UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE BOGOTÁ

FACULTAD DE INGENIERÍA
Vicedecanatura Académica
POSGRADOS

PROPOSAL SUBMISSION

DOCTORAL THESIS:	x	MASTER THESIS:	
MASTER FINAL WORK:		SPECIALIZATION FINAL WORK:	

1. **BIDDER:** Robinson Andrés Jaque Pirabán **ID:** 80190790
2. **PROGRAM:** Phylosophy Doctoral in Computer Science and Systems Engineering
3. **ADVISOR:** Fabio Augusto González Osorio
   DEPARTMENT: Computer Science and Industrial Engineering
4. **TITLE: Kernel Tensor Factorization**
5. **AREA:** Computer Science
6. **LINE OF RESEARCH:** Machine Learning
7. COMMENTARY WITH ADVISOR APROVAL
8. BIDDER SIGNATURE
9. SIGNATURE OF ADVISOR

Contents

1 Introduction

1.1 Basics of tensors

Tensors are multidimensional arrays, i.e. an N-way or N-order tensor is an element of tensor product of N vector spaces, each of which has its own coordinate system. A first order tensor is a vector, a second order tensor is a matrix, tensors of higher order are called high-order tensors. The order (ways or modes) of a tensor is the number of dimensions. Figure 1 represents a 3-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$.

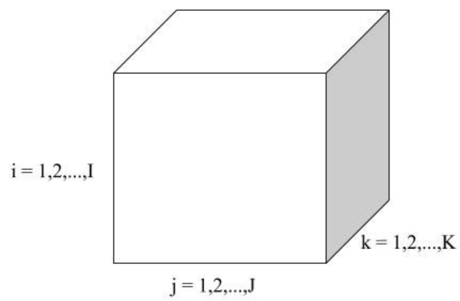
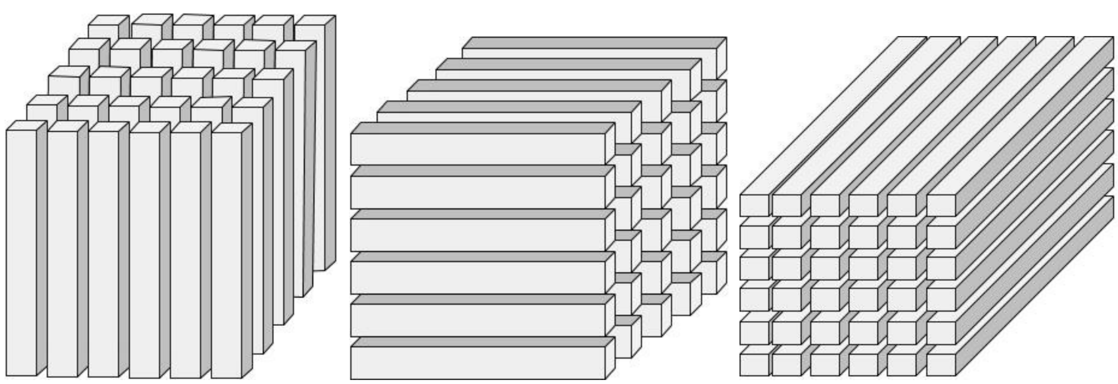


Figure 1: Third-order tensor

Fibers are defined by fixing every index by one. In a third-order tensor a column is a mode-1 fiber, denoted by $x_{:jk}$; a row is a mode-2 fiber, denoted by $x_{i:k}$; while a tube is a mode-3 fiber, denoted by $x_{i:k}$. 2 shows a fibers representation in 3rd-order tensor.

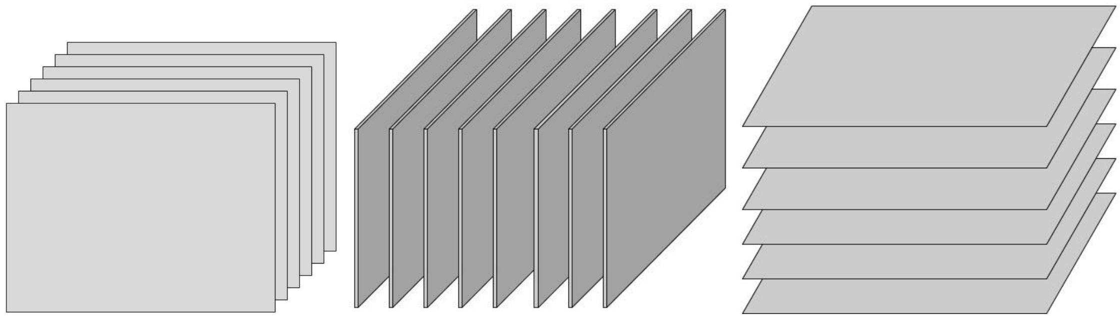


(a) columns (b) rows (c) tubes

Figure 2: 3rd-order tensor fibers

Slices are two-dimensional sections of a tensor defined by fixing two indexes. For instance, slices of 3rd-order tensor $\mathcal{X}$ are denoted by $X_{i::}$ (horizontal), $X_{:j:}$ (lateral) and $X_{::k}$ (frontal) and we ilustrate them in figure 3.

