Handwritten Digit Recognition Using Classifier Cooperation Schemes

Dusan Cakmakov[1], Dejan Gorgevik[2]

[1] 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

[2] University “Sv. Kiril i Metodij”, Faculty of Electrical Eng., Department of Computer and Information Technology, Karpos II bb, POBox 574, 1000 Skopje, Macedonia
dejan@etf.ukim.edu.mk

Abstract. Recent results in pattern recognition applications have shown that SVMs (Support Vector Machines) often have superior recognition rates in comparison to other classification methods. In this paper, the cooperation of three SVM classifiers for handwritten digit recognition, each using different feature family is examined. We investigate the advantages and weaknesses of various cooperation schemes based on classifier decision fusion using statistical reasoning. Although most of the used schemes are variations and adaptations of existing ones, such an extensive number of cooperation schemes have not been presented in the literature until now. The obtained results show that it is difficult to exceed the recognition rate of a single, well-tuned SVM classifier applied straightforwardly on all feature families as a single set. However, the classifier cooperation reduces the classifier complexity and need for samples, decreases classifier training time and sometimes improves the classifier performance.

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 recombined to produce more concise, more complete and/or more accurate situation description. In this paper, we discuss classification systems for handwritten digit recognition using three different feature families and SVM classifiers [1]. We start with a SVM classifier applied on all feature families as one set. Further, we used three SVM classifiers that work on the different feature families for the same digit image. As the feature sets “see” the same digit image from different points of view, we examined the possibility of decision fusion using statistical cooperation schemes. An extensive number of cooperation schemes were examined and corresponding recognition results are presented. Our aim was not to compete with the recognition rates of the other handwritten digit recognition systems e.g. [2], [3], but to compare the qualities of different feature families, corresponding SVM classifiers and their combination based on different classifier decision fusion.

The presented results show that it is difficult to achieve the recognition rate of a single optimized SVM classifier applied on the feature set that includes all feature families. However, the cooperation of individual classifiers designed for separate feature families reduce the classifier complexity and need for training samples, offering better opportunity to understand the role of the features in the recognition process.

2 The System Architecture

The recognition system is constructed around a modular architecture of feature extraction and digit classification units. The preprocessed isolated digit images are input for the feature extraction module, that transfers the extracted features toward SVM classifiers (see Fig. 1).

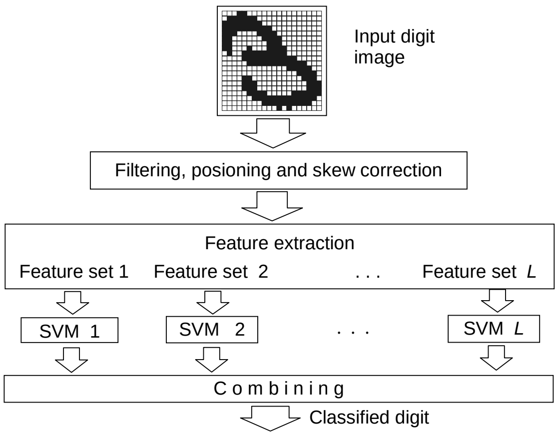

Fig. 1. The system architecture

Each image is centered in a square bounding box, and then slant correction is performed. The slant angle is estimated as the inclination of the line connecting the gravity centers of the top 25% part and the bottom 25% part of the image. Then a subpixel precision shear transformation is performed in order to remove the estimated inclination.

3 Feature Extraction

Three feature families were extracted from each digit image:

- contour profiles,
- ring-zones and
- Kirsch features.

Feature extraction was performed on the original unscaled images, after the slant correction.

The first feature family (FS1) is composed of 30 contour profile features (see Fig. 2). The image is scanned from left to right, top to bottom, right to left and bottom to top, respectively. The distance from the corresponding edge of the image to the first black pixel which the scanning line intersects, represent the contour profile feature on the first level. The distance to the first black pixel in the second black pixel run represent the contour profile features on the second level. Since not all of the character images were of the same size, the profile vectors were linearly rescaled in order to obtain 6 features from the left and right contour profiles and 5 features from the upper and lower profiles on the first level of the digit image. Finally, 4 features were extracted from the upper and the lower contour profiles of the second level.

