Gender Recognition Using a Min-Max Modular Support Vector Machine with Equal Clustering

Jun Luo and Bao-Liang Lu*

Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Rd., Shanghai 200240, China
{ljfootball, bllu}@sjtu.edu.cn

Abstract. Through task decomposition and module combination, min-max modular support vector machines (M^3-SVMs) can be successfully used for different pattern classification tasks. Based on an equal clustering algorithm, M^3-SVMs can divide the training data set of the original problem into several subsets with nearly equal number of samples, and combine them to a series of balanced subproblems which can be trained more efficiently and effectively. In this paper, we explore the use of M^3-SVMs with equal clustering method in gender recognition. The experimental results show that M^3-SVMs with equal clustering method can be successfully used for gender recognition and make the classification more efficient and accurate.

1 Introduction

Gender recognition is one of the most challenging problems among face recognition research areas. Nowadays, a lot of improvements have been achieved on this problem. As for classifiers, support vector machines (SVMs) have been successfully applied to solve the task and seem to be superior to other classifiers [1]. However, SVMs have to solve a quadratic optimization problem, and for a large-scale two-class problem such as gender classification it is rather difficult to improve classification accuracy.

In our previous work, a min-max modular support vector machine [2][3] is proposed to solve gender recognition problem [4] and also a novel clustering algorithm called equal clustering that can equally decompose training data sets is used to improve the performance of M^3-SVM [5]. In this paper, we explore the use of M^3-SVM with equal clustering method in gender recognition task. The gender recognition problem is a two-class pattern classification problem and the scale of training samples needs to be quite large in order to make the classification more accurate. Using M^3-SVM with equal clustering method, we can decompose the whole large problem of gender images into several balanced subproblems. Each individual subproblem becomes less complicated than the original problem and can be solved effectively and efficiently.

* To whome correspondence should be addressed. This work was supported by the National Natural Science Foundation of China under the grants NSFC 60375022 and NSFC 60473040.

J. Wang et al. (Eds.): ISNN 2006, LNCS 3972, pp. 210–215, 2006.

2 Min-Max Modular Support Vector Machine

The min-max modular neural network [6] is based on a 'divide-and-conquer' strategy and it can divide a large-scale problem into several independent smaller subproblems and then combine the results of each subproblem into a solution to the original problem. In essence, M^3-neural network is a general framework for machine learning.

M^3-SVM can easily solve multi-class problem and here for simplicity we just discuss the two-class problem. Given a training data set $\mathcal{S} = \mathcal{X}^+ \cup \mathcal{X}^-$, $\mathcal{X}^+$ and $\mathcal{X}^-$ denote the positive and negative training data sets for a two-class problem $\mathcal{T}$, respectively.

$$\mathcal{X}^+ = \{(x_i^+, +1)\}_{i=1}^{l^+}, \;\; \mathcal{X}^- = \{(x_i^-, -1)\}_{i=1}^{l^-} \tag{1}$$

where $x_i^+ \in \mathbf{R}^n$ and $x_i^- \in \mathbf{R}^n$ are the input vectors, and l^+ and l^- denote the total number of positive training data and the total number of negative training data of the two-class problem, respectively.

According to [2], M^3-SVMs consist of three steps. The first step is task decomposition. In this phase, $\mathcal{X}^+$ and $\mathcal{X}^-$ are decomposed into the following N^+ and N^- subsets, respectively.

$$\mathcal{X}^+ = \cup_{i=1}^{N^+} \mathcal{X}_i^+, \mathcal{X}_i^+ = \{(x_j^{+i}, +1)\}_{j=1}^{l_i^+}, i = 1, \ldots, N^+ \tag{2}$$

$$\mathcal{X}^- = \cup_{i=1}^{N^-} \mathcal{X}_i^-, \mathcal{X}_i^- = \{(x_j^{-i}, -1)\}_{j=1}^{l_i^-}, i = 1, \ldots, N^- \tag{3}$$

where $\cap_{i=1}^{N^+} \mathcal{X}_i^+ = \phi$, $\cap_{i=1}^{N^-} \mathcal{X}_i^- = \phi$, and ϕ denotes empty set as each subset is set to independent to others.

