EXPERIMENTS WITH CLASSIFIER COMBINING RULES IN FACE EXPRESSION RECOGNITION VIA SPARSE REPRESENTATION.

Fusion rulers for Face Expression Recognition

December 1, 2015

Pedro Diamel Marrero Fernandez

Universidade Federal de Pernambuco
Brasil **CIN**

TOPICS

INTRODUCTION

PAPERS

- Ouyang, Y., Sang, N., and Huang, R. (2015). Accurate and robust facial expressions recognition by fusing multiple sparse representation based classifiers. Neurocomputing, 149, 71-78.
- Kittler, J., Hatef, M., Duin, R. P., and Matas, J. (1998). On combining classifiers. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20(3), 226-239.
- Kuncheva, L. I., and Rodríguez, J. J. (2014). A weighted voting framework for classifiers ensembles. Knowledge and Information Systems, 38(2), 259-275.

MULTI CLASSIFIER SYSTEM

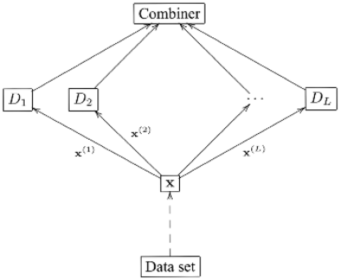

Figure: Approaches to building classifier ensambles

PROBABILISTIC SET-UP

DECISION PROFILE

Output of classifier $D_i(\mathbf{x})$

$$DP(\mathbf{x}) = \begin{bmatrix} d_{1,1}(\mathbf{x}) & \dots & d_{1,j}(\mathbf{x}) & \dots & d_{1,c}(\mathbf{x}) \\ \dots & & & & \\ d_{i,1}(\mathbf{x}) & \dots & d_{i,j}(\mathbf{x}) & \dots & d_{i,c}(\mathbf{x}) \\ \dots & & & & \\ d_{L,1}(\mathbf{x}) & \dots & d_{L,j}(\mathbf{x}) & \dots & d_{L,c}(\mathbf{x}) \end{bmatrix}.$$

Support from classifiers $D_1 \dots D_L$ for class j

Figure: Decision profile.

POSTERIOR PROBABILITY I

1. Crisp classifier: $\mu_D^i(x) \in \{0, 1\}, \sum_{i=1}^{c} \mu_D^i(x) = 1, \forall x \in \mathbb{R}^n$;
2. Fuzzy classifier: $\mu_D^i(x) \in [0, 1], \sum_{i=1}^{c} \mu_D^i(x) = 1, \forall x$; (Probabilistic interpretation of the outputs fall in this category)
3. Possibilistic classifier: $\mu_D^i(x) \in [0, 1], \sum_{i=1}^{c} \mu_D^i(x) > 0, \forall x$;

POSTERIOR PROBABILITY I

$$\mu_D^i = P(w_i|x), i = 1, 2..., c$$

POSTERIOR PROBABILITY II

$$P(w_k\ is\ the\ true\ class\,|s_1, s_2, ..., s_L), k = 1, ..., c,$$

POSTERIOR PROBABILITY II

$$P(w_k \; is \; the \; true \; class \, | s_1, s_2, ..., s_L), k = 1, ..., c,$$

Assume that the classifiers give their decisions independently:

$$P(w_k | s) = \frac{P(w_k)}{P(s)} \prod_i P(s_i | w_k)$$

THE COMBINING OF HARD CLASSIFIERS

EQUATION SUMMARY

- Majority vote (MV):

$$\log(P(w_k|s)) \propto \log(\frac{1-p}{p(c-1)}) \log(P(w_k)) + |I_+^k|$$

where $|I_+^k|$ is the number of votes for w_k.

- Weighted majority vote (WMV):

$$\log(P(w_k|s)) \propto \log(P(w_k)) + \sum_{i \in |I_+^k|} \theta_i + |I_+^k| \times \log(c-1)$$

where $\theta_i = \log(\frac{p_i}{1-p_i}), 0 < p_i < 1$

EQUATION SUMMARY

→ Recall combiner (REC):

$$\log(P(w_k|s)) \propto \log(P(w_k)) + \sum_i \log(1 - p_{ik})$$

$$+ \sum_{i \in |I_+^k|} v_{ik} + |I_+^k| \times \log(c-1).$$

where $v_{ik} = \log(\frac{p_{ik}}{1-p_{ik}}), 0 < p_{ik} < 1$

→ Naive Bayes combiner (NB):

$$\log(P(w_k|s)) \propto \log(P(w_k)) + \sum_i \log(p_{i,s_i,k}).$$

THE COMBINING OF SOFT CLASSIFIERS

EQUATION SUMMARY

- Product Rule (PR):

$$P^{(-(R-1))}(w_j)\prod_i P(w_j|x_i) = \max_k P^{(-(R-1))}(w_k)\prod_i P(w_k|x_i)$$

which under the assumption of equal priors, simplifies to the following:

