Unraveling Complex Temporal Patterns in EHRs via Robust Irregular Tensor Factorization

Yifei Ren, PhD[1], Linghui Zeng, MS[1], Jian Lou, PhD[1], Li Xiong, PhD[1], Joyce C Ho, PhD[1], Xiaoqian Jiang, PhD[2], Sivasubramanium V Bhavani, MD[1]
[1] Emory University, Atlanta, GA; [2] University of Texas Health Science Center

Abstract

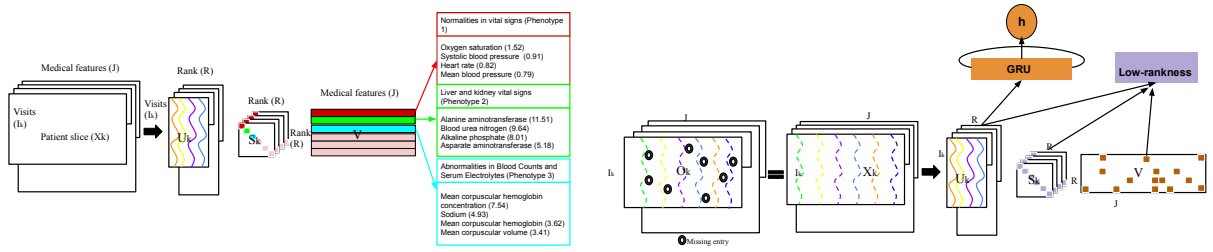
*Electronic health records (EHRs) contain diverse patient data with varying visit frequencies. While irregular tensor factorization techniques such as PARAFAC2 have been used for extracting meaningful medical concepts from EHRs, existing methods fail to capture non-linear and complex temporal patterns and struggle with missing entries. In this paper, we propose **REPAR**, an **R**NN **RE**gularized Robust **PAR**AFAC2 method to model complex temporal dependencies and enhance robustness in the presence of missing data. Our approach employs Recurrent Neural Networks (RNNs) for temporal regularization and a low-rank constraint for robustness, enabling precise patient subgroup identification and improved clinical decision-making in noisy EHR data. We design a hybrid optimization framework that handles multiple regularizations and various data types. REPAR is evaluated on 3 real-world EHR datasets, demonstrating improved reconstruction and robustness under missing data. Two case studies further showcase REPAR's ability to extract meaningful dynamic phenotypes and enhance phenotype predictability from noisy temporal EHRs.*

Introduction

Recent years have witnessed a global interest in mining electronic health records (EHRs) to improve healthcare and advance medical research [1, 2]. EHRs consist of detailed information such as diagnoses, laboratory test results, and medication prescriptions for large patient populations. However, directly using raw EHR data is challenging due to the complex structure associated with its longitudinal and multi-dimensional nature as well as the enormity of the data. Clinical scientists are interested in breaking apart heterogeneous syndromes into subgroups, i.e. *phenotypes*, such as diseases and disease subtypes, for better understandings of the differences in biological mechanisms and treatment responses, which could lead to more effective and precise treatment [3]. In medical contexts, the word "phenotype" refers to clinically relevant variations in morphology, physiology, or behavior. The analysis of phenotype plays a key role in clinical practice and medical research [4]. Therefore, raw EHR data are often mapped to phenotypes [5], which can be used for cohort (patient subgroup) selection and healthcare quality measurement, grouping patients with similar symptoms or treatment responses, enabling personalized care.

Sepsis Phenotype Example. Sepsis is a heterogeneous syndrome characterized by a dysregulated immunological response to infection that results in organ dysfunction and often death. Phenotyping in sepsis can help identify patient subgroups who have different clinical characteristics, prognoses, and treatment responses, thus enabling a precision approach to treatment. However, traditional clustering-based phenotyping methods [6, 7] cannot fully capture the complex temporal and inter-attribute dependencies of the data and are not robust to missing data.

EHR data often contains multiple types of information that are interconnected – imagine a multidimensional array (or tensor) where each dimension represents different aspects of patient care, such as symptoms, treatments, and time [8]. Tensor factorization breaks down a tensor into simpler components (rank-one tensors), enabling the extraction of meaningful patterns and relationships across dimensions for improved analysis and interpretation. Unlike traditional clustering-based approaches, tensor factorization-based computational phenotyping models not only cluster patients into subgroups, but also capture interactions between multiple attributes (e.g., specific procedures used to treat a disease) to extract concise and potentially more interpretable multiattribute patterns in latent spaces [9, 10, 11]. Several established methods exist for analyzing regular tensors where all dimensions are uniformly structured (e.g., data measurements at consistent time intervals), including Canonical Polyadic (CP) [12], Tucker [13], and Singular Value Decomposition (SVD) [14]. However, EHR data often present a challenge because patient visits and events occur at irregular time intervals. EHR records are also prone to missing observations due to variations in clinical practice that affect the frequency of measurements (e.g., a patient will have labs collected more often in intensive care units (ICUs) than in regular medical wards). Most existing methods may not capture the true patterns in the data effectively.



(a) PARAFAC2 overview

(b) REPAR overview

Figure 1: (a) PARAFAC2 breaks irregular patient data into three components: temporal evolution (U_k), health patterns (phenotypes) (V), and membership levels (S_k). This reveals important trends and interactions in the data. (b) REPAR enhances PARAFAC2 by using RNNs to model time-based health changes and a low-rank constraint to handle missing and noisy data more effectively.

