A Fuzzy Support Tensor Machines based on Support Vector Data Description

Ruiting Zhang[1]

Canvard college, Beijing Technology and Business University, 101118, Beijing ,China
E-mail: `rt_zhang@163.com`

Yuting Kang[2]

Canvard college, Beijing Technology and Business University, 101118, Beijing ,China
E-mail: `314174534@qq.com`

Most of the traditional machine learning algorithms are based on the vector, but in tensor space, Tensor learning is helpful to overcome the over-fitting problem than vector learning. In the meanwhile, these algorithms based on tensor require a smaller set of decision variables as compared to those approaches based on vector. Support tensor machine (STM) is a prevalent machine learning approach, and STM is widely applied to practical many problems. However, STM is sensitive to noise or outliers in the sample dataset. Fuzzy support tensor machine (FSTM) can partly overcome this shortcoming by assigning different fuzzy membership to different training samples. How to choose the appropriate fuzzy memberships is one of the important things for machine learning. At present, most of people define a fuzzy membership by introducing the distance between each point and its class centre. In this paper, We present a novel methods to compute fuzzy memberships based on support vector data description (SVDD), our experimental results on ORL database and Yale database show that the improved method can reduce the effect of outliers and improve the rate of classification accuracy.

---

[1]Speaker; Correspongding Author

1.Introduction

Nowadays, with the development of our society, a tremendous amount of data have continuously flooded into our views. The explosive growth of data has generated an urgent need for new techniques and new skills in the data mining areas. Machine learning is a important branch of data mining. One of the core tasks of machine learning is how to represent of data. High dimensional data is often encountered in the practical applications, Therefore, how to efficiently represent the image data has a fundamental effect on the classification. Most of the traditional learning algorithms are based on the vector space model, such as SVM[1].

However, in practice, like the gray images sequence and the gray images, a lot of objects need to be expressed in tensor, For example, the gray image (Figure.1) is the second order tensor. The color image is the third order tensor, with aspects consisting of column, row, and color. Therefore, human face recognition has received significant attention, and the robust face recognition systems is currently one of the most activity areas of research. Transforming the vector patterns to the corresponding matrix patterns before classification is very important to represent the images appropriately. In this way, it may destroy the data structural information or lead to high dimensional vectors. In recent years, because of the drawbacks, the machine learning algorithms based on tensor space has attracted great interest in the research areas. A set of algorithms have been extended to deal with tensors, for example, support tensor machine (STM) [2-6]. With using the tensor representation, the number of parameters estimated by the tensor-based learning can be greatly reduced. Therefore, the tensor-based learning algorithms are especially suitable for solving the small-sample-size (S3) problem, which the number of samples available for training is small and the number of input features used to represent the data is large. At the same time, involving high-dimensional data can also reduce the computational complexity observed in problems.

Figure 1: A gray image can be represented by a matrix

However, in many real applications areas, the training data is often contaminated by those outliers or noise. If we treat equally every sample may cause over-fitting problem. Moreover, some sample points are misplaced on the wrong side by accident. Fuzzy support vector machine (FSVM) [7-9] provides an effective method to deal with this problem how to classify the corrupted data correctly. In FSVM, different training sample points are associated with different memberships which represent different contributions to the learning of decision surface. It can partly reduce the effects of outliers by fuzzy membership functions. In other words, train sampling point with a lower membership value can be thought of as less representative of that class, while those with a higher membership value should be given more importance.

At present, one of the important things is to choose the appropriate fuzzy memberships. Considering the distance between each point and its class center, most of people define a fuzzy membership basing on the distance. In this paper, We present a novel methods to compute fuzzy

memberships based on support vector data description (SVDD). Experimental results reveal that the improved algorithm based on SVDD is more robust than the traditional STM and FSTM .

The rest of the paper is organized as follows: Section 2 provides an overview of Fuzzy membership and support vector data description (SVDD), In Section 3 we present a novel Fuzzy support tensor machine based on support vector data description (SVDD) algorithm, Section 4 discusses the experimental result, and we give a conclusion in Secton5.

2. Fuzzy Membership and Support Vector Data Description

2.1 Fuzzy Membership

In many real-word application community, due to over-fitting problem in SVMs, the training process is particularly sensitive to those sample points which are far away from their own class in the training dataset. It is very important to assign each data point in the training dataset with a membership in order to decrease the effect of those outliers or noises. Samples with a higher membership value can be thought of as more representative of that class, while those points with a lower membership value should be given less importance, so that those abnormal data with a low membership contribute to total error term decreases. In fact, this fuzzy membership value determines how important it is to classify a data sample correctly.

