Imputing Growth Snapshot Similarity in Early Childhood Development: A Tensor Decomposition Approach

Jennifer J. Schnur*, Ryan Karl*, Angélica García-Martínez†, Meng Jiang*, Nitesh V. Chawla*
*University of Notre Dame
Notre Dame, USA
{jschnur, rkarl, mjiang2, nchawla}@nd.edu
‡Un Kilo de Ayuda
México City, México
{agarciam}@unkilodeayuda.org.mx

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313188

***Abstract*—In this paper, we discuss a tensor decomposition method for imputing similarity scores between individual clinical pictures at predefined patient age intervals in order to construct a dynamic similarity network of patients with respect to early childhood anthropomorphic development. The method leverages Canonical Polyadic Decomposition (or PARAFAC) to compute missing Euclidean similarity scores between pairwise growth pictures, made up of height and weight measurements. We construct a tensor made up of serial affinity matrices to model how the similarities between different patients change over different trajectory snapshots. We intend to use this method to aid Un Kilo de Ayuda (UKA), a non-governmental organization located in Mexico that is made up of facilitators seeking to identify children at risk for malnutrition and suboptimal development. This tensor completion strategy will assist UKA with determining pairs of children with similar clinical pictures, so that they can better assist with selecting treatment strategies and ultimately build better programs tailored to specific families' needs.**

***Index Terms*—Tensor Decomposition, Tensor Factorization, Canonical Polyadic Decomposition, Data Imputation, Time-Series Data, Data Sparsity**

I. Introduction

The rate of generated health data has grown exponentially in recent years due to the expansion of electronic medical records, wearable technologies, and health tracking applications. As a result, healthcare providers are investing their resources into building platforms that can leverage this data to improve patient health. The trend in healthcare is shifting from cure to prevention. Hospitals and healthcare systems house useful repositories of big data (like patient records, test reports, medical images, etc.) that can be leveraged to cut the costs of healthcare, to improve reliability and efficiency, and to provide more effective treatments to patients [19]. Applying data science methods to health data has been proven to assist with such advances. Some success stories include using wearable technologies to monitor and prevent health problems, advancing pharmaceutical research to help find cures for diseases, and reducing hospital readmissions to cut healthcare costs, among many others [2, 7, 20]. Leveraging patient-patient similarity is the backbone behind many of these models. However, patients do not always comply with appointment schedules, and occasionally measurements are missed during routine checkups. These events leave gaps in patient records, which hinder machine learning methods that take these values into consideration when making predictions.

In this paper, we leverage similarities between patient trajectory snapshots to develop a dynamic similarity network of patients across a predefined age-span. Specifically, we intend to predict missing similarity scores between pairwise patients at any given age, only using existing similarity scores at other age snapshots via a tensor decomposition approach. This method has been tested on a child development database from Un Kilo de Ayuda (UKA), a Mexican non-governmental organization made up of facilitators who aim to identify children at risk of malnutrition and suboptimal development. By identifying children with similar growth development, UKA can better assist families to promote health and build better development programs tailored to specific children's needs.

One key challenge associated with this method is data sparsity, as many children may not have acquired measurements within the majority of age intervals (for our experiments sparsity is in excess of 85%). This sparsity increases the difficulty of constructing an accurate approximated tensor from the tensor factorization. Another key challenge associated with this method includes computational complexity. Tensors may become unwieldy as children are added to the system. We combat these challenges by creating smaller subsets of the data such that a tensor is constructed using children who meet a measurement threshold requirement (i.e. the number of measurements acquired by a given child must surpass a particular value in order for that child to be included in the tensor). In this way, we can reduce the sparsity of the tensor and the computational complexity of the factorization. Additionally, the data subsets will target the children who receive measurements most frequently and are therefore most active in the UKA program.

The experimental results associated with our naive baseline computational methods show that imputing similarity

scores between pairwise children with missing measurements requires further consideration than simply computing variations of random or average existing similarity scores. The experimental results of the tensor decomposition method, on the other hand, show that there is a clear advantage to leveraging the "knowledge of the pack" when computing similarity scores between pairwise children at varying age snapshots, as the tensor decomposition approach outperforms all naive baselines, including a baseline in which we fit growth curves to individual child trajectories and use values from these curves to predict missing similarity scores. Note that the development of our method and the code can be found at: https://github.com/RyanKarl/HealthBehaviourModeling.