Recently, PARAFAC2, an irregular tensor factorization method [15], has been popularized for extracting meaningful phenotypes from such temporal EHR (illustrated in Figure 1a). A scalable PARAFAC2 model was proposed in [16] to handle large and sparse data. Various constraints were introduced to improve the interpretability of the factor matrices for more meaningful phenotype extraction [17]. In addition, robust PARAFAC2 models have been studied in [1, 18] to handle missing and erroneous entries. Despite the advances in PARAFAC2, two fundamental challenges remain. First, existing PARAFAC2 models only impose linear and human-defined temporal regularization functions, which fail to fully capture the non-linear and complex temporal information in EHRs. Second, current models are designed only for a single data type – numeric or binary, and lack flexibility for generalized data types.

To address these limitations, we propose an **R**NN R**E**gularized Robust **PARA**FAC2 for Irregular Temporal Tensor Factorization, **REPAR**, to capture complex temporal dependencies and improve robustness, phenotype representations and predictability of tensor factorization. As shown in Figure 1b, REPAR starts with patient data that may have missing or inconsistent entries, O_k, and assumes that there is a complete underlying version of the data, X_k. To reconstruct this data, we use RNNs to model how patient health evolves over time and capture the complex temporal dependencies, along with low-rank constraints that simplify the data into its most important components, and then extract factor matrices $U_k, S_k, V^\top$ for further downstream analysis and phenotype interpretation. Our hypothesis is that by capturing the more complex temporal information, it will be more robust to missing entries in the original tensor (together with the low-rank constraint), and also achieve better predictability for the extracted phenotypes. This helps in better understanding patient subgroups and making predictions. In addition, we introduce a new optimization framework that can flexibly handle any smooth loss function, as opposed to the sole choice of the least square norm, to better suit input tensors with various data types.

In summary, we list our main contributions below:

1. We propose REPAR, a robust RNN and low-rank regularized PARAFAC2 tensor factorization method, to capture the complex temporal dependencies and enhance the robustness of the phenotype representations.
2. We propose a new hybrid optimization framework using stochastic gradient descent (SGD) and proximal average to handle multiple regularizations and a generalized loss function to support various data types.
3. We evaluate REPAR on three real-world temporal EHR datasets, which verify the improved recovery, phenotype representation, and predictability in the presence of missing values. Through downstream prediction analysis and interpretation of the dynamic subphenotype trajectory, we demonstrate that REPAR can robustly extract meaningful and high-predictability phenotypes that capture and distinguish different temporal patterns.

Related Work

SPARTan [16] scaled PARAFAC2 to large and sparse irregular tensors by introducing a sparsity aware computation module to reduce the per-iteration cost. Following SPARTan, COPA [17] introduced various constraints/regularizations

to improve the interpretability of the factor matrices. For example, COPA proposed the M-spline constraint [19] to the temporal evolution matrix, $\mathbf{U}_k$, to capture the temporal smoothness, non-negative constraint to $\mathbf{S}_k$ to ensure phenotype memberships are non-negative, and ℓ_1-norm regularization of $\mathbf{V}$ to induce sparse phenotype definitions. Despite their improvements in computational efficiency and output interpretability, COPA and SPARTan did not explicitly address the problem of missing entries in the input tensor, which severely limits them from more robust clinical usage.

REPAIR [1] and LogPar [18] address missing entries in PARAFAC2. Inspired by the robust low-rank tensor minimization (RLTM), the state-of-the-art mechanism for dealing with missing and error entries, REPAIR [1] separated the corrupted input tensor into a clean, completed tensor and an error tensor. Since the clean tensor is often low-rank, REPAIR added low-rank regularization (i.e., nuclear norm) on the clean tensor and sparsity regularization (ℓ_1-norm regularization) on the error tensor. It then proposed a novel two-phase optimization alternative direction method of multipliers (ADMM) approach to solve the low-rank regularized PARAFAC2 model. LogPar considered binary data with a one-class missing value scenario. LogPar modeled the binary irregular tensor with the Bernoulli distribution parameterized by an underlying real-valued tensor. Then they approximated the underlying tensor with a positive-unlabeled learning loss function to account for the missing values. However, both models are suitable for one type of data and cannot be easily adapted for composite regularization of the factor matrices.

tPARAFAC2 [20] tracks evolving patterns in incomplete temporal data. It extends PARAFAC2 with temporal smoothness regularization to capture time-evolving factors and incorporates statistical updates to handle missing values. However, tPARAFAC2's regularization assumes gradual and linear changes across time points, making it less suitable for capturing more complex, non-linear temporal patterns that can arise in certain real-world EHR datasets.

CNTF [21] treated each patient's data as an individual tensor, used CP to find the factor matrices and RNN to regularize the latent factor evolution. The RNN model was used to model the non-linear temporal dependency in patient progressions and can also integrate higher-order information. However, CNTF assumes interactions among modalities, which may not always be the case, as demonstrated by the empirical results in [17] and [22].