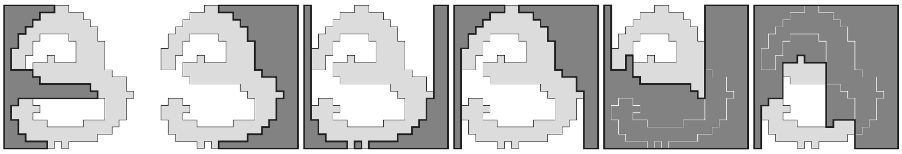

Fig. 2. Contour profiles of first and second level

The second group of 44 features (FS2) are extracted as pixel counts in rings zones around the gravity center of the image (see Fig. 3). We have used three rings, each divided in different number of equal zones. The outermost ring has a radius *r* equal to the distance from the gravity center to the furthest black pixel of the image. The first ring with radius $0.2 \cdot r$ provides 4 features and the second ring with radius $0.5 \cdot r$ provides 24 features. The last 16 features are provided from the outermost ring.

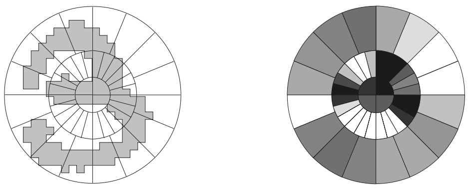

Fig. 3. Ring-zone features

The last group of 72 features (FS3) use Kirsch operator [4] to detect local directional information of the edges of the input pattern. Compared with chain code which also describes the edge direction, Kirsch edge detection is more robust even under noisy conditions.

The first black pixel which the scanning line intersects forms the first outermost periphery. The second black pixel which is the starting point of the second black pixel run forms the second outermost periphery (see Fig. 4). When the image is scanned in horizontal direction, the vertical and both diagonal Kirsch features are extracted at the outermost periphery. When the image is scanned in vertical direction, the horizontal and both diagonal Kirsch features are extracted at the outermost and second outermost periphery. This way, 3 Kirsch directional features are provided for each periphery pixel. The feature vectors are again linearly rescaled to 15 features coming from the left and right periphery each, 12 features coming from the first outermost top and bottom periphery each, and 9 features coming from the second outermost top and bottom peripheries.

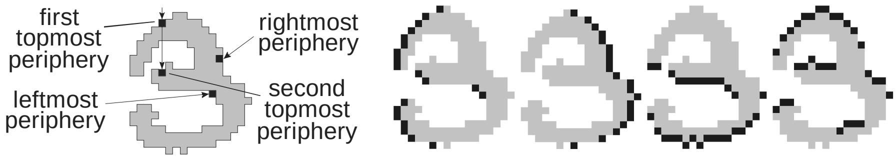

Fig. 4. Kirsch features

Kirsch feature extraction is performed on the grayscale digit images using subpixel precision. All parameters including the number of features by projection, the radiuses of rings for zone-pattern regions and the number of features coming from the outermost peripheries for Kirsch features are carefully chosen after several iterations using observations about their discriminative power. The features were preprocessed for zero mean and unit variance.

4 The Recognition Results

Our experiments were performed on an extract of the well-known NIST (National Institute of Standards and Technology) handwritten digit database. This database consists of 7 partitions denoted as: hsf_0, …, hsf_4, hsf_6 and hsf_7. Digit images from the hsf_0 partition were used for classifier training while the tuning of classifier parameters (kernel width σ and penalty C) was performed using the hsf_1 partition for validation. The final recognition rates were estimated on most difficult partition hsf_4. So, the samples in the test set belong to different writers from those in the learning set. In Fig. 5 a fragment of the NIST database is given.