$$\prod_i P(w_j|x_i) = \max_k \prod_i P(w_k|x_i)$$

- Sum Rule (SR):

$$(1-R)P(w_j)+\sum_i P(w_j|x_i) = \max_k [(1-R)P(w_k)+\sum_i P(w_k|x_i)]$$

which under the assumption of equal priors simplifies to the following:

$$\sum_i P(w_j|x_i) = \max_k \sum_i P(w_k|x_i)$$

ANOTHER APPROACH

TRAINABLE COMBINING OF CLASSIFIER

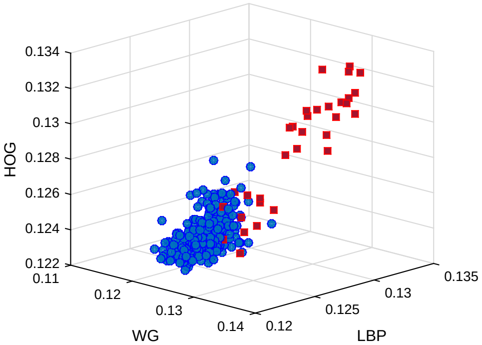

Figure: Decision profile of SRM of the neutral expression class output in feature subspace LBP, HoG and WG of the CK dataset. In red the true element in this class and blue not true elements in this class.

TRAINABLE COMBINING OF CLASSIFIER

Linear Opinion Pools (LOP):

$$P(w_k|s) = \sum_i \theta_{ki} P(w_k|s_i)$$

$$J = \sum_i \sum_k \theta_i \theta_k \sigma_{ik} - \lambda(\sum_i \theta_i - 1)$$

the solution minimizing J is:

$$\theta = \Sigma^{-1} I (I^T \Sigma^{-1} I)^{-1}$$

APPLICATION

SPARSE REPRESENTATION

Figure: Sparse representation of image CK++ dataset.

$$\alpha^{\star} = \arg\min_{\alpha \in \mathbb{R}^p} \|\alpha\|_0 \;\; s.t \; D\alpha = x$$

SPARSE REPRESENTATION

Figure: Sparse representation of image CK++ dataset.

$$\alpha^{\star} = \arg\min_{\alpha \in \mathbb{R}^p} \|\alpha\|_0 \ \ s.t\, D\alpha = x$$

FEATURE SUBSPACE

- → Histogram of Oriented Gradients (HoG).
- → Local Binary Patterns (LBP).
- → Gabor wavelet (GW)
- → Raw Image

CLASSIFICATION

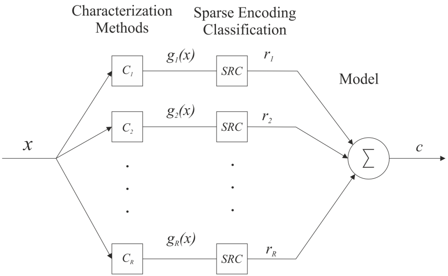

Figure: Framework for classification.

RESULT

A SIMULATION STUDY

Protocol:

- → Number of classes $c \in 2, 3, 4, 5, 10, 20, 50$;
- → Number of classifiers $L \in 2, 3, 4, 5, 10, 20, 50$;
- → Number of instances (labels) 500;
- → Number of runs 100.

A SIMULATION STUDY

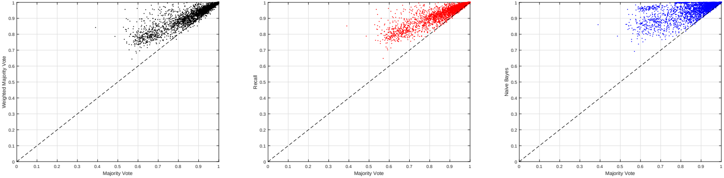

Figure: Relationship between the ensemble accuracies using the majority vote as the benchmark combiner. Each scatterplot contains 4900 ensembles points.

A SIMULATION STUDY

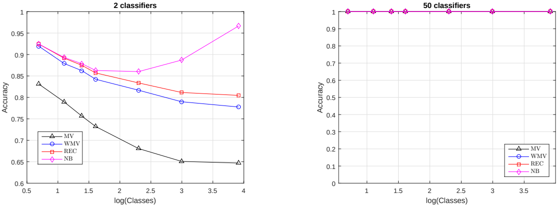

Figure: Ensemble accuracies of the 4 combiners as a function of log(c) (exact parameter estimates).

A SIMULATION STUDY

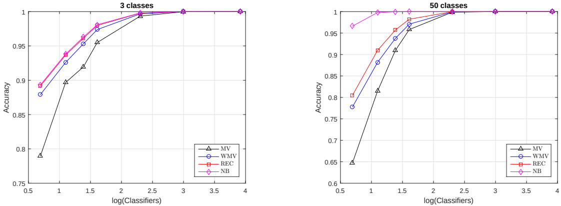

Figure: Ensemble accuracies of the 4 combiners as a function of log(L) (exact parameter estimates).

