

Gender Classification Based on Support Vector Machine with Automatic Confidence

Zheng Ji¹ and Bao-Liang Lu^{1,2,*}

¹Center for Brain-Like Computing and Machine Intelligence
Department of Computer Science and Engineering

²MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems
Shanghai Jiao Tong University
800 Dong Chuan Rd., Shanghai 200240, China
{jizheng, bllu}@sjtu.edu.cn

Abstract. In this paper, we propose a support vector machine with automatic confidence (SVMAC) for gender classification based on facial images. Namely, we explore how to incorporate confidence values introduced in each primitive training sample and compute these values automatically into machine learning. In the proposed SVMAC, we use both the labels of training samples and the label confidence which projected by a monotonic function as input. The main contribution of SVMAC is that the confidence value of each training sample is calculated by some common algorithms, such as SVM, Neural Network and so on, for gender classification. Experimental results demonstrate that SVMAC can improve classification accuracy dramatically.

Keywords: Support Vector Machine, Feature Extraction, Gender Classification.

1 Introduction

Due to its wide application in human-computer interaction, such as vital statistics, identity appraisal, visual surveillance and robot vision [1], gender classification based on facial images has attracted many researchers' attention.

One of the most challenging problems is to devise a proper classification algorithm to classify gender information of faces. In other words, we need to select a better classifier to improve the classification performance. Among all kinds of recognition algorithms [2] [3] [4] [5], support vector machine (SVM) is one of the most popular classification methods, providing a sound theoretic basis for constructing classification models with high generalization ability. Li and Lu [6] brought forward a framework based on multi-view gender classification where a trained layer support vector machine (LSVM) is utilized to recognize the angle of each facial image and classify its gender. To develop new machine learning algorithms and improve the performance of existing machine learning methods, it is very important for us to consider the problem how SVM can be ameliorated. Ji *et al* [7] proposed a support vector machine with confidence (SVMC) labeled manually. But before training, the confidence of each training sample

* Corresponding author.

must be labeled manually. Thus, when the number of training samples is very large we must spend much time in labeling the confidence. Furthermore, we can not guarantee all these labeled confidence values are reasonable. To explore how to label rational confidence for each training sample automatically, we propose the support vector machine with automatic confidence (SVMAC).

The remaining part of the chapter is organized as follows: Section 2 described the proposed SVMAC model in detail. Experiments are presented in Section 3. Some conclusions and discussions on future work are outlined in Section 4.

2 The Proposed SVMAC Model

The quadratic programming problems for the standard and soft margin forms of traditional SVM [8] [9] [10] [7] can be expressed as

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ \text{s.t.} \quad & \forall i, y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0 \end{aligned} \quad (1)$$

and

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + D \sum_i \xi_i^2 \\ \text{s.t.} \quad & \forall i, y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0 \end{aligned} \quad (2)$$

respectively. One way of incorporating confidence values is to re-scale the soft margin as follows,

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + D \sum_i \xi_i^2 \\ \text{s.t.} \quad & \forall i, y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq t(\pi_i) - \xi_i, \\ & \xi_i \geq 0 \end{aligned} \quad (3)$$

where $t(\pi_i)$ is a monotonic function to scale the confidence, namely

$$t(\pi_i) = h \cdot \pi_i, \quad \frac{1}{2} \leq \pi_i < 1 \quad (4)$$

where h is the scale parameter.

Existing work reported that the support vectors obtained by a support vector machine tend to be those training samples that people can not discriminate well [2] [11]. Based on this fact, we propose a modified support vector machine. First, we divide all the training samples into two disjointed subsets \mathcal{U} and \mathcal{V} ($\mathcal{X} = \mathcal{U} \cup \mathcal{V}$), which are later treated in a different way in the training algorithm. Then, we put the training samples in \mathcal{U} with confidence π_i less than 1, and the remaining training samples in \mathcal{V} with confidence π_i equal to 1. In essence, \mathcal{U} contains the training samples that tend to be support vectors after training. In the following, we denote the number of training samples in \mathcal{U} and \mathcal{V} by n_u and n_v , respectively.

