International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Tensors for Data Mining – A review of Real World Applications

Anupma Sinha
Department of CSE
Shri Venkateshwara University
sinhaanupma3@gmail.com

Dr. Anshu Srivastava
Department of CSE
Shri Venkateshwara University
anshuqrat114@gmail.com

***Abstract*--Tensors and tensor decompositions are very powerful and versatile tools that can model a wide variety of heterogeneous, multi aspect data. As a result, tensor decompositions, which extract useful latent information out of multi aspect data tensors, have witnessed increasing popularity and adoption by the data mining community. In the present work an overview of a very broad spectrum of applications where tensors have been instrumental in achieving state-of-the-art performance, ranging from social network analysis to brain data analysis, and from web mining to healthcare, has been carried out. Finally, a list of challenges and open problems that outline exciting future research directions have been dealt here.**

Key Words: Tensor Decomposition, Multiaspect data, Web Mining

1. Introduction:

Tensors are multidimensional extensions of matrices. Because of their ability to express multimodal or multiaspect data, they are very powerful tools in applications that inherently create such data. For instance, in online social networks, people tend to interact with each other in a variety of ways: they message each other, they post on each other's pages, and so on. All these different means of interaction are different aspects of the same social network of people, and can be modeled as a three-mode tensor, a "data cube," of (user, user, means of interaction). Given this tensor, there exists a rich variety of tools called tensor decompositions or factorizations that are able to extract meaningful, latent structure in the data. In data mining, tensor decompositions have been very popular and successful in achieving state-of-the-art performance. The list of applications where tensors have been successful ranges from social network analysis to brain data analysis, and from web mining and information retrieval to healthcare analytics. The present work emphasize the implications of the works summarized from a practitioner's point of view, in terms of describing both the decompositions and the volume and breadth of the applications that this survey contains. In this paper a review of the application of tensor decompositions in a broad variety of real-world applications, outlining the particular ways that tensors have been used in each case, has been dealt. Finally, the work has been concluded by highlighting a few open challenges in tensor decompositions that mark interesting future research directions.

2. Data Mining Applications

Tensors are very powerful and versatile tools, as demonstrated by the long list of their applications in data mining. In this section, we cover a wide spectrum of such applications: social and collaboration network analysis, web mining and web search, knowledge bases, information retrieval, topic modeling, brain data analysis, recommendation systems, urban computing, healthcare and medical applications, computer networks, speech and image processing, and computer vision. For each application, the topic dealt includes focus on what the problem formulation is, how a tensor is modelled, and which decomposition is used, and discuss the results.

2.1. Social and Collaboration Network Analysis

Social and collaboration network analysis can benefit from modelling data as tensors in various ways: when there exist multiple "views" of the network, can be expressed as a three-mode tensor with each frontal slice being an adjacency matrix of the network for a particular view. Furthermore, tensors have been used in modelling time-evolving networks, where each frontal slice of the tensor is a snapshot of the network for a particular point in time. Tensor applications in social networks date back to Acar et al. [2005], which is also one of the earliest tensor applications in data mining. In this particular work, the authors analyze IRC chat-room data and create a synthetic tensor data generator that mimics the properties of real chat-room data. The tensors are of the form (user, keyword, time), and for the purposes of demonstrating the utility of expressing the data as a tensor, the authors also create (user, keyword) and (user, time) matrices.

Kolda and Sun [2008] model the DBLP collaboration network as a tensor and use the Tucker decomposition to identify groups of authors who publish on similar topics and on similar conferences.

In addition to the social interactions, Lin et al. [2009] demonstrate that using the context behind those interactions can improve the accuracy in discovering communities. In this work, the authors define a "Metagraph," a graph that encodes the context of social interactions, and subsequently propose a Metagraph Factorization, which essentially boils down to a coupled tensor and matrix factorization, involving all the various tensors and matrices that capture relations in the Metagraph.

Papalexakis et al. [2012] apply the CP decomposition to a Facebook Wall posts database that forms a tensor and identify interesting patterns and their temporal evolution.

