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 Guo3

1The Hemodialysis Center, The Affiliated Wuxi People’s Hospital of Nanjing Medical University, 214000 Wuxi, China
2School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu 215009, China
3School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin 300072, China
4Department 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_dyy@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

Copyright © 2020 Xiaoyi Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 IncRNA-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
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 × m adjacent matrix \( Y \in \mathbb{R}^{n \times m} \). \( D = \{d_1, d_2, \cdots, d_n\} \) and \( S = \{s_1, s_2, \cdots, s_m\} \) are the drug and side effect sets, respectively. \( Y_{ij} = 1 \) 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:

\[
K_{\text{GIP-link}}(d_i, d_k) = \exp \left(-\gamma \| \mathbf{pr}_{d_i} - \mathbf{pr}_{d_k} \|_2^2 \right),
\]

where \( \gamma \) is the bandwidth of the Gaussian kernel. \( \gamma \) 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:

\[
K_{\text{COS-link}}(d_i, d_k) = \frac{\mathbf{pr}_{d_i}^T \mathbf{pr}_{d_k}}{\| \mathbf{pr}_{d_i} \| \| \mathbf{pr}_{d_k} \|}.
\]
A and side effect spaces, 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 $K_d$ and $K_s$ can be obtained as follows:

$$
K_d^* = \sum_{p=1}^{k_d} \beta_{p,d} K_{p,d}, \quad K_p,d \in \mathbb{R}^{m_d},
$$

$$
K_s^* = \sum_{q=1}^{k_s} \beta_{q,s} K_{q,s}, \quad K_{q,s} \in \mathbb{R}^{m_s},
$$

where $\beta_{p,d} = \{\beta_{1,d}, \beta_{2,d}, \ldots, \beta_{k_d,d}\}$ and $\beta_{q,s} = \{\beta_{1,s}, \beta_{2,s}, \ldots, \beta_{k_s,s}\}$ 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):

$$
A(P, Q) = \frac{\langle P, Q \rangle_F}{\|P\|_F \|Q\|_F},
$$

where $\|P\|_F = \sqrt{\text{Trace}(P^T P)}$ denotes the Frobenius norm. $(P, Q)_F = \text{Trace}(P^T Q)$ 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:

$$
\beta_{p,d} = \frac{A(K_{p,d}, K_{\text{ideal},d})}{\sum_{k=1}^{k_d} A(K_{k,d}, K_{\text{ideal},d})}, \quad p = 1, 2, \ldots, k_d,
$$

$$
\beta_{q,s} = \frac{A(K_{q,s}, K_{\text{ideal},s})}{\sum_{k=1}^{k_s} A(K_{k,s}, K_{\text{ideal},s})}, \quad q = 1, 2, \ldots, k_s,
$$

where $K_{\text{ideal},d} = Y_{\text{train}} Y_{\text{train}}^T \in \mathbb{R}^{m_d}$ and $K_{\text{ideal},s} = Y_{\text{train}}^T Y_{\text{train}} \in \mathbb{R}^{m_s}$ 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:

$$
K_d^* \approx AA^T, \quad A \in \mathbb{R}^{m_d},
$$

$$
K_s^* \approx BB^T, \quad B \in \mathbb{R}^{m_s},
$$

where $A$ and $B$ 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:

$$
\min \ J(\Theta) = \|Y_{\text{train}} - A \Theta B^T\|_F^2 + \lambda \|\Theta\|_F^2,
$$

where $\Theta \in \mathbb{R}^{r \times r}$ is the bijection matrix. $\lambda$ is the regularization coefficient of $\Theta$. In our study, $\lambda$ is set as 1.
Let \( \frac{\partial J}{\partial \Theta} = 0 \), so we can obtain functions as follows:

\[
\frac{\partial}{\partial \Theta} \left( \|Y_{\text{train}} - A\Theta B^T\|^2_F + \lambda \|\Theta\|^2_F \right) = 0,
\]

(10)

\[
-2A^T(Y_{\text{train}} - A\Theta B^T)B + 2\lambda \Theta = 0,
\]

(11)

\[
A^T A\Theta + \lambda \Theta (B^T B)^{-1} = A^T Y_{\text{train}} B (B^T B)^{-1},
\]

(12)

\[
A^T A\Theta + \lambda \Theta (B^T B)^{-1} = A^T Y_{\text{train}} (B^T)^{-1},
\]

(13)

\[
A^T A\Theta + \lambda \Theta (B^T B)^{-1} = A^T Y_{\text{train}} (B^T)^{-1},
\]

(14)

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

\[
Y^* = A\Theta B^T.
\]

(15)
three types of datasets are employed in our study. They are collected from the DrugBank [41], SIDER Effect Resource [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 the whole matrix. 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 \( K_{\text{GIP-link},d} \) and \( K_{\text{GIP-link},s} \) have better performance on Pauwels’s dataset (AUPR: 0.6557, AUC: 0.9398). The best parameters for the benchmark dataset are shown in Table 2. In this study, we use 5-fold Cross-Validation (5-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.