Proposed Method

In this section, we present the REPAR model and its optimization, which has the following appealing features that distinguish it from previous PARAFAC2 methods. (1) It accommodates a wide selection of regularizations, including statistical learning-based (e.g., l_1 norm and nuclear norm), deep learning-based (e.g., RNN regularization), and composite, to better capture the intrinsic nature of the irregular temporal EHR data. (2) It generalizes the loss functions from the sole choice of the least square norm to any smooth loss function, which better suits input tensors with various data types. We summarize popular loss functions for common data types in Table 1. (3) It introduces a new optimization framework geared to fully exploit the parallel computing capability of modern GPUs to boost efficiency. Furthermore, it incorporates Nesterov's momentum into the updates of REPAR to achieve faster convergence.

Problem Formulation

The goal of REPAR is to analyze patient data in a way that uncovers hidden patterns while handling missing or inconsistent information. We formalize the objective function for REPAR model in Definition 1. The *PARAFAC2 loss for* $\mathbf{O}$ ensures the reconstructed tensor closely approximates the original tensor. The *low-rank for* $\mathbf{O}$ enforces the underlying complete tensor to be separated from missing values and helps the model focus on the essential patterns in the data while filling in the gaps caused by missing entries. The *RNN loss* captures the temporal patterns in the data, and an *approximate uniqueness constraint* ensures tensor factorization uniqueness. For EHR phenotype discovery, various constraints can be imposed on the factorization matrices to yield meaningful and high-interpretability phenotypes. REPAR accommodates such interpretability-purposed constraints in Equation (1) including non-negativity on $\mathbf{S}$ via $c_1(\mathbf{S}_k)$ and sparsity on $\mathbf{V}$ via $c_2\|\mathbf{V}\|_1$. The process ensures that the results are clear, interpretable, and useful for identifying patient subgroups (phenotypes).

Table 1: Examples of Tensor Data Types and Loss Functions

Data Type	Loss Function
Binary	Positive Unlabeled loss [18]
Count	Poisson Loss [23]
Numerical	Least Square Loss
Strictly positive data	Rayleigh Loss [23]

Definition 1. *(REPAR objective function)*

$$\underset{\mathbf{Q}_k,\mathbf{H},\mathbf{S}_k,\mathbf{V}}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{(i,j)\in\Omega} \overbrace{L(\mathbf{O}_{ijk}, \{\mathbf{U}_k\mathbf{S}_k\mathbf{V}^\top\}_{ijk})}^{\textit{PARAFAC2 loss for } \mathbf{O}} + \overbrace{\sum_{k=1}^{K} \texttt{RNN}(\mathbf{U}_k)}^{\textit{RNN loss}} + \overbrace{\rho_1\|\mathbf{H}\|_* + \rho_2\|\mathbf{V}\|_* + \rho_3\|\mathbf{W}\|_*}^{\textit{low-rank for } \mathbf{O}}$$
$$+ \overbrace{\varrho_1 \sum_{k=1}^{K} \left(\|\mathbf{U}_k - \mathbf{Q}_k\mathbf{H}\|_F^2 + \varrho_2\|\mathbf{Q}_k^\top\mathbf{Q}_k - \mathbf{I}\|_F^2\right)}^{\textit{approximate uniqueness constraint}} + \overbrace{\sum_{k=1}^{K} c_1(\mathbf{S}_k) + c_2\|\mathbf{V}\|_1}^{\textit{interpretability constraint}} + \overbrace{\varrho_3\|\mathbf{W} - \overline{\mathbf{W}}\|_F^2}^{\textit{auxiliary } \overline{W} \textit{ for disentangling constraints}}, \tag{1}$$

$$\text{s.t. for } k = 1, ..., K, \overbrace{\mathbf{S}_k = \texttt{diag}(\overline{W}(k,:)), \mathbf{S}_k \textit{ is diagonal}}^{\textit{relation between } \mathbf{S},\mathbf{W}} \tag{2}$$

where $\mathbf{H}, \{\mathbf{S}_k\}, \mathbf{I} \in \mathbb{R}^{R\times R}$, $\mathbf{Q}_k \in \mathbb{R}^{I_k\times R}$, Ω *denotes the index of the non-missing entries,* c_1 *is the nonnegativity constraint, and* $c_2\|\mathbf{V}\|_1$ *is the sparsity penalty.*

Generalized PARAFAC2. The classic *PARAFAC2 loss for* $\mathbf{O}$ is the least squared loss as shown in Definition 2. EHRs encode various types of information such as simple yes/no answers (e.g., patient has diabetes), numeric measurements (e.g., blood pressure readings), and others might be counts (e.g., number of hospital visits). To accommodate these different data types, we extend PARAFAC2 to a more generalized form by introducing a general loss function $\sum_{k=1}^{K}\sum_{(i,j)\in\Omega} L(\mathbf{O}_{ijk}, \{\mathbf{U}_k\mathbf{S}_k\mathbf{V}^\top\}_{ijk})$ with any smooth loss function L, rather than limiting it to be the least squared loss. The capability to switch between various loss functions as shown in Table 1 allows REPAR to tailor to different input data types.

