Knowledge-Based Systems 96 (2016) 14–28

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

One-Class Support Tensor Machine

Yanyan Chen [a,b], Kuaini Wang [c], Ping Zhong [a,*]

[a] College of Science, China Agricultural University, Beijing 100083, China
[b] College of Applied Science and Technology, Beijing Union University, Beijing 102200, China
[c] College of Science, Xi'an Shiyou University, Xi'an 710065, China

ARTICLE INFO

Article history:
Received 16 June 2015
Revised 30 December 2015
Accepted 4 January 2016
Available online 11 January 2016

Keywords:
Support vector machine
Support tensor machine
One-class classification
High dimensional and small sample size problem

ABSTRACT

In fault diagnosis, face recognition, network anomaly detection, text classification and many other fields, we often encounter one-class classification problems. The traditional vector-based one-class classification algorithms represented by One-Class Support Vector Machine (OCSVM) have limitations when tensor is considered as input data. This work addresses one-class classification problem with tensor-based maximal margin classification paradigm. To this end, we formulate the One-Class Support Tensor Machine (OCSTM), which separates most samples of interested class from the origin in the tensor space, with maximal margin. The benefits of the proposed algorithm are twofold. First, the use of direct tensor representation helps to retain the data topology more efficiently. The second benefit is that tensor representation can greatly reduce the number of parameters. It helps overcome the overfitting problem caused mostly by vector-based algorithms and especially suits for high dimensional and small sample size problem. To solve the corresponding optimization problem in OCSTM, the alternating projection method is implemented, for it is simplified by solving a typical OCSVM optimization problem at each iteration. The efficiency of the proposed method is illustrated on both vector and tensor datasets. The experimental results indicate the validity of the new method.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Problem statement

In many practical application fields, we often encounter one-class classification problems, such as fault diagnosis, face recognition, the network anomaly detection and text classification etc. In one-class classification problems, usually, only one class is available, and others are either expensive to acquire or difficult to characterize. Support Vector Machine (SVM) [1] had been extended to one-class classification by Schölkopf et al. [2], termed as One-Class Support Vector Machine (OCSVM). OCSVM seeks for a hyperplane in the maximal margin sense, which separates most of the samples from the origin. Another approach derived from SVM to solve the one-class classification problem is the Support Vector Domain Description (SVDD) [3]. SVDD maps the data into a feature space and consists in determining the minimum volume enclosing ball which contains almost all the samples. At present, there are lots of researches on one-class classification problems [4–7], and all of which are based on vector space.

In pattern recognition, machine learning, computer vision and image processing and other fields, many of the original objects are represented as multidimensional arrays, namely tensor. For example, gray face image [8] is represented as a second order tensor (or matrix); color image and grayscale video sequence [9], gait contour sequence [10] and hyper spectral cube [11] are usually expressed as third order tensor; color video sequence is usually expressed as fourth order tensor. The traditional one-class classification methods represented by OCSVM failed when tensor is considered as input data. Although there are some methods converting tensor directly into vector, they may lead to structural information loss and data correlation damage [12]. All these reasons lead us to consider tensor representations and corresponding learning algorithms for one-class classification problems.

1.2. Related work

Based on SVM learning framework and the ideas of alternating projection and multiple linear algebra operations, Tao and Wu processed Supervised Tensor Learning (STL) framework [13], which took tensor as input data. Tensor representation can effectively reduce the overfitting problem in the traditional learning methods of vector representation. Based on STL framework and its bidirectional optimal projection algorithm, Tao and Cai proposed Support

* Corresponding author. Tel.: +86 13651156748.
E-mail address: zping@cau.edu.cn (P. Zhong).

http://dx.doi.org/10.1016/j.knosys.2016.01.007
0950-7051/© 2016 Elsevier B.V. All rights reserved.

Tensor Machine (STM) [14]. Since the parameters to be solved in tensor based classifier are far less than those in vector-based classifier, STM is especially suitable for small sample size problem. The experiment results also show that classification accuracy of STM is superior to that of the traditional SVM.

In recent years, there have been more and more researches on evolving the vector-based learning algorithm to tensor representation, and all have obtained good experimental results and applications. Cai et al. presented a linear Tensor Least Square(TLS) classifier with second order tensor, and applied STM for text classification [15]. Tao et al. extended the classic linear C-SVM [16] and ν-SVM [17] to the general tensor forms [12]. Zhang et al. generalized the vector-based learning algorithm Twin Support Vector Machine (TWSVM) to the tensor-based method Twin Support Tensor Machine (TWSTM), and implemented the classifier for micro calcification clusters detection [18]. By comparison with TWSVM, the tensor version reduces the overfitting problem significantly. Khemchandani et al. developed a least squares variant of STM, termed as Proximal Support Tensor Machine (PSTM) [19]. Kotsia et al. formulated the higher rank Support Tensor Machines (STMs) in which the separating hyperplane was constrained to be the sum of rank one tensors [20]. Hao et al. presented a novel linear Support Higher-order Tensor Machine (SHTM) which integrated the merits of linear C-SVM and tensor rank-one decomposition [21]. SHTM solved the problem of time-consuming caused by STM and was particularly effective for higher-order tensor. On the similar line, Liu et al. extended the concave-convex procedure-based Transductive Support Vector Machine (CCCP-TSVM) to tensor pattern CCCP-based Transductive Support Tensor Machine (CCCP-TSTM), in which the tensor rank-one decomposition was used to compute the inner product of the tensors [22].

