A Heuristically Perturbation of Dataset to Achieve a Diverse Ensemble of Classifiers

Hamid Parvin, Sajad Parvin, Zahra Rezaei, and Moslem Mohamadi

Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran
hamidparvin@mamasaniiau.ac.ir,
{s.parvin,rezaei,mohamadi}@iust.ac.ir

Abstract. Ensemble methods like Bagging and Boosting which combine the decisions of multiple hypotheses are among the strongest existing machine learning methods. The diversity of the members of an ensemble is known to be an important factor in determining its generalization error. We present a new method for generating ensembles, named CDEBMTE (Creation of Diverse Ensemble Based on Manipulation of Training Examples), that directly constructs diverse hypotheses using manipulation of training examples in three ways: (1) sub-sampling training examples, (2) decreasing/increasing error-prone training examples and (3) decreasing/increasing neighbor samples of error-prone training examples.

The technique is a simple, general meta-learner that can use any strong learner as a base classifier to build diverse committees. Experimental results using two well-known classifiers (1) decision-tree induction and (2) multilayer perceptron as two base learners demonstrate that this approach consistently achieves higher predictive accuracy than both the base classifier, Adaboost and Bagging. CDEBMTE also outperforms Adaboost more prominent when training data size is becomes larger.

We propose to show that CDEBMTE can be effectively used to achieve higher accuracy and to obtain better class membership probability estimates.

Experimental results using two well-known classifiers as two base learners demonstrate that this approach consistently achieves higher predictive accuracy than both the base classifier, Adaboost and Bagging. CDEBMTE also outperforms Adaboost more prominent when training data size is becomes larger.

Keywords: Classifier Ensemble, Diversity, Training Examples Manipulation.

1 Introduction

Nowadays, usage of recognition systems has found many applications in almost all fields [16-28]. There are many inherently different classifiers in the pattern recognition. It may be worthy to mention that usage of the broad versatility in kind of classifiers is not a trivial matter to be ignored. It means although a classification algorithm may obtain a good performance for a specific problem, it has not enough robustness for another problem. It has been always an ideal for pattern recognition

J.A. Carrasco-Ochoa et al. (Eds.): MCPR 2012, LNCS 7329, pp. 197–206, 2012.

communities to present an approach of how to select the best (or at least approximately the best) classifier for a specific problem. It has been shown that this desire is far from the reach; at least by employing simple classifiers as the main learner. A lot of research has been done to improve their performance. Most of these algorithms have provided a good performance for a specific problem, but they have not enough robustness for other problems.

Fortunately there is another way to cope with the problem. It is to use many learners instead of one learner and then to employ a democracy-wise method to firm the final decision of the committee (classifier). One of the major advances in inductive learning in the past decade was the development of ensemble or committee approaches that learn and retain multiple hypotheses and combine their decisions during classification [2]. For example, Boosting [3]-[4] is an ensemble method that learns a series of "weak" classifiers each one focusing on correcting the errors made by the previous one; and it is currently one of the best generic inductive classification methods [6]. Therefore, recent researches are directed to the combinational methods which have more power, robustness, resistance, accuracy and generality than simple method. Combination of Multiple Classifiers (CMC) can be considered as a general solution method for pattern recognition problems. Inputs of CMC are results of separate classifiers and output of CMC is their consensus decision [7].

Melville [10] presents a new meta-learner DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples), that uses an existing "strong" learner (one that provides high accuracy on the training data) to build an effective diverse committee in a simple, straightforward manner. This is accomplished by adding different randomly constructed examples to the training set when building new committee members. These artificially constructed examples are given category labels that disagree with the current decision of the committee, thereby easily and directly increasing diversity when a new classifier is trained on the augmented data and added to the committee.

There have been many methods developed to construct an ensemble. Some of these methods, such as Bagging and Boosting are *meta-learners* i.e. they can be applied to *any* base classifier. Other methods are specific to a particular base classifier. For example, Negative Correlation Learning [8] is used specifically to build committees of Neural Networks. We only deal with construction of general ensemble, i.e. *meta-learners*. Although a special base classifier may be better suited for a particular domain than others, a *general* ensemble approach that is independent of the particular base classifier is preferred in an unknown domain. So the aim of our research is to only propose a *meta-learner*.

