Neurocomputing 000 (2017) 1–9

Contents lists available at ScienceDirect

Neurocomputing

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

Online multilinear principal component analysis

Le Han [a], Zhen Wu [a], Kui Zeng [a], Xiaowei Yang [b,*]

[a] School of Mathematics, South China University of Technology, Guangzhou 510640, PR China
[b] School of Software Engineering, South China University of Technology, Guangzhou 510006, PR China

A R T I C L E I N F O

Article history:
Received 18 January 2017
Revised 22 July 2017
Accepted 26 August 2017
Available online xxx

Communicated by Dacheng Tao

Keywords:
Online
Multilinear principal component analysis
Dimension reduction

A B S T R A C T

Recently, the problem of extracting tensor object feature is studied and a very elegant solution, multilinear principal component analysis (MPCA), is proposed, which is motivated as a tool for tensor object dimension reduction and feature extraction by operating directly on the original tensor data. However, the original MPCA is an offline learning method and not suitable for processing online data since it generates the best projection matrices by learning on the whole training data set at once. In this study, we propose an online multilinear principal component analysis (OMPCA) algorithm and prove that the sequence generated by OMPCA converges to a stationary point of the total tensor scatter maximizing problem. Experiment results of an OMPCA-based support higher-order tensor machine for classification, show that OMPCA significantly lowers the time of dimension reduction with little sacrifice of recognition accuracy.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

With the rapid development in modern computer technology, tensor (the higher order generalization of vectors and matrices) data is becoming prevalent in many areas such as econometrics, chemometrics, psychometrics, computer vision, medical image analysis, web data mining and signal processing. For example, color images are three-dimensional (3D) objects with column, row and color modes [1]. MRI images [2], gray-level video sequences [3–6], gait silhouette sequences [7] and hyperspectral cube [8] are 3D data, color video sequences [9,10] are four-dimensional (4D) data. Generally speaking, a tensor is usually a multi-mode array and each mode corresponds to a feature space of tensor data. The tensor representation of data reflects the relationships between different features. If the multi-view is the multiple feature representation of the heterogeneous data in multi-view learning, each of the heterogeneous data examples is associated with multiple high-dimensional features [11]. Under this assumption, tensor can also be treated as the multi-view data.

Note that reshaping tensor objects into vectors breaks the natural structure and correlation in the original tensor data. More and more researchers focus on operating directly on the original tensor data. Within the last decade, researchers have suggested constructing multilinear models to extend the vector framework to tensor patterns [12–17]. In [13], Hao et al. presented a novel linear support higher-order tensor machine, combined with the tensor rank-one decomposition. AL-Shiha et al. [14] proposed a new supervised and unsupervised multilinear neighborhood preserving projection method for discriminative feature extraction by extending the original neighborhood preserving projection to its multilinear form. These tensor-based multilinear data analysis has shown that tensor models are capable of taking full advantage of the multilinear structures to provide better understanding and more precision. However, the tensor data contains large quantities of information redundancy and thus not all the features are important for classification and feature extraction [15–17]. Often, a dimension reduction scheme is needed to train a tensor model from a data set.

Dimension reduction is an attempt to transform a high-dimensional data set into a low-dimensional representation while retaining most of the underlying structure in the data [18]. Common algorithms of tensor dimension reduction include two-dimensional principal component analysis (2DPCA) [19], parallel factor model (Parafac) [20,21], Tucker decomposition [20,22], tensor canonical correlation analysis [11], multilinear principal component analysis (MPCA) [15] and so on [23–34]. In these algorithms, MPCA is the most popular one, which follows the classical PCA paradigm and determines a multilinear projection onto a tensor subspace of lower dimension. However, the original MPCA is not suitable to deal with online problems since whenever new additional samples are presented, MPCA system will have to repeat the learning process from the beginning. With the rapid increment of training data, its computational complexity will significantly increase. Motivated by this observation, in this study, we will propose an online multilinear principal component anal-

* Corresponding author.
E-mail address: xwyang@scut.edu.cn (X. Yang).

http://dx.doi.org/10.1016/j.neucom.2017.08.070
0925-2312/© 2017 Elsevier B.V. All rights reserved.

ysis (OMPCA) algorithm, which takes full use of the old training samples and sharply reduces the time of dimension reduction.

The rest of this paper is organized as follows. In Section 2, some notations of tensor and MPCA are introduced. The online multilinear principal component analysis algorithm is proposed in Section 3. Experiments can be found in Section 4. In Section 5 conclusions are given.

2. Preliminaries

In this section, we first introduce some notations and basic definitions used throughout the paper, and then briefly review the MPCA algorithm.

2.1. Notation and basic definitions

Except in some specified cases, lower-case bold letter, e.g., $\mathbf{x}$, represents a column vector, upper-case bold one, e.g., $\mathbf{X}$, represents a matrix, and calligraphic letter, e.g., $\mathcal{X}$, represents a tensor. Their elements are denoted by indices, which typically range from 1 to the capital letter of the index, e.g., $n = 1, \ldots, N$. In the following, we introduce some notations and definitions of the tensors in the area of multilinear algebra [20,35].