Different fuzzy membership functions have different effects on the FSVM or FSTM algorithm, At present, a prevalent methods for most people is to define a fuzzy membership basing on the distance between each point and its class center[9-10]. However, both of them ignore the affinity among sample points, which results in failing to differentiate valid samples from outliers or noise. for example, Figure 2 shows the affinity among sample points at two different classes. From Figure2 we can see that the distance between the point x and its class center in Fig.2(a) and Fig.2(b) is same. However, compared with x in Fig.2(b), the sample point x in Fig.2(a) is more like a valid sample point than outlier or noise. Motivated by the design of fuzzy membership[11], besides considering the distance, we should think about the affinity among training dataset and define the following new fuzzy membership function to reflect the importance of each training sample point for its own class more accurately.
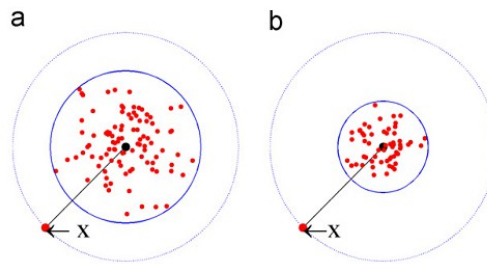
Figure 2: the difference of the affinity among sample points from two different classes

2.2 Support Vector Data Description

The support vector data description (SVDD) algorithm is developed by Tax and Duin to solve one-class classification problems[12]. Given a set of training samples $\{x_i, i = 1, 2, \cdots, l\}$. The mail idea of SVDD is to find a hypersphere of radius $R > 0$ and center a with a minimum volume containing most the target class samples. Then we can solve the following optimization problem:

$$\min \quad R^2 + C\sum_{i=1}^{l}\xi_i$$
$$\text{s.t.} \quad \|\Phi(x_i) - a\|^2 \le R^2 + \xi_i, \xi_i \ge 0, i = 1, 2, \cdots, l$$

(2.1)

Where ξ_i is a slack variable, C is a regularization parameter controls the tradeoff between the volume of the hepersphere and the errors, $\Phi(x)$ is a nonlinear function which maps x_i to a high dimensional feature space, By introducing Lagrange multiples and Applying the KKT conditions we solve the following dual problem:

$$\max \quad \sum_{i=1}^{l}\alpha_i K(x_i, x_i) - \sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j K(x_i, x_j)$$
$$\text{s.t.} \quad \sum_{i=1}^{l}\alpha_i = 1, 0 \le \alpha_i \le C, i = 1, 2, \cdots, l$$

(2.2)

Then we can compute R by any Support vector x_i with $\alpha_i > 0$:

$$R^2 = K(x_i, x_i) - 2\sum_{j=1}^{l}\alpha_j K(x_i, x_j) - \sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j K(x_i, x_j)$$

(2.3)

2.3 A New Fuzzy Membership Function

Based on the affinity among training sample points, by introducing the support vector data description, the fuzzy membership s_i is given by the following formula:

$$s_i = \begin{cases} 0.6 \times \dfrac{1 - d(x_i)/R}{1 + d(x_i)/R} + 0.4, & d(x_i) \le R \\ \\ 0.4 \times \dfrac{1}{1 + d(x_i) - R}, & d(x_i) > R \end{cases}$$

(2.4)

Where $R > 0$ is the radius of class, $d(X_i)$ is the distance between sample point X_i and its center.

3. Fuzzy Support Tensor Machine based on SVDD

3.1 Reformulate FSTM

Given the training samples: $\mathcal{T} = \{(X_1, y_1, s_1), \cdots, (X_l, y_l, s_l)\} \in (\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \times \mathcal{Y} \times s)^l$. The classification problem is modeled by the following programming:

$$\min_{u,v,b,\xi} \quad \frac{1}{2}\left\|uv^{\mathrm{T}}\right\|_{\mathrm{F}}^2 + C\sum_{i=1}^{l}s_i\xi_i,$$
$$\text{s.t.} \quad y_i(u^{\mathrm{T}}X_i v + b) \ge 1 - \xi_i, \xi_i \ge 0, i = 1, \cdots, l.$$

(3.1)

where $X_i \in \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ is a training sample, $y_i = \{+1, -1\}$ is the class label of X_i, and s_i [0,1] is the fuzzy membership

degree of X_i belonging to y_i, $C > 0$ is a regularization parameter and ξ_i is a slack variable, For solving the problem (3.1), we construct the Lagrangian function as follows:

$$L(u,v,b,\xi,\alpha,\beta) = \frac{1}{2}\left\|uv^{\mathrm{T}}\right\|_F^2 + C\sum_{i=1}^{l} s_i \xi_i - \sum_{i=1}^{l} \alpha_i [y_i(u^{\mathrm{T}} X_i v + b) - 1 + \xi_i] - \sum_{i=1}^{l} \beta_i \xi_i \tag{3.2}$$

