Computers in Biology and Medicine 78 (2016) 81–90

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Hidden discriminative features extraction for supervised high-order time series modeling

Ngoc Anh Thi Nguyen [a,b], Hyung-Jeong Yang [a,*], Sunhee Kim [c]

[a] Department of Computer Science, Chonnam National University, Gwangju 500-757, South Korea
[b] Faculty of Information Technology, University of Education, The University of Danang, VietNam
[c] Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, South Korea

ARTICLE INFO

Article history:
Received 2 March 2016
Received in revised form
24 August 2016
Accepted 25 August 2016

Keywords:
Electroencephalogram (EEG)
Microarray data
High-order time series
Discriminant analysis
Multi-way arrays
Tucker decomposition
Dimensionality reduction
Seizure prediction

ABSTRACT

In this paper, an orthogonal Tucker-decomposition-based extraction of high-order discriminative subspaces from a tensor-based time series data structure is presented, named as Tensor Discriminative Feature Extraction (TDFE). TDFE relies on the employment of category information for the maximization of the between-class scatter and the minimization of the within-class scatter to extract optimal hidden discriminative feature subspaces that are simultaneously spanned by every modality for supervised tensor modeling. In this context, the proposed tensor-decomposition method provides the following benefits: i) reduces dimensionality while robustly mining the underlying discriminative features, ii) results in effective interpretable features that lead to an improved classification and visualization, and iii) reduces the processing time during the training stage and the filtering of the projection by solving the generalized eigenvalue issue at each alternation step. Two real third-order tensor-structures of time series datasets (an epilepsy electroencephalogram (EEG) that is modeled as *channel* × *frequency bin* × *time frame* and a microarray data that is modeled as *gene* × *sample* × *time*) were used for the evaluation of the TDFE. The experiment results corroborate the advantages of the proposed method with averages of 98.26% and 89.63% for the classification accuracies of the epilepsy dataset and the microarray dataset, respectively. These performance averages represent an improvement on those of the matrix-based algorithms and recent tensor-based, discriminant-decomposition approaches; this is especially the case considering the small number of samples that are used in practice.

1. Introduction

Research on time series data has attracted considerable attention in the field of data mining, due to its prevalence in many high-impact applications, ranging from environmental monitoring, intelligent transportation systems, sensor networks, economy and finance, and motion capture to physiological signals in health care. Commonly, traditional time series is a collection of observations made chronologically that tend to be represented by a vector or a matrix to facilitate data processing and analysis [1]. Indeed, the nature of time series is such that it often generates massive amounts of data with multiple aspects and high dimensionality; the data can be denoted by a high-order time series or called a tensor time series [2–5].

The term "tensor" is used to represent the generalization of two-way data into high-order data and has become popular in modern sciences over recent years [1–8]. High-order time series is a collection of order finite these tensors and appears in a variety of high-impact applications in areas including neuroscience, bioinformatics, environmental monitoring, network systems, etc. These applications can be naturally represented as a sequence of tensors. For example, the multichannel time series of an electroencephalogram (EEG) in neuroscience data can be preserved as a third-order tensor time series that is modeled as *channel* × *frequency bin* × *time frame* [8–12]. Similarly, functional Magnetic Resonance Imaging (fMRI) in neurosciences is frequently organized as a third-order tensor time series comprising a three-dimensional (3-D) voxels evolving time frame [5–7]. In the bioinformatics domain, a multiple time series is assembled into a high-order time series form, *gene* × *sample* × *time* [13–18]. In addition, the time series of environmental data is modeled as *location* × *type* × *time*, the atmosphere data of climate modeling with *latitude* × *longitude* × *elevation grid points*, and network data are usually organized as network traffic with *source IP* × *destination*

* Corresponding author.
E-mail addresses: ntnanh@ued.udn.vn (N.A.T. Nguyen), hjyang@jnu.ac.kr (H.-J. Yang), sunheekim@korea.ac.kr (S. Kim).

http://dx.doi.org/10.1016/j.compbiomed.2016.08.018
0010-4825/

IP $\times$ *port number* $\times$ *time* [19–21].

The discovery of strong discriminative and interpretable features is salient for nearly all of the supervised recognition algorithms, and the development of a method that can exploit these hidden features while ensuring that the tensor representation of the original data input is preserved is a principal concern of researchers. With the increasing prevalence of such high-order time series, it has become necessary to move towards more versatile data analysis tools to extract useful and compact information for supervised modeling without disruption of the original multilinear structures. Consequently, tensor decomposition and factorization have been considered as parts of a new emerging exploratory technology of multidimensional data analysis for dimensionality reduction and feature extraction for such high-dimensional time series data in supervised learning [3,6–9]; this differs to the traditional approaches whereby feature-extraction methods are performed by unfolding or flattening the input sample tensors into the vectors. It is difficult, however, to interpret results and the specific information endorsed by the modalities is lost since the natural data structure is broken, and any correlation among the modes cannot be fully employed. Besides, further problems regarding the dimensional issue and the small sample size arise when the underlying original data structure is ignored [3,12,14]. Due to an ability to preserve multi-modal structures, a tensor analysis is far more desirable, since instead of performing a vectorization-derived feature extraction, it provides a natural and convenient representation that can be directly applied to the original tensor data [2].

