fourth IWFHR* pp. 255–264.