Tensor-Factorization-Based Phenotyping using Group Information: Case Study on the Efficacy of Statins

Jingyun Choi
Department of Computer Science and Engineering
Pohang University of Science and Technology
Pohang, Republic of Korea
jingyunc@postech.ac.kr

Yejin Kim
Department of Creative IT Engineering
Pohang University of Science and Technology
Pohang, Republic of Korea
yejin89@postech.ac.kr

Hun-Sung Kim
Department of Medical Informatics, College of Medicine
The Catholic University of Korea
Seoul, Republic of Korea
cadiz74@catholic.ac.kr

In Young Choi*
Department of Medical Informatics, College of Medicine
The Catholic University of Korea
Seoul, Republic of Korea
iychoi@catholic.ac.kr

Hwanjo Yu†
Department of Computer Science and Engineering
Pohang University of Science and Technology
Pohang, Republic of Korea
hwanjoyu@postech.ac.kr

ABSTRACT

To automatically extract medical concepts from raw electronic health records (EHRs), several applications based on machine learning techniques have been proposed. Among the various techniques, tensor factorization methods have attracted considerable attention because tensor representations can capture interactions among high-dimensional EHRs. Most of the existing tensor factorization methods for computational phenotyping are only designed to derive individual phenotypes that approximate the original data. However, deriving grouped phenotypes is desirable because patients form natural groups of interest (i.e., efficacy of treatment and disease categories). In this paper, we propose Supervised Non-negative Tensor Factorization with Multinomial Logistic Regression (SNTFL) to derive grouped phenotypes that are discriminative. We define a discriminative constraint to derive grouped phenotypes and jointly optimize a multinomial logistic regression during the tensor factorization process. Our case study on a hyperlipidemia dataset demonstrates that our proposed method obtains better discrimination on patient groups compared to the baselines and successfully discovers meaningful patient subgroups.

*Corresponding Author
†Corresponding Author

ACM-BCB '17, August 20-23, 2017, Boston, MA, USA

ACM ISBN 978-1-4503-4722-8/17/08…$15.00
https://doi.org/10.1145/3107411.3107423

CCS CONCEPTS

• **Applied computing** → **Health care information systems**; **Health informatics**; • **Computing methodologies** → *Factorization methods*;

KEYWORDS

Computational phenotyping, Joint learning, Representation learning

1 INTRODUCTION

Electronic health records (EHRs) include various types of information, such as patient demographics, medical histories, drug prescription histories, vital signs, immunizations, clinical notes, laboratory results and so on. In recent years, EHRs have become important sources of data, and many studies have used EHRs to solve healthcare problems, such as clinical decision making [29], predictive modeling [23, 35], patient management [15], and early disease detection [33]. However, the use of EHRs faces some challenges. First, EHRs contain heterogeneous (e.g., numerical laboratory results and clinical narrative notes), noisy, irregular and sparse data, which make analyses of EHRs difficult. The greatest challenge is that EHRs do not easily map to medical concepts that clinical researchers directly use or need. To address these challenges, we need to extract meaningful medical concepts that clinical researchers directly use from the raw EHR data. The process of transforming the raw EHR data into meaningful medical concepts (i.e., phenotypes) is called phenotyping. However, clinical researchers and domain experts must spend substantial time and effort to obtain specific medical concepts from the complicated EHRs [13, 26]. For example, in the eMERGE Network [26], collaborative efforts between clinical researchers and IT experts are required for phenotyping. To reduce this time and effort, computational phenotyping has been introduced and widely used in many studies [5, 11, 12, 14, 17, 18, 21, 24, 27, 31, 34, 36].

Computational phenotyping is the process of automatically extracting phenotypes from EHR data using machine learning techniques. Computational phenotyping uses various machine learning techniques, such as deep learning methods [5, 14, 18], probabilistic graphical models [27] and dimensionality reduction methods [11, 12, 17, 21, 31, 34, 36]. Deep learning methods are used to obtain phenotypes because such methods can obtain insights from the heterogeneous EHR data [5, 14, 18, 24]. For example, a modified deep learning method is applied to clinical time series data for discovering and detecting characteristic patterns of physiology [5]. A probabilistic graphical model is used to cluster patients' longitudinal trajectories [27]. Some studies have been conducted to develop phenotyping methods based on dimensionality reduction, such as matrix factorization [21, 36] and tensor factorization [11, 12, 17, 31, 34]. For example, PACIFIER [36], which densifies the extremely sparse EHR data by performing matrix completion, was proposed based on matrix factorization. However, matrix factorization has difficulty taking into account the multidimensional structure that would be lost when analyzing the multidimensional structure of the data by collapsing some of the modes [25]. In contrast to matrix factorization, tensor factorization can handle multidimensional structures and capture complex interactions among high-dimensional EHR data. Due to this advantage, computational phenotyping via tensor factorization has attracted considerable attention [11, 12, 17, 31, 34]. For example, Marble [12] extracts concise phenotypes and chronic disease phenotypes that represent the baseline characteristics of the overall population. Rubik [31] extracts phenotypes using additional auxiliary information, such as prior knowledge. These methods aim to derive "individual" phenotypes that approximate the original data as closely as possible.

However, it is desirable to derive "grouped" phenotypes because patients form natural groups of interest (i.e., efficacy of treatment and disease categories). For example, assume that patients belong to either group A, B, or C (i.e., high, moderate, and low efficacy of treatments, respectively). Each patient group has its own characteristics and consists of subgroups of its own patients. Therefore, we need to derive grouped phenotypes that have discriminative characteristics on each patient group rather than individual phenotypes. An example of grouped phenotypes is shown in Table 1. Phenotypes 1, 2, and 3 have discriminative characteristics on patient group A (i.e., high-efficacy group) compared to patient groups B and C (Table 1). This property also means that the high-efficacy group consists of subgroups that are represented as phenotypes 1, 2, and 3. Deriving grouped phenotypes rather than individual phenotypes is beneficial because such a group-based phenotyping can provide insights on subgroups within each patient group.

Table 1: Example of grouped phenotypes. Phenotypes 1, 2, and 3 have discriminative characteristics on patient group A compared to patient groups B and C.

Patient group A (e.g., high efficacy)	Patient group B (e.g., moderate efficacy)	Patient group C (e.g., low efficacy)
Phenotype 1	Phenotype 4	Phenotype 6
Phenotype 2	Phenotype 5	Phenotype 7
Phenotype 3		

Two methods have recently been proposed to derive discriminative phenotypes for better prediction performance [17, 34]. These methods can derive discriminative phenotypes when there are two patient groups. However, there is still a limitation: these methods cannot be used if there are more than three patient groups.

Therefore, this paper proposes Supervised Non-negative Tensor Factorization with Multinomial Logistic Regression (SNTFL) to derive grouped phenotypes that are discriminative regardless of the number of patient groups. Our main contributions are as follows:

(1) We propose a novel tensor factorization method, SNTFL, to derive grouped phenotypes that have discriminative characteristics on each patient group regardless of the number of patient groups.
(2) Our case study on a hyperlipidemia dataset demonstrates that our method discovers meaningful subgroups of patients.

The remainder of this paper is organized as follows. First, we introduce the preliminaries of tensor factorization and describe our proposed method. Next, we explain the data analysis procedure using SNTFL and evaluate our method on a hyperlipidemia dataset. Finally, we draw our conclusions.

2 PRELIMINARIES

We first present the preliminaries on tensors, and then we introduce tensor factorization and multinomial logistic regression.

