Шеста национална конференција со меѓународно учество ЕТАИ'2003
Sixth National Conference With International Participation ETAI'2003
Охрид, Република МАКЕДОНИЈА – Ohrid, Republic of MACEDONIA
17–20. IX 2003

I4–2

COOPERATION OF SUPPORT VECTOR MACHINES FOR HANDWRITTEN DIGIT RECOGNITION TROUGH PARTITIONING OF THE FEATURE SET

Dejan Gorgevik[1], Dusan Cakmakov[2]

[1] University "Sv. Kiril i Metodij", Faculty of Electrical Eng., Department of Computer Science and Information Technology, Karpos II bb, POBox 574, 1000 Skopje, Macedonia, dejan@etf.ukim.edu.mk
[2] University "Sv. Kiril i Metodij", Faculty of Mechanical Eng., Department of Mathematics and Computer Science, Karpos II bb, POBox 464, 1000 Skopje, Macedonia, dusan@mf.ukim.edu.mk

Abstract – **In this paper, various cooperation schemes of SVM (Support Vector Machine) classifiers applied on two feature sets for handwritten digit recognition are examined. We start with a feature set composed of structural and statistical features and corresponding SVM classifier applied on the complete feature set. Later, we investigate the various partitions of the feature set as well as the advantages and weaknesses of various decision fusion schemes applied on SVM classifiers designed for partitioned feature sets. The obtained results show that it is difficult to exceed the recognition rate of a single SVM classifier applied straightforwardly on the complete feature set. Additionally, we show that the partitioning of the feature set according to feature nature (structural and statistical features) is not always the best way for designing classifier cooperation schemes. These results impose need of special feature selection procedures for optimal partitioning of the feature set for classifier cooperation schemes.**

Index terms – **classification, committee, features, rejection, reliability**

1. INTRODUCTION

The classical paradigm for character recognition is concentrated around two steps, feature extraction, where an appropriate representation of the pattern is developed, and classification, where decision rules for separating pattern classes are defined. Combining features of different nature and the corresponding classifiers has been shown to be a promising approach in handwritten recognition systems [1], [2], [3], [4], [5]. Data from more than one source that are processed separately can often be profitably re-combined to produce more concise, more complete and/or more accurate situation description.

In this paper, classification systems for handwritten digit recognition using two different feature families and SVM classifiers [6] are examined. Following widely used terminology, our feature families are referenced as structural and statistical feature sets [7]. We start with a SVM classifier applied on both feature families as one feature set. These results serve as a basis for future investigations. Further, we used two SVM classifiers that work on the structural and statistical feature families and examined their cooperation using statistical decision fusion. In our terminology, we partition the complete feature set according to the feature nature into statistical and structural feature sets. Different statistical cooperation schemes were examined and corresponding recognition results are presented. In order to improve the system reliability, we introduced rejection criteria as a part of the classifier cooperation schemes.

The presented results show that it is difficult to achieve the recognition rate of a single classifier applied on the feature set that includes both feature families by combining the individual classifier decisions. However, the classifier cooperation schemes offer better possibility for fine tuning of the recognition versus the reliability tradeoff and reduce both, the classifier complexity and the need for samples.

Additionally, we tackle the problem of usefulness of partitioning of the feature set according to the nature of the obtained features. The researchers in the area of classifier combining tacitly agree that partitioning of the features according to their nature is an acceptable approach for designing pattern recognition systems based on classifier cooperation. The basic supposition is that different features (by their origin) can be considered as reasonably independent. They "see" the same pattern from different points of view and consequently, corresponding individual classifier decisions can be profitably recombined to produce more accurate recognition. Our results show that statistical and structural features, widely used for designing pattern recognition systems based on classifier cooperation schemes is not always the best way to partition the feature set. This result imposes need of special feature selection procedures for optimal partitioning of the feature set in cases of classifier decision fusion applied on different feature families.

Our goals in this paper are to examine usefulness of our feature extraction and selection technique, to study different classifier cooperation schemes and to investigate usefulness of partitioning of the feature set according to the nature of the features rather than to compete with the recognition rates of other handwritten digit recognition systems [8].

