Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 33–40
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Experimental study for the comparison of classifier combination methods

S.Y. Sohn[a,*], H.W. Shin[b]

[a]Department of Information and Industrial Engineering, Yonsei University, Seoul, Republic of Korea
[b]Samsung Economic Research Institute, Seoul, Republic of Korea

Received 7 June 2005; received in revised form 8 May 2006; accepted 29 June 2006

Abstract

In this paper, we compare the performances of classifier combination methods (bagging, modified random subspace method, classifier selection, parametric fusion) to logistic regression in consideration of various characteristics of input data. Four factors used to simulate the logistic model are: (a) combination function among input variables, (b) correlation between input variables, (c) variance of observation, and (d) training data set size. In view of typically unknown combination function among input variables, we use a Taguchi design to improve the practicality of our study results by letting it as an uncontrollable factor. Our experimental study results indicate the following: when training set size is large, performances of logistic regression and bagging are not significantly different. However, when training set size is small, the performance of logistic regression is worse than bagging. When training data set size is small and correlation is strong, both modified random subspace method and bagging perform better than the other three methods. When correlation is weak and variance is small, both parametric fusion and classifier selection algorithm appear to be the worst at our disappointment.

Keywords: Bagging; Random subspace method; Classifier selection; Parametric fusion

1. Introduction

Classification is an important problem in data mining. It has been studied extensively by the statistics and machine learning communities as a possible solution to the knowledge acquisition or knowledge extraction. One of the main issues of classification modeling is the improvement of classification accuracy. For that purpose, many researchers have recently placed considerable attention on the task of classifier combination methods. Among them three basic approaches are classifier ensemble, classifier selection, and parametric fusion.

Classifier ensemble algorithms combine the results of several individual classifiers [1–8]. However, most of ensemble algorithms do not take into account the local expertise of each classifier. This can mislead the consensus of multiple classifiers by overlooking the opinion of some better skilled classifiers in a specific region to which the given input belongs.

Sometimes, it is useful to decompose a complex problem into simpler subproblems and solve each subproblem one by one, instead of learning the global relation between input variables and target variable. Numerous approaches concerned with classifier selection by local region have been developed [9–13].

On the other hand, in an effort to find the best classifier, Shannon and Banks [14] suggested a parametric approach to combine a set of classification trees into a single final tree. They used the maximum likelihood estimate of the central tree which minimizes the distance from individual trees obtained based on bootstrap samples. We call this a parametric fusion.

Many empirical studies have been performed to compare various classifier combination methods [7,8,15,16]. Most of the comparison studies, however, are based on the real data sets instead of carefully designed experimental setting. Generalization of such results could be risky. Therefore, it is necessary to derive some rules by which one can choose a

* Corresponding author. Tel.: +82 2 361 4014; fax: +82 2 364 7807.
E-mail addresses: sohns@yonsei.ac.kr (S.Y. Sohn), won3@yonsei.ac.kr (H.W. Shin).

0031-3203/$30.00
doi:10.1016/j.patcog.2006.06.027

proper classifier combination methods based on the nature of data sets [17]. This kind of research is scarcely done regardless of its importance.

In this paper, we use Monte Carlo simulation in order to help a selection procedure for classifier combination methods based on various data characteristics.

The organization of this paper is as follows. In Section 2, we review the literature on the classifier combination methods in order to briefly explain the approaches to be compared in our study. In Section 3, we introduce the Taguchi experimental design for Monte Carlo simulation for comparison of classifier combination methods. In Section 4, we summarize the simulation results. Finally in Section 5, we discuss the implication of our results and suggest further study areas.

2. Classifier combination methods

Frequently used data classification approaches include neural networks, decision trees and logistic regression. Neural networks in general are known to provide relatively high classification accuracy for a non-linear model. Decision trees are frequently applied to the cases with categorical input variables and can be easily interpreted. Apparently, decision tree is the most popular base classifier for ensemble. However, ensemble algorithms based on the tree have been already actively investigated since 1994 using simulation and empirical studies [1]. Logistic regression is one of the classical parametric approaches used for classification.

