Image Classification Based on Support Tensor Machine

Xue Wang [1], Qingshan Wang [2], Yanyan Chen [1,*]

[1] College of Applied Science and Technology, Beijing Union University, Beijing, China
[2] Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing, China
* Correspondence: yykjtyanyan@buu.edu.cn

Abstract: In the traditional machine learning field, most classical machine learning algorithms are designed based on vector space data. However, in the problem of image classification, the data is directly expressed in tensor representation. In recent years, machine learning algorithms that using tensor as input data directly have drawn extensive attention from researchers and have achieved certain results. Direct use of tensor as input data can effectively maintain the structural information of the data, so that the information contained in the data can be fully utilized to improve the recognition effect; at the same time, the correlation model and algorithm with tensor as input data can effectively reduce the number of decision variables to be solved in the optimization problem, thus avoiding the problems such as over-fitting which are easy to occur in the traditional vector model during the learning process, which makes the tensor algorithm is especially suitable for high dimensional and small sample size problems. This paper mainly studies the two-class model and learning algorithm based on tensor theory and applies it to image classification problems. The Experimental results on human face images show that the classification accuracy of support tensor machine is better than that of support vector machine.

Keywords: tensor representation; support tensor machine; image classification

1. Introduction

In machine learning problems, the representation of the data is very important. It is a hot issue in the field of machine learning in recent years for data from different sources to be expressed in a form that is more conducive to application. In traditional machine learning methods, it is usual to use vectors as input data basically. However, in practical problems, many data need to be in tensor form for more accurate representation. Therefore, the new method of machine learning based on tensor data has been widely studied and applied, which has become a new research direction in the field of data mining [1,2].

As a machine learning method based on tensor data, the Support Tensor Machine (STM) was first proposed by Cai et al. in 2006 [3]. In recent years, great progress has been made in the research of its related theories and learning models, which has become a powerful means to solve the machine learning problem based on tensor data. The tensor-based machine learning method has the following two advantages compared with the vector-based machine learning method. The first is that the spatial and temporal information of the data is preserved to make full use of the information contained in the data. The second is that it is more suitable to solve the high dimensional and small sample size problems. Therefore, the machine learning method based on tensor can be well applied to the classical high dimensional and small sample size problems, such as face recognition and other image classification problems. Therefore, the machine learning method based on tensor can be well applied to the classical high dimensional and small sample size problems, such as face recognition and other image classification problems. In recent years, the research on tensor representation has received extensive attention [4-9]. At the same time, it has been successfully applied in related practical applications, including image classification, face recognition, scene classification, bioinformatics and so on [10-15]. It can be said that research work on tensor representation has become a hot topic in related research fields.

Since images can be naturally represented by tensor data, this paper studies machine learning models and algorithms based on tensor representation and applies them to image classification problems. Numerical experiment results show that the classification accuracy of the support tensor machine is higher than that of the traditional vector machine model, especially for the high dimensional and small sample size problem, the advantages of STM are more obvious.

2. Support Tensor Machines

2.1. STM Based on Optimal Projection

In 2006, Cai et al. extended the support vector machine to the tensor space model under the framework of Supervised Tensor Learning (STL). In tensor space, the linear discriminant function is:

$$f(\mathbf{X}) = \mathbf{u}^T \mathbf{X} \mathbf{v} + b, \quad \mathbf{u} \in R^{n_1}, \mathbf{v} \in R^{n_2} \tag{1}$$

Based on the classification idea of maximum interval, the decision hyperplane separates the two types of samples in the tensor space as far as possible, and the corresponding optimization problem is as follows:

$$\min_{w,b,\xi} \frac{1}{2}\left\|\mathbf{u}\mathbf{v}^T\right\|^2 + C\sum_{i=1}^{m}\xi_i$$
$$subject\ \ to\ \ y_i(\mathbf{u}^T\mathbf{X}_i\mathbf{v}+b) \ge 1-\xi_i \tag{2}$$
$$\xi_i \ge 0,\quad i=1,\dots,m.$$