2. THE SYSTEM ARCHITECTURE

The recognition system is constructed around a modular architecture of feature extraction and digit classification units. Preprocessed image is an input for the feature extraction module, which transfers the extracted features toward SVM classifiers (see Fig. 1).

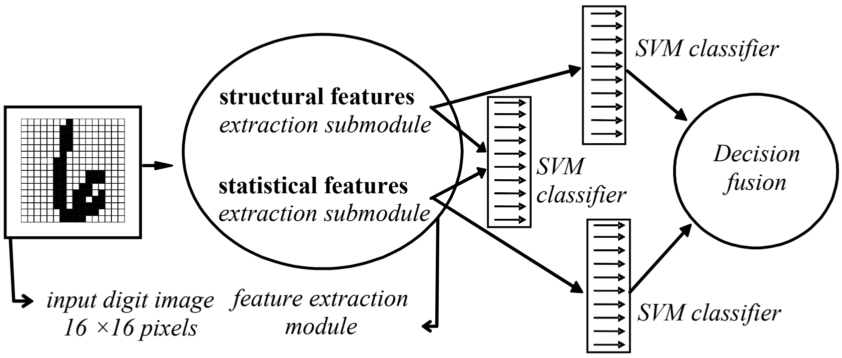


Fig. 1. The system architecture

From the digit images with resolution of 128x128 pixels, we have obtained 16x16 binary images on which the smoothing and centralizing preprocessing techniques have been applied. We have extracted 116 features that are classified as 54 structural and 62 statistical. The both feature families as one set are forwarded to the SVM classifier and obtained results are basis for future comparisons.

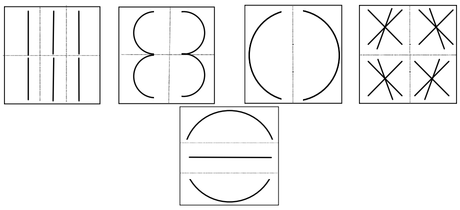
The structural and statistical feature sets are also forwarded to the separate SVM classifiers, and obtained results are combined using statistical cooperation schemes. On this level, rejection criteria are introduced and the corresponding system reliabilities are calculated.

3. THE FEATURE EXTRACTION MODULE

The structural feature set is a domain dependent set. Its nature and the techniques implemented for detection and extraction are strongly dependent of the nature of the objects to be recognized.

The first step in creating of the structural feature set is defining a reasonable set of elementary shape primitives for digit constructions. We have proposed 27 elementary primitives showed in Fig. 2. The digit image is searched for these primitives twice: firstly on the original digit image orientation, and secondly on the rotated digit image for 90°. So, the total number of primitives is 54, and that is the number of the elements in the structural feature set.

The detection and the extraction of the structural features are performed by dividing the image binary matrix into two, three, four and six subregions. The existing shape in each of those subregions is compared with the proposed primitives in the same sub-regions whose existence is expected.



6 4 2 12 3

Fig. 2. Image sub-regions and the elementary primitives

Thus, the structural feature is composed of 54 values of the calculated similarities [9] between the found shapes in the corresponding sub-regions and the corresponding elementary primitives.

The statistical feature set is composed of 62 features that give the pixel-based information presented by the densities of the lit pixels in various regions of the digit image. The first 54 statistical features are obtained from the projection histograms obtained by the vertical (16), horizontal (16) and two diagonal (22) projections (5 pixels left and right around the main diagonals). The last 8 features are obtained from the zone-pattern regions. This kind of features in different forms has been exploited in many pattern recognition systems [e.g. 10].

4. THE RECOGNITION RESULTS

The database for our experiments is an extraction of the NIST (National Institute of Standards and Technology) handwritten digit database. The total number of 23898 digit images is divided into two groups, 17952 images for the training phase and 5946 images for the test phase. The digits from the original database are rearranged so that digits in the test set belong to different writers from those in the learning set.

We have used a SVM classifier with Gaussian kernel. Because of the large number of samples, a more robust variation of SVM training software (Torch3) has been used [11].