When there are two possible classes $\Psi = \{w_0, w_1\}$, the posterior probability $P(w_1|x)$ for inputs x based on a logistic regression can be modeled as follows:

$$P(w_1|x_1, \ldots, x_K) = \frac{\exp(\hat{\beta}_0 + \sum_{k=1}^{K} x_k \hat{\beta}_k)}{1 + \exp(\hat{\beta}_0 + \sum_{k=1}^{K} x_k \hat{\beta}_k)}, \quad (1)$$

where $\hat{\beta}_k$ is the estimated logistic regression coefficient for input variable x_k [18].

In this research, we use such a logistic regression model as a base classifier and attempt to compare the performance of combination methods such as classifier ensembles (bagging, random subspace method), classifier selection, and parametric fusion.

2.1. Bagging

Bagging algorithm introduced by Breiman [1] votes classifiers generated by bootstrap samples. A bootstrap sample is generated by randomly sampling instances from the training set with replacement from training set T, and construct the classifier using each bootstrap sample. Then combine classifiers by simple majority vote. That is assign the most frequently predicted outcome as the final classification. The basic idea of bagging is to reduce the deviation of several classifiers by voting the classified results due to bootstrap resamples. In this paper, we choose bagging as one of ensemble methods and compare its performance to the others.

2.2. Modified random subspace method

Ho [5] proposed the random subspace method (RSM) in which input variables are randomly selected for training individual classifiers and the results of the classifiers are combined by the majority voting. We use a revised version of Ho's RSM where the training sample is replaced with bootstrap resample. It is called a modified RSM. This modified version has an advantage over the original approach in a way that the resulting classification rule can be more robust than that trained based on the same sample. We randomly select half of input variables for a logistic model training based on each bootstrap resample.

2.3. Classifier selection

The classifier selection approach has two possible training strategies. The first approach specifies the region first and then builds a responsible classifier for each region [9,16]. Each classifier is trained for the region for which it is responsible. In the second approach, all classifiers are trained based on whole training data set and the best classifier is found for each pre-specified region [10,13].

In this paper, we use the first approach due to the fact that it can be more easily applied to practitioners. More detailed description for the approach used in our analysis is as follows:

In order to find the expert classifier in a local region, divide a training data set into L clusters, $C_1, C_2, \ldots, C_L$, based on data characteristics, and find the cluster centroid, $v_1, v_2, \ldots, v_L$, as the arithmetic means of the observations in the respective clusters. Next, train individual classifiers $D_1, D_2, \ldots, D_L$ for corresponding clusters $C_1, C_2, \ldots, C_L$, respectively. Given unknown test observation i, find the nearest cluster center among $v_1, v_2, \ldots, v_L$, and then observation i is classified using the corresponding classifier which is responsible for that cluster.

2.4. Parametric fusion

Shannon and Banks [14] suggested a decision tree which is located in a minimum distance from all the individually fitted decision trees based on bootstrap resamples.

Since we use a logistic regression model as a base classifier, the following mathematical programming model is used to find a combined model P having a minimum error:

$$\text{Min} \sum_{b=1}^{B} (\hat{p}_b - p)^2, \quad (2)$$

where $\hat{P}_b$ is obtained based on $P(w_1|x)$ in (1), and B is the total number of bootstrap resamples ($b = 1, \ldots, B$).

Many researchers compared the performances of various classifier combination methods using real data sets. But the results may change according to given data characteristics. In order to generalize the results, a simulation study is necessary based on the controlled design.

3. Experimental design

We use a Monte-Carlo simulation based on a Taguchi design which accommodates not only controllable factors but also uncontrollable factor in the experiment.

3.1. Taguchi design

In our experiment, design factors are used to represent various data characteristics as well as the classifiers.

Five factors used in the simulation are: (a) combination function among input variables, (b) correlation between input variables, (c) variance of observation, (d) training set size, and (e) classification algorithms. Among these factors, (a) combination function is considered as a uncontrollable factor since it is not typically known from the given data.

