Hindawi
BioMed Research International
Volume 2020, Article ID 4675395, 11 pages
https://doi.org/10.1155/2020/4675395

Research Article

A Novel Triple Matrix Factorization Method for Detecting Drug-Side Effect Association Based on Kernel Target Alignment

Xiaoyi Guo,[1] Wei Zhou,[1] Yan Yu,[1] Yijie Ding,[2] Jijun Tang,[3,4] and Fei Guo[3]

[1]*The Hemodialysis Center, The Affiliated Wuxi People's Hospital of Nanjing Medical University, 214000 Wuxi, China*
[2]*School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu 215009, China*
[3]*School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin 300072, China*
[4]*Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA*

Correspondence should be addressed to Yan Yu; rush19830127@163.com, Yijie Ding; wuxi_dyj@163.com, and Fei Guo; fguo@tju.edu.cn

Received 15 March 2020; Accepted 8 April 2020; Published 29 May 2020

Guest Editor: Qin Ma

All drugs usually have side effects, which endanger the health of patients. To identify potential side effects of drugs, biological and pharmacological experiments are done but are expensive and time-consuming. So, computation-based methods have been developed to accurately and quickly predict side effects. To predict potential associations between drugs and side effects, we propose a novel method called the Triple Matrix Factorization- (TMF-) based model. TMF is built by the biprojection matrix and latent feature of kernels, which is based on Low Rank Approximation (LRA). LRA could construct a lower rank matrix to approximate the original matrix, which not only retains the characteristics of the original matrix but also reduces the storage space and computational complexity of the data. To fuse multivariate information, multiple kernel matrices are constructed and integrated via Kernel Target Alignment-based Multiple Kernel Learning (KTA-MKL) in drug and side effect space, respectively. Compared with other methods, our model achieves better performance on three benchmark datasets. The values of the Area Under the Precision-Recall curve (AUPR) are 0.677, 0.685, and 0.680 on three datasets, respectively.

1. Introduction

Drug treatment of patients' diseases may be accompanied by side effects, endangering the life and health of patients. Therefore, how to quickly and accurately find potential drug side effect information becomes an important step in the drug development process. The traditional methods to detect the side effects of drugs are usually biological and pharmacological experiments. These approaches often take a long time and huge capital investment. So, it is necessary to accurately and quickly predict the potential side effects of drugs through computation-based methods [1]. Most computation-based methods for predicting drug side effects used Machine Learning (ML) classification models to predict side effect categories by extracted features from the biochemical characteristics of drugs. ML has been widely used in the field of computational biology, containing potential disease-associated microRNAs [2, 3] or circRNAs [4], O-GlcNAcylation sites [5], prediction of DNA or RNA methylcytosine sites [6, 7], protein function identification [8–12], protein remote homology [13], analyzing microbiology [14], electron transport proteins [15], drug-target interactions [16], drug-side effect association [17, 18], protein-protein interactions [19, 20], and lncRNA-miRNA interactions [21].

Pauwels and Stoven develop a predictive model of drug-side effect association by Ordinary Canonical Correlation Analysis (OCCA) and Sparse Canonical Correlation Analysis (SCCA) [1, 22]. The input feature of OCCA and SCCA is extracted from chemical structures of drugs. Cheng and Wang proposed the Phenotypic Network Inference Model (PNIM) [23] to detect new potential drug-side effect associations. Mizutani and Stoven [24] utilized cooccurrence of drug profiles and protein interaction profiles to predict side effects. The Support Vector Machine (SVM) was used to build Adverse Drug Reaction (ADR) prediction, which is based on chemical structures, biological properties of drugs, and

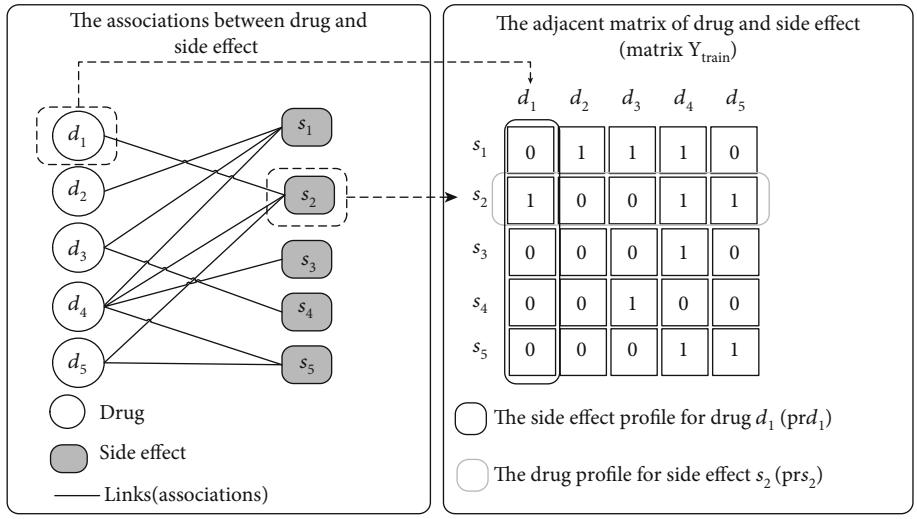


Figure 1: The schematic diagram of associations between drugs and side effects.

phenotypic characteristics [25]. Zhang et al. [26–28] built an ensemble method, which was based on the Integrated Neighborhood-Based Method (INBM) and Restricted Boltzmann Machine-Based Method (RBMBM). Matrix Factorization- (MF-) based methods have been widely used for link prediction in bipartite networks of systems biology. To predict drug-target interactions, Neighborhood Regularized Logistic Matrix Factorization (NRLMF) [29], Collaborative Matrix Factorization (CMF) [30], and Graph Regularized Matrix Factorization (GRMF) [31] were developed via the MF theory.

In our study, we develop a Triple Matrix Factorization- (TMF-) based model to identify the associations of drug and side effect. TMF employs the biprojection matrix and two latent feature matrices (from drug and side effect space) to estimate the strength of new drug-side effect associations. Latent feature matrices are built via Low Rank Approximation (LRA), which could construct a lower rank matrix to approximate the original matrix. To improve the performance of prediction, Kernel Target Alignment-based Multiple Kernel Learning (KTA-MKL) is used to integrate multiple kernel matrices in drug and side effect space, respectively. Our method can fuse multivariate information (multiple kernels) and obtain new associations through matrix projection. Compared with other existing methods, our model obtains better performance on three benchmark datasets.

2. Method

2.1. Problem Description. The dataset of drug-side effect association can be regarded as a bipartite network, which has n drugs and m side effects. The relationships of drug and side effect can be represented as a $n \times m$ adjacent matrix $\mathbf{Y} \in \mathbf{R}^{n \times m}$. $D = \{d_1, d_2, \cdots, d_n\}$ and $S = \{s_1, s_2, \cdots, s_m\}$ are the drug and side effect sets, respectively. $Y_{i,j} = 1$ denotes that drug d_i and side effect s_j are related; otherwise, it is 0. The associations between drugs and side effect terms are shown in Figure 1. The solid lines link the known drug-side effect associations. The hollow circles and filled squares are drugs and side effects, respectively. The prediction of new associations is a recommender task.

2.2. Drug Kernels and Side Effect Kernels. To predict the associations of drugs and side effects, we need to construct the relationship between drugs (or side effects). In this study, we build different kernels (similarity matrices) to describe the relationships of drugs (or side effects). In drug space, the fingerprint of 881 chemical substructures is employed to encode the drug chemical structure, which is shown in Figure 2. The fingerprint represents whether some substructures are present (1) or absent (0). What is more, the known links between drugs and side effect terms (a side effect profile for a specific drug) are also used to represent the information of the subjacent network, which is shown on the right side of Figure 1. In side effect space, the drug profile for a side effect also represents the subjacent network of side effects.

