Journal of Biomedical Informatics 88 (2018) 90–97

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Manifold regularized matrix factorization for drug-drug interaction prediction

Wen Zhang[a,b,*], Yanlin Chen[c], Dingfang Li[c], Xiang Yue[d]

[a] College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
[b] School of Computer Science, Wuhan University, Wuhan 430072, China
[c] School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
[d] Department of Computer Science and Engineering, The Ohio State University, OH 43210, USA

ARTICLE INFO

Keywords:
Manifold regularization
Drug-drug interaction prediction
Matrix completion

ABSTRACT

Drug-drug interaction (DDI) prediction is one of the most important tasks in drug discovery. Prediction of potential DDIs helps to reduce unexpected side effects in the lifecycle of drugs, and is important for the drug safety surveillance. Here, we formulate the drug-drug interaction prediction as a matrix completion task, and project drugs in the interaction space into a low-dimensional space. We consider drug features, i.e., substructures, targets, enzymes, transporters, pathways, indications, side effects, and off side effects, to calculate drug-drug similarities, and assume them as manifolds in feature spaces. In this paper, we present a novel computational method named "Manifold Regularized Matrix Factorization" (MRMF) to predict potential drug-drug interactions, by introducing the drug feature-based manifold regularization into the matrix factorization. In the computational experiments, the MRMF models, which utilize known drug-drug interactions and the drug feature-based manifold, produce the area under precision-recall curves (AUPR) up to 0.7963. We test manifold regularizations based on different drug features, and the MRMF models can produce robust performances. Compared with other state-of-the-art methods, the MRMF models can produce better performances in the cross validation and case study. The manifold regularization is the critical factor for the high-accuracy performances of our method. MRMF is promising and effective for the prediction of drug-drug interactions.

1. Introduction

Drug-drug interaction (DDI) prediction is an important task in drug discovery, which attracts great attention from both academy and industry [1]. Drug combinations have been a common therapy for elderly patients and cancer patients [2,3], but co-prescription of drugs may lead to a high risk of DDIs. Adverse drug-drug interactions usually cause adverse reactions and even the withdrawal of drugs [4–6]. Therefore, the DDI prediction not only helps to reduce health risks, but also prompts safe drug co-prescription.

Traditional DDI prediction methods include in vitro methods and clinical trials, but screening DDI candidates are time-consuming and costly. In recent years, machine-learning methods have been introduced to the DDI prediction. Here, we roughly categorize these methods as classification-based methods and similarity-based methods. Classification-based methods take DDI prediction as binary classification problems. Cami et al. [7] formulated presences or absences of DDI interactions between drugs as feature vectors, and then constructed a logistic regression model. Cheng et al. [8] proposed a heterogeneous network assisted inference (HNAI) method for the prediction of DDIs, which applied five prediction models (naive Bayes, decision tree, k-nearest neighbor, logistic regression and support vector machine) in the HNAI framework. Classification-based methods take interaction pairs as positive instances and other pairs as negative instances to train classification models. However, annotated non-interaction pairs may contain undetected or unobserved DDIs. Similarity-based methods assume that similar drugs may have same interactions. Gottlieb et al. [9] exploited seven different drug-drug similarity measures to infer potential DDIs. Vilar et al. [10,11] built two similarity-based models based on drug substructures and the interaction profile fingerprints, respectively. Li et al. [12] combined six types of drug-drug similarities into a likelihood ratio for predicting adverse effects of drug combinations. Park et al. [13] adopted random walk with restart algorithm to simulate signaling propagation from drug targets. Zhang et al. [14] proposed an

* Corresponding Author at: College of Informatics, Huazhong Agricultural University, Wuhan 430070, China.
E-mail addresses: zhangwen@whu.edu.cn, zhangwen@mail.hzau.edu.cn (W. Zhang), chenyanlin@whu.edu.cn (Y. Chen), dfli@whu.edu.cn (D. Li), tommy96@whu.edu.cn (X. Yue).

https://doi.org/10.1016/j.jbi.2018.11.005
Received 28 November 2017; Received in revised form 3 November 2018; Accepted 11 November 2018
Available online 13 November 2018
1532-0464/ © 2018 Elsevier Inc. All rights reserved.

