



Journal of
**Software
Engineering**

ISSN 1819-4311



Academic
Journals Inc.

www.academicjournals.com



Research Article

Selective Ensemble Learning Based on Greedy Reduction with Diversity

Yan Zhang, Danjv Lv and Ran Guo

School of Computer and Information, Southwest Forestry University, Kunming, China

Abstract

Background: Classifier selection is a crucial problem for ensemble learning to speed up the classifier prediction, reduce the storage space requirements and to further improve the classification accuracy. **Materials and Methods:** To select the best classifier subset from a pool of classifiers, the diversity metrics are exploited to evaluate the classifier for ensemble classification. Based on accuracy and differences between classifiers, this study presented AD and concurrency metrics to measure diversity of ensemble classifiers. With the greedy search method, the back-forward reduction with diversity (BRD) is proposed to delete the classifier with smallest diversity with evaluation function according to the diversity metrics. Experiments were conducted on ten data sets from UCI, remote sensing image and environmental audio data. Bagging, AdaBoost and MCS ensemble strategies were exploited in those experiments. **Results:** Six diversity metrics were involved in the BRD method to select the optimal subset individual classifiers to constitute the final ensemble. Ensemble selection method BRD outperformed Multiple Classifiers System (MCS) in classification performance and generalization. **Conclusion:** The BRD is effective method to ensemble selection to improve the performance and generalization in classification. Considering both accuracy and difference of classifiers, the concurrency and AD measure methods obtained the better classification results in ensemble selection than that of the other four diversity metrics including the Q-statistics, Kappa-statistic, correlation coefficient and dis diversity metrics.

Key words: Selective ensemble, diversity metrics, diversity matrix, ensemble methods

Received: May 18, 2016

Accepted: September 05, 2016

Published: December 15, 2016

Citation: Yan Zhang, Danjv Lv and Ran Guo, 2017. Selective ensemble learning based on greedy reduction with diversity. *J. Software Eng.*, 11: 60-65.

Corresponding Author: Yan Zhang, School of Computer and Information, Southwest Forestry University, Kunming, China

Copyright: © 2017 Yan Zhang *et al.* This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Ensemble learning methods^{1,2} train a set of base learners and combine their predictions to make the final decision. Due to their potential to greatly increase classification accuracy and generalization ability, the study of the methods for construction good ensembles has attracted a lot of attentions in machine learning community and successfully applied in many real world applications³⁻⁵.

Compared with the individual classifier, ensemble classifiers could obtain better performance but some disadvantages exist. With the increasing of the number of the base classifiers, training and predictive processing is time-consuming and the computational cost is high. Besides, not any ensemble classifier member is beneficial to the classification result. Zhou *et al.*⁶ firstly proposed the concept of selective ensemble learning, which is giving a set of trained individual classifiers, instead of combining all of them. Ensemble selection tries to select a subset of individual classifiers to comprise the final ensemble. In the selective ensemble learning, the storage resources required for storing the base classifiers reduces when the size of ensemble is smaller and for calculating outputs of individual classifiers, the efficiency in the training and prediction is improved. In the selective ensemble process, the weak or bad base classifiers are deleted. The generalization performance of selected ensemble may be even better than the ensemble consisting of all the given individual classifiers. The success of ensemble selection lies in how to select the optimal individual classifiers to improve performance with higher accuracy and stronger generalization. The hotspot attracts many researchers to focus the issue and presented some effective algorithm to select the base classifiers in ensemble methods⁷⁻¹².

MATERIALS AND METHODS

The central problem of ensemble selection research is how to design practical algorithms leading to smaller

ensembles without sacrificing or even improving the generalization performance contrasting to all-member ensembles.

Different type classifiers are trained by the same training data, respectively and combined with some strategies. The final classifier classifiers new data and gives their label by predictions. The ensemble methods comprise three phases:

- Production of multiple diverse base classifiers
- Ensemble selection or ensemble pruning
- Combination of the final base classifiers

The model of ensemble selection method process is shown as Fig. 1. The key step is the selection strategies.

The main idea of ensemble selection learning is that some are selected to build the ensemble classifier from an existing base classifier set based on some criteria, to speed up the classifier prediction, reduce the storage space requirements and to further improve the classification accuracy. In theory, the optimal classifier subset can be obtained through the exhaustive enumeration. But when the number of base classifiers is larger, its calculation is time-consuming and is not feasible in practice. Therefore, effective evaluation method to select the useful ones to construct the final ensemble the is the key step to ensemble selection.