The contributions of this paper include the following:

- We adapt affinity matrices to fit a tensor completion approach for patient-patient similarity imputation.
- We adapt methods typically used on behavioral datasets (such as recommendation systems) on a profile dataset (child health records).
- We show that a tensor decomposition imputation approach takes advantage of knowledge of the population and outperforms simpler methods that only consider knowledge of the individual.

II. Related Work

When it comes to multi-way data, researchers have come to understand that high-order tensors have clear advantages over standard matrices in terms of data representation, as tensors allow for less information loss, better uniqueness properties, and overall interpretability [6]. Tensor decomposition methods are often used within the contexts of image processing tasks, such as facial recognition [26], biomedical signal detection, such as epileptic seizure structures [1], multiple recommendation systems settings [9] [14], and social network analysis, such as modeling semantic relationships of tweets [23].

Within the field of healthcare analytics, tensor decomposition has become a reliable tool to uncover trends in electronic medical records (EMR) to improve the quality of care delivered by medical professionals. High dimensionality is a key issue within the EMR, as patients receive various vital measurements, lab tests, diagnoses, and procedure results. Methods such as TaGiTeD [29] have been created to capture high order interactions in EMR data by combining tensor decomposition with representation learning. Researchers have also used tensor decomposition to cluster patients with similar clinical profiles [22]. This strategy not only assists with diagnosis but also helps with treatment and prescription. Finally, tensor decomposition has been utilized to impute missing medical data in patient questionnaires [5]. Specifically, canonical polyadic decomposition (CPD) was used to predict survey responses that patients neglected to answer by learning the inherent collaborative relationship structure present in the existing data. To the best of our knowledge, our paper is the first to leverage CP decomposition for data imputation in the healthcare domain with a dataset with extreme levels of sparsity ($\geq 85\%$) and very few features (≤ 2).

Tensor decomposition research is a mature field with several notable results. A seminal paper in reconstructing missing values [11] demonstrated via robust experiments that using alternating least squares (ALS) is superior to computing the covariances between pairs of variables and applying these covariances in constructing a system of normal equations during optimization. Early work by [24] introduced dynamic tensor analysis (DTA) to provide a compact summary for high-order and high-dimensional data and reveal hidden correlations. DTA is scalable, space efficient, and fully automatic with no need for user defined parameters. Later, [13] built a method that is robust to the challenges of computing over high-dimensional, sparse data, where most of the entries of the tensor are zero. Their method, Memory-Efficient Tucker (MET), is based on Tucker decomposition, but after analyzing the available memory, is able to adaptively select the right execution strategy during the decomposition to achieve over 1000X space reduction without sacrificing speed. Also, [10] built a method that can be used to simultaneously recover the true low-rank tensor as well as the sparse corruption tensor even in the presence of arbitrarily large amounts of noise; this is similar to existing work [3].

The work of [15] developed an important parallelizable method ParCube for efficiently decomposing a tensor into sparse factors, that scales to very large datasets. ParCube provides theoretical guarantees for the algorithm's correctness, and has been extensively validated through experiments. Notably, the work of [25] designed a method to summarize high-order data cubes (tensors), and incrementally update these patterns over time. This framework for incremental tensor analysis (ITA) is able to efficiently compute a compact summary for high-order and high-dimensional data, and reveal hidden correlations for dynamic, streaming, and window-based tensors. Also of note is [21], which used related techniques to aid with tagging content on websites, and constructing recommender systems that can help to suggest a user the tags he might want to use for tagging a specific item. They present a factorization model PITF (Pairwise Interaction Tensor Factorization) which is a special case of the TD model with linear run-time both for learning and prediction. PITF explicitly models the pairwise interactions between users, items, and tags. See [12],[16] for a broader overview of this subject.

III. Problem Definition

The formal definition of the problem is as follows: Given the UKA database containing n children, $c_1, c_2, ..., c_n$, we define the height measurement (in centimeters) and the weight measurement (in kilograms) of child c_i at age a as $h_i^{(a)}$ and $w_i^{(a)}$, respectively, where $0 \leq a \leq 59$ months and $1 \leq i \leq n$.

