NeuroImage 58 (2011) 537–548

Contents lists available at ScienceDirect

NeuroImage

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

TWave: High-order analysis of functional MRI

Michael Barnathan [a], Vasileios Megalooikonomou [a,b,*], Christos Faloutsos [c], Scott Faro [d], Feroze B. Mohamed [d]

[a] Data Engineering Laboratory, Center for Information Science and Technology, Temple University, Philadelphia, USA
[b] Department of Computer Engineering and Informatics, University of Patras, Patras, Greece
[c] School of Computer Science, Carnegie Mellon University, Pittsburgh, USA
[d] Department of Radiology, Temple University School of Medicine, Philadelphia, USA

ARTICLE INFO

Article history:
Received 6 December 2010
Revised 23 May 2011
Accepted 17 June 2011
Available online 24 June 2011

Keywords:
Tensors
Wavelets
Clustering
Latent Semantic Analysis
Parallel Factor Analysis
fMRI

ABSTRACT

The traditional approach to functional image analysis models images as matrices of raw voxel intensity values. Although such a representation is widely utilized and heavily entrenched both within neuroimaging and in the wider data mining community, the strong interactions among space, time, and categorical modes such as subject and experimental task inherent in functional imaging yield a dataset with "high-order" structure, which matrix models are incapable of exploiting. Reasoning across all of these modes of data concurrently requires a high-order model capable of representing relationships between all modes of the data in tandem. We thus propose to model functional MRI data using *tensors*, which are high-order generalizations of matrices equivalent to multidimensional arrays or data cubes. However, several unique challenges exist in the high-order analysis of functional medical data: naïve tensor models are incapable of exploiting spatiotemporal locality patterns, standard tensor analysis techniques exhibit poor efficiency, and mixtures of numeric and categorical modes of data are very often present in neuroimaging experiments. Formulating the problem of image clustering as a form of Latent Semantic Analysis and using the WaveCluster algorithm as a baseline, we propose a comprehensive hybrid tensor and wavelet framework for clustering, concept discovery, and compression of functional medical images which successfully addresses these challenges. Our approach reduced runtime and dataset size on a 9.3 GB finger opposition motor task fMRI dataset by up to 98% while exhibiting improved spatiotemporal coherence relative to standard tensor, wavelet, and voxel-based approaches. Our clustering technique was capable of automatically differentiating between the frontal areas of the brain responsible for task-related habituation and the motor regions responsible for executing the motor task, in contrast to a widely used fMRI analysis program, SPM, which only detected the latter region. Furthermore, our approach discovered latent concepts suggestive of subject handedness nearly 100× faster than standard approaches. These results suggest that a high-order model is an integral component to accurate scalable functional neuroimaging.

Introduction

The traditional approach to data representation utilizes a matrix structure, with observations in the rows and features in the columns. Although this model is appropriate for many datasets, it is not always a natural representation because it assumes the existence of a single target variable and lacks a means of modeling dependencies between other features. Additionally, such a structure assumes that observed variables are scalar quantities by definition. This assumption may not be valid in certain domains, such as diffusion tensor imaging, where higher-order features predominate, or in domains which have strong spatiotemporal components, such as functional MRI.

Traditionally, these problems have been solved by reducing the features to scalars and fitting the dataset to a matrix structure. However, as well as potentially losing information, this strategy also employs a questionable approach from a philosophical standpoint: attempting to fit the data to an imprecise model rather than attempting to accurately model the existing structure of the data. Finally, while it may be possible to model dependencies between features by repeating the methodology multiple times, each with a different target variable, this yields suboptimal performance and may not be computationally feasible when real-time performance is required or when the dataset is very large.

To address these issues, we propose to model such datasets using *tensors*, which are generalizations of matrices corresponding to r-dimensional arrays, where r is known as the *order* of the tensor. Using a combination of wavelet and tensor analysis tools, we propose a novel framework for summarization, clustering, concept discovery,

* Corresponding author at: Vasileios Megalooikonomou, Data Engineering Laboratory, Center for Information Science and Technology, Temple University, 1805 N. Broad St., Philadelphia, PA, USA. Fax: +1 215 204 5082.
E-mail addresses: mbarnath@temple.edu (M. Barnathan), vasilis@temple.edu (V. Megalooikonomou), christos@cs.cmu.edu (C. Faloutsos), scott.faro@tuhs.temple.edu (S. Faro), feroze.mohamed@tuhs.temple.edu (F.B. Mohamed).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.06.043

and compression of high-order datasets, which we call TWave (Barnathan et al., 2010). Applying our technique to analysis of a large real-world digital opposition fMRI dataset, we compare the performance of TWave against voxelwise, SVD-based, wavelet-only, and tensor-only techniques and demonstrate our TWave method achieves superior results and reduces computation time vs. competing methodologies such as Latent Semantic Analysis.

Our approach has several advantages:

- Compression of the data model through grid quantization and wavelet preprocessing (inherent in the WaveCluster algorithm).
- Exploitation of spatial neighborhoods and local patterns.
- Efficiency up to two orders of magnitude faster than naive tensor approaches.
- The ability to identify noncontiguous clusters based on patterns in the projected space.
- Naturally fuzzy clustering based on similarities to discovered concepts.
- The projected space may reveal latent dataset concepts (our method revealed information about subject handedness in our dataset).

