A Modified Decision Templates Method for Persian Handwritten Digit Recognition

Mohammad Masoud Javidi [1], Fatemeh Sharifizadeh [2]

[1.] Department of Computer Science, Shahid Bahonar University of Kerman, Kerman, Iran. javidi@mail.uk.ac.ir
[2.] Department of Computer Science, Shahid Bahonar University of Kerman, Kerman, Iran.
fSharifizade@mail.uk.ac.ir

Abstract: In this paper a new method to combine multiple classifiers based on a static structure is presented. We establish our model based on decision templates (DT), as we do not only rely on similarity between a test sample x and c decision template matrices, moreover to make a decision about pattern x we construct q wrong decision templates, and compute likeness between pattern x and these matrices. We call this novel method Wrong Decision Templates (WDT). To evaluate our proposed model we use a very large dataset of Persian handwritten digit (HODA).The experimental results support our claim that constructing WDT matrices besides DT matrices, improves the performance of the conventional DT for Farsi handwritten digit recognition, such that the recognition rate of 98.16% is achieved, which has 60% decline of error rate with regard to DT method. Furthermore, Comparison other static combination methods indicates that the proposed model yields excellent recognition rate in handwritten digit recognition. Finally, the generalization capability of our proposed method is considered on two benchmark datasets from the UCI repository.

[Mohammad Masoud Javidi, Fatemeh Sharifizadeh. A Modified Decision Templates Method for Persian Handwritten Digit Recognition. Journal of American Science 2012;8(1):504-512]. (ISSN: 1545-1003). http://www.americanscience.org. 70

Keywords: Decision Templates; Handwritten Digit Recognition; Combining Classifiers; Wrong Decision Templates; Classifier Fusion

1. Introduction

In the last few decades, numerous methods have been proposed for machine recognition of handwritten characters, especially for the more popular languages such as English, Japanese and Chinese. In particular, handwritten numeral recognition has attracted much attention, and various techniques (pre-processing, feature extraction, and classification) have been proposed (Liu et al., 2003; Trier et al., 1996; Ho et al., 1994; Xu et al., 1991; Suen et al., 1990).

In contrast, very little research has been reported for the recognition of Persian (Arabic) handwritten digits (e.g. (Liu et al., 2009; Pan et al., 2009; Borji et al., 2008; Suen et al., 2006; Soltanzadeh et al., 2004; Amin, 1998). However, today research on Farsi (Persian) scripts and numerals is receiving increasing attention because of the automatic processing of handwritten data.

Combining classifiers is an approach which improves the performance of classification, particularly for complex problems such as those involving a limited number of patterns, high-dimensional feature sets, and highly overlapping classes (Ho et al., 1994; Soltanzadeh and Rahmati, 2004). There are two main strategies for combining classifiers: fusion and selection (Kuncheva, 2004) In fusion, we suppose that each ensemble member is trained on the whole feature space (Xu et al., 1992; Ng and Abramson, 1992; Kittler et al., 1998), whereas in selection, each member is assigned to learn a part of the feature space (Wood et al., 1997; Jacobs et al., 1991; Alpaydin et al., 1996; Haykin, 1998). Thus, in the former strategy, the final decision is made on the basis of the decisions of all members, while in the latter strategy, the final decision is made by aggregating the decisions of one or some of the experts (Kuncheva, 2004; Haykin, 1998).

One of the most popular methods of classifier fusion is the scheme of *Decision Templates* (DT), originally proposed by Kuncheva (Kuncheva et al., 2001). *DT* is a robust classifier fusion scheme that combines classifier outputs by comparing them with a characteristic template for each class. R.Ebrahimpour, F.Sharifizadeh, (Ebrahimpour and Sharifizadeh, 2009) used this method for Persian handwritten digit recognition.

In this paper, we propose a new fusion method which is essentially based on decision templates. In our model we do not rely only on the similarity between a test sample x and c decision template matrices, moreover, to make a decision about patterns we also purposely construct some wrong decision template matrices, and compute the likeness of pattern x with both decision template matrices and wrong decision template matrices.

The rest of this paper is organized as follows: in the next Section, we briefly describe our methods for feature extraction. We then review fusion methods in Section 3 and we describe the proposed model in detail in Section 4. Section 5 provides and illustrates the experimental results. Finally, Section 6 draws a conclusion and summarizes the paper.

2. Feature Extraction

The selection of a feature extraction method with a good discriminating power is probably the single most important stage for transforming the input space into the feature space. In order to avoid a high dimensional and redundant input space and to optimally design and train the experts, we first use the Characteristic Loci method and then Principle Component Analysis (PCA). Characteristic Loci is a robust feature extraction method much used in the literature of Persian handwritten digit recognition (Glucksman, 1967; Ebrahimi and Kabir, 2008; Knoll,1969). PCA is a common technique for extracting informative low dimensional patterns in data of high dimension, with no harmful loss of information content. It is basically a way of identifying patterns in data, along with their similarities and differences (Martinez and Kak, 2001; Manjunath, 2008).