The KKT necessary conditions for the optimality are:

$$u = \frac{\sum_{i=1}^{l} \alpha_i y_i X_i v}{v^{\mathrm{T}} v}, \tag{3.3}$$

$$v = \frac{\sum_{i=1}^{l} \alpha_i y_i u^T X_i}{u^{\mathrm{T}} u}, \tag{3.4}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0, \tag{3.5}$$

$$Cs_i - \alpha_i - \beta_i = 0, \quad i = 1, \cdots, l, \tag{3.6}$$

From equations (3.3) and (3.4), it is obviously that u and v rely on each other, and cannot be solved with traditional methods, so we can use alternating iterative algorithm[3,4]. We first fix u, Let $\beta_1 = \|u\|^2$, and $x_i = X_i^T u$, The optimization problem (3.1) is reduced to the following QPP:

$$\min_{v,b,\xi} \quad \frac{1}{2}\beta_1 v^T v + C\xi^T S\xi$$
$$\text{s.t.} \quad y_i(v^{\mathrm{T}} x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \ i = 1, \cdots, l. \tag{3.7}$$

It is clear that the new optimization problem (3.7) is identical to the standard FSVM optimization problem. Thus, we can use the same computational methods of FSVM to solve (3.7). Once v is obtained, Let $\beta_2 = \|v\|^2$, and $\tilde{x}_i = X_i v$, u can be computed by the following QPP:

$$\min_{v,b,\xi} \quad \frac{1}{2}\beta_2 u^T u + C\xi^T S\xi$$
$$\text{s.t.} \quad y_i(u^{\mathrm{T}} \tilde{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \ i = 1, \cdots, l. \tag{3.8}$$

Thus, u and v can be obtained by iteratively solving the optimization problems (3.7) and (3.8).

3.2 The Algorithm of FSTM

Inputs: the data set $\mathcal{T} = \{(X_1, y_1, s_1), \cdots, (X_l, y_l, s_l)\} \in (\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \times \mathcal{Y} \times s)^l$.

Outputs: the hyperplane $f(X) = \langle uv^T, X \rangle + b = 0$.

step1. Initialization: Let $u = (1, \cdots, 1)^T$, $\beta_1 = \|u\|^2$, and $x_i = X_i^T u$, $X = \left(x_1^T, x_2^T, \cdots, x_l^T\right)^T$;

step2. Computing v : v can be computed by solving the optimization problem (3.7);

step3. Computing u : let $\beta_2 = \|v\|^2$, and $\tilde{x}_i = X_i v$, $\tilde{X} = \left(\tilde{x}_1^T, \tilde{x}_2^T, \cdots, \tilde{x}_l^T\right)^T$, u can be computed by solving the optimization problem(3.8);

step4. Iteratively computing u and v : By step 2 and step 3, we can iteratively compute u and v until they tend to converge.

3.3 The Convergence analysis of FSTM algortihm

Since both the the optimization problems (3.7) and (3.8) are strictly convex, the solutions of FSTM are globally optimum, Specifically the solutions of equations (3.7) and (3.8) are globally optimum. Thus FSTM algorithm converges.

4. Experiments

In this section, in order to verify the effectiveness of the proposed FSTM, We compared the results of FSTM with the vector-based classification method FSVM and tensor-based classification method STM. All the algorithms have been implemented in MATLAB7.8 (R2009a) on Ubuntu running on a PC with system configuration Intel Core5 Duo (2.60GHz) with 2 GB of RAM. The data sets are taken from the UCI Repository. We randomly select our training images from the entire set and repeat the experiment 10 times. The parameter C is obtained by cross validation, and the range of the regularization constant C is from 2^{-6} to 2^6 with each step by multiplying 2. All experiments are primarily focused on the matrix data. We initialize both u and v are vectors of ones of appropriate dimensions.

In our experiments, we choose six database which are represent in Table 1. They are ORL face data1, ORL face data2, Yale face data1, Yale face data2 and Yale face data3. The AT&T laboratories in Cambridge provided the ORL database of the face images . It comprise 400 images of 40 individuals, with varying lighting, facial expression (closed or open eyes, smiling or not) and facial details (beard or gender, with glasses or no). The Yale database of the face images has been provided by Yale University. It contains 165 images about 15 individuals, where each class has 11 images. These images with facial expression (happy, sad, normal, sleepy, surprised, wink) and varying lighting condition (center-light, left-light). Take Yale face date3 for example, All images were normalized to a resolution of 100×100pixels with(10000)gray levels. One person facial expression from Yale database is displayed in Figure3. The detail information about the datasets is listed in Table 1.