In our simulation, we assume the following. There are 10 input variables x_k ($k = 1, \ldots, 10$) and outcome variable with two possible classes $\Psi = \{w_0, w_1\}$ where the 10 x_k are generated from a multivariate normal distribution $(0, \sigma^2\Omega)$. The true relationship between the output and the 10 input variables, $P(w_1|x)$, is assumed to follow a linear logistic model $f(x)$. Details regarding the levels used for each factor are as follows.

(a) *Combination function*: We consider two kinds of combination functions where the true activation function between input and output is assumed to be logistic function:

$$p(w_1|x) = \frac{\exp(f_i(x))}{1 + \exp(f_i(x))} \quad \text{for } i = 1, 2. \tag{3}$$

Here $f_i(x)$ takes one of the following two types of combination functions:

$$\begin{aligned} f_1(x) = {} & 0.03\log(x_1) - 0.09\log(x_2) + 0.07\log(x_3) \\ & - 0.01\log(x_4) + 0.02\log(x_5) - 0.16\log(x_6) \\ & + 0.01\log(x_7) - 0.21\log(x_8) + 0.31\log(x_9) \\ & - 0.11\log(x_{10}) + \varepsilon, \end{aligned}$$

$$\begin{aligned} f_2(x) = {} & \sin(x_1x_2) + 0.01(x_3 - 0.2)^2 + 0.01x_6(x_4 - x_5) \\ & + 0.02x_7 - 0.04(x_8 + x_9) + 0.8x_{10} + \varepsilon, \end{aligned}$$

where $\varepsilon \sim N(0, 1)$.

(b) *Correlation between input variables*: Generating the 10 input variables from multivariate normal distribution with mean 0 and variance–covariance matrix $\sigma^2\Omega$, we use three levels of correlation matrix Ω (weak, medium and large) as displayed in Tables 1–3. For the weak correlation, we use the range of individual correlation to be 0.05–0.3 while 0.4–0.6 for the medium, and 0.7–0.92 for large correlation.

(c) *Variance of input variables* σ^2: In order to assess the effects of variation in data sets on various classifier combination methods, we use three levels of σ^2 to be 1, 10, and 100. The resulting ranges of distributions of $P(w_1|x)$ are 0.37–0.62, 0.15–0.75, 0.05–0.95, respectively.

(d) *Size of training data set*: Sample size may play an important role in classification accuracy [19]. In the previous empirical research, Bauer and Kohavi [15] used training data sets ranging from 53 to 1620 times of the dimension of input variables in the base learning algorithms such as decision tree and Bayes classifier. Webb [20] used training set size to be 3 to 3488 times larger than that of input variables in decision tree. Raudys and Jain [19] used data size which corresponds to 1.4 times of the dimension of input variables in Euclidean distance (ED) classifier. However, logistic model fitting would fail with this size of data. In their study, other classifiers (Fisher's linear classifier, nearest neighbor classifier, etc.) used the training set size ranging from 4 to 100 times of the dimension of input variables.

We consider two levels of relative training data set size to be 5 and 100 times of input dimension, respectively:

- Small Training Data Set 50.
- Large Training Data Set 1000.

Then we use additional 50 observations for test data. All subset data sets for training and test are extracted from the identical original data set.

(e) *Classification methods*: We use five kinds of classification algorithms (bagging, modified random subspace method, classifier selection, parametric fusion, logistic regression) to compare their classification accuracy. Base algorithm is a logistic model. The number of bootstrap resamples used for ensemble algorithms is fifty based on Breiman [1,21]. For modified random subspace method, we randomly select half of input variables (five) for a logistic model training based on each bootstrap resample. For classifier selection, K-means clustering algorithm is used to find five clusters based on the ten input variables.

In summary, the design factors and their levels are summarized in Table 4.