We construct a third-order tensor, $\mathcal{T} \in \mathbb{R}^{n \times n \times 60}$, in which an entry in the tensor is given by:

$$\mathcal{T}_{i,j,a} = \begin{cases} s_{i,j}^{(a)} & \text{if } \exists\langle h_i^{(a)}, w_i^{(a)}\rangle \wedge \exists\langle h_j^{(a)}, w_j^{(a)}\rangle \\ NaN & \text{if } \nexists\langle h_i^{(a)}, w_i^{(a)}\rangle \vee \nexists\langle h_j^{(a)}, w_j^{(a)}\rangle, \end{cases} \quad (1)$$

where $s_{i,j}^{(a)}$ is the Euclidean similarity between the clinical pictures of children c_i and c_j at age a, with $0 \leq a \leq 59$ and $1 \leq i, j \leq n$. Euclidean similarity, $s_{i,j}^{(a)}$, is defined as:

$$s_{i,j}^{(a)} = \frac{1}{1 + \sqrt{(h_i^{(a)} - h_j^{(a)})^2 + (w_i^{(a)} - w_j^{(a)})^2}}. \quad (2)$$

Thus, the tensor $\mathcal{T}$ is made up of a sequence of 60 $n \times n$ affinity matrices between pairwise children at each age snapshot. We seek to predict the missing Euclidean similarity measures between pairwise children at each of these age snapshots. Formally, for $\mathcal{T}_{i,j,a} = NaN$, we will predict $\hat{\mathcal{T}}_{i,j,a} = \hat{s}_{i,j}^{(a)} \in (0, 1]$.

IV. Solutions

A. Tensor Factorization Methods

In order to impute missing similarity scores between pairwise children at each growth snapshot, we utilize tensor factorization methods in order to construct an approximation, $\hat{\mathcal{T}}$, of our original tensor $\mathcal{T}$. The method by which we construct this approximation is through canonical polyadic decomposition (CPD), which breaks down a single tensor into the sum of component rank-one tensors [18] [6]. This allows for a more compact representation of a rank R tensor, as follows:

$$\mathcal{T} \approx \hat{\mathcal{T}} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r, \quad (3)$$

$$\mathcal{T} \in \mathbb{R}^{I \times J \times K}, \boldsymbol{a} \in \mathbb{R}^I, \boldsymbol{b} \in \mathbb{R}^J, \boldsymbol{c} \in \mathbb{R}^K,$$

where $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ are rank-one tensors. We seek to minimize the difference between the original tensor and the approximation, $min_{\hat{\mathcal{T}}} ||\mathcal{T} - \hat{\mathcal{T}}||$. In particular, we utilize Non-Negative Canonical Polyadic Decomposition (NCPD) optimized with hierarchical alternating least squares (HALS), described in [4].

NCPD is particularly useful for this application since all values in our tensor $\mathcal{T}$, represent similarities between pairwise children, and thus must be greater than or equal to 0. HALS transforms the tensor decomposition method into a convex problem by reducing error with respect to a single mode of the tensor at a time while holding the others fixed. Furthermore, HALS has been demonstrated to be robust with respect to noise and suitable for large scale problems, such as this tensor decomposition [8]. We do not discuss the full details of the implementation and mathematical correctness of the method here, but more information can be found in the surveys [12, 17, 28].

For completeness, we discuss another tensor factorization method, known as Tucker decomposition, in which the original high-order rank-R tensor $\mathcal{T}$, is decomposed into multiple matrices and a single core tensor [18] [6], as shown:

$$\mathcal{T} \approx \hat{\mathcal{T}} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} \mathcal{G}_{pqr} a_r \circ b_r \circ c_r, \quad (4)$$

$$\mathcal{T} \in \mathbb{R}^{I \times J \times K}, \mathcal{G} \in \mathbb{R}^{P \times Q \times R},$$

$$\boldsymbol{a} \in \mathbb{R}^{I \times P}, \boldsymbol{b} \in \mathbb{R}^{J \times Q}, \boldsymbol{c} \in \mathbb{R}^{K \times R}.$$