Table 1
Summary of the benchmark dataset.

#	Features	Source	Dimension	#	Features	Source	Dimension
1	Substructures	Pubchem	881	5	Pathways	KEGG	253
2	Targets	DrugBank	780	6	Indications	SIDER	4897
3	Transporters	DrugBank	78	7	Side effects	SIDER	4897
4	Enzymes	DrugBank	129	8	Off side effects	OFFSIDES	9496

integrative label propagation framework by integrating substructures, side effects and off-label side effects. Zhang et al. [15] comprehensively studied a great number of similarity-based predictors and network missing link predictors. Ferdousi et al. [16] utilized drug biological information and the Russell-Rao similarity measure to discover new DDIs.

In this paper, we formulate the DDI prediction problem as a matrix completion task. The known DDIs can be represented as an adjacent matrix. Entries for observed interactions have the values "1"; other entries have the values "0", representing drug-drug pairs without known interaction information. Matrix completion is to estimate values for entries without known interaction information. The matrix factorization techniques are usually used for the matrix completion tasks, and solve many bioinformatics tasks [17,18]. In machine learning, traditional matrix factorization methods include singular value decomposition (SVD) [19], nonnegative matrix factorization (NMF) [20] and probabilistic matrix factorization (PMF) [21]. SVD decomposes an interaction matrix into three matrices, which provide the good approximation to the original matrix. NMF aims to decompose a non-negative matrix to two low-rank non-negative matrices. PMF decomposes a matrix to two low-rank matrices, and assume that two matrices follow Gaussian prior distribution. However, these traditional matrix factorization methods usually utilize the data structures of the matrix for predictions, but ignore background information of research topics.

Recently, studies have revealed that data usually lies on (or near to) a manifold, and nearby points are likely to have similar embedding, namely locally invariance. Many manifold learning algorithms are proposed to detect the underlying manifold structures, such as Locally Linear Embedding (LLE) [22], ISOMAP [23] and Laplacian Eigenmap [24]. These algorithms have demonstrated the usefulness of manifold structures and locally invariance.

Here, we incorporate the background information of drugs into the matrix factorization. We consider drug features, i.e., substructures, targets, enzymes, transporters, pathways, indications, side effects and off side effects, to calculate drug-drug similarities, and then assume them as the manifolds. We project drugs in the interaction space into a low-dimensional space, and introduce the manifold regularization. The manifold regularization is able to learn a manifold (in the low-dimensional space) on which the projected drugs are assumed to lie, and preserve the locally geometrical structures of the drug feature-based manifold. Therefore, we propose a method named "Manifold Regularized Matrix Factorization" (MRMF) to predict DDIs. In the computational experiments, the MRMF models, which utilize known drug-drug interactions and drug feature-based manifolds, produce the area under precision-recall curves (AUPR) up to 0.7963. We test manifold regularizations based on different drug features, and the MRMF models can produce robust performances. Compared with other state-of-the-art methods, MRMF models can produce better performances in the cross validation and case study. The manifold regularization is the critical factor for the high-accuracy performances of our method.

This paper has three contributions. We investigate the intrinsic relationship between drugs and assume drug feature-based manifold should be maintained, and then introduce manifold regularization and propose a novel matrix factorization method (MRMF) to predict potential DDIs. We develop an effective algorithm to minimize the objective function of MRMF. We consider different drug features and similarity measures to calculate different manifold regularization terms, and then construct MRMF models. Experimental results demonstrate that MRMF models can produce high-accuracy and robust performances.

2. Materials and method

2.1. Datasets

Several databases provide information about drugs and drug-drug interactions (DDIs). TWOSIDES [25] is a public database which contains drug-drug interaction-induced side effects. In this work, we use drug-drug interactions from the unsafe co-prescription in TWOSIDES. Pubchem database [26,27] can provide chemical structures. DrugBank database [28–31] is a comprehensive bioinformatics resource that includes targets, transporters and enzymes of drugs. KEGG database [32] is a collection of protein pathways that associated with drug targets. SIDER database [33] contains side effects and indications that compiled from public documents and package inserts. OFFSIDES database has the off side effects of drugs [25].