3. Classifier Fusion

Let $x \in R^n$ be a feature vector, $\{D_1,\ldots,D_L\}$ a set of classifiers and $\Omega = \{\omega_1,\ldots,\omega_c\}$ the set of class labels. We denote the output of the i'th classifier as $D_i(x) = [d_{i,1}(x),\ldots,d_{i,c}(x)]^T$, where $d_{i,j}(x)$ indicates the support that classifier D_i gives to the supposition that x comes from class ω_j. The L classifier outputs for an input pattern x can be arranged in a decision profile matrix (DP(x)) as shown in the Figure 1 (Kuncheva, 2004):

$$DP(x) = \begin{bmatrix} d_{11}(x) & \ldots d_{1j}(x) & \ldots & d_{1c}(x) \\ \vdots & \ddots \; \vdots & \ddots & \vdots \\ d_{i1}(x) & \ldots d_{ij}(x) & \ldots & d_{ic}(x) \\ \vdots & \ddots \; \vdots & \ddots & \vdots \\ d_{L1}(x) & \ldots d_{Lj}(x) & \ldots & d_{Lc}(x) \end{bmatrix}$$

Figure 1. Decision profile matrix for an input pattern x

There are two general approaches for using DP(x) to find the overall support for a class and subsequently label the input x in the class with the largest support. Some methods calculate the support for class i $(\mu_D^i(x))$ using only the i'th column of DP(x). Such methods are referred to as *class-conscious methods*. The alternative fusion approach is to use *all of DP*(x) to calculate the support for each class. Fusion methods in this category are called *class-indifferent* methods. In this paper, we base our novel model on Decision Templates, which is under the category of class indifferent methods. Since our proposed model is a modified form of DT, this method is briefly described in the subsection 3.1.

3. 1. Decision Templates

The idea of the decision templates combiner is to remember the most typical Decision Profile for each class ω_j, called the decision template, DT_j, and then compare it with the current decision profile DP(x) using a similarity measure S. The closest match will be labeled x.

Let $X = \{x_1,\ldots,x_n\}, x_i \in R^n$, be the training dataset that belongs to the class set $\Omega = \{\omega_1,\ldots,\omega_c\}$ and $D = \{D_1,\ldots,D_L\}$ be a set of classifiers.

Definition: The decision template DT_i for class i is the average of the decision profiles of the elements of the training set X, labeled in class i. Thus $DT_i(X)$ of class i is the $L \times c$ matrix $DT_i(X) = [dt_i(k,s)(X)]$ whose (k,s)'th element is computed by:

$$dt_i(k,s)(X) = \frac{\sum_{j=1}^{N} Ind(x_j,i)\, d_{k,s}(x_j)}{\sum_{j=1}^{N} Ind(x_j,i)}, k = 1,\ldots,L, s = 1\ldots c \quad (1)$$

where $Ind(x_j,i)$ is an indicator function with value 1 if pattern x_j belongs to class ω_i, and 0, otherwise (Kuncheva, 2004; Kuncheva, 2001). Henceforth we shall write DT_i instead of $DT_i(X)$.

After constructing DT matrices, in the testing phase, if $x \in R^n$ is submitted for classification, the DT scheme matches $DP(x)$ to $DT_i, i = 1,\ldots,c,$ and produces soft class labels by:

$$\mu_{D_{ens}}^i(x) = S(DT_i, DP(x)), i = 1,\ldots,c \quad (2)$$

where S is interpreted as a *similarity* measure. The higher the similarity between the decision profile of the current x (DP(x)) and the decision template for class i (DT_i), the higher the support for that class. Two measures of similarity are the: Squared Euclidean distance $(DT(E))$ and the Symmetric difference $(DT(S))$ (Kuncheva, 2001). Several studies have looked into possible applications of DT

(Dietrich et al., 2003; Dietrich et al., 2001; Kittler et al., 2002; Giacinto et al., 2003).

4. Proposed Model: Wrong Decision Templates

Our proposed technique is close to DT method. In the DT method only c decision template matrices corresponding to c classes are constructed, where the i' th matrix is constructed from all samples that belong to class i. However, some samples in class i might not have primitive structures and averaging from all samples in this class converge to the *template* which is caused to misplacing samples that have not primitive structure. Thus in the test phase some patterns, which belongs to the class c but do not have primitive structure, are wrongly classified. That is, when similarity between input sample and DT matrices is calculated, the structure of test sample might have more resemblance to the structure of another class. In other words, absence of primitive structures in some patterns in DT may lead to misclassification. For example, some pairs of numerals in Persian handwritten digits are more easily confused than other; such as 0-5, 2-3, 6-9. Some of such images are shown in Figure 2.

Input pattern				
True label	5	3	4	6
Wrong label	0	2	3	9

Figure 2. some misclassified samples of Persian numerals. Each image in the first row must be classified as it true label in the second row, however it misrecognized and classified with wrong label in the third row.

