Federated Tensor Factorization for Computational Phenotyping

Yejin Kim[1,3], **Jimeng Sun**[2], **Hwanjo Yu**[1], and **Xiaoqian Jiang**[3]

[1]Pohang University of Science and Technology, Pohang, Korea

[2]Georgia Institute of Technology, Atlanta, GA

[3]University of California, San Diego, La Jolla, CA

Abstract

Tensor factorization models offer an effective approach to convert massive electronic health records into meaningful clinical concepts (phenotypes) for data analysis. These models need a large amount of diverse samples to avoid population bias. An open challenge is how to derive phenotypes jointly across multiple hospitals, in which direct patient-level data sharing is not possible (e.g., due to institutional policies). In this paper, we developed a novel solution to enable federated tensor factorization for computational phenotyping without sharing patient-level data. We developed secure data harmonization and federated computation procedures based on alternating direction method of multipliers (ADMM). Using this method, the multiple hospitals iteratively update tensors and transfer secure summarized information to a central server, and the server aggregates the information to generate phenotypes. We demonstrated with real medical datasets that our method resembles the centralized training model (based on combined datasets) in terms of accuracy and phenotypes discovery while respecting privacy.

Keywords

ADMM; Federated approach; Phenotype

1 INTRODUCTION

Electronic health records (EHRs) become one of most important sources of information about patients, which provide insight into diagnoses [19] and prognoses [11], as well as assist in the development of cost-effective treatment and management programs [1, 12]. But meaningful use of EHRs is also accompanied with many challenges, for example, diversity of populations, heterogeneous of information, and data sparseness. The large degree of missing and erroneous records also complicates the interpretation and analysis of EHRs. Furthermore, clinical scientists are not used to the complex and high-dimensional EHR data [8, 21]. Instead, they are more accustomed to reasoning based on accurate and concise clinical concepts (or phenotypes) such as diseases and disease subtypes. Useful phenotypes

CCS CONCEPTS

•**Applied computing**→**Health informatics;** •**Computing methodologies**→*Factorization methods; Machine learning algorithms;*

should capture multiple aspects of the patients (e.g., diagnosis, medication and lab results) and be both sensitive and specific to the target patient population. Although some phenotypes can be easily concluded based on EHR data, a wide range of clinically important ones such as disease subtypes are not obtainable in a straightforward manner. The transformation from EHR data into useful phenotypes, or *phenotyping* is a fundamental challenge to learn from EHR data. Current approaches for translating EHR data into useful phenotypes are typically slow, manually intensive and limited in scope [4, 5]. Overcoming several disadvantages of the previous methods, tensor factorization methods have shown great potential in discovering meaningful phenotypes from complicated and heterogeneous health records [13, 14, 28].

Nevertheless, phenotypes developed from one hospital are often limited due to a small sample size and inherent population bias. Ideally, we would like to compute phenotypes on a large population with data combined from multiple hospitals. However, this will require healthcare data sharing and exchange, which are often impeded by policies due to the privacy concerns. For example, PCORnet data privacy guidance does not allow record-level research participant information sharing and it recommends a minimum count threshold (e.g., 10) for aggregate data sharing [25]. The same threshold is used in Informatics for Integrating Biology & the Bedsides (I2B2) [24], a famous system developed by National Center for Biomedical Computing based at Partners HealthCare. The real-world challenges motivate the development of a *federated phenotyping method* to learn phenotypes across multiple hospitals with mitigated privacy risks.

In the federated method, the hospitals perform most of computations, and a semi-trusted server supports the hospital by aggregating results from hospitals. The hospitals demand a certain form of summarized patient information (not patient-level data) anyhow for updating tensor. A challenge of the federated tensor factorization is that the summarized information can disclose the patient-level data. For example, an objective function of tensor factorization is $\|\mathscr{X} - \mathscr{O}\|^2$ where $\mathscr{X}$ is a tensor to be estimated using an observed tensor $\mathscr{O}$. Because the objective function is not linearly separable over hospitals, tensor factorization for each hospital inevitably demands the others patient-level data. Thus, hospitals should share summarized information that does not disclose the patient-level data but instead contains accurate phenotypes from the patient-level data.

However, sharing the summarized information raises another challenge when the data are distributed in many hospitals as a relatively small size, or when the data are unevenly distributed. Because of sampling error, noise in the summarized information can increase with small patient populations. Accuracy then can be decreased or unstable. Therefore, we need to ensure the robustness of summary information even with small sized or unevenly distributed samples.

In this paper, we develop federated **T**ensor factorization for p**ri**vacy preserving computational **p**henotyping (Trip), a new federated framework for tensor factorization over horizontally partitioned data (i.e., data are partitioned based on rows or patients). Our major contributions are the following:

i. **Accurate and fast federated method**: Trip is as accurate as centralized training model (based on combined datasets). The accuracy of Trip is robust on the patient size or distribution. Trip is fast compared to the centralized training model thanks to federated computation.

ii. **Rigorous privacy and security analysis**: Trip preserves the privacy of patient data by transferring summarized information. We prove that the summarized information does not disclose the patient data.

iii. **Phenotype discovery from real datasets**: Phenotypes that Trip discovers without sharing the patient-level data are the same phenotypes based on the combined data. Trip even discovers some phenotypes that individual hospital cannot discover due to biased and limited population.

2 RELATEDWORKS

Many privacy preserving data mining algorithms aim at constructing a global model only from aggregated statistics locally generated by participating institutions on their own data, without seeing others' data at a fine-grained level [22, 27]. More rigorous privacy criteria like *differential privacy* [10], which introduces noises, have been applied for several classification models through parameter or objective perturbations [6]. However, this is not desirable for computational phenotyping applications because noise can lead to "ghost" phenotypes, which do not exist in the original databases and might mislead healthcare providers with severe consequences. In this work, we will consider privacy protection like in the former privacy preserving data mining methods to compute phenotypes by only exchanging summary statistics, calculated by local participants.

Tensor factorization emerged as a promising solution for computational phenotyping thanks to its interpretability and Flexibility. In the medical context, tensor factorization has been adapted to enforce sparsity constraints [13], model interactions among groups of the same modality [14], and absorbing prior medical knowledge via customized regularization terms [28]. Our goal is to develop a federated tensor factorization framework to compute phenotypes in a privacy-preserving way. This is different from distributed tensor factorization models [7, 16] and grid tensor factorization models [9]. The latter assumes data spread across different but interconnected computer systems, in which the communication cost is negligible and data/computation can be arbitrarily reallocated to improve parallelization efficiency. In contrast, our Trip framework deals with data stored in separate sources (hospital at different locations) and requires the ability to go through policy barriers using accepted practices that respect privacy.

3 PRELIMINARIES

We first describe some preliminaries of tensor factorization, and summarize the notations and symbols in Table 1.

Definition 3.1: Outer product of N vectors $\mathbf{a}^{(1)} \circ \cdots \circ \mathbf{a}^{(N)}$ makes N-order rank-one tensor $\mathscr{X}$.

Definition 3.2: Kronecker product of two vectors $\mathbf{a} \in \mathbb{R}^{I_a \times 1}$ and $\mathbf{b} \in \mathbb{R}^{I_b \times 1}$ is

$$\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} a_1 \mathbf{b} \\ \vdots \\ a_{I_a} \mathbf{b} \end{bmatrix} \in \mathbb{R}^{I_a I_b \times 1},$$

Definition 3.3: Kharti-Rao product of two matrices $\mathbf{A} \in \mathbb{R}^{I_A \times R}$ and $\mathbf{B} \in \mathbb{R}^{I_B \times R}$ is $\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \cdots \mathbf{a}_R \otimes \mathbf{b}_R] \in \mathbb{R}^{I_A I_B \times R}$.