In our previous work [15], we complied a benchmark dataset from above databases. As summarized in Table 1, the dataset contains 48,584 DDIs between 548 drugs as well as 8 drug features (i.e. substructures, targets, enzymes, transporters, pathways, indications, side effects and off side effects). By using a feature, a drug can be represented as a feature vector. For example, a drug can be represented as an 881-dimensional substructure vector, whose dimension describe the presence or absence of corresponding substructures with value 1 or 0.

2.2. Drug-drug similarity

Three popular similarity measures, namely Jaccard similarity, cosine similarity and Gauss similarity are adopted in machine learning [34–36]. The similarity reflects pairwise relation of data points, and has been widely used in bioinformatics [37–41]. We denote feature vectors of drug d_i and drug d_j as $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, and three similarity measures are defined as follows.

Jaccard similarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is

$$S_{Jar}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{N_{11}}{N_{10} + N_{01} + N_{11}} \quad (1)$$

where N_{11} is the total number of elements where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ both have a value 1, N_{10} is the total number of elements where $\boldsymbol{x}_i$ has a value 1 and $\boldsymbol{x}_j$ has a value 0, and N_{01} is the total number of elements where $\boldsymbol{x}_i$ has a value 0 and $\boldsymbol{x}_j$ has a value 1.

Cosine similarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is

$$S_{cos}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{\boldsymbol{x}_i \cdot \boldsymbol{x}_j}{\|\boldsymbol{x}_i\|_2 \|\boldsymbol{x}_j\|_2} \quad 2$$

where $\|\cdot\|_2$ is the Euclidean norm.

Gauss similarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is

$$S_{Gau}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\sigma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2) \quad 3$$

where $\sigma(>0)$ is the bandwidth parameter, and we set $\sigma = 1/(\sum_{i=1}^{m} |\boldsymbol{x}_i|/m)$ as [42].

2.3. Manifold regularized matrix factorization

Given known drug-drug interactions (DDIs) between m drugs, the interactions can be represented as an $m \times m$ interaction matrix $A = (a_{ij})$, where $a_{ij} = 1$ if the ith drug interacts with the jth drug, and $a_{ij} = 0$ otherwise. In the matrix, the case $a_{ij} = 0$ means that interactions are not observed currently and this entry may contain ambiguous information. After formulating known interactions as the interaction matrix, we take the DDI prediction problem as a matrix completion task, which makes predictions for the entries without observed interactions.

For the matrix completion, we decompose the interaction matrix A into two low-rank matrices $X \in \mathbb{R}^{m\times k}$ and $Y \in \mathbb{R}^{m\times k}$, where $k \in \mathbb{N}$ is the dimension of latent feature vectors. In this way, drugs are projected from the high-dimensional interaction space into a low-dimensional space. The ideal matrix factorization is that $X \times Y^T$ is approximated to the interaction matrix A. The objective function is

$$\underset{X,Y}{\operatorname{argmin}} L = \frac{1}{2}||A - XY^T||_F^2 + \frac{\lambda}{2}(||X||_F^2 + ||Y||_F^2) = \frac{1}{2}\sum_{ij}(a_{ij} - \boldsymbol{x}_i\boldsymbol{y}_j^T)^2 + \frac{\lambda}{2}\left(\sum_i ||\boldsymbol{x}_i||_2^2 + \sum_j ||\boldsymbol{y}_j||_2^2\right) \quad (4)$$

where $||\cdot||_F$ is the Frobenius norm, $||\cdot||_2$ is the Euclidean norm, $\boldsymbol{x}_i$ is the ith row of X and $\boldsymbol{y}_j$ is the jth row of Y, $\lambda > 0$ is Tikhonov regularization parameter.