2.1 Notation

Table 2: Notation

Symbol	Description
a	scalar
$\mathbf{a}$	vector (bold lowercase character)
$\mathbf{A}$	matrix (bold uppercase character)
$\mathcal{X}$	tensor
$\mathbf{X}_{(k)}$	k-mode matricization of a tensor
R	number of latent factors
K	number of modes
C	number of classes
$\circ$	outer product of vectors
$\otimes$	Kronecker product
$\odot$	Khatri-Rao product
$\mathbf{A}^T$	transpose of $\mathbf{A}$
$\langle \mathbf{A}, \mathbf{B} \rangle$	sum of element-wise products of two matrices, where $\mathbf{A}$ and $\mathbf{B}$ have the same size

We summarize the notations used in this paper (Table 2). We use bold lowercase characters for vectors, bold uppercase characters for matrices and bold script font for tensors. We represent the i-th row of matrix $\mathbf{A}$ as $\mathbf{A}_{i*}$, the j-th column as $\mathbf{A}_{*j}$, and the (i, j)-th element as $\mathbf{A}_{ij}$. We represent the k-th element of vector $\mathbf{a}$ as $\mathbf{a}_k$. The definitions for the tensor and algebraic operations are as follows:

Definition 1. *Tensor. A tensor is a multidimensional array that can represent multiple attributes. The order of a tensor is the number of dimensions, and each dimension of a tensor is a mode. Vectors*

(first-order tensors) and matrices (second-order tensors) are special cases of tensors. A K-order tensor is denoted by $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_K}$, where I_k is the size of mode k. A K-order tensor is a rank-one tensor if it can be defined as the outer product of K vectors.

Definition 2. *k-mode matricization. A K-order tensor $\mathcal{X}$ can be unfolded into a matrix by reordering the elements of $\mathcal{X}$. This process is called the k-mode matricization, also referred to as unfolding, of $\mathcal{X}$. The k-mode matricization is denoted by $\boldsymbol{X}_{(k)} \in \mathbb{R}^{I_k \times (\prod_{j \neq k} I_j)}$. For example, given a third-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, the 1-mode matricization is $\boldsymbol{X}_{(1)} \in \mathbb{R}^{I_1 \times (I_2 I_3)}$.*

Definition 3. *Kronecker product. Given $\boldsymbol{A} \in \mathbb{R}^{I \times J}$ and $\boldsymbol{B} \in \mathbb{R}^{K \times L}$, the Kronecker product generates a matrix of size $(IK) \times (JL)$.*

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B} & \cdots & \boldsymbol{A}_{1J}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{I1}\boldsymbol{B} & \cdots & \boldsymbol{A}_{IJ}\boldsymbol{B} \end{bmatrix}$$

Definition 4. *Khatri-Rao product. Given $\boldsymbol{A} \in \mathbb{R}^{I \times R}$ and $\boldsymbol{B} \in \mathbb{R}^{J \times R}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same number of columns, the Khatri-Rao product generates a matrix of size $(IJ) \times R$.*

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{*1} \otimes \boldsymbol{B}_{*1} & \cdots & \boldsymbol{A}_{*R} \otimes \boldsymbol{B}_{*R} \end{bmatrix}$$

2.2 Tensor factorization

Tensor factorization is a dimensionality reduction method that is widely used in various fields, such as signal processing [19], computer vision [30] and bioinformatics [9], because it can analyze complex data sets that contain high dimensionality. Tensor factorization aims to approximate the observed tensor using the low-dimensional space. This process is equivalent to minimizing the difference between an observed tensor and a reconstructed tensor from the factor matrices.

CANDECOMP/PARAFAC (CP) decomposition [4, 10] is one of the most common tensor factorizations. CP decomposition approximates the observed tensor as a sum of rank-one tensors. Given a K-order tensor $\mathcal{X}$ and rank R, which indicates the number of latent factors, CP decomposition factorizes the K-order tensor $\mathcal{X}$ into K factor matrices, $\mathbf{U}^{(1)} \in \mathbb{R}^{I_1 \times R}, \ldots, \mathbf{U}^{(K)} \in \mathbb{R}^{I_K \times R}$. This process can be expressed as follows: $\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{U}^{(1)}_{*r} \circ \cdots \circ \mathbf{U}^{(K)}_{*r}$. Each factor matrix $\mathbf{U}^{(k)} \in \mathbb{R}^{I_k \times R}$ represents the latent feature vectors for the k-th mode of the observed tensor $\mathcal{X}$. For example, given a third-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, CP decomposition finds three factor matrices, $\mathbf{U}^{(1)} \in \mathbb{R}^{I_1 \times R}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{I_2 \times R}$, and $\mathbf{U}^{(3)} \in \mathbb{R}^{I_3 \times R}$, that approximate $\mathcal{X}$ (Figure 1).

2.3 Non-negative tensor factorization

Non-negative tensor factorization (NTF) decomposes a non-negative tensor into factor matrices that consist of non-negative elements. Given rank R and a K-order tensor $\mathcal{X}$, NTF minimizes the following objective function:

$$\min_{\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(K)}} \left\| \mathcal{X} - \sum_{r=1}^{R} \mathbf{U}^{(1)}_{*r} \circ \mathbf{U}^{(2)}_{*r} \circ \cdots \circ \mathbf{U}^{(K)}_{*r} \right\|_F^2 \quad (1)$$
$$\text{s.t.} \quad \mathbf{U}^{(k)} \geq 0, \text{ for } k = 1, \ldots, K.$$

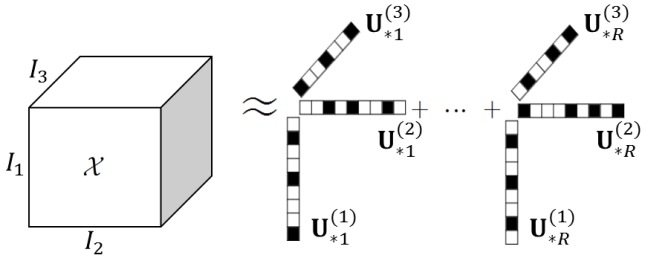


Figure 1: CP decomposition of a third-order tensor $\mathcal{X}$. The outer product of three vectors ($\mathbf{U}^{(1)}_{*r}$, $\mathbf{U}^{(2)}_{*r}$ and $\mathbf{U}^{(3)}_{*r}$) from factor matrices is a third-order rank-one tensor. The CP decomposition approximates $\mathcal{X}$ as a sum of R rank-one tensors.

This non-negativity constraint simplifies the interpretation of factor matrices when analyzing non-negative data [7].

2.4 Multinomial logistic regression

Multinomial logistic regression is an extension of binomial logistic regression for solving multiclass classification problems. In the multiclass classification problem, we are given N data points, $\{\mathbf{s}^1, \ldots, \mathbf{s}^N\}$, and their corresponding class labels, $\{y^1, \ldots, y^N\}$, $y^n \in \{1, \ldots, C\}$, where C is the total number of classes. Assume that each data point has d features and belongs to one group among C groups. The label information of the data points is represented by $\mathbf{Y} \in \mathbb{R}^{N \times C}$, which is an indicator matrix. $\mathbf{Y}_{nc}$ is set to 1 if the n-th data point belongs to the c-th group and to 0 otherwise. Let $\mathbf{W} \in \mathbb{R}^{d \times C}$ be the coefficient of the multinomial logistic regression. Given $\mathbf{W}$, the probability that the n-th data point $\mathbf{s}^n \in \mathbb{R}^{1 \times d}$ belongs to the c-th group is calculated as $\frac{exp(\mathbf{s}^n \mathbf{W}_{*c})}{\sum_{b=1}^{C} exp(\mathbf{s}^n \mathbf{W}_{*b})}$. $\mathbf{W}$ is estimated by minimizing the negative log-likelihood function:

$$-\sum_{n \in N} \sum_{c=1}^{C} \mathbf{Y}_{nc} log \frac{exp(\mathbf{s}^n \mathbf{W}_{*c})}{\sum_{b=1}^{C} exp(\mathbf{s}^n \mathbf{W}_{*b})}. \quad (2)$$

One of the most widely used methods for optimizing Eq. (2) is iterative reweighted least squares (IRLS) [22].

