2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

APTERA: Automatic PARAFAC2 Tensor Analysis

Ekta Gujral
University of California Riverside
Email: egujr001@ucr.edu

Evangelos E. Papalexakis
University of California Riverside
Email: epapalex@cs.ucr.edu

***Abstract*—In data mining, PARAFAC2 is a powerful and a multi-layer tensor decomposition method that is ideally suited for unsupervised modeling of data which forms "irregular" tensors, e.g., patient's diagnostic profiles, where each patient's recovery timeline does not necessarily align with other patients. In real-world applications, where no ground truth is available, how can we automatically choose how many components to analyze? Although extremely trivial, finding the number of components is very hard. So far, under traditional settings, to determine a reasonable number of components, when using PARAFAC2 data, is to compute decomposition with a different number of components and then analyze the outcome manually. This is an inefficient and time-consuming path, first, due to large data volume and second, the human evaluation makes the selection biased.**

In this paper, we introduce APTERA, a novel automatic PARAFAC2 tensor mining that is based on locating the L-curve corner. The automation of the PARAFAC2 model quality assessment helps both novice and qualified researchers to conduct detailed and advanced analysis. We extensively evaluate APTERA 's performance on synthetic data, outperforming existing state-of-the-art methods on this very hard problem. Finally, we apply APTERA to a variety of real-world datasets and demonstrate its robustness, scalability, and estimation reliability.

I. INTRODUCTION

Tensors are the generalization of vectors and matrices. They are ubiquitous (e.g. images, videos, and social networks) and ever-increasing in popularity. With the opportunity to handle large volumes and velocity of data as a result of recent technical developments, such as mobile connectivity , digital tools, biomedical technology, and modern medical testing techniques, we face multi-source and multi-view data [10] sets. Suppose, for example, that we are given health care record data, such as Centers for Medicare and Medicaid (CMS) [6], and we have information about patients who visited the hospital, or who got what kind of diagnosis in which visit, and when. Time modeling is difficult for the regular tensor factorization methods (e.g CP [4] and Tucker [22]), due to either data irregularity or time-shifted latent factor appearance of such data. Hence, such data is formulated as a 3-mode PARAFAC2 tensor [12]. PARAFAC2 decomposition is able to handle various chromatographic data and choosing the correct number of components allows it to separate each variability source by using spectral information. Consider amino acid data [16] where three compounds tyrosine, tryptophan and phenylalanine dissolved in phosphate-buffered water. In Figure (1), PARAFAC2 decomposition with rank-3 resembles the

IEEE/ACM ASONAM 2022, November 10-13, 2022
http://dx.doi.org/10.1145/XXXXXXX.XXXXXXX
978-1-6654-5661-6/22/$31.00 © 2022 IEEE

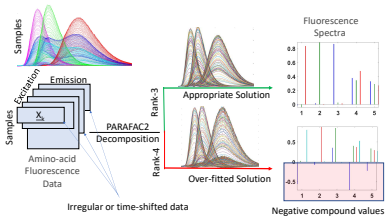

Figure. 1: Amino acid data PARAFAC2 decomposition.

pure spectra of tryptophan, tyrosine and phenylalanin. When PARAFAC2 decomposition with rank-4 is applied to this data, the fourth component does not resemble any of the compounds and in fact, it does not seem to reflect any chemical information. Therefore, it becomes very important to select the correct number of components to solve real-world problems.

In literature, one popular approach to find the rank of CP tensor is core consistency diagnostic (CORCONDIA) [3]. The CORCONDIA essentially assesses significant deviations from a super-diagonal core tensor. This would suggest that the CP decomposition is not optimal either because the selected rank is not correct, or the CP model cannot describe the data well enough. This approach is widely studied and explored among the tensor mining community. AutoTen [19] is a powerful method that uses CORCONDIA as a building block to provide unsupervised detection of multi-linear low-rank structure in tensors. Over the last few years, there has been various methods [21] proposed to find the number of component of fixed dimension tensor data. However, only one method namely Autochrome [15] estimates rank for irregular data. Unfortunately, this method uses various computation diagnostics that require the conversion of irregular data to regular data. This is expensive in terms of memory utilization.