The *norm* of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$ is analogous to the matrix Frobenius norm,



(a) Frontal slices (b) Horizontal slices (c) Vertical slices

Figure 3: 3rd-order tensor slices

i.e.

$$||\mathcal{X}|| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N}^2} \tag{1}$$

$\mathcal{X}$ is a *Rank-one* tensor if it is equal to the outer product of N vectors, i.e.,

$$\mathcal{X} = a^{(1)} \otimes a^{(2)} \otimes \ldots \otimes a^{(N)}$$

1.1.1 Unfolding and Folding Tensors

Unfolding is the process of *matricization* of a tensor. In other words, elements of a tensors are sorted to assemble a matrix. The mode-k unfolding of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$ is denoted by $X_{(k)} \in \mathbb{R}^{I_1 \times \prod_{k' \neq k} I_{k'}}$ and arrenges the mode-k tensor fibers as columns of resulting matrix. In addition, Kolda [41] presents a more general procedures of unfolding

Ding and Wei [28] present a fast algorithm for Hankel tensor-vector products. And [32] a method of fast linear transform algorithm synthesis for an arbitrary tensor.

2 Tensor Decomposition

Tensor decomposition originated with Hitchcock in 1927 [39], and the the multi-way model is attribuited to Cattell in 1944 [18].

Tensor works had attention in 60s with Tucker ([52], [53], [51]) and Carroll and Chang [17] and Harshman in 1970 [37] with applications in psychometrics. In 1981 Appellof and Davidson [9] used tensor decomposition in chemometrics which have been an popular field of application of tensor decomposition since then.

In last twenty years tensor decomposition applications have expanded to many fields such as signal processing, numberical linear algebra, computer vision, numerical analysis, neuroscience, data mining, graph analysis. Figure 4 shows seminal papers which opened broad application fields to tensor decomposition, plot also shows the amount of documents published about these works according to Scopus.

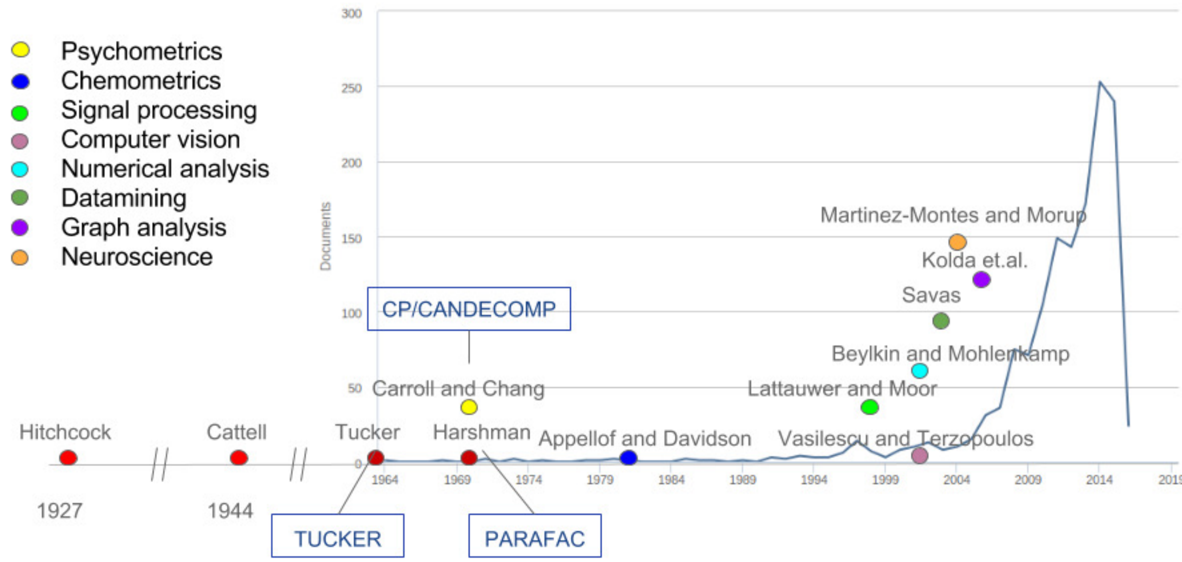


Figure 4: Time-line of tensor decomposition

Kolda [41], Acar [1] and [24] present an exhaustive and detailed review of fundamental decomposition methods and applications. Furthermore, [25] presents tensor properties as extension of estructural properties of matrices. On the other hand, Fanaee and Gama [34] introduce an interdisciplinary survey about tensor-based anomaly detection.

In following sections we explain some of basic methods to tensor decomposition which have been inspiration to many others methods propossed. Also, we summarize recently works and approaches of tensor decomposition on different application fields.

Chemometrics [19] A computationally efficient technique for the solution of multi-dimensional PBMs of granulation via tensor decomposition

Bioinformatics In [2] and [4] a multimodal problem is addressed, the authors formulate data fusion as a coupled matrix and tensor factorization problem and discuss its extension to a structure-revealing data fusion model in metabolomics.