In an ensemble, the output of several classifiers is useful to be participated in an ensemble only when they disagree on some inputs [5] and [14]. The measure of disagreement is called as the *diversity* of the ensemble. For regression problems, *mean squared error* is generally used to measure accuracy, and *variance* is used to measure diversity.

Melville and Mooney [9] have introduced a new meta-learner DECORATE that uses an existing learner to build an effective diverse committee in a simple, straightforward manner. This is accomplished by adding different randomly constructed examples to the training set when building new committee members. These artificially constructed examples are given category labels that *disagree* with the current decision of the committee, thereby easily and directly increasing diversity when a new classifier is trained on the augmented data and added to the committee.

In this paper, a new method to obtain diverse classifiers is proposed which uses manipulation of dataset structures. The proposed method is very similar to Adaptive Learning Classification [12] in which the learner does not treat all pattern equally. In this new method it is presented to make diversity in base classifiers (of type MultiLayer Perceptron (MLP) or Decision Tree (DT)). The effect of existing or non-existing of boundary instances is evaluated. Firstly, using a simple classification, the error-prone instances and their neighbors are detected. Using these subsets, different datasets of our main dataset are created. Then, several classifiers are trained on these datasets. They may be selected into or eliminated from the ensemble based on a diversity metric. This method that we have named "CDEBMTE: Creation of Diverse Ensemble Based on Manipulation of Training Examples" is described in section 3, accurately. The used decision-tree is J48 that is a Java implementation of C4.5 [13] introduced in [15]. Cross-validated learning curves support the hypothesis that CDEBMTE generally result in greater classification accuracy.

2 CDEBMTE: Creation of Diverse Ensemble Based on Manipulation of Training Examples

In CDEBMTE, using a simple classification, the error-prone samples and their neighbors which are called respectively EPS and NS are first detected. Using EPS and NS and also the original training set, different training sets are created from the original one. Then, several classifiers are trained on these datasets. Based on a diversity metric a subset of them is selected into an ensemble. Finally, a consensus function is employed to aggregate the results of these diverse classifiers.

2.1 Preparing Different Subsets of Training Dataset

To extract different subsets out of training set, we divide the primary Training Set (TS) in K random partition. Let us denote i-th partition by TS_i. Presume that cardinality of TS, $|TS|$, is denoted by N. By training a fixed number (denoted by H) of simple base classifiers on $TS \backslash TS_i$, we reach an ensemble that is expert for TS_i. Please note that these H classifiers are produced by a Bagging mechanism named Random Forest (RF). Breiman [1] proposes a variant of bagging which he calls RF. RF is a general class of ensemble building methods that uses decision tree as its base classifier. It has a parameter α that shows the ratio of sub-sampling. To be labeled a "random forest" an ensemble of decision trees should be built by generating independent identically distributed random vectors and use each vector to grow a decision tree. In this paper RF which is one of the well known versions of bagging classifier [7] is considered as Bagging method. It means when Bagging is used it refers to RF in the paper.

Let us denote the ensemble that is expert for TS_i by E_i. We test each data point in TS_i to find out how the data point has potential to be wrongly classified. We assign a number P_j to j-th data point in TS_i indicating how many classifiers out of H ones in the ensemble E_i, wrongly classify the j-th data point. The bigger P_i, the more error-prone the i-th data point. It is also worthy to mention that P_i is an integer number between 0 to H.

This procedure is repeated K times, choosing a different part for testing, each time. When N=K, the method is called the leave-one-out or U-method [7].