According to Eq. (3) for training subset \mathcal{U} and Eq. (1) for training subset \mathcal{V} , we can express the quadratic programming problem for soft margin form as follows:

$$\begin{aligned}
 \min_{\mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + D \sum_{i=1}^{n_u} \sigma_i^2 + C \sum_{i=1}^{n_v} \xi_i, \\
 \text{s.t.} \quad & \forall 1 \leq i \leq n_u, \\
 & y_i^u (\mathbf{w}^T \mathbf{u}_i + b) = t(\pi_i) - \sigma_i, \\
 & \forall 1 \leq i \leq n_v, \\
 & y_i^v (\mathbf{w}^T \mathbf{v}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0.
 \end{aligned} \tag{5}$$

Using the standard Lagrangian dual technique, we get the following dual form:

$$\begin{aligned}
 \min_{\lambda, \alpha} \quad & \frac{1}{2} \left(\sum_{i=1}^{n_u} \lambda_i y_i^u \mathbf{u}_i + \sum_{i=1}^{n_v} \alpha_i y_i^v \mathbf{v}_i \right)^T \\
 & \left(\sum_{i=1}^{n_u} \lambda_i y_i^u \mathbf{u}_i + \sum_{i=1}^{n_v} \alpha_i y_i^v \mathbf{v}_i \right) \\
 & - \sum_{i=1}^{n_u} t(\pi_i) \lambda_i + \frac{1}{4D} \sum_{i=1}^{n_u} \lambda_i^2 - \sum_{i=1}^{n_v} \alpha_i \\
 \text{s.t.} \quad & \forall 1 \leq i \leq n_u, \quad 0 \leq \lambda_i < +\infty, \\
 & \forall 1 \leq i \leq n_v, \quad 0 \leq \alpha_i \leq C, \\
 & \sum_{i=1}^{n_u} \lambda_i y_i^u + \sum_{i=1}^{n_v} \alpha_i y_i^v = 0.
 \end{aligned} \tag{6}$$

However, it is a considering problem how to label the confidence of training samples reasonably. In SVMC [7] the confidence of all the training samples is labeled manually. But if the number of the training samples is large, we must spend much time in labeling their confidence values. Moreover, we can not guarantee that all the labeled confidence values are reasonable because people’s action on making them is very subjective. Therefore, we suggest using logical methods to divide the training samples into the two sets \mathcal{U} and \mathcal{V} . The algorithm ALC of labeling the confidence is displayed in Algorithm 1.

As a matter of fact, the size of the distance between a training sample and decision boundary γ suggests whether the sample can be discriminated well. Obviously, the training sample which is far from the decision boundary can tend to be discriminated and should be appended into \mathcal{V} . Otherwise, it need to be added in \mathcal{U} . Therefore, the automatic mark is coincident with the manual label on the confidence of training samples.

Now we take into account the performance of SVMAC in some situation. According to the algorithm ALC and SVMAC defined in Eq. (5) that makes use of the method of probability statistics [7] [12] [13], we set the confidence values in \mathcal{U} less than 1. Those samples marked by small circles (green) are shown in Fig. 1, and the right figure of Fig. 2. We can observe that the decision boundary is changed if the confidence values of the support vectors in \mathcal{U} are assigned by employing the algorithm ALC where the sample set $\Gamma = \{(x_i, y_i) | 1 \leq i \leq N\}$ is trained by SVM. We can conclude that the movement of the decision boundary is identical to the one in SVMC [7].