There are four different types of measure functions, including Gaussian Interaction Profile (GIP) kernel [32–35], COsine Similarity (COS) [26], Correlation coefficient (Corr) [26], and Mutual Information (MI) [36, 37], which are employed to calculate the similarity between the drugs (or side effects).

For drug d_i and d_k, the GIP kernel is defined as follows:

$$\mathbf{K}_{\text{GIP-link},d}(d_i, d_k) = \exp\left(-\gamma \left\| \mathbf{pr}_{d_i} - \mathbf{pr}_{d_k} \right\|^2\right), \quad (1)$$

where γ is the bandwidth of the Gaussian kernel. γ is set as 1 in our study. $\mathbf{pr}_{d_i}$ and $\mathbf{pr}_{d_k}$ are the side effect profile of drug d_i and d_k, respectively.

The COS is defined as follows:

$$\mathbf{K}_{\text{COS-link},d}(d_i, d_k) = \frac{\mathbf{pr}_{d_i} \mathbf{pr}_{d_k}^T}{\left\|\mathbf{pr}_{d_i}\right\| \left\|\mathbf{pr}_{d_k}\right\|}. \quad (2)$$

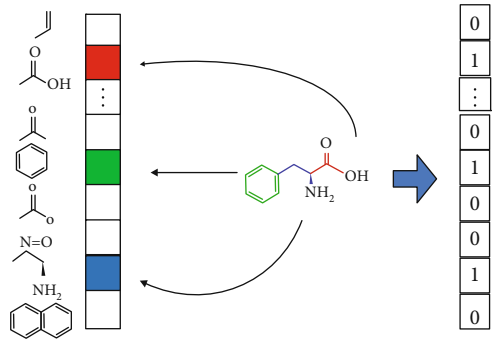


Figure 2: An example of the fingerprint vector.

Table 1: Summary of kernels for two feature spaces.

	Chemical fingerprint (drug space)	Side effect profiles (drug space)	Drug profiles (side effect space)
GIP	$\mathbf{K}_{\text{GIP-chem},d}$	$\mathbf{K}_{\text{GIP-link},d}$	$\mathbf{K}_{\text{GIP-link},s}$
COS	$\mathbf{K}_{\text{COS-chem},d}$	$\mathbf{K}_{\text{COS-link},d}$	$\mathbf{K}_{\text{COS-link},s}$
Corr	$\mathbf{K}_{\text{Corr-chem},d}$	$\mathbf{K}_{\text{Corr-link},d}$	$\mathbf{K}_{\text{Corr-link},s}$
MI	$\mathbf{K}_{\text{MI-chem},d}$	$\mathbf{K}_{\text{MI-link},d}$	$\mathbf{K}_{\text{MI-link},s}$

The Corr kernel is calculated as follows:

$$\mathbf{K}_{\text{Corr-link},d}(d_i, d_k) = \frac{\text{Cov}\left(\mathbf{pr}_{d_i}, \mathbf{pr}_{d_k}\right)}{\sqrt{\text{Var}\left(\mathbf{pr}_{d_i}\right)\text{Var}\left(\mathbf{pr}_{d_k}\right)}}. \tag{3}$$

In order to describe the degree of correlation between two random variables, we further use Mutual Information (MI) to measure the similarity between the two random variables:

$$\begin{aligned}\mathbf{K}_{\text{MI-link},d}(d_i, d_k) &= I\left(\mathbf{pr}_{d_i}, \mathbf{pr}_{d_k}\right) \\ &= \sum_{u=0}^{1}\sum_{v=0}^{1} f(u,v)\log\left(\frac{f(u,v)}{f(u)f(v)}\right),\end{aligned} \tag{4}$$

where $f(u)$ ($f(v)$) denotes the observed frequency of value u (v) in profile $\mathbf{pr}_{d_i}$ ($\mathbf{pr}_{d_k}$). $f(u,v)$ is the observed relative frequency. Similarly, the kernels of the fingerprint (drug space: $\mathbf{K}_{\text{GIP-chem},d}$, $\mathbf{K}_{\text{COS-chem},d}$, $\mathbf{K}_{\text{Corr-chem},d}$, and $\mathbf{K}_{\text{MI-chem},d}$) and drug profile of side effects (side effect space: $\mathbf{K}_{\text{GIP-link},s}$, $\mathbf{K}_{\text{COS-link},s}$, $\mathbf{K}_{\text{Corr-link},s}$, and $\mathbf{K}_{\text{MI-link},s}$) can be constructed via the above functions. The drug space has 8 kernels, and the side effect space has 4 kernels, which are listed in Table 1.

2.3. Kernel Target Alignment-Based Multiple Kernel Learning. In our study, the kernel sets for drug space $\mathbf{K}_d = \{\mathbf{K}_{1,d}, \mathbf{K}_{2,d}, \cdots, \mathbf{K}_{k_d,d}\}$ and side effect space $\mathbf{K}_s = \{\mathbf{K}_{1,s}, \mathbf{K}_{2,s}, \cdots, \mathbf{K}_{k_s,s}\}$ are combined via multiple kernel learning, respectively. k_d and k_s are the number of kernels in drug and side effect space, respectively. A heuristic approach of Kernel Target Alignment-based Multiple Kernel Learning (KTA-MKL) [38, 39] is employed to calculate the weights of each kernel. The optimal kernels of $\mathbf{K}_d^*$ and $\mathbf{K}_s^*$ can be obtained as follows:

$$\begin{aligned}\mathbf{K}_d^* &= \sum_{p=1}^{k_d}\beta_{p,d}\mathbf{K}_{p,d}, \quad \mathbf{K}_{p,d} \in \mathbf{R}^{n\times n}, \\ \mathbf{K}_s^* &= \sum_{q=1}^{k_s}\beta_{q,s}\mathbf{K}_{q,s}, \quad \mathbf{K}_{q,s} \in \mathbf{R}^{m\times m},\end{aligned} \tag{5}$$

where $\beta_d = \{\beta_{1,d}, \beta_{2,d}, \cdots, \beta_{k_d,d}\}$ and $\beta_s = \{\beta_{1,s}, \beta_{2,s}, \cdots, \beta_{k_s,s}\}$ are the weights of kernels in drug and side effect space, respectively. KTA-MKL estimates the weight of each kernel by COsine Similarity of matrices (drug space):

$$A(\mathbf{P}, \mathbf{Q}) = \frac{\langle \mathbf{P}, \mathbf{Q}\rangle_{\text{F}}}{\|\mathbf{P}\|_{\text{F}}\|\mathbf{Q}\|_{\text{F}}}, \tag{6}$$

where $\|\mathbf{P}\|_{\text{F}} = \sqrt{\langle \mathbf{P}, \mathbf{P}\rangle_{\text{F}}}$ denotes the Frobenius norm. $\langle \mathbf{P}, \mathbf{Q}\rangle_{\text{F}} = \text{Trace}(\mathbf{P}^T\mathbf{Q})$ is the Frobenius inner product. The value of kernel alignment can describe similarity of two kernels. KTA-MKL estimates the value between the ideal kernel matrix and the drug kernel (or side effect kernel) as follows:

$$\begin{aligned}\beta_{p,d} &= \frac{A\left(\mathbf{K}_{p,d}, \mathbf{K}_{\text{ideal},d}\right)}{\sum_{k=1}^{k_d} A\left(\mathbf{K}_{k,d}, \mathbf{K}_{\text{ideal},d}\right)}, \quad p = 1, 2, \cdots, k_d, \\ \beta_{q,s} &= \frac{A\left(\mathbf{K}_{q,s}, \mathbf{K}_{\text{ideal},s}\right)}{\sum_{k=1}^{k_s} A\left(\mathbf{K}_{k,s}, \mathbf{K}_{\text{ideal},s}\right)}, \quad q = 1, 2, \cdots, k_s,\end{aligned} \tag{7}$$

where $\mathbf{K}_{\text{ideal},d} = \mathbf{Y}_{\text{train}}\mathbf{Y}_{\text{train}}^T \in \mathbf{R}^{n\times n}$ and $\mathbf{K}_{\text{ideal},s} = \mathbf{Y}_{\text{train}}^T\mathbf{Y}_{\text{train}} \in \mathbf{R}^{m\times m}$ are the ideal kernels of drug and side effect, respectively, which are built via the training label (known associations).

