

A Novel Strategy for Designing Efficient Multiple Classifier

Rohit Singh¹, Sandeep Samal², and Tapobrata Lahiri^{3,*}

¹Wipro Technologies, K-312, 5th Block, Koramangala, Bangalore – 560095, India
rohit.singh@wipro.com

²Tata Consultancy Services, Bangalore
sandeep.sam@tcs.com

³Indian Institute of Information Technology, Allahabad – 211012, India
tlahiri@iiita.ac.in

Abstract. In this paper we have shown that systematic incorporation of decision from various classifiers following a simple decision decomposition rule, gives better decision in comparison to the existing multiple classifier systems. In our method each classifier were graded according to their effectiveness of providing more accurate results. This approach first utilizes the best classifier. If this classifier classifies the test sample into more than one class or fails to classify the test data then the feature next to the best is summoned to finish up the remaining part of the classification. The continuation of this process, along with the judicious selection of classifiers, yields better efficiency in identifying a single class for the test data. The results obtained after the experiments on a set of fingerprint images shows the effectiveness of our proposed classifier.

1 Introduction

Personal Identification systems based on fingerprints or facial images, diagnosis of diseases by analyzing the histopathological images, etc. are some applications where accuracy cannot be compromised with, as it may be a case of identifying an authorized person for access to critical or highly restrictive places, or it might be the case of saving the life of a patient through proper diagnosis. More often it is seen that a single classifier struggles to give a high accuracy and reliability level that some critical applications demands. As a result of this, a multiple classifier can be a viable solution for the accuracy and reliability constraints. Work has been going in this field from last decade. From the point of view of analysis, the classification scenarios can be of two types. In the first scenario, all the classifiers use the *same representation* of the input pattern. Here each classifier produces an estimate of the same aposteriori class probability. In the second scenario each classifier uses its *own representation* of the input pattern. They can be either sequential or pipelined [1], [7], or hierarchical [8], [9]. Other studies done in the gradual reduction of the set of possible classes are shown in [3], [4], [6]. The combination of ensembles of neural networks (based on

* Corresponding author.

different initializations) has been studied in the neural network literature [10], [11]. In another approach, a theoretical framework based on Bayesian decision rule has also been used [2]. But this approach has serious limitation. The assumptions made are unrealistic for most applications [13]. The first assumption is that joint probability distributions of measurements extracted by classifier are conditionally statistically independent. The second assumption is that posterior class probability does not deviate much from a priory probability.

The approach proposed by us, takes the result of the best classifier at any instant. The classifiers considered here use their *own representation* of the input pattern. This approach makes use of predicted class to classify a test image. The basic assumption is that the decision can be decomposed into multi-classifier space that can be thought as analogous to the decomposition of any general type of data into multidimensional space. Classifier that accurately classifies the maximum number of test data is chosen as the topmost level classifier regarding their usability. In this way the classifiers are organized into levels. The test data to be classified goes through these levels from top to bottom.

2 Underlying Principle – Decision Decomposition

The methodology proposed is based on ensemble of diverse classifiers working with fractal, ridge and wavelet features extracted from a fingerprint image database. For better comprehension of the underlying principle let us draw an analogy between data and decision.

Decomposition of any data into more than one dimension provides more detail and discriminatory information about this data. Let us take the help of one example of 2 dimensional Gel Electrophoresis (2D Gel) to isolate different proteins according to two-dimensional separation method. In figure 1(a) the separation of proteins has been shown as different separated bands according to their charge increasing from top to down. If we assume that each band represents only one protein, it may be a mistake. This is because two proteins of different masses may have same charge. Further horizontal separation of the proteins from each band according to their mass

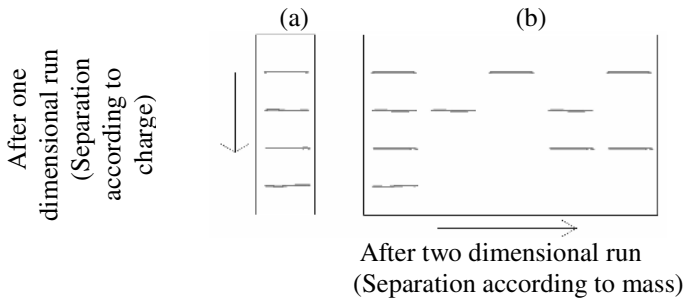


Fig. 1. (a) shows formation of protein bands on gel matrix in an ascending order of their charge and (b) shows further separation of proteins from their bands formed in (a) through their separation in an horizontally ascending order of mass

increasing from left to right, shows that this assumption holds true for the fourth band only. Other bands are further decomposed to give the idea of, more than one protein within each band as shown in figure 1(a).