Algorithm 1. ALC

Step 1: The sample set $\Gamma = \{(x_i, y_i) | 1 \leq i \leq N\}$ is trained by utilizing some algorithms such as SVM, Adaboost, Neural Network, and so on. Thus, a decision hyperplane γ , namely, a classifier can be obtained;

Step 2: Computing these distances between all the samples in Γ and the hyperplane γ and Obtaining the distance set $\Omega = \{d_i | \text{the distance between the } i\text{-th sample and } \gamma\}$;

Step 3: Given threshold value σ ,
for all i from 1 to N
 if $d_i < \sigma$,
 the sample (x_i, y_i) is added in \mathcal{U} ;
 else
 (x_i, y_i) is added in \mathcal{V} ;
 end if
end for

Step 4: The confidence values of the samples in \mathcal{V} are set to 1.0 while the ones in \mathcal{U} is projected onto the confidence space $[\frac{1}{2}, 1)$ according to the distances.

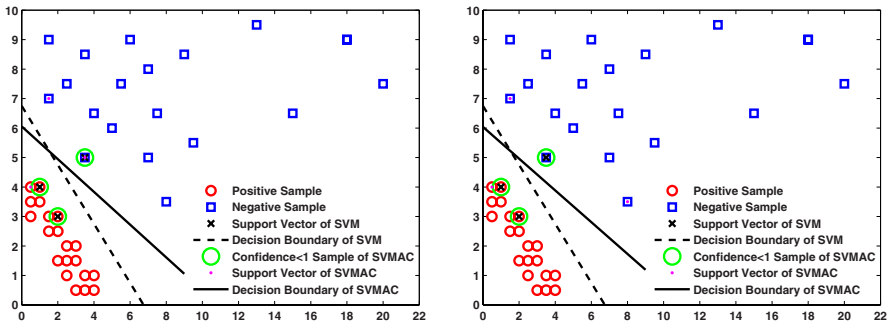


Fig. 1. Illustration of movement of decision boundary caused by the proposed SVMAC. The scale parameter h for SVMAC in Eq. (4) is set to 1.0 and 0.1 in the left and right figures, respectively.

We regard the support vectors obtained by means of SVM as the training samples close to noise. Therefore, we should assign them with confidence values less than 1. By training the proposed SVMAC on all the training samples with proper confidence values, we obtain the decision boundaries as shown in Fig. 1. From this figure, we can see that if the support vectors obtained by the traditional SVM are assigned with appropriate confidence values, some of them may be turned into non-support vectors after applying SVMAC. The decision boundary obtained by SVMAC can be regarded as a fitting achieved by training the samples in which some noise is removed. Therefore, the decision boundary obtained by SVMAC is superior to that obtained by traditional SVM. For example, the lower left training samples in Fig. 1 are much denser and closer to the boundary than the upper right training samples, the movement of separation boundary from the lower left corner to the upper right corner caused by the proposed SVMAC no doubt yields a better separation than that of traditional SVM.

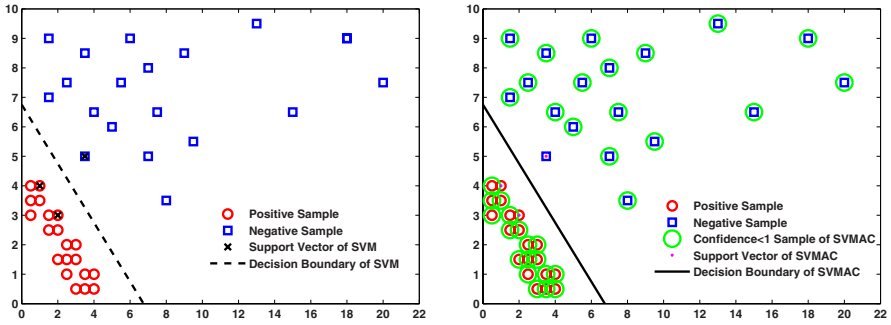


Fig. 2. Illustration of the decision boundaries formed by SVM (left) and SVMAC (right), where we only assign the confidence values (less than 1) to non-support vectors in \mathcal{V} for training SVMAC and do not consider any confidence values for support vectors in \mathcal{U}