Image processing [5] NNTF for facial expression recognition

Machine Learning [6] Tensor decompositions for learning latent variable models

Text mining [7] This paper describes a method for automatic detection of semantic relations between concept nodes of a networked ontological knowledge base by analyzing matrices of semantic-syntactic valences of words. These matrices are obtained by means of nonnegative factorization of tensors of syntactic compatibility of words.

Numerical analysis [8] use of the sum-factorization for the calculation of the integrals arising in Galerkin isogeometric analysis. While introducing very little change in an isogeometric code based on element-by-element quadrature and assembling, the sum-factorization approach, taking advantage of the tensor-product structure of splines or NURBS shape functions, significantly reduces the quadrature computational cost.

[13] Fast iterative solution of the Bethe-Salpeter eigenvalue problem using low-rank and QTT tensor approximation.

[21] Blind identification of a second order Volterra-Hammerstein series using cumulant cubic tensor analysis.

Neuroscience [10] Decomposition of brain diffusion imaging data uncovers latent schizophrenias with distinct patterns of white matter anisotropy, using NNTF to clustering.

Signal processing Cichocki et.al. [22] sum up tensor decomposition approaches for signal processing problems.

Barker and Virtanen [11] deal with monaural sound source separation problem using NNTF of modulation spectrograms.

[15] Necessity to manually assign the NTF components to audio sources in order to be able to enforce prior information on the sources during the estimation process, Automatic Allocation of NTF Components for User-Guided Audio Source Separation

[36] propose a shifted 2D non-negative tensor factorisation algorithm which extends non-negative matrix factor 2D deconvolution to the multi-channel case. The use of this algorithm for multi-channel sound source separation of pitched instruments is demonstrated.

Other applications [33] Discovering and Characterizing Mobility Patterns in Urban Spaces: A Study of Manhattan Taxi Data. by using non-negative tensor factorization (NTF), we are able to cluster human behavior based on spatio-temporal dimensions. Second, for understanding these clusters, we propose to use HypTrails, a Bayesian approach for expressing and comparing hypotheses about human trails.

[35] NTF factorization for household electrical seasonal consumption disaggregation

2.1 Tensor Factorization Methods

2.1.1 Canonica Polyadic Decomposition / PARAFAC

Canonica Polyadic (CP) decomposition ([41], [45]), CANDECOMP [17] or PARAFAC [37] decompose a tensor as a finite sum of rank-one tensors. For instance, given a third order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, CP decomposition express it as

$$\mathcal{X} \approx \sum_{r=1}^{R} a_r \otimes b_r \otimes c_r \tag{2}$$

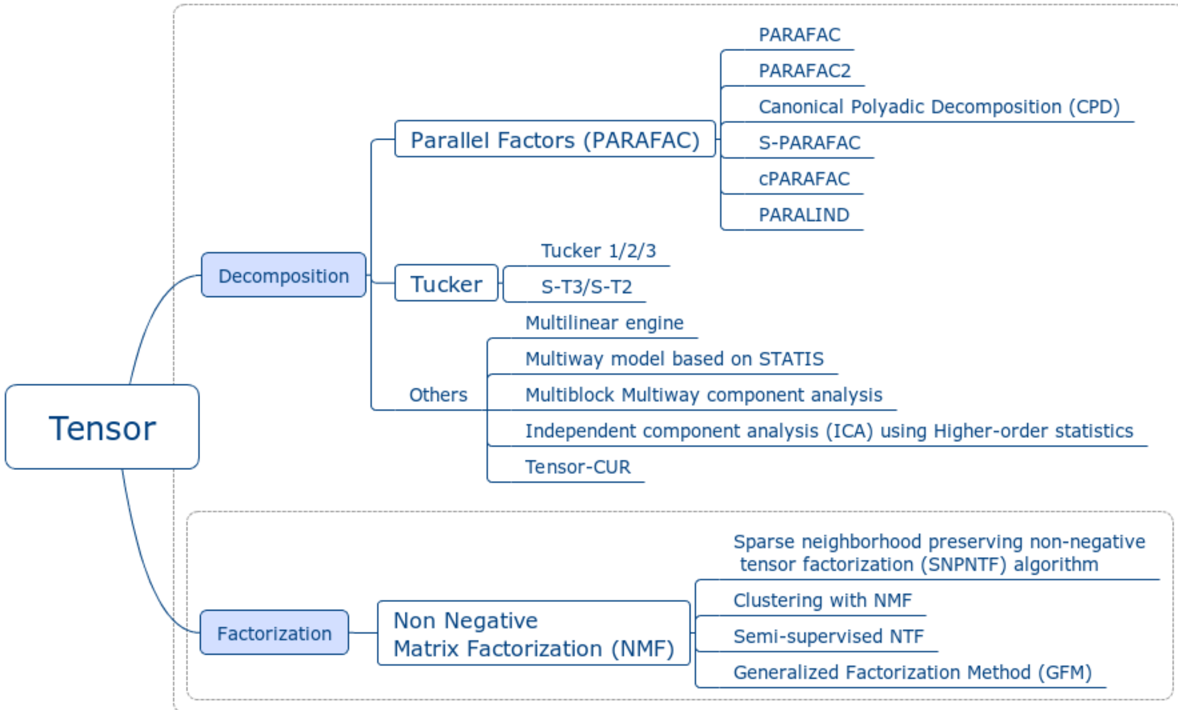


Figure 5: Tensor decomposition methods

where $a_r \in \mathbb{R}^I$, $b_r \in \mathbb{R}^J$, $c_r \in \mathbb{R}^K$ and R is a positive integer.