where $\mathcal{G}$ is the core tensor and $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ are matrices. While Tucker approximation is especially useful for dimensionality reduction and estimation of signal subspaces, we favor CP decomposition due to known disadvantages of Tucker decomposition [27]. Tucker is generally considered to be slower than CP. Also, the decomposition from Tucker is invariant to rotations in the factor matrices, and when using Tucker, the core tensor becomes difficult to visualize. Since we intend to deploy the tensor decomposition method within the UKA organization, it is crucial to maintain visualization of latent factors and interpretability of the model. Also, for our application, there is no need to generate the core tensor as is done during a Tucker decomposition. As a result, we favor CP decomposition for generating the tensor approximation, as this method aims at identifying latent factors that are directly meaningful in constructing the dynamic similarity network of early child development pictures.

V. Experiments

In this section we outline the dataset specifications, validation settings, evaluation methods, and experimental results associated with the initial baseline methods and the CP decomposition approach.

A. DATA SOURCES

The database used in these experiments comes from Un Kilo de Ayuda (UKA), a non-governmental organization based in Mexico, seeking to promote early childhood development by tracking child growth and assisting families without convenient access to healthcare services gain perspective on their children's health. The data consists of $486,926$ unique children with over 3 million total measurement records spanning the years $2000 - 2018$. Measurements primarily include height (centimeters) and weight (kilograms) associated with a given measurement date. Since not all children necessarily receive both a height and a weight measurement on the same date due to UKA workflow, we generated a subset of the data in which a height and a weight measurement was recorded on the same date for a given child. The resulting subset consists of $444,142$ unique children, $227,465$ of which are male, while $216,677$ are female. The average number of combined height and weight measurements per child is 2.95, meaning that on average, a given child acquires approximately 3 height and weight measurements on the same date within the first 5 years life through the UKA system. Note that we normalize the height and weight measurements to values in the range of [0,1] using min-max normalization.

In order to combat the sparsity associated with generating tensors, given the low measurement frequency as described above, we constructed tensors by taking subsets of the UKA data such that a child is included in a given tensor only if the number of measurements that the child has acquired has surpassed a given threshold. Specifically, we constructed 4 tensors that include children who have acquired at least 14, 16, 18, and 20 measurements, respectively. Table I outlines

TABLE I: Tensor Sparsity Statistics

Tensor (Min # Measurements per Child)	Number of Children Included	Tensor Size (Total Entries)	Non-Empty Entries	Tensor Sparsity
14	634	24,117,360	1,948,323	91.92%
16	302	5,472,240	546,143	90.02%
18	108	699,840	88,436	87.36%
20	46	126,960	18,804	85.19%

the subsets derived from the UKA database according to a minimum number of measurements required per unique child along with the corresponding tensor size and sparsity, given that a tensor consists of 60 $n \times n$ similarity matrices, one for each age snapshot (0 to 59 months).

B. Naive Baselines

For our naive baselines, we implemented four methods to analyze the difficulty of predicting children's pairwise similarity scores without explicitly knowing weight and height measurements at any given age:

- Baseline 0 computes the overall average similarity score for children of the same age (rounded to the nearest month) and uses this value as the prediction for missing similarity scores between pairwise children.
- Baseline 1 uses a similarity score of 0.5 as a prediction for missing similarity scores between pairwise children.
- Baseline 2 uses a randomly selected number between 0 and 1 as a prediction for a missing similarity score between pairwise children.
- Baseline 3 fits a logarithmic curve to each unique child's height and weight trajectories over 0-59 months of life. Similarity scores between pairwise children are computed at a given age by extrapolating height and weight values from these curves at this age value.
- Baseline 4 computes the overall average for children of the same age (rounded to the nearest month) within the same residential location.

C. Validation Settings

We utilized leave-p-out validation, similar to holdout validation, where we held out a percentage of data for testing our method following training. For each constructed tensor, we conducted CP decomposition experiments using four validation settings with respect to known tensor entries, specifically 1) 90% training and 10% testing 2) 80% training and 20% testing 3) 70% training and 30% testing and 4) 60% training and 40% testing. We also experimented with four different training error tolerances in order to fit the approximated tensor generated from the CP decomposition to the original tensor with respect to known entries and analyze when potential overfitting may occur. Specifically, we use tolerances 1×10^{-3}, 1×10^{-5}, 1×10^{-7}, and 1×10^{-9}.

