Fusion of Combination Rules of an Ensemble of MLP Classifiers for Improved Recognition Accuracy of Handprinted Bangla Numerals

U. Bhattacharya and B. B. Chaudhuri
CVPR Unit, Indian Statistical Institute, Kolkata-108, India
{ujjwal,bbc}@isical.ac.in

Abstract

In handwritten character recognition problem, the input images are often affected by distortions and noise. Thus such images at different resolutions include different variations in the input data. In the present work, we considered wavelet transform to obtain multi-resolution representation of each input character image. At each resolution level, we considered three MLPs with different numbers of nodes in their hidden layers and combined the outputs produced by all the MLPs of the whole ensemble by using weighted sum rule, product rule and majority voting. The set of misclassified samples produced by one combination rule is neither a subset nor a superset of a similar set produced by another rule. So, majority voting has been used for the second and final round to produce final outputs after combining the results of the three combinations of the first stage. The proposed approach produced 99.10% correct recognition rate on the test set of Bangla (a major Indian script) numeral database.

1 Introduction

Many diverse algorithms/schemes for handwritten character recognition [1] exist and each of these has its own merits and demerits. An important aspect of a handwriting recognition scheme is the selection of a good feature set and a large number of feature extraction methods are available in the literature [2]. So, it seems justified to investigate how an existing feature extraction method(s) can be used along with an intelligent classification strategy to achieve both speed and acceptable recognition accuracies in different scripts. Additionally, a feature set which is independent of any particular script has special importance with respect to multiscript countries like India.

In the present work, we consider a script independent feature set for recognition of handwritten numerals. Daubechies [3] wavelet transform has been used to compute features at five different resolution levels – images at two finer and another two coarser resolutions in addition to the original resolution have been obtained. Examples of handwritten numeral whose true class cannot be identified in its original or coarser resolution levels but it may be identifiable at a higher resolution (similar to seeing through a magnifying glass) is shown in Fig. 1. For a given numeral image, images at higher resolutions (equivalent to magnification by 200% and 400%) are obtained by interpolating the input image using wavelet composition method. MLPs are considered as classifiers at each resolution levels.

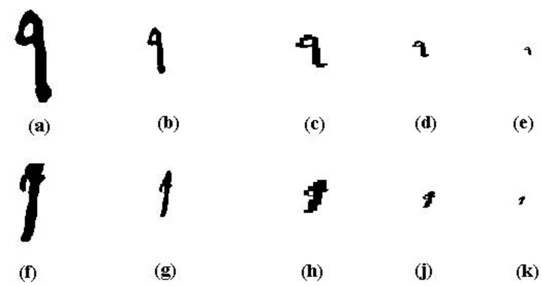

Figure 1. $(a)-(e)$ **Images of a handwritten English '9' at different resolutions.** $(f)-(k)$ **Images of a handwritten English '1' at different resolutions.** $(c),(h)$ **Original resolutions.**

Since the wavelet transform provides an invariant interpretation of a character image at different resolution levels characterizing different physical structures of the character, this feature extraction scheme intuitively seems to be very effective for handwritten character recognition tasks. An early work in which a wavelet based approach had been considered for recognition of handwritten characters is found in [4]. In [5] a study of a few combination rules on the classification results at different resolutions had been reported.

For a given input data set, the classification accuracy provided by an MLP classifier, largely depends on the number of nodes in its hidden layer(s). During extensive simulation runs, we observed that the set of correctly recognized samples corresponding to any particular choice of hidden layer size is neither a superset nor a subset of the similar set corresponding to another choice of hidden layer size.

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 **IEEE**

Similar observations are also made when resolutions of input images are different but the sizes of the hidden layers are the same.

The above observations indicate that variations with respect to resolution of input images and hidden layer size of MLP classifier encode different characteristics of the input data. On the other hand, it has been already established that an ensemble of classifiers usually provide improved accuracies in pattern recognition applications. Based on these facts, we considered an ensemble of MLP classifiers at each resolution level and finally fusion of outputs of the whole set of MLP classifiers are done. A number of classifier fusion methods are available in the literature. One popular approach for classifier fusion deals with the outputs of different members of the classifier ensemble. Intuitively, it is understood that such fusion of an ensemble of classifiers performs better than a single classifier when each member in the ensemble includes certain diversities and no single member can perform perfectly.