Domanov [29] shows relaxed uniqueness conditions and algebraic algorithm for Canonical polyadic decomposition, as well as a reduction to generalized eigenvalue decomposition [30] and uniqueness properties [31] of third-order tensors.

2.1.2 TUCKER Decomposition

Tucker decomposition was introduced by Tucker ([52], [53]). It is also named N-mode PCA [40], High-order SVD (HOSVD) [26] or N-mode SVD [54].

The Tucker decomposition is a form of higher-order PCA [41]. It descomposes a tensor into a core tensor $\mathcal{G}$ multiplied by a matrix algong each mode. For instance, given a third order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, Tucker decomposition express it as

$$\mathcal{X} \approx \mathcal{G} \times_1 A \times_2 B \times_3 C \tag{3}$$

Where $\times_k$ is the mode-k product, $A \in \mathbb{R}^{I \times P}$, $B \in \mathbb{R}^{J \times Q}$, $C \in \mathbb{R}^{K \times R}$ are the factor matrices (usually orthogonals) and can be interpreted as the principal components for each mode. $\mathcal{G} \in \mathbb{R}^{P \times Q \times R}$ is the core tensor and its entries show the interactions between the different components.

2.1.3 Non-negative Tensor Factorization

Non-negative Matrix Factorization The general problem of non-negative matrix factorization (NMF) is to decompose a matrix $X \in \mathbb{R}_{\geq 0}^{n \times l}$ into two matrix factors: basis $W \in \mathbb{R}_{\geq 0}^{n \times k}$ and coefficients $H \in \mathbb{R}_{\geq 0}^{k \times l}$, i,e.

$$X \cong WH \tag{4}$$

The factorization problem can be seem as an optimization problem:

$$\min_{W,H} d(X, WH) \tag{5}$$

where $d(,)$ is a distance or divergence function and the problem could have different types of restrictions. For instance, if $d(,)$ is the Euclidean Distance and there are not restrictions, the problem is solved by finding the SVD; if X, W and H are restricted to be positive, then the problem is solved by NMF. An comprehensive survey of NMF variants and algorithms is finded in [55]. One NMF approach is Symmetric-NMF [27], (SNMF) which produces a factorization:

$$(X_{lxn}^{T} X_{nxl}) = H_{lxk} H_{kxl}^{T} \tag{6}$$

An important characteristic of this version of NMF is that it is amenable to be used as a kernel method. This is discussed in the next subsection.

Non-negative Tensor Factorization An natural extension of nonnegative matrix factorization with high-order arrays is nonnegative n-dimensional tensor factorization (n-NTF). This kind of generalization is indeed not trivial since NTF possesses many new properties varying from NMF ([38], [47]).

First, the data to be processed in NMF are vectors in essence. However, in some applications the original data may not be vectors, and the vectorization might result in some undesirable problems. For instance, the vectorization of image data, which is two dimensional, will lose the local spatial and structural information. Second, one of the core concerns in NMF is the uniqueness issue, and to remedy the ill-posedness some strong constraints have to be imposed. Nevertheless, tensor factorization will be unique under only some weak conditions. Besides, the uniqueness of the solution will be enhanced as the tensor order increases[55].

There are generally two types of NTF model—NTD [47] and more restricted NPARAFAC [38], whose main difference lies in the core factor tensor. As for the solution, there are some feasible approaches. For example, NTF can be restated as regular NMF by matricizing the array [56], [47]. Or the alternating iteration method can be utilized directly on the outer product definition of tensors [38], [49], [12]. Similarly, SED, GKLD and other forms of divergence can also be used as the objective functions [12], [23], [57]. And some specific update models can adopt the existing conclusions in NMF.

2.1.4 Other methods

The tensor decomposition addressed in [14] may be seen as a generalization of Singular Value Decomposition of matrices. They consider general multilinear and multihomogeneous tensors. Then, they show how to reduce the problem to a truncated moment matrix problem and give a new criterion for flat extension of Quasi-Hankel matrices.

2.2 Kernel methods

In contrast with traditional learning techniques, kernel methods do not need a vectorial representation of data. Instead, they use a kernel function. Therefore, kernel methods are naturally applied to unstructured, or complex structured, data such as texts, strings, trees and images [50].