After the decomposition, every two subsets from $\mathcal{X}^+$ and $\mathcal{X}^-$ are chosen to form $N^+ \times N^-$ smaller subproblems as follows:

$$(\mathcal{T}^{(i,j)})^+ = \mathcal{X}_i^+, \, (\mathcal{T}^{(i,j)})^- = \mathcal{X}_j^- \tag{4}$$

where $(\mathcal{T}^{(i,j)})^+$ and $(\mathcal{T}^{(i,j)})^-$ denote the positive training data set and the negative training data set for the subproblem $\mathcal{T}^{(i,j)}$. And then all of the subproblems can be trained by SVMs in a massively parallel way.

The last step is to integrate all small trained modules to get a solution to the original problem using the minimization and maximization principles. The $N^+ \times N^-$ smaller SVMs are integrated into a M^3-SVM with N^+ MIN units and one MAX unit as follows,

$$\mathcal{T}^i(x) = \min_{j=1}^{N^-} \mathcal{T}^{(i,j)}(x) \;\text{ for } i = 1, \ldots, N^+ \text{ and } \mathcal{T}(x) = \max_{i=1}^{N^+} \mathcal{T}^i(x) \tag{5}$$

where $\mathcal{T}^{(i,j)}(x)$ denotes the transfer function of the trained SVM corresponding to the two-class subproblem $\mathcal{T}^{(i,j)}$, and $\mathcal{T}^i(x)$ denotes the transfer function of a combination of N^- SVMs integrated by the MIN unit.

3 Equal Clustering Task Decomposition Method

As mentioned above, the first step of M^3-SVM is task decomposition. Based on the previous work, several task decomposition strategies have been applied to solve gender recognition problem [4]. However, when training data is not identically distributed and no useful prior knowledge information can be found, the effectiveness of these two data partition methods will be unstable. In order to handle this problem, based on the algorithm "GeoClust" [7], we have proposed the equal clustering method in our previous work [5].

The way of generating reasonable clusters needs to solve the unconstrained nonlinear programming problem as follows:

$$\underset{c_1\, c_2\, \dots\, c_m}{\text{Minimize}} : h = \max_{i=1}^{m} |W_i - \overline{W}|,\ N = \sum_{i=1}^{m} W_i \tag{6}$$

where c_i and W_i denote the center vector and the number of samples in the ith cluster, respectively, m denotes the number of clusters, and $\overline{W}$ is the mean number of samples per cluster, i.e, $\overline{W} = \lfloor \frac{N}{m} \rfloor$.

Actually, the equal clustering algorithm is to generate spatially localized clusters that contain nearly equal number of samples to keep load balanced, so that it might catch local probability distribution of the training data set. This strategy has been evaluated on several benchmark data sets such as Banana data set and Letter Recognition data set [5], and its advantages have been demonstrated.

4 Experiments

4.1 Experimental Setup

In this section, we present experimental results on the CAS-PEAL face database [8] to compare M^3-SVMs using equal clustering method with the traditional SVMs and M^3-SVMs using random partition strategy. Here, there are several parameters for equal clustering algorithm and in our experiments we set *maxiter* $= 6000$, $\alpha = 0.01 \times 10^{-\lfloor \frac{(m-1)}{10} \rfloor}$, $l = 3$, and $\varepsilon = \lfloor \frac{N}{10m} \rfloor$. Here N denotes the number of training samples and m denotes the number of partition.

The training data sets include $2,670$ male and female samples respectively with 12 kinds of face poses. For each pose we select equal number of male and female samples. In order to enhance the performance, all images have been preprocessed and scaled into 65×75 images as shown in Fig.1. Here gray scale vector is chosen as feature and the dimension of each input vector is 4875. The probe sets represent face images with various pose degree ranged from $-30°$ to $30°$. In order to ensure the credibility of the conclusions, all experiments are repeated three times and the average is taken.

