World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:3, No:9, 2009

Combining Diverse Neural Classifiers for Complex Problem Solving: An ECOC Approach

R. Ebrahimpour, M. Abbasnezhad Arabi, and H. Babamiri Moghaddam

***Abstract*—**Combining classifiers is a useful method for solving complex problems in machine learning. The ECOC (Error Correcting Output Codes) method has been widely used for designing combining classifiers with an emphasis on the diversity of classifiers. In this paper, in contrast to the standard ECOC approach in which individual classifiers are chosen homogeneously, classifiers are selected according to the complexity of the corresponding binary problem. We use SATIMAGE database (containing 6 classes) for our experiments. The recognition error rate in our proposed method is %10.37 which indicates a considerable improvement in comparison with the conventional ECOC and stack generalization methods.

***Keywords*—**Error correcting output code, combining classifiers, neural networks.

I. Introduction

DIVIDE and conquer is a common approach in solving complex machine learning problems and more specifically, sophisticated classification problems. In this principle, a complicated problem is divided into a number of simple problems each of which is solved by a simple classifier e.g. a binary classifier. The result is then achieved by combining the solutions to each simple problem yielding an increase in the efficiency, recognition rate and the reliability of the system [1].

Obviously, one of the factors that necessitates divide and conquer approach is the existence of numerous classes in a problem resulting in an extensive complexity. The idea is to map the input into another space i.e. feature space where the classes are more likely to be separated. In such a space, classifiers will be able to provide decision boundaries to distinguish the regions for different classes.

When dealing with a combining problem, it is necessary to have at hand a number of independent classifiers as well as a mathematical framework to combine the solutions to these classifiers. The independence criteria for the classifiers are defined so that they yield the same results for correctly classified patterns and different results when a misclassification occurs. Thus, correlation reduction between classifiers seems inevitable. Note that those patterns that all the classifiers fail to classify them correctly are not classifiable using the combining system [2]. In fact, in such situations unclassified region(s) exist. To resolve the unclassified regions, some probabilistic or distance measurement mechanisms might be employed.

Here, we aim to solve a 6-class classification problem by training 31 Multilayer perceptrons neural networks as binary classifiers with different learning parameters and structures. In contrast to the standard ECOC [3,4,5] approach in which individual classifiers are chosen homogeneously, classifiers are selected according to the complexity of the corresponding binary problem. These classifiers are then combined in the test phase to construct the decision functions using the minimum distance method.

The remainder of this paper is organized as follows. In section II, we discuss the most common correlation reduction procedures. In section III, we outline the error correcting output code method. The classifier combining algorithm is described in section IV. The results of the proposed method performed on the SATIMAGE database are reported in Section V which is followed by the conclusions in section VI.

II. Correlation Reduction Techniques

There are a couple of practices to perform correlation reduction. These techniques either modify the structure of the learner or alter the input pattern representation or the corresponding training set visible to each binary classifier.

The most common approaches are listed below.

A. Employing a Different Learning Machine

In this method, the learning procedure is altered either by using different learning algorithms or by using certain algorithms with different complexities and different learning parameters [2].

B. Different Representation of the Patterns

This method utilizes different representations of input

Reza Ebrahimpour is with Department of Electrical Engineering, Shahid Rajaee University, Tehran, Iran (e-mail: rebrahimpour@srttu.edu) and School of Cognitive Sciences, Institute for Studies on Theoretical Physics and Mathematics, Niavaran, Tehran, P.O. Box 19395-5746, Iran (e-mail: ebrahimpour@ipm.ir).

Mehdi Abbasnezhad Arabi is with Department of Electrical Engineering, Shahid Rajaee University, Tehran, Iran (e-mail: m_abasnejad@yahoo.com)

H. Babamiri Moghaddam is with Department of Electrical Engineering, Shariaty University, Tehran, Iran (e-mail: hasina_moghaddam@yahoo.com)

feature set or feature set partitioning to divide the sub-features between classifiers so as to create independent classifiers [6].

C. Partitioning the Training Set

Every learning machine contains a number of parameters whose values are set during the training period. For a given structure and an identical representation of patterns, different training sets will yield different recognition and generalization abilities for the learner. In fact, parameter selection or the so-called model selection will vary from classifier to classifier since they have been exposed to different and independent training patterns. Consequently, partitioning the training patterns help reduce the correlation between binary classifiers [7].