A number of classifier fusion methods are available in the literature. One popular approach for classifier fusion deals with the outputs of different members of the classifier ensemble. Intuitively, it is understood that such fusion of an ensemble of classifiers performs better than a single classifier when each member in the ensemble includes certain diversities and no single member can perform perfectly. In the present work, we considered three approaches such as weighted sum rule, product rule and majority voting for combination of MLP outputs. Among these three combination approaches, product and sum rules are more popular than the majority voting scheme. However, in general, none of them can individually be considered as the best. So, we considered a second level of fusion of the results obtained after initial fusion. At this second level, the majority voting scheme, the simplest combination strategy, has been considered.

Simulation results of the above described approach have been obtained on a database of handwritten Bangla (the second most popular script in the Indian subcontinent) numerals. A few samples from this database is shown in Fig. 2. We obtained 99.10% correct classification and 0.2% rejection on a test set of 5000 Bangla numerals. The speed of recognition is more than 60 numerals per second on a PIV desktop computer.

The rest of this article is organized as follows. In Section 2 we describe wavelet transform-based multiresolution features. MLP classifiers and their combination strategies are described in Section 3. The details of the present recognition approach has been detailed in Section 4. Simulation results are reported in Section 5. Section 6 concludes the present article.

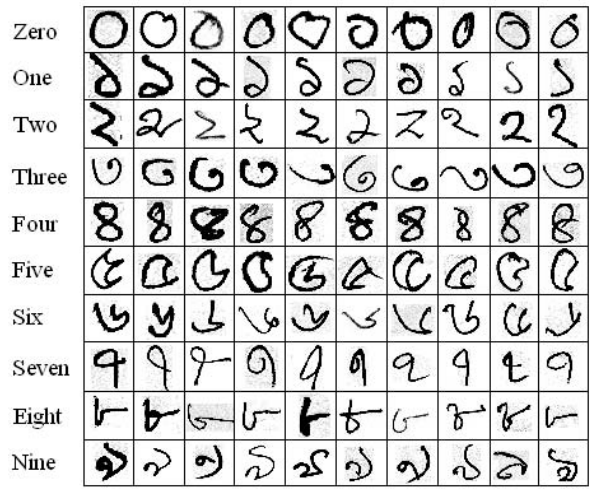

Figure 2. Typical sample set of handwritten Bangla numerals

2 Wavelet descriptor for multiresolution representation of pixel image

A wavelet transform is orthogonal and operates on an input vector whose length is an integral power of two. This is a fast linear operation. It generates a vector which is of the same length but numerically different from the input vector. Wavelet transform can be viewed as a rotation in function space, from the input domain to a different domain. The basis functions of the wavelet domain are called wavelets. Wavelets are quite localized both in space and in frequency.

There exist infinitely many possible sets of wavelets and different sets of wavelets make different trade-offs between how compactly they are localized in space and how smooth they are. A wavelet transform is usually implemented by a binary tree of filters. The art of finding good wavelets lies in the design of these set of filters which achieve the above trade-offs and also make the perfect reconstruction of the original signal possible.

The working principle of a wavelet transform is as follows. An input signal x is split into a lowpass or smooth component x_0 and a highpass or detail component x_1 respectively by a lowpass filter L and a highpass filter H. Both of these two components are down-sampled in the ratio 2:1. The lowpass component x_0 is then split further into x_{00} and x_{01} by using the above filters for the second time and these are again down-sampled in the ratio 2:1. This process (pyramidal algorithm [6]) of splitting and down-sampling is continued as far as required or a trivial size of the smooth...smooth component (usually 2) is reached.