During the past decade, many effective ensemble selection methods have been proposed. Those methods can be classified into three categories^{2,13}:

- **Ordering-based selection:** The base classifiers are ordered based on some criterion. The final ensemble will only include classifiers in the front-part
- **Clustering-based selection:** Those methods try to identify a number of representative prototype individual classifiers to constitute the final ensemble
- **Optimization-based selection:** Those methods consider the ensemble selection problem as an optimization problem, which tries to find the subset of individual classifiers that maximizes or minimizes a classifier related to the generalization ability of the final ensemble

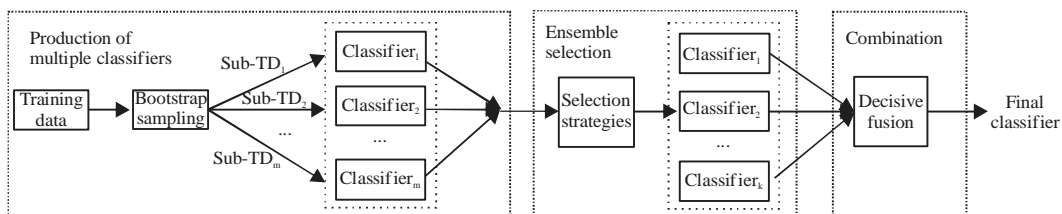


Fig. 1: Model of ensemble selection method process

Some metrics for diversity: Kuncheva and Whitaker¹⁴ compared 10 statistics that can be applied to the measurement of diversity. They look at 4 statistics that are averaged pair-wise results and 6 that are non-pair-wise result. In this study, the 4 pair-wise statistics are exploited and other two metrics are presented.

To measure ensemble diversity, a classical approach is to measure the pairwise dissimilarity between 2 learners and then average all the pairwise measurements for the overall diversity⁴. Given a data set $D = \{(x_1, y_1), \dots, (x_m, y_m)\}$, for 2 classifier h_i and h_j , some characters are denoted:

$$a = \sum_{(k=1)}^m I(h_i(x_k) = y_k \cap h_j(x_k) = y_k) \quad (1)$$

$$b = \sum_{(k=1)}^m I(h_i(x_k) = y_k \cap h_j(x_k) \neq y_k) \quad (2)$$

$$c = \sum_{(k=1)}^m I(h_i(x_k) \neq y_k \cap h_j(x_k) = y_k) \quad (3)$$

$$d = \sum_{(k=1)}^m I(h_i(x_k) \neq y_k \cap h_j(x_k) \neq y_k) \quad (4)$$

• **Disagreement measure:**

$$dis_{ij} = \frac{b+c}{m} \quad (5)$$

The value dis_{ij} is in $[0, 1]$ as well as the larger the value, the large the diversity.

• **Q-statistic:**

$$Q_{ij} = \frac{ad-bc}{ad+bc} \quad (6)$$

The value Q_{ij} is in the range of $[-1, 1]$:

$$Q_{ij} \begin{cases} = 0, & h_i \text{ and } h_j \text{ are independent} \\ > 0, & h_i \text{ and } h_j \text{ make similar predictions} \\ < 0, & h_i \text{ and } h_j \text{ make different predictions} \end{cases}$$

• **Correlation coefficient:**

$$\rho_{ij} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(c+d)(b+d)}} \quad (7)$$

This is a classic statistic for measuring the correlation between two binary vectors. The value indicates that the smaller the measurement, the larger the diversity.

• **Kappa-statistic:**

It is defined as :

$$Kp = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad (8)$$

where, θ_1 and θ_2 are the probabilities that the two classifiers agree and agree by chance, respectively. The probabilities for h_i and h_j can be estimated on the data set D:

$$\theta_1 = \frac{a+d}{m} \quad (9)$$

$$\theta_2 = \frac{(a+b)(a+c) + (c+d)(b+d)}{m^2} \quad (10)$$

The value indicates that the smaller the measurement, the larger the diversity.

• **Accuracy and diversity (AD):** In order to simplify the calculating process, the metric Accuracy and Diversity (AD) can be proposed to take both accuracy and diversity of two classifiers into account. The metric AD is defined as:

$$AD_{ij} = \frac{\alpha(a-d) + \beta(c+b)}{m} \quad (11)$$

where, a, b, c, d and m are the same definition over above. The first term of numerator defines the accuracy of both 2 classifiers, while the second term is the difference between them and α, β are the coefficient to trade-off the accuracy and diversity. Their value is in range $[0, 1]$ and $\alpha + \beta = 1$. The diversity increases as the value of AD_{ij} metric becomes larger.