III. Error Correcting Output Code (ECOC)

ECOC is an information theoretic concept that seeks to distinguish among different signals corrupted as a result of noise in a transmission channel. The main idea is based on adding some redundant cases in possible output set which do not match with any of the acceptable labels. If one of these cases appears in the output, the system realizes occurrence of an error.

Error detection procedure in ECOC method can be summarized as follows. Assume that we want to transmit a binary variable x through a transmission channel. The level of an electronic signal may be distorted because of noise. We may represent this signal with one bit, and then compare the received signal with a certain threshold and assign it to either class "a" or class "b". When noise changes the level of signal in such a way that it exceeds the threshold, an assignment error will occur. One should note that although representing the signal in more bits would increase the redundancy of system and lower the efficiency, it can provide us with the ability of error correction.

A. ECOC in Classification

Assume that Z is a $k \times b$ code matrix with binary elements where k is the number of classes and b is the number of binary classifiers. Each row of Z is a code word that is used as a label for one of the classes and each column is a map to convert the multi-class problem to binary sub- problems. In training phase, binary classifiers are constructed over these sub-problems. In the test phase, for a given pattern x, the set of experts provide an output vector $Y=[Y_1, Y_2, ..., Y_b]$. The distance of this vector from the class label i is defined as the following aggregation:

$$L_i = \sum_{j=1}^{b} \left| Z_{ij} - Y_j \right|. \tag{1}$$

In decision making, pattern x is assigned to the class with the minimum distance.

Moreover, increasing the Hamming distance between the rows contributes to better performance of the decision-making process. Likewise, increasing the Hamming distance between columns will affect the local error covariance.

IV. The Procedure of Combining Classifiers

Various approaches have been introduced for classifier combining such as voting, maximum distance rule, minimum distance rule, averaging and Dempster-Shafer. The main purpose of combining classifiers is to minimize the reconstruction error.

In ECOC, several methods have been proposed for reconstruction error minimization including least squares, centroid algorithm, linear combining and the minimum distance technique.

Obviously, the higher the distance between the class labels, the less the error and the output of the classifiers (Y) will judge with less sensitivity toward classifiers' error (i.e. the distance between L_i and L_j). Suppose that the classifiers provide the posterior probability $Y=[Y_1, Y_2, ..., Y_b]^T$ of the main class members. Classes are constructed over the columns of Z. Hence, in the matrix equation form $y = z^T.q$ where k is the number of classes and $q=[q_1, q_2, ..., q_k]$ is the initial probability of each class. Substituting the value of Y_j in L we obtain:

$$L_i = \sum_{j=1}^{b} \left| \sum_{k=1}^{L} q_L Z_{ij} - Z_{ij} \right|. \tag{2}$$

Hence, if we separate the cases with $L= i$ and considering the fact that

$$\sum_{L=1}^{k} q_i = 1$$

We have:

$$\begin{aligned} L_i &= \sum_{j=1}^{b} \left| q_i \;.Z_{ij} - Z_{ij} + \sum_{i \neq L} q_L \;.Z_{ij} \right| \\ &= \sum_{j=1}^{b} \left| q_i.Z_{ij} - (\sum_{i \neq L} q_L + q_i)\;.Z_{ij} + \sum_{i \neq L} q_L.Z_{ij} \right| \\ &= \sum_{j=1}^{b} \left| \sum_{i \neq L} q_L (Z_{ij} - Z_{Lj}) \right| \\ &= \sum_{i \neq L} q_L \;. \sum_{j=1}^{b} \left| Z_{ij} - Z_{iL} \right| \end{aligned} \tag{3}$$

where i is the number of classes and L_i is the distance between vector Y and the class labels. In the test phase, test patterns are assigned to a class with minimum distance i.e. ($MIN(L_i)$).

V. Experimental Results

To inspect the ECOC method in this classification problem, we use 31 multi-layer perceptrons with a single hidden layer as in [3,4,5]. We employ the BP algorithm for training the

MLPs and the initial weights are randomly chosen in the range of [-1, 1].

Our experiment is conducted in two steps. In the first step, the proposed ECOC method is applied. The appropriate number of neurons in the hidden layer for each expert is found using an iterative algorithm. The values are in the range of 10 to 24 and are shown in Fig. 1. We stop training each individual binary classifier as soon as it attains a minimum recognition rate of %90. Note that the difference between the 31 neural networks is in the number of neurons in the hidden layer only.