Definition 3.4: *Matricization* is to reshape the tensor into a matrix by unfolding elements of the tensor. Mode-n matricization of tensor $\mathscr{O}$ is denoted as $\mathbf{O}_{(n)}$.

Tensor factorization is a dimensionality reduction approach that represents the original tensor as a lower dimensional latent matrix. The CANDECOMP/PARAFAC (CP) [3] model is the most common tensor factorization, which approximates the original tensor $\mathscr{O}$ as $\mathscr{X}$, a linear combination of R rank-one tensors that are made from outer product of N vectors. That is, CP tensor factorization is represented as

$$O \approx \mathscr{X} = \sum_{r=1}^{R} \mathbf{A}^{(1)}(:, r) \circ \cdots \circ \mathbf{A}^{(N)}(:, r),$$

where $\mathbf{A}^{(n)}(:, r)$ refers to the rth column of $\mathbf{A}^{(n)}$. Here, $\mathbf{A}^{(n)}$ is the nth factor matrix. R is referred as the rank of the $\mathscr{X}$. The columns from factor matrices represent latent concepts that describe the data as lower dimensions.

Tensor factorization for phenotyping is to compute a factorized tensor $\mathscr{X}$ that contains latent medical concepts from data (or observed tensor $\mathscr{O}$). $\mathscr{X}$ consists of the R most prevalent phenotypes. The nth factor matrix, $\mathbf{A}^{(n)}$ defines the elements from the mode n to comprise the phenotypes. That is, rth phenotype consists of rth column of factor matrices [13].

The objective function of the tensor factorization with regularization terms for pairwise distinct constraints [28] is formulated as

$$\min_{\mathscr{X}} \Psi = \| \mathscr{X} - O \|_F^2 + \sum_{n=1}^{N} \frac{\lambda}{2} \| \mathbf{I} - \mathbf{A}^{(n)T} \mathbf{A}^{(n)} \|_F^2. \tag{1}$$

It is rewritten with respect to mode-n matricization

$$\min_{\mathbf{A}^{(n)}} \Psi = \| \mathbf{A}^{(n)} \Pi^{(n)T} - \mathbf{O}_{(n)} \|_F^2 + \frac{\lambda}{2} \| \mathbf{I} - \mathbf{A}^{(n)T} \mathbf{A}^{(n)} \|_F^2. \tag{2}$$

where $\Pi^{(n)} = \mathbf{A}^{(N)} \odot \ldots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \ldots \odot \mathbf{A}^{(1)}$. This is our decomposition goal in the rest of this paper. Solving the problem (1) while preserving privacy is technically challenging because the tensor residual term $\mathscr{X} - \mathscr{O}$ inherently contains other hospitals' patient data that involve sensitive information.

4 FEDERATED TENSOR FACTORIZATION

We first provide a general overview of the Trip and then formulate the problem with iterative updating rules for optimization.

4.1 Overview

Trip is a federated tensor factorization for horizontally partitioned patient data. We assume the data are horizontally partitioned along patient mode, that is, hospitals have their own patient data on the same medical features (Figs. 1, 2). Let us assume that there are K hospitals and a central server, where the server distributes most decomposition computation to hospitals and aggregates intermediate results from them. We assume *Honest-but-Curious* adversary model, in which the server and hospitals are curious on data of others but do not maliciously manipulate intermediate results [18].

A local observed tensor $\mathscr{O}_k$ is the local patient data in hospital k (Fig. 2); a local factorized tensor $\mathscr{X}_k$ is the factorized tensor generated by local observed tensor in hospital k, $\mathscr{X}_k$ has N modes for the set of patient and medical features (eg. medication, diagnosis). In this case, N = 3 because we have modes for patient, medication, and diagnosis. The horizontally partitioned patient mode of each $\mathscr{X}_k$ is generated from distinct set of patients whose size is I_{1k}. For simplicity, first mode (n = 1) always denotes patient mode. On the other hand, $N-1$ medical features modes that hospitals share of each $\mathscr{X}_k$ is generated from the same set of N − 1 medical features whose size is I_n, (n = 2,…, N). For example, diagnosis and medication can be the feature modes. The size of $\mathscr{X}_k$ is $I_{1k} \times I_2 \times \cdots \times I_N$, $\forall k$.

We assume that factor matrix on feature modes of the local factorized tensor $\mathscr{X}_k$ is the same for all the hospitals. By assuming that, all hospitals are enforced to share the same phenotypes. Also, the objective function Ψ in Eq. (1) can be linearly separable on hospitals; consequently hospitals can update their local factorized tensor indirectly using other hospitals' patient data while respecting privacy.

The local factorized tensor $\mathscr{X}_k$ is computed as following steps: first, in patient mode, hospital k (k = 1, …, K) computes local factor matrix independently (step 1) in Fig. 1. For feature modes, hospital k computes the local factor matrices (step 2) and send them together with the Lagrangian multipliers to the server (step 3). The server then generates harmonized factor matrix (global factor matrix) by combining all the local factor matrices with Lagrangian multipliers (step 4). After receiving the global factor matrix (step 5), hospital k updates the Lagrangian multipliers (step 6). Hospitals and the server repeat the procedures until the local factor matrices are converged. During the procedures, the global factor matrices can retain phenotypes from local factor matrices without directly using the local patient data.

4.2 Formulation

We first formulate separable objective function on hospitals for federated tensor factorization. The objective function for tensor factorization, Ψ in Eq. (1) is reformulated with respect to the local factorized tensor.

$\mathcal{X}_k$ is decomposed into factor matrices $\mathbf{A}_k^{(1)} \in \mathbb{R}^{I_{1k} \times R}$ (patient mode) and $\mathbf{A}_k^{(n)} \in \mathbb{R}^{I_n \times R}$, n

2 (feature modes). We assume that the local factor matrices of feature modes $\mathbf{A}_k^{(n)}$ from all hospitals are equal to the global factor matrix (Fig. 2), i.e.,

$$\mathbf{A}^{(n)} = \mathbf{A}_1^{(n)} = \mathbf{A}_2^{(n)} = \ldots = \mathbf{A}_K^{(n)}, n \geq 2. \quad (3)$$

This assumption is reasonable because all hospitals aim to have the same phenotypes and share them with others. By assuming Eq. (3), the horizontal concatenation of the local factor matrices of patient mode $\mathbf{A}_k^{(1)}$ forms the (global) factor matrix $\mathbf{A}^{(1)}$ (Fig. 2):

$$\mathbf{A}^{(1)} = \begin{bmatrix} \mathbf{A}_1^{(1)}; \\ \vdots \\ \mathbf{A}_K^{(1)} \end{bmatrix}. \quad (4)$$