3 PROPOSED METHOD

We propose a novel tensor factorization method, SNTFL, to derive grouped phenotypes that are discriminative. We first i) formulate SNTFL by incorporating a discriminative constraint to derive grouped phenotypes. Then, we ii) jointly optimize a multinomial logistic regression during the tensor factorization process. Finally, we iii) obtain grouped phenotypes and latent representations of patients to the phenotypes.

3.1 Problem formulation

SNTFL is an extension of NTF for deriving grouped phenotypes that have discriminative characteristics on each patient group. Our objective is to improve the discrimination among patient groups; thus, we use the group information (i.e., labels) of patients to achieve this goal. First, given a K-order tensor $\mathcal{X}$, we assume that the first mode of $\mathcal{X}$ denotes the patient. Then, each row of the patient factor matrix $\mathbf{U}^{(1)}$ represents the latent representation of a patient. The

main idea is to exploit the group information of patients for making the rows of $\mathbf{U}^{(1)}$ discriminative on the labels. When approximating $\mathcal{X}$, factor matrices influence each other. If each row of $\mathbf{U}^{(1)}$ has discriminative characteristics, then the columns (i.e., the latent feature vectors) of other factor matrices also contain discriminative characteristics. These columns from all factor matrices except $\mathbf{U}^{(1)}$ are used as phenotypes such that we can derive grouped phenotypes that are discriminative. Therefore, we should minimize the difference between the observed K-order tensor and a reconstructed tensor from the factor matrices while simultaneously generating the discriminative latent representations of patients. We formulate the objective function of SNTFL as follows:

$$\min_{\mathbf{U}^{(1)},\ldots,\mathbf{U}^{(K)}} \frac{1}{2}\left\|\mathcal{X}-\sum_{r=1}^{R}\mathbf{U}^{(1)}_{*r}\circ\mathbf{U}^{(2)}_{*r}\circ\cdots\circ\mathbf{U}^{(K)}_{*r}\right\|_F^2+\lambda\Omega(\mathbf{U}^{(1)}) \quad \text{s.t.} \quad \mathbf{U}^{(k)}\geq 0, \text{ for } k=1,\ldots,K, \tag{3}$$

where $\Omega(\mathbf{U}^{(1)})$ is a discriminative constraint to derive grouped phenotypes that are discriminative and λ controls the weight of the discriminative constraint.

We can rewrite the objective function of SNTFL in terms of the unfolded tensor as

$$\min_{\mathbf{U}^{(k)}} \frac{1}{2}\left\|\mathbf{X}_{(k)}-\mathbf{U}^{(k)}\mathbf{G}^T\right\|_F^2+\lambda\Omega(\mathbf{U}^{(1)}) \quad \text{s.t.} \quad \mathbf{U}^{(k)}\geq 0, \text{ for } k=1,\ldots,K, \tag{4}$$

where $\mathbf{G}=\mathbf{U}^{(K)}\odot\cdots\odot\mathbf{U}^{(k+1)}\odot\mathbf{U}^{(k-1)}\odot\cdots\odot\mathbf{U}^{(1)}$.

Next, we define the discriminative constraint using a multinomial logistic regression. Assume that we have N patients and that each patient belongs to one group among C groups. The latent representation for the n-th patient is $\mathbf{U}^{(1)}_{n*}\in\mathbb{R}^{1\times R}$. To represent the group information of the n-th patient, we use an indicator matrix $\mathbf{Y}\in\mathbb{R}^{N\times C}$. $\mathbf{Y}_{nc}$ is set to 1 if the n-th patient belongs to the c-th group and to 0 otherwise. Let $\mathbf{P}\in\mathbb{R}^{N\times C}$ refer to probabilities, where each element denotes the probability that a certain patient belongs to one group among C groups. We incorporate the multinomial logistic regression in the tensor factorization process. Let $\mathbf{W}\in\mathbb{R}^{R\times C}$ be the coefficient of the multinomial logistic regression. Given $\mathbf{W}$, the probability that the n-th patient belongs to the c-th group is calculated as

$$\mathbf{P}_{nc}=\frac{exp(\mathbf{U}^{(1)}_{n*}\mathbf{W}_{*c})}{\sum_{b=1}^{C}exp(\mathbf{U}^{(1)}_{n*}\mathbf{W}_{*b})}. \tag{5}$$

Using Eq. (5), we can define the discriminative constraint as

$$\Omega(\mathbf{U}^{(1)})=-\sum_{n\in N}\sum_{c=1}^{C}\mathbf{Y}_{nc}log\mathbf{P}_{nc}. \tag{6}$$

Based on this discriminative constraint, the latent representations of patients should be more discriminative. Taken together, SNTFL minimizes the difference between the observed K-order tensor and the reconstructed tensor from the factor matrices; simultaneously, it minimizes the discriminative constraint to obtain the discriminative latent representations of patients and to improve discrimination among patient groups.

3.2 Joint optimization

We apply an alternating minimization approach to solve the optimization problem in Eq. (3). When solving a subproblem to update the k-th factor matrix $\mathbf{U}^{(k)}$, the other factor matrices are fixed. The objective function in Eq. (3) also contains a standard bound-constrained optimization problem. For this reason, we adopt the projected gradient method [20] to obtain factor matrices that consist of non-negative elements.

For the patient mode, we optimize each row of $\mathbf{U}^{(1)}$ (Algorithm 1) because each latent representation corresponds to the different group information. The gradient of the objective function with respect to $\mathbf{U}^{(1)}_{n*}$ is

$$\nabla_{\mathbf{U}^{(1)}_{n*}}=-(\mathbf{X}_{(1)})_{n*}\mathbf{G}+\mathbf{U}^{(1)}_{n*}\mathbf{G}^T\mathbf{G}-\lambda\sum_{c=1}^{C}\mathbf{Y}_{nc}(\mathbf{W}^T_{*c}-\mathbf{P}_{n*}\mathbf{W}^T). \tag{7}$$

To update $\mathbf{U}^{(1)}_{n*}$, we adopt the projected gradient method:

$$\mathbf{U}^{(1)}_{n*}\leftarrow P[\mathbf{U}^{(1)}_{n*}-\alpha\nabla_{\mathbf{U}^{(1)}_{n*}}], \tag{8}$$

where $P[\mathbf{A}]$ is a function that projects all the negative elements of $\mathbf{A}\in\mathbb{R}^{I\times J}$ to zero with $P[\mathbf{A}]=\begin{cases}\mathbf{A}_{ij} & \text{if } \mathbf{A}_{ij}>0\\ 0 & \text{otherwise}\end{cases}$, $i\in\{1,\ldots,I\}$ and $j\in\{1,\ldots,J\}$. α is the step size for the update rule. We use the Armijo rule [3] to adaptively determine α, which is adjusted until the following condition is satisfied:

$$(1-\sigma)\langle\nabla_{\mathbf{U}^{(1)}_{n*}},\triangle_n\rangle+\frac{1}{2}\langle\triangle_n,\triangle_n\nabla^2_{\mathbf{U}^{(1)}_{n*}}\rangle\leq 0, \tag{9}$$

where $\nabla^2_{\mathbf{U}^{(1)}_{n*}}=\mathbf{G}^T\mathbf{G}-\lambda\sum_{c=1}^{C}\mathbf{Y}_{nc}(\mathbf{W}(\mathbf{P}^T_{n*}\mathbf{P}_{n*}-diag\{\mathbf{P}_{n*}\})\mathbf{W}^T)$ and $\triangle_n=P[\mathbf{U}^{(1)}_{n*}-\alpha\nabla_{\mathbf{U}^{(1)}_{n*}}]-\mathbf{U}^{(1)}_{n*}$. σ is the parameter of this condition. A common choice of σ is 0.01 [20].