Informally, a kernel function measures the similarity of two objects. Formally, a kernel function, $k : X \times X \to \mathbb{R}$, maps pairs (x, y) of objects in a set X, the problem space, to the reals. A kernel function implicitly generates a map, $\Phi : X \to F$, where F corresponds to a Hilbert space called the feature space. The dot product in F is calculated by k, specifically $k(x, y) =< \Phi(x), \Phi(y) >_F$. Given an appropriate kernel function, complex patterns in the problem space may correspond to simpler patterns in the feature space. For instance, non-linear patterns in the problem space may correspond to linear patterns in the feature space.

Both k-means and SNMF have kernelized versions, which receive as input a kernel matrix instead of a set of sample represented by feature vectors. The kernel version of k-means is called, unsurprisingly, kernel k-means (KKM). In the case of SNMF, the kernelized version works as follows.

SNMF starts with an initial estimation of the matrix factor H and iteratively update it using the updating equation:

$$H_{i,k} = H_{i,k}(1 - \beta + \beta \frac{((X^T X)H)_{i,k}}{(HH^T H)_{i,k}})$$

The kernel version of the algorithm is obtained by using a kernel matrix K instead of the expression $(X^T X)$, where K is an $l \times l$ matrix with $K_{i,j} = k(x_i, x_j)$. There are different types of kernels some of them general and some of them specifically defined for different types of data. The most popular general kernels are the linear kernel

$$k(x, y) =< x, y >, \tag{7}$$

the polynomial kernel

$$k(x, y) = p(< x, y >),$$

where $p(\,)$ is a polynomial with positive coefficients, and the Gaussian (or RBF) kernel

$$k(x, y) = e^{\frac{\|x-y\|^2}{2\sigma^2}}. \tag{8}$$

The cluster centroids estimated by the kernel versions of both algorithms are in the feature space and correspond to the points $C_j = \frac{1}{n}\sum_{x_i \in C_j} \Phi(x_i)$. However, we are interested on the pre-image in the original space of this centroids, i.e., points $\hat{C}_j$ such that $\Phi(\hat{C}_j) = C_j$. However, it is possible that a exact pre-image may not even exist, so we look for the $\hat{C}_j$ that minimizes the following objective function:$\min_{\hat{C}_j} \left\| \hat{C}_j - C_j \right\|^2$. According to Kwok et al. [?], the optimum C_j can be found by iterating the following fixed-point formula:

$$\hat{C}_j^{t+1} = \frac{\sum_{i=1}^{N} \exp(\frac{-||\hat{C}_j^t - x_i||)}{s}) x_i}{\sum_{i=1}^{N} \exp(\frac{-||\hat{C}_j^t - x_i||}{s})} \tag{9}$$

3 Kernel Non-negative Matrix Factorization

Kernel Non-negative Matrix Factorization (KNMF) can be naturally derivated of convex NMF ([42], [43] and [48]). Given a kernel function $\phi : x \in X \to \phi(x) \in F$, mapping for N elements $\phi(X) = [\phi(x_1), \ldots \phi(x_N)]$. Then, KNMF can be defined as

$$\phi(X) \cong \phi(X) W H^T \tag{10}$$

Therefore, the cost function to minimize is

$$||\phi(X) - \phi(X) W H^T||_F^2 = tr(K) - 2tr(H^T K W) + tr(W^T K W H^T H) \tag{11}$$

Where kernel $K = \phi^T(X) phi(X)$

3.1 General problems

Usually, tensor factorization address the folowing problems independent of the application: blind source separation, tensor completion.

3.2 Tensor completion

In tensor completion, an given incomplete tensor is given, i.e. some of its entries are missing and we should complete that. Following Ji Lu et.al [44] notation, low rank matrix completion is noted as

$$\begin{aligned} &\min_X \ \text{rank}(X) \\ &\text{s.t. } X_\Omega = M_\Omega \end{aligned} \tag{12}$$

where Ω is an index set, then X_Ω is coping entries of X in the indexes Ω and missed entries $\hat{\Omega}$ would be 0

The missing entries in X are determined in order to minimize the matrix X rank. i.e. a non convex optimization problem since rank is nonconvex.

Frequently, trace norm (or nuclear norm) $||\cdot||_*$ is used to approximate the rank of matrices.

Trace norm is the tighest convex envelop for the matrices rank.

$$\begin{aligned} &\min_X ||X||_* \\ &\text{s.t. } X_\Omega = M_\Omega \end{aligned} \tag{13}$$

Since tensor is a generalization of the matrix concept, we generalize the optimization problem as

$$\begin{aligned} &\min_{\mathcal{X}} ||\mathcal{X}||_* \\ &\text{s.t. } \mathcal{X}_\Omega = \mathcal{T}_\Omega \end{aligned} \tag{14}$$

Where $\mathcal{X}$ and $\mathcal{T}$ are n-order tensors with identical size.

Acar et.al [3] presents a scalable tensor factorization method to deal with completion problem using PARAFAC method. Cao [16] propose a new tensor completion model via folded-concave penalty for estimating missing values in tensor data.To solve the resulting nonconvex optimization problem, we develop a local linear approximation augmented Lagrange multiplier (LLA-ALM) algorithm which combines a two-step LLA strategy to search a local optimum of the proposed model efficiently. They show numerical results in image and video data sets and compare with nuclear norm penalization method in order of demostrate its advantage in terms of the accuracy and robustness.