Accordingly, we represent the global factorized tensor $\mathcal{X}$ in Eq. (1) with respect to the local factorized tensor $\mathcal{X}_k$ (Fig. 2) as

$$\mathcal{X} = \begin{bmatrix} \mathcal{X}_1; \\ \vdots \\ \mathcal{X}_K \end{bmatrix} = \begin{bmatrix} \sum_r \mathbf{A}_1^{(1)}(:,r) \circ \mathbf{A}^{(2)}(:,r) \circ \cdots \circ \mathbf{A}^{(N)}(:,r); \\ \vdots \\ \sum_r \mathbf{A}_K^{(1)}(:,r) \circ \mathbf{A}^{(2)}(:,r) \circ \cdots \circ \mathbf{A}^{(N)}(:,r) \end{bmatrix},$$

and we can make the objective function Ψ linearly separable on k as $\|\mathcal{X} - O\|_F^2 = \sum_{k=1}^K \|\mathcal{X}_k - O_k\|_F^2$. The optimization problem for tensor factorization is reformulated with respect to local tensors:

$$\min_{\mathcal{X}_k} \Psi = \sum_{k=1}^K \|\mathcal{X}_k - O_k\|_F^2 + \sum_{n=2}^N \frac{\lambda}{2} \|\mathbf{I} - \mathbf{B}^{(n)T} \mathbf{A}^{(n)}\|_F^2$$
$$s.t. \mathbf{A}^{(n)} = \mathbf{A}_k^{(n)} \ n \geq 2, \forall k$$
$$\mathbf{B}^{(n)} = \mathbf{A}^{(n)} \ n \geq 2. \quad (5)$$

Here, the non-convex second term $\|\mathbf{I} - \mathbf{A}^{(n)T} \mathbf{A}^{(n)}\|_F^2$ in Eq. (1) is replaced to a convex term $\|\mathbf{I} - \mathbf{B}^{(n)T} \mathbf{A}^{(n)}\|_F^2$ using $\mathbf{B}^{(n)}$ such that $\mathbf{A}^{(n)} = \mathbf{B}^{(n)}$. We assume that the pairwise constraint is

only applied to the feature modes. This assumption is reasonable because phenotypes are defined as only combination of medical features in feature modes.

Augmented Lagrangian function $\mathscr{L}$ for the new optimization problem (5) is

$$\mathscr{L}=\Psi+\sum_{k=1}^{K}\sum_{n=2}^{N}\left[\left(\mathbf{A}^{(n)}-\mathbf{A}_k^{(n)}\right)^T\mathbf{H}_k^{(n)}+\frac{\omega}{2}\|\mathbf{A}^{(n)}-\mathbf{A}_k^{(n)}\|_F^2\right]+\sum_{n=2}^{N}\left[\left(\mathbf{B}^{(n)}-\mathbf{A}^{(n)}\right)^T\mathbf{Y}^{(n)}+\frac{\mu}{2}\|\mathbf{B}^{(n)}-\mathbf{A}^{(n)}\|_F^2\right]$$

where $\mathbf{H}_k^{(n)}$ and $\mathbf{Y}^{(n)}$ are the Lagrangian multipliers. The penalty terms that are multiplied by parameter ω and μ help $\mathscr{L}$ to improve the convergence property (i.e., *method of multiplier*) [23] during federated optimization in Section 4.3.

4.3 Federated optimization

The optimization problem (5) is then solved via consensus alternating direction method of multipliers (ADMM) [2], which decomposes the original problem into sub-problems using auxiliary variables and ensures convergence to a stationary point even with nonconvex problem [15]. Our problem is decomposed to sub-problems for hospitals with respect to the local factor matrices. Individual components of the local factor matrices are iteratively updated while other local factor matrices are fixed. Once all hospitals update the local factor matrices, server updates the global factor matrix and send it back to every hospital. Hospitals and server repeat this procedure until the local factor matrices converge before maximum iteration.

4.3.1 Patient mode—Because the factor matrix for patient mode does not need to be shared, each hospital updates the local factor matrix without sharing the intermediate results. The local matricized residual tensor on patient mode is

$$(\mathscr{X}_k-O_k)_{(1)}=\mathbf{A}_k^{(1)}\Pi^{(1)T}-\mathbf{O}_{(1)k}\in\mathbb{R}^{I_{1k}\times(I_2\cdots I_N)}.$$

Horizontal concatenation of the local matricized residual tensors $\mathbf{A}_k^{(1)}\Pi^{(1)T}-\mathbf{O}_{(1)k}$ from K hospitals becomes the global matricized residual tensor $\mathbf{A}^{(1)}\Pi^{(1)T}-\mathbf{O}_{(1)}$. To compute $\mathbf{A}_k^{(1)}$, we separate the first term in Ψ in Eq. (1) to each hospital as

$$\|\mathbf{A}^{(1)}\Pi^{(1)T}-\mathbf{O}_{(1)}\|_F^2=\sum_{k=1}^{K}\|\mathbf{A}_k^{(1)}\Pi^{(1)T}-\mathbf{O}_{(1)k}\|_F^2. \tag{6}$$

By setting derivatives of Ψ with respect to $\mathbf{A}_k^{(1)}$ to zero, a closed form solution for updating $\mathbf{A}_k^{(1)}$ is

$$\mathbf{A}_k^{(1)} = \{\mathbf{O}_{(1)k}\Pi^{(1)}\}\{\Pi^{(1)T}\Pi^{(1)}\}^{-1}. \quad (7)$$

4.3.2 Feature modes—Hospitals update the local factor matrices using the global factor matrix, and server makes the global factor matrix by aggregating the intermediate local factor matrices from hospitals in turn.

Update the local factor matrices: The local matricized residual tensor on feature modes is

$$(\mathcal{X}_k - O_k)_{(n)} = \mathbf{A}^{(n)}\Pi_k^{(n)T} - \mathbf{O}_{(n)k} \in \mathbb{R}^{I_n \times (I_N \cdots I_{n+1} I_{n-1} \cdots I_{1k})}$$

where $\Pi_k^{(n)} = \mathbf{A}_k^{(N)} \odot \ldots \odot \mathbf{A}_k^{(n+1)} \odot \mathbf{A}_k^{(n-1)} \odot \ldots \odot \mathbf{A}_k^{(1)}$, $n \geq 2$. Contrast to patient mode, vertical concatenation of the local matricized residual tensors $\mathbf{A}^{(n)}\Pi_k^{(n)T} - \mathbf{O}_{(n)k}$ becomes the global matricized residual tensor $\mathbf{A}^{(n)}\Pi^{(n)T} - \mathbf{O}_{(n)}$. The first term in Ψ becomes

$$\left\|\mathbf{A}^{(n)}\Pi^{(n)T} - \mathbf{O}_{(n)}\right\|_F^2 = \sum_{k=1}^K \left\|\mathbf{A}_k^{(n)}\Pi_k^{(n)T} - \mathbf{O}_{(n)k}\right\|_F^2 \quad (8)$$

with $\mathbf{A}^{(n)} = \mathbf{A}_k^{(n)}$. The closed form solution for $\mathbf{A}_k^{(n)}$ is

$$\mathbf{A}_k^{(n)} = \{\mathbf{O}_{(n)k}\Pi_k^{(n)} + \omega\mathbf{A}^{(n)} + \mathbf{H}_k^{(n)}\}\{\Pi_k^{(n)T}\Pi_k^{(n)} + \omega\mathbf{I}\}^{-1}. \quad (9)$$

This closed form solution updates the local factor matrices using the both local observed tensor $\mathbf{O}_{(n)k}$ and global factor matrix $\mathbf{A}^{(n)}$. That is, each hospital uses both their patient data and the common phenotypes from others to update their local phenotypes. Now, hospitals send the local information $\mathbf{A}_k^{(n)}$ and $\mathbf{H}_k^{(n)}$ to server for following updates on the global factor matrix.