Because some digits do not have primitive structure and due to writing habits like the filled loop or cursive writing they might be misclassified by our base classifiers. So we use *characteristic template* for these samples.

In our model, after calculating DT matrices, to increasing efficiency and recognition rate in the recognition system, we construct a confusion matrix and *Wrong Decision Templates* (WDT) matrices. To form each WDT matrix, the average of the decision profiles of the elements of the training set X whose true class is i, and which are assigned to class j is calculated. Clearly, take a decision about class of test sample from both of DT and WDT matrices is *far more effective* than using only DT matrices to decision making about a class of test sample.

With different number of WDT matrices, the different shapes of some classes that do not have a common shape are included. In some cases, samples that are not classified in their own class are less than to configure WDT. Thus, depending on the dataset, the number of WDT matrices can vary. In addition it must be noted that since WDT matrices are only computed for wrong data, the computational load is not significant.

Let $X=\{x_1,\ldots,x_n\}, x_i \in R^n$, be the training dataset that belongs to the class set $\Omega=\{\omega_1,\ldots,\omega_c\}$, and $D=\{D_1,\ldots,D_L\}$ be a set of classifiers. The confusion matrix for all classifiers is constructed (Catherine et al., 2002). We continue the procedure as follow:

1. Train base classifiers with X. Construct Decision Profile matrices for X with the outputs of base classifiers.
2. Construct DT matrices in its usual manner.
3. Select the best classifier D and then form its confusion matrix (CM).
4. Using the CM, construct DT for samples that are classified in a particular group by this classifier, in spite of the fact that they are really belong to another class. In fact, DT matrices are constructed for wrongly classified samples. Each one of such matrices will be called a Wrong Decision Template (WDT). The $WDT_r, r=1,...,q,$ is the average of the decision profiles of the elements of the training set X whose true class is i, but are assigned by D to class j. Thus WDT_r is the $L \times c$ matrix. $WDT_r(X)=[wdt_r(k,s)(X)]$ whose $(k,s)'th$ element is computed by:

$$wdt_r(k,s)(X)=\frac{\sum_{t=1}^{N} Ind(x_t,i,j)d_{k,s}(x_t)}{\sum_{t=1}^{N} Ind(x_t,i,j)},$$

$$k=1,...,L, s=1,...,c \quad (3)$$

   where $Ind(x_t,i,j)$ is an indicator function with values 0 and 1. If pattern x_t belongs to class i, but is assigned to class j by D, then the value of indicator function is 1, otherwise its' value is 0. $d_{k,s}(x_t)$ in Eq.(3) is the $(k,s)'th$ value of Decision profile matrix of pattern x_t.
5. After constructing the DT and WDT matrices, in the test phase, when $x \in R^n$ is submitted for classification, the similarity between $DP(x)$ and DT_i, $i=1,\ldots,c$, as well as WDT_r, $r=1,\ldots,q$, is calculated. The match with the highest similarity is given the class label of current test pattern x.