In the second step, the standard ECOC technique is used. The MLPs in this step have identical structures. These networks are homogeneous with respect to the number of hidden layers as well as the number of epochs. In order to compare the two ECOC methods, we train the neural networks for the standard method with minimum, average and maximum number of 10, 17 and 24 neurons respectively.

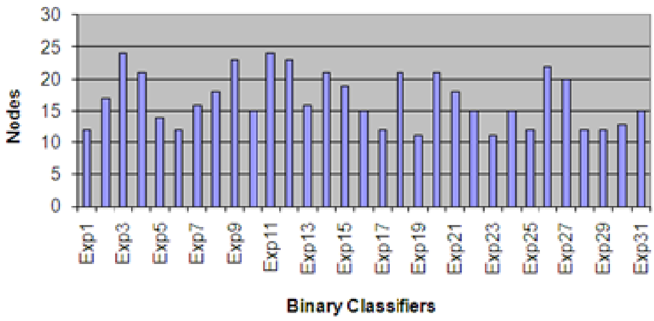
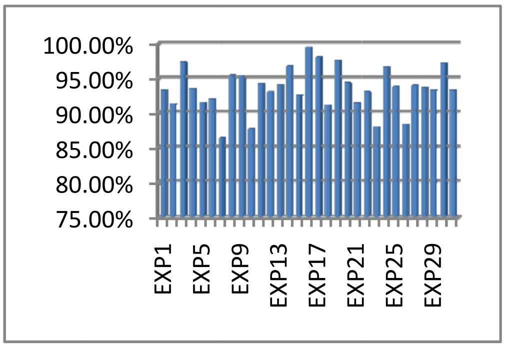


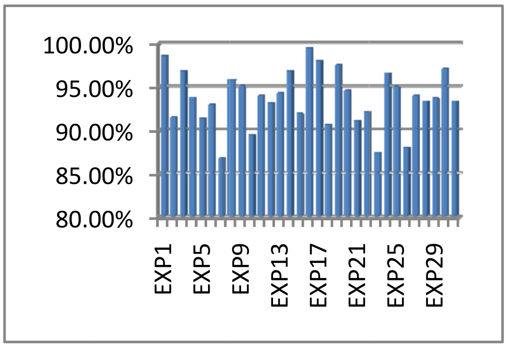
Fig. 1 The number of the hidden layer nodes for the binary classifiers in the suggested method

The problem solving procedure is almost the same in the two methods and is as follows. First, the input patterns are fed into the binary classifiers and an initial recognition of the classes is achieved. Then, the results of the neural networks are combined using the minimum distance method. The recognition rates for the classifiers of the conventional and the proposed ECOC method are shown in Fig. 2-A through 2-D.

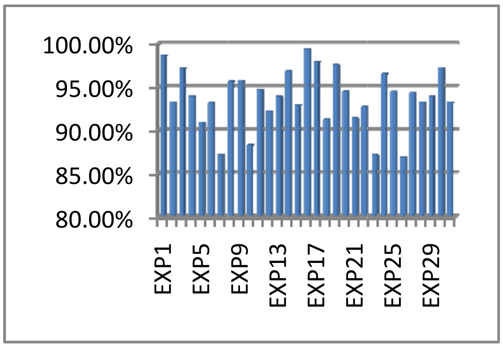
A-Standard ECOC (n=10)



B-Standard ECOC (n=24)



C-Standard ECOC (n=17)



D-Modified ECOC

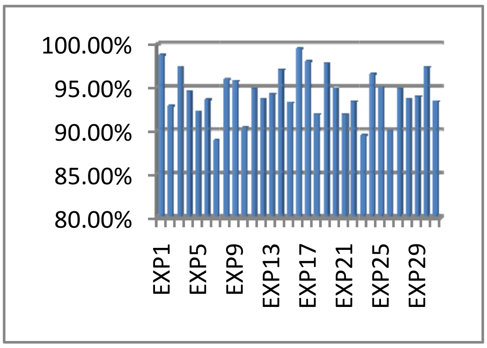


Fig. 2 The recognition rates of the homogeneous experts

The SATIMAGE database contains 6439 patterns with 36 features, 4290 patterns for training and 2145 patterns for testing. If a neural network is used for solving the problem, the output vector dimension will be 4290×6 for training and 2145×6 for the testing phase. A recognition rate of %83 was achieved in this experimentation. Taking advantage of the ECOC method, we convert the 6439×36 input into a 6439×31 matrix. Then each of the columns of this matrix is fed into one of the 31 classifiers to obtain 1×2145 and 1×4290 vectors. Next, the output matrices are compared to the ideal outputs and the recognition rate of each classifier is calculated.