Update the global factor matrix: Server updates the global factor matrix based on the local information. The objective function is

$$\min_{\mathbf{A}^{(n)}} \frac{\lambda}{2}\left\|\mathbf{I} - \mathbf{B}^{(n)T}\mathbf{A}^{(n)}\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{A}^{(n)} - \mathbf{B}^{(n)} - \mathbf{Y}^{(n)}/\mu\right\|_F^2 + \frac{\omega}{2}\sum_{k=1}^K \left\|\mathbf{A}^{(n)} - \mathbf{A}_k^{(n)} + \mathbf{H}_k^{(n)}/\omega\right\|_F^2$$

that also uses the pairwise constraint. $\mathbf{A}^{(n)}$ is updated to be similar with $\mathbf{A}_k^{(n)}$ in the third term. That is, the global phenotypes are made to be similar with all other hospitals'

phenotypes. By derivatives of this function with respect to $\mathbf{A}^{(n)}$, we derive the following closed form solution:

$$\mathbf{A}^{(n)}=\{(\mu+K\omega)\mathbf{I}+\lambda\mathbf{B}^{(n)}\mathbf{B}^{(n)T}\}^{-1}\cdot\{\lambda\mathbf{B}^{(n)}+\mu\mathbf{B}^{(n)}+\boldsymbol{Y}^{(n)}+\omega\sum_{k}\mathbf{A}_{k}^{(n)}-\sum_{k}\mathbf{H}_{k}^{(n)}\}. \tag{10}$$

Now, server sends the global information $\mathbf{A}^{(n)}$ to hospitals for the next iteration. Server updates $\mathbf{B}^{(n)}$ by

$$\mathbf{B}^{(n)}=\mathbf{A}^{(n)}+\frac{1}{\mu}\boldsymbol{Y}^{(n)}. \tag{11}$$

Algorithm 1

Trip

1	**Input**: $\mathcal{O}, \lambda, \omega, \mu$
2	Initialize $\mathbf{A}_{k}^{(n)}, \mathbf{H}_{k}^{(n)}, \mathbf{Y}^{(n)}$.
3	**repeat**
4	// Update patient mode $n=1$
5	Hospitals set $\mathbf{A}_{k}^{(1)} \, \forall k$ (Eq. 7).
6	**for** $n=2, …, N$ **do** //Update feature modes
7	Hospitals set and send $\mathbf{A}_{k}^{(n)} \, \forall k$ (Eq. 9).
8	Server sets and sends $\mathbf{A}^{(n)}$ (Eq. 10).
9	Server sets $\mathbf{B}^{(n)}$ and $\mathbf{Y}^{(n)}$ (Eq.11, 12).
10	Hospitals set and send $\mathbf{H}_{k}^{(n)} \, \forall k$ (Eq. 13).
11	**end for**
12	**until** Converged

<u>Update Lagrangian multipliers:</u> Finally, server updates Lagrangian multipliers as

$$\boldsymbol{Y}^{(n)}=\boldsymbol{Y}^{(n)}+\mu(\mathbf{B}^{(n)}-\mathbf{A}^{(n)}). \tag{12}$$

Hospitals also updates local Lagrangian multipliers as

$$\mathbf{H}_{k}^{(n)}=\mathbf{H}_{k}^{(n)}+\omega(\mathbf{A}^{(n)}-\mathbf{A}_{k}^{(n)}) \tag{13}$$

to adjust the gap between local and global factor matrices. The entire procedures of updating the tensors are summarized in Algorithm 1.

4.4 Convergence proof

We prove that our federated tensor factorization (5) converges. Due to limited space, detailed proof of inequality (17) and (22) can be found in our technical report [17] or [2]. For each n = 2, $\cdots$, N, let us denote

$$
\begin{aligned}
x \quad &= [\mathbf{A}_1^{(n)}(:), \ldots, \mathbf{A}_K^{(n)}(:)] \in \mathbb{R}^{I_n R \times K}, \\
z \quad &= [\mathbf{A}^{(n)}(:), \ldots, \mathbf{A}^{(n)}(:)] \in \mathbb{R}^{I_n R \times K}, \\
y \quad &= [\mathbf{H}_1^{(n)}(:), \ldots, \mathbf{H}_K^{(n)}(:)] \in \mathbb{R}^{I_n R \times K}, \text{ and} \\
r^t \quad &= x^t - z^t,
\end{aligned} \tag{14}
$$

for vectorized local factor matrices, global factor matrix, Lagrangian multipliers, and residual at iteration t, respectively. Then $\mathcal{L}$ is rewritten as

$$
\mathcal{L}(x, z, y) = f(x) + g(z) + y^T(x - z) + (\omega/2)\|x - z\|^2 \tag{15}
$$

where $f(x) = \sum_{k=1}^{K} \|\mathbf{A}_k^{(n)}\Pi_k^{(n)T} - \mathbf{O}_{(n)k}\|^2$ and

$g(z) = \frac{\lambda}{2}\|\mathbf{I} - \mathbf{B}^{(n)T}\mathbf{A}^{(n)}\|^2 + (\mathbf{B}^{(n)} - \mathbf{A}^{(n)})^T \mathbf{Y}^{(n)} + \frac{\mu}{2}\|\mathbf{B}^{(n)} - \mathbf{A}^{(n)}\|^2$. Let (x^*, z^*, y^*) be a saddle point, and define

$$
V^t = (1/\omega)\|y^t - y^*\|^2 + \omega\|z^t - z^*\|^2. \tag{16}
$$

V^t decreases in each iteration (proof in [17]) because

$$
V^{t+1} \le V^t - \omega\|r^{t+1}\|^2 - \omega\|z^{t+1} - z^t\|^2. \tag{17}
$$

Adding the inequality (17) through $t = 0$ to ∞ gives

$$
\omega \sum_{t=0}^{\infty} \left(\|r^{t+1}\|^2 + \|z^{t+1} - z^t\|^2 \right) \le V^0, \tag{18}
$$

which implies $r^t \to 0$ and $z^t \to z^{t+1}$ as $t \to \infty$.

Now, we define $p^t = f(x^t) + g(z^t)$ and show p^t converges. Because (x^*, z^*, y^*) is a saddle point,

$$\mathscr{L}(x^*, z^*, y^*) \leq \mathscr{L}(x^{t+1}, z^{t+1}, y^*). \quad (19)$$

That is, using $x^* = z^*$ at the saddle point,

$$p^* \leq p^{t+1} + y^{*T} r^{t+1} + (\omega/2)\|r^{t+1}\|^2, \quad (20)$$

which implies that upper bound of $p^* - p^{t+1}$ is

$$p^* - p^{t+1} \leq y^{*T} r^{t+1}. \quad (21)$$

Lower bound of $p^* - p^{t+1}$ (proof in [17]) is

$$p^* - p^{t+1} \geq (y^{t+1})^T r^{t+1} + \omega (z^{t+1} - z^t)^T (r^{t+1} + (z^{t+1} - z^*)). \quad (22)$$

The upper and lower bounds go to zero because $r^t \to 0$ and $z^t \to z^{t+1}$ as $t \to \infty$, i.e., $\lim_{t\to\infty} p^t = p^*$. Thus, the objective function Ψ of our federated optimization converges.