Figure 3: Eleven facial samples from a subject within the Yale database

Data sets	Data size	Training size	Testing size	Class	Atrribute
ORL face data1	400	2	18	40	32×32
ORL face data2	400	2	18	40	64×64
Yale face data1	165	2	20	15	32×32
Yale face data2	165	2	20	15	64×64
Yale face data3	165	2	20	15	100×100

Table1: Description of data sets, from left to right are original data set name, data size, training size, testing size ,data class and data attribution.

The experimental results are summarized in Table2, where the recognition performance denotes the mean value of 10 times testing results and plus or minus the standard deviation. In other words, Table 2 shows the mean accuracy rates and standard deviations of all algorithms in our experiments deal with different database. From the Table2, it can be seen that FSTM based on SVDD outperforms FSVM and STM in most experiments.

Table2 The mean accuracy rates(%) and standard deviations of FSTM, FSVM and STM corresponding to the data sets(Black font denotes best Accuracy)

Data sets	Matrix Size	STM	FSVM	FSTM
ORL face data1	32×32	79.31±2.66	81.25±11.16	**82.55±2.81**
ORL face data2	64×64	81.15±5.36	**84.41±15.92**	83.26±3.25
Yale face data1	32×32	80.12±3.43	81.22±13.24	**82.49±3.37**
Yale face data2	64×64	82.07±5.42	86.38±11.29	**86.65±2.74**
Yale face data3	100×100	84.68±7.51	89.56±12.33	**90.27±3.63**

5. Conclusion

In this paper, considering the affinity among training samples, we firstly define the fuzzy membership of training sample points which to compute based on SVDD in tensor space, and propose an improved tensor-based method---FSTM algorithm to learn better from datasets in the presence of outliers or noises. For solving the small-sample-size (S3) problems, the tensor representation always performs better than the vector representation. This is due to the fact that the number of parameters estimated by STM is much less than that of estimated by standard SVM. The similar results hold true for the FSTM algorithm. The above several numerical experiments show that the tensor-based methods have more advantages than vector-based methods for small-sample-size (S3) problems.

References

[1]V. Vapnik, *The nature of statistic learning theory*. New York: Springer-Verlag, New York, pp1-150, (1995)

[2]I. Kotsia, W. W. Guo, I. Patras, *Higher rank support tensor machines for visual recognition*,Patter Recognition. 45,no.1,pp4192-4203 (2012).

[3]D. Cai, X. F. He, J. W. Han, *Learning with tensor representation*, Department of Computer Science Technical Report No.2716, University of Illinois at Urbana-Champaign (UIUCDCS-R-2006-2716), (2006)

[4]D. Cai, X. F. He, J. R. Wen, J. Han, W. Y. Ma, *Support tensor machines for text categorization*, Department of Computer Science Technical Report No.2714, University of Illinois at Urbana-Champaign (UIUCDCS-R-2006-2714), (2006)

[5]Tao D, Li X, Hu W, Maybank SJ, Wu X , *Supervised tensor learning*. Knowl Inf Syst,no.13, pp 450-457, (2005)

[6] R. Khemchandani, A. Karpatne, S. Chandra, *Proximal support tensor machines*, International Journal of Machine Learning and Cybernetics, no.4,pp703-712, (2013)

[7]C.F. Lin and S.D. Wang, *Fuzzy support vector machines*, IEEE Trans. on Neural Networks, vol. 13, no. 2, pp. 464-471, (2002)

[8]T. Inoue and S. Abe, *Fuzzy support vector machines for pattern classification*, Proceedings of IJCNN'01, Washington DC, vol. 2, pp. 1449-1454, (2001)

[9]H-P Huang and Y-H Liu, *Fuzzy support vector machine for pattern recognition and data mining*, International Journal of Fuzzy Systems, vol. 4, no. 3, pp. 826-835, (2002)

[10]Wenjuan An, Mangui Liang, *Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises*, Neurocomputing110, pp.101-110,(2013)

[11] X. Zhang, X. Xiao, G. Xu, *Fuzzy support vector machine based on affinity among samples*, J. Software 17 (5) (2006) 951–958.

[12] D. Tax, R.P. Duin, *Support vector domain description*, Pattern Recognition Lett. 20 (11–13) pp. 1191–1199, (1999)

PoS(ISCC2015)041