The following two popular tensor decompositions have been used for most applications: Parallel Factor Analysis (PARAFAC), also known as “Canonical Decomposition (CP)”, and Tucker. A number of CP and Tucker extensions have also been applied successfully [3,6–9,11,14,18,22]. However, the limitation of the previously mentioned extracted features is its lack of category information that is useful for classification scenarios; these features can obtain as much information as possible for class discrimination that should be extracted to boost the classification efficiency. An algorithm for a tensor-discriminant analysis should therefore be developed so that information that can model the differences between classes of data during the classification process can be employed. For this purpose, the tensor linear Laplacian discriminant analysis (TLLDA) [23] and the local tensor discriminant analysis (LTDA) [24] have been applied recently so that class-label information can be used as the prior information of classification scenarios.

In this paper, we present an orthogonal Tucker decomposition that is composed of a high-order discriminant analysis, whereby class-label information is incorporated in order to more effectively capture and model the differences of the input tensor data between classes. This proposed method exploits high-order discriminant constraints that are directly imposed on hidden factors of the Tucker decomposition which discovers category information between classes to reinforce the tensor-based time series structures, named as Tensor Discriminative Feature Extraction (TDFE) The proposed method can effectively manage the small sample sizes, whereby a reduction that facilitates the extraction of small sets of discriminative and interpretable features is enabled.

In particular, the proposed TDFE tensor-decomposition method is applied for supervised learning of tensor time series domain. Firstly, the proposed method is performed in order to extract high-order subspaces features from training tensor inputs during the dimension-reduction step. In this step, the significant features from the reduced core tensor that express the correlation among the basic components are vectorized into a feature vector that denotes training features. In the feature-extraction step, the testing tensor inputs are projected onto a feature subspace that is spanned by the estimated basic factors of every modality of the training data. The Fisher information-ranking score is then applied to remove the discriminatively redundant features from the vectorized extracted features. Lastly, a supervised learning classifier for classification is built by comparing the training feature vector with the testing feature vector. With this scheme, the tensor-structure-preservation modeling is evaluated to show the generalities with the real high-order time sereis of EEG-epilepsy data; namely, the seizure and normal classes, and microarrays with favorable and problematic responder prediction capabilities. The experiment results present the averages of 98.26% and 89.63% for the classification accuracies of the epilepsy dataset and the microarray dataset, respectively. This demonstrates an improvement to the results of the matrix-based algorithms and recent tensor-based, discriminant-decomposition approaches.

The remainder of this paper is organized as follows. Basic knowledge regarding tensor algebra and Tucker decomposition material are briefly introduced in Section 2, along with a brief overview of the previous works on feature extraction for time series mining of epilepsy and microarray tensor-based datasets. In Section 3, the effectiveness of the orthogonal Tucker that is based on discriminant analyses is presented in terms of a high-order feature extraction. Section 4 presents the details of the experiments and evaluations and in Section 5, the paper is summarized and the possible directions for future research are discussed.

2. Preliminaries and related works

2.1. Fundamentals of tensor algebra: notation, operation

Tensors, also known as “multiway arrays”, represent a higher-order generalization of vectors and matrices. The number of dimensions, or “ways” or “modes”, determines the order of a tensor [2,8,25]. An *N*th-order tensor (*N*-way with $N \geq 3$) is denoted by boldface Euler-script letters, whereby $\mathcal{X} = \left\{ x_{i_1 i_2 \cdots i_N} \right\} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and the values of the *N* indices typically range from 1 to their capital version. For example, as shown in Fig. 1, a third-order tensor that is represented as $\mathcal{X} = \left\{ x_{ijk} \right\} \in \mathbb{R}^{I \times J \times K}$ has three indices. This third-order tensor $\mathcal{X} = \left\{ x_{ijk} \right\} \in \mathbb{R}^{I \times J \times K}$ comprises horizontal, lateral, and frontal slices that are denoted by $\mathcal{X}_{i::}$, $\mathcal{X}_{:j:}$, and $\mathcal{X}_{::k}$, respectively. Tubes or vectors at position (*i*, *j*) along mode-3, mode-2, and mode-1 are denoted by $\mathcal{X}_{i::}$, $\mathcal{X}_{:j:}$, and $\mathcal{X}_{::k}$, respectively. A matrix, or a two-way tensor, is denoted by a boldface

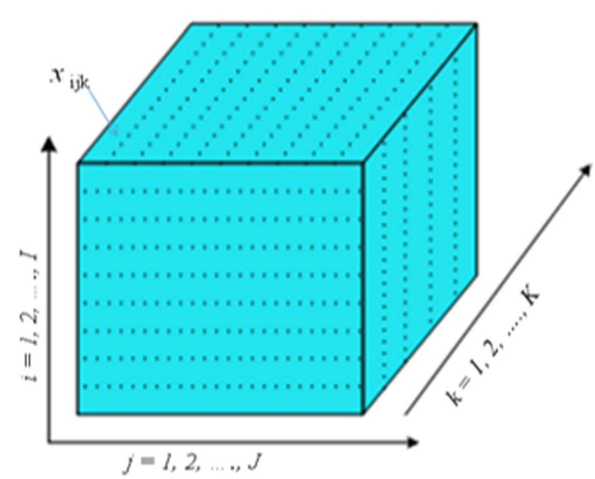

Fig. 1. Third-order as $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ with the element x_{ijk}.