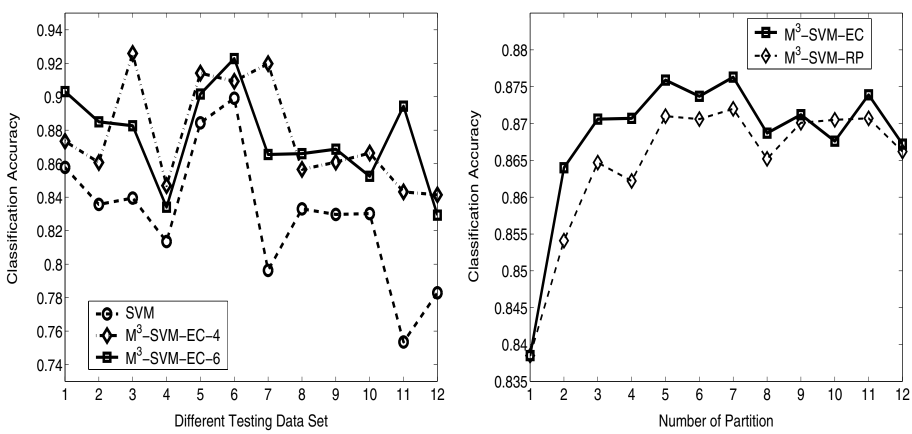
In all figures below, 'RP' means random partition method, and 'EC' means equal clustering method. For convenience, M^3-SVM-RP denotes min-max modular support vector machine using random partition strategy, while M^3-SVM-EC denotes the one using equal clustering strategy.

Fig. 1. A subject face instance. Six of poses including looking middle with 0 degree, looking up with 0 degree, looking up with left 22 degrees, looking up with right 22 degrees, looking middle with right 22 degrees and looking down with right 22 degrees.

4.2 Experimental Results

From Fig.2(a), we can see the advantage of M^3-SVM-EC over traditional SVM. First of all, M^3-SVM-EC can improve the classification accuracy by a high degree on gender recognition problem. In Table 1 we only present two division examples. The results of these two M^3-SVMs divided into 4 (M^3-SVM-EC-4) and 6 parts (M^3-SVM-EC-6) respectively, are both quite better than traditional SVM classifiers. Furthermore, from the results we can also see that the number of partition can't determine the final result. Meanwhile, Fig.2(b) illustrates that M^3-SVM with equal clustering method takes higher generalization accuracy than M^3-SVM with random partition method in most of time.

Moreover, as the equal clustering can make each modular have nearly same number of samples, M^3-SVM can run in a more efficient way than that with random partition method. From Fig.3(a) we can see that M^3-SVM with equal



(a) (b)

Fig. 2. The comparative results of M^3-SVMs based on equal clustering method and SVMs and M^3-SVMs based on random partition method. (a) Results of SVM and M^3-SVM divided into 4 and 6 parts with equal clustering method; (b) Results of M^3-SVMs with random partition and equal clustering method on total test data set.

Table 1. The problem statistics and the results based on different classifiers. The name of test data set is composed of three parts: 'POSE' denotes the face pose; 'PD', 'PM', and 'PU' denote person facing down, middle and up, respectively; the number denotes the pose angle.

Test Data	Female	Male	Total	Train	Test	SVM	M^3-SVM-EC-4	M^3-SVM-EC-6
POSE_PD00	445	595	1040	400	640	85.78%	87.34%	**90.31%**
POSE_PD15	846	1032	1878	600	1278	83.57%	86.07%	**88.50%**
POSE_PD22	44	158	202	40	162	83.95%	**92.59%**	88.27%
POSE_PD30	846	1032	1878	800	1078	81.35%	**84.69%**	83.40%
POSE_PM00	445	595	1040	400	640	88.44%	**91.41%**	90.16%
POSE_PM15	844	1032	1876	400	1476	89.91%	90.92%	**92.28%**
POSE_PM22	44	158	202	40	162	79.63%	**91.98%**	86.55%
POSE_PM30	844	1032	1876	600	1276	83.30%	85.66%	**86.60%**
POSE_PU00	445	595	1040	400	640	82.97%	86.09%	**86.87%**
POSE_PU15	445	595	1040	400	640	83.02%	**86.64%**	85.25%
POSE_PU22	44	158	202	60	142	75.35%	84.32%	**89.44%**
POSE_PU30	846	1032	1878	800	1078	78.29%	**84.14%**	82.93%