Fig. 5. A fragment from the NIST database

We used SVMs with Gaussian kernel because it provided better recognition rates than linear, polynomial or sigmoidal kernel. Because of the large number of samples we have used SVMTorch that is a more robust variation of SVM training software library [5].

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

The recognition rates of different classifier cooperation schemes applied on the above described 3 feature sets: FS1, FS2 and FS3 are given in Table 1. In the second column the corresponding cooperation scheme is given, followed by the recognition rate and the rank of the cooperation scheme when combining classifiers are trained using 1000, 2000, 5000, 10000, 30000 and all 53449 available samples.

The first 3 rows show recognition rates of each feature set individually. The row a) gives the recognition results of a single optimized SVM classifier applied on the three feature sets as a whole. The next row b) gives the recognition rate of a hypothetical cooperation scheme that knows to choose the right class if it is predicted by at least one of the member classifiers. This is the theoretical upper bound of the recognition rates achievable by classifier decision fusion.

The cooperation schemes 1-4 are voting schemes including variations of the Borda count that is a generalization of the majority vote [6]. The 5-12 cooperation schemes use various averages, the maximum of the sum, product, maximum and the minimum of the corresponding pairs of the classifier outputs respectively to make the final decision [7]. The Dempster Rule [8] and many variations [9] are given in rows 13-35. The naive Bayes cooperation scheme given in rows 36-37 uses the confusion matrices of member classifiers to estimate the certainty of the classifier decisions [10]. The fuzzy integration 38-39 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) [11]. We have also used a variety of decision templates schemes 40-62 described elsewhere [12], [9]. The generalized committee prediction and its variations 63-67 are based on a weighted combination of the predictions of the member classifiers [13]. Cooperation scheme 76 uses linear regression to make decision fusion. In the cooperation scheme 69 the 4 individual SVM outputs (40 features) are input to another SVM classifier. This kind of cooperation is also known as classification task [8].

Table 1 shows that the cooperation 69 (svmcmb) has unbeatable recognition rate in all cases. However, this method is most complex because it needs additional classifier and additional samples for its training. Because of additional number of samples used in the training, this method sometimes outperforms even the "oracle method".

Increasing the number of training samples, indeed increase recognition rates of individual classifiers and their cooperation. On the other hand, increasing recognition rates of individual classifiers increase their correlation that reduces the possibility for improvement of cooperation recognition rates.

Voting cooperation schemes (1-4) are among worst because they use most limited information of member classifiers, ignoring useful information about second choices, reliability of the choice, distribution of the choices for different classes, etc.

The simplest cooperation schemes (5-12) as we expected, have average recognition rates and should be used in not demanding applications.

Table 1. Recognition rates (%) of combining SVM classifiers for 3 feature families (FS1, FS2, FS3) and different sizes of learning set (1000, 2000, 5000, 10000, 30000 and all 53449 samples). R stands for rank