A manifold [43] is a topological space that locally resembles Euclidean space near each point. In other words, each point of a n-dimensional manifold has a neighborhood that is homeomorphic to the n-dimensional Euclidean space. As mentioned in Section 2.2, we can calculate drug-drug similarities in feature spaces, and take the similarities as manifolds in feature spaces. Then, we assume that drugs may keep manifolds, and introduce the manifold regularization, which means that the drugs approximately keep manifolds in the low-dimensional space. The manifold regularizations for drugs in the low-dimensional space are

$$L_{reg}^{row} = \frac{1}{2}\sum_{ij}||\boldsymbol{x}_i - \boldsymbol{x}_j||_2^2 s_{ij},\ L_{reg}^{col} = \frac{1}{2}\sum_{ij}||\boldsymbol{y}_i - \boldsymbol{y}_j||_2^2 s_{ij} \quad (5)$$

where s_{ij} is the similarity between the ith drug and the jth drug. By combining (4) and (5), we propose the objective function of the Manifold Regularized Matrix Factorization (MRMF) as follows:

$$\begin{aligned}\underset{X,Y}{\operatorname{argmin}} L_{mix} &= L + \mu(L_{reg}^{row} + L_{reg}^{col}) \\ &= \tfrac{1}{2}\textstyle\sum_{ij}(a_{ij} - \boldsymbol{x}_i\boldsymbol{y}_j^T)^2 + \tfrac{\lambda}{2}\left(\sum_i ||\boldsymbol{x}_i||_2^2 + \sum_j ||\boldsymbol{y}_j||_2^2\right) \\ &\quad + \tfrac{\mu}{2}\left(\textstyle\sum_{ij}||\boldsymbol{x}_i - \boldsymbol{x}_j||_2^2 s_{ij} + \sum_{ij}||\boldsymbol{y}_i - \boldsymbol{y}_j||_2^2 s_{ij}\right)\end{aligned} \quad (6)$$

where $\mu > 0$ is manifold regularization parameter.

2.4. Optimization algorithm

To estimate the latent feature matrices X and Y, we propose an alternating decent method which can minimize the objective function of MRMF effectively. The alternating descent method rotates between fixing $\boldsymbol{x}_i$ and fixing $\boldsymbol{y}_j$. When all $\boldsymbol{x}_i$ or $\boldsymbol{y}_j$ are fixed, the optimization problem in Eq. (6) is transformed as a least-squares problem, and thus we can estimate $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$. First of all, we initialize $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ randomly, and then calculate partial derivatives of objective function L_{mix} with respect to $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$,

$$\begin{aligned}\nabla_{\boldsymbol{x}_i} L_{mix} &= \sum_j(\boldsymbol{x}_i\boldsymbol{y}_j^T - a_{ij})\boldsymbol{y}_j + \lambda\boldsymbol{x}_i + \mu\left(\sum_j(\boldsymbol{x}_i - \boldsymbol{x}_j)s_{ij} - \sum_j(\boldsymbol{x}_j - \boldsymbol{x}_i)s_{ji}\right) \\ &= \boldsymbol{x}_i\left(\textstyle\sum_j \boldsymbol{y}_j^T\boldsymbol{y}_j + \lambda I + \mu\left(\sum_j s_{ij} + \sum_j s_{ji}\right)I\right) - \sum_j a_{ij}\boldsymbol{y}_j - \mu\sum_j(s_{ij} + s_{ji})\boldsymbol{x}_j\end{aligned} \quad (7)$$

$$\begin{aligned}\nabla_{\boldsymbol{y}_j} L_{mix} &= \sum_i(\boldsymbol{y}_j\boldsymbol{x}_i^T - a_{ij})\boldsymbol{x}_i + \lambda\boldsymbol{y}_j + \mu\left(\sum_i(\boldsymbol{y}_j - \boldsymbol{y}_i)s_{ji} - \sum_i(\boldsymbol{y}_i - \boldsymbol{y}_j)s_{ij}\right) \\ &= \boldsymbol{y}_j\left(\textstyle\sum_i \boldsymbol{x}_i^T\boldsymbol{x}_i + \lambda I + \mu\left(\sum_i s_{ji} + \sum_i s_{ij}\right)I\right) - \sum_i a_{ij}\boldsymbol{x}_i - \mu\sum_i(s_{ji} + s_{ij})\boldsymbol{y}_i\end{aligned} \quad (8)$$