Algorithm 1 Subproblem 1 solver to update $\mathbf{U}^{(1)}$

Input: $\mathbf{U}^{(1)}$, $\mathbf{G}$, maximum iterations l_{max}, weight λ, group information $\mathbf{Y}$
Output: $\mathbf{U}^{(1)}$

```
for l = 1 : l_max do
    if convergence then
        break
    end if
    for n = 1 : N do
        Set ∇_{U^(1)_{n*}} using Eq. (7)
        U^(1)_{n*} ← P[U^(1)_{n*} − α∇_{U^(1)_{n*}}]
    end for
end for
return U^(1)
```

If $k\neq 1$, then the gradient of the objective function with respect to $\mathbf{U}^{(k)}$ is

$$\nabla_{\mathbf{U}^{(k)}}=-\mathbf{X}_{(k)}\mathbf{G}+\mathbf{U}^{(k)}\mathbf{G}^T\mathbf{G}. \tag{10}$$

With the gradient $\nabla_{\mathbf{U}^{(k)}}$, $\mathbf{U}^{(k)}$ is updated (Algorithm 2) as

$$\mathbf{U}^{(k)}\leftarrow P[\mathbf{U}^{(k)}-\alpha\nabla_{\mathbf{U}^{(k)}}], \tag{11}$$

where $P[\mathbf{A}]$ is a function that projects all the negative elements of $\mathbf{A} \in \mathbb{R}^{I \times J}$ to zero with $P[\mathbf{A}] = \begin{cases} \mathbf{A}_{ij} & \text{if } \mathbf{A}_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$, $i \in \{1, \ldots, I\}$ and $j \in \{1, \ldots, J\}$. α is the step size. The condition of the Armijo rule for selecting α is

$$(1-\sigma)\langle \nabla_{\mathbf{U}^{(K)}}, \triangle \rangle + \frac{1}{2}\langle \triangle, \triangle \mathbf{G}^T \mathbf{G} \rangle \leq 0, \tag{12}$$

where $\triangle = P[\mathbf{U}^{(k)} - \alpha \nabla_{\mathbf{U}^{(k)}}] - \mathbf{U}^{(k)}$.

Algorithm 2 Subproblem 2 solver to update $\mathbf{U}^{(k)}$ $(k \neq 1)$

Input: $\mathbf{U}^{(k)}$, $\mathbf{G}$, maximum iterations l_{max}
Output: $\mathbf{U}^{(k)}$

```
for l = 1 : l_max do
    if convergence then
        break
    end if
    Set ∇_{U^(k)} using Eq. (10)
    U^(k) ← P[U^(k) − α∇_{U^(k)}]
end for
return U^(k)
```

The entire algorithm for SNTFL is presented in Algorithm 3. SNTFL takes four inputs: a K-order tensor $\mathcal{X}$, rank R, weight λ of the discriminative constraint and group information $\mathbf{Y}$ of patients. The algorithm starts by initializing the factor matrices with non-negative values. After initialization, a subproblem to update the k-th factor matrix $\mathbf{U}^{(k)}$ is solved with the other fixed factor matrices. In the case of the patient mode, the first mode of $\mathcal{X}$, Algorithm 1 is used to update $\mathbf{U}^{(1)}$ due to the discriminative constraint. Otherwise, Algorithm 2 is used. After updating $\mathbf{U}^{(1)}$, a common multinomial logistic regression is optimized to update the coefficient $\mathbf{W}$. We use IRLS [22] to optimize the multinomial logistic regression. Our proposed alternating optimization algorithm using the projected gradient method decreases the value of the objective function in Eq. (3) until convergence is reached.

Algorithm 3 Optimization algorithm for SNTFL

Input: tensor $\mathcal{X}$, rank R, weight λ, group information $\mathbf{Y}$
Output: $\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(K)}$

```
Initialize U^(1), ..., U^(K) with non-negative values
repeat
    for k = 1 : K do
        G = U^(K) ⊙ ··· ⊙ U^(k+1) ⊙ U^(k−1) ⊙ ··· ⊙ U^(1)
        if k = 1 then
            U^(1) ← solve subproblem 1 using Algorithm 1
            Update W for multinomial logistic regression
        else
            U^(k) ← solve subproblem 2 using Algorithm 2
        end if
    end for
until convergence
return U^(1), ..., U^(K)
```

3.3 Phenotype discovery using SNTFL

For our case study, we use a hyperlipidemia dataset from EHRs from Seoul St. Mary's Hospital [16]. Each patient is classified into three groups according to the efficacy of a statin (i.e., high-efficacy group, moderate-efficacy group and low-efficacy group). This group information of patients is used for the discriminative constraint. We construct a third-order tensor with patient, trait (i.e., patient's demographics and past medical history before taking a statin) and statin modes. The tensor values are the numbers of co-occurrences between trait and statin for each patient. The constructed tensor is decomposed using SNTFL (Figure 2).

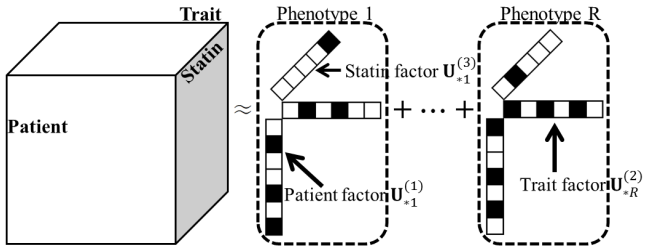


Figure 2: Generating phenotypes using SNTFL. Each phenotype consists of the patient factor $\mathbf{U}^{(1)}_{*r}$, the trait factor $\mathbf{U}^{(2)}_{*r}$ and the statin factor $\mathbf{U}^{(3)}_{*r}$.

We define a phenotype as a set of traits and an associated statin that can occur together in a patient. $\mathbf{U}^{(2)}_{*r}$ represents how much the certain traits are involved in the r-th phenotype. $\mathbf{U}^{(3)}_{*r}$ represents the certain statin that is involved in the r-th phenotype. For example, the r-th phenotype is defined using $\mathbf{U}^{(2)}_{*r}$ and $\mathbf{U}^{(3)}_{*r}$. The patient factor matrix $\mathbf{U}^{(1)}$ is used as the latent representation of patients. Each row of $\mathbf{U}^{(1)}$ expresses the latent representation of a patient. $\mathbf{U}^{(1)}_{n*}$ represents the degree to which the n-th patient is involved in R phenotypes. After decomposing the observed tensor, a multinomial logistic regression is trained on $\mathbf{U}^{(1)}$ to define grouped phenotypes. According to the magnitudes of the coefficients in the multinomial logistic regression, the phenotypes of each group are defined. We use MATLAB software and MATLAB Tensor Toolbox Version 2.5 [2] from Sandia National Laboratories to represent tensors and to compute tensor operations.