The first and simplest possible orthogonal wavelet system is the Haar wavelet (Thesis of A. Haar, 1909). However, Daubechies [3] constructed a set of orthonormal wavelet basis function that are perhaps the most elegant. These wavelets are compactly supported in the time-domain and have good frequency domain decay. This describes the reason behind our choice of Daubechies wavelet transform. The simplest member of this family of wavelets is the Daubechies-4 wavelet which has only four coefficients

$$l_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, l_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, l_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, l_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$

The above coefficients form the lowpass or smoothing filter L and another set of four coefficients

$$h_0 = l_3, h_1 = -l_2, h_2 = l_1 \text{ and } h_3 = -l_0$$

form the highpass filter H. (In signal processing contexts L and H are called quadrature mirror filters.)

A simple extension of the above principle to multidimensional arrays is possible. A wavelet transform of an image, a 2-dimensional array, is easily obtained by transforming the array on its row index (for all values of its column indices), then on its column. Each transformation corresponds to multiplication by an orthogonal matrix and by matrix associativity, the result is independent of the order in which the row or column are transformed.

The layout of application of wavelet transform recursively on an image is shown in Figure 3. The successive application of the transform produces an increasingly smoother version of the original image.

(a) (b) (c) (d) (e)

Figure 3. (a) **Layout of wavelet decomposition (L $\rightarrow$ low-pass filter, H $\rightarrow$ high-pass filter, $j \rightarrow$ level no.);** (b) **Original image of a Bangla handwritten numeral;** (c) **The size normalized(32×32) image;** (d) **and** (e) **Smooth...smooth components of wavelet decompositions at resolution levels 16×16 and 8×8 respectively.**

3 MLP

3.1 Ensemble of MLP classifiers

An ensemble of classifiers [7] is usually used to increase the robustness and performance of classification. An ensemble of MLP classifiers is often practical and provide effective solutions for difficult pattern recognition tasks. The motivation for such systems may be derived from an empirical observation that specialized MLP architectures are superior in different cases, or it may follow from the nature of the application. In several other cases, the motivation for using such an ensemble is to avoid the trial and error approach for making a near optimal choice of hidden layer size(s). In this approach, we abandon the attempt to find the best architecture of an MLP classifier, and instead, attempted to use a number of choices in a smart way.

3.2 Combination of multiple MLP classifiers

Depending on which of the two output information levels a combination is based on, the problems of combining multiple MLP classifiers can be summarized in the following two types:

- Given K individual MLP classifiers, C_i, $i = 1, \ldots, K$, each of which assigns a label y_{ij} to a given input x, producing an event $C_i(x) = y_{ij}$, the problem is to combine these events to give x one definite class label y_j, where j denotes one of the possible classes.
- The combination is based on the output information at the measurement level. For an input x, each C_i produces a real vector $Y_i = [y_{i1}, \ldots, y_{iN}]$ (where y_{ij} denotes the degree that C_i considers that x has label y_j), the problem is to use these events $C_i(x) = Y_i, i = 1, \ldots, K$ to design a combination strategy Q with $Q(x) = y_j$, where j denotes one of the possible classes.

Since the main reason for combining an ensemble of MLPs is to improve their recognition performance, we considered MLPs with varying architectures, and inputs at different resolutions.

In the literature, there exists a variety of methods [8, 9] for the combination of results obtained from multiple classifiers. Tax et al. [10] have presented a detailed study of these combination rules. In a few handwritten character recognition approaches such as [11] multiple classifiers were used for improved recognition accuracies.

In the present article, we considered five different resolution levels of an input character image and three different choices of hidden layer size at each resolution level. The outputs by the fifteen members of the above ensemble of MLP classifiers have been combined using three different rules – weighted sum rule, product rule and majority voting. The three class labels provided by these three combination approaches have been combined at the second level to obtain final classification.

4 Recognition scheme

Initially, the input image of a handwritten numeral is binarized temporarily using a simple histogram-based thresholding technique. The binary image is used to compute

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

its bounding box (smallest possible rectangle enclosing the image). The bounding box containing the gray-level image is first normalized to the size 32×32 using the moment method [12]. Wavelet decomposition algorithm is applied to this normalized image recursively for two times to obtain an 8×8 smooth. . .smooth approximation of the original image as the final decomposition. Here, we also obtain a 16×16 smooth. . .smooth approximation of the original image at an intermediate stage. The 16×16 and 8×8 approximations of the size normalized sample in Fig.1(c) are shown in Figs.1(d) and 1(e) respectively. We also consider inverse wavelet transform to magnify the size normalized image to 64×64 using three other null matrices of the same size. Similarly, it is magnified to 128×128 also. These two magnified versions of the image in Fig. 1(c) are shown in Figs. 1(b) and 1(a) respectively. All these versions of the original image are gray-valued images. We used Otsu's thresholding technique [13] for binarization of each of the above images.