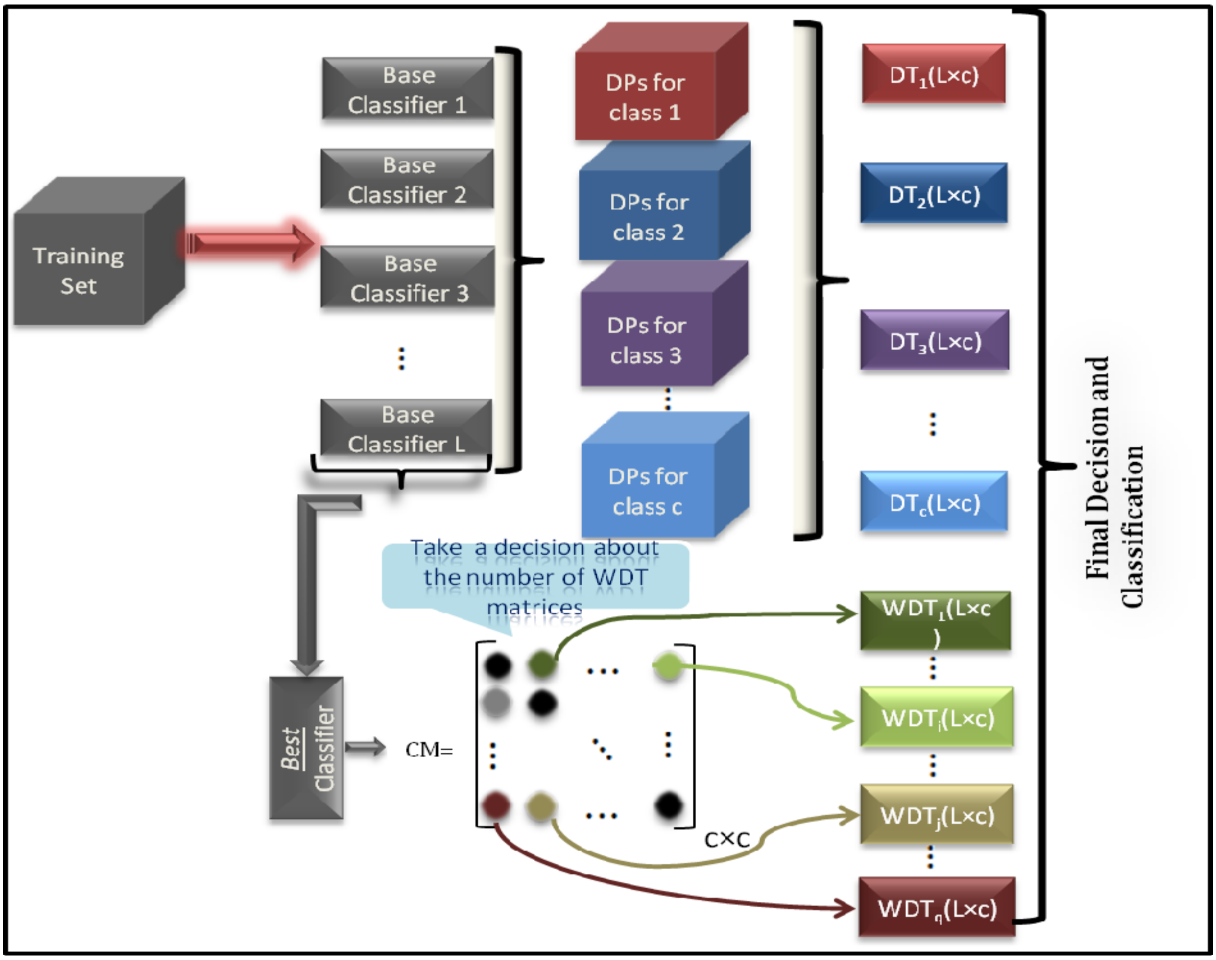

Figure 3. The structure of our proposed method.

Figure 3 shows structure for the whole process of our proposed method from a general view of operation.

5. Experimental Results

To evaluate the performance of our proposed method and also exhibit the advantage of using it in recognition of Persian handwritten digits, we carried out several experiments on the public domain datasets. In addition, we conducted performance evaluation comparisons for our proposed method with two datasets from the UCI repository. In these cases WDT outperforms DT.

5.1. Experiments on Two Persian Handwritten Digits Dataset

In this study, first we evaluate our recognition method on a very large dataset of Persian handwritten digit (HODA) (Khosravi and Kabir, 2007). This dataset is described in 5.1.1. In section 5.2 the generalization capability of WDT method is asserted on two datasets from the UCI repository. Each database is divided into training, validation, and test sets, which includes approximately 58%, 17%, and 25% of the available data respectively.

5.1.1. The Hoda Dataset

Khosravi and Kabir have gathered a very large corpus of Farsi handwritten digits in 2007. Binary images of 102,352 digits were extracted from about 12,000 registration forms of two types, filled by bachelor and senior high school students and then these forms were scanned at 200 dpi with a high speed scanner. The preprocessing, finding areas of interest and digit extraction, was performed and this Farsi digit dataset is divided into a set of 60,000 samples used for training and a set of 20,000 samples for testing.

The samples in this dataset are very accurate and simple, because the registration forms were scrupulously filled for the university entrance examinations and students pay great attention when completing such forms. Thus to provide a benchmark for evaluating our method for Persian handwritten digit recognition we extracted samples from HODA database which are harder to recognize. Using K-nearest neighbors method with $K = 6$ we selected the data which were classified into more than three classes. Some samples of 10 classes are shown in Figure 4. We finally ended up with the following subset for our experiments:

Training set:	12400 digits
Validation set:	3677 digits
Test set:	5360 digits

Figure 4. Samples of the subset of HODA Farsi dataset

5.1.2. Experiments

In the first stage of feature extraction stage we use the Characteristic Loci method and a feature vector with 81 components is extracted from each image. Then, in order to decrease computational load and to achieve high accuracy, dimensionality reduction was performed using PCA. In the first stage the number of PCA components must be specified. We used a MLP with 35 hidden neurons and 10 output nodes to specify the number of PCA components. Table 1 shows the error rate of the MLP computed with different number of PCA Components for the validation set of the subset of the HODA dataset.

Table 1. Error rates of the MLP with different number of PCA components for the subset of HODA dataset

Number of input neurons	30	40	50	60	70
Error rate (%)	7.26	6.31	**5.71**	5.73	5.85

A 50-dimensional subspace was found to be optimal for the subset of HODA dataset. Here these global eigenspaces were used in all subsequent experiments.

For this experiment we used a set of 4 classifiers. We will now shortly discuss the set of basic classifiers.

We use the MLP as the base classifiers with one hidden layer, with the connecting weights estimated by the error back-propagation, BP, algorithm minimizing the squared error criterion. 50 input nodes for the PCA components were used for a single MLP on the subset of HODA dataset, and 10 output nodes corresponded with the ten digits.

The MLP has learning parameters, such as number of hidden neurons, number of epochs and the learning rate. To find the best parameter values, we adjusted the parameters on the training set, and tested them on the validation set. Parameters that gave the best results on the validation set were used for classifying in the testing phase.