To fill the gap, we propose a novel method APTERA to estimate the rank of irregular 'PARAFAC2' data that discover the number of components (interchangeably rank) through higher-order singular values decomposition (HOSVD).

II. BACKGROUND

In this section, we provide the necessary background for notations and tensor operations. Then, we briefly discuss the related work regarding the PARAFAC2 decomposition for tensor factorization and rank estimation method available in the literature. Table (I) contains the symbols used throughout the paper.

Symbols	Definition
$\underline{\mathbf{X}}, \mathbf{X}, \mathbf{x}, x$	Tensor, Matrix, Column vector, Scalar
$\mathbf{X}^T, \mathbf{X}^{-1}, \mathbf{X}^\dagger$	Transpose, Inverse, Pseudo-inverse
$diag(\mathbf{X})$	Extract diagonal of matrix $\mathbf{X}$
$\underline{\mathbf{X}}_k$	shorthand for $\underline{\mathbf{X}}(:,:,k)$ (k-th frontal slice of $\underline{\mathbf{X}}$)
$\underline{\mathbf{X}}^{(n)}, \mathbf{X}^{(n)}$	mode-n matricization of $\underline{\mathbf{X}}$, matrix $\mathbf{X}$ at mode-n
$\|\mathbf{A}\|_F, \|\mathbf{a}\|_2$	Frobenius norm, ℓ_2 norm
$\circ, \circledast, \otimes, \odot$	Outer, Hadmard, Kronecker and Khatri-Rao product
OoM	Out of Memory
MTTKRP	Matricized tensor times Khatri-Rao product[17]

TABLE I: Table of symbols and their description

A. Brief Introduction to PARAFAC2 Tensor Decomposition methods

The PARAFAC2 model was first developed by Harshman [12] to handle the situation where the number of observations (row dimension) in each $\underline{\mathbf{X}}_k$ may vary e.g study of phonetics. In his work, Harshman described a way to factorize multiple matrices simultaneously given that one factor was not exactly the same in all those matrices. This can be solved by imposing orthogonality constraints on a linear transformation as a coupling relationship between the similar factors to ensure identifiability. Hence, the PARAFAC2 model for 3-mode tensor $\underline{\mathbf{X}}_k \in \mathbb{R}^{I_k \times J}$ is given by:

$$\begin{aligned}\mathcal{L} =& \text{argmin}_{\mathbf{U},\mathbf{S},\mathbf{V}} \frac{1}{2} ||\underline{\mathbf{X}}_k - \mathbf{U}_k \mathbf{S}_k \mathbf{V}^T||_2^F \quad \forall k \\ & \text{subject to} \quad \mathbf{U}_k = \mathbf{Q}_k * \mathbf{H} \text{ and } \mathbf{Q}_k \mathbf{Q}_k^T = \mathbf{I}_r\end{aligned} \tag{1}$$

where $\mathbf{U}_k \in \mathbb{R}^{I_k \times R}$ are coupled matrices, $\mathbf{H} \in \mathbb{R}^{R \times R}$ is coefficients matrix, $\mathbf{Q}_k \in \mathbb{R}^{I_k \times R}$ are left-orthogonal coupling matrices to ensure uniqueness of factors and $\mathbf{W} = \mathbf{S}_k \in \mathbb{R}^{R \times R}$ is set of diagonal matrix. The Equ (1) in form of orthogonal form can be re-written as :

$$\begin{aligned}\mathcal{L} =& \text{argmin}_{\mathbf{Q}} \frac{1}{2} ||\underline{\mathbf{X}}_k - \mathbf{Q}_k \mathbf{H} \mathbf{W} \mathbf{V}^T||_2^F \quad \forall k \\ & \text{subject to} \quad \mathbf{Q}_k \mathbf{Q}_k^T = \mathbf{I}_r\end{aligned} \tag{2}$$