Table 1. Description of training and test data based on facial images

Data Set Description		Total	Male	Female	Training	Test
CAS-PEAL	PD00	1040	595	445	311*2	418
	PD15	939	516	423	296*2	347
	PD30	939	516	423	296*2	347
	PM00	1039	595	444	310*2	419
	PM15	938	516	422	295*2	348
	PM30	938	516	422	295*2	348
	PU00	1040	595	445	311*2	418
	PU15	939	516	423	296*2	347
	PU30	939	516	423	296*2	347
TOTAL		8751	4881	3870	5412	3339
FERET	PM00	992	589	403	282*2	428
BCMI	PM00	1045	529	516	361*2	323
TOTAL	TOTAL	10788	5999	4789	6698	4090

From the angle of the confidence, the decision boundaries in Fig. 1 are the most superior boundaries. However, the decision boundaries produced by traditional SVM and the proposed SVMAC are the same as shown in Fig. 2, where only the non-support vectors in \mathcal{V} are assigned with confidence values less than 1 and none of support vectors in \mathcal{U} is assigned with confidence value. From this figure, we can see that the confidence values less than 1 assigned to non-support vectors in \mathcal{V} don't affect the decision boundary. In other words, after some non-support vectors in SVM are marked by the confidence values less than 1 through using ALC, the whole classification accuracy will not be decreased.

3 Experiments

To evaluate the performance of SVMAC, we select the gender classification problem based on multi-view facial images in the CAS-PEAL face database [14] and frontal

Table 2. Description of the mean accuracy denoted by *MA* (%) and the standard deviation defined as *SD* (%) caused by SVM and SVMAC with RBF kernel used by the first row and the second row in each group in turn on $F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8$ corresponding to MLGBP-CCL, MLGBP-LDA, LGBP-CCL, LGBP-LDA, MLBP, LBP, Gabor, and gray respectively, where CAS-PEAL, FERET and BCMI are represented by C, F, and B in front of these description names, and m (the number of window blocks) is ranged from 5×5 to 10×10

Description	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
	<i>MA/SD</i>	<i>MA/SD</i>	<i>MA/SD</i>	<i>MA/SD</i>	<i>MA/SD</i>	<i>MA/SD</i>	<i>MA/SD</i>	<i>MA/SD</i>
C-PD00	99.1/0.4	99.9/0.2	96.7/11.4	99.6/0.4	95.5/2.9	94.2/10.2	90.6/16.3	92.5/4.9
	99.3/0.3	100.0/0.0	97.4/4.8	99.9/0.2	96.4/1.1	95.3/4.4	92.6/15.1	93.8/4.3
C-PD15	99.3/1.1	100.0/0.0	97.4/7.5	99.7/0.1	94.9/1.1	93.5/3.0	88.7/26.0	91.3/1.1
	99.4/0.9	100.0/0.0	97.7/7.3	99.8/0.1	95.6/0.9	94.6/2.1	90.5/17.2	92.6/2.0
C-PD30	98.9/0.4	99.9/0.2	96.4/6.2	99.8/0.4	92.7/2.0	92.2/3.2	89.9/13.1	89.4/1.5
	99.1/0.6	100.0/0.0	96.8/4.8	100.0/0.0	93.3/2.1	93.2/1.9	90.4/8.4	90.6/2.5
C-PM00	99.6/0.2	100.0/0.0	97.3/10.6	100.0/0.0	96.1/2.7	94.5/10.4	91.6/26.3	94.6/6.0
	99.6/0.3	100.0/0.0	97.5/10.8	100.0/0.0	96.6/3.3	95.5/8.5	92.7/16.7	95.4/3.6
C-PM15	99.6/0.2	99.9/0.1	97.1/2.9	99.7/0.3	95.1/0.9	94.2/14.6	92.7/30.2	94.5/7.4
	99.7/0.8	100.0/0.0	97.3/2.7	99.9/0.2	95.7/2.5	94.7/15.3	93.2/23.4	95.0/6.9
C-PM30	98.9/0.3	99.9/0.1	96.3/6.8	99.7/0.5	93.7/9.1	92.4/15.3	91.9/17.4	92.1/1.5
	99.4/0.2	100.0/0.0	96.9/2.8	100.0/0.0	95.1/6.0	93.2/11.3	93.1/8.3	92.6/2.0
C-PU00	99.4/0.5	100.0/0.0	96.7/5.4	99.9/0.1	95.4/4.2	94.5/4.7	90.2/33.2	89.0/28.4
	99.6/0.5	100.0/0.0	97.6/5.6	100.0/0.0	96.3/2.4	95.3/4.8	91.6/18.4	92.0/17.9
C-PU15	99.3/0.2	100.0/0.0	97.6/1.3	99.9/0.1	96.0/4.2	95.6/1.8	92.6/8.9	89.3/16.5
	99.9/0.2	100.0/0.0	98.3/2.5	100.0/0.0	96.6/0.8	96.0/2.5	93.1/6.5	90.2/18.8
C-PU30	98.8/1.4	100.0/0.0	96.9/8.6	99.9/0.2	94.0/4.0	92.8/5.1	89.0/21.1	88.2/9.1
	99.2/0.5	100.0/0.0	97.0/6.3	100.0/0.0	95.0/1.2	93.6/5.5	90.2/13.0	89.6/5.9
F-PM00	96.8/2.3	99.6/0.2	94.6/11.6	98.8/2.3	93.7/1.3	92.3/9.0	90.4/9.8	89.5/3.7
	97.2/2.4	99.7/0.2	95.1/8.8	99.1/0.6	93.8/1.0	93.1/3.9	91.6/15.1	91.2/4.6
B-PM00	98.8/1.2	100.0/0.0	97.3/8.3	99.4/1.4	96.3/2.0	96.2/3.2	93.4/10.2	95.4/2.1
	99.0/0.7	100.0/0.0	98.1/0.7	99.7/0.2	97.2/0.9	97.3/0.9	95.1/6.1	95.7/1.2