In this experiment, the learning rate for the MLP was 0.25 on the subset of HODA dataset. For diversifying base classifiers, the weights of MLP neural networks are initially set to small random values. In addition, different topologies for base classifiers are assumed.

For each of the ten classes Figure 5 illustrates the performance of each expert on the subset of HODA dataset. In this Figure the left most bar in each classifier corresponds to digit 0, class 1, and the right most bar point out to digit 9, class 10.

As mentioned, a confusion matrix can be used to realize the distribution of errors across the classes. Table 2 shows the confusion matrix of the recognition results for the best MLP on the subset of HODA dataset. For instance, two of the most misrecognized digits belong to digits 3 and 6 (See Table 2). As shown in Table 2, 90 images of digit 2 are mistaken for digit 3.

The results of our proposed method using different number of wrong decision template matrices are presented in Tables 3, where classification accuracy is shown. We only display the performance on the test sets, which have not been seen during training of either the individual classifiers or the second level fusion models. In each column the results for various number of WDT matrices is shown. Each result is the average of ten times testing on the subset of HODA dataset.

The left half section of the Table 3 deals with the Decision Templates method applied on all 4 base classifiers. In this method, in an ordinary manner, we calculate 10 decision template matrices to 10 ten classes. Decisions are made based on a Euclidean distance similarity measure. In the entire right half of the Table 3, the results of Wrong Decision Templates method (WDT) with different number of WDT matrices are shown.

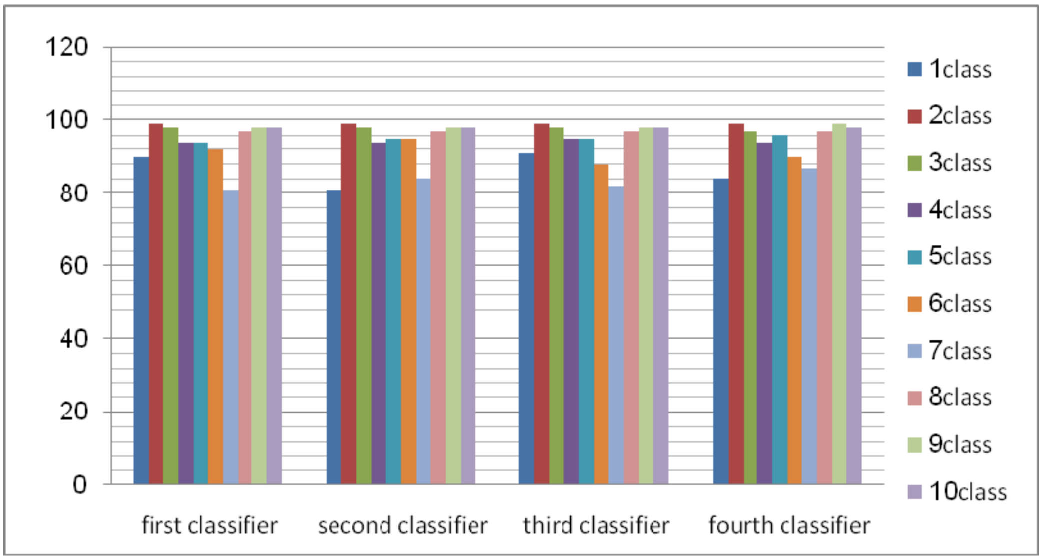

Figure 5. Recognition rates averaged over ten test runs on the subset of HODA dataset. The bars denote the average recognition rates of experts, broken down by 10 classes.

Table 2. Confusion matrix of best classifier for the 10 classes of the handwritten digit recognition on the subset of HODA dataset.

Class No.	0	1	2	3	4	5	6	7	8	9
0	**547**	0	0	0	0	30	0	0	0	0
1	1	**1715**	7	0	0	0	1	0	0	3
2	0	5	**2826**	9	16	0	6	0	0	2
3	1	0	<u>90</u>	**1965**	44	0	0	0	0	3
4	1	2	17	18	**1566**	2	2	0	0	0
5	38	0	0	0	59	**780**	25	0	14	0
6	3	11	222	7	33	1	**2450**	9	0	36
7	0	1	21	5	0	0	2	**858**	0	0
8	0	0	1	0	0	0	0	0	**384**	1
9	1	15	22	3	3	2	6	0	0	**1876**

Table 3. Recognition rate of WDT with different number of WDT besides result of DT method.

Technique	**Decision Templates Method**	**Wrong Decision Templates Method**						
Number of WDT	0	3	6	9	11	15	17	19
Recognition Rate (%) on the Subset of HODA Dataset	95.46	96.63	96.91	97.64	97.75	98.04	98.15	<u>**98.16**</u>

As described in last section, after calculating DT matrices, to increase the efficiency and recognition rates in our system, WDT matrices are constructed. For instance, according to Table 2, the network mistakes 90 samples of digit 3 for digit 2. We construct a specific WDT matrix (according to Eq. (11)) for these 90 misclassified samples of training set.

