Combining SVM Classifiers for Handwritten Digit Recognition

Dejan Gorgevik
Ss. Cyril and Methodius University, Faculty of Electrical Eng., Department of Computer Science and Engineering, Skopje, Macedonia
dejan@etf.ukim.edu.mk

Dusan Cakmakov
Ss. Cyril and Methodius University, Faculty of Mechanical Eng., Department of Mathematics and Computer Science, Skopje, Macedonia
dusan@ereb.mf.ukim.edu.mk

Abstract

In this paper, we investigate the advantages and weaknesses of various decision fusion schemes using statistical and rule-based reasoning. The cooperation schemes are applied on two SVM (Support Vector Machine) classifiers performing classification task on two feature families referenced as structural and statistical features. The obtained results show that it is difficult to exceed the recognition rate of a single classifier applied straightforwardly on both feature families as one set. The rule based cooperation schemes enable an easy and efficient implementation of various rejection criteria. On the other hand, the statistical cooperation schemes provide higher recognition rates and offer possibility for fine-tuning of the recognition versus the reliability tradeoff.

1. Introduction

Combining features of different nature and the corresponding classifiers has been shown to be a promising approach in many pattern recognition applications. 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, we discuss classification systems for handwritten digit recognition using two different feature families and SVM classifiers [1]. Our feature families are referenced as structural and statistical feature sets [2], and they differ (especially structural features) from the feature sets with the same reference used in other systems for handwritten character recognition [3][4]. We started with a SVM classifier applied on both feature families as one set. Further, we examined two SVM classifiers that worked separately on the different feature families for the same digit image. We analyzed the possibility of their cooperation using various statistical and rule-based cooperation schemes. In order to improve the system reliability, we introduced rejection criteria as a part of the classifier decision fusion. Our aim was not to compete with the recognition rates of the other handwritten digit recognition systems [5], but to compare the qualities of different feature families, corresponding SVM classifiers and their combination based on different statistical and rule-based decision fusion.

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 (Figure 1).

From the digit images with resolution of 128×128 pixels, we have obtained 16×16 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.

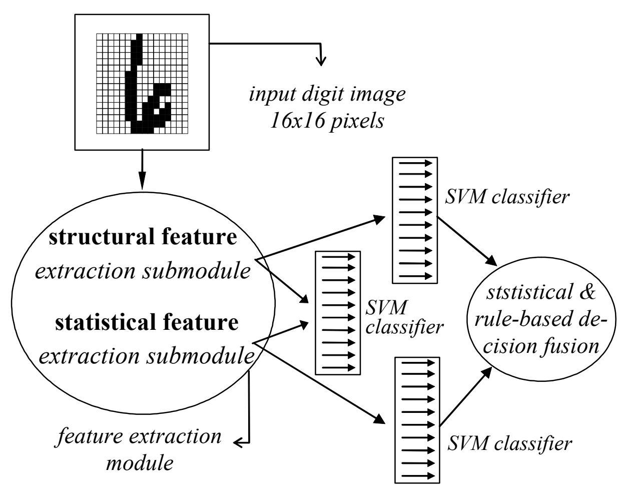

Figure 1. The system architecture

The structural and statistical feature sets are also forwarded to the separate SVM classifiers, and obtained

1051-4651/02 $17.00 (c) 2002 IEEE

results are combined using statistical and rule-based reasoning. On this level, rejection criteria are introduced and the corresponding system reliabilities are calculated.

3. The data base and feature extraction

The database for our experiments is an extraction of the NIST (National Institute of Standards and Technology) segmented handwritten digit database. The digit images are in 128×128 gray level pixels presented with real numbers in [-1, 1] interval. 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 in order that digits in the test set belong to different writers from those in the learning set.

To create the structural feature set we have defined a set of elementary shape primitives for digit constructions. We have proposed 27 elementary primitives shown in Figure 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.

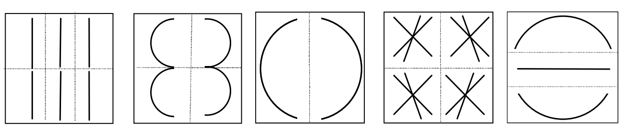

Figure 2. Image sub-regions and elementary primitives

The existing shape in each of those sub-regions is compared with the referent, idealized primitives in the same sub-regions whose existence is expected. The similarity measure between the found shape and the primitive is based on differences of changes of angles along both shapes, normalized to take values between 0 and 1. This similarity measure is a simplified variation of the curve matching technique described in [6].

The statistical feature set is composed of 62 features that give the pixel-based information in the terms of density of the lit pixels in various digit image regions. The first 54 statistical features are obtained from the projection histograms issued from the vertical (16), horizontal (16) and two diagonal (22) projections (with 5 pixels left and right around the main diagonals). The last 8 features are obtained from the zone-pattern regions shown in Figure 3.