Definition 1 (Tensor)**.** A tensor, also known as Nth-order tensor, multidimensionality array, N-way or N-mode array, is a higher-order generalization of a vector (first-order tensor) and a matrix (second-order tensor), and denoted as $\mathcal{X} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ where N is the order of $\mathcal{X}$, also called way or mode. The element of $\mathcal{X}$ is denoted by $x_{i_1,i_2,\cdots,i_N}$, $i_n \in \{1, 2, \cdots, I_n\}$, $1 \le n \le N$.

Definition 2 (Kronecker product)**.** If $\mathbf{X}$ is a $m \times n$ matrix and $\mathbf{Y}$ is a $p \times q$ matrix, then the Kronecker product $\mathbf{X} \otimes \mathbf{Y}$ is the $mp \times nq$ block matrix:

$$\mathbf{X} \otimes \mathbf{Y} = \begin{bmatrix} x_{11}\mathbf{Y} & \cdots & x_{1n}\mathbf{Y} \\ \vdots & \ddots & \vdots \\ x_{m1}\mathbf{Y} & \cdots & x_{mn}\mathbf{Y} \end{bmatrix}. \tag{1}$$

Definition 3 (Inner product)**.** The inner product of two same-sized tensors $\mathcal{X}, \mathcal{Y} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as the sum of the products of their entries, i.e.,

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1,i_2,\cdots,i_N} y_{i_1,i_2,\cdots,i_N}. \tag{2}$$

Definition 4 (n-mode product)**.** The n-mode product of a tensor $\mathcal{X} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ and a matrix $\mathbf{U} \in R^{J_n \times I_n}$, denoted by $\mathcal{X} \times_n \mathbf{U}$, is a tensor in $R^{I_1 \times I_2 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$ given by

$$(\mathcal{X} \times_n \mathbf{U})_{i_1,i_2,\cdots,i_{n-1},j_n,i_{n+1},\cdots,i_N} = \sum_{i_n=1}^{I_n} x_{i_1,i_2,\cdots,i_N} u_{j_n,i_n}. \tag{3}$$

Remark: Given a tensor $\mathcal{X} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ and two matrices $\mathbf{U} \in R^{J_n \times I_n}, \mathbf{V} \in R^{J_m \times I_m}$, one has $(\mathcal{X} \times_n \mathbf{U}) \times_m \mathbf{V} = (\mathcal{X} \times_m \mathbf{V}) \times_n \mathbf{U} = \mathcal{X} \times_n \mathbf{U} \times_m \mathbf{V}$.

Definition 5 (Frobenius norm)**.** The Frobenius norm of a tensor $\mathcal{X} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ is the square root of the sum of the squares of all its elements, i.e.,

$$\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle} = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} \mathcal{X}^2_{i_1,i_2,\cdots,i_N}}. \tag{4}$$

Definition 6 (Tensor rank-one decomposition)**.** If a tensor $\mathcal{X} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ can be written as

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{x}_r^{(1)} \circ \mathbf{x}_r^{(2)} \circ \cdots \circ \mathbf{x}_r^{(N)} = \sum_{r=1}^{R} \prod_{n=1}^{N} \circ \mathbf{x}_r^{(n)}, \tag{5}$$

we call (5) tensor rank-one decomposition with length R, also known as CP decomposition.

Definition 7 (Total tensor scatter)**.** Let $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$ be a set of M tensor samples in $R^{I_1 \times I_2 \times \cdots \times I_N}$. The total scatter of these tensors is defined as:

$$\Psi_{\mathcal{X}} = \sum_{m=1}^{M} \left\| \mathcal{X}_m - \bar{\mathcal{X}} \right\|_F^2, \tag{6}$$

where $\bar{\mathcal{X}}$ is the mean tensor defined as $\bar{\mathcal{X}} = \frac{1}{M} \sum_{i=1}^{M} x_i$.

Definition 8 (Mode-n matricization of a tensor)**.** The mode-n matricization of a tensor $\mathcal{X} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ is denoted by $\mathbf{X}_{(n)} \in R^{I_n \times \prod_{i=1, i \neq n}^{N} I_i}$.