To solve Eq (2), most common method is Alternating Least Square (ALS) that updates $\mathbf{Q}_k$ by fixing other factor matrices i.e $\mathbf{H}, \mathbf{W}$, and $\mathbf{V}$. The orthogonal coupling matrix $\mathbf{Q}_k$ can be obtained by Singular Value decomposition (SVD) of $(\mathbf{H}\mathbf{W}\mathbf{V}^T \underline{\mathbf{X}}_k^T) = [\mathbf{P}_n, \Sigma_n, \mathbf{Z}_n^T]$. With $\mathbf{Q}_k^T = \mathbf{P}_n \mathbf{Z}_n^T$ fixed, the rest of factors can be obtained as:

$$\begin{aligned}\mathcal{L} =& \text{argmin}_{\mathbf{H},\mathbf{W},\mathbf{V}} \frac{1}{2} ||\mathbf{Q}_k \underline{\mathbf{X}}_k - \mathbf{H}\mathbf{W}\mathbf{V}^T||_2^F \, s.t. \mathbf{Q}_k \mathbf{Q}_k^T = \mathbf{I}_r \\ & \text{argmin}_{\mathbf{H},\mathbf{W},\mathbf{V}} \frac{1}{2} ||\underline{\mathbf{Y}} - \mathbf{H}\mathbf{W}\mathbf{V}^T||_2^F\end{aligned} \tag{3}$$

The Eq. (2) is equivalent of solving CP decomposition of $\underline{\mathbf{Y}}$ using ALS method. The author [20] proposed method namely Scalable PARAFAC2 for large and sparse tensors. The speed up of the process is obtained by modifying core computational kernel. We use improved version of PARAFAC2 decomposition [20] for our method.

B. Brief Introduction to Automatic Tensor Mining

As outlined in the introduction, rank detection and low-rank structure discovery are very hard problems, and there are currently no general-purpose methods that can achieve these tasks efficiently. There is very limited work done for PARAFAC2 data rank estimation. There exists a method named Autochrome [15] which uses PARAFAC2 decomposition for estimating the rank of tensor data. The method is based on a number of model diagnostics (quality criteria) collected from models with different numbers of factors. They combining these diagnostics to assess what are the appropriate number of components of data. However. this method is limited to gas chromatography–mass spectrometry data and also various diagnostics computations require regular CP tensor as input instead of the irregular tensor.

To our best knowledge, there is no work in the literature that deals with the reveling a number of components of irregular data with PARAFAC2 decomposition without using expensive computations of Core Consistency Diagnostics and not limited to a specific type of data. To fill the gap, we propose a scalable and efficient method that reveals the number of components of the PARAFAC2 model.

III. Proposed Method: Aptera

In data mining applications (e.g. chromatography, health care), we are given a very large irregular multi-layer data which is required to analyze by domain researchers, and we are asked to identify various useful patterns that could potentially help to grow the business or provide valuable insights about data. Most of the time, this analysis is done unsupervised as collecting ground truth is extremely expensive and requires human intervention. Unfortunately, it is not straightforward to determine the proper number of components for PARAFAC2 tensors. Since CORCONDIA based methods have instabilities in the quality estimations[21] and, therefore, we propose a new method for finding the structure in PARAFAC2 tensor data using the L-corner approach that reduces the human intervention and trial-and-error fine-tuning. Our proposed method consists of three steps as described below.