Definition 2. *(Classic PARAFAC2 model [15])*

$$\underset{\{\mathbf{U}_k\},\{\mathbf{S}_k\},\mathbf{V}}{\operatorname{argmin}} \sum_{k=1}^{K} \frac{1}{2}\|\mathbf{O}_k - \mathbf{U}_k\mathbf{S}_k\mathbf{V}^\top\|_F^2,$$

s.t. $\mathbf{U}_k = \mathbf{Q}_k\mathbf{H}, \mathbf{Q}_k^\top\mathbf{Q}_k = \mathbf{I}, \mathbf{S}_k$ *is diagonal, where* $\mathbf{Q}_k \in \mathbf{R}^{I_k\times R}$ *is orthogonal,* $\mathbf{I}_k \in \mathbf{R}^{R\times R}$ *is the identity matrix and* R *is the target rank of the PARAFAC2 factorization.*

RNN regularization. In order to model the temporal dependency in phenotype progression, we regard each patient's temporal evolution matrix $\mathbf{U}_k \in \mathbb{R}^{I_k\times R}$ as a multivariate time series with each variable describing the progression of the corresponding phenotype for patient k. For each timestamp t, we use the RNN model to predict $\mathbf{U}_k^t$ given the values from previous timestamps and minimize the Mean Square Error (MSE) between the real and predicted value.

A key feature of our model is that the RNN regularization is jointly optimized with the PARAFAC2 tensor factorization to enforce the patient temporal evolution matrix is consistent with the temporal dependency captured by RNN as well as to recover the temporal tensor. RNN regularization inherently captures long-term dependencies even though it only regularizes the temporal evaluation matrix.

Low-rank for O. RLTM [24] can recover a tensor with missing values via a low-rank regularization function. The idea is to separate the underlying completed tensor from the corrupted tensor where the completed part is often low-rank. As we have previously shown, adding low-rank on $\mathbf{O}$ via nuclear norm constraints on the factor matrices $\mathbf{H}, \mathbf{V}, \mathbf{W}$ can improve robustness to missing entries [1].

Algorithm 1 Optimization Framework for REPAR

Input: Input tensor $\mathbf{O}$; Model parameters ρ_0-ρ_3, ϱ_1-ϱ_3; Optimization parameters η's; Interpretability constraint types c_1, c_2, c_3 and `RNN` sub-model; Initial rank estimation R.

1: **while** Not reach convergence criteria **do**
2: Update $\{\mathbf{U}_k\}$, $\{\mathbf{Q}_k\}$ by SGD with momentum;
3: Update `RNN` by SGD with momentum;
4: Update $\mathbf{W}, \mathbf{S}_k, \mathbf{H}$ by Proximal SGD with momentum;
5: Update $\mathbf{V}$ by Proximal averaging SGD with momentum;
6: **end while**

Output: Phenotype factor matrices $\{\mathbf{U}_k\} = \{\mathbf{Q}_k\mathbf{H}\}, \{\mathbf{S}_k\}, \mathbf{V}$.

Optimization

To solve the optimization problem in (1), REPAR follows an alternative optimization strategy (shown in Algorithm 1) where we optimize one variable individually using SGD with all other variables fixed. The variables to be optimized can be categorized into three groups according to whether the subproblems are purely smooth (i.e., differentiable), proximal mapping-based smooth, or multiple non-smooth (i.e., non-differentiable). In particular, when dealing with multiple non-smooth regularized subproblems, we introduce the proximal average-based technique as a replacement for the existing optimization approach adopted in the previous PARAFAC2 works [1, 17]. As a result, REPAR can take advantage of the parallel computing feature of GPU to boost efficiency. Furthermore, to improve convergence speed, we incorporate Nesterov's momentum, a method that helps the model learn faster and avoid getting stuck on less optimal solutions, into the update of PARAFAC2 using SGD.

Experiments

Datasets

We use three datasets to test REPAR, which represent different types of data, including binary, categorical, and numeric. Note that only REPAR changes the loss functions throughout the use of the three datasets.

MIMIC-III [25]: The ICU dataset is collected between 2001 and 2012. We keep records of patients with at least 10 hospital visits and construct a three-mode tensor. We select 405 medical NDC codes and 202 diagnosis codes that have the highest frequency as in [11].

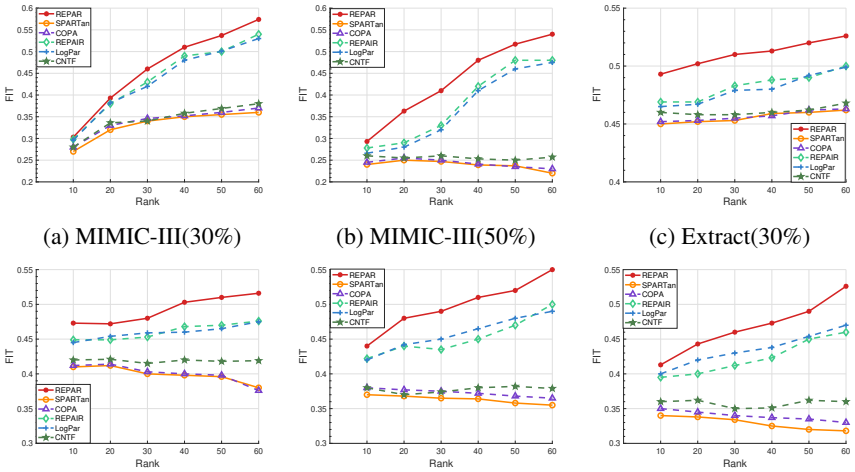
MIMIC-EXTRACT [26]: MIMIC-Extract is an open-source pipeline for transforming raw EHR data in MIMIC-III into data frames that are directly usable in common machine-learning pipelines. We use the vitals labs mean table, which contains 34,472 patients with 104 features (vital signs and laboratory measurements).