The best result of WDT methods in Table 3 is underlined. Note that when number of WDTs is 0, we have the Decision templates method. So in our experiment the Decision templates method is compared with our proposed method.

Table 3 shows that the recognition rate is improve as the number of WDT matrices increased from 0 to 19. As discussed in section 4, using *WDT matrices* to calculate similarity between input pattern and both DT matrices and several WDT matrices is *far more effective* than take decision about class of test sample from only 10 DT matrices. As shown, the recognition rate of Wrong Decision Templates is superior to that of decision templates method.

With our dataset adding more WDT matrices is not good, because the numbers of samples that are not classified in their own class are less than to configure WDT or samples are noisy and it is normal to misrecognized. So we ignore adding more WDT matrices because adding more WDT does not yield discernible effect. As shown in Table 3 when the number of WDT matrices changes from 15 to 19 on the subset of HODA dataset the recognition rate has much less growth.

Essentially using WDT enables the model to learn from its misclassifications, such that the final decision that is made according to the WDT method corrects the mistakes were made by base classifiers and DT. For example decision template corresponds to class 7, digit 6, is shown in Figure.6.a, it is obvious that all of base classifiers have about 80% recognition rate for samples of class 7. Thus, to take into consideration Table 2.b, we use e.g. 2 WDT matrices to learn even from misclassification. Figure.6.b, 6.c shows WDT matrices for samples of class 7 that are mistaken with class 3 and class 5, respectively.

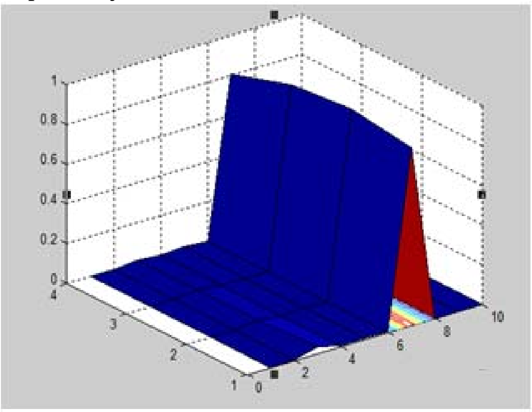

Figure 6.a. DT matrix of class 7 corresponds to digit 6. All of base classifiers have 80% recognition rate on the subset of HODA dataset.

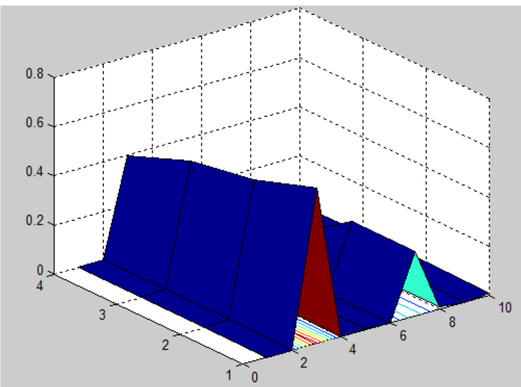

Figure 6.b. considering the confusion matrix of best classifier of subset of HODA dataset, WDT matrix of class 7 that are mistaken with class 3 is constructed.

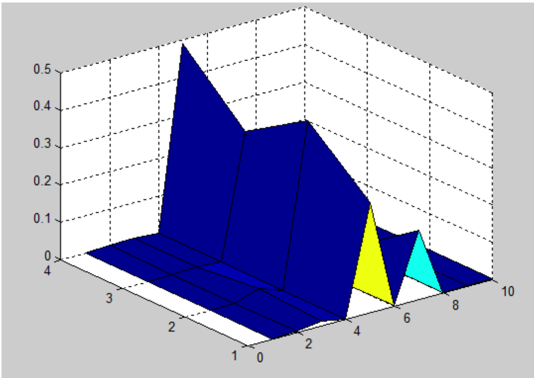

Figure 6.c. considering the confusion matrix of best classifier of subset of HODA dataset, WDT matrix of class 7 that are mistaken with class 5 is constructed.

We would like to compare the performance of the proposed method with respect to other static combination strategies in the literature of Persian handwritten digit recognition. These methods were implemented under the same condition as the previous experiment. The results are tabulated in Table 4. In each row various learning rates for base classifiers is applied. The highest recognition rate of each row is typed in bold. And maximum result in class indifferent and class conscious methods are underlined. Each result is the average of ten times testing the corresponding model. In WDT method, the best case i.e., 19 WDT matrices are used.

The left half section of the Table 4 deals with the class indifferent methods applied on all 4 base classifiers. It is clearly that the recognition rate of the proposed model is higher than for those of the other class indifferent methods. In fact, the performance of WDT is 2.7% and 1.65% better than DT and Stack Generalization methods, respectively, when the learning rate is 0.4.

In the entire right half of the Table 4 the results of class conscious methods are shown. The best results over the 4 combining rules are underlined.