A. PARAFAC2 decomposition

Here, we solve R_{max}-component PARAFAC2 decompositions as given in Equ. (4) by using random initialization. For each decomposition, we keep same initial parameters i.e. number of maximum iterations, tolerance for convergence etc.

$$\begin{aligned}\mathcal{L} =& \sum_{k=1}^{K} \underset{\mathbf{Q}_k}{\text{argmin}} \frac{1}{2} ||\mathbf{X}_k - \mathbf{Q}_k \mathbf{H} \mathbf{W} \mathbf{V}^T||_F^2 \quad \forall k \in [1, K] \\ & \text{subject to} \quad \mathbf{Q}_k \mathbf{Q}_k^T = \mathbf{I}_{R_{max}}\end{aligned} \tag{4}$$

Due to the irregular nature of the first mode of PARAFAC2 data, we use its resultant latent factors to create CP tensors using the Khatri-Rao product on factors $\underline{\mathbf{Y}} = (\mathbf{H} \odot \mathbf{V} \odot \mathbf{W}) \in \mathbb{R}^{R_{max} \times J \times K}$. This gives us a flexibility to use any existing method to discover the rank of the reconstructed tensor. Unfortunately, CORCONDIA based methods like AutoTen [19], Autochrome [15] get confused because the input, i.e the CP tensor, is created using outcome of PARAFAC2 decomposition instead of actual data which could have a different number of components. For example, consider the PARAFAC2 data has total of 10 components and we factorize this data

with $R_{max} = 20$. When we provide the CP tensor with $R_{max} = 20$ to CORCONDIA based methods, it is highly likely possible that Core Consistency diagnostic metric is close to 100% at $R_{max} = 20$, because it can trivially produce "super-diagonal" core. To overcome such instabilities, we use multi-linear orthogonal projections via Higher Order Singular Value Decomposition (HOSVD) for discovering the number of component.

B. Formation of L-curve using Pareto Optimal Truncation

The Singular Value Decomposition (SVD) gives the best low-rank approximation of a matrix. In the sense of multi-linear rank, a generalization of the SVD is the higher-order SVD (HOSVD). Nowadays, it is better known with the effort of de Lathauwer et al. [8], who analyzed the structure of core tensor and proposed to use multi-linearity to discover the rank of the tensor. Motivated by this, we compute HOSVD of $\underline{\mathbf{Y}}$ as given in Equ. (5).

$$[\underline{\mathbf{G}}, \mathbf{A}, \sigma] = HOSVD(\underline{\mathbf{Y}}) \tag{5}$$

where $\underline{\mathbf{G}}$ is decomposed core tensor, $\mathbf{A}$ is set of matrices for each dimension and σ is set of n-mode multi-linear (interchangeably higher-order) non-negative singular values which appear in decreasing order. We can reconstruct 3-mode CP tensor using $\underline{\mathbf{G}}$ and $\mathbf{A}$ as given below Equ. (6).

$$\underline{\mathbf{Y}} = \underline{\mathbf{G}} \times \mathbf{A}^1 \times \mathbf{A}^2 \times \mathbf{A}^3 \tag{6}$$

Selecting the appropriate degree of compression is equivalent to estimating the rank of the tensor. Though the best rank approximation is NP-hard, a satisfying result can always be estimated by choosing a proper degree of truncation. Here, we use Pareto optimal truncation [1], [13] based on the upper bound on the singular values. For any possible 3-mode tensor dimensions, the corresponding relative error E can be defined as

$$vec(E_{rjk}) = \sum_{r=1}^{R_{max}} \sigma\{1\}(r) + \sum_{j=1}^{J} \sigma\{2\}(j) + \sum_{k=1}^{K} \sigma\{3\}(k) \tag{7}$$

$$E(n) = E_{rjk} = \frac{\sqrt{E_{rjk}}}{||\sigma\{1\}||} \tag{8}$$

where $n \in \{1, 2, 3, \ldots, R_{max}JK\}$ is linearized index pairs of R_{max}, J and K e.g. $(n = 1) \leftarrow [r = 1, j = 1, k = 1]$. Next, we define the points on the 2D plane with possible tensor dimension $\mathbf{d}$ as:

$$P(n) = \sqrt{(x(n))^2 + (y(n))^2}, \quad x(n) = ||d(n)||; \quad y(n) = E(n) \tag{9}$$