For output combining, the output vector (real output) of each classifier forms a new 31×2145 matrix. This is followed by calculating the absolute difference between the new matrix and the real output from the 31×6 ECOC matrix which is produced by exhaustive code generation:

$$L_{jk} = \sum_{k=1}^{2145} \sum_{j=1}^{6} \sum_{i=1}^{31} \left| real\_out_{ik} - ECOC_{ij} \right| \quad (4)$$

where L_{jk} is a 6×2145 matrix and j represents labels of the class and k denotes the number of the corresponding pattern. We find the minimum element of each column. The row number represents the corresponding class to which each pattern belongs. After assigning the patterns to classes, we compare the outputs with their matching target values and then determine the overall recognition rate.

In Table I, the proposed ECOC is compared to the standard ECOC and a single MLP. We observe that the suggested ECOC demonstrates better average recognition rate. One should note that the number of training epochs was fixed to 500 for all the models.

TABLE I
THE RECOGNITION RATES OF THE DIFFERENT MODELS

Model	**Modified ECOC**	**Standard ECOC**	**Standard ECOC**	**Standard ECOC**	**MLP**
Number of Hidden Layer Neurons	**10-24**	**10**	**17**	**24**	**15**
Ave. Recognition Rate (%)	**89.60**	**88.34**	**88.71**	**88.90**	**83.00**

VI. CONCLUSION

Our survey showed that taking advantage of heterogeneous classifier combining increases both the efficiency and the recognition rate. Another benefit of the suggested model, as observed in Table II, is that the recognition error is more uniformly distributed over the classes as compared to the standard technique. In addition, the proposed ECOC method outperforms the other methods such as stack generalization with recognition rate of %88.41[8], homogeneous ECOC with recognition rate of %87.45[9] and MLP. It is also observed that classifier combining is advantageous only when the error rate of the classifiers is trivial. Moreover, each individual classifier should demonstrate different error patterns as compared to the other classifiers. Employing diverse types of binary classifiers in the proposed method, achieved from neural networks with different initial weights and different hidden layer structures, is believed to decrease the recognition error rate.

TABLE II
CONFUSION MATRICES FOR THE DIFFERENT MODELS

A-Standard ECOC (n=10)

	class 1	**class 2**	**class 3**	**class 4**	**class 5**	**class 6**
Image 1	**488**	**0**	**11**	**0**	**2**	**0**
Image 2	**0**	**223**	**0**	**1**	**5**	**0**
Image 3	**5**	**0**	**445**	**11**	**0**	**3**
Image 4	**2**	**2**	**51**	**129**	**3**	**23**
Image 5	**18**	**4**	**1**	**1**	**187**	**27**
Image 6	**0**	**0**	**10**	**57**	**16**	**420**

B-Standard ECOC (n=17)

	class 1	**class 2**	**class 3**	**class 4**	**class 5**	**class 6**
Image 1	**486**	**0**	**13**	**0**	**2**	**0**
Image 2	**0**	**223**	**0**	**1**	**5**	**0**
Image 3	**2**	**0**	**449**	**11**	**1**	**1**
Image 4	**1**	**2**	**49**	**132**	**3**	**23**
Image 5	**15**	**4**	**0**	**1**	**195**	**23**
Image 6	**0**	**0**	**10**	**62**	**15**	**419**

C-Standard ECOC (n=24)

	class 1	**class 2**	**class 3**	**class 4**	**class 5**	**class 6**
Image 1	**487**	**0**	**12**	**0**	**2**	**0**
Image 2	**0**	**223**	**0**	**1**	**5**	**0**
Image 3	**2**	**0**	**448**	**11**	**1**	**2**
Image 4	**1**	**2**	**50**	**130**	**3**	**24**
Image 5	**15**	**4**	**0**	**2**	**192**	**25**
Image 6	**0**	**0**	**10**	**60**	**14**	**419**

D-Modified ECOC

	class 1	**class 2**	**class 3**	**class 4**	**class 5**	**class 6**
Image 1	**488**	**0**	**10**	**0**	**3**	**0**
Image 2	**0**	**224**	**0**	**1**	**4**	**0**
Image 3	**2**	**0**	**448**	**12**	**1**	**1**
Image 4	**1**	**2**	**44**	**138**	**3**	**22**
Image 5	**15**	**4**	**0**	**1**	**204**	**14**
Image 6	**0**	**0**	**13**	**55**	**13**	**422**

REFERENCES

[1] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall, Englewood Cliffs, NJ, 1999.