4.5 Privacy analysis

In our *Honest-but-Curious* adversary scenario, privacy of patient data is preserved because patient-level data are not disclosed to the both server and hospitals. The server and hospitals cannot access to unintended fine-grained local information. The local data are only accessible to the corresponding hospital. The server also cannot indirectly learn patient data from the local factor matrices. After receiving $\mathbf{A}_k^{(n)}$, the server might try to do reverse-engineering through Eq. (9). However, server cannot access to $\Pi_k^{(n)}$ because $\mathbf{A}_k^{(1)}$ from patient mode is not shared. If server accesses to $\Pi_k^{(n)}$ by any chance as $\mathbf{A}_k^{(1)}$ is leaked, reverse-engineering cannot still restore patient-level data. That is, the matricized unknown observed tensor (patient data) has an equation in form of $\mathbf{A}_k^{(n)} = \mathbf{O}_{(n)k} \Pi_k$ after removing all the known values in Eq. (9) for simplicity. The size of the unknown information in $\mathbf{O}_{(n)k}$ is $I_n \times (I_{1k} \cdots I_{n-1} I_{n+1} \cdots I_N)$, and the size of $\Pi_k^{(n)}$ and $\mathbf{A}_k^{(n)}$ is $(I_{1k} \cdots I_{n-1} I_{n+1} \cdots I_N) \times r$ and $I_n \times r$, respectively. Element-wise computation generates only $I_n \cdot r$ equations for the unknown $I_1 \cdots I_N$ values. Server cannot recover the unknown values from the $I_n \cdot r$ equations that is less than the number of unknown values (r is always selected as $I_n \cdot r \ll I_1 \cdots I_N$).

Hospitals also cannot learn other hospitals' data from the global factor matrix. If hospital k' knows all the information of Eq. (10) for global factor matrix by any chance, hospital k' cannot restore other hospitals' local factor matrix $\mathbf{A}_k^{(n)}$. If the hospital k' can access to $\mathbf{A}_k^{(n)}$ by any chance, $\mathbf{A}_k^{(n)}$ has still insufficient information to recover the data as shown in the case of server.

4.6 Secure alignment of feature modes

In Section 4.2, we first assume that hospitals have the same element set for each feature in feature modes, but in practice, hospitals may have different elements. For example, Hospital 1 and Hospital 2 have set of diagnosis: Y_1 and Y_2 (Fig. 3), but each index of Y_1 and Y_2 refers to a different element. In this case, before concatenating the local tensors, the index of feature modes should refer to the same element among hospitals. Thus, we introduce a secure alignment method for feature modes by which hospitals do not reveal the elements they have and get an integrated and sorted view on the elements of feature modes. This secure alignment enables hospitals to have only the position of its elements without knowing other hospitals' elements as like Y_1 and Y_2 are aligned to make the index from two sets refer to the same element (Fig. 3). For each feature mode, hospital k assigns integer values to the set of elements Y_k (eg. ICD9 codes). Hospitals use polynomial properties of set intersection [18]:

<u>Lemma 4.1:</u> *A polynomial function of y that represents set of elements $Y_k = \{y_{ik}\}_{i=1}^{I}$ at hospital k is $f_k(y) = (y - y_{1k})(y - y_{2k}) \cdots (y - y_{Ik}) = \sum_{i=0}^{I} a_i y^i$. A y_{ik} is an element of Y_k ($y_{ik} \in Y_k$) if and only if $f_k(y_{ik}) = 0$.*

<u>Lemma 4.2:</u> *A polynomial function that represents intersection of Y_k and $Y_{k'}$ ($Y_k \cap Y_{k'}$) is $f_k * r + f_{k'} * s$ where r, s are polynomial functions with gcd(r, s) = 1. Given $f_k * r + f_{k'} * s$, one cannot learn individual elements on Y_1 and Y_2 other than elements in $Y_1 \cap Y_2$.*

Hospitals express Y_k as a polynomial function (or in short polynomial) f_k by Lemma 4.1. To prevent the factorization of the polynomial, hospital k multiplies a term $r = (y - a)$ to f_k ($= f_k * r$), where a is a random prime number that is selected with overwhelming probability that the a does not represent any element from Y_k. For simplicity, $f_k * r$ is denoted as f_k. Because computing the polynomials with large $|Y_k|$ can cause computational overhead, hospitals compute the polynomials' modulus (denoted as %) by a random prime number P ($P > y_k$) instead of the polynomial itself, i.e., hospitals compute $f_k \% p = \left[\sum_{i=0}^{I} (a_i \% P) y^i \right] \% P$ by equivalence of modulus operation and use it instead of f_k.

Then server receives $f_k \% P$ from hospitals. To find a pairwise intersection between hospital k and k', server computes a pairwise sum of polynomials as $[f_k \% P + f_{k'} \% P] \% P = [f_k + f_{k'}] \% P$, which refers to the polynomial for intersection between hospital k and k'. Server repeats this procedure for every pair of k and k' ($k' \neq k$), and send the $K - 1$ polynomials to each hospital. Hospital k then checks whether its element $y_k \in Y_k$ is in the pairwise

intersection of hospital k and other hospital k', that is, if $[f_k(y_k) + f_{k'}(y_k)]\% P = 0$, then $y_k \in Y_k \cap Y_{k'}$ by Lemma 4.2.

By combining all the pairwise intersection with $K - 1$ hospitals, hospital k checks whether the element y_k is in the intersection of all the K hospitals. For example, combining the pairwise intersection of $f_1(y_1) + f_2(y_1) = 0$ (i.e., $y_1 \in Y_1 \cap Y_2$) and $f_1(y_1) + f_3(y_1) \neq 0$ (i.e., $y_1 \notin Y_1 \cap Y_3$) gives $y_1 \in Y_1 \cap Y_2 \cap Y_3^c$. After obtaining 2^{K-1} intersections with Y_k, hospital k sends the size of 2^{K-1} intersections to server. Server collects the size of intersection from all the K hospitals and obtain the size of all the $2^K - 1$ combinations of intersections (the number of combinations is two cases, whether in or out, for every Y_k except one case when $Y_1^c \cap \ldots \cap Y_K^c$). Finally, hospitals receive the size of $2^K - 1$ intersection, and align their elements according to the size information. Hospitals have the same order of these intersections such as $Y_1 \cap Y_2, Y_1 \cap Y_2^c, Y_1^c \cap Y_2$ (Fig. 3). The elements within the intersections are sorted. Thus, all hospitals have the same size and order of elements for every feature mode.

5 EXPERIMENTS

We evaluate Trip by measuring computational performance (accuracy and time) and deriving meaningful phenotypes. We compare Trip with two baselines:

i. **Central model:** Ordinary tensor factorization method for phenotyping. Regardless of privacy problem that concerns data sharing, this model runs on a central server where all the patient data are combined [28].

ii. **Local model:** We devise an intuitive local model, by which hospitals run the central model at their sides and send the final factor matrices of feature modes to server. Server averages the factor matrices and sends the averaged factor matrices back to hospitals without iterative updating like Trip. Because each column in factor matrices can represent different phenotypes over hospitals, before averaging the matrices, server sorts the columns of each hospital's factor matrix so that all hospitals have the same phenotypes at each column. For all feature modes n, server first chooses a pivot hospital k_p and computes cosine similarity between every pair of r_p th and r th column from factor matrix of hospital k_p and other hospitals k as $\text{similarity} = \cos(\mathbf{A}_{k_p}^{(n)}(:, r_p), \mathbf{A}_k^{(n)}(:, r))$ where $\forall k \neq k_p, \forall r \neq r_p$. Server then finds the most similar combination of r_p and r for all pairs. Finding the best combination that matches multiple items (columns) to multiple items can be solved in polynomial time by Hungarian method [20]. Finally, server changes the order of columns in $\mathbf{A}_k^{(n)}$ according to the combination so that each column from $\mathbf{A}_{k_p}^{(n)}$ and $\mathbf{A}_k^{(n)}$ refer to the same phenotype.