Therefore, we have a total of four controllable factors ($2 \times 3 \times 3 \times 5$) as well as one uncontrollable factor (combination function) with two levels each. We put controllable factors in a $2 \times 3 \times 3 \times 5$ inner array and this array is replicated two times according to an outer array formed by the levels of uncontrollable factor. And this whole procedure is

Table 1
Weak correlation between input variables Ω

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1	0.07	0.27	0.2	0.13	0.01	0.18	0.11	0.24	0.08
x_2	0.07	1	0.07	0.16	0.27	0.16	0.24	0.05	0.20	0.13
x_3	0.27	0.07	1	0.18	0.09	0.24	0.14	0.29	0.05	0.16
x_4	0.20	0.16	0.18	1	0.05	0.21	0.27	0.13	0.01	0.21
x_5	0.13	0.27	0.09	0.05	1	0.02	0.03	0.19	0.12	0.04
x_6	0.01	0.16	0.24	0.21	0.02	1	0.28	0.21	0.16	0.21
x_7	0.18	0.24	0.14	0.27	0.03	0.28	1	0.12	0.27	0.26
x_8	0.11	0.05	0.29	0.13	0.19	0.21	0.12	1	0.27	0.03
x_9	0.24	0.20	0.05	0.01	0.12	0.16	0.27	0.27	1	0.07
x_{10}	0.08	0.13	0.16	0.21	0.04	0.21	0.26	0.03	0.07	1

Table 2
Medium correlation between input variables Ω

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1	0.41	0.53	0.39	0.6	0.59	0.39	0.45	0.41	0.45
x_2	0.41	1	0.60	0.60	0.45	0.51	0.59	0.41	0.52	0.51
x_3	0.53	0.60	1	0.47	0.51	0.42	0.47	0.42	0.38	0.46
x_4	0.39	0.60	0.47	1	0.39	0.47	0.51	0.42	0.58	0.56
x_5	0.6	0.45	0.51	0.39	1	0.38	0.42	0.51	0.57	0.36
x_6	0.59	0.51	0.42	0.47	0.38	1	0.44	0.41	0.47	0.45
x_7	0.39	0.59	0.47	0.51	0.42	0.44	1	0.60	0.39	0.51
x_8	0.45	0.41	0.42	0.42	0.51	0.41	0.60	1	0.38	0.39
x_9	0.41	0.52	0.38	0.58	0.57	0.47	0.39	0.38	1	0.47
x_{10}	0.45	0.51	0.46	0.56	0.36	0.45	0.51	0.39	0.47	1

Table 3
Large correlation between input variables Ω

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1	0.81	0.75	0.79	0.72	0.92	0.77	0.82	0.86	0.81
x_2	0.81	1	0.92	0.71	0.74	0.86	0.72	0.74	0.92	0.77
x_3	0.75	0.92	1	0.72	0.75	0.71	0.88	0.72	0.71	0.88
x_4	0.79	0.81	0.72	1	0.75	0.77	0.71	0.81	0.77	0.91
x_5	0.72	0.74	0.75	0.75	1	0.81	0.83	0.81	0.71	0.76
x_6	0.92	0.86	0.71	0.77	0.81	1	0.71	0.75	0.75	0.74
x_7	0.77	0.72	0.88	0.71	0.83	0.71	1	0.85	0.81	0.76
x_8	0.82	0.74	0.72	0.81	0.81	0.75	0.85	1	0.77	0.78
x_9	0.86	0.92	0.71	0.77	0.71	0.75	0.81	0.77	1	0.71
x_{10}	0.81	0.77	0.88	0.91	0.76	0.74	0.76	0.78	0.71	1