where $\mathbf{d}$ is a vector of multi-indices and represents as $d(1) \leftarrow [r = 1, j = 1, k = 1]$, $d(2) \leftarrow [r = 2, j = 1, k = 1]$, and so on. Finally, using equation 9, $P(2) = 0.5534$ for tensor dimension $[r = 1, j = 1, k = 2]$. Now, we sort the points P and update residual norm (x) and solution norm (y) accordingly. By eliminating the P values that do not satisfy the monotonic condition, we can get a Pareto front end [11]. Having realized the important roles played by the norms of the solution y and norms of the residual x, it is quite natural to plot these two quantities versus each other, i.e., a trade-off curve. This is precisely the L-curve that can be utilized for estimation of the rank of the tensor. Due to space limitations, pseudocode of computing Pareto front end will be provided in supplementary material.

Dataset	Dimension	Components
Syn-I	$200 \times 500 \times 1000$	5 (Synthetic)
Amino Acid	$5 \times 201 \times 61$	3 (See[16])
Wine-GCMS	$2700 \times 200 \times 44$	4 (See [16])
EU-Core	$986 \times 986 \times 827$	28 (See [23])
CMS	$250 \times 1K \times 98K$	NA

TABLE II: Details for the datasets.

C. Rank Estimation with L-curve Corner

In this step, we use the L-curve corner method [7] to estimate the number of components of a tensor. To improve the efficiency of the method, we can adapt a triangle method [5] that uses geometric properties like the angle and direction of the triangle to estimate the L-curve curvature. Although, above process gives estimated rank for each dimension, but note that PARAFAC2 requires only a single rank value. Therefore, we report minimum rank predicted across tensor regular modes.

IV. Experimental Evaluation

We implemented APTERA in Matlab, using the Tensor Toolbox [2], which supports efficient computations for sparse and dense tensors. We use the public implementation for the algorithm of [21], [15], and we make our code publicly available[1].

A. Synthetic Data Description

A first step in evaluating our method is to check its performance on simulated data whose rank and factors can be predefined. We create synthetic tensors by generating two-factor matrices with R columns each, where their elements are drawn as Gaussian with unit variance. Then, these are normalized column-wise using the l2 norm. The set of factor matrices for irregular mode is created in such a way that it retains the property of orthogonality. By considering these matrices as the PARAFAC2 factor matrices, therefore, the rank of the PARAFAC2 tensor will be exactly R. We considered a setup with 1000 subjects, 500 feature variables, and a maximum of 200 observations for each subject with rank-5. Also, we deformed the generated tensor data by an additive noise tensor that has the rank higher than 5 but has norm $2\times$ less than actual synthetic data.

B. Real Data Description

We evaluate the performance of the proposed method APTERA for the real datasets to assess the practicality in real-world scenarios. For this reason, in our experiments we includes real data sets as shown in Table (II). Details of datasets and full paper can be found at [9].

[1] http://www.cs.ucr.edu/~egujr001/ucr/madlab/src/aptera.zip

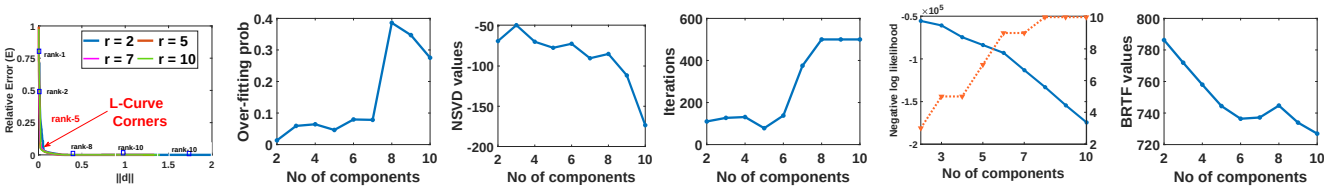

Figure. 2: Baselines comparison on the synthetic dataset. From left to right represents, (a) our proposed method APTERA, (b) Autochrome, (c) NSVD based, (d) Iteration based, (e) Tucker ARD based, and (f) BRTF based method. Also, r indicates no of components.