5.1 Accuracy and Time

We use a large publicly available dataset MIMIC-III containing de-identified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012 [26]. MIMIC-III includes

information such as demographics, laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality. We construct a 3-order tensor with patients, laboratory test results, and medication. The tensor value is the number of co-occurrences of abnormal lab results and medication from the same patient within specific time window. We generate four datasets as setting the time window as 3 hours, 6 hours, 1 day, or 7 days, and have the number of nonzero values of around 15 million (M), 25M, 40M, and 50M, respectively. The size for 7-day-window tensor (MIMIC-III 50M) is 38,035 patients by 3,229 medications by 304 lab results. Because duplicated co-occurrence can be counted with large time window, we set the maximum value of count as three, which is a median of 1-day-window tensor (MIMIC-III 40M). The count value larger than three is truncated to three.

We evaluate accuracy and time of Trip compared to two baseline models by varying the number of nonzero values, hospitals, and skewness (for unevenly distributed patients). We measure accuracy using root mean square error (RMSE) between the factorized tensor and the observed tensor. We also measure time elapsed by adding time for computation, communication, and alignment. Because Trip and local model distribute computation of local tensors to hospitals, we consider the computation time on the local tensors as the largest computation time on one hospital. The communication time is measured as the total number of communicated bytes between server and hospitals divided by data transfer rate of 15 MB/sec. The communication time for central model is time for transferring the local patient data to server. We repeat the evaluation ten times and average them. We run the models until it converges before maximum iteration 100. The rank is set to ten. λ is set to 10^{-2} after trying 10^{-3}, 10^{-2}, 10^{-1}, 1, and 10.

5.1.1 Number of nonzeros—We use the four MIMIC-III datasets that have 15M, 25M, 40M, and 50M nonzero values. We assume that MIMIC-III datasets are distributed in three hospitals, on which the patients are randomly distributed as the same size. As a result, Trip has low RMSE as much as central model and resembles central model for all the four datasets (Fig. 4a). For MIMIC-III 15M, RMSE from central model converges to 1.4404. Similarly, the RMSE from Trip starts to stable at around 50 iterations and converges to 1.4409 (Fig. 4c). Both of the RMSE from central model and Trip are significantly smaller than that of local model (1.5957). MIMICIII 50M dataset also shows similar convergence. RMSE from Trip starts to stable at around 60 iterations and converges to 1.8482, which is also similar to RMSE from central model, 1.8479 (Fig. 4d). MIMIC-III 25M shows the RMSE of 1.4955 from Trip, 1.4947 from central model, 1.6867 from local model. MIMIC-III 40M shows the RMSE of 1.7913 from Trip, 1.7903 from central model, 2.0037 from local model. Convergence results on MIMIC-III 25M and 40M can be found in our technical report [17].

In addition, total time elapsed for Trip (and local model) is much faster (half less) than central model in all datasets (Fig. 4b). Trip reduces computation time by distributing updating procedures to de-centralized hospitals; consequently, Trip reduces total time elapsed although sacrificing communication and alignment time. For the datasets of 15M, 25M, 40M, and 50M, computation time from Trip is 3,152, 4,183, 5,796, and 7,125 seconds, and computation time from central model is 5,266, 8,068, 11,661, and 15,105 seconds, which take majority of total time. Communication time from Trip is around 100.7 seconds

for all the cases, and communication time from central model is 30.1, 53.9, 85.0, and 114.4 seconds. Note that Trip saves not only computation time but also communication time with large dataset (MIMIC-III 50M). Alignment time for Trip takes 22.6, 26.1, 29.3, and 59.0 seconds, which is negligible compared to computation time. Based on those observations, we can see that Trip efficiently resembles central model without no cost for privacy even at reduced time owing to distributed computation.

5.1.2 Number of hospitals—Using MIMIC-III 15M dataset, we partition the patients evenly into one to five hospitals. RMSE with one hospital refers to RMSE of central model. We observed that RMSE of Trip is stable as the number of hospitals increases, and is similar to RMSE of central model 1.4404, whereas RMSE of local model increase (Fig. 5a) with large variance. That is, compared to local model, in which local factorized tensors are diverged, Trip is robust on the finely split data. It means that phenotypes from Trip are accurate and not biased even with many small sized patient data.

Total time of Trip and local model are significantly faster than that of central model. As the number of hospitals increases and the patient data are spread more, the total time of Trip and local model decrease (Fig. 5c). Specifically, computation time for Trip and local model decrease because more hospitals distribute the computation, and communication time for Trip slightly increases, whereas communication time for local model is negligibly short. Alignment time is negligible for both Trip and local model.

5.1.3 Skewness—We partition the patients in MIMIC-III 15M unevenly in three hospitals. One hospital takes 1/3 (evenly distributed), 0.5, 0.7, and 0.9 of patients, and the other two hospitals take the remaining patients evenly. Note that elements in feature mode are still overlapped enough among hospitals. We observed that RMSE of Trip is stable although patients are distributed unevenly, whereas RMSE of local model is higher than that of Trip with large variance. Factorized tensor of local model can be inaccurate because the local factorized tensor from a small sized hospital can be biased and far different from others' results. However, the hospital can benefit from Trip by overcoming this bias and producing a generalized results.

Total time of Trip and local model increase (Fig. 5d) as the skewness increases. Time for computation increases because computational overhead occurs on one hospital with large data, and time for communication and alignment does not increase significantly.

5.2 Phenotype discovery

We use de-identified EHRs dataset from University of California, San Diego (UCSD) Medical Center with 8,022 patients by 748 medications by 299 diagnoses. Specifically, it is from two hospitals that have 4,703 patients (UCSD1) and 3,319 patients (UCSD2). We construct a 3-order tensor with patient, medication, and diagnosis mode with around 1.6 million of non-zero elements. The value of tensor is the number of co-occurrences of medication and diagnosis event from the same patient at the same visit.

We discover phenotypes from Trip and compare them with phenotypes from central model and individual central model of two hospitals in UCSD (i.e., run central model

independently at UCSD1 and UCSD2). λ is set 1 to derive more distinct phenotypes than those from MIMIC-III. A domain expert summarizes the factorized tensor into clinically meaningful phenotypes. The phenotypes consist of set of diagnoses and its corresponding medications. Due to limited space, medication factors in phenotypes are omitted and can be found in our project website [17].

As a result, Trip discovers unbiased and hidden phenotypes compared to the phenotypes from two individual central models (UCSD1, UCSD2). The phenotypes from Trip contain top-ranked phenotypes from UCSD1 and UCSD2, and are similar to phenotypes from combined central model, UCSD1+UCSD2 (Table 2). The phenotypes from Trip consist of top five phenotypes from UCSD1 and top four from UCSD2. The phenotypes from Trip are also the same with phenotypes from central model except gastrointestinal complaints and neurogenic bladder. Without our federated model, the two individual hospitals could derive biased phenotypes that are only fitted to the local data. It means that Trip can effectively resemble central model without cost for privacy.

In addition, Trip discovers a new phenotype, *sickle cell/chronic pain crisis*, that is contained in neither of UCSD1 and UCSD2. This phenotype consists of diagnoses related to *sickle cell diseases* or *chronic pain crisis* and corresponding medications (Table 3). Based on physician's judgement, this phenotype is clinically meaningful in that sickle cell disease usually accompanies chronic pain such as constipation, back/neck pain, headache, (pruritic disorder, insomnia, and wheezing. *sickle cell/chronic pain crisis* is not dominant in individual hospital but is dominant in overall perspective. Note that RMSE of Trip is low as much as RMSE of central model while reducing total time (Table 4), and RMSEs of two individual UCSD datasets are lower than others because those two use separated small local datasets. Also, note that communication time of the central model is due to transferring the data.