2.4. Triple Matrix Factorization-Based Model. Inspired by MF [29–31, 40], the similarity between drugs (or side effects) can be approximated by the inner product of two drug (or side effect) features as follows:

$$\begin{aligned}\mathbf{K}_d^* &\approx \mathbf{A}\mathbf{A}^T, \quad \mathbf{A} \in \mathbf{R}^{n\times r_d}, \\ \mathbf{K}_s^* &\approx \mathbf{B}\mathbf{B}^T, \quad \mathbf{B} \in \mathbf{R}^{m\times r_s},\end{aligned} \tag{8}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the matrices of Low Rank Approximation and r_d and r_s are the dimensions of the latent feature space in drug and side effect space, respectively. The objective function of TMF is defined as follows:

$$\min\ J(\mathbf{\Theta}) = \left\|\mathbf{Y}_{\text{train}} - \mathbf{A}\mathbf{\Theta}\mathbf{B}^T\right\|_{\text{F}}^2 + \lambda\|\mathbf{\Theta}\|_{\text{F}}^2, \tag{9}$$

where $\mathbf{\Theta} \in \mathbf{R}^{r_d\times r_s}$ is the biprojection matrix. λ is the regularization coefficient of $\mathbf{\Theta}$. In our study, λ is set as 1.

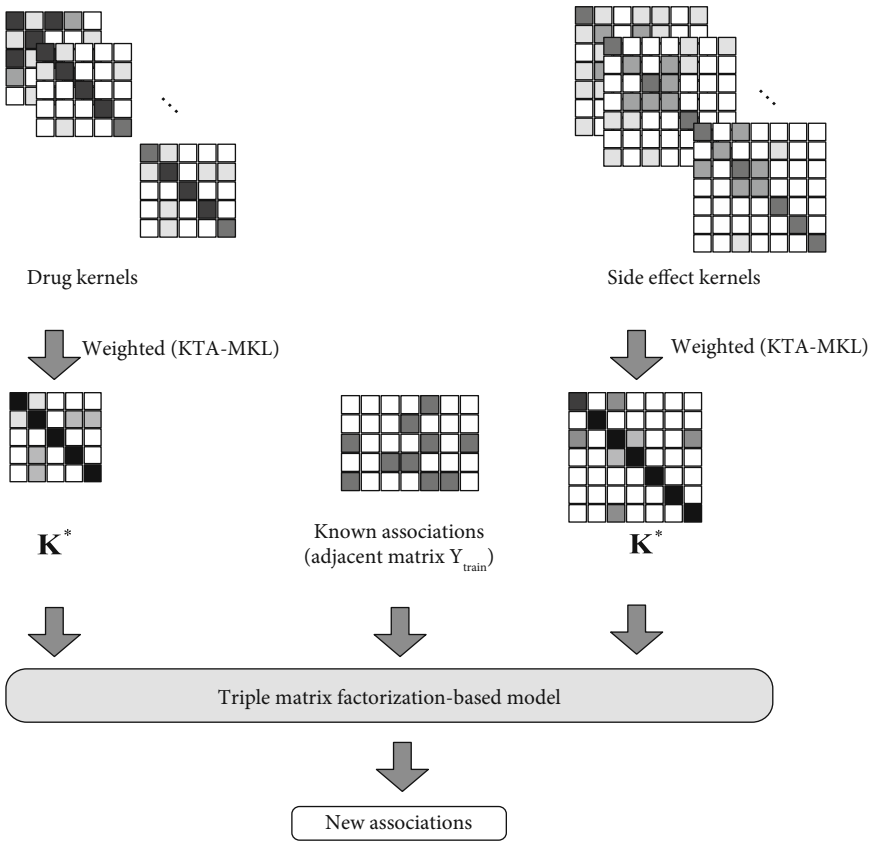


Figure 3: Overview of our method.

Require: A training matrix $\mathbf{Y}_{\text{train}} \in \mathbf{R}^{n\times m}$ (known associations), the fingerprint vector $\mathbf{f}_{d_i}^{\text{chem}} \in \mathbf{R}^{1\times 881}(1 \le i \le n)$ for the drug; Two parameters: the r_d and r_s for TMF;
Ensure: The prediction of $\mathbf{Y}^* \in \mathbf{R}^{n\times m}$;
1: Constructing the drug and side effect kernels, which are listed in Table 1;
2: Utilizing Equation (7) (KTA-MKL) to obtain the weights β_d and β_s for drug and side effect, respectively;
3: Building $\mathbf{K}_d^*$ and $\mathbf{K}_s^*$ via Equation (5), respectively;
4: Calculating $\mathbf{A} \in \mathbf{R}^{n\times r_d}$ and $\mathbf{B} \in \mathbf{R}^{m\times r_s}$ by Singular Value Decomposition (SVD);
5: Solving Equation (14) (TMF) to estimate $\boldsymbol{\Theta}$;
6: Calculating $\mathbf{Y}^* = \mathbf{A}\boldsymbol{\Theta}\mathbf{B}^T$;

Algorithm 1: Algorithm of our method.

Let $\partial J/\partial \boldsymbol{\Theta} = 0$, so we can obtain functions as follows:

$$\frac{\partial\left(\left\|\mathbf{Y}_{\text{train}} - \mathbf{A}\boldsymbol{\Theta}\mathbf{B}^T\right\|_{\text{F}}^2 + \lambda\|\boldsymbol{\Theta}\|_{\text{F}}^2\right)}{\partial \boldsymbol{\Theta}} = 0, \tag{10}$$

$$-2\mathbf{A}^T\left(\mathbf{Y}_{\text{train}} - \mathbf{A}\boldsymbol{\Theta}\mathbf{B}^T\right)\mathbf{B} + 2\lambda\boldsymbol{\Theta} = 0, \tag{11}$$

$$\mathbf{A}^T\mathbf{A}\boldsymbol{\Theta}\mathbf{B}^T\mathbf{B} + \lambda\boldsymbol{\Theta} = \mathbf{A}^T\mathbf{Y}_{\text{train}}\mathbf{B}, \tag{12}$$

$$\mathbf{A}^T\mathbf{A}\boldsymbol{\Theta} + \lambda\boldsymbol{\Theta}\left(\mathbf{B}^T\mathbf{B}\right)^{-1} = \mathbf{A}^T\mathbf{Y}_{\text{train}}\mathbf{B}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}, \tag{13}$$

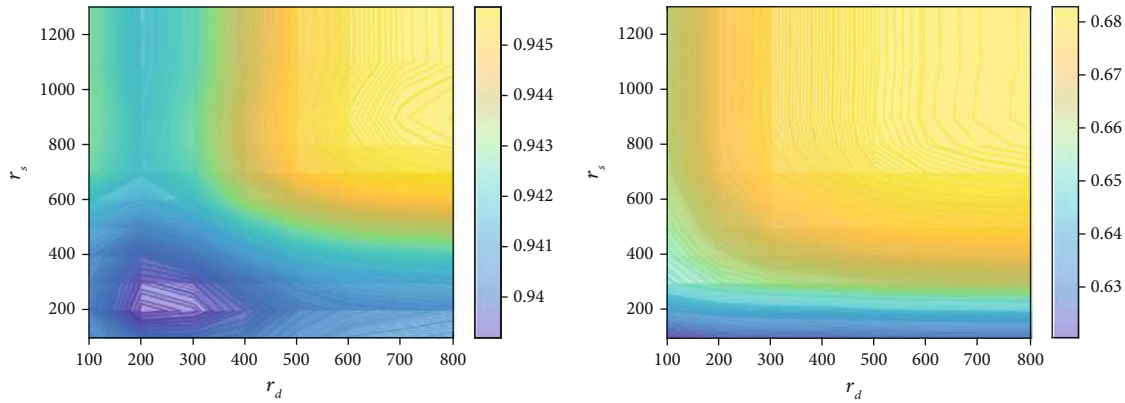
$$\mathbf{A}^T\mathbf{A}\boldsymbol{\Theta} + \lambda\boldsymbol{\Theta}\left(\mathbf{B}^T\mathbf{B}\right)^{-1} = \mathbf{A}^T\mathbf{Y}_{\text{train}}\left(\mathbf{B}^T\right)^{-1}, \tag{14}$$

Table 2: Three benchmark datasets.