The above figure gives the idea that more is the attributes (or dimensions) better is the discrimination among classes. Drawing analogy from this experiment we can say that one classifier may initially give some cluster or assembly of classes that can be further separated with the use of other classifier. This approach forms the basis of our work.

Suppose that there are N different classes

$$\{C_i\}_{i=1}^N \quad (1)$$

each of which is represented by a set of data

$$\{d_j\}_{j=1}^M \quad (2)$$

Also suppose that we have a test data d_t whose class is unknown. The proposed algorithm can be generalized as follows:

Step1: First we design P number of classifiers. In this design different classifiers are characterized by different type of logical features extracted from the same data (say image) while their classification rule (distance metric, nature of class boundary) is kept as same.

Step2: Next the efficiency of each classifier is tested. The classifiers' levels are fixed according to the descending order of their efficiency value.

Step3: We start our hunt for appropriate class against the test data using Classifier of level 1, say, CL1. If the test data falls within the overlapping boundary of more than one classes, go to step 4, else if, it falls within a single non-overlapping class boundary, it is assigned as the appropriate class against the test data.

Step4: If the number of overlapping classes defining the test data are

$$\{C_k\}_{k=1}^K \quad (3)$$

we take the help of Classifier of level 2, say, CL2 and focus our attention on K number of overlapping classes only. If, with CL2 test data falls within the overlapping boundary of more than one classes again (say L number of classes where $L \leq K$) repeat step 4 with next level of classifier and so on.

In our methodology we have applied the above Multiple Classifier System by Decision Decomposition (MCSDD). However, as this algorithm is very strict regarding rejection of wrongly classified data and also because of the problem of comparison of this rule with the existing Multiple Classifier Systems (MCS) we have also applied a modified and somewhat flexible rule to compare it with existing MCS.

3 Implementation

3.1 Database and Software Used

The experiment was carried out in Matlab 7.0 environment with 168 fingerprint images, i.e. 8 images taken from each of the 21 persons (i.e., classes). The images

were downloaded from Biometric System Lab., University of Bologna – ITALY (<http://www.csr.unibo.it/research/biolab/>). Out of 8 images per classes, 6 were used as training data while the rest 2 were kept for testing purpose.

3.2 Designing Classifiers

We designed three classifiers based on three different input feature sets that are multi-fractal parameter, third level decomposed wavelet coefficients using ‘haar’ wavelet and ridge features.

3.3 Extraction of Multi-fractal Parameter

The multi-fractal parameter extraction from different intensity plane was carried out in a number of steps. Firstly the RGB image is converted to gray scale one. From the gray-scale image (say, I), n numbers of binary images $\{B_i\}_{i=1}^n$ are obtained by splitting the intensity level of the image at $(n + 1)$ equally spaced different intensity intervals applying the following rule.

$$B_i = 1 \text{ if } g_i \leq I \leq g_{i+1}. \quad (4)$$

where g_i is the i -th gray level. Otherwise,

$$B_i = 0. \quad (5)$$

Thus the pixels of the i -th binary image B_i having gray value 1 are marked as occupied for the corresponding intensity interval. The well-known Box-counting algorithm [6] was then applied to find the fractal dimension D_i for B_i . In practice we have chosen the value of n equals to 8 and the uppermost and lowermost gray levels g_U and g_L respectively as:

$$g_U = 1.4 \times g_m \text{ and } g_L = 0.6 \times g_m. \quad (6)$$

where g_m is the mean gray value of I .