	cooperation	1000	R	2000	R	5000	R	10000	R	30000	R	all	R
	FS1	87.3853		89.9857		91.5476		93.1931		94.8829		95.3398	
	FS2	86.5123		89.6003		90.3557		92.4087		94.7840		95.1454	
	FS3	89.9226		91.5203		92.4240		94.1002		95.7133		96.1003	
a)	Single SVM	92.7514		93.9911		94.2758		95.9076		97.1899		97.3843	
b)	oracle	95.4387		96.3629		96.6664		97.5105		98.2505		98.4909	
1	vote	90.6012	64	92.4974	66	93.2493	67	94.7925	66	96.3084	60	96.6698	51
2	borda	91.0957	54	92.9850	55	93.6944	59	95.2017	49	96.5573	44	96.9597	6
3	bks	91.9500	27	93.1692	44	94.1445	8	94.9698	62	96.1532	65	96.4038	65
4	bksv	92.1120	14	93.3329	20	94.2861	5	95.1915	50	96.2964	61	96.5727	59
5	avg	91.5646	35	93.3585	18	94.0320	11	95.3944	9	96.6681	23	96.9461	8
6	prod	91.1264	53	92.9646	57	93.8581	43	95.3296	19	96.7670	7	96.9086	15
7	harm	90.9218	59	92.6645	65	93.6910	60	95.2171	46	96.6562	25	96.8182	40
8	cprod	90.5296	66	92.7889	62	93.7643	56	95.1386	55	96.6289	34	96.8898	24
9	maxmax	89.9567	68	92.3678	67	93.3226	65	94.6151	67	96.2402	62	96.5334	62
10	minmax	90.5143	67	92.1768	68	93.3209	66	94.9101	64	96.4942	49	96.5880	54
11	med	91.0804	55	93.0805	51	93.6910	60	95.1318	56	96.5692	41	96.8864	26
12	davg	91.5578	36	93.3636	17	94.0166	13	95.4183	7	96.6494	28	96.9222	12
13	demp	91.9278	28	93.1453	48	93.8853	40	95.1761	53	96.5795	38	96.8455	32
14	dempp1	91.3191	45	93.2050	40	93.8785	41	95.2921	29	96.6426	29	96.8352	36
15	dempp2	91.2441	49	93.2272	37	93.9075	35	95.2733	33	96.6698	22	96.8932	22
16	dempp3	90.6029	63	92.9066	60	93.8103	53	95.2648	34	96.6920	19	96.9512	7
17	dempp4	91.3344	44	93.2374	35	94.0013	16	95.3245	20	96.7500	9	96.9086	15
18	dempi1	91.3651	42	93.2561	34	93.9399	31	95.3552	15	96.7040	17	96.8949	21
19	dempi2	91.1793	51	93.1658	45	93.8188	52	95.2034	48	96.5914	37	96.7619	45
20	dempc	90.7581	60	93.0362	54	93.8103	53	95.2921	29	96.7398	11	96.9324	9
21	dempmk	91.2219	50	93.1914	42	93.8905	38	95.2409	44	96.6409	30	96.8728	27
22	dempch	91.1673	52	93.1334	49	93.9570	28	95.2648	34	96.7346	12	96.9239	11
23	dempas	92.1785	11	93.3005	26	93.9007	36	95.3688	12	96.7739	4	97.0143	4
24	dempchi	91.3992	41	93.2647	30	93.8888	39	95.3006	26	96.6716	21	96.8438	33
25	dempchi2	91.4777	40	93.3482	19	93.9979	18	95.3620	14	96.7142	14	96.8983	20
26	dempbc	91.2833	46	93.2050	40	93.8734	42	95.3023	25	96.6528	27	96.8335	38
27	dempht	91.6175	33	93.4062	16	93.9859	22	95.3910	10	96.6835	20	96.8932	22
28	dempchr	92.2075	8	93.4505	11	93.9467	29	95.3688	12	96.7517	8	97.0313	3
29	dempchr2	91.6448	32	93.4386	12	94.0814	10	95.4115	8	96.7193	13	96.9069	17
30	dempjac	91.2458	48	93.1538	47	93.8257	51	95.2478	39	96.6340	31	96.8216	39
31	dempper	91.3617	43	93.2306	36	93.8922	37	95.3211	21	96.6630	24	96.8352	36
32	dempse	89.8339	69	93.5869	8	93.9996	17	95.1830	51	96.4038	56	96.6903	49
33	dempfr	92.0148	20	93.7080	7	93.9945	20	95.1778	52	96.3783	58	96.5692	60
34	dempm	91.6005	34	93.1555	46	93.8581	43	94.9937	61	96.5744	40	96.7466	46
35	dempmc	92.4019	5	93.936	5	94.1002	9	95.2085	47	96.2350	63	96.4107	64