Chen [20] propose a method to deal with the completion problem when the number of missing entries increases, since factorization schemes may overfit the model because of incorrectly predefined ranks, while completion schemes may fail to interpret the model factors. Hence, they present an approach to complete the missing entries and simultaneously capture the underlying model structure that combines a rank minimization technique with Tucker model decomposition. Moreover, as the model structure is implicitly included in the Tucker model, they use factor priors, which are usually known a priori in real-world tensor objects, to characterize the underlying joint-manifold drawn from the model factors.

4 Problem Statement

Matrix and analogous tensor factorization is central to different important tasks in machine learning and information retrieval such as: clustering, latent topic analysis, recommendation, blind source separation, completion, among others. The widespread use of multisensor technology and the emergence of big data sets have highlighted the limitations of standard flat-view matrix models and the necessity to move toward more versatile data analysis tools [22]. Conventional methods preprocess multiway data by arranging them into a matrix, which might lose the original multiway structure of the data [55]. Hence, tensors address multimodal or multiview data, and tensor factorization brings tools to perform common machine learning tasks.

On the other hand, kernel methods are ubiquitous in machine learning, performance of these methods have been broadly demostrated, and there are evidence between some types of kernels and robustness. However, there are few exploration of kernel tensor factorization approaches.

A satisfactory solution of this general challenge requires to answer some particular**research question**:

- How incorporate kernel methods in tensor factorization in order to deal with multiway data?

An answer to the question derivate in a method which factorize a given tensor in the feature space induced by a Kernel function. Naturally, to inquire about the effects of decompose tensors in an space induced by a kernel function open space to evaluate the proposal method performance as well as its capabilities dealing with multimodal data, scalability and robustness to noise and outliers.

5 Goals

The general problem addressed by this research proposal is the design of non-supervised learning algorithms, in particular tensor factorization algorithms, applied in the space induced by a kernel function.

General objective

To build an kernel-based tensor factorization algorithm for unsupervised learning.

Specific objectives

(a) To design an efficient and effective kernel tensor factorization algorithms.

(b) To Implement algorithm.

(c) To evaluate the impact, in terms of robustness and accuracy, of using particular kernels in kernel-based methods for tensor factorization.

(d) To evaluate the performance of the proposed algorithms in both synthetic and real world datasets, and its impact in particular learning tasks: completion, clustering and blind source separation.

(e) To evaluate scalability comparing algorithms' complexity and parallelization prospectives.

6 Jusification

Due to the rich characteristics of natural processes and environments, it is rare that a single acquisition method provides complete understanding thereof. Information about a phenomenon or a system of interest can be obtained from different types of instruments, measurement techniques, experimental setups, and other types of sources. The increasing availability of multiple data sets that contain information, obtained using different acquisition methods, about the same system, introduces new degrees of freedom that raise questions beyond those related to analyzing each data set separately.

By treating that natural high-order array as a matrix, information is lost since lack the original multiway structure of the data.

Tensor factorizations have several advantages over two-way matrix factorizations including uniqueness of the optimal solution and component identification even when most of the data is missing [46].

multiway decomposition techniques explicitly exploit the multiway structure that is lost when collapsing some of the modes of the tensor in order to analyze the data by regular matrix factorization approaches.

Tensor decompositions are in frequent use today in a variety of fields ranging from psychology, chemometrics, signal processing, bioinformatics, neuroscience, web mining, and computer vision to mention but a few.

Matrix and analogous tensor factorization is central to different important tasks in machine learning and information retrieval such as: clustering, latent topic analysis, recommendation, tensor completion, blind source separation, among others.

7 Methodology

Each specific objective is to be attained through the following approaches:

(a) Design and implementation of Kernel Tensor Factorization algorithm: An kernel matrix factorization methods will be adapted to use tensorial kernels. Efficient implementations will be developed exploiting particularities of the used kernels. In particular, an on-line learning strategy for matrix factorization will be designed.

(b) Strategy to evaluate robustness and accuracy: In order to evaluete algorithm capabilities to deal with noise and outliers, a collection of synthetic and real datasets with different degrees of contamination will be assembled. An experimental setup for algorithm evaluation will be designed taking into account different factors (kernel type, contamination degree, etc) and the robustness measures as influence curve and/or the breakdown bound adapted to tensor factorization.

(c) Performance evaluation in specific tasks: Algorithm will be tested in specific machine learning task, such as completion, clustering and blind source separation on synthetic and real data sets.

(d) Evaluation of algorithm scalability: An theoretical time and space complexity bound will be assesed to the algorithm. Furthermore, an parallelizable version of algorithm will be suggested.

References

References

[1] E. Acar and B. Yener. Unsupervised Multiway Data Analysis: A Literature Survey. *IEEE Transactions on Knowledge and Data Engineering*, 21(1):6–20, jan 2009.