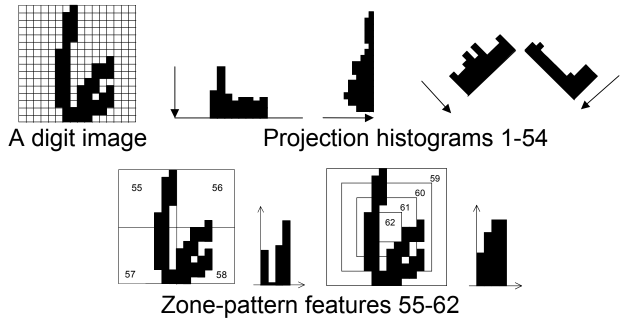

Figure 3. Projection histograms and zone-pattern features

Each of the numerical values of the 62 statistical features represents the filled up percentage of the projection histograms. So, the statistical features have values between 0 and 1.

4. The Recognition Result

We have used a SVM classifier with Gaussian kernel. The outputs of this classifier applied on our samples fall in [-8, 5] interval. Let us denote the classifier outputs in descending order by $O_1, O_2, ..., O_{10}$ ($O_1 \geq O_2 \geq ... \geq O_{10}$). The rejection criterion is 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.

Recog. is the classifier recognition rate. RRecog., Miscl. and Rejec. denote the recognition, misclassification and rejection rates for reliability of at least 99%. Reliab. denotes the reliability that is calculated as Reliability = Recognition/(100% – Rejection). The results show that the statistical feature set has stronger discrimination power and provides better recognition rates. However, the recognition rate on the statistical feature set is more than 0.7 percent lower than the

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

SVM with Gaussian kernel	Recog.	T_1,T_2	RRecog.	Miscl.	Rejec.	Reliab.
Statistical features	96.80% (5756)	0.1 0.96	92.23% (5484)	0.92% (55)	6.48% (407)	99.01%
Structural features	94.92% (5644)	0.4 1.5	81.48% (4845)	0.81% (48)	17.71% (1053)	99.02%
Structural + Statistical features	97.53% (5799)	-0.4 0.72	94.80% (5637)	0.96% (57)	4.24% (252)	99.00%

recognition rate of the classifier applied to the complete feature set.

4.1. The Statistical Decision Fusion

The statistical decision fusion is built around two SVM classifiers performing classification separately on the structural and statistical feature sets. In Table 2, the recognition rates using various statistical cooperation schemes are presented. We 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 [7]. The Dempster rule considers the fuzziness of the classifier votes by giving less confidence to less certain votes [8]. The naive Bayes cooperation scheme uses the confusion matrices of member classifiers to estimate the certainty of the classifier decisions [8]. The Borda count cooperation method is a generalization of the majority vote [9]. The fuzzy integration is based on searching for the maximal grade of agreement between the objective evidence (provided by the sorted classifier outputs for class *i*) and the expectation (the fuzzy measure values of both classifiers) [10]. We have also used one of the decision templates approaches described elsewhere [11]. The generalized committee prediction is based on a weighted combination of the predictions of the member classifiers [12].

A few results in Table 2 deserve attention. The best recognition rates (>97.7%) are obtained by five of the cooperation schemes. Let us note that these results are about 0.2% higher than the recognition rate of the SVM that uses both feature families as one feature set (Table 1). The best recognition rates with reliability of 99% are provided by the schemes 10 (Generalized Committee) and 2 (Product). These results are also noticeably better than the corresponding results shown in Table 1. Generally speaking, the statistical cooperation schemes slightly improved recognition rates and reliabilities in comparison to the classifier that utilizes simple integration of the both feature families in one feature set.

4.2. The Rule-Based Decision Fusion

Let us denote by *a1, a2* and *a3* the first, the second and the third choice of the structural feature classifier, and by *b1, b2* and *b3* the first, the second and the third choice of the statistical feature classifier for a given pattern. Our experiments showed that the inclusion of additional choices (after the third) provides insignificant recognition rate improvement. The results of classifier outputs based on various rule-based cooperation schemes are evaluated and given in Table 3.

Four results in Table 3 deserve attention. Best reliability is obtained by rule 1 (consensus) but the recognition rate is relatively weak. A good compromise is provided by rules 3 and 6, where we choose the first decision *b1* of the statistical feature classifier as a final decision *c*, if it is among the "top two" decisions (*a1*, *a2*) in the rule 3 and among the "top three" decisions (*a1*, *a2*, *a3*) in the rule 6 of the structural feature classifier. It seems that in this case the structural feature classifier gives a safety rule for the right decision. The reliabilities of 98.41% and 97.92% by recognition rates of 95.80% and 96.45% are noticeable results, better than some previous attempts using the same feature sets [2].

On the other hand, best recognition rate is provided by the relatively complex rule 8. Unfortunately, this rule produces high misclassification rate that results in lower reliability. Let us notice that the recognition rates achieved by rule-based cooperation schemes is still about 0.3% lower than the recognition rate of the SVM that uses both feature families as one feature set (Table 1). They are also noticeably lower than the recognition rates of the statistical cooperation schemes (Table 2).