The outputs of this classifier applied on our samples fall in [-8, 5] interval. Let us denote the classifier outputs in descending order by O_1, O_2, ..., O_{10} ($O_1 \geq O_2 \geq ... \geq O_{10}$). We have used a rejection criterion based on the top two classifier outputs. All the samples with highest value of classifier outputs that is smaller than a certain threshold T_1 ($O_1 < T_1$) or for which the difference between the top two classifier outputs is smaller than a certain threshold T_2 ($O_1 - O_2 < T_2$) are rejected. Varying these thresholds to obtain reliability of at least 99% we have obtained the results shown in Table 1.

Table 1. *Recognition rates on the structural, statistical and both feature families*

SVM (Gaussian kernel)	**Recog** (%)	**T_1, T_2**	**RRecog.** (%)	**Miscl.** (%)	**Reject.** (%)	**Reliab.** (%)
Statistical features	97.01	-0.533, 0.9988	92.40	0.92	6.68	99.01
Structural features	94.92	0.382, 1.4626	81.97	0.81	17.22	99.02
Structural + Statistical features	97.73	0.0056, 0.577	94.76	0.94	4.29	99.02

Recog. is the classifier recognition rate. RRecog., Miscl. and Reject. denote the recognition, misclassification and rejection rates for reliability of at least 99% provided by the rejection criterion using the corresponding values of T_1 and T_2. Reliab. denotes the reliability that is calculated as Reliab. = RRecog./(100%–Reject.). These results show that the statistical feature set has stronger discrimination power and provides better recognition rate. However, the recognition rate of the statistical feature set is more then 0.7 percent lower then the recognition rate of the classifier applied to the complete feature set.

4.1 Decision Fusion on Statistical and Structural Feature Sets

The classifier cooperation schemes are built around two SVM classifiers performing classification separately on structural and statistical feature families. In Table 2, the recognition rates using various statistical cooperation schemes are presented. We have used the same rejection criterion as in Table 1, and suitable values for T_1 and T_2 were chosen in order to achieve reliability of at least 99%.

The decision fusion methods: Product, Dempster Rule, Fuzzy Integral, and Decision Templates require possibilistic outputs. To map the original output values to [0, 1] interval we used the mapping $1/(1+e^{-x})$.

In order to make the final decision, first four cooperation schemes use the maximum of the sum, the maximum of the product, the maximum of the maximum and the maximum of the minimum of the corresponding pairs of the classifier outputs [12]. The Dempster rule considers the fuzziness of the classifier votes by giving less confidence to less certain votes [13]. The naive Bayes cooperation scheme uses the confusion matrices of member classifiers to estimate the certainty of the classifier decisions [13]. The

Borda count cooperation method is a generalization of the majority vote [14]. The fuzzy integration is based on searching for the maximal grade of agreement between the objective evidence (provided by the sorted classifier outputs for specific class) and the expectation (the fuzzy measure values of both classifiers) [15]. We have also used several decision template approaches described elsewhere [16]. The dynamic average approach uses dynamic weights that are adjusted to be proportional to the certainties of the respective classifier outputs rather than globally chosen weights as in generalized committee [17]. The generalized committee prediction is based on a weighted sum of the predictions of the member classifiers [18]. The modified generalized committee differs from the generalized committee only by the chosen weights that are different for every class output. In other words, we used a vector of adjusted weights rather than one common weight per classifier.

A few results in Table 2 deserve attention. The best recognition rates (Recog.>97.70%) are obtained by five of the cooperation schemes. Let us note that these results are almost identical to the recognition rate of the SVM that uses both feature families as one feature set (see Table 1).

On the other hand, the best recognition rates with reliability of 99% are provided by the schemes 21 (Generalized Committee) and 2 (Product). These results are noticeably better than the corresponding results shown in Table 1. Generally speaking, the classifier cooperation schemes with rejection criteria offer improved recognition rates in comparison to the classifier that utilizes the both feature families in one feature set.

4.2 Decision Fusion on Random Partitions of the Feature Set

To examine usefulness of grouping features according to their "nature" we performed a few random partitioning of the complete feature set in different relations between structural and statistical features.