where I is the $k \times k$ identity matrix. We further calculate the second derivatives of L_{mix} with respect to $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$,

$$\begin{aligned}\nabla_{\boldsymbol{x}_i}^2 L_{mix} &= \sum_j \boldsymbol{y}_j^T\boldsymbol{y}_j + \lambda I + \mu\left(\sum_j s_{ij} + \sum_j s_{ji}\right)I = Y^TY + \lambda I \\ &\quad + \mu\left(\sum_j s_{ij} + \sum_j s_{ji}\right)I\end{aligned} \quad (9)$$

$$\begin{aligned}\nabla_{\boldsymbol{y}_j}^2 L_{mix} &= \sum_i \boldsymbol{x}_i^T\boldsymbol{x}_i + \lambda I + \mu\left(\sum_i s_{ji} + \sum_i s_{ij}\right)I \\ &= X^TX + \lambda I + \mu\left(\sum_i s_{ji} + \sum_i s_{ij}\right)I\end{aligned} \quad (10)$$

Note that symmetric matrices $\nabla_{\boldsymbol{x}_i}^2 L_{mix}$ and $\nabla_{\boldsymbol{y}_j}^2 L_{mix}$ are positive definite, we can develop alternating update rules of $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ by using Newton's method [44],

$$\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i - \nabla_{\boldsymbol{x}_i} L_{mix}(\nabla_{\boldsymbol{x}_i}^2 L_{mix})^{-1} \quad (11)$$

$$\boldsymbol{y}_j \leftarrow \boldsymbol{y}_j - \nabla_{\boldsymbol{y}_j} L_{mix}(\nabla_{\boldsymbol{y}_j}^2 L_{mix})^{-1} \quad (12)$$

and we rewrite the update rules (11) and (12) as follows,

$$\boldsymbol{x}_i \leftarrow \left(\sum_j a_{ij}\boldsymbol{y}_j + \mu\sum_j(s_{ij} + s_{ji})\boldsymbol{x}_j\right)\left(Y^TY + \lambda I + \mu\left(\sum_j s_{ij} + \sum_j s_{ji}\right)I\right)^{-1} \quad (13)$$

$$\boldsymbol{y}_j \leftarrow \left(\sum_i a_{ij}\boldsymbol{x}_i + \mu\sum_i(s_{ji} + s_{ij})\boldsymbol{y}_i\right)\left(X^TX + \lambda I + \mu\left(\sum_i s_{ji} + \sum_i s_{ij}\right)I\right)^{-1} \quad (14)$$

The update (13) and (14) will be repeated until latent feature matrices X and Y converges.

There are two advantages of the alternating descent method. First, our method utilizes Newton's method to solve the least-squares problem, which ensures that each update decreases Eq. (6) until convergence [44]. Although we have to calculate the inverse matrix for each latent feature vector in (13) and (14), the inverse matrices are constant matrices for $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ respectively, and they can be precomputed. More importantly, we can compute each $\boldsymbol{x}_i$ independently and compute each $\boldsymbol{y}_j$ independently, and this allows for massive parallelization of the optimization algorithm.