• **Concurrency:**

$$CON_{ij} = 2 \times a - 2 \times d \quad (12)$$

This metric is derived from the concurrency of a model h_t with respect to a sub-ensemble S in literature Banfield⁹, which based on the correctness of both the ensemble and the classifier with regard to a thinning set. In this study, the concurrency metric is defined as between classifier h_i and h_j .

```

Algorithm BRD (Back-forward reduction with diversity)
Input: Training set L, test set T, original base classifiers C = {c1, c2, ..., cn},
Output: New base classifiers, S = {c1*, c2*, ..., cn*}, n < t

1. For i = 1 to t do
2. Li = Bootstrap(L);
3. Train each ci ∈ C with Li;
4. End For
5. Calculate the diversity matrix DM for C based on measure method;
6. Though DM, di =  $\frac{1}{t} \sum_{j=1}^t dm_{ij}$  (i = 1, ..., n);
7. S = C;
8. E0 = Error(Ensemble(S));
9. while S ≠ ∅ do
10. h = arg minhk ∈ S f(S, dk);
11. E1 = Error(Ensemble(S-{h}));
12. If E1 < E0 then
13. E0 = E1; S = S-{h};
14. End If
15. End While
16. Return S
    
```

Fig. 2: Description of back-forward reduction with diversity process

The h_i is rewarded for obtaining a correct decision, the rewarded more for obtaining a correct decision when the classifier h_j is incorrect. Both h_i and h_j are incorrect, there is a penalization. For CON_{ij} metric, that is the larger the value, the large the diversity.

Measure matrix: According to the methods to measure the diversity between two different classifiers, the Diversity Matrix (DM) can be defined as:

$$DM = \begin{bmatrix} dm_{11} & \dots & dm_{1m} \\ \vdots & \ddots & \vdots \\ dm_{m1} & \dots & dm_{mm} \end{bmatrix} \quad (13)$$

where, $dm_{ii} = 0$ ($i = 1, \dots, m$), $dm_{ij} = dm_{ji}$ ($i, j = 1, \dots, m$).

Each dm_{ij} ($i \neq j$) denotes the diversity two classifiers h_i and h_j . The pair-wise statistic metric can be obtained as the value for dm_{ij} ($i \neq j$), so can the dm_{ji} . So DM is a symmetry matrix.

For a classifier h_i in ensemble, its diversity can be simply acquired by:

$$d_i = \frac{1}{m} \sum_{(j=1)}^m dm_{ij} \quad (14)$$

Classifiers selection on elimination with diversity: Greedy ensemble selection algorithm attempts to find the globally

best subset of classifiers by taking local greedy decision for changing the current subset. In this study, backward elimination is exploited in the greedy selection. Firstly, the current classifier subset S is initialized to the complete base classifier set H. Then, at each iteration, the classifier $h_t \in S$ that optimizes the evaluation function f will be removed from S to improve the classification accuracy. With the idea of greedy methods, the evaluation function f selects the classifier h_t which has the smallest diversity in the current subset S. The iterative process will not stop until the error of ensemble subset S start to increase. This process is named as back-forward reduction with diversity (BRD). The description of BRD is illustrated in Fig. 2.

RESULTS AND DISCUSSION

In order to compare the performance of the above methods, some experiments are carried out to show the results of the classification. Eight UCI datasets¹⁵, a remote sensing image and environmental audio data are used in the experiments. The dataset TM_RS comes from the remote sensed data, including six kinds of type of land, such as forest, grass, arable land, sand, bare land and others. The environmental audio data, named as EnAudio, is acquired from network and field recording, with 8k sampling rate, 16 bits and mono-track. It includes five classes, the sound of birds, frogs, wind, rain and thunder. The audio length amounts to almost 10 min. In the pre-processing, silence and noise are removed. Information on these data is given in Table 1.

Method of experiment: The single classifier includes the traditional classifying methods such as decision tree J48, neural network and Radius Basic Function (RBF), while Bagging, AdaBoost and Multiple Classifier System (MCS) are involved in the ensemble strategies. Ten base classifiers are used as candidates in the ensemble selection. The WEKA machine learning library¹⁶ was used as the source of learning algorithms. Four multilayer perceptrons (MLPs), 4 decision trees (J48), one Radium Basis Function (RBF) classifier and one Naïve Bayes classifier were trained. The different parameters used to train the algorithms for MPLs and decision trees are listed in Table 2 and 3. The rest of the parameters were left unchanged in their default values.

Ensemble classifiers: Table 4 are illustrated the results of the classification accuracy of three ensemble strategies. The J48, MPLs and RBF are base classifier for Bagging and AdaBoost respectively and the default parameters are explored in the J48 and MPLs. The number of iterations is 10. The MCS is

ensemble of 10 individual classifiers, whose result is obtained by majority voting. The Overall Accuracy (OA) of three ensemble methods is listed in Table 4.