Table 2. *Various statistical cooperation schemes and corresponding recognition rates*

#	Cooperation schemes	Recog. (%)	T_1, T_2	RRecog. (%)	Miscl. (%)	Reject. (%)	Reliab. (%)
1.	Average	97.70	0.453, 0.140	95.34	0.96	3.70	99.00
2.	Product	97.73	0.376, 0.039	95.39	0.96	3.65	99.01
3.	Max-Max	97.09	0.590, 0.197	93.51	0.92	5.57	99.02
4.	Min-Max	97.23	0.322, 0.122	94.80	0.92	4.27	99.03
5.	Borda count	97.70	0.907, 0.281	95.34	0.96	3.70	99.00
6.	Naive Bayes	97.24	0.919, 0.860	93.98	0.92	5.10	99.03
7.	Dempster	97.78	0.217, 0.081	94.82	0.94	4.24	99.02
8.	Fuzzy Integral	97.09	0.590, 0.197	93.51	0.92	5.57	99.02
9.	Decision Templates P1	97.66	0.518, 0.097	94.57	0.94	4.49	99.01
10.	Decision Templates P2	97.68	0.857, 0.033	94.90	0.92	4.17	99.03
11.	Decision Templates P3	97.61	0.805, 0.030	95.06	0.94	4.00	99.02
12.	Decision Templates P4	97.56	0.497, 0.107	94.82	0.92	4.25	99.03
13.	Decision Templates I1	97.66	0.645, 0.097	95.19	0.96	3.85	99.00
14.	Decision Templates I2	97.60	0.932, 0.016	94.38	0.92	4.69	99.03
15.	Decision Templates I3	97.56	0.827, 0.015	94.94	0.94	4.12	99.02
16.	Decision Templates I4	96.99	0.489, 0.183	93.53	0.92	5.55	99.02
17.	Decision Templates I5	97.09	0.352, 0.216	93.02	0.91	6.07	99.03
18.	Decision Templates C	97.12	0.577, 0.195	93.58	0.92	5.50	99.02
19.	Decision Templates E	97.71	0.967, 0.022	95.16	0.94	3.90	99.02
20.	Dynamic average	97.70	0.204, 0.083	95.24	0.94	3.82	99.02
21.	Generalized Committee	97.73	0.221, 0.073	95.41	0.94	3.65	99.02
22.	Modified Generalized Committee	97.76	0.216, 0.074	95.27	0.94	3.78	99.02

Ten pairs of randomly partitioned feature subsets were created by combining 25 randomly chosen features from the structural feature set and 29 randomly chosen features from the statistical feature set to form a mixed feature set containing 54 features. The remaining 29 features from the structural feature set and the remaining 33 features from the statistical feature set were combined in another feature set containing 62 features. By partitioning and joining the parts of the structural and statistical feature sets this way, we have obtained two feature sets with the same number of features as the original structural and statistical feature sets. The new feature subsets contain features from both feature families in relations that correspond to structural and statistical feature families 25:29 ≈ 29:33 ≈ 54:62. By this partitioning we were able to use the same system to conduct the experiments, i.e. to avoid most of disturbances that could be implied by the altered system.

For each pair of obtained feature subsets a pair of SVM classifiers was trained to perform classification over a single mixed feature set. The outputs of each pair of classifiers were combined using the same classifier cooperation schemes given in Table 2. In Table 3, average recognition rates over the ten random partitioning of the feature set are presented.

The obtained results show that mixing the features from the "weaker" structural and "stronger" statistical feature set, the performances of the classifier trained on these mixed feature sets perform more comparably with 0.65% average difference in recognition performance and 2.69% when applying rejection criteria for reliability of 99%. The slight advantage of the feature set that contains 62 features is expected, because of the larger number of features used to represent the sample.

The recognition results obtained by random partitioning of the feature set (Table 3) are not worse, but also not better than the recognition results obtained by partitioning of the feature set according to the feature nature (Table 2). Let us note that in Table 2 five recognition rates exceeds 97.70%, while in Table 3 there are only two. On the other hand, the best recognition rate comes from the modified generalized committee in Table 3 (97.82%). By applying rejection criteria in the cooperation schemes, the situation is quite similar in opposite way. Now, in Table 2 we have five recognition rates over 95.30%, while there are four in Table 3. The best result is obtained by the generalized committee in Table 2 (95.41%).