Table 1. Different data combinations and reasons of its usages

Classes of Producing Training Subset	Feature of Resultant Subsets
1. κ* TS	To create the base classifiers
2. κ*TS∪μ*NS	Classification by high complex boundaries with more concentration on crucial points
3. κ*TS∪λ*EPS∪μ*NS	Classification by complex boundaries with more concentration on error prone and crucial points
4. κ*(TS\EPS) ∪μ*NS	Classification by high simple boundaries with more concentration on crucial points
5. κ*TS∪λ*EPS	Classification by high complex boundaries with more concentration on error prone data samples
6. κ*(TS\EPS)	Classification by very simple boundaries

```
TS: Training Set
FinalEnsemble={}
Function ProduceFinalEnsemble(E)
pre_acc=0;
pre_div=0;
 For i=1 to |E|
   tempEnsemble= Ci∪FinalEnsemble;
   div=D(tempEnsemble);
   acc=TEST(tempEnsemble, TS);
   If (acc≥pre_acc)&(div≥pre_div) Then
      FinalEnsemble=tempEnsemble;
      pre_acc=acc;
      pre_div=div;
 End.
Return FinalEnsemble
```

Fig. 1. Pseudo-code of CDEBMTE algorithm

2.2 Creating an Ensemble of Diverse Classifiers

Using more diverse classifiers is one of the most important factors controlling the performance of combinational classifiers. Based on three sets of data, TS, EPS and NS, in this step, a number of classifiers are heuristically retrained. As shown in Table 1, a number of combinations are used to create diversity in the ensemble. Also,

Table 1 expresses the reasons of their usage. It is worthy to be mentioned that κ, λ and μ are three parameters that are to be set. The domains of these parameters are {0.2, 0.4, 0.6, 0.8, 1}. For example, κ**TS* means that a sub-sampling from *TS* that contains a κ percept of *TS*; and κ**TS*∪λ**EPS*∪μ**NS* means that a union of three sub-samplings: (1) κ percept from *TS*, (2) λ percept from *EPS* and (3) μ percept from *NS*.

Retraining of classifiers, according to combinations of Table 1, results the classifiers so that each of them concentrates on special aspect of data. This can result in very good diversity in the ensemble. In other hand, although the accuracy of each classifier may not be significantly better than a simple classifier, it can yield to satisfactory diverse base classifiers.

In this study, six classes of producing a subset are defined. In first class of producing subsets, by sliding parameter κ in all possible values in its domain, we obtain 5 training subsets. Equivalently in second class, by sliding parameters κ and μ in all possible values in their domains, we obtain 5*5 training subsets. In 3rd, 4th, 5th and 6th classes, we respectively reach 125 (5*5*5), 25, 25 and 5 training subsets by sliding their parameters.

2.3 Diversity of an Ensemble

To define a meaningful diversity metric first we define a between-classifier diversity metric over one example. As equation (1) shows, the between-classifier diversity metric over one example is denoted by $d(C_i, C_j, x_k)$.

$$d(C_i, C_j, x_k) = \begin{cases} 0 & C_i(x_k) = C_j(x_k) \\ 1 & C_i(x_k) \neq C_j(x_k) \end{cases}. \tag{1}$$

In the equation (1), C_i stands for *i*-th classifier from the ensemble and $C_i(x_k)$ stands for its output over an exemplary sample x_k. Based on the equation (1), we define a diversity metric for an ensemble according to equation (2).

$$D(E) = \frac{1}{|E|^2 * |TS|} \sum_{x_k \in TS} \sum_{C_i \in E} \sum_{C_j \in E} d(C_i, C_j, x_k). \tag{2}$$

where E stands for an ensemble of classifiers. $D(E)$ is the diversity of ensemble E.

2.4 CDEBMTE

In CDEBMTE, a rough ensemble is first generated and then pruning it into the final ensemble. A classifier is trained on each subset defined on the training data. We train each classifier on each subset defined on the training data, thereby forcing them to differ from the each other. Therefore these classifiers as an ensemble can be considered a diverse one. While enforcing the diversity into the rough initial ensemble, we still want to extract the most diverse subset of classifiers out of the rough initial ensemble. We do this by eliminating a new classifier if adding it to the existing ensemble decreases its accuracy. This process is repeated until we exceed the maximum number of iterations. The pseudo code of the CDEBMTE algorithm is presented in Fig 1.