Anandkumar et al. [2014] show that orthogonal CP can be used to estimate latent factor models such as Latent Dirichlet Allocation (LDA) and Mixed Membership Block Models (MMBMs) using themethod ofmoments.

Huang et al. [2013] use orthogonal CP for recovering a Mixed Membership Block Model that detects overlapping communities in social networks and apply it to a Facebook social network dataset (two-mode). This is a very interesting alternative perspective because the authors use the tensor decomposition to identify a model on a two-mode graph and do not use the tensor decomposition on the graph itself.

Schein et al. [2015] use dyadic events between countries in order to discover multilateral relations among them. In particular, they propose a Bayesian version of the CP decomposition, postulating a Poisson distribution on the data, which has been shown to be more effective when dealing with sparse, count data.

2.2. Web Mining and Web Search

Kolda et al. [2005] extend Kleinberg's HITS algorithm for authoritativeness and hubness scores of webpages, including context information on the link. In particular, for each link, they use the anchor text as the context. This creates a three-mode tensor of (webpage, webpage, anchor text), and the CP decomposition gives the authoritativeness and hubness scores (**A** denotes the authorities and **B** the hubs, and **C** encodes the topic of each set of hubs and authorities). This, in contrast to plain HITS (which can be seen as an SVD of the hyperlink matrix), provides more intuitive interpretation. The authors demonstrate the superiority of the approach in identifying topically coherent groups of webpages by applying it to a custom crawl of the web, emulating what a commercial search engine does.

Sun et al. [2005] personalize web search by using historic clic-through data of users. They construct a (user, query, page) tensor that records the clicks of a user to a particular result of a query, and they use HOSVD to take a low-rank factorization of the click-through tensor. By reconstructing the tensor from its HOSVD, they fill in missing values, which can then be used as personalized result recommendations for a particular user.

Agrawal et al. [2015a] model the comparison between the results of different search engines using tensors. For a set of queries, they create a (query, keyword, date, search engine) tensor and use the CP decomposition to create latent representations of search engines in the same space. They apply this tool to compare Google and Bing web search and find that for popular queries, the two search engines have high overlap.

Subsequently, Agrawal et al. [2015b] apply the same methodology to compare Google and Twitter-based search, finding that the overlap is lower, showing the potential for social-media-based web search.

2.3. Knowledge Bases, Information Retrieval, and Topic Modelling

Chew et al. [2007] tackle the problem of Cross-language Information Retrieval, where we have parallel documents in different languages and we need to identify latent topics as in Latent Semantic Analysis (LSA); in a nutshell, LSA refers to taking a low-rank Singular Value Decomposition of a (term, document) matrix and exploiting the low-rank structure to identify synonyms.

Kang et al. [2012] and Papalexakis et al. [2012] apply the CP decomposition to data coming from the Read the Web project. In particular, the data are in the form (noun-phrase, noun-phrase, context-phrase). Using the CP decomposition, the authors identify latent topics that are semantically and contextually coherent, coming from each one of the rank-one components of the decomposition. Furthermore, the factor matrices of the CP decomposition define latent embeddings of the noun-phrases to a concept space; the authors find similar noun-phrases in that concept space, which are essentially "contextual synonyms," for example, noun-phrases that can be used in the same semantics/context.

Jeon et al. [2015] apply the Tucker decomposition to a particular snapshot of the Freebase knowledge base that contains entities and relations aboutmusic. The authors use the factor matrices of Tucker as latent embeddings (as in the case of the CP factor matrices) and identify semantically coherent latent concept entities and relations.

Chang et al. [2014] use a RESCAL inspired decomposition to model and analyze knowledge base data. The new additions to the model are type constraints: each entity of a knowledge base has a known type (e.g., "person"); therefore, these type constraints are explicitly included in the model by excluding triples of entity-relation-entity with incompatible types from the optimization. This both saves computation and produces a more accurate result.