PhysioNet Sepsis Dataset [27]: PhysioNet 2019 Early Prediction of Sepsis from Clinical Data Challenge is an open-access dataset. It contains 20,336 patients with 40 time-dependent variables such as HR, O2Sat, Temp, etc. Since most of the features are extremely sparse, we select 6 dense features and then discretize the variables using criteria in [28].

Methods for Comparison

We compare REPAR[1] with 5 baseline methods: 1) SPARTan [16] and 2) COPA [17] are two state-of-the-art PARAFAC2 methods with different temporal regularizations, 3) REPAIR [1] and 4) LogPar [18] are two state-of-the-art robust PARAFAC2 methods to handle missing entries, 5) CNTF [21] represents each patient data separately and performs RNN regularized *regular* tensor factorization. The key difference between CNTF and REPAR is that CNTF treats each patient data separately as a *regular* tensor and ignores the interactions or correlations among patient groups, which can lead to low prediction accuracy as demonstrated by the empirical results in [17] and [22].

[1] https://github.com/Emory-AIMS/REPAR



(a) MIMIC-III(30%) (b) MIMIC-III(50%) (c) Extract(30%)

(d) Extract(50%) (e) Sepsis(30%) (f) Sepsis(50%)

Figure 2: FIT Scores Across Datasets: Comparison of FIT scores for REPAR and baseline models (SPARTan, COPA, REPAIR, LogPar, CNTF) on datasets with 30% and 50% missing data. REPAR consistently achieves higher scores, demonstrating superior robustness and effectiveness in handling missing entries and capturing temporal patterns.

Experiment Results

Tensor Factorization Robustness

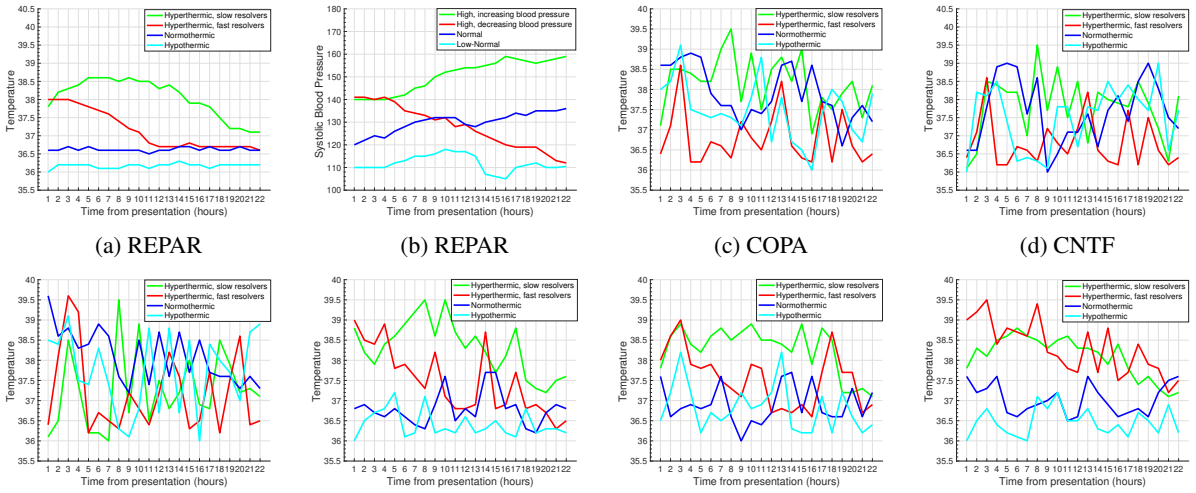
To test the robustness of REPAR model against missing entries, we randomly omit values into the two datasets. The original uncorrupted tensor, denoted as $\{\mathbf{G}_k\}$, serves as the ground truth. We adopt the $FIT \in (-\infty, 1]$ score [29], a measure of how well the model fits the data, as the quality measure (the higher the better):

$$FIT = 1 - \frac{\sum_{k=1}^{K} \|\mathbf{G}_k - \mathbf{U}_k\mathbf{S}_k\mathbf{V}^T\|^2}{\sum_{k=1}^{K} \|\mathbf{G}_k\|^2}. \tag{3}$$

We run each set for 5 different random initializations and report the average FIT. We set the missing ratio to 30% and 50% as we observe that the baselines' FIT scores drop significantly at higher missing ratios (30% to 50%) in [1], then we test model completion performance under different target ranks, R, from 10 to 60. As Figure 2 shows, REPAR outperforms all the other baseline methods on all datasets under both missing ratio settings. In particular, REPAR achieves a FIT score of 0.574 and 0.524 on MIMIC-III when the missing ratio equals 30% and 50% respectively, a 10% relative improvement when compared to the best baseline model REPAIR. REPAR shows the same outstanding performance with 7% and 10% improvement to the best baseline model for the MIMIC-EXTRACT (REPAIR) and Sepsis (LogPar) datasets respectively. LogPar and REPAIR perform better than CNTF, COPA, and SPARTan thanks to their regularizations to handle missing entries. CNTF and COPA perform slightly better than SPARTan because of the temporal smoothness regularization. LogPar outperforms REPAIR on the Sepsis dataset but is worse on MIMIC-III and MIMIC-EXTRACT. This demonstrates the importance of appropriately tuning the loss function as Sepsis is a binary dataset. Hence the non-negative positive-unlabeled loss in LogPar is more suitable for such data. Comparing different missing ratios, REPAR demonstrates superior robustness when the missing ratio increases to 50% from 30%. Though REPAIR and LogPar have low-rank regularization, they do not model the non-linear temporal information, which is key to achieving better robustness via missing value recovery.