As you can understand from Fig 1, a greedy algorithm is employed to find the most diverse subset of the classifiers existing in the ensemble. The CDEBMTE algorithm incrementally examines possibility of adding the next classifier to the pool of current ensemble by considering two factors. First it temporarily adds the next classifier to the current ensemble. If both accuracy and diversity of the current ensemble increases over TS, the search algorithm accepts the current ensemble; else the temporarily new added classifier is removed from current ensemble.

3 Experimental Results

To evaluate the performance of CDEBMTE we ran experiments on 18 representative datasets from the UCI repository [11]. The datasets are summarized in Table 2. Note that the datasets vary in the numbers of training examples, classes, numeric and nominal attributes; thus providing a diverse testbed.

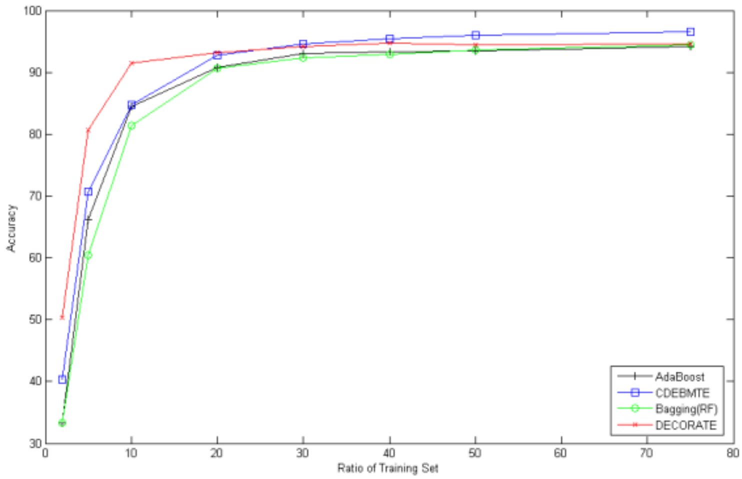
The performance of each learning algorithm is evaluated using 10 complete runs of 10-fold cross-validation (except 3 monk problems). In each 10-fold cross-validation, each dataset is randomly split into 10 equal-size segments and results are averaged over 10 trials. For each trial, one segment is set aside for testing, while the remaining data is available for training.

The size ensemble in each algorithm is considered same with CDEBMTE algorithm. It means after running of CDEBMTE algorithm, the size of final ensemble is extracted and is feed as the ensemble size to other algorithms.

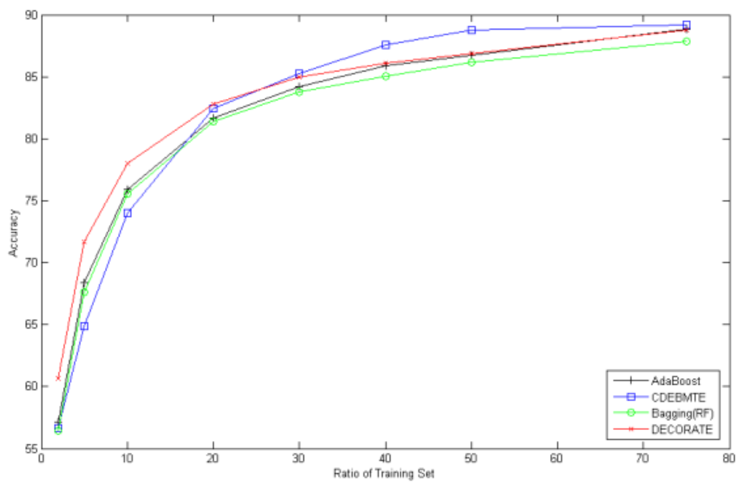
Table 2. Summary of datasets. The training and test sets in all three Monk's problems are predetermined.

Dataset Name	Cases	Classes	Attributes
anneal	898	6	39
audio	226	6	69
autos	205	6	25
breast-w	699	2	9
credit-a	690	2	15
glass	214	6	9
heart-c	303	2	13
hepatitis	155	2	19
colic	368	2	22
iris	150	3	4
labor	57	2	16
lymph	148	4	18
monk	432	2	6
soybean	683	19	35
splice	3190	3	62
wine	178	3	13

Fig 2-a depicts the performance of CDEBMTE comparing to the performances of DECORATE, AdaBoost and Bagging over Iris dataset. In Fig 2-a the accuracy of different ensemble methods in terms of different ratios of TS is presented.