For each train-test validation setting and training error tolerance outlined above, we conducted 10 runs of the CP decomposition algorithm in which tensor entries used for training and testing were randomly chosen from existing tensor entries. Note that error and standard deviation was averaged for all trials. The results of all experiments are provided in Figure 1 and Table II.

D. Evaluation Methods

In order to measure the accuracy of our methods, we utilize the root mean squared error (RMSE) and mean absolute error (MAE) to determine the overall difference between the observed similarities and the predicted similarities of the children's clinical pictures. These measures are defined as

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y_i} - y_i)^2}{n}} \tag{5}$$

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |\hat{y_i} - y_i| \tag{6}$$

where y_i represents an observed similarity and $\hat{y_i}$ represents a predicted similarity between a pair of children, while n represents the total number of observed similarities. The MAE and RMSE will determine how close our approximated tensor computation gets to the original tensor, with respect to the known entries.

E. Experimental Results

In this section, we provide a comparison between the results of our baseline computational methods and the CP decomposition approach with respect to predicted similarity scores for pairwise children who possess ground truth measurements. In Table II, we computed the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) between the predicted values outputted from the methods and the actual ground truth for these values. The results show that the CP decomposition approach outperforms all experimental baselines in predicting similarity scores between pairwise children with varying train-test holdout settings for every tensor size. In particular, the CP decomposition method outperforms Baseline 3, in which height and weight values were extracted from individual curves fit to unique children's growth trajectories and then used to compute similarity scores. This result implies that there is an advantage to leveraging the "knowledge of the pack" that CPD provides over "knowledge of the individual" that the curve-fitting provides. This is further supported by the fact that baselines 1 and 4, that do try to leverage "knowledge of the pack" by computing average similarities of groups of children, outperform baseline 3.

Another interesting finding from these experiments involves the relationship between tensor size and the RMSE and MAE results. From Table II and Figure 1, it is evident that as tensor size (and therefore sparsity) increases, RMSE and