[2] Evrim Acar, Rasmus Bro, and Age K. Smilde. Data Fusion in Metabolomics Using Coupled Matrix and Tensor Factorizations. *Proceedings of the IEEE*, 103(9):1602–1620, sep 2015.

[3] Evrim Acar, Daniel M. Dunlavy, Tamara G. Kolda, and Morten Mørup. Scalable tensor factorizations for incomplete data. *Chemometrics and Intelligent Laboratory Systems*, 106(1):41–56, mar 2011.

[4] Evrim Acar, Evangelos E Papalexakis, Gözde Gürdeniz, Morten A Rasmussen, Anders J Lawaetz, Mathias Nilsson, and Rasmus Bro. Structure-revealing data fusion. *BMC bioinformatics*, 15(1):239, jan 2014.

[5] Gaoyun An, Shuai Liu, and Qiuqi Ruan. A sparse neighborhood preserving non-negative tensor factorization algorithm for facial expression recognition. *Pattern Analysis and Applications*, aug 2015.

[6] Anima Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. pages 2773–2832, oct 2012.

[7] A. V. Anisimov, O. O. Marchenko, and T. G. Vozniuk. Determining Semantic Valences of Ontology Concepts by Means of Nonnegative Factorization of Tensors of Large Text Corpora. *Cybernetics and Systems Analysis*, 50(3):327–337, jun 2014.

[8] P. Antolin, A. Buffa, F. Calabrò, M. Martinelli, and G. Sangalli. Efficient matrix computation for tensor-product isogeometric analysis: The use of sum factorization. *Computer Methods in Applied Mechanics and Engineering*, 285:817–828, mar 2015.

[9] C. J. Appellof and E. R. Davidson. Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents. *Analytical Chemistry*, 53(13):2053–2056, nov 1981.

[10] Javier Arnedo, Daniel Mamah, David A Baranger, Michael P Harms, Deanna M Barch, Dragan M Svrakic, Gabriel A de Erausquin, C Robert Cloninger, and Igor Zwir. Decomposition of

brain diffusion imaging data uncovers latent schizophrenias with distinct patterns of white matter anisotropy. *NeuroImage*, 120:43–54, oct 2015.

[11] Tom Barker and Tuomas Virtanen. Semi-supervised non-negative tensor factorisation of modulation spectrograms for monaural speech separation. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 3556–3561. IEEE, jul 2014.

[12] E Benetos and C Kotropoulos. Non-Negative Tensor Factorization Applied to Music Genre Classification. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(8):1955–1967, nov 2010.

[13] Peter Benner, Venera Khoromskaia, and Boris N. Khoromskij. A reduced basis approach for calculation of the Bethe–Salpeter excitation energies by using low-rank tensor factorisations. *Molecular Physics*, pages 1–14, feb 2016.

[14] A. Bernardi, J. Brachat, P. Comon, and B. Mourrain. General tensor decomposition, moment matrices and applications. *Journal of Symbolic Computation*, 52:51–71, may 2013.

[15] Cagdas Bilen, Alexey Ozerov, and Patrick Pérez. Automatic Allocation of NTF Components for User-Guided Audio Source Separation, mar 2016.

[16] Wenfei Cao, Yao Wang, Can Yang, Xiangyu Chang, Zhi Han, and Zongben Xu. Folded-concave penalization approaches to tensor completion. *Neurocomputing*, 152:261–273, mar 2015.

[17] J. Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3):283–319, sep 1970.

[18] Raymond B. Cattell. "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika*, 9(4):267–283, dec 1944.

[19] Anwesha Chaudhury, Ivan Oseledets, and Rohit Ramachandran. A computationally efficient technique for the solution of multidimensional PBMs of granulation via tensor decomposition. *Computers & Chemical Engineering*, 61:234–244, feb 2014.

[20] Yi-Lei Chen, Chiou-Ting Hsu, and Hong-Yuan Mark Liao. Simultaneous tensor decomposition and completion using factor priors. *IEEE transactions on pattern analysis and machine intelligence*, 36(3):577–91, mar 2014.

[21] I. Cherif and F. Fnaiech. Blind identification of a second order Volterra-Hammerstein series using cumulant cubic tensor analysis. In *2008 34th Annual Conference of IEEE Industrial Electronics*, pages 1851–1856. IEEE, nov 2008.

[22] Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and HUY ANH PHAN. Tensor Decompositions for Signal Processing Applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32(2):145–163, mar 2015.

[23] Andrzej Cichocki, Rafal Zdunek, Seungjin Choi, Robert Plemmons, and Shun-ichi Amari. Non-Negative Tensor Factorization using Alpha and Beta Divergences. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 3, pages III–1393–III–1396. IEEE, 2007.

[24] Pierre Comon. Tensors : A brief introduction. *IEEE Signal Processing Magazine*, 31(3):44–53, may 2014.

[25] Stephane Dartois. Tensor Models: extending the matrix models structures and methods. page 13, mar 2016.

[26] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A Multilinear Singular Value Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, jan 2000.

[27] CHQ Ding, X He, and HD Simon. On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering. In *2005 SIAM International Conference on Data Mining - SDM*, 2005.