(a)



(b)

Fig. 2. Accuracy of different ensemble methods in terms of different ratios of TS. (a) Accuracy over Iris dataset (b) Averaged accuracy over all 18 datasets.

Fig 2-b depicts the performance of CDEBMTE comparing to the performances of DECORATE, AdaBoost and Bagging averaged over all 18 datasets. In Fig 2-b the averaged accuracy of different ensemble methods in terms of different ratios of TS is presented. As it is depicted, DECORATE outperforms other methods considerably when the size of *TS* is small. It is due to adding artificial data samples to train set. It means while the number of training set is not enough to properly the other methods learn the models of the classes, DECORATE increases the size of training set to learn the models of the classes better.

Table 3. Experimental results

Dataset Name	Bagging (RF)	Boosting (Arc-X4)	CDEBMTE
Wine	95.47	96.12	96.31
Iris	95.87	96.16	96.22
Monk 1	87.22	98.06	98.48
Monk 2	86.16	87.33	87.51
Monk 3	96.27	97.66	97.76

For *TS* sizes above 50 percent, DECORATE falls gradually in comparing with AdaBoost. It is finally slightly placed in rank three after AdaBoost method. While CDEBMTE sliding improves as *TS* size becomes greater. It is finally placed in rank one above all methods.

The detailed achieved results some prominent ensemble methods are presented in Table 3. In obtaining the results of Table 3, the ratio of sampling from train set in Bagging, Boosting and CDEBMTE algorithms is 70%. CDEBMTE outperforms obviously from prominent ensemble methods, Bagging and Boosting, in some of datasets according Table 3. It is again more outstanding especially in artificial dataset 2 where two of classes are composed by two clusters.

4 Conclusions

In this paper, a new method to improve performance of combinational classifier systems, CDEBMTE, is proposed. CDEBMTE is based on increasing the diversity of ensemble. First different datasets are extracted from training dataset. Then it trains a number of classifiers over them. Finally it selects a subset of the trained classifiers.

The proposed ensemble methodology, CDEBMTE, is examined on some datasets and it shows considerable improvements in comparison with some of the most competent ensemble methods.

Paper also shows that emphasizing on crucial data can cause improvement in diversity. Also we showed that usage of different datasets causes to quite diverse classifiers. It also shows that CDEBMTE can be effectively used to achieve higher accuracy and to obtain better class membership probability estimates in comparison with the most powerful methods.

References

1. Breiman, L.: Random forests. Machine Learning 45, 5–32 (2001)
2. Dietterich, T.G.: An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine Learning 40(2), 139–157 (2000)