TABLE II: Error Score Comparison

Method	Minimum Measure-ments/Child	60% Train		70% Train		80% Train		90% Train	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
CPD	**14**	**0.01239±0.00038**	**0.00181±6.79E-05**	**0.01072±0.00043**	**0.00136±6.56E-05**	**0.00863±0.00037**	**0.00089±4.69E-05**	**0.00591±0.00022**	**0.00043±1.98E-05**
Baseline 0	14	0.01650±0.00023	0.00289±0.00073	0.01814±0.00013	0.00275±0.00043	0.02187±0.00033	0.00269±0.00011	0.03189±0.00026	0.00281±0.00046
Baseline 1	14	2.88291±0.00005	0.49635±0.00057	3.27354±0.00003	0.50642±0.00025	4.01942±0.00021	0.50226±0.00014	5.72176±0.00030	0.50407±0.00003
Baseline 2	14	2.93610±0.00002	0.49798±0.00057	3.30157±0.00020	0.50514±0.00031	4.09572±0.00001	0.50591±0.00072	5.76821±0.00021	0.49378±0.00033
Baseline 3	14	0.81402±0.04264	5.41842±0.08281	0.68028±0.02510	3.94524±0.05652	0.52938±0.02805	2.49025±0.04945	0.35142±0.01271	1.09818±0.03532
Baseline 4	14	0.15105±0.02470	0.05806±0.00317	0.15273±0.02057	0.05578±0.00396	0.15051±0.08718	0.05674±0.00372	0.14942±0.03260	0.05621±0.00313
CPD	**16**	**0.01311±0.00045**	**0.00211±9.16E-05**	**0.01122±0.00046**	**0.00156±7.83E-05**	**0.00911±0.00021**	**0.00103±2.98E-05**	**0.00634±0.00012**	**0.00051±1.11E-05**
Baseline 0	16	0.01642±0.00793	0.00280±0.00060	0.02039±0.00600	0.00302±0.00411	0.02404±0.00623	0.00289±0.00241	0.03349±0.00597	0.00287±0.00055
Baseline 1	16	2.88291±0.00098	0.49870±0.00003	3.26701±0.00406	0.50815±0.00057	3.98634±0.00190	0.50131±0.00068	5.66567±0.00725	0.50518±0.00017
Baseline 2	16	2.92469±0.00247	0.49912±0.00083	3.29663±0.00170	0.50093±0.00020	4.11176±0.00165	0.50224±0.00054	5.78615±0.00702	0.49279±0.00080
Baseline 3	16	0.55654±0.02094	2.67566±0.04130	0.47192±0.01514	1.98017±0.03160	0.37838±0.02097	1.21901±0.02810	0.27040±0.00768	0.67500±0.02205
Baseline 4	16	0.18564±0.04403	0.05936±0.05220	0.18416±0.04928	0.05944±0.00696	0.17978±0.05857	0.05874±0.00856	0.17820±0.06163	0.05916±0.00789
CPD	**18**	**0.01264±0.00021**	**0.00221±7.77E-05**	**0.01082±0.00027**	**0.00164±4.44E-05**	**0.00896±0.00028**	**0.00111±4.03E-05**	**0.00626±0.00017**	**0.00055±1.96E-05**
Baseline 0	18	0.01599±0.00301	0.00280±0.00066	0.01777±0.00022	0.00269±0.00013	0.02234±0.00058	0.00275±0.00055	0.03371±0.00521	0.00296±0.00049
Baseline 1	18	2.86401±0.00805	0.49790±0.00082	3.27980±0.00079	0.50452±0.00011	4.02051±0.00811	0.50277±0.00019	5.66567±0.00537	0.50453±0.00029
Baseline 2	18	2.90985±0.00127	0.50254±0.00075	3.30987±0.00393	0.50133±0.00023	4.14868±0.00663	0.49791±0.00032	5.78557±0.00146	0.49537±0.00008
Baseline 3	18	0.32475±0.02623	0.93287±0.02810	0.27213±0.03422	0.65890±0.03680	0.20888±0.00824	0.44290±0.02416	0.14111±0.01589	0.18504±0.01938
Baseline 4	18	0.21945±0.04686	0.06379±0.00812	0.21303±0.04958	0.06329±0.01055	0.21720±0.04974	0.06329±0.01055	0.21407±0.05936	0.06285±0.01174
CPD	**20**	**0.01387±0.00036**	**0.00258±5.77E-05**	**0.01195±0.00022**	**0.00194±5.08E-05**	**0.00961±0.00027**	**0.00127±4.26E-05**	**0.00685±0.00019**	**0.00064±1.63E-05**
Baseline 0	20	0.01647±0.00170	0.00288±0.00049	0.01925±0.00165	0.00292±0.00078	0.02305±0.00702	0.00284±0.00042	0.03294±0.00040	0.00289±0.00061
Baseline 1	20	2.85601±0.00526	0.49985±0.00082	3.29462±0.00285	0.49987±0.00033	4.05989±0.00051	0.49982±0.00011	5.68329±0.00730	0.49983±0.00031
Baseline 2	20	2.88991±0.00072	0.49880±0.00038	3.33286±0.00401	0.49869±0.00030	4.10882±0.00423	0.49886±0.00037	5.74821±0.00422	0.49856±0.00054
Baseline 3	20	0.20881±0.03411	0.37661±0.03163	0.17756±0.03999	0.27186±0.03181	0.12746±0.00934	0.15968±0.01865	0.08006±0.00837	0.06574±0.01066
Baseline 4	20	0.33257±0.09769	0.09019±0.02859	0.33297±0.09777	0.09026±0.02713	0.32784±0.11877	0.09072±0.02938	0.32885±0.11397	0.09026±0.02878