Multi-Attribute Phenotype Trajectories with Clinical Expert Adjudication

Compared to traditional clustering based phenotyping methods, REPAR not only can group similar patients but also discover novel multi-attribute phenotypes which is difficult using the other methods. We demonstrate the effectiveness



(a) REPAR (b) REPAR (c) COPA (d) CNTF

(e) SPARTan (f) REPAIR (g) LogPar (h) REPAR-1 (ablation study)

Figure 3: Multi-attribute Phenotype Trajectory: (a-b) REPAR identifies four distinct subgroups based on temperature and systolic blood pressure trajectories, adjudicated by clinical experts. (c-g) Other methods struggle to uncover distinct temporal patterns or generate noisier phenotype trajectories, demonstrating lower clinical interpretability.

of REPAR to find multi-attribute dynamic phenotype trajectories with the missing ratio of 30% using the Sepsis dataset. For our visualization, we focused on patients who had 36 observations over time and set the model to identify 4 distinct patient subgroups. We selected two key measurements, temperature and systolic blood pressure. We then assigned each patient to the phenotype best matched using the highest membership value in $\mathbf{S}$. We tracked the temporal trajectory of the average measurement values for each subgroup without post-processing the extracted phenotypes.

As shown in Figure 3a, REPAR identifies four subgroups that have clearly different temporal patterns in both temperature and systolic blood pressure. Each of the 4 phenotypes was reviewed by a critical care expert. The clinical expert matched the four groups to previously validated temperature trajectory based subgroups: 1) hyperthermic slow resolver which exhibits a high temperature and slow decreasing trend as time increases, 2) hyperthermic fast resolver which exhibits a high temperature and a fast decreasing trend as time increases, 3) normothermic with normal temperature, and 4) hypothermic with low temperature. This corresponds to the temperature trajectory based patient subgroups (phenotypes) identified in the recent work by clinical experts [7]. Correspondingly, these four groups also exhibit different systolic blood pressure temporal patterns, namely: 1) high, increasing blood pressure, 2) high, decreasing blood pressure, 3) normal, and 4) low-normal as shown in Figure 3b. This demonstrates the ability of REPAR for identifying multi-attribute temporal phenotypes, which the single-attribute trajectory based methods can not capture.

Phenotype Presentation with Clinical Expert Adjudication

We present the representative phenotypes discovered by REPAR on MIMIC-III and MIMIC-EXTRACT in Table 2a and 2b with the missing ratio set to 30%. We set the rank to 3. For MIMIC-III, which contains counts of medical events, we identified the most frequent features associated with each subgroup (i.e., the top-weighted features from each column of V). For MIMIC-EXTRACT, a tensor with numerical measurements like vital signs, we identified the most important vital signs using the top-weighted features from each column of V. We then use the $\mathbf{S}$ matrix to find the patient subgroup of each phenotype and calculate the average value of the vital signs shown in the "Average value" column. Similar to the phenotype trajectories, we do not post-process these extracted phenotypes.

A critical care expert reviewed and adjudicated the medical concepts that apply to the extracted phenotype presentations. For example, the phenotype "Abnormalities in vital signs" corresponds to patients with low oxygen saturation, high temperature, high systolic blood pressure, and high heart rate. The "metabolic syndrome" phenotype references patients with high glucose and high blood pressure.

Table 2: Phenotypes discovered by REPAR. The red color corresponds to diagnosis and blue color corresponds to medications.

(a) MIMIC-III

Cardiovascular Disturbances	**Weight**	**Average Value**
Cardiovascular syph NEC	2.12	2.6
Coronary atherosclerosis of autologous vein bypass graft	2.03	3.4
Atrial fibrillation	1.94	5.9
Metoprolol Tartrate	1.84	1.4
Labetalol	1.73	3.8
Acetaminophen	1.64	4.9
Electrolyte Disturbances	**Weight**	**Average Value**
Electrolyte and fluid disorders not elsewhere classified	6.74	9.1
Functional diarrhea	5.79	8.4
Vomiting alone	4.15	14.3
Potassium Chloride	3.39	3.9
Magnesium Sulfate	2.18	10.1
Hydromorphone	1.12	2.3
Gastrointestinal Disturbances	**Weight**	**Average Value**
Hemorrhage of gastrointestinal tract, unspecified	19.2	2.4
Gastrointestinal vessel anomaly	17.3	3.5
Malignant neoplasm of body of stomach	16.5	4.7
Ipratropium Bromide Neb	13.9	2.1
Fentanyl Citrate	7.9	3.7
Lactulose	3.91	4.6