Each of the above 128×128, 64×64, 32×32, 16×16, 8×8 binarized images of an input numeral are fed to the input layers of 3 MLP networks whose hidden layer sizes are respectively $\frac{1}{8}$, $\frac{1}{4}$ and $\frac{1}{2}$ of the size of the respective input layers. The responses at the output layers of these 15 MLPs are separately combined using the sum and product rules and a voting scheme. For the weighted sum and product rules, the responses at the output layers are directly used for combination purpose. In these cases, the responses of 15 MLPs added or multiplied class-wise and the resulting values corresponding to all of the 10 classes are compared and decision about the output class is taken in favour of the maximum such values. However, for the voting scheme, different responses at the nodes of the output layers of each of the fifteen MLPs are compared and for any particular MLP, the output node with maximum value determines the recognized class of the input numeral by that MLP. Thus, the majority voting scheme decides the particular class as the output in favor of which maximum number of MLPs responded. Here, it is possible that no clear majority is obtained. In such cases, the input numeral is said to be rejected by the voting scheme.

Finally, it is required to combine the results delivered by the three combination rules. For this, we consider the majority voting scheme for the second time. In this final level, an input numeral is decided to be rejected if the responses of the three combining rules differ from each other. On the other hand, if at least two combining strategies classify the input numeral into the same category, the majority voting scheme recognizes it to belong to this class.

5 Experimental results

The authors are not aware of the availability of any standard database of handwritten Bangla numerals. So, such a database had been developed with the help of a group of University students. Our database consists of 12,938 isolated handwritten Bangla numerals. These data had been collected from different sections of the population of West Bengal, India and this includes variations with respect to age, sex, education, place of origin, income group and profession. Since there appears variation in the writing style of a single individual at different points of time, each individual has been approached on maximum 4 different occasions for the sample.

Table 1. Final Confusion Matrix on Bangla Numeral Database

	০	১	২	৩	৪	৫	৬	৭	৮	৯
০	**99.60**	0.20	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00
১	0.20	**98.80**	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.20
২	0.00	0.40	**98.60**	0.00	0.20	0.00	0.00	0.00	0.20	0.20
৩	0.20	0.00	0.00	**99.20**	0.00	0.20	0.20	0.00	0.20	0.00
৪	0.00	0.20	0.20	0.20	**99.40**	0.00	0.00	0.00	0.00	0.00
৫	0.40	0.00	0.00	0.20	0.00	**98.60**	0.40	0.00	0.00	0.00
৬	0.00	0.00	0.00	0.20	0.00	0.40	**98.80**	0.00	0.00	0.00
৭	0.20	0.00	0.00	0.00	0.20	0.00	0.40	**99.20**	0.00	0.00
৮	0.00	0.20	0.00	0.00	0.20	0.00	0.00	0.00	**99.60**	0.00
৯	0.00	0.40	0.20	0.00	0.00	0.00	0.00	0.00	0.00	**99.20**

The whole set of 12,938 samples of Bangla numerals, is composed of three subsets called training set, validation set and test set consisting respectively of 6938 , 1000 and 5000 samples. Validation data sets have been used for termination of training of respective MLP classifiers.

The best performance with respect to each of the three choices of hidden layers have been obtained when 32×32 images are fed to the input layer of respective MLPs. So, for combination of MLP outputs using weighted sum rule, we selected 0.6, 0.6, 2.6, 0.6 and 0.6 as the weights of the output vectors of MLPs with input layer sizes 128×128, 64×64, 32×32, 16×16 and 8×8 respectively. During our extensive simulations, we considered quite a few such sets of weights and the above set of weights was found to be the best. Consideration of the maximum weight in favour of the original resolution level is justified by the fact that it presents the image before any smoothing or interpolation and also the recognition performance of the concerned MLPs are the highest among the MLPs with same ratio of input and hidden layer sizes.