By introducing positive Lagrangian multipliers $\alpha_i, \beta_i \ge 0, i=1,\dots,m$, the Lagrangian function of optimization problem (2) is:

$$L(\mathbf{u},\mathbf{v},b,\xi,\alpha,\beta)=\frac{1}{2}\left\|\mathbf{u}\mathbf{v}^T\right\|^2 + C\sum_i \xi_i - \sum_i \alpha_i y_i(\mathbf{u}^T\mathbf{X}_i\mathbf{v}+b-1+\xi) - \sum_i \beta_i\xi_i \tag{3}$$

According to KKT conditions, the partial derivatives of L with respect to $\mathbf{u}, \mathbf{v}, b, \xi_i$, are set to zero, and the following formula can be obtained:

$$\mathbf{u} = \frac{\sum_i \alpha_i y_i \mathbf{X}_i \mathbf{v}}{\mathbf{v}^T\mathbf{v}} \tag{4}$$

$$\mathbf{v} = \frac{\sum_i \alpha_i y_i \mathbf{u}^T \mathbf{X}_i}{\mathbf{u}^T\mathbf{u}} \tag{5}$$

$$\sum_i \alpha_i y_i = 0 \tag{6}$$

$$C-\alpha_i-\beta_i=0,\ i=1,\dots,m \tag{7}$$

From (4) and (5), it can be seen that $\mathbf{u}$ and $\mathbf{v}$ are dependent on each other, and cannot be solved independently. Therefore, Cai et al. introduced an alternative projection algorithm and proposed a simple and effective method to solve this optimization problem.

Firstly fix $\mathbf{u}$ and let $\mu_1 = \left\|\mathbf{u}\right\|^2$, $\mathbf{x}_i = \mathbf{X}_i^T\mathbf{u}$. Equations (2) can be optimized as follows:

$$\min_{w,b,\xi} \frac{1}{2}\mu_1\left\|\mathbf{v}\right\|^2 + C\sum_{i=1}^{m}\xi_i$$
$$subject\ \ to\ \ y_i(\mathbf{v}^T\mathbf{x}_i+b) \ge 1-\xi_i \tag{8}$$
$$\xi_i \ge 0,\quad i=1,\dots,m.$$

Obviously, optimization problem (8) is a standard support vector machine optimization problem, Therefore, using the SVM algorithm can solve it. Similarly, While $\mathbf{v}$ is calculated, let $\mu_2 = \left\|\mathbf{v}\right\|^2$ and $\tilde{\mathbf{x}}_i = \mathbf{X}_i\mathbf{v}$, $\mathbf{u}$ can be solved according to the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}\mu_2\left\|\mathbf{u}\right\|^2 + C\sum_{i=1}^{m}\xi_i$$
$$subject\ \ to\ \ y_i(\mathbf{u}^T\mathbf{x}_i+b) \ge 1-\xi_i \tag{9}$$
$$\xi_i \ge 0,\quad i=1,\dots,m.$$

Accordingly, you can get $\mathbf{v}$ and $\mathbf{u}$ by solving the optimization problems (8) and (9) iteratively.

2.2. Algorithm of STM

Input: The training sample $\mathbf{X}_i \in R^{n_1} \otimes R^{n_2} (i=1,\dots,m)$, parameter $\mathbf{u}$, testing samples $\mathbf{Xt}_j \in R^{n_1} \otimes R^{n_2} (j=1,\dots,t)$.

Output: The optimal parameter variable in the decision function $\mathbf{u} \in R^{n_1}$, $\mathbf{v} \in R^{n_2}$ and b, the class label of the testing sample.

Step 1. Initialization: let $\mathbf{u} = (1,\dots,1)^T$;

Step 2. Calculate $\mathbf{v}$. Let $\mu_1 = \left\|\mathbf{u}\right\|^2$, by solving optimization problems:

$$\min_{\alpha} \frac{1}{2\mu_1}\sum_{i,j=1}^{m}\alpha_i\alpha_j\mathbf{u}^T(\mathbf{X}_i \cdot \mathbf{X}_j)\mathbf{u}$$
$$s.t.\ \ 0 \le \alpha_i \le \frac{1}{vl} \tag{10}$$
$$\sum_{i=1}^{l}\alpha_i = 1, i=1,\dots,m$$