EXPERIMENTS WITH REAL DATA

Protocol: We used $L = 100$ decision tree classifiers and 10-fold cross validation. All experiments were run in Matlab. The accuracy of each ensemble is the average across the 100 testing results.

EXPERIMENTS WITH REAL DATA

Table: Ensemble Accuracies With the 4 Combiners.

Data set	MV	WMV	REC	NB
Balance	$0.836 + 0.040$	$0.838 + 0.039$	$0.827 + 0.053$	$0.827 + 0.667$
Breast-w	$0.962 + 0.021$	$0.962 + 0.020$	$0.963 + 0.019$	$0.963 + 0.468$
Breast-y	$0.712 + 0.080$	$0.713 + 0.076$	$0.616 + 0.093$	$0.616 + 0.476$

EXPERIMENT IN APPLICATION OF EXPRESSION RECOGNITION.

Protocol: To evaluate this work in FER problems, we used public DBs, i.e., Extended Cohn-Kanade(CK+). In all experiment were using person-independent FER scenarios, where the subjects in training set were completely different from the subjects within test set (i.e., the subjects used for training procedure cannot be used for testing phase).

EXPERIMENT IN APPLICATION OF EXPRESSION RECOGNITION.

Table: Average Test Accuracy for All Combined and All Feature Subspaces.

		Combining of Soft Classifiers					Combining of Crisp Classifiers				Trainable Classifiers
		RP	RS	RMX	RMI	RMD	RM	WMV	REC	NB	LOP
L/H	μ	**0.891**	**0.891**	0.888	0.875	0.891	0.874	0.856	0.885	0.883	**0.891**
	σ	0.216	0.216	0.226	0.228	0.216	0.240	0.240	0.230	0.229	0.216
R/H	μ	0.834	0.834	0.820	0.820	0.834	0.830	0.843	**0.869**	0.843	0.840
	σ	0.256	0.256	0.257	0.276	0.256	0.242	0.260	0.233	0.274	0.260
R/L	μ	0.842	0.842	0.826	0.827	0.842	0.827	0.832	**0.851**	0.841	0.842
	σ	0.248	0.248	0.262	0.255	0.248	0.258	0.271	0.260	0.274	0.244
R/L/H	μ	0.845	0.843	0.847	0.837	0.843	0.870	0.855	**0.876**	0.847	0.867
	σ	0.259	0.258	0.248	0.254	0.258	0.228	0.246	0.236	0.266	0.240
R/W	μ	0.901	0.901	0.893	0.862	0.901	0.866	0.920	**0.933**	0.891	0.906
	σ	0.189	0.189	0.190	0.221	0.189	0.215	0.163	0.136	0.218	0.183
R/W/H	μ	0.911	0.911	0.896	0.863	0.911	0.901	0.894	0.916	**0.925**	0.916
	σ	0.182	0.182	0.189	0.235	0.182	0.191	0.192	0.169	0.162	0.177
R/W/L	μ	0.908	0.908	0.895	0.868	0.908	0.904	0.892	0.913	0.887	**0.918**
	σ	0.183	0.183	0.188	0.226	0.183	0.194	0.203	0.191	0.224	0.177
R/W/L/H	μ	0.913	0.913	0.900	0.871	0.913	0.897	0.897	0.913	**0.917**	0.914
	σ	0.183	0.183	0.187	0.228	0.183	0.209	0.213	0.188	0.196	0.181
W/H	μ	0.924	0.924	0.925	0.902	0.924	0.932	0.882	**0.935**	0.921	0.923
	σ	0.167	0.167	0.164	0.197	0.167	0.144	0.216	0.150	0.196	0.174
W/L	μ	0.935	0.935	0.930	0.894	0.935	0.941	0.892	**0.939**	0.933	0.931
	σ	0.157	0.157	0.138	0.216	0.157	0.130	0.202	0.152	0.170	0.158
W/L/H	μ	0.919	0.919	0.930	0.908	0.919	0.918	0.898	0.911	**0.935**	0.925
	σ	0.187	0.187	0.163	0.196	0.187	0.191	0.212	0.199	0.172	0.171

EXPERIMENT IN APPLICATION OF EXPRESSION RECOGNITION.

Table: Compared to Individual Methods ($e * 1000^{-1}$).

	Combining											Individual			
Methods	L/H	R/H	R/L	R/L/H	**R/W**	**R/W/H**	R/W/L	R/W/L/H	**W/H**	**W/L**	**W/L/H**	R	L	H	W
Accuracy	891 216	869 233	851 260	876 236	933 136	925 162	918 177	917 196	939 152	939 152	935 172	767 244	914 137	875 196	892 184

CONCLUSION

- The combining of hard classifiers shows the best results for facial expression recognition problem via SR.
- Combining classifiers trained on different feature sets is very useful, especially when in these feature set probabilities are well estimated by the classifier.
- The trainable classifier combination may have good results for such problems.
- The best results were obtained using: WG + LBP + SR + REC.

Thank you...