Datasets	Drugs	Side effects	Associations
Pauwels's dataset	888	1385	61,102
Mizutani's dataset	658	1339	49,051
Liu's dataset	832	1385	59,205

where Equation (14) is a Sylvester equation. The final prediction can be constructed by

$$\mathbf{Y}^* = \mathbf{A}\boldsymbol{\Theta}\mathbf{B}^T. \tag{15}$$



(a) AUPR values on Mizutani (under different r_d and r_s)

(b) AUC values on Mizutani (under different r_d and r_s)

Figure 4: The AUC and AUPR values (under different r_d and r_s).

Table 3: The performance of different kernels via 5-fold Cross-Validation.

Models	Pauwels's dataset		Mizutani's dataset		Liu's dataset	
	AUPR	AUC	AUPR	AUC	AUPR	AUC
$\mathbf{K}_{\text{GIP-chem},d}$ & $\mathbf{K}_{\text{GIP-link},s}$ [a]	0.4420	0.8950	0.4735	0.9148	0.4718	0.9145
$\mathbf{K}_{\text{COS-chem},d}$ & $\mathbf{K}_{\text{COS-link},s}$ [a]	0.4892	0.8994	0.5343	0.9070	0.5224	0.9067
$\mathbf{K}_{\text{Corr-chem},d}$ & $\mathbf{K}_{\text{Corr-link},s}$ [a]	0.4994	0.8981	0.5217	0.9005	0.5143	0.9026
$\mathbf{K}_{\text{MI-chem},d}$ & $\mathbf{K}_{\text{MI-link},s}$ [a]	0.4978	0.9079	0.5591	0.9214	0.5529	0.9238
$\mathbf{K}_{\text{GIP-link},d}$ & $\mathbf{K}_{\text{GIP-link},s}$ [b]	0.6254	0.9300	0.6623	0.9376	0.6574	0.9398
$\mathbf{K}_{\text{COS-link},d}$ & $\mathbf{K}_{\text{COS-link},s}$ [b]	0.5861	0.9035	0.6324	0.9090	0.6252	0.9087
$\mathbf{K}_{\text{Corr-link},d}$ & $\mathbf{K}_{\text{Corr-link},s}$ [b]	0.5833	0.8999	0.6123	0.9014	0.6047	0.9013
$\mathbf{K}_{\text{MI-link},d}$ & $\mathbf{K}_{\text{MI-link},s}$ [b]	0.6557	0.9428	0.6615	0.9369	0.6587	0.9408
Mean weighted [c]	0.6598	0.9353	0.6724	0.9280	0.6651	0.9285
KTA-MKL [c]	0.6765	0.9434	0.6847	0.9409	0.6801	0.9426

[a]The TMF uses the drug fingerprint and drug profile for side effects. [b]The TMF uses the side effect profile for drugs and drug profile for side effects. [c]The TMF uses the drug fingerprint, side effect profile for drugs, and drug profile for side effects.

The overview of our proposed method is shown in Figure 3 and Algorithm 1.

3. Result

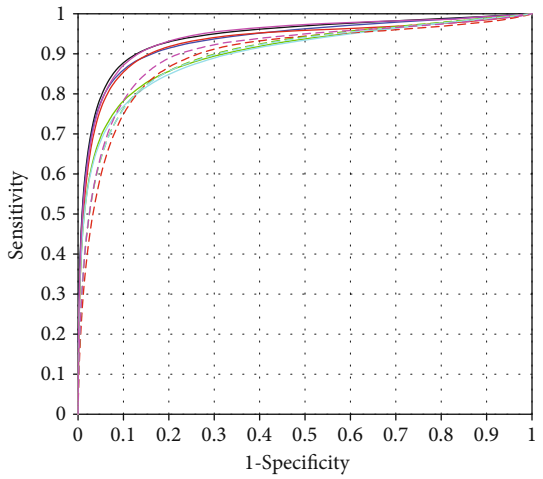
In this section, we employed benchmark dataset to evaluate our approach and compared it with other existing methods.

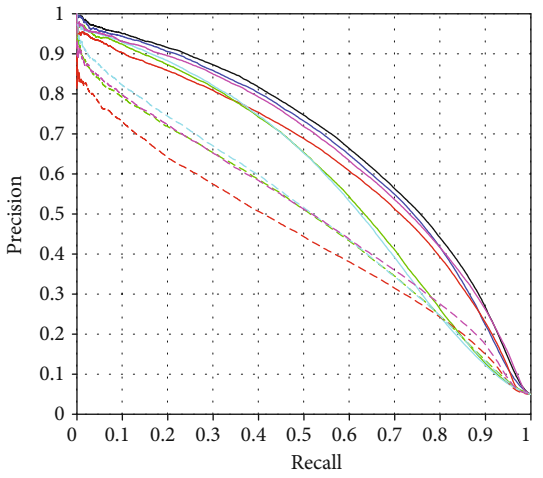
3.1. Datasets. In order to test the performance of our model, three types of datasets are employed in our study. They are Pauwels's dataset, Liu's dataset, and Mizutani's dataset, which are collected from the DrugBank [41], SIDe Effect Resource (SIDER) [42], KEGG DRUG [43], and PubChem-Compound [44, 45]. Table 2 lists benchmark datasets of this study.

3.2. Evaluation Measurements. The training adjacent matrix can be obtained via randomly setting known associations as 0. In this study, we use 5-fold Cross-Validation (5-CV) and 5-fold local Cross-Validation (5 local CV) to test our method. 5-CV randomly sets known associations as 0 in the whole matrix. 5 local CV is employed to evaluate the prediction of new drugs, which do not have any side effect information. 5 local CV sets some rows of the adjacent matrix as 0 to test related drugs. The Area Under the Precision-Recall curve (AUPR) and Area Under the receiver operating Characteristic curve (AUC) are utilized to evaluate the performance of prediction.

3.3. Selecting Optimal Parameters. In this section, we use the grid search method to get the optimal r_d and r_s. We test different values of and from 100 to the max value with the step of 100. The results of the grid search method are shown in Figure 4 (on Mizutani's dataset by 5-CV). $r_d = 700$ and $r_s = 800$ are the best parameters (AUPR) on Mizutani's dataset. In Figure 4, the lower value of AURP and AUC is blue, and the higher value is yellow. On the other two datasets, we use the same parameters of r_d and r_s.