Table 4
Design factors and levels

Factors	Levels		
Combination function	$f_1(x)$	$f_2(x)$	
Correlation between input variables	0.05–0.3	0.4–0.6	0.72–0.92
Variance of observation (σ^2)	1	10	100
Classification algorithm	Bagging, Modified RSM	Parametric fusion, Classifier selection	Logistic regression
Training set size	5 times of input variable	100 times of input variables	

again repeated 10 times by generating random errors in (1). Therefore, we have a total of 1800 runs and at each run the signal-to-noise ratio (S/N) of classification accuracy is observed as a response so that one can select the best robust classification method. Since larger accuracy is better, we take the following signal to noise ratio [22,23]:

$$SN_i = \frac{1}{n}\sum_{j=1}^{n} SN_{ij} = \frac{1}{n}\sum_{j=1}^{n}\left[-10\log\left\{\frac{1}{2}\sum_{k=1}^{2}\frac{1}{y_{ijk}^2}\right\}\right], \quad (4)$$

where i is each treatment ($i = 1, \ldots, 90$), j is the replication ($j = 1, \ldots, 10$), k is the type of combination functions and y is classification accuracy.

We use the ANOVA of SN ratio for the Taguchi design in order to identify significant factor effects. From this experimentation, we are not only interested in the four kinds of main effects (correlation, variance, classification method, and training set size) on the classification accuracy but also some interaction effects.

3.2. Hypotheses

The main research hypotheses and reasons behind those are given as H_1–H_4.

(H_1) When the training set size is large, performance of logistic regression would be better than the other three methods.

Generally, classifier combination methods would not be able to take advantage of combining when data size is large. Some of the classifier combination methods may loose the information due to the fact that they use partial variables and observations. So single classifiers such as logistic regression would be a better choice when data size is large.

(H_2) When the training set size is small, performance of bagging and modified RSM would be better than the other three methods.

In case of small data size, bagging is known as an accurate and stable classifier combination method because it combines individual classifiers based on bootstrap resamples. Also, modified RSM is designated to perform well relatively in case of extremely small data by using only the half of the input variables in each bootstrap resample.

(H_3) When the training set size is small and the correlation is strong, performance of the modified RSM would be better than the other three methods.

It is natural to expect that the performance of the modified RSM would be better when the data set size is small and the correlation between input variables is high, because it uses reduced variable dimension in each bootstrap resample.

(H_4) When the variance of input data is large, classifier selection would perform better than the rest of them due to their robustness. On the contrary, when the variance of input data is small, parametric fusion would perform better than the other four methods.

We expect that the classifier selection would be better than the other four methods when applied to the data with large variance because it makes local expert classifiers which are specialized in specific area of wide input space. On the contrary, parametric fusion would be better off when the variance is small, because it will help fitting of meta level parameter.

4. Results of Monte Carlo simulation

Results of Monte Carlo simulation based on Taguchi design are obtained in the form of S/N ratio as given in Eq. (4). We use ANOVA to find significant factors. Results are obtained at 5% significance level (see Table 5). Also, Duncan test is conducted for multiple comparison of classifiers according to data characteristic (see Appendices A and B).

Since we are interested in any effects related to classification algorithm (F_1), we try to explain significant effects with highest order involving (F_1). That is classification method × training set size × correlation ($F_1 \times F_2 \times F_3$) and classification method × correlation × variance ($F_1 \times F_3 \times F_4$).

First, the interaction effect ($F_1 \times F_2 \times F_3$) is related to H_1–H_3. As shown in Appendix A, when training set size is large, performance of logistic regression is not significantly different from that of bagging (H_1). However when training set size is small, logistic regression is worse than bagging (H_2). These results present bagging is effective for small training sample because combining result based on bootstrap can complement unstable classifiers. In terms of H_3, when small data has strong correlation between variables, modified RSM performs better than single classifier, classifier selection, parametric fusion. Also, in this case, modified RSM has no significantly different performance from bagging. It would be due to the fact that both modified RSM and bagging use bootstrap sampling from small data. Also, in case of modified RSM, strong correlation between variables appear to compromise partial loss of input information. The result of H_3 agrees with study by Ref. [8].