As shown in Table 4, combining the results of base classifiers with class indifferent methods is far more effective than combining the results with class conscious methods. Clearly using all classifier outputs to calculate the final support for each class is more useful than other fusion methods that use only the support for that particular class to make their decision.

Also, even when base classifiers are not good; our proposed model has better result than other fusion methods and this shows robustness of our model which is a modified version of DT model.

The accuracy of the combinations in our experimental result in the second row of Table 4 are not very high compared to recognition rate on the subset of HODA dataset reported in first row of same Table. This experiment is performed, because it does not confer special attention on designing the individual first level classifiers and it is accomplished to showing partial views to the difference between result of first level base classifiers and result of fusion methods. Table 5 reveals these differences. In each row various values of learning rate for base classifiers is applied and in the first row the most excellent individual classifiers are employed in the ensemble; however, in the second row the base classifiers are weaker than of base classifiers in the first row. Values are the average (display only recognition rate on the test set) and standard deviation of ten times testing the corresponding model. Fourfold cross validation exhibits 400 epochs is adequate for the subset of HODA dataset.

Comparison between difference of recognition rate of base classifiers and combining methods exhibits that in utilizing optimized base classifiers the result of fusion methods are not much varies in contrast with using ordinary base classifiers.

5.2 Experiment on the Two Benchmarked Datasets

We also conducted experimentation on two datasets from the UCI repository. Information about these datasets is shown in Table 6.

Table 4. Comparison between recognition rate (%) of the proposed method and other static combination methods

	Class indifferent				Class conscious			
	Fusion Method	**DT**	**WDT**	**Stack generalization**	**MIN**	**MAX**	**Average**	**Product**
Subset of HODA Dataset	**Learning Rate=0.4**	95.46	**98.16**	96.51	94.32	94.91	94.53	94.89
	Learning Rate=0.1	94.16	**97.05**	95.26	93.47	93.97	93.52	93.93

Table 5. Recognition rates of base classifiers beside the best result of fusion methods on the HODA dataset.

Recognition Rate of Base Classifiers					Best result of fusion methods (%)	
Learning rate	Classifier 1 (25 hidden neurons)	Classifier 2 (30hidden neurons)	Classifier 3 (35hidden neurons)	Classifier 4 (40 hidden neurons)	Best result of class indifferent methods	Best result of class conscious methods
LR=0.4	94.11,0.63	93.98,0.54	94.17,1.46	94.06,0.87	98.16	94.91
LR=0.1	93.17,0.98	93.49,0.85	93.54,0.93	93.26,1.03	97.05	93.97

Table 6. summaries of two UCI datasets

Dataset	**Size**	**Attributes**	**Class**
Vehicle	946	18	4
Pima Indian Diabetes	768	8	2

To evaluate the generalization capability of our proposed method, WDT, we compared it with DT method and also with single MLP network. Four identical MLP networks, with different initial weight and hidden neurons, for both DT and WDT methods are used. The results are reported in Table 7.

Table 7. Recognition rate over two UCI datasets. The results are average of ten times testing

Technique	**Single MLP**	**DT**	**WDT**
Recognition Rate(%) on Vehicle Dataset	77.10	78.15	**80.52**
Recognition Rate(%) on Pima Indian Diabetes Dataset	72.46	73.68	**79.93**

Table 7 elucidates the ability of our utilized combining method versus stand alone MLP and DT methods in increasing the recognition rate. So these results affirm that our proposed method is a robust method in the literature of static combining methods.

6. Conclusion

In this paper, a new method for multiple classifiers systems is presented that is based on DT, such that in the training phase c DT matrices and q WDT matrices are constructed. In the test phase, to make decision on a test sample and its corresponding class, similarity between the sample and both DT and WDT matrices is measured. Experiments on our proposed model, WDT, with a real world dataset in Persian handwritten digit recognition task revealed the recognition rate of 95.46%, 98.16% for conventional DT and WDT, respectively. Comparison with other related fusion methods in the literature of static combining methods also demonstrated that WDT is a rich combining method for Persian handwritten digit recognition. Testing our proposed method on two benchmarked datasets presents efficiency and robustness of this method as a fusion method in classification systems.

Corresponding Author:
Dr. Mohammad Masoud Javidi
Department of Computer Science
Shahid Bahonar University of Kerman,
Kerman, Iran.
E-mail: javidi@mail.uk.ac.ir