2.4. Brain Data Analysis

Acar et al. [2007] analyze electroencephalogram (EEG) data from patients with epilepsy in order to localize the origin of the seizure. To that end, they model the EEG data using a three-mode (time samples, scales, electrodes) tensor (after preprocessing the EEG measurements via a wavelet transformation). In order to analyze the EEG tensor, they use the CP decomposition: when they identify a potential seizure (which has signatures on the time and frequency domains), they use the factor vector of the third mode (the "electrodes" mode) to localize that activity. In some cases, the data contain artifacts that may shadow the seizures from the CP decomposition (such as activity caused by the movement of the eyes) and therefore have to be removed. In order to remove those artifacts, the authors use the Tucker3 decomposition, which can capture the subspace variation for each mode better than CP (due to its increased degrees of freedom, which in turn make it harder to interpret).

Davidson et al. [2013] have MRI measurements over time and wish to estimate regions of the brain and the connections between them. They model the data as a tensor (x-coordinates, y-coordinates, time) and assume that using a CP decomposition, each rank-one tensor gives a particular region of the brain (the first two modes in space and the third in time). In order to guide the decomposition to find the right structure in the brain, they use linear constraints for the spatial factors of the decomposition according to groups of voxels in the brain that are known to behave in a coherent manner. Applying those constraints, the authors are able to detect nodes and the network between those nodes with higher accuracy.

He et al. [2014] extend supervised learning models such as Support Vector Machines to operate on tensors as opposed to vectors or points. They leverage the fact that data such as MRI brain scans have an inherent tensor structure that should be exploited and propose a Kernel SVM that uses the CP decomposition of the data points as a compact representation that preserves the structure.

2.5. Recommendation Systems

One of the first attempts to apply tensors to collaborative filtering and recommendation systems is Xiong et al. [2010]. The authors propose to extend Bayesian Probabilistic Matrix Factorization (BPMF) which is widely used in Recommendation Systems in the case where we have temporal information. They propose a Bayesian Probabilistic Tensor Factorization (BPTF) that is based on the CP model.

In a similar spirit as earlier and around the same time, Karatzoglou et al. [2010] propose to use context (such as time) in traditional user-item recommendation scenarios by modeling the data as a tensor.

Rendle [2010] introduces Factorization Machines, a generalization of Support Vector Machines that parameterizes the data internally using a factorization model instead of using the raw data. The Factorization Machine can have degree 2 or higher. In the case of degree 2, the internal factorization is a bilinear matrix factorization, whereas for higher degrees, it uses a CP model.

Finally, Pantraki and Kotropoulos [2015] work on image and tag recommendation on Flickr. They model the data as a multiset where we have three matrices: (image, feature), (image, tag keyword), and (image, user). Since the data do not strictly form a tensor, the authors apply PARAFAC2 to this dataset and demonstrate its ability to obtain a joint low-rank representation of this dataset, which can provide high-quality image and tag recommendations.

2.6. Urban Computing

Urban Computing refers to a class of applications that study human activity and mobility within a city, with the ultimate goal to improve the livability of an urban environment. Mu et al. [2011] use historical data from a metropolitan area in order to forecast areas with potential future crime activity. They model the data as a four-mode tensor, where the first three modes are (longitude, latitude, time) and the fourth mode consists of features such as residential burglary information, social events, and offender data. They use a form of Tucker decomposition to obtain a low-rank representation of that tensor, and they use that representation for linear discriminant analysis to predict future crime activity.

In Wang et al. [2014], the problem the authors solve is the one of estimating the travel time of a particular trajectory in a city road network. In order to do so, they use real GPS data of travel times by a set of drivers. The issue that arises with these types of data is the high degree of sparsity, since many road segments may have not been traversed at all (or sometimes ever); thus, trying to estimate travel times for all road segments this way may result in inaccurate measurements. To alleviate this data sparsity, the authors propose to use historical data for those GPS trajectories, as well as side information about time slots and road segments, to fill in missing travel time values.

Zheng et al. [2014] analyze noise complaint data from New York City in order to identify the major sources of noise pollution in the city for different times of the day and the week. The issue with the human-generated complaints is that they result in very sparse data where some regions are overrepresented and some are underrepresented or not present at all.