4 EXPERIMENTAL DESIGN

4.1 Dataset and preprocessing

We used a hyperlipidemia dataset from EHRs from Seoul St. Mary's Hospital between January 1, 2010, and December 31, 2012 [16]. This dataset includes patient demographics, past diagnoses and prescription histories before taking a statin, lab test results at the first and second visits (3 to 6 months) for statin prescription, and statins. We selected 4,970 patients who took one statin from among 17 statins and who had low-density lipoprotein cholesterol (LDL-C) levels at the first visit and second visit. We calculated the percent reduction in LDL-C (△LDL-C) between the first and second visits using $100\% \times \frac{(first\ visit - second\ visit)}{first\ visit}$. The patients belonged

to one of three groups according to the efficacy of a statin: high-efficacy (△LDL-C>50%), moderate-efficacy (10%≤△LDL-C≤50%), and low-efficacy (△LDL-C<10%) groups. The proportion of each group was imbalanced – 25.94% (1,289), 53.14% (2,641) and 20.92% (1,040) for each group, respectively; the majority of patients belonged to the moderate-efficacy group. To balance the number of patients in each group, we randomly selected the same number of patients from each group (i.e., 1,040 patients from each group, for 3,120 patients in total). We represented the data as the number of co-occurrences between trait (i.e., patient's demographics and past diagnoses/prescription histories before taking a statin) and statin of each patient. This co-occurrence is a natural representation to describe the interaction between a trait and statin. For example, patient 1, who received atorvastatin (10 mg), was a woman in her 50s and diagnosed with diabetes mellitus prior to taking a statin (Figure 3). We binned numerical attributes (e.g., LDL-C and age) and transformed it to a binary type (0 for absence and 1 for presence). Then, the trait consisted of 144 attributes: gender (2), binned age group (7), diagnosis histories (17), prescription histories (7) and binned lab test results (111). Thus, the constructed tensor size for the analysis was 3,120 patients by 144 traits by 17 statins.

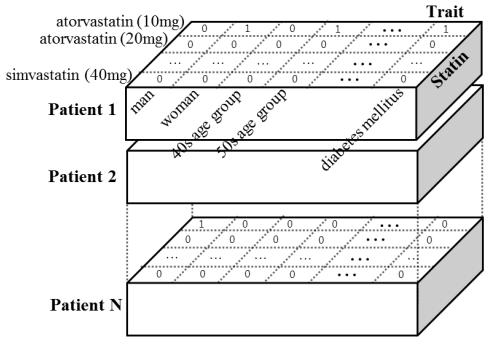


Figure 3: Constructed tensor from EHR data.

4.2 Analysis procedure

We summarize the design of our experiment in Figure 4. We used the constructed tensor that had a size of 3,120 patients by 144 traits by 17 statins (step 1), derived phenotypes using SNTFL with the training set (step 2), generated latent representations of new patients using the test set (step 3), and predicted the new patient's group using a multinomial logistic regression (step 4). We present the phenotype discovery (analysis 1) and prediction evaluation (analysis 2) results.

Specifically, to generate a new patient's latent representation (i.e., each row of new patient factor matrix $\tilde{\mathbf{U}}^{(1)}$), we projected the new patient's data $\tilde{\mathcal{X}}$ onto the space of derived phenotypes, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$. Given $\tilde{\mathcal{X}}$, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$, we calculated the $\tilde{\mathbf{U}}^{(1)}$ that best approximated $\tilde{\mathcal{X}} \approx \sum_{r=1}^{R} \tilde{\mathbf{U}}^{(1)}_{*r} \circ \mathbf{U}^{(2)}_{*r} \circ \mathbf{U}^{(3)}_{*r}$. Finally, we used the rows of $\tilde{\mathbf{U}}^{(1)}$ as features to predict the new patient's group (step 4-2). The prediction model was trained on $\mathbf{U}^{(1)}$, which was generated from the training set (step 4-1).

4.3 Measures and baselines

We used two types of area under the classification receiver operating characteristic curves (AUCs) as metrics to evaluate the prediction results of clinical diagnosis: MacroAUC, which is the average AUC of all the classes, and MicroAUC, which is the globally calculated AUC regardless of classes. We compared our proposed method with the following three baseline methods.

(1) NTF: Basic CP model [4, 10].
(2) CP-APR: Non-negative CP alternating Poisson regression model [6].
(3) SNTFL ($\lambda = 0$) : SNTFL with $\lambda = 0$.

All the methods were repeated ten times for stratified random sampling. We randomly selected 2,184 of all patients as the training set (70%) and the remainder as the test set (30%). We report the average Macro AUC and Micro AUC after ten repetitions.

5 RESULTS AND DISCUSSION

5.1 Tuning the weight of discriminative constraint

We studied the effect of λ to determine how it affected the prediction result of SNTFL. We first varied the number R of phenotypes from 30 to 70, and then we performed SNTFL with various values of λ. For each R, we varied λ from 0 to 0.006 to evaluate the sensitivity of SNTFL to the discriminative constraint. We calculated the average MicroAUC over 10 trials (Figure 5). The average MicroAUC first increased, reached its maximum value when $\lambda = 0.005$ and then decreased. When R was 70, the average MicroAUC was the greatest with $\lambda = 0.005$. The best average MicroAUC was 0.7592 when $R = 70$ and $\lambda = 0.005$.

5.2 Phenotype discovery

Using SNTFL with $\lambda = 0.005$ and $R = 50$, we derived 50 phenotypes. Among these phenotypes, we first filtered out phenotypes that were not statistically significant (i.e., p-values < 0.05 of a multinomial logistic regression). Then, a physician (HSK) removed phenotypes that were not clinically meaningful. To observe discriminative characteristics on each patient group and obtain insights on subgroups within each patient group, we present representative phenotypes from the 19 selected phenotypes (Figures 6, 7 and 8). For each patient group, three representative phenotypes with the largest coefficient $\mathbf{W}$ of the multinomial logistic regression are reported. We visualized the involvement of the traits and reported an associated statin on each phenotype. For the interpretability, we removed small non-zero values of the trait factor matrix of less than 0.01 and the statin factor matrix of less than 0.001.

For the high-efficacy group, the three phenotypes were as follows (Figure 6):

(1) "Men over 60 with metabolic diseases" (Figure 6a)
(2) "Women over 70 with metabolic diseases" (Figure 6b)
(3) "Mid-aged men with heart disease and prediabetes" (Figure 6c)

The first dominant phenotype in the high-efficacy group is the "men over 60 with metabolic diseases" phenotype (Figure 6a), which means a patient group in which male patients over 60 are diagnosed with metabolic diseases (i.e., diabetes mellitus, hypertension and