References

1. Liu CL, Nakashima K, Sako H, Fujisawa H. Handwritten digit recognition: benchmarking of state-of the-art techniques. Pattern Recognition 2003; 36:2271-2285.
2. Trier OD, Jain AK, Taxt T. Feature extraction methods for character recognition a survey. Pattern Recognition 1996; 29(4): 641-662.
3. Ho TK, Hull JJ, Srihari SN. Decision combination in multiple classifier systems. IEEE Trans Pattern Anal Mach Intell 1994; 16(1):66-75.
4. Xu L , Krzyzak A , Suen CY. Associative switch for combining multiple classifiers. In: Int. Joint Conf. on Neural Networks 1991; 1:43-48.
5. Suen CY, Nadal C, Mai TA, Legault R, Lam L. Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts. In:Proc.IWFHR 1990; 131-143.
6. Liu CL, Suen CY. A new benchmark on the recognition of handwritten Bangla and Farsi numeral characters. Pattern Recognition 2009; 3287-3295.
7. Pan WM, Bui TD, Suen CY. Isolated Handwritten Farsi Numerals Recognition Using Sparse and Over- Complete Representations, icdar. 10th International Conference on Document Analysis and Recognition 2009; 586-590.
8. Borji A, Hamidi, Mahmoudi F. Robust Handwritten Character Recognition with Features Inspired by Visual Ventral Stream. Neural Processing Letters 2008; 28(2): 97-111.
9. Suen CY, Izadinia S, Sadri J, Solimanpour F. Farsi script recognition: a survey, in: Proceedings of the Summit on Arabic and Chinese Handwriting Recognition. University of Maryland, College Park, MD, 2006; 101-110.
10. Soltanzadeh H, Rahmati M. Recognition of Persian handwritten digits using image profiles of multiple orientations. Pattern Recognition Letters 2004; 25: 1569-1576.
11. Amin A. Off-line Arabic character recognition: the state of the art. Pattern Recognition 1998; 31: 517-530.
12. Ghaderi R. Arranging simple neural networks to solve complex classification problems. Ph.D. Thesis, Surrey University, 2000.
13. Kuncheva LI. Combining Pattern Classifiers: Methods and algorithms. Published by John Wiley Sons. Inc., 2004.
14. Xu L, Krzyzak A, Suen CY. Methods of combining multiple classifiers and their application to handwriting recognition. IEEE Trans. Systems Man Cybernet 1992; 22: 418-435.
15. Ng KC, Abramson B. Consensus diagnosis: a simulation study. IEEE Trans. Systems Man Cybernet 1992. 22: 916-928.
16. Kittler J, Hatef M, Duin RPW, Matas J. On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell 1998; 20 (3): 226-239.
17. Woods K, Kegelmeyer WP, Bowyer K. Combination of multiple classifiers using local accuracy estimates. IEEE Trans. Pattern Anal. Mach. Intell 1997. 19: 405-410.
18. Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE. Adaptive mixtures of local experts. Neural Comput 1991; 3: 79-87.
19. Alpaydin E, Jordan MI. Local linear perceptrons for classification. IEEE Trans. Neural Networks1996; 7(3): 788-792.
20. Haykin S. Neural Networks A Comprehensive Foundation. Second ed., Prentice-Hall 1998.
21. Kuncheva LI, Bezdek JC, Duin RPW. Decision Templates for Multiple Classifier Fusion: An Experimental Comparison. Pattern Recognition 2001; 34(2): 299-314,.
22. Ebrahimpour R, Sharifizadeh F. Persian Handwritten Digit Recognition with Classifier Fusion: Class Conscious versus Class Indifferent Approaches. World Academy of Science, Engineering and Technology 2009; 57: 552-559,.
23. Glucksman HA. Classification of mixed-font alphabets by characteristic loci. IEEE Comput. Conf 1967; 138-141,.
24. Ebrahimi A, Kabir E. A pictorial dictionary for printed Farsi sub words. Pattern Recognition Letters 2008; 29: 656-663.
25. Knoll AL. Expriments with characteristic loci for recognition of handprinted characters. IEEE Trans. Comput 1969; 18: 366-372.
26. Martinez A, Kak A. PCA versus LDA. IEEE Trans Pattern Anal Mach Intell 2001; 23(2): 228-233.
27. Aradhya VNM, Kumar GH, Noushath S. Multilingual OCR system for South Indian scripts and English documents: An approach based on Fourier transform and principal component analysis. Engineering Applications of Artificial Intelligence 2008; 21: 658-668.
28. Dietrich C, Palm G, Schwenker F. Decision templates for the classification of bioacoustic time series, Information Fusion 2003; 4:101-109.
29. Dietrich C, Schwenker F, Palm G. Classification of time series utilizing temporal and decision fusion. Second International Workshop on Multiple Classifier Systems 2001; Lecture Notes in Computer Science; 2096:378-387.
30. Kittler J, Balette M, Czyz J, Roli F, Vanderdorpe L. Decision level fusion of intramodal personal identity verification experts. 2nd International Workshop on Multiple Classifier Systems; Lecture Notes in Computer Science 2002; 2364: 314- 324.
31. Giacinto G, Roli F, Didaci L. Fusion of multiple classifiers for intrusion detection in computer networks. Pattern Recognition Letters 2003; 24:1795-1803.
32. Shipp CA, Kuncheva LI. Relationship between Combination methods and measures of diversity in combining classifiers. Information Fusion 2002. 3:135-148.
33. Khosravi H, Kabir E. Introducing a very large dataset of handwritten Farsi digits and a study on their varieties. Pattern Recognition Letters 2007; 28:1133-1141.

12/22/2011