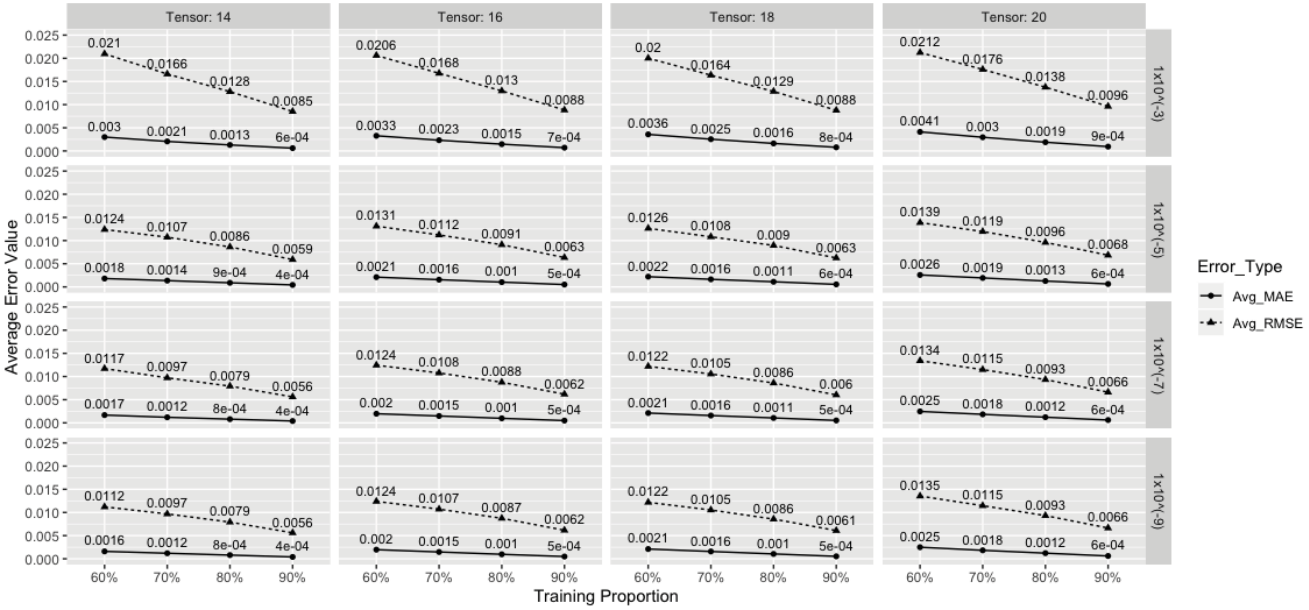

Fig. 1: CP Decomposition Average RMSE and MAE Results

MAE decrease. We attribute this finding to be the result of CP decomposition relying on a richer set of patterns when determining latent relationships between child growth and age. More explicitly, when more children are introduced to the tensor, the decomposition is more successful at capturing the latent characteristics available within the similarity network, and therefore the reconstruction more accurately represents true similarities between pairwise children at varying age snapshots. One might expect that tensor sparsity would negatively impact a tensor decomposition approach; however, the opposite result is found.

Finally, and expectedly, the experimental findings show that as we increase the proportion of training data and decrease the acceptable training error tolerance, we obtain a more successful tensor reconstruction, as evidenced by lower RMSE and MAE. However, as we decrease the maximum error tolerance allowed for tensor reconstruction during the fitting process, we find that RMSE and MAE decreases are minimal once the error value of 1×10^{-9} is reached. This implies that overfitting begins to occur at this threshold and suggests we may be able to utilize regularization to further improve the method. We leave this for future work.

VI. Conclusion

In this paper, we presented a CP decomposition approach for imputing missing similarity scores between pairwise children at varying age snapshots with respect to the children's growth trajectories. The CP decomposition approach outperforms all

tested baselines, including a baseline in which individual curves are fit to unique children's height and weight trajectories over the age-span from birth to 59 months. Furthermore, the CP decomposition approach proved to be most successful when more children (and therefore more growth patterns) were introduced to the tensor. These results show that leveraging "knowledge of the pack" proves to be a better strategy than leveraging "knowledge of the individual" when imputing these similarity scores. Given the completed, approximated tensor generated from the CP decomposition method, we have constructed a dynamic similarity network that relates unique children based on their early development patterns. The similarity network will assist UKA workers by identifying children at risk for a decline in health, and providing families with resources that have helped children with similar developmental tendencies. More specifically, this technique will help UKA fill in the gaps for missing data in child health profiles to better understand if their health trajectory indicates they are at risk and need intervention from a social worker or a medical specialist. This way, poor health conditions can be better identified and addressed before they lead to chronic ailments that persist into adulthood.