3.4. Performance of Different Kernels. We evaluate the performance of multiple kernels and single kernel on three datasets. The results of prediction are listed in Table 3 and Figure 5. Obviously, the kernels of $\mathbf{K}_{\text{MI-link},d}$ and $\mathbf{K}_{\text{MI-link},s}$ have better performance on Pauwels's dataset (AUPR: 0.6557, AUC: 0.9079), Mizutani's dataset (AUPR: 0.6615, AUC: 0.9369),

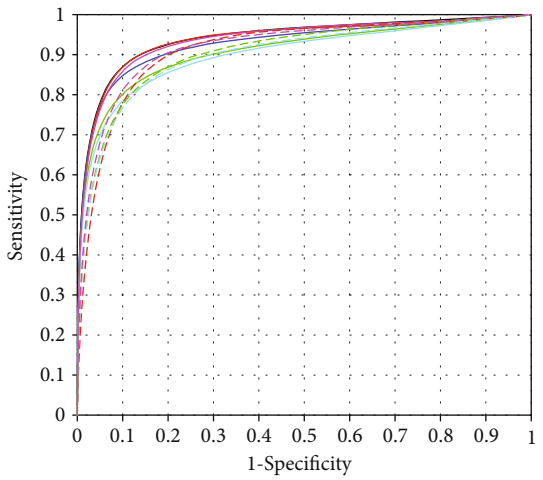


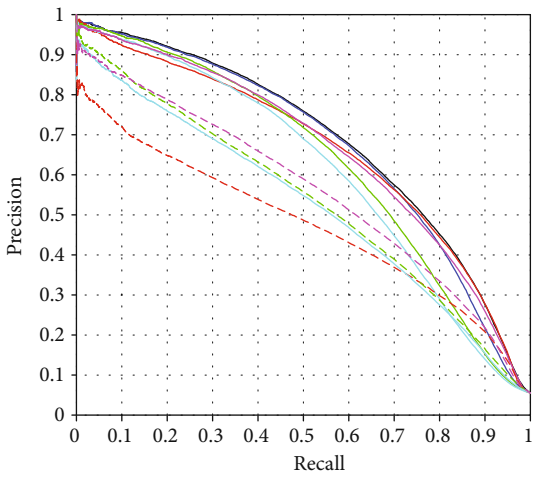


Weighted with KTA
Mean weighted
Drug($K_{GIP\text{-}link,d}$)&side effect($K_{GIP\text{-}link,s}$)
Drug($K_{COS\text{-}link,d}$)&side effect($K_{COS\text{-}link,s}$)
Drug($K_{Corr\text{-}link,d}$)&side effect($K_{Corr\text{-}link,s}$)
Drug($K_{MI\text{-}link,d}$)&side effect($K_{MI\text{-}link,s}$)
Drug($K_{GIP\text{-}chem,d}$)&side effect($K_{GIP\text{-}link,s}$)
Drug($K_{COS\text{-}chem,d}$)&side effect($K_{COS\text{-}link,s}$)
Drug($K_{Corr\text{-}chem,d}$)&side effect($K_{Corr\text{-}link,s}$)
Drug($K_{MI\text{-}chem,d}$)&side effect($K_{MI\text{-}link,s}$)

(a) ROC curves on Pauwels's dataset

(b) PR curves on Pauwels's dataset



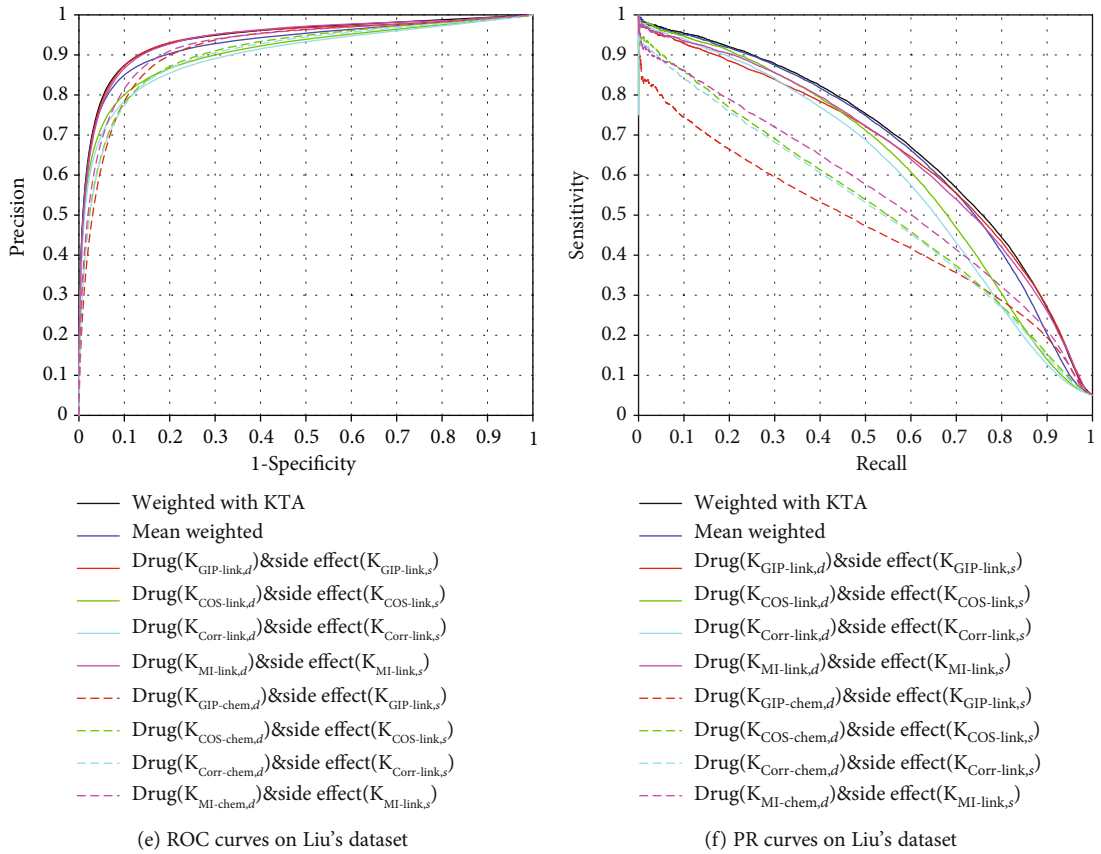


Weighted with KTA
Mean weighted
Drug($K_{GIP\text{-}link,d}$)&side effect($K_{GIP\text{-}link,s}$)
Drug($K_{COS\text{-}link,d}$)&side effect($K_{COS\text{-}link,s}$)
Drug($K_{Corr\text{-}link,d}$)&side effect($K_{Corr\text{-}link,s}$)
Drug($K_{MI\text{-}link,d}$)&side effect($K_{MI\text{-}link,s}$)
Drug($K_{GIP\text{-}chem,d}$)&side effect($K_{GIP\text{-}link,s}$)
Drug($K_{COS\text{-}chem,d}$)&side effect($K_{COS\text{-}link,s}$)
Drug($K_{Corr\text{-}chem,d}$)&side effect($K_{Corr\text{-}link,s}$)
Drug($K_{MI\text{-}chem,d}$)&side effect($K_{MI\text{-}link,s}$)

(c) ROC curves on Mizutani's dataset

(d) PR curves on Mizutani's dataset

Figure 5: Continued.



(e) ROC curves on Liu's dataset (f) PR curves on Liu's dataset

Figure 5: The ROC and PR curves of different models (single kernel and multiple kernels).

and Liu's dataset (AUPR: 0.6587, AUC: 0.9408). In addition, the KTA-MKL model achieves the best results on Pauwels's dataset (AUPR: 0.6765, AUC: 0.9434), Mizutani's dataset (AUPR: 0.6847, AUC: 0.9409), and Liu's dataset (AUPR: 0.6801, AUC: 0.9426), respectively. KTA-MKL could combine kernels from different sources via the heuristic method, which is better mean weighted.

In Table 4, we list the weight of each kernel on three datasets. We can find that the weights of $\mathbf{K}_{MI-link,d}$ and $\mathbf{K}_{MI-link,s}$ are the highest than other kernels. At the same time, their performance is also the best. KTA-MKL could reduce bias of kernels by the low weights.