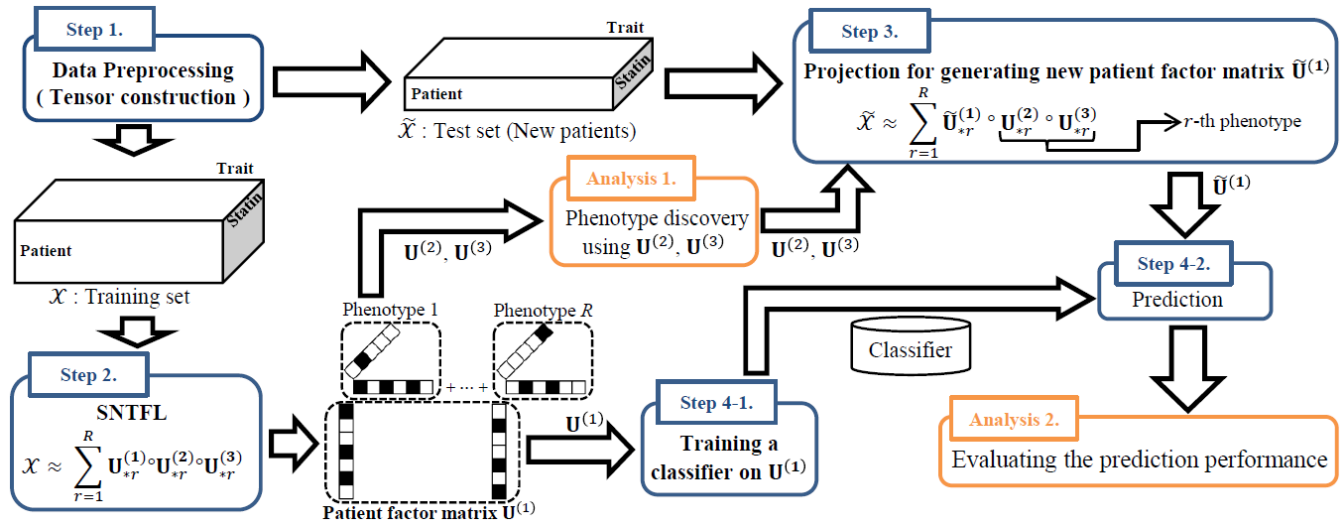


Figure 4: Design of the experiment. We first constructed a tensor (step 1), derived phenotypes using SNTFL (step 2), generated latent representations of new patients (step 3), and predicted the new patient's group using a multinomial logistic regression (step 4). We present the phenotype discovery (analysis 1) and prediction evaluation results (analysis 2).

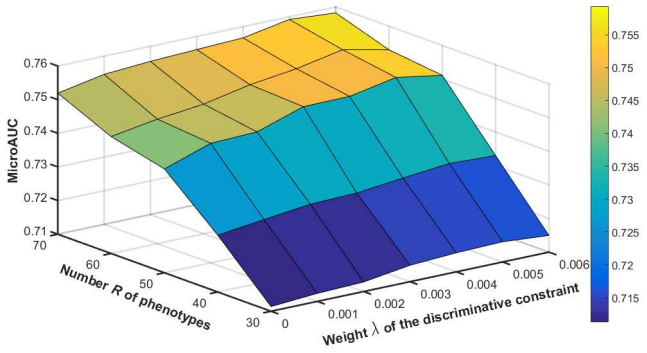


Figure 5: Effect of the weight λ on the average MicroAUC of SNTFL with R from 30 to 70 over 10 trials.

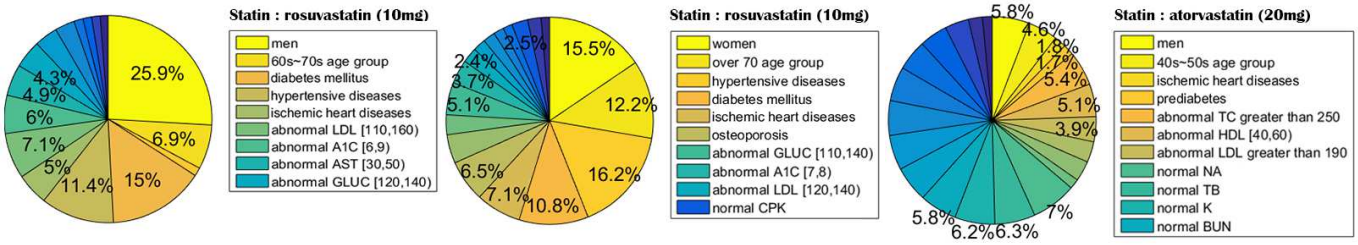
heart disease) before taking a statin, have abnormal lab results (i.e., LDL, A1C, AST and glucose) at the first visit, and take rosuvastatin (10 mg). The abnormal results of the A1C and glucose tests, which are used to screen for and diagnose diabetes [1], support the fact that the patient is diagnosed with diabetes prior to taking a statin. This finding is inconsistent with statin guidelines and statin therapy. According to the domain knowledge, rosuvastatin (10 mg) is known to reduce LDL-C levels by up to 50% [1, 28]. If we want to lower LDL-C levels by more than 50%, rosuvastatin (20~40 mg) is recommended for patients according to statin guidelines [1, 28]. However, patients who take high doses of statins may be more likely to develop kidney problems [8]. Therefore, this result is very valuable because we can recommend that elderly male patients with metabolic diseases take a relatively low dose of rosuvastatin when they are highly associated with this phenotype.

The second phenotype is the "women over 70 with metabolic diseases" phenotype (Figure 6b). This group consists of female patients over 70. Patients with metabolic diseases have abnormal levels in the A1C and glucose tests, which are associated with diabetes [1], and take rosuvastatin (10 mg). Similar to the "men over 60 with metabolic diseases" phenotype (Figure 6a), this result is also not consistent with statin therapy [1, 28]. Although patients take rosuvastatin (10 mg), their LDL-C levels decrease by more than 50%. From this finding, we can recommend that elderly female patients with metabolic diseases take rosuvastatin (10 mg) rather than rosuvastatin (20~40 mg) if they are highly associated with this result.

Third, the "mid-aged men with heart disease and prediabetes" phenotype (Figure 6c) represents a patient group that consists of mid-aged men with prediabetes and heart disease. After taking atorvastatin (20 mg), which is known to lower LDL-C levels by up to 50% [1, 28], their LDL-C levels decrease by more than 50%. In the case of atorvastatin, high doses (40~80 mg) are recommended to lower LDL-C levels by more than 50% according to statin guidelines [1, 28]. However, we can recommend that patients take a relatively low dose of atorvastatin when they are highly associated with this phenotype.

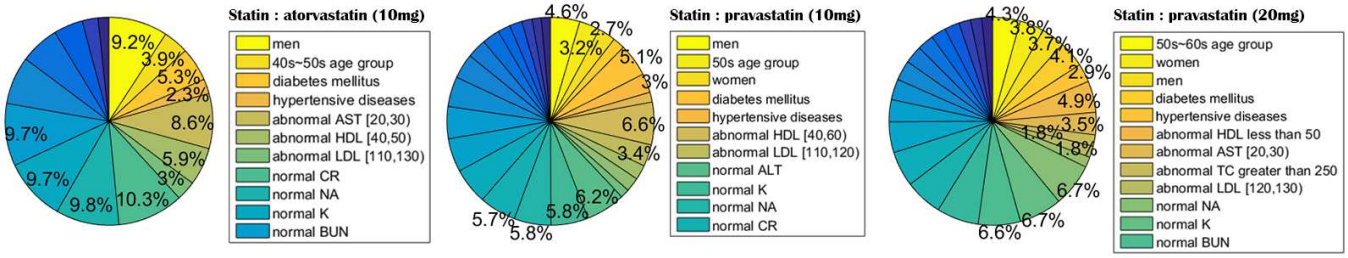
We also reported three phenotypes of the moderate-efficacy group (Figure 7), as follows:

(1) "Mid-aged men with diabetes mellitus and hypertension" (Figure 7a)
(2) "Mid-aged patients with diabetes mellitus and hypertension" (Figure 7b)
(3) "Patients over 50 with diabetes mellitus and hypertension" (Figure 7c)



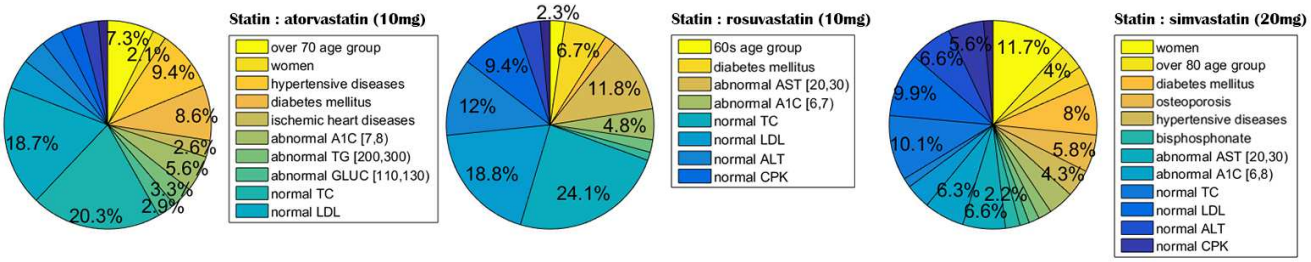
(a) Men over 60 with metabolic diseases. (b) Women over 70 with metabolic diseases. (c) Mid-aged men with heart disease and pre-diabetes.