Background

Tensor tools

Tensors are defined within the context of data mining as multidimensional arrays. The number of indices required to index the tensor is referred to as the *order* of the tensor, while each individual dimension is referred to as a *mode*. The number of elements defined on each mode is referred to as the mode's *dimensionality*. The dimensionality of a tensor is written in the same manner as the dimensionality of a matrix; for example, $20 \times 50 \times 10$.

Tensors represent generalizations of scalars, vectors, and matrices, which are tensors of orders 0, 1, and 2, respectively. Tensors and the notion of order are illustrated in Table 1.

An important operation applicable to our analysis is the *tensor product*. This product generalizes not from matrix multiplication, but from the Kronecker product operation defined on matrices, which is given as follows.

Given an $m \times n$ matrix $\mathbf{A}$ and a $p \times q$ matrix $\mathbf{B}$, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is defined by the following $mp \times nq$ block matrix:

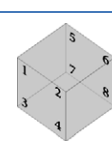
$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m,1}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{pmatrix}$$

The tensor product is similar, but the result is another tensor rather than a block matrix. Specifically, if given order r and s tensors $\mathcal{A}$ and $\mathcal{B}$, their tensor product $\mathcal{A} \otimes \mathcal{B}$ is a tensor of order $r+s$ defined as follows:

$$(\mathcal{A} \otimes \mathcal{B})_{i_1, i_2, \ldots, i_r, j_1, j_2, \ldots, j_s} = \mathcal{A}_{i_1, i_2, \ldots, i_r} * \mathcal{B}_{j_1, j_2, \ldots, j_s}$$

For example, the procedure of taking a tensor product is shown in Fig. 1, with arrows representing the direction of multiplication.

Table 1
Scalars, vectors, matrices, tensors, and their orders.

1	[1 2 3 4]	$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$	
0	1	2	*r*

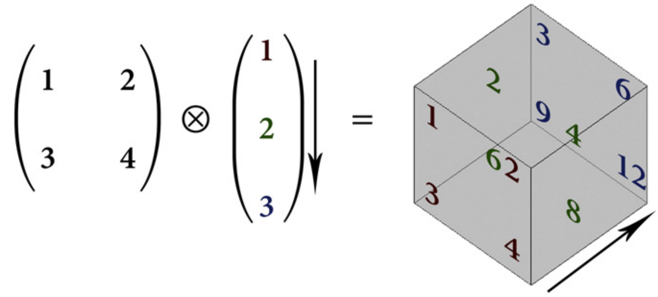


Fig. 1. Graphical example of a tensor product.

The operation known as the Khatri–Rao product is useful in the computation of several tensor decompositions and is defined in terms of the Kronecker product. Let $\mathbf{A}$ be a $p \times n$ matrix and $\mathbf{B}$ be a $q \times n$ matrix. Their Khatri–Rao product $\mathbf{A} \odot \mathbf{B}$ is as follows:

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \mathbf{a}_2 \otimes \mathbf{b}_2, \ldots, \mathbf{a}_n \otimes \mathbf{b}_n]$$

Singular value decomposition (SVD) is a unique matrix factorization by which an $m \times n$ matrix is decomposed into two projection matrices and a core matrix, as follows:

$$\mathbf{A} = \mathbf{U} \times \Sigma \times \mathbf{V}^{\mathbf{T}}$$

where $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{U}$ is an $m \times r$ column-orthonormal projection matrix, $\mathbf{V}$ is an $n \times r$ column-orthonormal projection matrix, and Σ is a diagonal $r \times r$ *core matrix*, where r is the rank of the projection.

Singular value decomposition has a wide variety of applications: for example, truncation of the SVD coefficients provides an optimal low-rank approximation (i.e. minimizes the Frobenius norm). This indicates a close relationship between principal component analysis (PCA) and SVD.

SVD is also used to discover the rank of a matrix, find the pseudoinverse, and solve least squares minimization problems. Additionally, the solution to SVD may be used in an unsupervised summarization technique known as Latent Semantic Analysis (LSA) (Deerwester et al., 1999). In this technique, $\mathbf{A}$ is treated as a term-document matrix. Here, singular value decomposition automatically derives a user-specified number of latent *concepts* from the given terms which form a basis for the rows and columns of the matrix. The projection matrices $\mathbf{U}$ and $\mathbf{V}$ then contain term-to-concept and document-to-concept similarities, respectively. Thus, SVD can be used to provide simple yet powerful automatic data summarization. This technique may be naturally viewed as a form of co-clustering, in which the rows and columns of a matrix cluster to the same space. An alternative graphical interpretation exists, in which clusters represent shared “waypoints” through which edges pass between vertices. Use of the eigendecomposition or SVD is also common in a graphical context, where it is known as *spectral graph theory*; here a common technique is to cluster on the eigenvector corresponding to the second smallest eigenvalue of the Laplacian matrix, thereby partitioning vertices along edges which are likely to be minimal cuts. This technique is known as *Fiedler retrieval*. It is also possible to project new query vectors into the space defined by the SVD, known as *folding in*; this enables recommendation as the query projects to the same space as both the rows and columns and can be assessed using a distance metric.

The natural extension of singular value decomposition to tensors is known as *high-order singular value decomposition*, or HOSVD. This decomposition, in turn, is a special case of the *Tucker decomposition*, which is capable of concurrent data co-clustering across every mode of a tensor. Formally, let $\mathcal{A}$ be a tensor of order r; i.e. $\mathcal{A} \in \Re^{d_1 \times d_2 \times \ldots \times d_r}$. We may then define the Tucker