Table 1: Characteristics of the used datasets

Dataset	Train No.	Test No.	Attributes	Classes
Diabetes	384	192	8	2
Heart	120	68	13	2
Ionosphere	263	88	34	2
Tic-tac-toe	144	240	9	2
Sonar	104	52	60	2
Vehicle	424	212	18	4
Vote	163	109	16	2
Wine	90	45	13	3
TM_RS	586	2347	6	6
EnAudio	684	2284	23	5

Table 2: Different parameter in MLPs

Algorithm	Parameters	
	M	H
MPLs_1	0.3	4
MPLs_2	0.3	8
MPLs_3	0.4	4
MPLs_4	0.4	8

Table 3: Different parameter in decision trees

Algorithm	Parameters	
	M	H
J48_1	0.25	4
J48_2	0.25	8
J48_3	0.40	4
J48_4	0.40	8

Table 4: Classification accuracy for three ensemble methods on each dataset

Dataset	Bagging			AdaBoost			
	J48	MLPs	RBF	J48	MLPs	RBF	MCS
Diabetes	0.7365	0.7510	0.7609	0.7089	0.7328	0.7508	0.7667
Heart	0.8103	0.8044	0.8206	0.7956	0.7897	0.8147	0.8206
Ionosphere	0.9432	0.9091	0.9318	0.9091	0.9318	0.9432	0.9659
Tic-tac-toe	0.8385	0.9983	0.6817	0.7420	0.9933	0.7125	0.8821
Sonar	0.8038	0.7808	0.7885	0.7923	0.8173	0.8232	0.8212
Vehicle	0.7193	0.8009	0.6920	0.7396	0.8005	0.6821	0.7821
Vote	0.8569	0.82844	0.8312	0.8193	0.8239	0.8202	0.8477
Wine	0.9778	0.9956	0.9867	0.9600	0.9889	0.9889	0.9867
TM_RS	0.9350	0.9469	0.9278	0.9326	0.9408	0.9355	0.9428
EnAudio	0.9125	0.9287	0.9317	0.9204	0.9301	0.9444	0.9278

Table 5: Classification results of reduction with diversity in ensemble selection

Dataset	MCS	BRD-Q	BRD-Kp	BRD-Cor	BRD-DIS	BRD-AD	BRD-Con
Diabetes	0.7667	0.7745	0.7766	0.7766	0.7698	0.7677	0.7783
Heart	0.8206	0.8206	0.8294	0.8294	0.8232	0.8294	0.8236
Ionosphere	0.9659	0.9659	0.9659	0.9659	0.9773	0.9773	0.9773
Tic-tac-toe	0.8821	0.9067	0.8879	0.8879	0.8821	0.9383	0.9571
Sonar	0.8212	0.8442	0.8365	0.8365	0.8365	0.8308	0.8303
Vehicle	0.7821	0.7859	0.7859	0.7873	0.7849	0.7854	0.7863
Vote	0.8477	0.8624	0.8486	0.8486	0.8514	0.8541	0.8541
Wine	0.9867	0.9911	0.9911	0.9911	0.9889	0.9867	0.9931
TM_RS	0.9428	0.9470	0.9479	0.9479	0.9456	0.9489	0.9545
EnAudio	0.9278	0.9295	0.9347	0.9347	0.9295	0.9297	0.9326

As the results indicate, three datasets obtained best accuracy by Bagging with MLPs; one by Bagging with J48, two obtained best accuracy by AdaBoost with MLPs, one by AdaBoost with RBF and three by MCS with 10 base classifiers. Overall, the MCS could acquire better classification performance.

Select diverse base classifiers for ensemble: In order to make best use of the larger difference among the base classifiers involved in the Ensemble learning, 6 metrics including Q-statistics, Kappa-statistic, correlation coefficient, disagreement, AD and concurrency are exploited to measure the diversity of classifiers. The search method uses the back-forward reduction with diversity (BRD). So, classification results are summarized in the Table 5.

In Table 5 the 2nd column is the accuracy of MCS in all members ensembles, while the 3rd column to the last one show the results of ensemble with subset of classifiers, which selected by BRD including 6 different measure diversity metrics, respectively. For each dataset, the highest accuracy is in bold font.