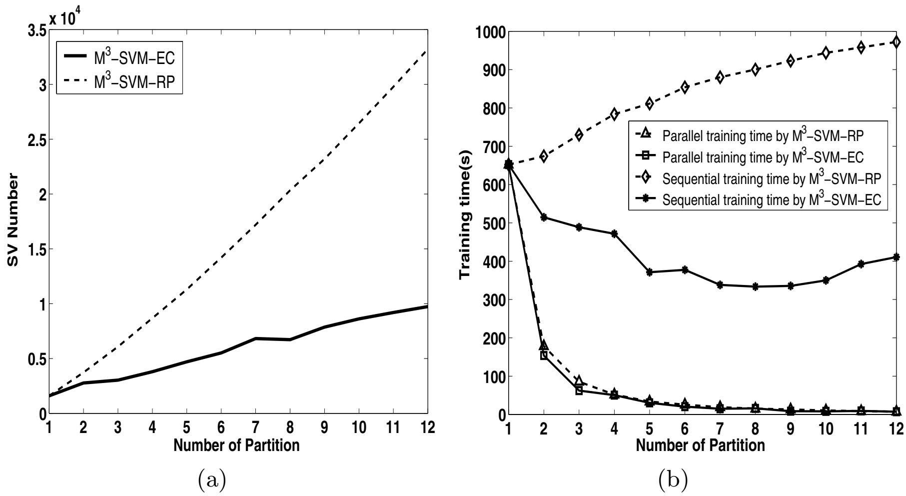


Fig. 3. The time performance of M^3-SVMs using equal clustering method and random partition method.(a)The number of SVs produced;(b)The training time of two task decomposition methods.

clustering method generates less support vectors than that with random partition method does, because it can make training data more separable. Actually, the smaller the number of support vectors is, the less the training time and the cost of realization of M^3-SVM are. Even though the data partitioning costs more preprocessing time, M^3-SVM-EC's less support vectors and more generalization accuracy can compensate for it and it takes less training time than M^3-SVM-RP does in both sequential and parallel modes as shown in Fig.3(b).

5 Conclusion

In this paper, M^3-SVMs with the equal clustering method are applied to solve gender recognition problem and some comparisons have been done to show the advantage of this method over traditional SVMs and M^3-SVMs with random partition method. From our experiments it can be seen that M^3-SVMs with equal clustering can improve the performance of gender recognition and be more efficient.

Acknowledgements

The authors thank Yi-Min Wen for providing the source programme of equal clustering algorithm.

References

1. Moghaddam, B., Yang, M.H.: Gender Classification with Support Vector Machines. In:Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, (2000) 306-311
2. Lu, B.L., Wang, K.A., Utiyama, M., Isahara, H.: A Part-versus-part Method for Massively Parallel Training of Support Vector Machines. In: Proceedings of IJCNN'04, Budapast, (2004) 735-740
3. Wang, K.A., Lu, B.L.: Task Decomposition Using Geometric Relation for Min-Max Modular SVM. In: ISNN 2005: Lecture Notes in Computer Science, vol.3496. Springer-Verlag, Berlin Heidelberg (2005) 887-892
4. Lian, H.C., Lu, B.L., Takikawa, E., Hosoi, S.: Gender Recognition Using A Min-max Modular Support Vector Machine. In: ICNC 2005: Lecture Notes in Computer Science, vol.3611. Springer-Verlag, Berlin Heidelberg (2005) 438-441
5. Wen, Y.M., Lu, B.L., Zhao, H.: Equal Clustering Makes Min-Max Modular Support Vector Machines More Efficient. In: ICONIP'2005, Taipei, (2005)
6. Lu, B.L., Ito, M.: Task Decomposition and Module Combination Based on Class Relations: A Modular Neural Network for Pattern Classification. IEEE Transcations on Neural Network 10(5) (1999) 1244-1256
7. Choudhury, A., Nair, C.P., Keane, A.J.: A Data Parallel Approach for Large-Scale Gaussian Process Modeling. In: Proceedings of the Second SIAM International Conference on Data Mining, (2002)
8. Gao, W., Cao, B., Shan, S.G., Zhou, D.L., Zhang, X.H., Zhao, D.B.: The CAS-PEAL Large-Scale Chinese Face Database and Evaluation Protocols, Technical Report No. JDL_TR_04_FR_001, Joint Research & Development Laboratory, CAS, 2004