3.4 Extraction of Wavelet Coefficients

The wavelet coefficients are obtained from the converted gray image. Two dimensional discrete wavelet transformations is applied using *dwt2* function of MATLAB with ‘haar’ wavelet. It gives the output as approximation and detail coefficients. The approximation coefficient of third level has been considered as our second feature.

3.5 Extraction of Ridge Feature

For extraction of ridge features following steps were executed [12]:

Step1: We determine a reference point and region of interest for the fingerprint image.

Step2: Tessellate the region of interest around the reference point.

Step3: The region of interest is then filtered in eight different directions using a bank of Gabor filters.

Step4: We then compute the average absolute deviation from the mean (AAD) of gray value in individual sectors in filtered images, which defines the feature vector.

4 Classification Rule

After the feature parameters are obtained they were assigned to three classifiers as mentioned in the previous section. A clustering algorithm is used for each level of classifiers starting from level 1 up to the level that gives single class decision subjected to some modifications and flexibility as discussed in the latter section. The general steps are as follows:

Step1: Choose CL1

Step2: Find class boundary of each class. For this, first find the average feature value of say i-th class and keep it as class center, C_i of that class. Also find the distance of the maximally distant feature point from C_i , r_i and keep it as the class radius assuming each class as sphere.

Step3: Before presenting a test data to the classifier, extract its feature parameters and consider it as the feature point, T.

Step4: Find the Euclidian distances between T and all the class centers, $\{d_i\}_{i=1}^{21}$ for 21 classes.

Step5: Count the number of classes J for which the following expression holds true

$$d_i \leq r_i$$

If
 the classifier level is not the end level (i.e., the 3rd level in our case),
 Then,
 if J = zero or $J \geq 2$, chose the next level classifier.
 else if J = 1, stop classification and assign the corresponding class to the test data.
 Else if
 the classifier level is the 3rd level,
 Then,
 if J = 1, assign the corresponding class to the test data.
 else reject the test data for classification

Step6: Repeat step 2 to 5 for next level classifiers till the final decision about the test data (whether to be accepted to a particular class or fully rejected) is obtained.

However, the above rule is very stringent and also it requires a large amount of data per class for getting an accurate class boundary. Hence, we have incorporated the following modifications in our algorithm.

If, at the last level of classifier (refer step 6 above),
 J = zero or $J \geq 2$, Go To step 1
 Assign the test data to the j-th class for which,
 $d_j = \text{minimum of } \{d_i\}_{i=1}^{21}$

Figure 2 (a), (b) and (c) shows the three classification criteria discussed above

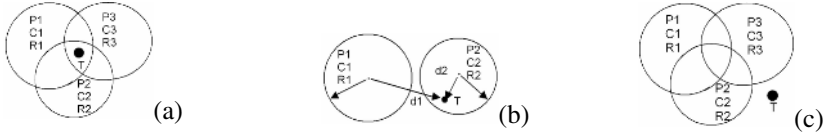


Fig. 2. (a): Test feature point lying in more than one classes (P, C and R are the respective person, class center and class radius). (b): Test feature point lying in one class and satisfying radius test i.e. $d_{T2} < R_2$. (c): Test feature point lying outside all the classes (the case of rejection).

5 Benchmarking

For benchmarking purpose we compare the result of our approach to that of Kittler’s sum rule, which is considered to be best among all combination schemes. For this we have to find out the efficiency of the Kittler’s sum rule [2].

Kittler’s sum rule: assign test feature point, θ to class w_j if,

$$(1 - R)P(w_j) + \sum_{i=1}^R P(w_j | x_i) = \max_{k=1}^m \left[(1 - R)P(w_k) + \sum_{i=1}^R P(w_k | x_i) \right] \tag{7}$$

where, R is number of features, m is the number of classes, x_i is the i-th classifier based on the feature, x_i , $P(w_j)$ = prior probability of the class w_j for raw data, and $P(w_k|x_i)$ = aposteriori probability of the class w_k given the feature, x_i

6 Result and Discussions

Table 1 shows that, when the wavelets, ridges and multi-fractal classifiers were used independent of each other; the recognition accuracy produced was very low at 30.95%, 21.43 and 19.05% respectively. But on combining the three classifiers through our above-described method, the recognition accuracy increased to 80.95% while the recognition accuracy of the Kittler’s sum rule also increased to 57%. This shows that the performance of a pattern recognition system can be improved significantly by multiple classifier systems proposed by us. It also shows that the judicious selection and combination of classifier can increase the efficiency of the recognition system by many folds.