face pictures in FERET¹ and BCMI² databases as a benchmark problem, respectively, and make some comparative studies. The total 10788 different-pose facial images are organized into 11 groups in each of which the numbers of training and test samples are 70% and 30% of the whole group, respectively (See Table 1).

In this paper, we use gray, Gabor, local binary pattern (LBP) [15] [16], multi-resolution local binary pattern (MLBP) [17], local Gabor binary pattern (LGBP) [18], and multi-resolution local Gabor binary pattern (MLGBP) approaches to extract the features of each facial image. Thereinto, the MLGBP feature as input of SVM [8] [9] [10] and SVMAC classifiers is derived by combining multi-resolution analysis, Gabor characteristic and uniform LBP histograms. All experiments were performed on a Pentium fourfold CPU (2.83GHz) PC with 8GB RAM.

¹ <http://www.frvt.org/FERET/default.htm>

² BCMI face database is set up and packed up by the Center for Brain-Like Computing and Machine Intelligence Shanghai Jiao Tong University, Shanghai, China.

From Table 2, we conclude that the average classification accuracy caused by SVMAC is higher than SVM on the same face feature, and the standard deviation of classification precision brought by SVMAC is lower than SVM. What's the more, the performance improvement of SVMAC is obvious for Gray, Gabor, LBP and MLBP features and the maximum improvement accuracy obtained between SVMAC and SVM reaches 3.0%. But for MLGBP-CCL, LGBP-CCL, MLGBP-LDA and LGBP-LDA features, because all the accuracies are very high, SVMAC improve the classification performance a little only. These indicate that SVMAC improves the classification performance compared to traditional SVM, where the parameter C in Eqs. (1) and (6) is consistent. In addition, we observe that the classification performance is also dependent on the distributions of training samples. Generally speaking, there are two kinds of sample distributes. One is dense and the other is sparse, such as in Fig. 1. In this situation, if the confidence values less than 1 are set for the support vector samples, the decision boundary obtained by SVMAC will favor the sparse samples in comparison with traditional SVM. Consequently, from the experimental results and theoretical analysis, by modifying the confidence values of the support vector samples, we can separate the data samples more reasonably.