3. Freund, Y., Schapire, R.E.: A decision–theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55, 119–139 (1997)
4. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: Saitta, L. (ed.) Proceedings of the Thirteenth International Conference on Machine Learning (ICML 1996). Morgan Kaufmann (1996)
5. Hansen, L.K., Salamon, P.: Neural network ensembles. IEEE Transaction on Pattern Analysis and Machine Intelligence 12, 993–1001 (1990)
6. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, New York (2001)
7. Kuncheva, L.I.: Combining Pattern Classifiers, Methods and Algorithms. Wiley, New York (2005)
8. Liu, Y., Yao, X.: Ensemble learning via negative correlation. Neural Networks 12 (1999)
9. Melville, P., Mooney, R.: Constructing Diverse Classifier Ensembles Using Artificial Training Examples. In: Proc. of the IJCAI, vol. I, pp. 505–510 (2003)
10. Melville, P.: Creating Diverse Ensemble Classifiers (2006)
11. Newman, C.B.D.J., Hettich, S., Merz, C.: UCI repository of machine learning databases (1998), http://www.ics.uci.edu/~mlearn/MLSummary.html
12. Qiao, X., Liu, Y.: Adaptive Weighted Learning for Unbalanced Multicategory Classification. Biometrics, 159–168 (2009)
13. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
14. Tumer, K., Ghosh, J.: Error correlation and error reduction in ensemble classifiers. Connection Science 8(3-4), 385–403 (1996)
15. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, San Francisco (1999)
16. Daryabari, M., Minaei-Bidgoli, B., Parvin, H.: Localizing Program Logical Errors Using Extraction of Knowledge from Invariants. In: Pardalos, P.M., Rebennack, S. (eds.) SEA 2011. LNCS, vol. 6630, pp. 124–135. Springer, Heidelberg (2011)
17. Fouladgar, H., Minaei-Bidgoli, B., Parvin, H.: On Possibility of Conditional Invariant Detection. In: König, A., Dengel, A., Hinkelmann, K., Kise, K., Howlett, R.J., Jain, L.C. (eds.) KES 2011, Part II. LNCS, vol. 6882, pp. 214–224. Springer, Heidelberg (2011)
18. Minaei-Bidgoli, B., Parvin, H., Alinejad-Rokny, H., Alizadeh, H., Punch, W.F.: Effects of resampling method and adaptation on clustering ensemble efficacy, Online (2011)
19. Parvin, H., Minaei-Bidgoli, B.: Linkage Learning Based on Local Optima. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 163–172. Springer, Heidelberg (2011)
20. Parvin, H., Helmi, H., Minaei-Bidgoli, B., Alinejad-Rokny, H., Shirgahi, H.: Linkage Learning Based on Differences in Local Optimums of Building Blocks with One Optima. International Journal of the Physical Sciences 6(14), 3419–3425 (2011)
21. Parvin, H., Minaei-Bidgoli, B., Alizadeh, H.: A New Clustering Algorithm with the Convergence Proof. In: König, A., Dengel, A., Hinkelmann, K., Kise, K., Howlett, R.J., Jain, L.C. (eds.) KES 2011, Part I. LNCS, vol. 6881, pp. 21–31. Springer, Heidelberg (2011)
22. Parvin, H., Minaei, B., Alizadeh, H., Beigi, A.: A Novel Classifier Ensemble Method Based on Class Weightening in Huge Dataset. In: Liu, D., Zhang, H., Polycarpou, M., Alippi, C., He, H. (eds.) ISNN 2011, Part II. LNCS, vol. 6676, pp. 144–150. Springer, Heidelberg (2011)
23. Parvin, H., Minaei-Bidgoli, B., Alizadeh, H.: Detection of Cancer Patients Using an Innovative Method for Learning at Imbalanced Datasets. In: Yao, J., Ramanna, S., Wang, G., Suraj, Z. (eds.) RSKT 2011. LNCS, vol. 6954, pp. 376–381. Springer, Heidelberg (2011)

24. Parvin, H., Minaei-Bidgoli, B., Ghaffarian, H.: An Innovative Feature Selection Using Fuzzy Entropy. In: Liu, D. (ed.) ISNN 2011, Part III. LNCS, vol. 6677, pp. 576–585. Springer, Heidelberg (2011)
25. Parvin, H., Minaei, B., Parvin, S.: A Metric to Evaluate a Cluster by Eliminating Effect of Complement Cluster. In: Bach, J., Edelkamp, S. (eds.) KI 2011. LNCS, vol. 7006, pp. 246–254. Springer, Heidelberg (2011)
26. Parvin, H., Minaei-Bidgoli, B., Ghatei, S., Alinejad-Rokny, H.: An Innovative Combination of Particle Swarm Optimization, Learning Automaton and Great Deluge Algorithms for Dynamic Environments. International Journal of the Physical Sciences 6(22), 5121–5127 (2011)
27. Parvin, H., Minaei, B., Karshenas, H., Beigi, A.: A New N-gram Feature Extraction-Selection Method for Malicious Code. In: Dobnikar, A., Lotrič, U., Šter, B. (eds.) ICANNGA 2011, Part II. LNCS, vol. 6594, pp. 98–107. Springer, Heidelberg (2011)
28. Qodmanan, H.R., Nasiri, M., Minaei-Bidgoli, B.: Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence. Expert Systems with Applications 38(1), 288–298 (2011)