[28] Weiyang Ding, Liqun Qi, and Yimin Wei. Fast Hankel tensor-vector product and its application to exponential data fitting. *Numerical Linear Algebra with Applications*, 22(5):814–832, oct 2015.

[29] I Domanov and L De Lathauwer. Canonical polyadic decomposition of third-order tensors: relaxed uniqueness conditions and algebraic algorithm. *arXiv preprint arXiv:1501.07251*, 2015.

[30] I Domanov and LD Lathauwer. Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition. *SIAM Journal on Matrix Analysis and Applications*, 2014.

[31] Ignat Domanov and Lieven De Lathauwer. On the Uniqueness of the Canonical Polyadic Decomposition of Third-Order Tensors—Part II: Uniqueness of the Overall Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 34(3):876–903, jul 2013.

[32] Pavel Dourbal. Synthesis of fast multiplication algorithms for arbitrary tensors. page 79, feb 2016.

[33] Lisette Espín-Noboa, Florian Lemmerich, Philipp Singer, and Markus Strohmaier. Discovering and Characterizing Mobility Patterns in Urban Spaces: A Study of Manhattan Taxi Data. *arXiv preprint*, jan 2016.

[34] Hadi Fanaee-T and Joao Gama. Tensor-based anomaly detection: An interdisciplinary survey. *Knowledge-Based Systems*, feb 2016.

[35] M Figueiredo. Exploring the performance of non-negative multi-way factorization for household electrical seasonal consumption disaggregation. *Neural Networks (IJCNN) ...*, 2014.

[36] D. Fitzgerald, M. Cranitch, and E. Coyle. Shifted 2D Non-negative Tensor Factorisation, 2006.

[37] RA Harshman. Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 1970.

[38] T. Hazan, S. Polak, and A. Shashua. Sparse image coding using a 3D non-negative tensor factorization. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 50–57 Vol. 1. IEEE, 2005.

[39] Frank L. Hitchcock. The Expression of a Tensor or a Polyadic as a Sum of Products. *Journal of Mathematics and Physics*, 6(1-4):164–189, apr 1927.

[40] Arie Kapteyn, Heinz Neudecker, and Tom Wansbeek. An approach ton-mode components analysis. *Psychometrika*, 51(2):269–275, jun 1986.

[41] Tamara G. Kolda and Brett W. Bader. Tensor Decompositions and Applications. *SIAM Review*, 51(3):455–500, aug 2009.

[42] Brian Kulis, Mátyás Sustik, and Inderjit Dhillon. Learning low-rank kernel matrices. In *Proceedings of the 23rd international conference on Machine learning - ICML '06*, pages 505–512, New York, New York, USA, jun 2006. ACM Press.

[43] Yang Li, Yangzhou Du, and Xueyin Lin. Kernel-based multifactor analysis for image synthesis and recognition. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 114–119 Vol. 1. IEEE, 2005.

[44] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–20, jan 2013.

[45] J Möcks. Topographic components model for event-related potentials and some biophysical considerations. *IEEE transactions on bio-medical engineering*, 35(6):482–4, jun 1988.

[46] Morten Mørup. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):24–40, jan 2011.

[47] Morten Mørup, Lars Kai Hansen, and Sidse M Arnfred. Algorithms for sparse nonnegative Tucker decompositions. *Neural computation*, 20(8):2112–31, aug 2008.

[48] Roman Rosipal, Mark Girolami, Leonard J. Trejo, and Andrzej Cichocki. Kernel PCA for Feature Extraction and De-Noising in Nonlinear Regression. *Neural Computing & Applications*, 10(3):231–243, dec 2001.

[49] Amnon Shashua and Tamir Hazan. Non-negative tensor factorization with applications to statistics and computer vision. In *Proceedings of the 22nd international conference on Machine learning - ICML '05*, pages 792–799, New York, New York, USA, aug 2005. ACM Press.

[50] John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[51] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, sep 1966.

[52] LR Tucker. Implications of factor analysis of three-way matrices for measurement of change. *Problems in measuring change*, 1963.

[53] LR Tucker. The extension of factor analysis to three-dimensional matrices. *Contributions to mathematical psicology*, 1964.

[54] M. Alex O. Vasilescu and Demetri Terzopoulos. Multilinear Analysis of Image Ensembles: TensorFaces. In Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen, editors, *Computer Vision — ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, Berlin, Heidelberg, apr 2002. Springer Berlin Heidelberg.

[55] Yu-Xiong Wang and Yu-Jin Zhang. Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Transactions on Knowledge and Data Engineering*, 25(6):1336–1353, jun 2013.

[56] Max Welling and Markus Weber. Positive tensor factorization. *Pattern Recognition Letters*, 22(12):1255–1261, oct 2001.

[57] S Zafeiriou and M Petrou. Nonnegative tensor factorization as an alternative Csiszar–Tusnady procedure: algorithms, convergence, probabilistic interpretations and novel probabilistic tensor. *Data Mining and Knowledge Discovery*, 2011.