REFERENCES

[1] E. Acar, C. Bingol, H. Bingol, R. Bro, and B. Yener. Multiway analysis of epilepsy tensors, 08 2007.

[2] A. Chen. Apple's researchkit generates reliable health data — at least for asthma patients. 2017.

[3] Y. Chen et al. Low-rank matrix recovery from errors and erasures. *IEEE Transactions on Information Theory*, 59 (7):4324–4337, 2013.

[4] A. Cichocki et al. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 92(3):708–721, 2009.

[5] J. Dauwels, L. Garg, A. Earnest, and L. K. Pang. Tensor factorization for missing data imputation in medical questionnaires. In *2012 IEEE ICASSP)*, pages 2109–2112. IEEE, 2012.

[6] L. De Lathauwer. A survey of tensor methods. In *2009 IEEE International Symposium on Circuits and Systems*, pages 2773–2776, May 2009. doi: 10.1109/ISCAS.2009.5118377.

[7] E. C. Declan Butler and E. C. Hayden. How ebola-vaccine success could reshape clinical-trial policy. 2015.

[8] B. Erichson and A. Williams. Tensortools: Fitting and visualizing tensor decompositions in python. URL `https://github.com/ahwillia/tensortools`.

[9] E. Frolov and I. Oseledets. Tensor methods and recommender systems, Feb 2017. ISSN 1942-4787.

[10] Q. Gu, H. Gui, and J. Han. Robust tensor decomposition with gross corruption. In *Advances in Neural Information Processing Systems*, pages 1422–1430, 2014.

[11] Y. Haitovsky. Missing data in regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(1):67–82, 1968.

[12] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[13] T. G. Kolda and J. Sun. Scalable tensor decompositions for multi-aspect data mining. In *2008 Eighth IEEE international conference on data mining*, pages 363–372. IEEE, 2008.

[14] S. Kutty, L. Chen, and R. Nayak. A people-to-people recommendation system using tensor space models. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 187–192, 2012.

[15] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos. Parcube: Sparse parallelizable candecomp-parafac tensor decomposition. *ACM Trans. Knowl. Discov. Data*, 10 (1), July 2015. ISSN 1556-4681. doi: 10.1145/2729980. URL `https://doi.org/10.1145/2729980`.

[16] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos. Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM Trans. Intell. Syst. Technol.*, 8(2), Oct. 2016. ISSN 2157-6904. doi: 10.1145/2915921. URL `https://doi.org/10.1145/2915921`.

[17] K. Petersen, M. Pedersen, et al. The matrix cookbook, vol. 7. *Technical University of Denmark*, 15, 2008.

[18] S. Rabanser, O. Shchur, and S. Günnemann. Introduction to tensor decompositions and their applications in machine learning. *ArXiv*, abs/1711.10781, 2017.

[19] D. Ramesh, P. Suraj, and L. Saini. Big data analytics in healthcare: A survey approach. 01 2016. doi: 10.1109/MicroCom.2016.7522520.

[20] L. Rao. How this startup is trying to upend health insurance. 2015.

[21] S. Rendle and L. Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation, 2010.

[22] M. Ruffini, R. Gavalda, and E. Limón. Clustering patients with tensor decomposition. *arXiv preprint arXiv:1708.08994*, 2017.

[23] H. Shan, A. Banerjee, and R. Natarajan. Probabilistic tensor factorization for tensor completion. 2011.

[24] J. Sun, D. Tao, and C. Faloutsos. Beyond streams and graphs: Dynamic tensor analysis, 2006. URL `https://doi.org/10.1145/1150402.1150445`.

[25] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, and C. Faloutsos. Incremental tensor analysis: Theory and applications. *ACM TKDD*, 2(3):1–37, 2008.

[26] M. A. O. Vasilescu and D. Terzopoulos. Multilinear subspace analysis of image ensembles. In *IEEE CVPR*, volume 2, pages II–93, June 2003.

[27] A. Williams. Notes on tensor decompositions. 2016.

[28] A. Williams. Solving least-squares regression with missing data. *Its Neuronal Blog*, 2018.

[29] K. Yang et al. Tagited: Predictive task guided tensor decomposition for representation learning from electronic health records, 2017.