4 Conclusions and Future Work

We have proposed a novel support vector machine with automatic confidence, i.e., SVMAC. The most important advantage of this presented SVMAC over traditional SVM is that some explicit human prior knowledge estimated by the algorithm ALC on training samples can be easily incorporated into learning. We have derived the quadratic programming problem for SVMAC and analyzed its performance theoretically. Experimental results on a gender classification problem based on facial images indicate that this proposed method can improve classification accuracy dramatically. As future work, we would like to give a bound for the improvement on classification accuracy about SVMAC and apply it to other real-world pattern classification problems, such as text classification, age estimation and object recognition.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant No. 60773090 and Grant No. 90820018), the National Basic Research Program of China (Grant No. 2009CB320901), and the National High-Tech Research Program of China (Grant No. 2008AA02Z315).

References

1. Schmitt, L.A., Gruver, W.A., Ansari, A.: A robot vision system based on two-dimensional object-oriented models. *IEEE Transactions on Systems, Man and Cybernetics* 16(4), 582–589 (1986)
2. Graf, A., Wichmann, F., Bühlhoff, H., Schölkopf, B.: Classification of faces in man and machine. *Neural Computation* 18(1), 143–165 (2005)

3. Doumpos, M., Zopounidis, C., Golfopoulou, V.: Additive support vector machines for pattern classification. *IEEE Transactions on System, Man, and Cybernetics* 37(3), 540–550 (2007)
4. Peng, J., Heisterkamp, D.R., Dai, H.K.: Lda/svm driven nearest neighbor classification. *IEEE Transactions on Neural Networks* 14(4), 158–163 (2003)
5. Chen, J.H., Chen, C.S.: Reducing svm classification time using multiple mirror classifiers. *IEEE Transactions on System, Man, and Cybernetics, Part B* 34(2), 1173–1183 (2004)
6. Li, J., Lu, B.L.: A framework for multi-view gender classification. In: Ishikawa, M., Doya, K., Miyamoto, H., Yamakawa, T. (eds.) *ICONIP 2007, Part I. LNCS*, vol. 4984, pp. 973–982. Springer, Heidelberg (2007)
7. Ji, Z., Yang, W.Y., Wu, S., Lu, B.L.: Encoding human knowledge for visual pattern recognition. In: *The 5th International Symposium on Neural Networks* (2008)
8. Cristianini, N., Shawe-Taylor, J.: *An introduction to support vector machines and other kernel-based learning methods*. Publishing House of Electronics Industry (2004)
9. Sloin, A., Burshtein, D.: Support vector machine training for improved hidden markov modeling. *IEEE Transactions on Signal Processing* 56(1), 172–188 (2008)
10. Williams, P., Li, S., Feng, J., Wu, S.: A geometrical method to improve performance of the support vector machine. *IEEE Transactions on Neural Networks* 18(3), 942–947 (2007)
11. Feng, J., Williams, P.: The generalization error of the symmetric and scaled support vector machine. *IEEE Transactions on Neural Networks* 12(5), 1255–1260 (2001)
12. Osuna, E., Freund, R., Girosi, F.: An improved training algorithm for support vector machines. *Neural Networks for Signal Processing*, 276–285 (1997)
13. Vapnik, V.N.: *Statistical learning theory*. Wiley, New York (1998)
14. Gao, W., Cao, B., Shan, S.G., et al.: The cas-peal large-scale chinese face database and baseline evaluations. Technical report of JDL (2004), http://www.jdl.ac.cn/~peal/peal_tr.pdf
15. Ojala, T., Pietikainen, M., Maeopaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE transactions on Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)
16. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J(G.) (eds.) *ECCV 2004. LNCS*, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
17. Lian, H.C., Lu, B.L.: Multi-view gender classification using multi-resolution local binary patterns and support vector machines. *International Journal of Neural Systems* 17(6), 479–487 (2007)
18. Xia, B., Sun, H., Lu, B.L.: Multi-view gender classification based on local gabor binary mapping pattern and support vector machines. *IEEE International Joint Conference on Neural Networks*, 3388–3395 (2008)