Classification of Biomedical Data Using Enhanced Classifier Fusion Model

[1]Pradeep Kumar Mallick, [2]Bibhu Prasad Mohanty, [3]Asish Patro, [4]Subhashree Rout
[1]Department of Computer Science and Engineering,
Balasore College of Engineering and Technology, Odisha
[2]Dept. of Information & communication Technology, FM university, Balasore
[3,4]Department of Computer Science and Application, Utkal University, Odisha.

***Abstract*—** Classification consists of assigning a class label to a set of unclassified cases; To achieve more accuracy in every level of the input data classifier fusion is the solution. Selection of a suitable classifier in classifier fusion is a tedious task. In the proposed model, the output of the three classifiers is fed to the dynamic classifier fusion technique. This model will use each classifier for every individual data. We have used principal component analysis (PCA) to deal with issues of high dimensionality in biomedical classification. Three types of classification techniques on microarray data like Multi Layer Perceptron (MLP), FLANN and PSO- FLANN have been implemented and compared; it has been observed that MLP is showing better result. We have also proposed a model for classifier fusion, where the model will choose the relevant classifiers according to the different region of datasets.

Keywords- Principal Component Analysis; Classification; Classifier fusion; FLANN; PSO-FLANN; MLP

I. INTRODUCTION

Classification is the process of assigning unknown input patterns of data to some known classes based on their properties [1-2]. For a long time many research area in designing classifiers have focused improving efficiency, accuracy and reliability of classifier for a wide range of applications. Fusing output of more classifier is an alternative method to build more reliable classifier [3, 7]. It is well known that in many situations combining output of several classifier leads to improved classification result. This occurs because each classifier produces error on different area of input space. In other words, subset of input space that each classifier labels correctly will differ from one classification to another. This implies that by using information from more than one classifier, it is probable that the better overall accuracy can be obtained for a given problem. On the other hand, instead of picking up just one classifier a better approach would be to use more than one classifier while averaging there output. The new classifier might not be better than the single best classifier but it will distinguish or eliminate the risk of picking an inadequate single classifier. For any pattern classification, increase in data size, number of classes, and dimension of feature space and inter class separability effect the performance of any classifier. A single classifier is generally unable to handle wide variability and scalability of data. In many problem domains most modern technique of pattern classification uses a combination of classifier and fused decision provided by the some often using any selected set of appropriate features for the task.

Combining classifier is a thrust research area based on both statistical pattern recognition and machine learning. It is also known as committee of learners, mixtures of experts, classifiers ensemble, multiple classifier system, consensus theory etc. By having a number of different classifier it is wise to use them in a combination in the hope of increasing the overall accuracy and efficiency.

The layout of this paper is as follows; the related work on classifier fusion is given in section II. Section III deals with preliminary concepts related to this paper. Our proposed model is described in section IV. The experimental evaluation is given in section V. Finally, section VI deals with the conclusion and future work.

II. RELATED WORK

Zhenyu Chen et al. [1] proposed a multiple kernel SVM based data mining system. Multiple tasks, including feature selection, data fusion, class prediction, decision rule extraction, associated rule extraction and sub class discover, are incorporated in an integrated framework. All-AML leukemia data set is used to demonstrate the performance of the system.

Esma Kilic et al. [2] investigate two kinds of classifier system which are capable of estimating how much to weight each base classifier dynamically, during the calculation of the overall output for a given test data instance: 1. In "referee: based system", a referee is associated with each classifier which learns the area of expertise of its associated classifier and weights it accordingly. 2. Each referee in referee base system learns a two class problem where as a getting system learns an *L*-class problem assigning the input to one of *L* base classifiers. The study shows that, by using well trained selection unit, we can get as high accuracy as using all the base classifiers with drastic decrease in the number of base classifiers used, and improve accuracy.

Nicolas Garcia-Pedrajas et al. [3] The author suggest that there are two different task in the area: one, training and construction of ensemble of classifiers, with each one being able to solve a multi class problem; the other task is the fusion of binary classifiers, with each one solving a different two-class problem to construct a multi-class classifier. The paper presents a study of the different class binarization methods for the various standards multi class classification problems.

Hui-Min Feng et al. [4] introduced a fuse model, which compare comprehensively four fuzzy integrals in multiple classifier fusions and hope to give the foundation for selecting choquet integral. According the theoretical and experimental analysis the paper gives the conclusion that choquet is the best suitable for classifier fusion.