Figure 6: Three representative phenotypes of the high-efficacy group (GLUC: glucose, TC: total cholesterol, NA: sodium, TB: total bilirubin, K: potassium, BUN: blood urea nitrogen).



(a) Mid-aged men with diabetes mellitus and hypertension. (b) Mid-aged patients with diabetes mellitus and hypertension. (c) Patients over 50 with diabetes mellitus and hypertension.

Figure 7: Three representative phenotypes of the moderate-efficacy group (CR: creatinine, NA: sodium, K: potassium, BUN: blood urea nitrogen, TC: total cholesterol).



(a) Women over 70 within normal LDL levels. (b) Patients over 60 within normal LDL levels. (c) Women over 80 within normal LDL levels.

Figure 8: Three representative phenotypes of the low-efficacy group (GLUC: glucose, TC: total cholesterol).

Each subgroup within the moderate-efficacy group takes atorvastatin (10 mg), pravastatin (10 mg) and pravastatin (20 mg), respectively. According to statin therapy, atorvastatin (10 mg) [1, 28] is expected to reduce LDL-C levels by up to 50%. Additionally, pravastatin (10 mg) and pravastatin (20 mg) are known to lower LDL-C levels by up to 30% [28]. Each phenotype's expected efficacy is consistent with statin therapy.

Finally, we reported three phenotypes of the low-efficacy group (Figure 8), as follows:

(1) "Women over 70 within normal LDL-C level" (Figure 8a)
(2) "Patients over 60 within normal LDL-C level" (Figure 8b)
(3) "Women over 80 within normal LDL-C level" (Figure 8c)

Phenotypes within the low-efficacy group represent patient groups in which patients have a normal LDL-C level (LDL-C<100 mg/dL) [32] prior to taking a certain statin. Consequently, unlike statin therapy, these phenotypes show that the efficacy is lower than the efficacy expected according to statin guidelines regardless of the traits of patients and types of prescribed statins.

These results demonstrate the various advantages of using our proposed method. First, we can easily observe discriminative characteristics on each patient group. In contrast to the high-efficacy group (Figure 6) and the moderate-efficacy group (Figure 7), phenotypes within the low-efficacy group (Figure 8) represent patients within their normal LDL-C level (LDL-C<100 mg/dL) and TC level (TC<200 mg/dL) [32] before taking a statin. Compared to the phenotypes of the moderate-efficacy group (Figure 7), the phenotypes within the high-efficacy group (Figure 6) are associated with a specific condition, such as metabolic diseases, and represent the patients whose percent reduction in LDL-C is higher than expected. According to statin guidelines, taking atorvastatin (40~80 mg) or rosuvastatin (20~40 mg) is recommended to reduce more than 50 percent of △LDL-C [1, 28]. However, patients who take high doses of statins may be more likely to develop kidney problems, such as acute kidney injury [8]. From our findings, we can recommend a low dosage of statin, such as atorvastatin (20 mg) or rosuvastatin (10 mg), rather than atorvastatin (40~80 mg) or rosuvastatin (20~40 mg) when patients are highly associated with particular phenotypes (Figures 6a, 6b and 6c) of the high-efficacy group. The phenotypes of the moderate-efficacy group represent the patients whose percent reduction in LDL-C is in accordance with statin therapy [1, 28]. According to the results, our proposed method can discover the high efficacy of a statin in specific conditions, such as metabolic diseases, as well as the expected efficacy of a statin, as known according to statin guidelines. Moreover, we can easily obtain insights on subgroups within each patient group. For example, we observed two different subgroups of patients with metabolic diseases within the high-efficacy group (Figures 6a and 6b).

However, it is impossible to conduct a numerical analysis for the efficacy of each statin because we cannot conclude that a statin only has an influence on reducing LDL-C levels. Therefore, a clinical study must be conducted to determine how much each statin decreases LDL-C levels rather than data-driven approaches such as our proposed method.

5.3 Accuracy

To evaluate the extent to which our proposed method obtains better discrimination on several groups compared to the baselines, the derived phenotypes and the latent representations of patients were evaluated on a classification task of predicting a new patient's group. We compared SNTFL (λ = 0.005) to three baselines with varying R. The results indicated that the latent representations of SNTFL (λ = 0.005) obtained better diagnostic accuracy than the baselines (Figures 9, 10). For all values of R, SNTFL (λ = 0.005) achieved the best prediction result in the two metrics.

In terms of MicroAUC (Figure 9), the best average MicroAUC was 0.7592 using SNTFL (λ = 0.005, R = 70) but only 0.7148 using CP-APR (R = 70); i.e., SNTFL achieved an increase of almost 0.0444. For NTF (R = 70) and SNTFL (λ = 0, R = 70), the best average MicroAUC was 0.7388 and 0.7519, respectively. When R was 50, a significant difference in the average MicroAUC between SNTFL (λ = 0) and SNTFL (λ = 0.005) was observed; i.e., the average MicroAUC of SNTFL (λ = 0.005) was approximately 0.0125 higher than the average MicroAUC of SNTFL (λ = 0).

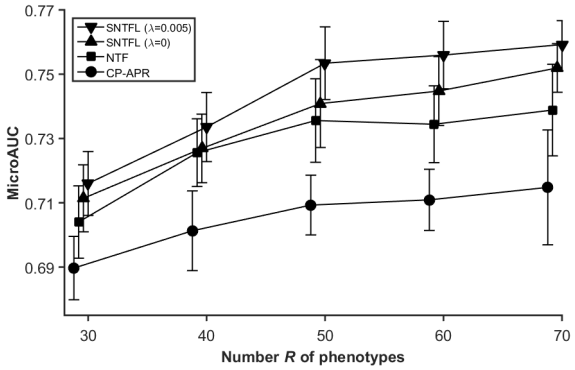


Figure 9: Effect of the number R of phenotypes on the average MicroAUC over 10 trials. Bars: ±1 s.d., n = 10.

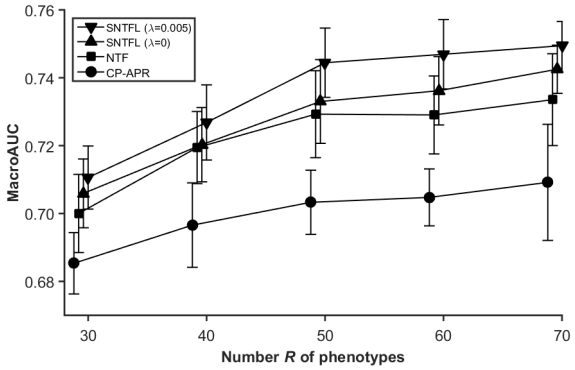


Figure 10: Effect of the number R of phenotypes on the average MacroAUC over 10 trials. Bars: ±1 s.d., n = 10.

The best average MacroAUC was 0.7494 using SNTFL (λ = 0.005, R = 70) (Figure 10). This result was 0.0402 higher than the best average MacroAUC of CP-APR (R = 70). For NTF (R = 70) and SNTFL (λ = 0, R = 70), the best average MacroAUC was 0.7335 and 0.7425, respectively.