Table 3. *Average recognition rates of cooperation schemes on 10 random partitioning of the feature set*

#	Cooperation schemes	Recog. (%)	RRecog. (%)	Miscl. (%)	Reject. (%)	Reliab. (%)
	Set 1 (54 features)	96.02 (σ=0.25)	89.20 (σ=1.37)	0.89	9.91	99.00
	Set 2 (62 features)	96.67 (σ=0.18)	91.89 (σ=0.98)	0.91	7.20	99.00
1.	Average	97.67 (σ=0.10)	95.32 (σ=0.37)	0.94	3.74	99.00
2.	Product	97.69 (σ=0.10)	95.31 (σ=0.35)	0.95	3.75	99.01
3.	Max-Max	97.38 (σ=0.09)	94.51 (σ=0.62)	0.93	4.56	99.02
4.	Min-Max	97.36 (σ=0.14)	94.45 (σ=0.42)	0.94	4.61	99.03
5.	Borda count	97.67 (σ=0.10)	95.32 (σ=0.37)	0.94	3.74	99.00
6.	Naive Bayes	97.24 (σ=0.12)	74.60 (σ=15.15)	0.65	24.75	99.03
7.	Dempster	97.63 (σ=0.10)	95.10 (σ=0.45)	0.94	3.96	99.02
8.	Fuzzy Integral	97.38 (σ=0.09)	94.51 (σ=0.62)	0.93	4.56	99.02
9.	Decision Templates P1	97.61 (σ=0.08)	95.10 (σ=0.35)	0.94	3.96	99.01
10.	Decision Templates P2	97.63 (σ=0.07)	95.09 (σ=0.34)	0.94	3.97	99.03
11.	Decision Templates P3	97.66 (σ=0.08)	95.24 (σ=0.41)	0.94	3.83	99.02
12.	Decision Templates P4	97.43 (σ=0.10)	94.58 (σ=0.54)	0.94	4.49	99.03
13.	Decision Templates I1	97.60 (σ=0.08)	95.07 (σ=0.32)	0.93	4.00	99.00
14.	Decision Templates I2	97.57 (σ=0.09)	94.96 (σ=0.45)	0.94	4.10	99.03
15.	Decision Templates I3	97.64 (σ=0.11)	95.21 (σ=0.40)	0.94	3.85	99.02
16.	Decision Templates I4	97.13 (σ=0.12)	94.20 (σ=0.60)	0.93	4.88	99.02
17.	Decision Templates I5	97.38 (σ=0.09)	94.45 (σ=0.62)	0.93	4.63	99.03
18.	Decision Templates C	97.37 (σ=0.10)	94.44 (σ=0.62)	0.93	4.63	99.02
19.	Decision Templates E	97.63 (σ=0.10)	95.21 (σ=0.31)	0.94	3.86	99.02
20.	Dynamic average	97.65 (σ=0.10)	95.30 (σ=0.35)	0.95	3.75	99.02
21.	Generalized Committee	97.75 (σ=0.07)	95.34 (σ=0.31)	0.94	3.72	99.02
22.	Modified Generalized Committee	97.82 (σ=0.12)	95.34 (σ=0.38)	0.94	3.72	99.02

5. CONCLUSION

In this paper, we address some issues in designing high reliability system for hand-written digit recognition using SVM classifiers. We used two different feature families referenced as structural and statistical features. Decision level fusion is performed using statistical cooperation schemes. To improve the system reliability, we introduced rejection criteria in decision fusion schemes.

The presented results show that it is difficult to achieve the recognition rate of the single classifier applied on the complete feature set by statistical decision fusion applied on the individual classifier outputs. However, classifier cooperation schemes reduce classifier complexity, slightly improve recognition rates and enable fine tuning of the recognition versus the reliability tradeoff. Additionally, we compare decision fusions of the classifiers designed for different partitioning of the feature set including random partitioning and partitioning by feature nature and show that "natural" partitioning of the feature set according to feature nature is not the best way to apply classifier decision fusion.