3.5. Comparison with Existing Methods. To evaluate the performance of the TMF model, we compare it with other methods. The results are listed in Table 5. Obviously, our method (TMF) achieves the best results on Pauwels's dataset (AUPR: 0.677), Mizutani's dataset (AUPR: 0.685), and Liu's dataset (AUPR: 0.680). Zhang et al.'s work (ensemble model) [26] obtained the good performance of AUPRs (0.660, 0.666, and 0.661). The best AUCs (0.954, 0.950, and 0.953) are achieved by Neighborhood Regularized Logistic Matrix Factorization (NRLMF) [29], which is also based on Matrix Factorization (MF). The results of other MF-based models, including Collaborative Matrix Factorization (CMF) [30] and Graph Regularized Matrix Factorization (GRMF) [31], are competitive. Local and Global Consistency (LGC) [18] is our previous work. LGC obtains the second best results of AUPR (0.668, 0.673, and 0.670) on three datasets, respectively.

Table 4: The kernel weights on three datasets.

Kernel	Pauwels's dataset	Mizutani's dataset	Liu's dataset
$\mathbf{K}_{GIP-chem,d}$	0.1159	0.1168	0.1167
$\mathbf{K}_{COS-chem,d}$	0.1224	0.1226	0.1226
$\mathbf{K}_{Corr-chem,d}$	0.1200	0.1203	0.1203
$\mathbf{K}_{MI-chem,d}$	0.1113	0.1122	0.1116
$\mathbf{K}_{GIP-link,d}$	0.0596	0.0621	0.0613
$\mathbf{K}_{COS-link,d}$	0.1538	0.1533	0.1528
$\mathbf{K}_{Corr-link,d}$	0.1507	0.1498	0.1497
$\mathbf{K}_{MI-link,d}$	0.1664	0.1628	0.1650
$\mathbf{K}_{GIP-link,s}$	0.0151	0.0173	0.0152
$\mathbf{K}_{COS-link,s}$	0.3286	0.3374	0.3380
$\mathbf{K}_{Corr-link,s}$	0.2909	0.2865	0.2855
$\mathbf{K}_{MI-link,s}$	0.3654	0.3588	0.3613

3.6. Local CV and Case Study. In some cases, certain drugs are new and have no information of side effects. The 5 local CV is employed to test the performance of the side effect prediction for new drugs. In this section, we also compare TMF

Table 5: Comparison to existing methods via 5-fold Cross-Validation.

Datasets	Methods	AUPR	AUC
Pauwels	Pauwels's method[a]	0.389 ± N/A	0.897 ± N/A
	Liu's method[a]	0.345 ± N/A	0.920 ± N/A
	Cheng's method[a]	0.588 ± N/A	0.922 ± N/A
	RBMBM[a] [26]	0.612 ± N/A	0.941 ± N/A
	INBM[a] [26]	0.641 ± N/A	0.934 ± N/A
	Ensemble model[a] [26]	0.660 ± N/A	0.949 ± N/A
	CMF[b]	0.646 ± 0.007	0.939 ± 0.005
	GRMF[b]	0.643 ± 0.006	0.937 ± 0.005
	NRLMF[b]	0.654 ± 0.005	0.954 ± 0.005
	LGC[b]	0.668 ± 0.008	0.952 ± 0.007
	Our method	0.677 ± 0.004	0.943 ± 0.003
Mizutani	Mizutani's method[a]	0.412 ± N/A	0.890 ± N/A
	Liu's method[a]	0.366 ± N/A	0.918 ± N/A
	Cheng's method[a]	0.599 ± N/A	0.923 ± N/A
	RBMBM[a] [26]	0.619 ± N/A	0.939 ± N/A
	INBM[a] [26]	0.646 ± N/A	0.932 ± N/A
	Ensemble model[a] [26]	0.666 ± N/A	0.946 ± N/A
	CMF[b]	0.645 ± 0.005	0.938 ± 0.006
	GRMF[b]	0.646 ± 0.007	0.937 ± 0.007
	NRLMF[b]	0.660 ± 0.006	0.950 ± 0.005
	LGC[b]	0.673 ± 0.007	0.948 ± 0.007
	Our method	0.685 ± 0.006	0.941 ± 0.008
Liu	Liu's method[a]	0.278 ± N/A	0.907 ± N/A
	Cheng's method[a]	0.592 ± N/A	0.922 ± N/A
	RBMBM[a] [26]	0.616 ± N/A	0.941 ± N/A
	INBM[a] [26]	0.641 ± N/A	0.934 ± N/A
	Ensemble model[a] [26]	0.661 ± N/A	0.948 ± N/A
	CMF[b]	0.649 ± 0.006	0.938 ± 0.005
	GRMF[b]	0.650 ± 0.007	0.938 ± 0.008
	NRLMF[b]	0.656 ± 0.005	0.953 ± 0.006
	LGC[b]	0.670 ± 0.008	0.951 ± 0.007
	Our method	0.680 ± 0.005	0.943 ± 0.006

[a]Results are derived from [26]. [b]Results are derived from [18].

with other MF-based models, including NRLMF, CMF, and GRMF. The results are listed in Table 6 and Figure 6.

The proposed method (TMF) achieves the best results of AUPRs on Pauwels's dataset (AUPR: 0.392), Mizutani's dataset (AUPR: 0.399), and Liu's dataset (AUPR: 0.401). Other MF-based models also are still comparable with our results. NRLMF obtains AUPRs of 0.374, 0.390, and 0.398 on three datasets, respectively.

To predict the side effects of a new drug, our model calculates the strength of associations between the new drug and all existing side effects. The predictive strength scores of TMF will be ranked by descending order. The higher the value of the score, the higher the possibility of associations. In this section, we discuss two cases (drug caffeine and captopril on Mizutani's dataset) of top 10 associations predicted. The details are listed in Tables 7 and 8. Results are checked by the masked associations between drug caffeine (or captopril) and side effects.

Table 6: Comparison with MF-based models via 5-fold local Cross-Validation.