These results indicate that the classification model of SNTFL was more predictive than the classification model of the baselines. This means that the grouped phenotypes and the latent representations from SNTFL have discriminative characteristics on each patient group rather than from the baselines. From these results, we infer that using the group information of patients and optimizing a multinomial logistic regression during the tensor factorization process made the phenotypes and latent representations discriminative. Conversely, the baselines obtained inferior accuracy because the baselines aimed only to derive individual phenotypes that approximate the original data as closely as possible.

6 CONCLUSION

In this paper, a novel tensor factorization method, named SNTFL, is proposed to derive grouped phenotypes that are discriminative regardless of the number of patient groups. We define a discriminative constraint to improve discrimination among patient groups and design an algorithm that optimizes a multinomial logistic regression during the tensor factorization process. To evaluate the effectiveness of SNTFL, we conduct a case study on a hyperlipidemia dataset. Consequently, our method derives grouped phenotypes that have discriminative characteristics on each patient group and achieves better diagnostic accuracy than the baselines. Moreover, SNTFL successfully discovers meaningful patient subgroups.

ACKNOWLEDGMENTS

This work was partly supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea (No. HC15C1362), and Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.2014-0-00147, Basic Software Research in Human-level Lifelong Machine Learning (Machine Learning Center)).

REFERENCES

[1] American Diabetes Association. 2017. Standards of Medical Care in Diabetes-2017 Abridged for Primary Care Providers. *Clinical Diabetes* 35, 1 (2017), 5–26.
[2] B. W. Bader, T. G. Kolda, et al. 2012. MATLAB Tensor Toolbox Version 2.5. Available online. (January 2012). http://www.sandia.gov/~tgkolda/TensorToolbox/
[3] D. P. Bertsekas. 1999. Nonlinear programming. In *Belmont: Athena scientific*. 1–60.
[4] J. D. Carroll and J. J. Chang. 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika* 35, 3 (1970), 283–319.
[5] Z. Che, D. Kale, W. Li, M. T. Bahadori, and Y. Liu. 2015. Deep computational phenotyping. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM.* 507–516.
[6] E. C. Chi and T.G. Kolda. 2012. On tensors, sparsity, and nonnegative factorizations. *SIAM J. Matrix Anal. Appl.* 33, 4 (2012), 1272–1299.
[7] A. Cichock, A. H. Phan R. Zdunek, and S.I. Amari. 2009. Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. *John Wiley & Sons* (2009).
[8] C. R. Dormuth and et al. 2013. Use of high potency statins and rates of admission for acute kidney injury: multicenter, retrospective observational analysis of administrative databases. *Bmj* 346, f880 (2013).
[9] M. Dyrby, D. Baunsgaard, R. Bro, and S. B. Engelsen. 2005. Multiway chemometric analysis of the metabolic response to toxins monitored by NMR. *Chemometrics and Intelligent Laboratory Systems* 76, 1 (2005), 79–89.
[10] R. A. Harshman. 1970. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. In *UCLA Work Papers Phonet*, Vol. 16. 1–84.
[11] J. C. Ho, J. Ghosh, S. R. Steinhubl, W. F. Stewart, J. C. Denny, B. A. Malin, and J. Sun. 2014. Limestone: High-throughput candidate phenotype generation via tensor factorization. *Journal of biomedical informatics* 52, 6 (2014), 199–211.
[12] J. C. Ho, J. Ghosh, and J. Sun. 2014. Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.* 115–124.
[13] G. Hripcsak and D. J. Albers. 2013. Next-generation phenotyping of electronic health records. Journal of the American Medical Informatics Association. *Journal of the American Medical Informatics Association* 20, 1 (2013), 117–121.
[14] D. Kale, Z. Che, Y. Liu, and R. Wetzel. 2014. Computational discovery of physiomes in critically ill children using deep learning. In *DMMI Workshop, AMIA*.
[15] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani. 2011. Risk prediction models for hospital readmission: a systematic review. *Jama* 306, 15 (2011), 1688–1698.
[16] H. S. Kim, H. Lee, B. Park, S. Park, H. Kim, S. H. Lee, and et al. 2016. Comparative analysis of the efficacy of low-and moderate-intensity statins in Korea. *International journal of clinical pharmacology and therapeutics* 54, 11 (2016), 864.
[17] Y. Kim, R. El-Kareh, J. Sun, H. Yu, and X. Jiang. 2017. Discriminative and distinct phenotyping by constrained tensor factorization. *Scientific Reports* 7 (2017).
[18] T. A. Lasko, J. C. Denny, and M. A. Levy. 2013. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PloS one* 8, 6 (2013), e66341.
[19] L. D. Lathauwer and B. D. Moor. 1998. From matrix to tensor: Multilinear algebra and signal processing. In *Institute of Mathematics and Its Applications Conference Series*. 1–16.
[20] C. J. Lin. 2007. Projected gradient methods for nonnegative matrix factorization. *Neural computation* 19, 10 (2007), 2756–2779.
[21] C. Liu, F. Wang, J. Hu, and H. Xiong. 2015. Temporal phenotyping from longitudinal electronic health records: A graph based framework. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM.* 705–714.
[22] T. P. Minka. 2003. A comparison of numerical optimizers for logistic regression. *Unpublished draft* (2003).
[23] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley. 2016. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports* 6 (2016).
[24] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley. 2017. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics* (2017), bbx044.
[25] M. Mørup. 2011. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1, 1 (2011), 24–40.
[26] K. M. Newton, P. L. Peissig, A. N. Kho, S. J. Bielinski, R. L. Berg, V. Choudhary, M. Basford, C. G. Chute, I. J. Kullo, R. Li, J. A. Pacheco, L. V. Rasmussen, L. Spangler, and J. C. Denny. 2013. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *JAMIA* 20, e1 (2013), e147–e154.
[27] P. Schulam, F. Wigley, and S. Saria. 2015. Clustering Longitudinal Clinical Marker Trajectories from Electronic Health Data: Applications to Phenotyping and Endotype Discovery. In *AAAI*. 2956–2964.
[28] N. J. Stone and et al. 2014. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults. *Circulation* 129, 25 suppl 2 (2014), S1–S45.
[29] K. B. Wagholikar, K. L. MacLaughlin, M. R. Henry, R. A. Greenes, R. A. Hankey, H. Liu, and R. Chaudhry. 2005. Clinical decision support with automated text processing for cervical cancer screening. *Journal of the American Medical Informatics Association* 19, 5 (2005), 833–839.
[30] H. Wang and N. Ahuja. 2004. Compact representation of multidimensional data using tensor rank-one decomposition. *vectors* (2004).
[31] Y. Wang, R. Chen, J. Ghosh, J. C. Denny, A. Kho, Y. Chen, B. A. Malin, and J. Sun. 2015. Rubik: Knowledge guided tensor factorization and completion for health data analytics. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM.* 1265–1274.
[32] Williams and Lippincott. 2002. Third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. *Circulation* 106, 25 (2002), 3143–3143.
[33] J. Wu, J. Roy, and W. F. Stewart. 2010. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. *Medical care* 48, 6 (2010), S106–S113.
[34] K. Yang, X. Li, H. Liu, J. Mei, G. Xie, J. Zhao, and F. Wang. 2017. TaGiTeD: Predictive Task Guided Tensor Decomposition for Representation Learning from Electronic Health Records. In *Thirty-First AAAI Conference on Artificial Intelligence*.
[35] J. Zhou, J. Sun, Y. Liu, J. Hu, and J. Ye. 2013. Patient risk prediction model via top-k stability selection. In *SIAM conference on data mining (SIAM)*.
[36] J. Zhou, F. Wang, J. Hu, and J. Ye. 2014. From micro to macro: data driven phenotyping by densification of longitudinal electronic medical records. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.* 135–144.