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

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

<table>
<thead>
<tr>
<th>Models</th>
<th>Pauwels’s dataset</th>
<th>Mizutani’s dataset</th>
<th>Liu’s dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUPR</td>
<td>AUC</td>
<td>AUPR</td>
</tr>
<tr>
<td>( K_{\text{GIP-chem},d} ) &amp; ( K_{\text{GIP-link},s} )</td>
<td>0.4420</td>
<td>0.8950</td>
<td>0.4735</td>
</tr>
<tr>
<td>( K_{\text{GIP-chem},d} ) &amp; ( K_{\text{GIP-link},s} )</td>
<td>0.4892</td>
<td>0.8994</td>
<td>0.5343</td>
</tr>
<tr>
<td>( K_{\text{GIP-chem},s} ) &amp; ( K_{\text{GIP-link},d} )</td>
<td>0.4994</td>
<td>0.8981</td>
<td>0.5217</td>
</tr>
<tr>
<td>( K_{\text{GIP-chem},d} ) &amp; ( K_{\text{GIP-link},s} )</td>
<td>0.4978</td>
<td>0.9079</td>
<td>0.5591</td>
</tr>
<tr>
<td>( K_{\text{GIP-link},d} ) &amp; ( K_{\text{GIP-link},s} )</td>
<td>0.6254</td>
<td>0.9300</td>
<td>0.6623</td>
</tr>
<tr>
<td>( K_{\text{GIP-chem},d} ) &amp; ( K_{\text{GIP-link},s} )</td>
<td>0.5861</td>
<td>0.9035</td>
<td>0.6324</td>
</tr>
<tr>
<td>( K_{\text{GIP-chem},s} ) &amp; ( K_{\text{GIP-link},d} )</td>
<td>0.5833</td>
<td>0.8999</td>
<td>0.6123</td>
</tr>
<tr>
<td>( K_{\text{GIP-chem},d} ) &amp; ( K_{\text{GIP-link},s} )</td>
<td>0.6557</td>
<td>0.9428</td>
<td>0.6615</td>
</tr>
<tr>
<td>Mean weighted( ^{a} )</td>
<td>0.6598</td>
<td>0.9353</td>
<td>0.6724</td>
</tr>
<tr>
<td>KTA-MKL( ^{c} )</td>
<td>0.6765</td>
<td>0.9434</td>
<td>0.6847</td>
</tr>
</tbody>
</table>

\( ^{a} \)The TMF uses the drug fingerprint and drug profile for side effects. \( ^{c} \)The TMF uses the drug fingerprint, 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.
Figure 5: Continued.
the KTA-MKL model achieves the best results on Pauwels’ 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 $K_{MI-link,a}$ and $K_{MI-link,z}$ 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’ dataset (AUPR: 0.677), Mizutani’s dataset (AUPR: 0.680), 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.

### 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...
The results are listed in Table 6 and Figure 6. 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.

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 effects.
Figure 6: The ROC and PR curves of different methods via 5 local CV.
The authors declare that they have no conflicts of interest.

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

<table>
<thead>
<tr>
<th>Side effect</th>
<th>Score</th>
<th>Ranks</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diarrhea</td>
<td>0.3992</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Diabetic neuropathy</td>
<td>0.3893</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Varicocele</td>
<td>0.3844</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>Gynecomastia</td>
<td>0.3815</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>Conjunctivitis</td>
<td>0.3794</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>Telangiectasia</td>
<td>0.3737</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>Lump</td>
<td>0.3663</td>
<td>7</td>
<td>Yes</td>
</tr>
<tr>
<td>Dysthnesia</td>
<td>0.3638</td>
<td>8</td>
<td>No</td>
</tr>
<tr>
<td>Palpitations</td>
<td>0.3632</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>Fecal incontinence</td>
<td>0.3563</td>
<td>10</td>
<td>Yes</td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Pauwels</th>
<th>Mizutani</th>
<th>Liu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our method</td>
<td>977</td>
<td>873</td>
<td>929</td>
</tr>
<tr>
<td>LGC [18]</td>
<td>1290</td>
<td>1170</td>
<td>1211</td>
</tr>
<tr>
<td>CMF [18]</td>
<td>910</td>
<td>757</td>
<td>846</td>
</tr>
<tr>
<td>GRMF [18]</td>
<td>1360</td>
<td>1175</td>
<td>1282</td>
</tr>
<tr>
<td>NRLMF [18]</td>
<td>1966</td>
<td>1250</td>
<td>1911</td>
</tr>
<tr>
<td>Ensemble model [26]</td>
<td>4330</td>
<td>2715</td>
<td>3611</td>
</tr>
</tbody>
</table>

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**


[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.