6 CONCLUSIONS

This paper presents Trip, a federated tensor factorization for computational phenotyping without sharing patient-level data. We developed secure data harmonization and privacy-preserving computation procedures based on ADMM, and analyzed that Trip ensure the confidentiality of patient-level data. Experimental results on data from MIMIC-III and UCSD medical center demonstrated that our framework resembles the central model very well. Trip is also accurate even with small or skewly distributed patient data, and fast compared to the central model. We also showed that Trip discovers phenotypes as the central model with combined patient data does, which are unbiased or not discovered (hidden) phenotypes from each hospital. As a result, Trip can help derive useful phenotypes from EHR data to overcome policy barriers due to the privacy concerns. We plan to apply it to much larger scale datasets in the future and facilitate the discovery of novel and important "phenotypes" to support clinical research and precision medicine.

Acknowledgments

References

1. Bennett C, Doub T, Selove R. EHRs connect research and practice: Where predictive modeling, artificial intelligence, and clinical decision support intersect. Health Policy and Technology. Jun; 2012 1(2):105–114.

2. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning. 2011; 3(1):1–122.

3. Carroll JD, Chang J-J. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. Psychometrika. 1970; 35(3):283–319.

4. Carroll, RJ., Eyler, AE., Denny, JC. Naïve electronic health record phenotype identification for rheumatoid arthritis. Proceedings of the American Medical Informatics Association Annual Symposium; Jan. 2011; p. 189-196.

5. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, Pacheco JA, Boomershine CS, Lasko TA, Xu H, Karlson EW, Perez RG, Gainer VS, Murphy SN, Ruderman EM, Pope RM, Plenge RM, Kho AN, Liao KP, Denny JC. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. Journal of the American Medical Informatics Association. Jun; 2012 19(1e):e162–e169. [PubMed: 22374935]

6. Chaudhuri K, Monteleoni C, Sarwate A. Differentially Private Empirical Risk Minimization. Journal of Machine Learning Research (JMLR). Jul.2011 12:1069–1109. [PubMed: 21892342]

7. Choi JH, Vishwanathan S. Dfacto: Distributed factorization of tensors. Advances in Neural Information Processing Systems. 2014:1296–1304.

8. Cimino JJ. From data to knowledge through concept-oriented terminologies: Experience with the Medical Entities Dictionary. Journal of the American Medical Informatics Association. May; 2000 7(3):288–297. [PubMed: 10833166]

9. De Almeida, AL., Kibangou, AY. Distributed large-scale tensor decomposition. Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on; IEEE; 2014. p. 26-30.

10. Dwork C. Differential privacy. International Colloquium on Automata, Languages and Programming. 2006; 4052(d):1–12.

11. Ebadollahi S, Sun J, Gotz D, Hu J, Sow D, Neti C. Predicting patient's trajectory of physiological data using temporal trends in similar patients: A system for near-term prognostics. AMIA Annual Symposium Proceedings. 2010; 2010:192–196. [PubMed: 21346967]

12. Greengard S. A new model for healthcare. Communications of the ACM. 2013; 56(2):17– 19.

13. Ho JC, Ghosh J, Steinhubl SR, Stewart WF, Denny JC, Malin BA, Sun J. Limestone: High-throughput candidate phenotype generation via tensor factorization. Journal of biomedical informatics. 2014; 52:199–211. [PubMed: 25038555]

14. Ho, JC., Ghosh, J., Sun, J. Marble: High-throughput Phenotyping from Electronic Health Records via Sparse Nonnegative Tensor Factorization. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14; New York, NY, USA. ACM; 2014. p. 115-124.

15. Hong M, Luo Z-Q, Razaviyayn M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. SIAM Journal on Optimization. 2016; 26(1):337–364.

16. Kang, U., Papalexakis, E., Harpale, A., Faloutsos, C. Gigatensor: scaling tensor analysis up by 100 times-algorithms and discoveries. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining; ACM; 2012. p. 316-324.

17. Kim, Y., Sun, J., Yu, H., Jiang, X. Federated Tensor Factorization for Computational Phenotyping. 2016. http://dm.postech.ac.kr/trip
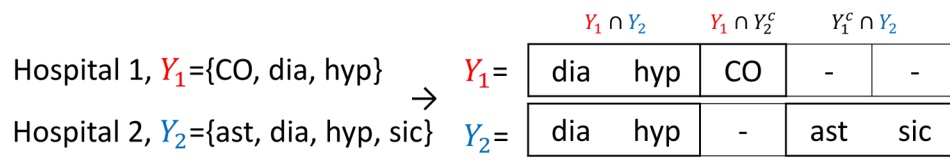
18. Kissner, L., Song, D. Advances in Cryptology–CRYPTO 2005. Springer; 2005. Privacy-preserving set operations; p. 241-257.

19. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in Medicine. Aug; 2001 23(1):89–109. [PubMed: 11470218]

20. Kuhn HW. The hungarian method for the assignment problem. Naval research logistics quarterly. 1955; 2(1–2):83–97.

21. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. M.D. computing: computers in medical practice. Sep; 1991 8(5):300–315. [PubMed: 1749340]

22. Li Y, Jiang X, Wang S, Xiong H, Ohno-Machado L. Vertical grid logistic regression (vertigo). Journal of the American Medical Informatics Association. 2015:ocv146.

23. Lin Z, Liu R, Su Z. Linearized alternating direction method with adaptive penalty for low-rank representation. Advances in neural information processing systems. 2011:612–620.

24. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association. 2010; 17(2):124–130. [PubMed: 20190053]

25. PCORnet Data Privacy Task Force. Technical Approaches for Protecting Privacy in the PCORnet Distributed Research. 2015.

26. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L-W, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG. Multiparameter intelligent monitoring in intensive care ii (mimic-ii): A public-access intensive care unit database. Critical Care Medicine. May.2011 39:952–960. [PubMed: 21283005]

27. Vaidya J, Yu H, Jiang X. Privacy-preserving svm classification. Knowledge and Information Systems. 2008; 14(2):161–178.