Finally, the work of Zhang et al. [2015] addresses the problem of exploring, analyzing, and estimating drivers' refueling behavior in an urban setting for better planning (e.g., placement of gas stations) and recommendation of nearby gas stations with minimal wait time.

2.7. Healthcare and Medical Applications

Ho et al. [2014] use a tensor-based technique to automatically derive phenotype candidates from electronic health records. Intuitively, the problem is to automatically identify groups of patients who have similar diagnoses and have undergone similar procedures. In order to do that, they propose a tensor decomposition based on the CP model, where each phenotype candidate is a rank-one component of the CP decomposition of a (patient, diagnosis, procedure) tensor.

The work of Perros et al. [2015] is the first datamining application of the Hierarchical Tucker (H-Tucker) model of 3.4. In particular, the authors propose a sparse version of H-Tucker and apply it to a disease phenotyping problem, much like the one addressed by Ho et al. [2014]. The difference is that here, the authors use co-occurrence data of patients exhibiting the same disease within a large database ofmedical records.

Mohammadi et al. [2016] consider the problem of network alignment with a goal of preserving triangles across the aligned graphs. This has applications in comparative interactomics in biology.

2.8. Computer Networks

Maruhashi et al. [2011] use the CP decomposition to analyze network traffic data from Lawrence Berkeley National Labs (LBNL) forming a tensor of (source IP, destination IP, port #, timestamp) where each entry indicates a connection between the two IP addresses on a given port for a given time tick. Using the factor matrix corresponding to the "timestamp" mode, the authors propose a spike detection algorithm on the temporal profile of the latent components of the decomposition, identifying anomalous connections in the data.

Finally, Mao et al. [2014] analyze two types of network data: (1) Honeynet Data of (source IP, destination IP, timestamp) and (2) Intrusion Detection System (IDS) logs of (event type, timestamp, target IP). They apply the CP decomposition, and using clustering methods on the factor matrices, for different temporal resolutions, they are able to identify anomalous events, outperforming state-of-the-art IDS systems.

2.9. Speech and Image Processing and Computer Vision

Nion et al. [2010] use the CP decomposition to conduct Blind Source Separation (BSS). BSS is the problem of estimating a set of signals that are mixed by an unknown channel (hence "blind"), using solely the information on the receiver's end as measured by a set of sensors. Liu et al. [2013] propose a tensor-based method for completing missing values in series of images. In order to do so, they define the trace norm for the tensor case and extend matrix completion algorithms that use the matrix trace norm. The authors compare their proposed method against Tucker, CP, and doing SVD on each tensor slice separately, and they demonstrate that their proposed method achieves superior performance.

Finally, Tao et al. [2008] propose a Bayesian version of HOSVD and use it in order to tackle model 3D face data. In particular, using their proposed decomposition in ensembles of 3D face images, they compute a parsimonious representation of the 3F faces in the form of core tensor, which captures the latent characteristics of the 3D faces, which is sufficient to later reconstruct any particular face with a given expression.

3. Conclusions And Future Challenges

Tensor decompositions are very versatile and powerful tools, ubiquitous in data mining applications. They have been successfully integrated in a rich variety of real world applications, and due to the fact that they can express and exploit higher-order relations in the data, they tend to outperform approaches that ignore such structure. Furthermore, recent advances in scaling up tensor decompositions have employed practitioners with a strong arsenal of tools that can be applied to many big multiaspect data problems. The success that tensors have experienced in data mining during the last few years by no means indicates that all challenges and open problems have been addressed. Quite to the contrary, there exist challenges, some of which are summarized below, which delineate very exciting future research directions:

Modeling space and time: What is the best way to exploit spatial or temporal structure that exists in the data? Is there a generic way to incorporate such modifications in a tensor model and enable it to handle spatiotemporal data effectively? Furthermore, another open problem in spatiotemporal data modeling is selecting the right granularity for the space and time modes.