Get the optimal solution α^* and calculate:

$$\mathbf{v} = \frac{1}{\mu_1}\sum_{i=1}^{l}\alpha_i^*\mathbf{X}_i^T\mathbf{u} 。$$

Step 3. Calculate $\mathbf{u}$. Let $\mu_2 = \left\|\mathbf{v}\right\|^2$, through solving optimization problems：

$$\min_{\alpha} \frac{1}{2\mu_2}\sum_{i,j=1}^{m}\hat{\alpha}_i\hat{\alpha}_j\mathbf{v}^T(\mathbf{X}_i \cdot \mathbf{X}_j)\mathbf{v}$$
$$s.t.\ \ 0 \le \hat{\alpha}_i \le \frac{1}{vl} \tag{11}$$
$$\sum_{i=1}^{l}\hat{\alpha}_i = 1, i=1,\dots,m$$

Get the optimal solution $\hat{\alpha}^*$ and calculate:

$$\mathbf{u} = \frac{1}{\mu_2}\sum_{i=1}^{l}\hat{\alpha}_i^*\mathbf{X}_i\mathbf{v} 。$$

Step 4. Repeat step 2~3 until the termination condition is met: the maximum number of iterations is reached, or the convergence condition is satisfied:

$$\left\|\mathbf{u}_i - \mathbf{u}_{i-1}\right\| \le tolerance \tag{12}$$

Step 5. Get the discriminant function and calculate the class label of the test sample:

$$f(\mathbf{Xt}_j) = \mathrm{sgn}(\mathbf{u}^T\mathbf{Xt}_j\mathbf{v}+b) \tag{13}$$

Step 6. End.

3. Experimental Results

In order to verify the effectiveness of STM algorithm in image classification problem, this section focuses on the tensor data of the real world: face image classification. Taking the face image of the ORL public dataset as the object, the classification accuracy of the STM and SVM algorithms is compared, and the number of training samples with the performance of the algorithms are evaluated.

3.1. Experiment Preparation

The ORL dataset consists of 40 people's face pictures (as shown in Figure 1). Each person has 10 different images, each of which is a 23×28 256-level grayscale image, all of which are directly stored as 2nd-order tensors. Firstly, make the face images in ORL database normalized, and all features are normalized to [0,1]. Since the experiment focused on the classification performance of the tensor algorithm in the high dimensional and small sample size problems, no cropping, compression or other operations to reduce the number of features were performed on the data.

In this paper, the classification accuracy ACC (the number of correctly classified samples/the number of total classified samples) is used as the evaluation index to evaluate the performance of the classifier. STM and SVM algorithms involve a parameter: penalty factor C, so this paper adopts the idea of 10-fold cross validation to select the optimal parameter.

The final result is expressed as the average value of 10-fold cross validation, and the parameter selection range is: $C=\{2^{-8},2^{-7},2^{-6},2^{-5},2^{-4},2^{-3},2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}, 2^{3}, 2^{4}, 2^{5}, 2^{6}, 2^{7}, 2^{8}\}$. The implements of all the Numerical experiments are in MATLAB 7.0 version on a PC, whose operating system is Microsoft Windows 8.1 professional x64 and processor's memory capacity is 2GB.

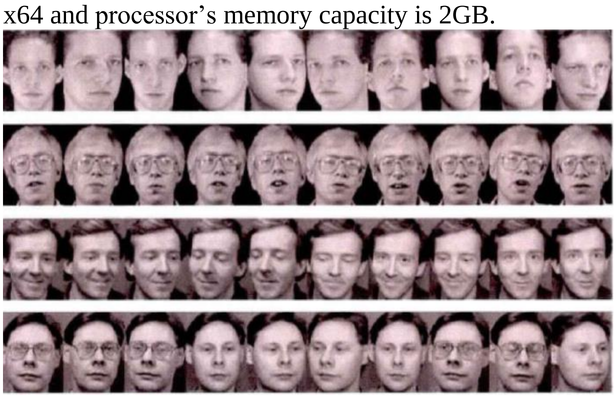

Figure 1. Diagram of tensor data.