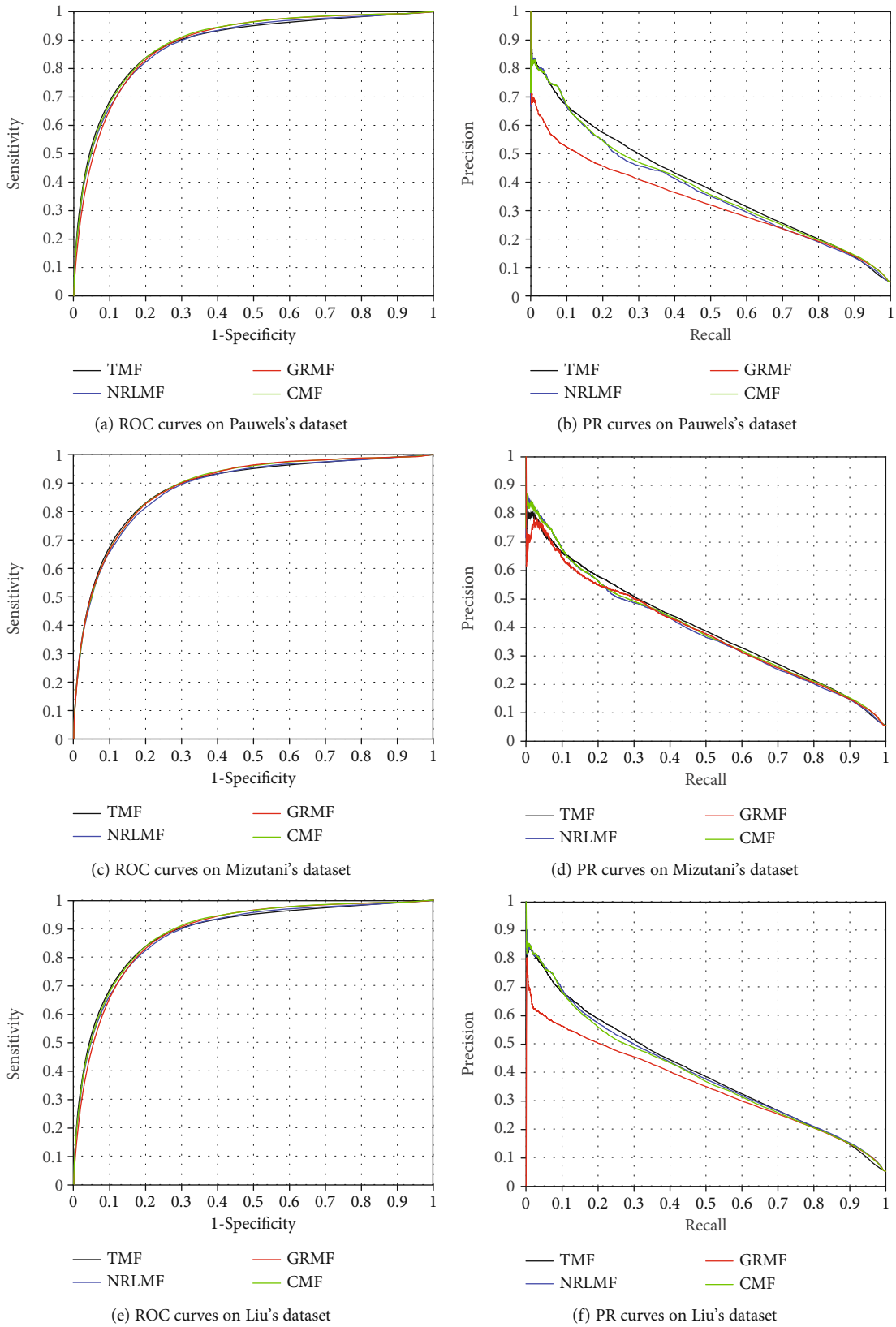
Datasets	Methods	AUPR	AUC
Pauwels	CMF*	0.382 ± 0.006	0.894 ± 0.004
	GRMF*	0.358 ± 0.008	0.883 ± 0.005
	NRLMF*	0.374 ± 0.007	0.886 ± 0.004
	Our method	0.392 ± 0.008	0.889 ± 0.004
Mizutani	CMF*	0.395 ± 0.005	0.889 ± 0.004
	GRMF*	0.392 ± 0.008	0.890 ± 0.006
	NRLMF*	0.390 ± 0.006	0.882 ± 0.005
	Our method	0.399 ± 0.013	0.886 ± 0.003
Liu	CMF*	0.393 ± 0.007	0.894 ± 0.005
	GRMF*	0.379 ± 0.008	0.895 ± 0.006
	NRLMF*	0.398 ± 0.006	0.897 ± 0.004
	Our method	0.401 ± 0.015	0.891 ± 0.004

*Results are derived from [18].

3.7. Running Time. We evaluate the performance for predictive models of running time. The results of test are listed in Table 9. The running time of CMF is less than our method (TMF), LGC, GRMF, and NRLMF on Pauwels's dataset (910 seconds), Mizutani's dataset (757 seconds), and Liu's dataset (846 seconds). TMF costs 977, 873, and 929 seconds, which are less than the ensemble model [26].

4. Conclusion and Discussion

In this study, we develop a Triple Matrix Factorization-based model to predict the associations between drugs and side effect terms. In drug space, several kernels are constructed from the chemical substructure fingerprint and known side effect-associated subnet. The side effect kernels are built from the known drug-associated subnet. The kernel functions include GIP, COS, Corr, and MI. Above kernels are combined by KTA-MKL in drug and side effect space, respectively. The integrated kernel matrices (including drug and side effect) are Low Rank Approximation in the TMF model. Our model (TMF) is tested on three benchmark datasets of drug-side effect association. Compared with other excellent methods, TMF achieves the best results (5-CV) on Pauwels's dataset (AUPR: 0.677), Mizutani's dataset (AUPR: 0.685), and Liu's dataset (AUPR: 0.680), respectively. In addition, our model is also compared with CMF, GRMF, and NRLMF under 5 local CV. The best AUPRs are achieved on Pauwels's dataset (AUPR: 0.392), Mizutani's dataset (AUPR: 0.399), and Liu's dataset (AUPR: 0.401). However, our method does not consider the topological relationship of drugs or side



(a) ROC curves on Pauwels's dataset

(b) PR curves on Pauwels's dataset

(c) ROC curves on Mizutani's dataset

(d) PR curves on Mizutani's dataset

(e) ROC curves on Liu's dataset

(f) PR curves on Liu's dataset

Figure 6: The ROC and PR curves of different methods via 5 local CV.

Table 7: Top 10 ranks of predictive side effects for drug caffeine.

Side effect	Score	Ranks	Confirmed
Diarrhea	0.3992	1	Yes
Diabetic neuropathy	0.3893	2	Yes
Varicocele	0.3844	3	Yes
Gynecomastia	0.3815	4	Yes
Conjunctivitis	0.3794	5	Yes
Telangiectasia	0.3737	6	No
Lump	0.3663	7	Yes
Dyskinesia	0.3638	8	No
Palpitations	0.3632	9	No
Fecal incontinence	0.3563	10	Yes

Table 8: Top 10 ranks of predictive side effects for drug captopril.

Side effect	Score	Ranks	Confirmed
Diarrhea	0.4150	1	No
Diabetic neuropathy	0.4043	2	Yes
Varicocele	0.4004	3	Yes
Conjunctivitis	0.3973	4	Yes
Gynecomastia	0.3938	5	Yes
Myoglobinuria	0.3885	6	No
Esophageal varices	0.3854	7	Yes
Lump	0.3806	8	Yes
Palpitations	0.3770	9	No
Eclampsia	0.3674	10	Yes

Table 9: The running time (seconds) via 5-fold Cross-Validation.

Model	Pauwels	Mizutani	Liu
Our method	977	873	929
LGC [18]	1290	1170	1211
CMF [18]	910	757	846
GRMF [18]	1360	1175	1282
NRLMF [18]	1966	1250	1911
Ensemble model [26]	4330	2715	3611

effects. In the future, a graph- or hypergraph-embedded MF-based model will be developed to improve the predictive performance of drug-side effect association.

Data Availability

The datasets, codes and corresponding results are available at https://figshare.com/s/10ee9c07123304a0ef82.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by a grant from the National Science Foundation of China (NSFC 61772362, 61902271, and 61972280) and the Natural Science Research of Jiangsu Higher Education Institutions of China (19KJB520014). The authors also thank professor Wen Zhang for kindly providing the datasets on his website.

References

[1] E. Pauwels, V. Stoven, and Y. Yamanishi, "Predicting drug side-effect profiles: a chemical fragment-based approach," *BMC Bioinformatics*, vol. 12, no. 1, pp. 169–181, 2011.

[2] L. Jiang, Y. Xiao, Y. Ding, J. Tang, and F. Guo, "FKL-Spa-LapRLS: an accurate method for identifying human microRNA-disease association," *BMC Genomics*, vol. 19, Supplement 10, p. 911, 2018.

[3] X. Zeng, L. Liu, L. Lü, and Q. Zou, "Prediction of potential disease-associated microRNAs using structural perturbation method," *Bioinformatics*, vol. 34, no. 14, pp. 2425–2432, 2018.

[4] Q. Zhao, Y. Yang, G. Ren, E. Ge, and C. Fan, "Integrating bipartite network projection and KATZ measure to identify novel circRNA-disease associations," *IEEE Transactions on NanoBioscience*, vol. 18, no. 4, pp. 578–584, 2019.

[5] C. Jia, Y. Zuo, and Q. Zou, "O-GlcNAcPRED-II: an integrated classification algorithm for identifying O-GlcNAcylation sites based on fuzzy undersampling and a K-means PCA oversampling technique," *Bioinformatics*, vol. 34, no. 12, pp. 2029–2036, 2018.

[6] L. Wei, S. Luan, L. A. E. Nagai, R. Su, and Q. Zou, "Exploring sequence-based features for the improved prediction of DNA N4-methylcytosine sites in multiple species," *Bioinformatics*, vol. 35, no. 8, pp. 1326–1333, 2019.