In the five fine-to-coarse resolution levels, the correct classification percentages with respect to hidden layers, whose sizes are half of the respective input layers, are respectively 96.6%, 96.6%, 97.1%, 95.58% and 94.76% on the test dataset. Fusion of outputs of fifteen MLPs at the first level using weighted sum rule provided 98.2%

recognition accuracy. Fusion of combination rules at the second level of combinations of recognition results provided 99.1% final true classification and 0.2% samples have been rejected. Thus we achieved only 0.7% misclassification which is comparable to the existing state-of-the art techniques. The final confusion matrices corresponding to Bangla in Table 1.

6 Conclusions

The potential of wavelets in image compression is established and its applicability in various other image processing problems is being explored. On the other hand, during the last decade, combinations of classifiers to achieve higher classification accuracies have been studied with great interest. In this paper we have considered three different rules for combinations of fifteen MLP classifiers with different hidden layer sizes and trained using features at different resolution levels. Integration of the results of three combination rules have been done using the majority voting scheme. This novel strategy of using integration of multiple classifier combiners helps to improve recognition accuracy without any significant increase in computation. The proposed approach is independent of the script. Its recognition accuracy on Bangla numeral database is comparable to the state-of-the-art techniques. Also, this is fast enough for its implementation in real-life applications. Finally, the wavelet based features are also not affected in the presence of moderate noise or discontinuity or small changes in orientation.

References

[1] Plamondon, R., Srihari, S. N.: On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. IEEE Trans. Patt. Anal. and Mach. Intell. 22(2000) 63-84.

[2] Trier, O. D., Jain, A. K., Taxt, T.: Feature Extraction Methods for Character Recognition - A Survey. attern Recognition. 29(1996) 641 - 662.

[3] Daubechies, I.: The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. on Information Theory. 36(1990) 961-1005.

[4] Wunsch, P., Laine, A. F.: Wavelet Descriptors for Multiresolution Recognition of Handprinted Characters. Pattern Recognition. 28(1995), 1237-1249.

[5] U. Bhattacharya, S. Vajda, A. Mallick, B. B. Chaudhuri and A. Belad: On the Choice of Training Set, Architecture and Combination Rule of Multiple MLP Classifiers for Multiresolution Recognition of Handwritten Characters, Proceedings of 9th Int. Workshop on Frontiers in Handwriting Recognition, (Eds. F. Kimura and H. Fujisawa), Tokyo, October 26-29, 2004, pp. 19-424.

[6] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation", *IEEE Trans. on Pattern Anal. and Machine Int.*, vol. 11(7), 1989, pp 674 -693.

[7] Kittler, J., Hatef, M., Duin, R. P. W., Matas, J.: On combining classifiers. IEEE Trans. PAMI. 20 (1998), 226 - 239.

[8] C. Y. Suen, C. Nadal, T. A. Mai, R. Regault and L. Lam, "Computer Recognition of Unconstrained Handwritten Numerals", *Proc. IEEE*, vol. 80, 1992, pp. 1162-1180.

[9] T. K. Ho, J. J. Hull and S. N. Srihari, "Decision Combination in Multiple Classifier Systems", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 16, 1994, pp. 66-75.

[10] Tax, D. M. J., Breukelen, M. V., Duin, R. P. W., Kittler, J.: Combining multiple classifiers by averaging or by multiplying?. Pattern Recognition. 33 (2000), 1475 - 1485.

[11] F. Kimura and M. Sridhar, "Handwritten Numeral Recognition Based on Multiple Algorithms", *Pattern Recognition*, vol. 24, 1991.

[12] Casey, R. G.: Moment normalization of handprinted characters. IBM J. Res. Develop., 14(1970) 548-557.

[13] N. Otsu, A Threshold Selection Method from Grey-Level Histograms, IEEE Trans. Systems, Man, and Cybernetics, vol. 9, 1979, pp. 377-393.