Jiangtao huang et al. [5] proposed a new multiple classifier fusion method integrated classifier selection and classifier combination. This paper based on interval-valued fuzzy permutation. Firstly, normalize all classifier posterior probabilities using the priory knowledge of corresponding classifier recognition rate. Secondly, convert decision matrix for multiple classifier system into interval-valued fuzzy decision matrix. Thirdly, determine the grade of possibility of each class for input sample for multiple classifier system. Finally, selects the best classifier in current pattern recognition task using interval valued fuzzy permutation. The experiments have shown that the new multiple classifier fusion approach using interval-valued fuzzy permutation can prov$_{id}$e much better accuracy compare to independent classifier and some other fusion methods.

Hazem M. Ei-Bakry [6] proposed an efficient algorithm for pattern detection using combine classifier and data fusion. In this paper efficient neural network for face detection are presented. Such classifier are designed based on cross co-relation in the frequency domain between the input matrix and the input weight of neural network, this approach is developed to reduce the computation steps required by these ENN's for the searching process.

Mangai UG, et al. [7] has done a survey of decision fusion and feature training strategies for pattern classification. Most of this technique use the databases from the UCI repository, vistexture, speech and medical image for exhibiting there performance. The author also proposed a framework which uses decision and feature fusion for a better classification result. The result is presented using three benchmark dataset selected from the UCI repository.

Likuncheva et al. [8] proposed a simple rule for adapting the class combiner to the application. C decision templates are estimated with the same training set that is used for the set of classifier. These templates are then matched to the decision profile of new incoming objects by some similarity measure. The author compared eleven version of the proposed model with fourteen other techniques for classifier fusion on the Sitimage Phoneme data set from the database ELENA.

Matteore et al. [9] evaluated the performances of three basic ensembles to integrate six different sources of high dimension bio molecular data. They also studied the performances resulting from the application simple greedy classifier selection scheme.

Albert H.R.Ko et al. [10] proposed a pair-wise fusion matrix transformation, which produces reliable probabilities for the use of classifier combination and can be amalgamaped with most existent fusion function for combining classifier. The PFM pair wise fusion requires crisp class label outputs from classifiers, and is suitable for high-class problem, or problems with training samples. The experimental results suggest that the performance of a PFM can be a notch above that of simple majority voting rule, and a PFM can work on problem where a behavior-knowledge space might not be applicable.

III. Preliminaries

A. *Principal component analysis* (PCA): is a mathematical procedure that uses an orthogonal transformation to convert a set of observation [14] of possibly co related variables into a set of values uncorrelated variables called principal components.

B. *Functional Link Artificial Neural Network* (FLANN): is a mathematical model or computational model that is inspired by structural and/or functional aspects of biological neural networks [12]. It consist of an inter connected group of artificial neurons and it processes information using a model.

C. *Particle Swarm Optimization* (PSO): is a stochastic based search algorithm widely used to find the optimum solution introduced by Kennedy and Eberthart [11] in 1995. PSO has used in this paper to update the weight of a FLANN model. PSO as an optimization tool provides a population-based search procedure in which individuals called particles change their position (state) with time.

The velocity V_{id} and X_{id} position of the i^{th} particle are updated as follows:

$$V_{id}=V_{id}+c_1*rand1_{id}*(pbest_{id}-X_{id})+c_2*rand2_{id}*(gbest_{id}-X_{id}) \quad (1)$$

$$X_{id} = X_{id} + V_{id} \quad (2)$$

Where, X_i is the position and V_i is the velocity of the particle. *pbest* is the best previous position yielding the best fitness value for the i^{th} particle and *gbest* is the best position discovered by the whole population. c_1 and c_2 and are the acceleration constants reflecting the weighting of stochastic acceleration terms that pull each particle toward *pbest* and *gbest* positions respectively. $rand1_{id}$ and $rand2_{id}$ are two random numbers in range of (0,1).

D. *PSO-FLANN***:** in our work the FLANN model is trained with biologically inspired soft computing [13] technique i.e. PSO. Weight of the model has been updated by the PSO.Each of the input vector passed through the model, error has calculated and the parameters have been updated by PSO.

E. *Multi Layer Perceptron* **(**MLP): MLP is a network of simple neurons called perceptrons [15]. The basic concept of a single perceptron was introduced by Rosenblatt in 1958. The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output through some nonlinear activation function. Mathematically this can be written as

$$y = \Psi\left(\sum_{i=1}^{n} w_i x_i + b\right) = \Psi(w^T x + b) \quad (3)$$

Where, w denotes the vector of weights, x is the vector of inputs, b is the bias and Ψ is the activation function.

F. *Classification*: is the process of assigning unknown input pattern of data to some known classes based on their properties [12]. Data classification is two step process, in training phase a classifier is build based upon predetermined set of data classes or concepts, in test phase the data are used to estimate the accuracy of the classification rules.