Table 1. (continues)

	cooperation	1000	R	2000	R	5000	R	10000	R	30000	R	all	R
36	pprod	91.2594	47	93.0498	53	93.9126	34	95.3705	11	96.7432	10	96.9188	13
37	bayes	91.5322	38	93.1231	50	93.7387	58	95.0789	59	96.3646	59	96.7892	41
38	fi	90.6797	61	92.9134	59	93.5187	62	94.9647	63	96.4226	51	96.6698	51
39	fic	90.5688	65	92.8145	61	93.4488	63	94.9084	65	96.4840	50	96.7142	47
40	dtp1	92.1001	15	93.3141	21	93.8478	45	95.2426	40	96.4158	52	96.5863	55
41	dtp2	92.1580	12	93.4096	14	93.9314	32	95.2614	36	96.5215	46	96.7841	42
42	dtp3	91.7215	30	93.4215	13	94.0303	12	95.3501	18	96.5761	39	96.8728	27
43	dti1	92.0745	19	93.4829	10	93.9894	21	95.2563	38	96.5965	35	96.8694	29
44	dti2	91.5442	37	92.7293	63	93.4369	64	95.1165	57	96.0986	67	95.9008	68
45	dti3	90.9252	58	93.0771	52	93.8427	49	95.2972	27	96.5931	36	96.8540	31
46	dte	91.9892	23	93.2834	28	93.9723	24	95.2938	28	96.6545	26	96.9069	17
47	dtmnk	92.1580	12	93.4096	14	93.9314	32	95.2614	36	96.5215	46	96.7841	42
48	dtch	90.9713	57	92.0813	69	92.9646	68	94.3440	69	96.1072	66	96.4414	63
49	dtcan	92.0012	21	92.7241	64	92.8861	69	94.3645	68	94.6237	69	94.0337	69
50	dtas	91.9585	25	93.2613	32	93.9689	25	95.3518	16	96.6323	32	96.8438	33
51	dtchi	92.2314	6	93.2988	27	93.9962	19	95.3091	23	96.5658	42	96.7057	48
52	dtchi2	92.2160	7	93.3039	25	94.0047	15	95.3108	22	96.5624	43	96.7636	44
53	dtbc	92.1001	15	93.3141	21	93.8478	45	95.2426	40	96.4158	52	96.5863	55
54	dthl	92.1819	10	93.1879	43	93.9450	30	95.2836	31	96.5334	45	96.6392	53
55	dtchr	91.9585	25	93.2613	32	93.9689	25	95.3518	16	96.6323	32	96.8438	33
56	dtchr2	92.1904	9	93.2186	39	93.9621	27	95.2836	31	96.5215	46	96.6869	50
57	dtjac	92.1001	15	93.3141	21	93.8478	45	95.2426	40	96.4158	52	96.5863	55
58	dtper	92.1001	15	93.3141	21	93.8478	45	95.2426	40	96.4158	52	96.5863	55
59	dtse	91.7164	31	93.2630	31	93.7933	55	95.0022	60	96.2350	63	96.2896	66
60	dtfr	92.0012	21	93.7319	6	94.0166	13	95.1540	54	96.3919	57	96.5624	61
61	dtm	93.0686	3	94.1786	3	94.3594	3	95.6451	2	96.7687	6	96.8898	24
62	dtmc	93.3636	2	94.2724	2	94.4208	2	95.1165	57	95.9520	68	96.1839	67
63	epw	90.6353	62	92.9424	58	93.8376	50	95.2324	45	96.7091	15	96.9171	14
64	gc	91.0378	56	92.9663	56	93.7626	57	95.3074	24	96.6988	18	96.8574	30
65	mgc	91.7352	29	93.2766	29	93.9859	22	95.4609	6	96.7091	15	96.9307	10
66	ogc	91.9602	24	93.5784	9	94.1599	7	95.5342	4	96.8301	2	97.0723	2
67	omgc	91.4964	39	93.2203	38	94.2622	6	95.4728	5	96.7773	3	97.0075	5
68	mlr	92.9458	4	94.1258	4	94.3389	4	95.6280	3	96.7739	4	96.9052	19
69	svmcmb	97.1814	1	97.2888	1	97.2990	1	97.3792	1	97.7765	1	97.8788	1

It is interesting that Dempster Rule and its variations (13-35) have in average better recognition rates than decision templates schemes (40-62).