Dataset	Dimension	Components	Noisy Components	Density (%)	APTERA Predicted Rank	Autochrome Predicted Rank	NSVD Predicted Rank
Syn-I	$200 \times 500 \times 1000$	5	10	100	5 (0%)	7 (+40%)	7(+40%)
Syn-II	$250 \times 750 \times 1500$	10	20	75	9 (-10%)	15 (+50%)	17(+70%)
Syn-III	$500 \times 1000 \times 2000$	15	30	50	16(+6%)	27 (+80%)	23 (+53%)
Syn-IV	$750 \times 1500 \times 1000$	5	10	100	5 (0%)	7 (+40%)	6 (+40%)
Syn-V	$1000 \times 2000 \times 1500$	10	20	75	11(+10%)	18 (+80%)	20 (+100%)
Syn-VI	$2000 \times 2000 \times 2000$	15	30	50	14(+10%)	22 (+46%)	7 (-53%)

TABLE III: Experiment results for various synthetic data for multiple feature variations. We report predicted number of components and its deviation from actual number of components.

C. Baselines

In this experiment, five baselines **AutoChrome** [15], **Iteration based** [14], **NSVD based** [21],**Tucker ARD based** [18] and **BRTF Based** [24] have been used as to evaluate the performance. Note that in literature, AutoChrome and Iteration based method is directly applicable for PARAFAC2 decomposition. To compare with CP/Tucker decomposition based methods, we converted the tensor data from irregular to regular format by appending zeros.

D. Rank Structure of synthetic data

For our synthetic dataset, we observe in Figure (2) that APTERA presents a quite distinct L-curve corner at rank-5 for all given $2-10$ components which is the correct answer. On the other hand, even though AutoChrome and NSVD seems to approximate a region around 7 components, it struggles to give a definitive answer and leaves open the possibility of up to 8 or more components. Both, Tucker ARD and BRTF based methods not able to provide certain solution for synthetic data. Interestingly, even iteration based baseline seems to be working better than AutoChrome and NSVD, showing a subtle indication at 5 components.

Furthermore, to evaluate the robustness of the APTERA, we alter the tensor data features i.e. level of noise, number of components, density and size. The experiment results are provided in the table (III). It is observed that for tensor data mixed with high level of additive noisy tensor rank, Autochrome performance is declined as compared to APTERA.

E. Rank Structure of Real Datasets

While APTERA performs reasonable on synthetic tensor data, indicating a L-curve corner exactly where the predefined number of components is, in order to evaluate its practicality in real-world scenarios, it is also important to research its performance and behavior on real-world data. For this reason, we analyze a range of real data sets and performance is provided in Table (IV). Due to space limitations, we explain results of Centers for Medicare and Medicaid (CMS) data only.

Centers for Medicare and Medicaid (CMS) data files were created to allow researchers to gain familiarity using Medicare

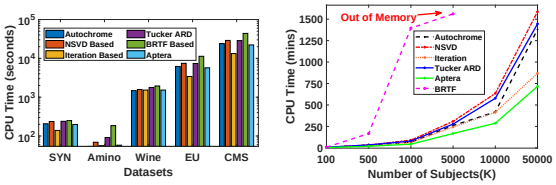

Figure. 3: (a) Computation time of rank estimation for synthetic and real data.(b) Scalability Analysis on synthetic data.