Secondly, in terms of $F_1 \times F_3 \times F_4$, when correlation is weak and variance is small, both classifier selection and parametric fusion combining are the worst choice (H_4). It is observed that the accuracy of combining parametric fusion drops when the estimated parameter detects any peculiar outlier among one or more bootstrap samples. In case of

Table 5
Analysis of variance for Taguchi design

Source of variation	Degree of freedom	Sum of square	Mean square	F-value	P-value
F_1	4	178.76	44.69	126.82	0.0001
F_2	1	55.13	55.13	156.47	0.0001
$F_1 \times F_2$	4	3.81	0.95	2.71	0.0294
F_3	2	2.13	1.06	3.03	0.0487
$F_1 \times F_3$	8	15.63	1.95	5.55	0.0001
$F_2 \times F_3$	2	1.71	0.85	2.43	0.0886
$F_1 \times F_2 \times F_3$	8	6.84	0.85	2.43	0.0135
F_4	2	3.24	1.62	4.60	0.0103
$F_1 \times F_4$	8	10.12	1.26	3.59	0.0004
$F_2 \times F_4$	2	0.32	0.16	0.46	0.6329
$F_1 \times F_2 \times F_4$	8	1.93	0.24	0.69	0.7032
$F_3 \times F_4$	4	47.12	11.78	33.44	0.0001
$F_1 \times F_3 \times F_4$	16	32.33	2.02	5.74	0.0001
$F_2 \times F_3 \times F_4$	4	3.34	0.83	2.38	0.0507
$F_1 \times F_2 \times F_3 \times F_4$	16	3.42	0.21	0.61	0.8790

F_1: Classification algorithm; F_2: training set size; F_3: correlation of variables; F_4: variance of input data.

classifier selection, we found that four clusters are the best fit in terms of accuracy for our test data. In general, however, classifier selection shows low performance. It would be due to the fact that some clusters consist of only one class.

5. Discussion

In this research we compare the classification accuracy of logistic regression, bagging, modified RSM, classifier selection, and parametric fusion under various combinations of data characteristics. Four factors ((a) combination function, (b) correlation between input variables, (c) variance of observation, and (d) training set size) were used to generate data while the combination function was used as a uncontrollable factor in Taguchi design. Results of the Taguchi experiment indicated the following at 5% significance level. First, training set size, correlation among input data, and variance are significant factors for classification accuracy. Secondly, when the training data is small and there is strong correlation between input variables, both bagging and modified RSM perform better than the other three methods. Thirdly, when correlation is weak and variance is small, both classifier selection and parametric fusion methods appear to be the worst.

In order to utilize the lessons we learnt from this experimentation, in practice, one can do a preliminary analysis of data at hand and check a training set size in relation to the number of input variable, and correlation as well as variation. Training set size and three levels of correlation as well as those of variance used in our simulation can be used as good indicators to selecting an appropriate classification algorithm.

In this comparison, we use a logistic model as a basis. It can be extended to other classification algorithms such as neural network and decision tree. Results of ensemble, classifier selection, combining classifier based on these other individual algorithms may be different from what we obtained in this study. Also, different distribution may have an effect on classification accuracy in ensemble algorithm. This simulation result is based on the generated data which can be extended by adjusting control parameters such as the number of classes and number of classification features as well. They can be used as the other factors in the design of experiment. Such extension is left for further study areas.

Acknowledgment

This work was supported by Grant No. (R04-2002-000-20003-0) from the Basic Research Program of the Korea Science & Engineering Foundation.

Appendix A

The classification method × training data size × correlation ($F_1 \times F_2 \times F_3$) is given in Table A1.

Appendix B

The classification method × correlation × variance ($F_1 \times F_3 \times F_4$) is given in Table B1.