[2] R.Ghaderi, Arranging simple neural network to solve complex classification problem, PhD thesis, Surrey University, 2000.

[3] Dietterich, T.G., G. Bakiri, “Error-correcting output codes: A general method for improving multiclass inductive learning programs”, Proc. of the 9th National Conference on Artificial Intelligence, AAAI-91. AAAI Press, pp. 572–577, 1991.

[4] T. Windeatt, R. Ghaderi, “Binary labeling and decision level fusion”, Inform. Fusion 2, 103–112, 2001.

[5] T. Windeatt, R. Ghaderi, “Coding and decoding strategies for multi-class learning problems”, Inform. Fusion 4, 11–21, 2003.

[6] J. Kittler, S. A. Hojjatoleslami, T. Windeatt." Weighting Factors in Multiple Expert Fusion", In Proc. Of British Machine vision Conference BMVC 97, pp. 42-50, Essex University, Essex U.K, 1997.

[7] T.Windeatt , R. Ghaderi, "Dynamic Weighting Factors for Decision Combining", Proc. of IEE Int. Conf. On Data Fusion, Great Malvern, UK, pp. 123-130, October 1998.

[8] N. Hatami, R. Ebrahimpour, " Combining Multiple Classifiers: Diversify with Boosting and Combining by Stacking" International Journal of Computer Science and Network Security (IJCSNS), Vol.7, No.1, pp. 127-131, 2007.

[9] N. Hatami, R. Ebrahimpour, R. Ghaderi, " Error Correcting Output Codes with Weighted Minimum Distance Decoding by Genetic Algorithm" Proc. of the 3th Information and Knowledge Technology (IKT07), Ferdowsi University of Mashad, Mashad, Iran, Nov. 27-29, 2007.

[10] J.Kitter , R.Ghaderi ,T.Windeatt , J.mtas “Face verification via Error correcting output codes “ International Journal of Image and Vision Computing Vol.21,pp.1163-1169,2003.

[11] T. Windeatt and R. Ghaderi. Binary codes for multi-class decision combining. Volume 4051, pages 23–34, Florida,USA, April 2000. 14th Annual International Conferenceof Society of Photo-Optical Instrumentation Engineers (SPIE).

[12] R. Ghaderi, T. Windeatt, Least squares and estimation measures via error correcting output code, in: 2nd International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, Springer-Verlag, July 2001, pp. 148–157.

[13] T.G. Dietterich, Ensemble methods in machine learning, in: J. Kittler, F. Roli (Eds.), Multiple Classifier Systems, MCS2000,Cagliari, Italy, Springer Lecture Notes in Computer Science, 2000,pp. 1–15.

[14] T.G Dietterich, G. Bakiri, Error-correcting output codes: A general method for improving multiclass inductive learning programs, in: Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), AAAI Press, 1991, pp. 572–577.

[15] T.G. Dietterich, G. Bakiri, Solving multi-class learning problems via error-correcting output codes, Journal of Artificial Intelligence Research 2 (1995) 263–286.

[16] E.B. Kong, T.G. Diettrich, Error-correcting output coding corrects bias and variance, in: M. Kaufmann (Ed.), 12th Interna-20 T. Windeatt, R. Ghaderi / Information Fusion 4 (2003) 11–21tional Conference of Machine Learning, San Fransisco, 1995, pp.313–321.

Mehdi Abbasnezhad Arabi received the BS degree in electronics engineering from Shahid Rajaee University in 2007. His research interests are neural networks, pattern recognition, multiple classifier systems, computational vision andmodels of human vision.

Reza Ebrahimpour was born in Mahallat, Iran, in July 1977. He received the BS degree in electronics engineering from Mazandaran University , Mazandaran, Iran and the MS degree in biomedical engineering from Tarbiat Modarres University,Tehran, Iran, in 1999 and 2001,respectively. He received his PhD degree in July 2007 from the School of Cognitive Science,Institute for Studies on Theoretical Physics andMathematics, where he worked on view-independent face recognition with Mixture of Experts. His research interests include human and machine vision, neural networks, pattern recognition.