Finally, the prediction matrix is the product of two low-rank matrices X and Y, as:

$$A^p = XY^T \quad (15)$$

In the prediction matrix A^p, scores for entries without observed interactions indicate the probabilities of novel DDIs. The optimization

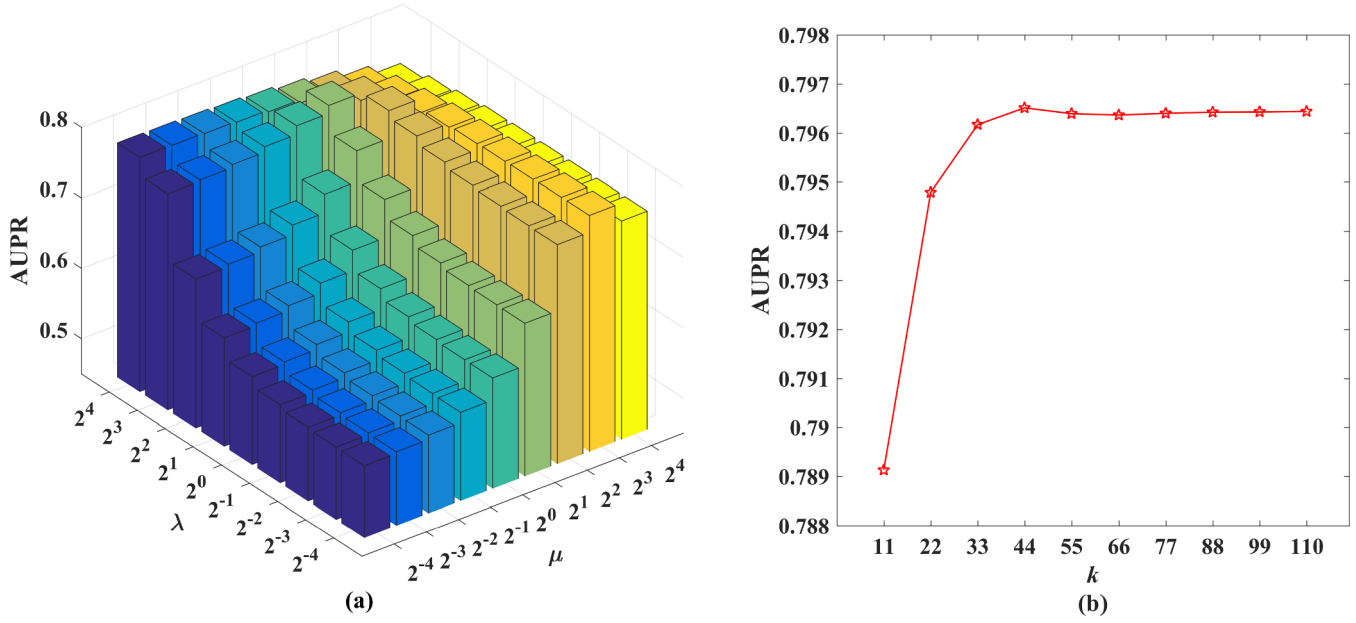

Fig. 1. The influences of parameters on MRMF models.

algorithm is described in Algorithm 1.

Algorithm1. Optimization algorithm of Manifold Regularized Matrix Factorization

input: drug-drug interaction matrix $A = (a_{ij}) \in \mathbb{R}^{m \times m}$;
drug similarity matrix $S = (s_{ij}) \in \mathbb{R}^{m \times m}$;
dimension of latent feature vector $k \in \mathbb{N}$;
Tikhonov regularization parameter $\lambda > 0$;
manifold regularization parameter $\mu > 0$.
output: prediction matrix A^p.
initialize drug latent matrices $X \in \mathbb{R}^{m \times k}$ and $Y \in \mathbb{R}^{m \times k}$ randomly
while X and Y have not converged **do**
  for each row vector $\boldsymbol{x}_i (1 \le i \le m)$ *of* X **do**
    $\boldsymbol{x}_i \leftarrow \left(\sum_j a_{ij}\boldsymbol{y}_j + \mu \sum_j (s_{ij} + s_{ji})\boldsymbol{x}_j\right)\left(Y^T Y + \lambda I + \mu\left(\sum_j s_{ij} + \sum_j s_{ji}\right)I\right)^{-1}$
  end for
  for each row vector $\boldsymbol{y}_j (1 \le j \le m)$ *of* Y **do**
    $\boldsymbol{y}_j \leftarrow (\sum_i a_{ij}\boldsymbol{x}_i + \mu \sum_i (s_{ji} + s_{ij})\boldsymbol{y}_i)(X^T X + \lambda I + \mu(\sum_i s_{ji} + \sum_i s_{ij})I)^{-1}$
end for
end while
calculate the prediction matrix $A^p = XY^T$