(b) MIMIC-EXTRACT

Abnormalities in Vital Signs	**Weight**	**Average Value**
Oxygen Saturation	0.91	85
Temperature	0.82	38.5
Systolic Blood Pressure	0.79	153.1
Heart Rate	0.61	115
Metabolic Syndrome	**Weight**	**Average Value**
Glucose	7.54	150
Systolic Blood Pressure	4.93	153.1
Mean Blood Pressure	3.41	95
Diastolic Blood Pressure	2.73	82
Abnormalities in Blood Counts and Serum Electrolytes	**Weight**	**Average Value**
Mean Corpuscular Hemoglobin Concentration	2.14	30.8
Sodium	1.41	135.2
Mean Corpuscular Hemoglobin	1.30	30.8
Chloride	1.03	103

Downstream Prediction Analysis

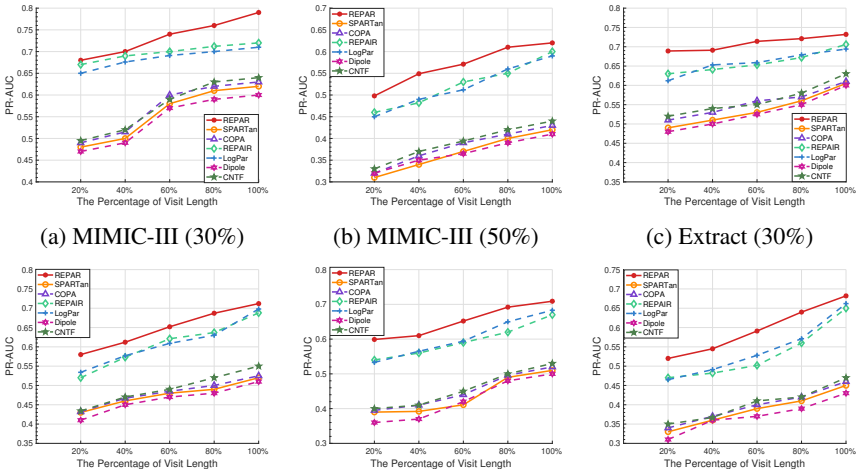
We further evaluate the derived phenotypes' predictability power via a downstream prediction task. We predict in-hospital mortality on MIMIC-III and MIMIC-EXTRACT datasets using the in-hospital death flag, and predict if a patient will have sepsis on PhysioNet Sepsis dataset. We split the data with a proportion of 8:2 as training and test sets and use PR-AUC (Area Under the Precision-Recall Curve) score to evaluate the predictive power. A logistic regression model is trained on the patient importance membership matrix $\mathbf{S}_k$. Besides tensor models, we also include a non-tensor prediction model. We adopt Dipole [30], an attention-based RNN model for diagnosis prediction. We vary the percentage of visit length used for prediction to mimic this real-world setting.

As shown in Figure 4, REPAR outperforms all the other methods. When the visit length ratio is 60%, REPAR outperforms the best baseline methods by 8%, 9% and 16% in Figure 4b, 4d, and 4f respectively. This demonstrates strong predictability even with missing values. Due to the RNN regularization, REPAR can learn the hidden non-linear temporal dependency and patterns, which can be used to improve missing value recovery and predictability. CNTF, COPA and SPARTan perform worse than REPAR, REPAIR, and LogPar due to the lack of low-rank regularization to handle missing values. The non-tensor based Dipole model performs the worst on all 3 datasets because the input data is sparse and noisy, demonstrating the benefit of tensor factorization in filtering out the noise and identifying the important and inherent features from the data.

Conclusion

We proposed REPAR, an RNN-Regularized Robust PARAFAC2 for Irregular Temporal Tensor Factorization model for capturing the complex temporal relationships and addressing the missing values in temporal EHR. REPAR has a new hybrid optimization framework using stochastic gradient descent and proximal average that can handle multiple regularizations and generalized loss functions. We conducted extensive experiments on EHR datasets to demonstrate that REPAR can robustly extract meaningful and high predictability phenotypes in the presence of missing data. The results are verified by the clinical expert on our team to discover meaningful medical concepts and identify patient subgroups for complex disease such as sepsis. Future work directions may include validating these phenotypes across multiple hospitals to ensure they are consistently meaningful; jointly optimizing tensor factorization and various prediction tasks together to further enhance predictability; incorporating more advanced architectures like Transformers to capture more complex temporal interactions, and integrating REPAR into hospital decision-support systems.

REPAR's ability to uncover complex temporal phenotypes has significant practical implications. For instance, in ICUs, REPAR could assist in real-time monitoring of patient trajectories to identify those at risk of deteriorating. Its robustness to missing data is particularly valuable in clinical settings where measurement frequencies vary, ensuring



(a) MIMIC-III (30%) (b) MIMIC-III (50%) (c) Extract (30%)

(d) Extract (50%) (e) Sepsis(30%) (f) Sepsis (50%)

Figure 4: PR-AUC for downstream prediction tasks on the different datasets with the missing ratio in parenthesis (missing percentage %). For MIMIC-III and MIMIC-Extract, the classification task is mortality prediction while Sepsis is sepsis prediction.

reliable insights even in sparse datasets. Additionally, the phenotypes identified by REPAR can inform personalized treatment plans, enabling precision medicine approaches for complex diseases like sepsis and chronic conditions requiring long-term management.

Acknowledgements

This work was supported by National Science Foundation award CNS-2124104; National Institute of Health awards R01LM013712, U01CA274576 and GM144867.