3.2. Analysis of Experimental Results

The face image data is directly stored in the form of tensors. In the traditional vector algorithm, the conversion process from tensor data to vector data is easy to lose the structural information of data, resulting in the incompleteness of data features and affecting the recognition rate of the image. The algorithm based on the support tensor machine is applied to the image classification problem, and the tensor is directly used as the input data, which avoids the loss of data structure information. Therefore, this experiment compares the parameters of STM and SVM to compare their advantages and disadvantages.

Since STM and SVM are both two-class machine learning models, two categories are randomly selected for each experiment as positive and negative training samples. For simplicity of description, Table 1 lists 10 randomly selected training samples. For example, (39, 29) represents that the current set of experiments consists of two images of people numbered 39 and 29, in which 39 is a positive sample and 29 is a negative sample. Since the tensor representation algorithm is more suitable for the high dimensional and small sample size problems, in order to verify the effectiveness of the algorithm, This experiment selected the small sample training set, which contains the number of training samples of 4, 6, 7, 10 and 12 respectively. In order to ensure the reliability of the results, the training set was divided randomly for 10 times, and the final results were expressed as the average of 10 experiments.

Table 1 shows the average classification accuracy of the STM and SVM algorithms on different training sets when the training sample is 4, and the highest value is marked in bold. It can be clearly seen that in the 10 groups of randomly generated experiments, 7 groups of data show that the classification accuracy of SVM is lower than that of STM, and the other 3 groups of data show that the classification accuracy of STM is lower than that of SVM. It can been see from the Table 1 that for the classification of small sample images, the classification accuracy of STM algorithm is higher than that of SVM algorithm.

Table 1. Average ACCs of STM and SVM in different training sets when the number of training samples is 4.

Training Set	**Algorithm**	**ACC**	**Training Set**	**Algorithm**	**ACC**
(39,29)	STM	**0.99167**	(3,4)	STM	**0.91233**
	SVM	0.9375		SVM	0.78419
(36,13)	STM	**0.91897**	(10,2)	STM	0.9875
	SVM	0.91882		SVM	**1**
(21,10)	STM	**0.90995**	(2,19)	STM	**1**
	SVM	0.9099		SVM	**1**
(18,33)	STM	0.90915	(1,18)	STM	**0.9375**
	SVM	**0.90916**		SVM	**0.9375**
(35,37)	STM	**0.89521**	(18,19)	STM	**1**
	SVM	0.86743		SVM	**1**

Due to limited space, This paper uses the three experimental results (39, 29), (35, 37) and (3, 4) to demonstrate the relationship between the classification accuracy of the two algorithms and the number of training samples, as shown in Fig. 2-4. It can be seen when the number of training samples is small, the classification accuracy rate of STM is obviously higher than that of SVM. As the number of samples increases, the SVM classification accuracy rate is roughly similar to the SVM classification accuracy rate. This also shows that the classification algorithm based on tensor representation is particularly suitable for high dimensional and small sample size problems.

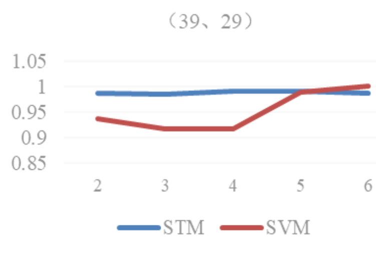

Figure 2. The ACCs of STM varies with the number of training samples on training set (39, 29).

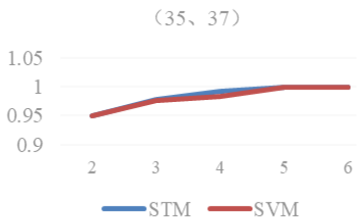

Figure 3. The ACCs of STM and SVM varies with the number of training samples on training set (35, 37).

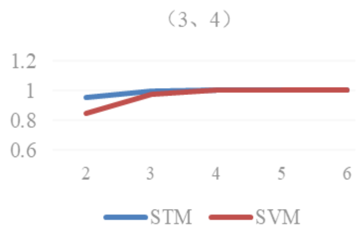

Figure 4. The ACCs of STM and SVM varies with the number of training samples on training set (3, 4).