As to the nonlinear cases, there are several studies on kernel methods for tensors. Signoretto et al. elaborated on a possible framework to extend the flexibility of tensor-based models with kernel-based techniques, which gave a constructive definition to the feature space of infinite dimensional tensors [23]. He et al. introduced a new scheme to design structure-preserving kernels for supervised tensor learning [24]. For specified order of tensors, there are some even simple and convenient calculation kernel methods, i.e. matrix kernel function for second order tensor [25,26] and K_{3rd} kernel function for third order tensor [27].

All the relevant literatures show that the tensor representation of the data can remain the natural structure and correlation of the data, and avoid information missing, over-fitting and curse of dimensionality. Therefore, the studies of the tensor have attracted more and more attentions, and the researchers have extended the framework of the support vector machine to the tensor patterns and have proposed many tensor-based learning machines.

1.3. Research contribution

Utilizing the advantages of STL framework and one-class Support Vector Machine, we derive a tensor-based one-class classification algorithm, named One-Class Support Tensor Machine (OCSTM). The main idea of OCSTM is to find a hyperplane in tensor space (or tensor feature space), which can separate most samples of the interest class from the origin with the maximal margin principle. Since the matrix kernel function is relatively simple, we first present our OCSTM with second order tensor. Then we generalize a framework of OCSTM for high order tensors. Based on STL framework and its bidirectional optimal projection algorithm, the corresponding optimization problem in OCSTM is solved in an iterative manner, where at each iteration the parameters corresponding to the projections are estimated by solving a typical OCSVM optimization problem. This new model takes tensor as input data, so that it can utilize the structural information which is presented in multidimensional features of an object. Since tensor representation can greatly reduce the number of parameters estimated by OCSVM, our OCSTM helps overcome the overfitting problem encountered mostly in vector-based algorithms and especially suits for high dimensional and small sample size problem.

To evaluate the new algorithm, two kinds of datasets are considered: vector-based and tensor-based datasets. For vector-based datasets, since the vector data points need to be converted to tensor form, we first discuss on how to choose the proper tensor size. Then we test the classification performance of our OCSTM by comparison with OCSVM. Since tensor representation is particular suitable for high dimensional and small sample size cases, we detail the performance of each classifier with different small sample sizes. We also discuss on time cost and overfitting problem on vector-based dataset. The considered tensor-based dataset is human face image, which can be represented as a second order tensor and always be joined lines into vector in traditional vector learning algorithms. The experiments in this part indicate that tensor representation can reserve the important structure attributes of the image, which vector-based methods cannot qualify. All the experimental results indicate the validity and advantage of the new OCSTM.

The rest of this paper is organized as follows: In Section 2, we give a brief overview of Support Tensor Machine and the kernel function for tensor. The One-Class Support Tensor Machine, is described in Section 3. The experimental results on vector datasets and tensor datasets are presented in Sections 4 and 5. In Section 6, we generalize a framework of OCSTM for high order tensors. Finally, we provide some concluding remarks and suggestions for future work in Section 7.

2. Brief overview of Support Tensor Machine

2.1. Support Tensor Machine

Support Tensor Machine, which was developed by Cai and Tao [13,14], is a tensor generalization of SVM in the tensor space. Suppose the training samples $\{\mathbf{X}_i, y_i\} (i = 1, \ldots, l)$, where $\mathbf{X}_i \in \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ is the data point in 2nd-order tensor space, $y_i \in \{-1, 1\}$ is the class associated with $\mathbf{X}_i$, $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ are two vector spaces. STM tries to find the following linear classifier in the tensor space:

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

such that the two classes can be separated with maximal margin.

The optimization problem of linear STM can be stated as:

$$\begin{aligned} \min_{\mathbf{u}\in\mathbb{R}^{n_1},\mathbf{v}\in\mathbb{R}^{n_2},b\in\mathbb{R},\xi\in\mathbb{R}^l} \quad & \frac{1}{2}\|\mathbf{u}\mathbf{v}^T\|^2 + C\sum_{i=1}^{l}\xi_i \\ \text{s.t.} \quad & y_i(\mathbf{u}^T\mathbf{X}_i\mathbf{v} + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \ldots, l \end{aligned} \tag{2}$$

In order to solve the optimization problem (2), Cai and Tao describe a simple yet effective computational method. To fix $\mathbf{u}$ at first, let $\beta_1 = \|\mathbf{u}\|^2$ and $\mathbf{x}_i = \mathbf{X}_i{}^T\mathbf{u}$. Thus, the optimization problem (2) is identical to the standard SVM optimization problem with variable $\mathbf{v}$:

$$\begin{aligned} \min_{\mathbf{v}\in\mathbb{R}^{n_2},b\in\mathbb{R},\xi\in\mathbb{R}^l} \quad & \frac{1}{2}\beta_1\|\mathbf{v}\|^2 + C\sum_{i=1}^{l}\xi_i \\ \text{s.t.} \quad & y_i(\mathbf{v}^T\mathbf{x}_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \ldots, l \end{aligned} \tag{3}$$

It is clear that the optimization problem (3) can be solved by using the same computational methods of SVM. While $\mathbf{v}$ is obtained, let $\beta_2 = \|\mathbf{v}\|^2$ and $\widetilde{\mathbf{x}}_i = \mathbf{X}_i\mathbf{v}$, optimization problem (2) is the

Download English Version:

https://daneshyari.com/en/article/403455

Download Persian Version:

https://daneshyari.com/article/403455

Daneshyari.com