claims data while protecting beneficiary privacy. The CMS data contains multiple files per year. The file contains synthesized data taken from a 5% random sample of Medicare beneficiaries in 2008 and their claims from 2008 to 2010. We decompose PARAFAC2 tensor with rank between $R = 2$ to $R = 50$. Our aim is to estimate appropriate rank to find clinically-meaningful groups of features. For this data, BRTF based and Autochrome baselines unable to proceed due to out of memory after computing parafac2 decompositions. Iteration based baseline and our method APTERA, estimated 11 and 9 components, respectively. NSVD does not provide any estimation of rank for this data. We observe one of the the component (or cluster) in which most of the patients has respiratory disease. These are the patients with high utilization ($> 50\%$), multiple clinical visits (avg 67) and high severity (death rate 8-10%). Most of the patients share ICD-9 code 492 (Emphysema), 496 (Chronic airway obstruction) and 511 (Pleurisy). These codes are characterized by obstruction of airflow that interferes with normal breathing. Phenotype of top 3 components discovered by APTERA based on high factor values is provided in table (V). The codes are decoded in readable format and corresponds to diagnosis or examination. We do not perform any additional post-processing on these results.

F. Run Time Analysis

Figure (3(a)), shows the time taken by each method for synthetic and real dataset. We remark that our proposed method is faster that most baselines except iteration based method where only PARAFAC2 decomposition is considered and no further computations are considered to find rank.

Methods	Wine	Amino	EUCore	CMS	Wine	Amino	EUCore	CMS
$R_o \rightarrow$	4	3	28	–	Percent Deviation (%)			
Autochrome	4	2	26	OoM	**0.00**	-33.33	-7.15	–
NSVD	7	6	13 and 35	50	75.00	100.00	-53.57	–
Iterations	3	**3**	25	11	-25.00	**0.00**	-10.71	–
Tucker ARD	6	**3**	33	2	50.00	**0.00**	17.78	–
BRTF	3	2	49	OoM	-25.00	-33.33	75.00	–
APTERA	**4**	**3**	29	9	**0.00**	**0.00**	**3.57**	–

TABLE IV: Performance of APTERA for rank estimation. Numbers where our proposed method outperforms other baselines are bolded. The negative sign indicates solution is under-fitted and positive values (> 0 for deviation) indicates over-fitted solution.

C1: Congenital Anomalies	C2: Neurological Disorders	C3: Leukemia
Perinatal conditions	Epilepsy	Infections
Cardiac anomalies	Paralysis	Anemia
Club foots	Developmental disorders	Immunity disorders
Short gestation	Tingling	Swollen lymph nodes
Low birth weight	Memory loss	Nosebleeds

TABLE V: Phenotype of top 3 components discovered by APTERA.

G. Scalability Analysis

We also evaluate the scalability of our algorithm on synthetic dataset in terms of time needed for increasing load of input users (K). We report run time for single execution for each method.A PARAFAC2 tensors $\underline{\mathbf{X}} \in \mathbb{R}^{100\times100\times[100-50K]}$ are decomposed with fixed target rank $R = 10$. Figure (3(b)) indicated that all methods seem to scale fairly well with the data size except BRTF method. The time needed by APTERA increases very linearly with increase in non-zero elements. Our proposed method APTERA, successfully estimate the rank of the large PARAFAC2 tensors in reasonable time as shown in Figure (3(b)) and is up to average $15-20\%$ faster than baseline methods except iteration based method (APTERA slower 18%) where only decomposition is performed. We remark the favorable scalability properties of APTERA, rendering it practical to use for large tensors.

V. CONCLUSION

In this paper, we work towards an automatic, PARAFAC2 tensor mining algorithm that minimizes human intervention. We encourage reproducibility by making our code publicly available. Our main contributions are:

- **Algorithm:** We proposed a new scalable method called APTERA for discovering low-rank structure in irregular data, which is based on the finding l-curve corner of higher order singular values.
- **Evaluation:** We evaluate our method on synthetic data, showing their robustness compared to the baselines, as well as a wide variety of real datasets.

VI. ACKNOWLEDGMENTS

The research is partially supported by National Science Foundation CDS&E under grant no. OAC-1808591 and CAREER grant no. IIS-2046086. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding parties.