4. Conclusions and Future Work

The machine learning algorithm based on tensor theory has received extensive attention in recent years, which is mainly because the learning method using tensor as input data directly has the following advantages over the traditional vector learning algorithm: Direct use of tensors as input data can effectively maintain the structural information of the data, so that the information contained in the data can be fully utilized to improve the recognition effect; The relevant model and algorithm with tensor as input data can effectively reduce the number of decision variables to be solved in the optimization problems, so as to avoid the over-fitting problem that is prone to occur in the learning process of traditional vector model, which makes the tensor algorithm especially suitable for the high dimensional and small sample size problems.

This paper mainly studies the classification algorithm based on tensor data and its application in image classification. Experiments were carried out using the ORL face image datasets to verify the advantages of support tensor machine in the image classification problems. The greatest advantage of tensor representation based learning algorithm is that it can maintain the integrity of tensor data, especially for the high dimensional and small sample size problems, the classification accuracy of support tensor machine is better than that of support vector machine.

Although the number of decision variables that the classifier based on tensor theory needs to solve is far less than the vector classifier, however, the tensor model used in this paper is based on the STL framework, which uses the alternating projection algorithm. The training time depends on the iteration times of the alternating projection algorithm, which is often much higher than the vector model. Therefore, it is an urgent problem to optimize the algorithm of classification model based on tensor and improve the training efficiency. In addition, the in-depth study of tensor kernel method enables the support tensor machine model to be used for nonlinear classification problems, and further research is needed.

References

[1] D. Tao, X. Li, W. Hu. Supervised tensor learning. *Knowledge and Information Systems*, **2015**, 13(1): 450-457.

[2] D. Tao, X. Li, X. Wu. Supervised tensor learning. *Knowledge and Information Systems*, **2007**, 13(1): 1-42.

[3] D. Cai, X. He, J. Han. Learning with tensor representation. UIUCDCS-R-2006-2716. Department of Computer Science, University of Illinois at Urbana-Champaign, 2006.

[4] D. Cai, X. He, J.R. Wen, J. Han, W.Y. Ma. Support Tensor Machines for Text Categorization. *International Journal of Academic Research in Business and Social Sciences*, **2002**, 2(12): 2222-6990.

[5] I. Kotsia, W. Guo and I. Patras. Higher rank Support Tensor Machines for visual recognition. *Pattern Recognition*, **2012**, 4192-4203.

[6] X. Zhang, X. Gao, Y. Wang. Twin Support Tensor Machines for MCS Detection. *Journal of Electronics (China)*, **2009**, 26(3): 318-325.

[7] Y.Y. Chen, K.N. Wang, P. Zhong. One-class support tensor machine. *Knowledge-Based Systems*, **2016**, 96: 14-28.

[8] Y.Y. Chen, K.N. Wang, P. Zhong. One-class support higher order tensor machine classifier. *Applied Intelligence*, **2017**, 7: 1-9.

[9] L. He, X. Kong, P.S. Yu. Dusk: A dual structure-preserving kernel for supervised tensor learning with applications to neuroimages. *Proc. SIAM Int. Conf. Data Min.*, **2014**, 127-135.

[10] B.K. Cao, L.F. He, X.K. Wei, M.Q. Yu, S. Philip, Klumpp, Heide, Leow, D. Alex. T-BNE: Tensor-based brain network embedding. In SDM, 2017.

[11] Cichocki, Andrzej, Mandic, Danilo, De Lathauwer, E.Li, G.U. Zhou, Q.B. Zhao, Caiafa, Cesar, Phan, Huy Anh. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, **2015**, 32(2): 145-163.

[12] T.J. Guo, L. Han, L.F. He, X.W. Yang. A ga-based feature selection and parameter optimization for linear support higher-order tensor machine. *Neurocomputing*, **2014**, 144: 408–416.

[13] C.T. Lu, L.F. He, W.X. Shao, B.K. Cao and S. Philip. Multilinear factorization machines for multi-task multi-view learning. In WSDM, 2017, 701–709, ACM.

[14] B. Krawczyk, M. Galar, M. Woźniak. Dynamic ensemble selection for multi-class classification with one-class classifiers. *Pattern Recognition*, **2015**, 83: 34-51.

[15] R. Pilarczyk, X. Chang, W. Skarbek. Human Face Expressions from Images-2D Face Geometry and 3D Face Local Motion versus Deep Neural Features. Unpublished, 2019.