References

1. Ren Y, Lou J, Xiong L, Ho JC. Robust Irregular Tensor Factorization and Completion for Temporal Health Data Analysis. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management; 2020. p. 1295-304.
2. Yadav P, Steinbach M, Kumar V, Simon G. Mining Electronic Health Records: A Survey. ACM Computing Surveys. 2017;50(6):1-40.
3. Miravitlles M, Calle M, Soler-Cataluña JJ. Clinical Phenotypes of COPD: Identification, Definition and Implications for Guidelines. Archivos de Bronconeumología (English Edition). 2012;48(3):86-98.
4. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. npj Digital Medicine. 2019 09;2:1-11.
5. Kirby JC, Speltz P, Rasmussen LV, Basford M, Gottesman O, Peissig PL, et al. PheKB: A catalog and workflow for creating electronic phenotype algorithms for transportability. Journal of the American Medical Informatics Association. 2016 03;23:ocv202.
6. Seymour CW, Kennedy JN, Wang S, Chang CCH, Elliott CF, Xu Z, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA. 2019;321(20):2003-17.
7. Bhavani SV, Carey KA, Gilbert ER, Afshar M, Verhoef PA, Churpek MM. Identifying Novel Sepsis Subphenotypes Using Temperature Trajectories. American Journal of Respiratory and Critical Care Medicine. 2019;200(3):327-35.
8. Kolda TG, Bader BW. Tensor Decompositions and Applications. SIAM Rev. 2009;51:455-500. Available from: https://api.semanticscholar.org/CorpusID:16074195.
9. Ho JC, Ghosh J, Steinhubl SR, Stewart WF, Denny JC, Malin BA, et al. Limestone: High-throughput Candidate Phenotype Generation via Tensor Factorization. Journal of Biomedical Informatics. 2014;52:199-211.

10. Ho JC, Ghosh J, Sun J. 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; 2014. p. 115-24.
11. Kim Y, El-Kareh R, Sun J, Yu H, Jiang X. Discriminative and Distinct Phenotyping by Constrained Tensor Factorization. Scientific reports. 2017;7(1):1114.
12. Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. Psychometrika. 1970;35:283-319.
13. Snyder CW, Law HG, Pamment PR. Calculation of Tucker's three-mode common factor analysis. Behavior Research Methods and Instrumentation. 1979;11:609-11.
14. Kilmer ME, Braman K, Hao N, Hoover RC. Third-Order Tensors as Operators on Matrices: A Theoretical and Computational Framework with Applications in Imaging. SIAM Journal on Matrix Analysis and Applications. 2013;34(1):148-72.
15. Harshman RA. Parafac2: Mathematical and Technical Notes. UCLA Working Papers in Phonetics. 1972:30-44.
16. Perros I, Papalexakis EE, Wang F, Vuduc R, Searles E, Thompson M, et al. SPARTan: Scalable PARAFAC2 for large & sparse data. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017. p. 375-84.
17. Afshar A, Perros I, Papalexakis EE, Searles E, Ho J, Sun J. COPA: Constrained PARAFAC2 for Sparse & Large Datasets. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management; 2018. p. 793-802.
18. Yin K, Afshar A, Ho JC, Cheung WK, Zhang C, Sun J. LogPar: Logistic PARAFAC2 Factorization for Temporal Binary Data with Missing Values. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2020. p. 1625-35.
19. Ramsay JO. Monotone Regression Splines in Action. Statistical Science. 1988:425-41.
20. Chatzis C, Schenker C, Pfeffer M, Acar E. tPARAFAC2: Tracking evolving patterns in (incomplete) temporal data. arXiv preprint arXiv:240701356. 2024.
21. Yin K, Qian D, Cheung WK, Fung BC, Poon J. Learning Phenotypes and Dynamic Patient Representations via RNN Regularized Collective Non-Negative Tensor Factorization. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33; 2019. p. 1246-53.
22. Purushotham S, Meng C, Che Z, Liu Y. Benchmark of Deep Learning Models on Large Healthcare MIMIC Datasets. Journal of Biomedical Informatics. 2017;83:112-34.
23. Hong D, Kolda TG, Duersch JA. Generalized Canonical Polyadic Tensor Decomposition. SIAM Review. 2020;62(1):133–163.
24. Goldfarb D, Qin Z. Robust low-rank tensor recovery: Models and algorithms. SIAM Journal on Matrix Analysis and Applications. 2014;35(1):225-53.
25. Johnson AE, Pollard TJ, Shen L, Lehman LwH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific Data. 2016;3(1):1-9.
26. Wang S, McDermott MB, Chauhan G, Ghassemi M, Hughes MC, Naumann T. MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III. In: Proceedings of the ACM conference on health, inference, and learning; 2020. p. 222–235.
27. Reyna MA, Josef CS, Jeter R, Shashikumar SP, Westover MB, Nemati S, et al. Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. Critical Care Medicine. 2019;48(2).
28. Comstedt P, Storgaard M, Lassen AT. The Systemic Inflammatory Response Syndrome (SIRS) in acutely hospitalised medical patients: a cohort study. Scandinavian journal of trauma, resuscitation and emergency medicine. 2009;17:1-6.
29. Bro R, Andersson CA, Kiers HA. PARAFAC2—Part II. Modeling chromatographic data with retention time shifts. Journal of Chemometrics. 1999.
30. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining; 2017. p. 1903-11.