IV. PROPOSED MODEL

Feature extraction aims to reduce the computational cost of feature measurement, increase the classifier efficiency and allow greater classification accuracy based on the process of deriving new features from the original features. In the feature extraction the whole feature space is projected into another dimensional space for a better analysis of the features. The dimension of a data set can be reduced by principal components analysis, linear discriminant analysis, factor analysis, independent component analysis, etc.

In the model (figure 1) dimension of microarray data (MAD) has been reduced using PCA, the data has classified using three classifiers, FLANN, MLP and PSO-FLANN, and the result of the three classification techniques has fused together using dynamic classifier fusion (DCF) to have the better accuracy of the model.

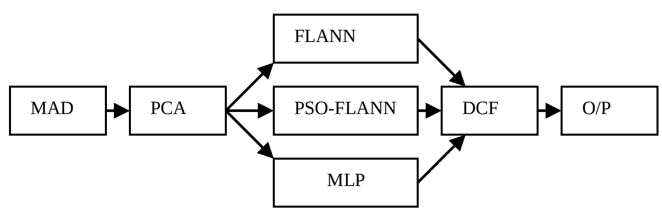

Figure 1. Proposed model

V. EXPERIMENTAL EVALUATION

The Breast Cancer data set have been used for experimental evaluation. The dimension of this data set is 98 X 1,213. After implementing PCA the original data set has reduced to 98 X 97. The breast cancer data set contains three class level 1-11 features used for class level 1, 12-62 for class level 2 and rest are belong to class level 3. The input vector for the classification contains 98 genes with 97 conditions using PCA and also the classification is observed on the original data set (without PCA). The weight of each classifier initially chosen by using random function. The mean square error graph (MSE) is used as the cost function of the classification techniques. The learning of weights has been updated by Least Mean Square (LMS) in FLANN; PSO is used in PSO-FLANN and back propagation algorithm used in MLP. The result of three classifiers is fused together using DCF. DCF chooses the classifier according to the best class regions. In MLP classifier with hidden layer 4, $\eta = 0.4$, $\alpha = 0.7$ (both are the learning parameters in MLP) and umber of iteration 1000 gives confusion matrix table 4. In FLANN classifier $\mu = 0.0009$ is taken and with 1000 iterations and the confusion matrix is shown in table 2. In PSO-FLANN with p_{num}=10, $\mu = 0.001$, $c1$=0.4 and $c2$= 0.3 with 1000 iteration gives confusion matrix as shown in table 3. DCF as shown in table 5, is achieved by fusing the table 2, 3, 4. From the table 5, it is noted that MLP gives better result as compared to FLANN and PSO-FLANN. It has been observed that using DCF percentage of accuracy increased to 78.787%.The classification result of the data set is also obtained without using PCA shown in table 1

Table 1. Classification results obtained from testing using FLANN, PSO-FLANN and MLP for Breast Cancer data set without using PCA.

Classification Techniques	FLANN	PSO-FLANN	MLP
Accuracy in %	33.33	41.49	38.23

Table 2. Classification results obtained from testing using FLANN and PCA for Breast Cancer data set

Classified observation	Cluster $_1$	Cluster $_2$	Cluster $_3$
Class 1	4	1	1
Class 2	4	10	3
Class 3	3	0	7
Cumulative	11	11	11

Table 3. Classification result obtained from testing using PSO-FLANN and PCA for Breast Cancer data set.

Classified observation	Cluster $_1$	Cluster $_2$	Cluster $_3$
Class 1	3	1	0
Class 2	4	9	0
Class 3	3	1	11
Cumulative	11	11	11

Table 4. Classification result obtained from testing using MLP and PCA for Breast Cancer Data set.

Classified observation	Cluster $_1$	Cluster $_2$	Cluster $_3$
Class 1	5	0	0
Class 2	3	10	1
Class 3	3	1	10
Cumulative	11	11	11

Table 5. DCF Result obtained from Fusion of FLANN, MLP and PSO-FLANN

Classified observation	Cluster 1	Cluster 2	Cluster 3
Class 1	5	0	0
Class 2	3	10	0
Class 3	3	1	11
Cumulative	11	11	11

Table 6. Comparison of classification result

Name of data set	Percentage of accuracy			
	FLANN	PSO-FLANN	MLP	DCF
Breast Cancer	63.6364	69.696	75.757	78.787