Table 2. Various statistical cooperation schemes and corresponding recognition rates

#	Cooperation schemes	Recog.	T_1, T_2	RRecog.	Miscl.	Rejec.	Reliab.
1.	Average	**97.71%**	-0.05, 0.65	94.87%	0.96%	4.17%	99.00%
2.	Product	**97.70%**	0.15, 0.13	**95.16%**	0.96%	3.88%	99.00%
3.	Max-Max	97.07%	0.4, 0.87	93.41%	0.94%	5.65%	99.00%
4.	Min-Max	97.29%	-0.3, 0.09	93.47%	0.94%	5.58%	99.00%
5.	Dempster	**97.73%**	-0.257, 0.025	94.95%	0.96%	4.09%	99.00%
6.	Naive Bayes	96.92%	0.8, 0.865	93.74%	0.94%	5.31%	99.01%
7.	Borda count	96.80%	18, 2	93.12%	0.79%	6.09%	99.16%
8.	Fuzzy Integral	97.07%	0.585, 0.2	93.58%	0.94%	5.48%	99.00%
9.	Decision templates	**97.70%**	0.88, 0.034	94.69%	0.96%	4.36%	99.00%
10.	Generalized Committee	**97.78%**	0.514, 0.05	**95.34%**	0.96%	3.70%	99.00%

Table 3. Various rule-based cooperation schemes and corresponding recognition rates

#	Rule-based cooperation schemes	Recog.	Miscl.	Rejec.	Reliab.
1.	**if** *a1=b1* **then** *c=a1* **else** REJECT	93.12%	0.79%	6.09%	**99.16%**
2.	**if** *a1=b1* **or** *a1=b2* **then** *c=a1* **else** REJECT	94.48%	2.56%	2.96%	97.37%
3.	**if** *b1=a1* **or** *b1=a2* **then** *c=b1* **else** REJECT	**95.80%**	1.55%	2.66%	**98.41%**
4.	**if** *a1=b1* **or** *a1=b2* **then** *c=a1* **elseif** *b1=a2* **then** *c=b1* **else** REJECT	95.90%	2.79%	1.31%	97.17%
5.	**if** *a1=b1* **or** *a1=b2* **or** *a1=b3* **then** *c=a1* **else** REJECT	94.77%	3.50%	1.73%	96.44%
6.	**if** *b1=a1* **or** *b1=a2* **or** *b1=a3* **then** *c=b1* **else** REJECT	**96.45%**	2.05%	1.50%	**97.92%**
7.	**if** *a1=b1* **or** *a1=b2* **or** *a1=b3* **then** *c=a1* **elseif** *b1=a2* **or** *b1=a3* **then** *c=b1* **else** REJECT	95.88%	3.73%	0.39%	96.25%
8.	**if** *b1=a1* **or** *b1=a2* **or** *b1=a3* **then** *c=b1* **elseif** *a1=b2* **or** *a1=b3* **then** *c=a1* **else** REJECT	**97.24%**	2.37%	0.39%	97.62%

5. Conclusion

In this paper, we addressed 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 and rule-based cooperation schemes. To improve the system reliability, we introduced rejection criteria in our decision fusion 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 using statistical or rule based decision fusion. However, the statistical cooperation schemes slightly improve recognition rates and offer possibility for fine-tuning of the recognition versus the reliability tradeoff. The rule-based cooperation schemes enable an easy implementation of rejection criteria. Additionally, the cooperation of separate classifiers designed for separate feature families reduces classifier complexity and offers better possibilities to understand the role of the features in the recognition process.

References

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

[2] Radevski V. and Bennani Y., "Reliability control in committee classifier environment", *Int. joint conference on neural networks, IJCNN 11*, Como, Italy, Vol. III, 2000, pp. 561-565.

[3] Duerr B., Haettich W., Tropf H., and Winkler G., "A combination of statistical and syntactical pattern recognition applied to classification of unconstrained handwritten numerals", *Pattern Recognition*, Vol. 12, 1980, pp. 189-199.

[4] Heutte L., Moreau J. V., Paquet T., Lecourtier Y., and Olivier C., "Combining structural and statistical features for the recognition of handwritten characters", *Proc. of the 13th Int. Conf. on Pattern Recognition*, 1996, pp. B74.4.

[5] LeCun Y., Jackel L. D., Bottou L., Brunot A., Cortes C., Denker J. S., Drucker H., Guyon I., Muller U. A., Sackinger E., Simard P., and Vapnik V., "Comparison of learning algorithms for handwritten digit recognition", In F. Fogelman and P. Gallinari, editors, *International Conference on Artificial Neural Networks*, Paris, 1995, pp. 53-60.

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

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

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

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

[10] Cho, S.B. and Kim, J.H., "Combining multiple neural networks by fuzzy integral and robust classification", *IEEE Transactions on System, Man and Cybernetics*, Vol. 20, No. 2, 1995, pp. 380-384.

[11] Kuncheva, L.I., Bezdek, J.C., and Duin, P.W., "Decision templates for multiple classifier fusion: an experimental comparison", *Pattern Recognition*, Vol. 34, No. 2, 2001, pp. 299-314.

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

1051-4651/02 $17.00 (c) 2002 IEEE