[7] Q. Zou, P. Xing, L. Wei, and B. Liu, "Gene2vec: gene subsequence embedding for prediction of mammalian N6-methyladenosine sites from mRNA," *RNA*, vol. 25, no. 2, pp. 205–218, 2019.

[8] Y. Ding, J. Tang, and F. Guo, "Protein crystallization identification via fuzzy model on linear neighborhood representation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 1, 2019.

[9] Y. Wang, Y. Ding, J. Tang, Y. Dai, and F. Guo, "CrystalM: a multi-view fusion approach for protein crystallization prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 1, 2019.

[10] H. Wang, Y. Ding, J. Tang, and F. Guo, "Identification of membrane protein types via multivariate information fusion with Hilbert–Schmidt independence criterion," *Neurocomputing*, vol. 383, pp. 257–269, 2020.

[11] Y. Shen, Y. Ding, J. Tang, Q. Zou, and F. Guo, "Critical evaluation of web-based prediction tools for human protein subcellular localization," *Briefings in Bioinformatics*, 2019.

[12] L. Wei, Y. Ding, R. Su, J. Tang, and Q. Zou, "Prediction of human protein subcellular localization using deep learning," *Journal of Parallel and Distributed Computing*, vol. 117, pp. 212–217, 2018.

[13] B. Liu, S. Jiang, and Q. Zou, "HITS-PR-HHblits: protein remote homology detection by combining PageRank and hyperlink-induced topic search," *Briefings in Bioinformatics*, vol. 21, no. 1, pp. 298–308, 2020.

[14] K. Qu, F. Guo, X. Liu, Y. Lin, and Q. Zou, "Application of machine learning in microbiology," *Frontiers in Microbiology*, vol. 10, p. 827, 2019.

[15] X. Ru, L. Li, and Q. Zou, "Incorporating distance-based top-n-gram and random forest to identify electron transport proteins," *Journal of Proteome Research*, vol. 18, no. 7, pp. 2931–2939, 2019.

[16] Y. Ding, J. Tang, and F. Guo, "Identification of drug-target interactions via fuzzy bipartite local model," *Neural Computing and Applications*, 2019.

[17] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via multiple information integration with centered kernel alignment," *Neurocomputing*, vol. 325, pp. 211–224, 2019.

[18] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via semi-supervised model and multiple kernel learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2619–2632, 2019.

[19] Y. Ding, J. Tang, and F. Guo, "Predicting protein-protein interactions via multivariate mutual information of protein sequences," *BMC Bioinformatics*, vol. 17, no. 1, p. 398, 2016.

[20] Y. Ding, J. Tang, and F. Guo, "Identification of protein-protein interactions via a novel matrix-based sequence representation model with amino acid contact information," *International Journal of Molecular Sciences*, vol. 17, no. 10, p. 1623, 2016.

[21] H. Liu, G. Ren, H. Chen, Q. Liu, Y. Yang, and Q. Zhao, "Predicting lncRNA–miRNA interactions based on logistic matrix factorization with neighborhood regularized," *Knowledge-Based Systems*, vol. 191, p. 105261, 2020.

[22] Y. Yamanishi, E. Pauwels, and M. Kotera, "Drug side-effect prediction based on the integration of chemical and biological spaces," *Journal of Chemical Information and Modeling*, vol. 52, no. 12, pp. 3284–3292, 2012.

[23] F. Cheng, W. Li, X. Wang et al., "Adverse drug events: database construction and in silico prediction," *Journal of Chemical Information and Modeling*, vol. 53, no. 4, pp. 744–752, 2013.

[24] S. Mizutani, E. Pauwels, V. Stoven, S. Goto, and Y. Yamanishi, "Relating drug-protein interaction network with drug side effects," *Bioinformatics*, vol. 28, no. 18, pp. i522–i528, 2012.

[25] M. Liu, Y. Wu, Y. Chen et al., "Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs," *Journal of the American Medical Informatics Association*, vol. 19, no. 1, pp. 28–35, 2012.

[26] W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, and W. Xiao, "Predicting potential side effects of drugs by recommender methods and ensemble learning," *Neurocomputing*, vol. 173, no. P3, pp. 979–987, 2016.

[27] W. Zhang, F. Liu, L. Luo, and J. Zhang, "Predicting drug side effects by multi-label learning and ensemble learning," *BMC Bioinformatics*, vol. 16, no. 1, pp. 365–375, 2015.

[28] W. Zhang, X. Yue, F. Liu, Y. Chen, S. Tu, and X. Zhang, "A unified frame of predicting side effects of drugs by using linear neighborhood similarity," *BMC Systems Biology*, vol. 11, no. 6, pp. 23–34, 2017.

[29] Y. Liu, M. Wu, C. Miao, P. Zhao, and X.-L. Li, "Neighborhood regularized logistic matrix factorization for drug-target interaction prediction," *PLoS Computational Biology*, vol. 12, no. 2, article e1004760, 2016.

[30] X. Zheng, H. Ding, H. Mamitsuka, and S. Zhu, "Collaborative matrix factorization with multiple similarities for predicting drug-target interactions," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1025–1033, 2013.

[31] A. Ezzat, P. Zhao, M. Wu, X.-L. Li, and C.-K. Kwoh, "Drug-target interaction prediction with graph regularized matrix factorization," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, no. 3, pp. 646–656, 2016.

[32] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori, "Gaussian interaction profile kernels for predicting drug–target interaction," *Bioinformatics*, vol. 27, no. 21, pp. 3036–3043, 2011.

[33] A. C. A. Nascimento, R. B. C. Prudêncio, and I. G. Costa, "A multiple kernel learning algorithm for drug-target interaction prediction," *BMC Bioinformatics*, vol. 17, no. 1, pp. 46–61, 2016.

[34] T. van Laarhoven and E. Marchiori, "Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile," *Plos One*, vol. 8, no. 6, article e66952, 2013.

[35] W. Zhang, Y. Chen, and D. Li, "Drug-target interaction prediction through label propagation with linear neighborhood information," *Molecules*, vol. 22, no. 12, pp. 2056–2069, 2017.

[36] N. J. Cerf and C. Adami, "Information theory of quantum entanglement and measurement," *Physica D Nonlinear Phenomena*, vol. 120, no. 1-2, pp. 62–81, 1998.

[37] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[38] N. Cristianini, J. Kandola, A. Elisseeff, and J. Shawe-Taylor, "On kernel-target alignment," *Advances in Neural Information Processing Systems*, vol. 179, no. 5, pp. 367–373, 2001.

[39] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine Learning Research*, vol. 5, no. 1, pp. 27–72, 2002.

[40] J.-Y. Shi, A.-Q. Zhang, S.-W. Zhang, K.-T. Mao, and S.-M. Yiu, "A unified solution for different scenarios of predicting drug-target interactions via triple matrix factorization," *BMC Systems Biology*, vol. 12, no. S9, Supplement 9, p. 136, 2018.

[41] D. S. Wishart, "Drugbank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic Acids Research*, vol. 34, no. Database issue, pp. 668–672, 2006.

[42] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, and P. Bork, "A side effect resource to capture phenotypic effects of drugs," *Molecular Systems Biology*, vol. 6, no. 1, p. 343, 2010.

[43] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, "KEGG for representation and analysis of molecular networks involving diseases and drugs," *Nucleic Acids Research*, vol. 38, Supplement 1, pp. D355–D360, 2010.

[44] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant, "PubChem: a public information system for analyzing bioactivities of small molecules," *Nucleic Acids Research*, vol. 37, no. Web Server, pp. W623–W633, 2009.

[45] Q. Li, T. Cheng, Y. Wang, and S. H. Bryant, "PubChem as a public resource for drug discovery," *Drug Discovery Today*, vol. 15, no. 23-24, pp. 1052–1057, 2010.