Table A1

Duncan grouping					Mean of S/N ratio	Classification method	Data size	Correlation
			A		41.14	Bagging	Large	Strong
	B		A		41.01	Bagging	Large	Medium
	B		A	C	41.01	Logistic	Large	Medium
	B		A	C	40.95	RSM	Small	Strong
	B	D	A	C	40.93	Logistic	Large	Weak
	B	D	A	C	40.85	RSM	Large	Weak
E	B	D	A	C	40.80	Bagging	Small	Strong
E	B	D		C	40.78	RSM	Large	Strong
E	B	D		C	40.77	Bagging	Large	Weak
E	B	D		C	40.69	RSM	Large	Medium
E	B	D		C	40.68	Bagging	Small	Medium
E		D	F	C	40.66	Logistic	Large	Strong
E		D	F		40.58	Bagging	Small	Weak
E	G		F		40.46	Logistic	Small	Medium
	G		F	H	40.38	Logistic	Small	Weak
	G		I	H	40.17	Classifier selection	Large	Medium
	G		I	H	40.15	Subspace	Small	Medium
	G		I	H	40.14	Parametric fusion	Large	Strong
			I	H	40.14	Classifier selection	Large	Weak
			I	H	40.06	Classifier selection	Large	Strong
			I		40.04	Logistic	Small	Strong
			I		40.04	Parametric fusion	Large	Weak
			I		40.01	RSM	Small	Weak
			I		39.92	Parametric fusion	Large	Medium
			J		39.62	Classifier selection	Small	Medium
			J		39.52	Classifier selection	Small	Strong
			J		39.50	Classifier selection	Small	Weak
			J		39.50	Parametric fusion	Small	Strong
			J		39.37	Parametric fusion	Small	Weak
			J		39.31	Parametric fusion	Small	Medium

Means with the same letter are not significantly different.

Table B1

Duncan grouping					Mean of S/N ratio	Classification method	Correlation	Variance
			A		41.57	Logistic	Medium	Medium
			A		41.45	Bagging	Medium	Medium
	B		A		41.33	Bagging	Weak	Small
	B		A	C	41.19	Bagging	Strong	Large
	B		D	C	41.04	Logistic	Weak	Small
	B		D	C	41.02	RSM	Weak	Small
	E		D	C	40.88	Bagging	Strong	Small
	E		D	C	40.88	RSM	Strong	Small
	E		D	C	40.86	RSM	Strong	Medium
	E	F	D	C	40.85	RSM	Strong	Large
	E	F	D	C	40.84	Bagging	Strong	Medium
G	E	F	D	C	40.78	Logistic	Strong	Large
G	E	F	D	H	40.70	Bagging	Medium	Small
G	E	F	D	H	40.69	RSM	Medium	Medium
G	E	F	I	H	40.55	RSM	Medium	Large
G	E	F	I	H	40.50	Logistic	Weak	Large
G	E	F	I	H	40.47	RSM	Weak	Large
G	J	F	I	H	40.43	Logistic	Medium	Large
G	J	F	I	H	40.43	Logistic	Weak	Medium
G	J	K	I	H	40.38	Bagging	Medium	Large
G	J	K	I	H	40.36	Bagging	Weak	Large
G	J	K	I	H	40.35	Bagging	Weak	Medium
	J	K	I	H	40.31	Logistic	Strong	Small
	J	K	I	L	40.20	Logistic	Medium	Small
	J	K	M	L	40.04	Classifier selection	Medium	Medium
	J	K	M	L	40.02	RSM	Medium	Small
	J	K	M	L	40.02	Classifier selection	Weak	Large

Table B1 (*Continued*)

Duncan grouping				Mean of S/N ratio	Classification method	Correlation	Variance
J	K	M	L	40.02	Parametric fusion	Strong	Large
N	K	M	L	39.96	Classifier selection	Strong	Medium
N	K	M	L	39.95	Logistic	Strong	Medium
N		M	L	39.89	Parametric fusion	Strong	Medium
N		M	L	39.89	Classifier selection	Weak	Medium
N		M	L	39.88	Classifier selection	Strong	Large
N		M	L	39.88	Classifier selection	Medium	Large
N		M	L	39.82	Parametric fusion	Medium	Large
N		M	L	39.81	Parametric fusion	Weak	Small
N		M	L	39.80	RSM	Weak	Medium
N		M	L	39.77	Classifier selection	Medium	Small
N		M		39.74	Parametric fusion	Weak	Medium
N		M		39.74	Parametric fusion	Medium	Medium
N		O		39.57	Parametric fusion	Weak	Large
N		O		39.55	Parametric fusion	Strong	Small
N		O		39.55	Classifier selection	Weak	Small
N		O		39.54	Classifier selection	Strong	Small
		O		39.29	Parametric fusion	Medium	Small

Means with the same letter are not significantly different.