28. Wang, Y., Chen, R., Ghosh, J., Denny, JC., Kho, A., Chen, Y., Malin, BA., Sun, J. Rubik: Knowledge guided tensor factorization and completion for health data analytics. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM; 2015. p. 1265-1274.
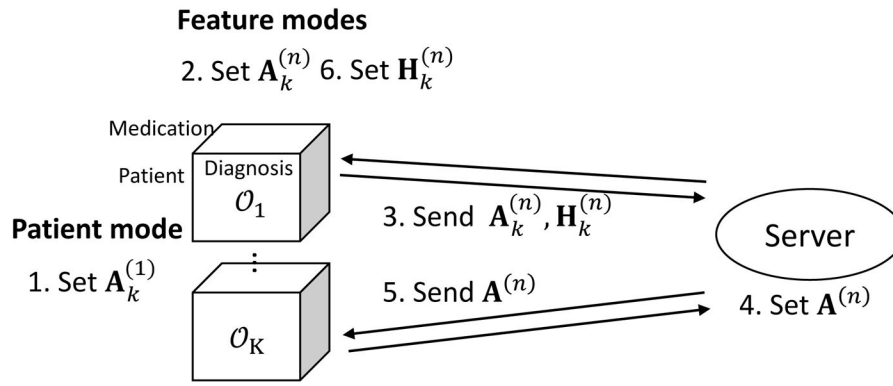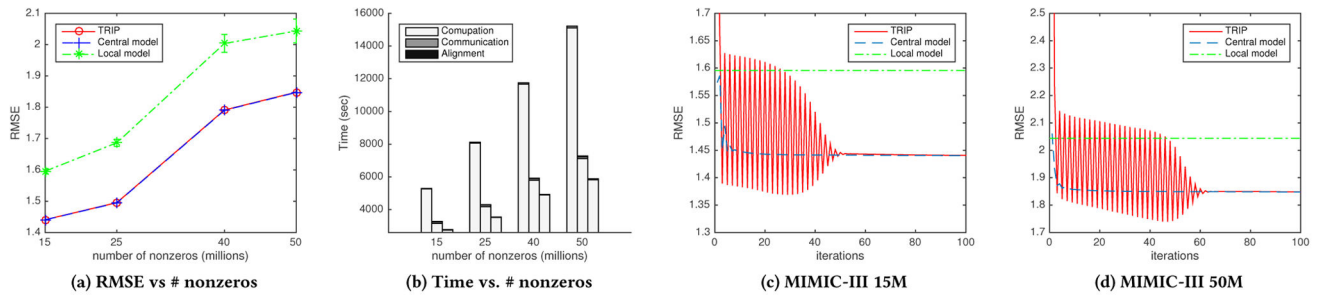
Figure 1.
Process of federated tensor factorization.
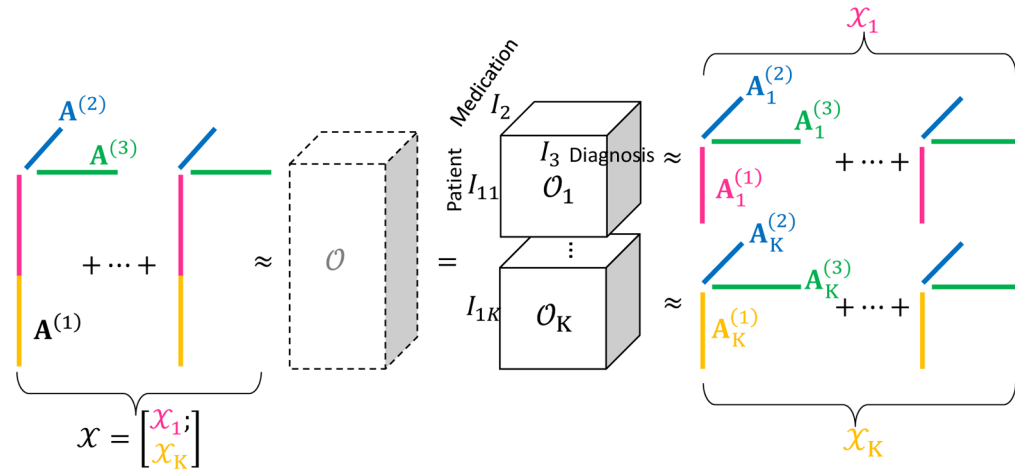
Figure 2.
Equivalence between tensor factorization with respect to each local tensor $\mathcal{O}_k$ and tensor factorization with respect to global tensor $\mathcal{O}$. Without $\mathcal{O}$, tensor factorization that is globally optimal across hospitals can be achieved via local tensor factorization.

	$Y_1 \cap Y_2$		$Y_1 \cap Y_2^c$	$Y_1^c \cap Y_2$	
Hospital 1, Y_1={CO, dia, hyp} → Y_1=	dia	hyp	CO	-	-
Hospital 2, Y_2={ast, dia, hyp, sic} Y_2=	dia	hyp	-	ast	sic

Figure 3.
Example of secure alignment on feature mode.

Figure 4.
RMSE and total time over the number of nonzeros (Fig. 4a, 4b). The first, second, and third stacked bars in Fig. 4b refer to central model, Trip, and local model, respectively. RMSE of Trip, central model, and local model over iteration (Fig. 4c, 4d).
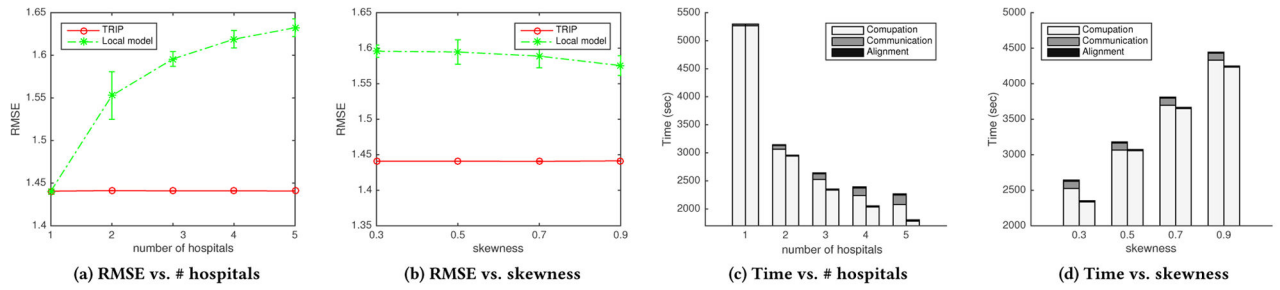
Figure 5.
RMSE over the number of hospitals (Fig. 5a) and skewness (Fig. 5b). Total time over the number of hospitals (Fig. 5c) and skewness (Fig. 5d). The former and latter stacked bars in Fig. 5c, 5d refer to Trip and local model, respectively.

Table 1

Notations and symbols

∘	outer product
⊙	Khatri-Rao product
R	number of ranks
N	number of modes (order)
K	number of hospitals
A, **B**	matrix
$\mathscr{X}, \mathscr{O}$	tensor
$\mathbf{O}_{(n)}$	matricized tensor of $\mathscr{O}$ on nth mode

Table 2

Phenotypes from Trip, central model on UCSD1+UCSD2, UCSD1, and UCSD2. Some phenotypes appear in both UCSD hospitals

Rank	TRIP	UCSD1+UCSD2	UCSD1	UCSD2
1				
2				
3				
4				
5				Diabetes
6			Non-specific complaints	
7			COPD w/o exacerbation	Anxiety with gastrointestinal complaints
8		Non-specific complaints	Acute on chronic pain	Cystic fibrosis with pneumonia
9	Sickle cell/chronic pain crisis	Sickle cell/chronic pain crisis	COPD with Pneumonia	Non-specific complaints
10	Non-specific complaints		Anxiety with hypertension	Lymphoma

Table 3

Detailed phenotype of *sickle cell/chronic pain crisis*

Diagnosis
Sickle cell disease NOS, Hb-SS disease with crisis, Constipation NOS, Pruritic disorder NOS, Generalized pain, Headache, Insomnia, Chronic pain syndrome, Wheezing
Medication
Hydroxyurea, Deferasirox, Docusate, Diphenhydramine, Hydromorphone, Acetaminophen, Zolpidem, Folic Acid, Baclofen

Table 4

RMSE and time [sec] elapsed. Central model is run on UCSD1,2, UCSD1, UCSD2.

	Trip	**UCSD1,2**	**UCSD1**	**UCSD2**
RMSE	1.2304	1.2327	1.2267	1.1778
Computation	446.5	656.9	432.8	314.6
Communication	15.1196	3.4533	0	0
Alignment	1.2052	0	0	0