6. REFERENCES

[1] F. Kimura, M. Shridar: Handwritten Numerical Recognition Based on Multiple Algorithms, *Pattern Recognition*, Vol. 24, No. 10, 1991, pp. 969–983.

[2] Y. S. Huang, C. Y. Suen: A Method of Combining Multiple Classifiers – A Neural Network Approach. *Proc. 12th Int. Conf. Pattern Recognition and Computer Vision*, Jerusalem, 1994, pp. 473–475.

[3] T. Suzuki, H. Nishida, Y. Nakajima, H. Yamagata, M. Tachikawa, G. Sato: A Handwritten Character Recognition System by Efficient Combination of Multiple Classifiers, *Int. Association for Pattern Recognition Workshop on Document Analysis Systems, World Scientific*, Singapore, 1995, pp. 169–187.

[4] S. Yamaguchi, T. Tsutsumida, F. Kimura, A. Iwata: Study on Multi-Expert Systems for Handprinted Numeral Recognition, In: Downton, A.C., Impedovo, S. (eds.): *Progress in Handwriting Recognition. World Scientific*, Singapore, 1997, pp. 285–292.

[5] J. Dahmen, D. Keysers, H. Ney: Combined Classification of Handwritten Digits Using the 'Virtual Test Sample Method'. In: Kittler, J., Roli, F. (eds.): Multiple Classifier Systems, 2nd Int. Workshop, MCS 2001 Cambridge, UK, *Lecture Notes in Computer Science Vol. 2096*, Springer-Verlag, 2001, pp. 109–118.

[6] C. Burges: A Tutorial on Support Vector Machines for Pattern Recognition, *Knowledge Discovery and Data Mining*, Vol. 2, 1998, pp. 1–47.

[7] V. Radevski, Y. Bennani: Reliability Control in Committee Classifier Environment, *Int. Joint Conference on Neural Networks*, IJCNN 11, Como, Italy, Vol. III., 2000, pp. 561–565.

[8] Y. LeCun, L. D. Jackel, L. Bottou, A. Brunot, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Muller, E. Sackinger, P. Simard, V. Vapnik: Comparison of Learning Algorithms for Handwritten Digit Recognition. In: Fogelman, F., Gallinari, P. (eds.): *Int. Conf. Artificial Neural Networks*, Paris, 1995, pp. 53–60.

[9] D. Cakmakov: Curve Matching Using Turning Functions. *Proc. of the Int. Conf. on Signal and Image Processing SIP'98*, Las Vegas, 1998, pp. 588–592.

[10] G. Burel, I. Pottier, J. Y. Catros: Recognition of Handwritten Digits by Image Processing and Neural Network, *Proc. Of the Int. Joint Conf. on Neural Networks*, IJCNN 3, 1992, pp. 666–671.

[11] R. Collobert, S. Bengio, and J. Mariéthoz. Torch: a modular machine learning software library. Technical Report IDIAP-RR 02–46, IDIAP, 2002.

[12] J. Kittler, M. Hatef, R. P. W. Duin, J. Matas: On Combining Classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, 1998, pp. 226–239.

[13] L. Xu, A. Krzyzak, C. Y. Suen: Methods of combining multiple classifiers and their application to handwritten recognition, *IEEE Transactions on System, Man and Cybernetics*, Vol. 22, 1992, pp. 418–435.

[14] T. K. Ho, J. J. Hull, S. N. Srihari: Decision Combination in Multiple Classifier Systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, 1994, pp. 66–75.

[15] S. B. Cho, J. H. Kim: Combining Multiple Neural Networks by Fuzzy Integral and Robust Classification, *IEEE Transactions on System, Man and Cybernetics*, Vol. 20, No. 3, 1995, pp. 380–384.

[16] L. I. Kuncheva, J. C. Bezdek, P. W. Duin: Decision Templates for Multiple Classifier Fusion: An Experimental Comparison, *Pattern Recognition*, Vol. 34, No. 2, 2001, pp. 299–314.

[17] A. D. Jimenez, N. Walsh: Dynamically weighted ensemble neural networks for classification, Proceedings of the 1998 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK 1998.

[18] C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995, pp. 364–369.