References

[1] L. Breiman, Bagging predictors, Technical Report No. 421, University of California at Berkeley, 1994.

[2] L. Breiman, Bias, variance, and arcing classifiers, Technical Report No. 460, University of California at Berkeley, 1996.

[3] C.I. Christodoulou, C.S. Pattichis, Combining neural classifications in EMG diagnosis, in: Proceeding of the Conference on EUFIT, 1998, pp. 1837–1841.

[4] T.G. Dietterich, An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization, Mach. Learn. 40 (2) (2000) 139–158.

[5] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 21 (11) (1998) 832–844.

[6] J. Kittler, M. Hatef, R.P.W. Duin, J. Mtas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.

[7] M. Skurichina, R.P.W. Duin, Bagging for linear classifiers, Pattern Recognition 31 (7) (1998) 909–930.

[8] M. Skurichina, R.P.W. Duin, Bagging, boosting and the random subspace method for linear classifiers, Pattern Anal. Appl. 5 (2002) 121–135.

[9] J. Cao, M. Ahmadi, M. Shridhar, Recognition of handwritten numerical with multiple feature and multistage classifier, Pattern Recognition 28 (2) (1995) 153–160.

[10] G. Giacinto, F. Roli, Dynamic classifier selection based on multiple classifier behaviour, Pattern Recognition 34 (2001) 1879–1881.

[11] L.I. Kuncheva, Cluster-and-selection method for classifier combination, Proceedings of the Fourth International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies (KES'2000), 2000, pp. 185–188.

[12] R. Liu, B. Yuan, Multiple classifiers combination by clustering and selection, Inform. Fusion 2 (3) (2001) 163–168.

[13] K. Woods, W.P. Kegelmeyer, K. Bowyer, Combination of multiple classifiers using local accuracy estimates, IEEE Trans. Pattern Anal. Mach. Intell. 19 (4) (1997) 405–410.

[14] W.D. Shannon, D. Banks, Combining classification trees using MLE, Statist. Med. 18 (1999) 727–740.

[15] E. Bauer, R. Kohavi, An empirical comparison of voting classification algorithms: bagging, boosting, and variants, Mach. Learn. 36 (1999) 105–139.

[16] S.Y. Sohn, Meta analysis of classification algorithm for pattern recognition, IEEE Trans. Pattern Anal. Mach. Intell. 21 (11) (1999) 1137–1144.

[17] S.Y. Sohn, S.H. Lee, Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea, Safety Science 41 (1) (2003) 1–14.

[18] K.L. Mcfadden, Predicting pilot-error incidents of US airline pilots using logistic, Appl. Ergon. 28 (3) (1997) 209–212.

[19] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Trans. Pattern Anal. Mach. Intell. 13 (3) (1991) 252–264.

[20] G.I. Webb, Multiboosting: a technique for combining boosting and wagging, Mach. Learn. 40 (2000) 159–196.

[21] L. Breiman, Arcing classifiers, Technical Report No. 486, University of California at Berkeley, 1996.

[22] D.H. Chen, Y.N. Sun, A self-learning segmentation framework—the Taguchi approach, Computerized Med. Imaging Graph. 24 (2000) 283–296.

[23] J.F.C. Khaw, B.S. Kim, L.E.N. Lim, Optimal Design of Neural Networks Using the Taguchi method, Neurocomputing 7 (1995) 225–245.

About the Author—SO YOUNG SOHN is a professor of Industrial Engineering. Her current research areas include Data Mining, Technology Management, and Quality and Reliability Data Analysis. Detailed publication information can be found at http://isl.yonsei.ac.kr.

About the Author—HYUNG WON SHIN is a senior researcher of the strategic management department of the Samsung Economic Research Institute (SERI). His research interests include classifier combination and its application to data mining.