Table 1. shows some representative results of applying the proposed methodology to a set of test images

Number of Queries	Efficiency of Wavelet based Classifier (in %)	Efficiency of Ridge based Classifier (in %)	Efficiency of multi-fractal based Classifier (in %)	Efficiency of proposed multiple classifier (in %)	Efficiency of Kittler’s Sum Rule (in %)
42	30.95	21.43	19.05	80.95	57.14

7 Conclusions

The problem of combining classifiers, which use different representations of the patterns to be classified was studied. We have developed a decision decomposition based framework for utilizing decision obtained from multiple classifiers. Our proposed MCSDD showed that passing the input pattern through several classifier levels does analogous discriminatory function based on multidimensional data. The multiple classifier system designed here does not degrade the quality of any classifier and rather fully utilize the quality of each individual classifier.

Our result shows that MCSDD has far better efficiency than the existing multiple classifier systems. Incorporation of more features and large database is expected to enhance the classification efficiency further.

Acknowledgement

We gratefully acknowledge the financial support received by the corresponding author for continuing this work in the form of Grant-in-aid from ILTP Cooperation between India (DST) and Russia (RAS) for the Indo-Russian collaborative project.

References

1. Pudil, P.; Novovicova, J.; Blaha, S. and Kittler, J., "Multistage Pattern Recognition with Reject Option," Proc. 11th IAPR Int'l Conf. Pattern Recognition, Conf. B: Pattern Recognition Methodology and Systems, vol. 2, 1992, pp. 92-95.
2. Kittler, J.; Hatef, M.; Duin, R.P.W.; Matas, J.; "On combining classifiers.," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 20 , Issue: 3, March 1998 ,pp:226 – 239.
3. Denisov, D.A. and Dudkin, A.K.; "Model-Based Chromosome Recognition via Hypotheses Construction/Verification," Pattern Recognition Letters, vol. 15, no. 3, 1994, pp. 299- 307.
4. Fairhurst, M.C. and Abdel Wahab, H.M.S., "An Interactive Two-Level Architecture for a Memory Network Pattern Classifier," Pattern Recognition Letters, vol. 11, no. 8, 1990, pp. 537-540.
5. Feder, "Fractals", Plenum Press, New York, 1988.
6. Kimura, F. and Shridhar, M., "Handwritten Numerical Recognition Based on Multiple Algorithms," Pattern Recognition, vol. 24, no. 10, 1991, pp. 969-983.
7. El-Shishini, H.; Abdel-Mottaleb, M.S.; El-Raey, M. and Shoukry, A. "A Multistage Algorithm for Fast Classification of Patterns," Pattern Recognition Letters, vol. 10, no. 4, 1989, pp. 211-215.
8. Kurzynski, M.W., "On the Identity of Optimal Strategies for Multistage Classifiers," Pattern Recognition Letters, vol. 10, no. 1, 1989, pp. 39-46.
9. Zhou, J.Y. and Pavlidis, T. "Discrimination of Characters by a Multi-Stage Recognition Process," Pattern Recognition, vol. 27, no. 11, 1994, pp. 1,539-1,549.
10. Hashem, S. and Schmeiser, B. "Improving Model Accuracy Using Optimal Linear Combinations of Trained Neural Networks," IEEE Trans. Neural

11. Networks, vol. 6, no. 3, 1995, pp. 792-794.
12. Lahiri T. and Samal S., "A novel technique for making multiple classifier based decision", Proc. WSEAS International Conference on MATHEMATICAL BIOLOGY and ECOLOGY, Corfu, Greece, August 17-19, 2004
13. Jain, A.K, Prabhakar, S. Hong, L. and Pankanti, S., 2000 "Filterbank-Based Fingerprint Matching" IEEE Transaction on Image Processing, Vol. 9, No. 5, May 2000