2.2. MPCA

Given a set of M tensor objects $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$ with $\mathcal{X}_m \in R^{I_1 \times I_2 \times \cdots \times I_N}, m = 1, \ldots, M$ for training. Denote the column orthonormal matrix $\mathbf{U}^{(n)} \in R^{I_n \times P_n}, n = 1, 2, \ldots, N$ where $I_n \ge P_n$. For $m = 1, \ldots, M$, let $\mathcal{Y}_m = \mathcal{X}_m \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \cdots \times_N \mathbf{U}^{(N)^T}$ be the projection of $\mathcal{X}_m$ onto the tensor subspace $R^{P_1 \times P_2 \times \cdots \times P_N}$. MPCA aims to determine the N projection matrices $\{\mathbf{U}^{(n)} \in R^{I_n \times P_n}, I_n \ge P_n, n = 1, 2, \ldots, N\}$ by solving the total tensor scatter maximizing problem [15]:

$$\max_{\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}} \quad \Psi\left(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}\right)$$
$$s.t. \quad \mathbf{U}^{(n)^T} \mathbf{U}^{(n)} = \mathbf{I}^{(n)}, n = 1, 2, \ldots, N \tag{7}$$

where $\Psi\left(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}\right) = \sum_{m=1}^{M} \|(\mathcal{X}_m - \bar{\mathcal{X}}) \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \cdots \times_N \mathbf{U}^{(N)^T}\|_F^2$ and $\mathbf{I}^{(n)} \in R^{P_n \times P_n}$ is a unit matrix.

Note that it is difficult to find the global optimal solution of the problem since it is nonconvex. In [15], the problem (7) is solved cyclically over each $\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(N)}$ while fixing the remaining blocks at their last updated values. In details, given the current iteration $\left(\mathbf{U}_k^{(1)}, \ldots, \mathbf{U}_k^{(N)}\right)$, generate the next iteration $(\mathbf{U}_{k+1}^{(1)}, \ldots, \mathbf{U}_{k+1}^{(N)})$ according to the iteration

$$\mathbf{U}_{k+1}^{(n)} \in \underset{\mathbf{U}^{(n)^T}\mathbf{U}^{(n)} = \mathbf{I}^{(n)}}{\operatorname{argmax}} \Psi\left(\mathbf{U}_{k+1}^{(1)}, \ldots, \mathbf{U}_{k+1}^{(n-1)}, \mathbf{U}^{(n)}, \mathbf{U}_k^{(n+1)}, \ldots, \mathbf{U}_k^{(N)}\right), \quad n = 1, \ldots, N. \tag{8}$$

To get the optimal solution of (8), we rewrite the function $\Psi(\mathbf{U}_{k+1}^{(1)}, \ldots, \mathbf{U}_{k+1}^{(n-1)}, \mathbf{U}^{(n)}, \mathbf{U}_k^{(n+1)}, \ldots, \mathbf{U}_k^{(N)})$ as an inner product $\langle \mathbf{U}^{(n)}, \Phi_{M,k}^{(n)} \mathbf{U}^{(n)} \rangle$, where $\Phi_{M,k}^{(n)} := \sum_{m=1}^{M} (\mathbf{X}_{m(n)} - \overline{\mathbf{X}}_{(n)}) \widehat{\mathbf{U}}_k^{(n)} \widehat{\mathbf{U}}_k^{(n)T} (\mathbf{X}_{m(n)} - \overline{\mathbf{X}}_{(n)})^T$ is a symmetric matrix with $\widehat{\mathbf{U}}_k^{(n)} := \mathbf{U}_k^{(n+1)} \otimes \cdots \otimes \mathbf{U}_k^{(N)} \otimes \mathbf{U}_{k+1}^{(1)} \otimes \cdots \otimes \mathbf{U}_{k+1}^{(n-1)}$. Furtherly, it is natural to get the optimal solution of (8) which consists of the P_n eigenvectors of the matrix $\Phi_{M,k}^{(n)}$ corresponding to the largest P_n eigenvalues. Then, MPCA can be shown as follows.

MPCA
Initial: centralize the given dataset $\{x_1, x_2, \ldots, x_M\}$ and obtain an initial point $(\mathbf{U}_0^{(1)}, \ldots, \mathbf{U}_0^{(N)})$.
For $k = 0, 1, 2, \ldots$
  For $n = 1, 2, \ldots, N$
    $$\mathbf{U}_{k+1}^{(n)} \in \underset{\mathbf{U}^{(n)^T}\mathbf{U}^{(n)} = \mathbf{I}^{(n)}}{\operatorname{argmax}} \langle \mathbf{U}^{(n)}, \Phi_{M,k}^{(n)} \mathbf{U}^{(n)} \rangle \tag{9}$$
End for
If $\Psi(\mathbf{U}_{k+1}^{(1)}, \ldots, \mathbf{U}_{k+1}^{(N)}) - \Psi(\mathbf{U}_k^{(1)}, \ldots, \mathbf{U}_k^{(N)}) < \eta$, break and output the feature tensors $\{\mathcal{Y}_m = \mathcal{X}_m \times_1 \mathbf{U}_{k+1}^{(1)^T} \times_2 \mathbf{U}_{k+1}^{(2)^T} \cdots \times_N \mathbf{U}_{k+1}^{(N)^T}, m = 1, \ldots, M\}$.
End for

Download English Version:

https://daneshyari.com/en/article/6864949

Download Persian Version:

https://daneshyari.com/article/6864949

Daneshyari.com