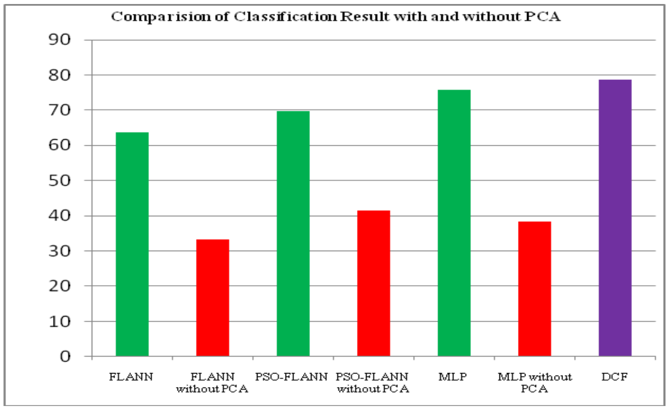

Figure 2. Comparison of classification results with dynamic classifier fusion (DCF) for Breast Cancer data set.

VI. Conclusion And Future Work

Ensemble learning or classifier fusion is an ever growing field, with a wide scope of inter disciplinary research over the fields of computer science, mathematics, statistics and machine learning. In the future, one can expect rich concepts from widely varying areas such as information theory, optimization theory, rough fuzzy sets, soft computing, evolutionary computation etc., to contribute and enrich this problem domain in the field of pattern recognition. This paper processes an efficient dynamic classifier fusion. The input features are extracted using PCA technique. The classifiers are designed using simple LMS, Back propagation and PSO algorithms. MLP gives better performance compared to FLANN and PSO-FLANN, where as the dynamic classifier fusion enhance the accuracy of both the classifier. In future we can take diversify classifier to achieve more accuracy.

References

[1] Z. Chen, J. Li, L.W. Wei, W. Xu, Y. Shi, “Multiple kernel SVM based multiple-task oriented data mining system for gene expression data analysis”, Expert Systems with Applications, Vol-38, pp.12151-12159, 2011.

[2] E. Kilic, E. Alpaydin, “Learning the areas of expertise of classifiers in an ensemble”, Proc. Computer Science,Vol-3, pp.74-82, 2011.

[3] N. Garcia, B. Pedrajas, B. Ortiz, “An empirical study of binary classifier fusion methods for multi class classification” , Information fusion ,Vol-12 , pp.111-130, 2011.

[4] H. M. Feng, X. F. Li, J.F. Chen, “A comparative study of four fuzzy integrals for classifier fusion”, Proc. of Int. Conf. on Machine learning and Cybernetics, pp.332-338, 2010.

[5] J. Huange, M. Wang, B. Gu, Z. Chen, “Multiple classifier combination based on interval-valued fuzzy permutation”, Journal of Computational Information Systems, Vol-6, pp.1759-1768, 2010.

[6] M. Hazem, E. Bakry, “An efficient algorithm for pattern detection using combined classifiers and data fusion”, Information Fusion, Vol-11, pp.133-148, 2010.

[7] U.G. Mangai, S. Samanta, S. Das, P.R. Chowdhury, “A survey of decision fusion and feature training strategies for pattern classification”, IETE Technical Review , Vol-27, pp.293-307, 2010.

[8] L. I. Kuncheva, J.C. Bezbek, R.P.W Duin, “Decision template for multiple classifier fusion: an experimental comparison”, Pattern Recognition, Vol-34, pp.299-314, 2010.

[9] M. Re, G. Valentini, “An ensemble based data fusion for gene function prediction, Multiple Classifier Systems”, Springer, pp.448-457, 2009.

[10] H.R. Albert, R. Ko, R. Sabourin, A. S. Britto, L. Oliveira, “Pair wise fusion matrix for combining classifiers”, Pattern Recognition, Vol-40, pp. 2198-2210, 2007.

[11] J. Kennedy, R. Eberhart, “Particle Swarm Optimization”, Proc. of IEEE Int. Conf. on Neural Networks, pp.1942-1948, 1995.

[12] A.M. Sarhan, “Cancer classification based on micro array gene expression data using DCT and ANN”, Proc. Of Int. Conf. on General of Theoretical and Applied Information Technology, pp. 208-216, 2009.

[13] R. Kumar, M.S.B. Saithij, S. Vaddadi, S.V.K.K. Anoop, “An intelligent functional link artificial neural network for channel equalization”, Proc. of Int. Conf. on Signal Processing Robotics and Automation, pp. 240-245 2009.

[14] E.Peterson, “Partitioning large –sample microarray –based gene expression profile using principal component analysis”, Computer Methods Programming in Biomedicine, pp107-109 2003.

[15] W.Chen, S.Chen, C.Lin, “A Speech Recognition Method Based on The Sequential Multi-layer Perceptrons”, Neural Networks, Vol-9, pp.655-699, 1996