The naive Bayes cooperation schemes (36-37) are relatively good choice while the fuzzy integration (38-39) shows weak results.

The generalized committee prediction and its variations (63-67), together with linear regression (68) are among the best methods and should be considered as serious candidates for implementation in any pattern recognition application based on classifier cooperation.

5 Conclusion

In this paper, the cooperation of three feature families for handwritten digit recognition using SVM classifiers is examined. We investigate an extensive number of cooperation schemes based on classifier decision fusion.

The presented results show that it is difficult to achieve the recognition rate of a single SVM applied on the feature set that includes all feature families by combining the individual SVM decisions. In our experiments only one of the cooperation schemes exceeded the recognition rate of a single SVM classifier. These results impose the crucial question: whether the methods for classifier cooperation are still needed [14] or pattern recognition tasks could be better solved by a single, well-optimized SVM classifier. However, the classifier cooperation reduces the classifier complexity, need for samples, and sometimes can increase the classifier performance.

References

1. Burges, C.: A Tutorial on Support Vector Machines for Pattern Recognition, Knowledge Discovery and Data Mining, Vol. 2 (1998) 1–47
2. LeCun, Y., Jackel, L. D., Bottou, L., Brunot, A., Cortes, C., Denker, J. S., Drucker, I. Guyon, H., Muller, U. A., Sackinger, E., Simard, P., 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) 53–60
3. Gorgevik, D., Cakmakov, D.: An Efficient Three-Stage Classifier for Handwritten Digit Recognition, Proc. of 16th Int. Conference on Pattern Recognition, Vol. 4, IEEE Computer Society, Cambridge, UK (2004) 507–510
4. Pratt, W. K.: Digital Image Processing: PIKS Inside, Third Ed., John Wiley & Sons (2001)
5. Collobert, R., Bengio, S., Mariéthoz, J.: Torch: a modular machine learning software library, Technical Report IDIAP-RR 02-46, Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP), CH-1920 Martigny, Switzerland (2002) (www.torch.ch)
6. Ho, T.K., Hull, J.J., Srihari, S.N.: Decision Combination in Multiple Classifier Systems, IEEE Tran. on Pattern Analysis and Machine Intelligence, Vol. 16, No. 1 (1994) 66–75
7. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On Combining Classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3 (1998) 226–239
8. Schürmann, J.: Pattern Classification: A Unified View of Statistical and Neural Approaches, Johm Wiley & Sons, Inc. (1996)
9. Gorgevik, D.: Classifier Combining for Handwritten Digit Recognition, Ph.D. dissertation, Faculty of Electrical Engineering, Skopje, Macedonia (2004).
10. Xu, L., Krzyzak, A., Suen, C.Y.: Methods of combining multiple classifiers and their application to handwritten recognition, IEEE Transactions on System, Man and Cybernetics, Vol. 22 (1992) 418–435
11. Cho, S.B., Kim, J.H.: Combining multiple neural networks by fuzzy integral and robust classification, IEEE Tran. on System, Man and Cyber., Vol. 20, No. 3 (1995) 380–384.
12. Kuncheva, L.I., Bezdek, J.C., Duin, P.W.: Decision templates for multiple classifier fusion: an experimental comparison, Pattern Recognition, Vol. 34, No. 2 (2001) 299–314
13. Bishop, C.M.: Neural Networks for Pattern Recognition, Clarendon Press, Oxford (1995).
14. Kittler, J.: A Framework for Classifier Fusion - Is It Still Needed, in F. J. Ferri, J. M. Inesta, A. Amin and P. Pudil, Eds., Advances in Pattern Recognition, Lecture Notes in Computer Science, Vol. 1876, Springer-Verlag (2000) 45–56