The results indicate that it is better to ensemble subset classifiers instead of all of the entire base classifiers. The back-forward reduction with diversity (BRD method) outperforms the MCS in all datasets. In the six measure diversity methods, different metrics have different performance. The BRD-Con, which is concurrency measure, in 5 out of 10 datasets, obtains the highest classification

accuracy. And using BRD-AD, 2 datasets wins the highest accuracy and outperforms the DIS measure in all cases. The BRD-Q, BRD-Kp and BRD-Cor have the same accuracy in most datasets, especially BRD-Kp and BRD-Cor. Thus, the diversity measure methods, considering both accuracy and differences, concurrency and AD are the most effective methods in the ensemble selection. Compared with other metrics, the DIS is not optimal in the BRD method.

CONCLUSION AND FUTURE RECOMMENDATIONS

Selective ensemble learning trains a number of base classifiers and selects part of their outcomes according to a certain rule to assemble. How to select the optimal classifiers involved in the ensemble classification is key factor in ensemble selection performance. In this study Accuracy and Diversity (AD) and concurrency metrics are proposed to measure the optimal classifiers from the entire base classifiers as well as the Q-statistics, Kappa-statistic, correlation coefficient, diversity metrics. In the experiments, ensemble methods including Bagging, AdaBoost, MCS and Back-forward reduction diversity are carried on ten datasets. According to the analysis of experimental results, the proposed ensemble selection method BRD outperformed MCS in classification performance and generalization. Especially, considering both accuracy and difference of classifiers, the concurrency and AD measure methods obtained the better classification results in ensemble selection.

In this study, the initial base classifiers with the random setting the parameters, are not in big scale and only four kinds of the learning methods are explored in the experiments. In the further study, the effective search method will be focus on and the number of ensemble base classifiers is increased to analyze the performance and the effectiveness.

ACKNOWLEDGMENT

This study was supported by the national natural science foundation of China under the Grant No. 61462078.

REFERENCES

1. Dietterich, T.G., 1997. Machine-learning research: Four current directions. *AI Mag.*, 18: 97-136.
2. Zhou, Z.H., 2012. *Ensemble Methods Foundations and Algorithms*. CRC Press, USA., ISBN: 9781439830031, Pages: 236.
3. Zhang, C. and Y. Ma, 2012. *Ensemble Machine Learning: Methods and Application*. Springer, New York, ISBN: 9781441993250, Pages: 332.
4. Kuncheva, L.I., 2004. *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley and Sons, New York, ISBN: 9780471660255, Pages: 300.
5. Rokach, L., 2010. *Pattern Classification Using Ensemble Methods*. World Scientific, Singapore, ISBN: 9789814271073, Pages: 225.
6. Zhou, Z.H., J. Wu and W. Tang, 2002. Ensembling neural networks: Many could be better than all. *Artif. Intell.*, 137: 239-263.
7. Bi, K., X.D. Wang, X. Yao and J.D. Zhou, 2014. Adaptively selective ensemble algorithm based on bagging and confusion matrix. *Acta Electronica Sinica*, 42: 711-716.
8. Zhang, C.X. and J.S. Zhang, 2011. A survey of selective ensemble learning algorithms. *Chin. J. Comput.*, 34: 1399-1410.
9. Banfield, R.E., L.O. Hall, K.W. Bowyer and W.P. Kegelmeyer, 2005. Ensemble diversity measures and their application to thinning. *Inform. Fusion*, 6: 49-62.
10. Caruana, R., A. Niculescu-Mizil, G. Crew and A. Ksikes, 2004. Ensemble selection from libraries of models. *Proceedings of the 21st International Conference on Machine Learning*, July 2004, Banff, Canada, pp: 18-26.
11. Zhang, M., F. Di and J. Liu, 2015. Universal steganalysis based on selective ensemble classifier. *J. Sichuan Univ. (Eng. Sci. Edn.)*, 47: 36-41.
12. Partalas I., G. Tsoumakas and I. Vlahavas, 2008. Focused ensemble selection: A diversity-based method for greedy ensemble selection. *Proceedings of the 18th European Conference on Artificial Intelligence*, July 21-25, 2008, Patras, Greece, pp: 117-121.
13. Tsoumakas, G., I. Partalas and I. Vlahavas, 2008. A taxonomy and short review of ensemble selection. *Proceedings of the 18th European Conference on Artificial Intelligence*, July 21-25, 2008, Patras, Greece, pp: 1-6.
14. Kuncheva, L.I. and C.J. Whitaker, 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach. Learn.*, 51: 181-207.
15. Blake, C.L. and C.J. Merz, 1998. *UCI Repository of Machine Learning Databases*. 1st Edn., University of California, Irvine, CA.
16. Witten, I.H., E. Frank and A.H. Mark, 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd Edn., Morgan Kaufmann, San Francisco, CA., USA.