Unsupervised model selection: In a wide variety of applications, ground truth is not easy to obtain; however, one need to have unsupervised means of understanding which tensor decomposition is more appropriate (e.g., CP vs. Tucker vs. DEDICOM etc.) and, given a decomposition, what model order is most appropriate for the data.

Dealing with high-order data: Many real-world applications involve data that can be represented as very high-order tensors.

Connections with Heterogeneous Information Networks: In data mining, there exists a very rich line of work on Heterogeneous Information Networks (HINs), which are graphs between different types of nodes, connected with various types of edges. An HIN can be represented as a tensor, and in fact, a multiview social network is such an HIN. Outlining connections between such works and tensor decompositions is a very interesting future direction that aims toward unifying different data mining approaches.

References

[1] Tamara G. Kolda and Jimeng Sun. 2008. Scalable tensor decompositions for multi-aspect data mining. In 8th IEEE International Conference on Data Mining, 2008 (ICDM'08). IEEE, 363–372.

[2] Evrim Acar, Canan Aykut-Bingol, Haluk Bingol, Rasmus Bro, and B¨ ulent Yener. 2007. Multiway analysis of epilepsy tensors. Bioinformatics 23, 13 (2007), i10–i18.

[3] Evrim Acar, Evangelos E. Papalexakis, Morten A. Rasmussen, Anders J. Lawaetz, Mathias Nilsson, and Rasmus Bro. 2014. Structure-revealing data fusion. BMC Bioinformatics 15, 1 (2014), 239.

[4] Rakesh Agrawal, Behzad Golshan, and Evangelos Papalexakis. 2015a. A study of distinctiveness in web results of two search engines. In 24th International Conference on World Wide Web, Web Science Track.

[5] Rakesh Agrawal, Behzad Golshan, and Evangelos Papalexakis. 2015b. Whither social networks for web search? In 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Sydney, Australia.

[6] J. Douglas Carroll and Jih-Jie Chang. 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. Psychometrika 35, 3 (1970), 283–319.

[7] Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and Huy Anh Phan. 2015. Tensor decompositions for signal processing applications: From twoway to multiway component analysis. IEEE Signal Processing Magazine 32, 2 (2015), 145–163.

[8] Ian Davidson, Sean Gilpin, Owen Carmichael, and Peter Walker. 2013. Network discovery via constrained tensor analysis of fMRI data. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 194–202.

[9] Joyce C. Ho, Joydeep Ghosh, and Jimeng Sun. 2014. Marble: High-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 115–124.

[10] Furong Huang, Sergiy Matusevych, Anima Anandkumar, Nikos Karampatziakis, and Paul Mineiro. 2014. Distributed latent Dirichlet allocation via tensor factorization. Neural Information Processing Systems (NIPS) Optimization Workshop 2014.

[11] Ji Liu, Przemyslaw Musialski, PeterWonka, and Jieping Ye. 2013. Tensor completion for estimating missing values in visual data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35, 1 (2013), 208–220.

[12] Shahin Mohammadi, David Gleich, Tamara Kolda, and Ananth Grama. 2016. Triangular alignment (TAME): A tensor-based approach for higher-order network alignment.

IEEE/ACM Transactions on Computational Biology and Bioinformatics (2016).
[13] Steffen Rendle. 2010. Factorization machines. In 2010 IEEE 10th International Conference on Data Mining (ICDM'10). IEEE, 995–1000.
[14] Jian-Tao Sun, Hua-Jun Zeng, Huan Liu, Yuchang Lu, and Zheng Chen. 2005. CubeSVD: A novel approach to personalized web search. In Proceedings of the 14th International Conference on World Wide Web. ACM, 382–390.
[15] Q. Zhao, L. Zhang, and A. Cichocki. 2015. Bayesian CP factorization of incomplete tensors with automatic rank determination. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37 (2015), 1751–1763.
[16] Yu Zheng, Tong Liu, Yilun Wang, Yanmin Zhu, Yanchi Liu, and Eric Chang. 2014. Diagnosing New York City's noises with ubiquitous data. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 715–725.