REFERENCES

[1] A. H. Ali and M. Nazir. Finding a pareto optimal solution for a multi-objective problem of managing radio resources: A qos aware algorithm. *Wireless Personal Communications*, 107(4):1661–1685, 2019.

[2] B. Bader and T. Kolda. Matlab tensor toolbox version 2.2. *Albuquerque, NM, USA: Sandia National Laboratories*, 2007.

[3] R. Bro and H. A. Kiers. A new efficient method for determining the number of components in parafac models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(5):274–286, 2003.

[4] J. D. Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35:283–319, 1970.

[5] J. L. Castellanos, S. Gómez, and V. Guerra. The triangle method for finding the corner of the l-curve. *Applied Numerical Mathematics*, 43(4):359–373, 2002.

[6] CMS. Data. https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs, 2008.

[7] A. Cultrera and L. Callegaro. A simple algorithm to find the l-curve corner in the regularization of inverse problems. *arXiv preprint arXiv:1608.04571*, 2016.

[8] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors. *SIAM journal on Matrix Analysis and Applications*, 21(4):1324–1342, 2000.

[9] E. Gujral. *Modeling and Mining Multi-Aspect Graphs With Scalable Streaming Tensor Decomposition*. University of California, Riverside, 2021.

[10] E. Gujral and E. E. Papalexakis. Semi-supervised multi-aspect community detection. In *Proceedings of the 2018 SIAM SDM*, pages 702–710. SIAM, 2018.

[11] P. Hansen. The l-curve and its use in the numerical treatment of inverse problems. computational inverse problems in electrocardiology. *Advances in Computational Bioengineering*, 5:119, 2001.

[12] R. A. Harshman. Parafac2: Mathematical and technical notes. *UCLA working papers in phonetics*, 22:122215, 1972.

[13] L. He, H. Wang, and M. Zhang. Identification of underwater propeller noise by low-rank approximation of cyclic spectrum. In *OCEANS 2018 MTS/IEEE Charleston*, pages 1–6. IEEE, 2018.

[14] J. C. Hoggard and R. E. Synovec. Parallel factor analysis (parafac) of target analytes in gc× gc- tofms data: automated selection of a model with an appropriate number of factors. *Analytical chemistry*, 79, 2007.

[15] L. G. Johnsen, J. M. Amigo, T. Skov, and R. Bro. Automated resolution of overlapping peaks in chromatographic data. *J. Chemometrics*, 28(2):71–82, 2014.

[16] H. A. Kiers. A three–step algorithm for candecomp/parafac analysis of large data sets with multicollinearity. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 12(3):155–171, 1998.

[17] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51:455–500, 2009.

[18] M. Mørup and L. K. Hansen. Automatic relevance determination for multi-way models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 23(7-8):352–363, 2009.

[19] E. E. Papalexakis. Automatic unsupervised tensor mining with quality assessment. In *Proceedings of the 2016 SIAM SDM*, pages 711–719. SIAM, 2016.

[20] I. Perros, E. E. Papalexakis, F. Wang, R. Vuduc, E. Searles, M. Thompson, and J. Sun. Spartan: Scalable parafac2 for large & sparse data. In *Proceedings of the 23rd ACM SIGKDD Int. Conf. on KDD*, pages 375–384. ACM, 2017.

[21] Y. Tsitsikas and E. E. Papalexakis. Nsvd: N ormalized s ingular v alue d eviation reveals number of latent factors in tensor decomposition. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, pages 667–675. SIAM, 2020.

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

[23] H. Yin, A. R. Benson, J. Leskovec, and D. F. Gleich. Local higher-order graph clustering. In *Proceedings of the 23rd ACM SIGKDD*, pages 555–564, 2017.

[24] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S.-I. Amari. Bayesian robust tensor factorization for incomplete multiway data. *IEEE transactions on NN and learning systems*, pages 736–748, 2015.