3. Results and discussion

3.1. Evaluation metrics

In the paper, we adopt 5-fold cross validation (5-CV) to evaluate prediction models. Specifically, known interactions are randomly divided into 5 subsets with equal size. Each subset is left out as the test set in turn, while remaining subsets are used as the training set to build prediction models. We repeat the procedure until each subset is ever tested. The results of prediction models are averaged over 20 independent runs of 5-CV to avoid bias of data split.

In this work, we use recall, precision, accuracy, F-measure, AUC and AUPR to evaluate prediction models:

$$\text{recall}=\frac{\text{TP}}{\text{TP+FN}}$$

$$\text{precision}=\frac{\text{TP}}{\text{TP+FP}}$$

$$\text{accuracy}=\frac{\text{TP+TN}}{\text{TP+FP+TN+FN}}$$

$$\text{F-measure}=\frac{2\times\text{recall}\times\text{precision}}{\text{recall+precision}}$$

where TP, FP, TN and FN are the numbers of true positives, false positives, true negatives and false negatives, respectively. The AUC and AUPR are the most popular evaluation metrics. AUC is the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate (TPR) versus the false positive rate (FPR). AUPR is the area under the precision-recall curve, which plots ratio of true positives among predicted positives for each recall rate. In the benchmark dataset, non-interaction pairs take a large proportion of all drug pairs, thus we adopt AUPR as the primary metric for it heavily punishes non-interactions.

3.2. Performances of MRMF models

3.2.1. Parameter settings and consensus of prediction matrices

Manifold Regularized Matrix Factorization (MRMF) method has three parameters: the dimension of latent feature vector k, the Tikhonov regularization parameter λ and the manifold regularization parameter μ. Three parameters may influence performances of MRMF models, and we discuss how to set the parameters in this section.

As mentioned in Section 2.1, the benchmark dataset contains eight drug features, which bring diverse information about drugs. Different features can lead to different manifold regularization terms. Among these features, substructures are usually considered as the most important information for drugs. Therefore, we use substructures to build MRMF models, and then study the influences of parameters. Here, we use 5-fold cross validation (5-CV) to evaluate our prediction models, and adopt AUPR as the evaluation metric.

In general, MRMF decomposes DDI matrix into two low-rank latent feature matrices, thus the dimension of latent feature vector k should be less than the row number of the interaction matrix. For this reason, we consider combinations of the following values: {2\%, 4\%, 6\%, 8\%, 10\%, 12\%, 14\%, 16\%, 18\%, 20\%} of the matrix's row number (drug number, $m = 548$) for k,

Table 2
The differences of prediction matrices based on different initial feature matrices.

F-norm	A_1^p A_1^p	A_2^p	A_3^p	A_4^p	A_5^p	A_6^p	A_7^p	A_8^p	A_9^p	A_{10}^p
A_1^p	0	3.8284	3.4059	3.8292	4.7877	3.9755	4.0696	4.3932	4.6944	3.6442
A_2^p	3.8284	0	3.8349	4.7469	5.4738	4.4292	5.0089	4.7787	4.581	3.8684
A_3^p	3.4059	3.8349	0	3.5452	4.6198	4.2028	3.9689	3.7429	4.0836	3.7849
A_4^p	3.8292	4.7469	3.5452	0	4.632	4.8914	4.4046	4.0214	4.2849	3.904
A_5^p	4.7877	5.4738	4.6198	4.632	0	5.2724	5.7368	4.7556	3.9766	4.1658
A_6^p	3.9755	4.4292	4.2028	4.8914	5.2724	0	5.0825	4.7529	5.2118	3.1552
A_7^p	4.0696	5.0089	3.9689	4.4046	5.7368	5.0825	0	4.0124	5.0255	4.8835
A_8^p	4.3932	4.7787	3.7429	4.0214	4.7556	4.7529	4.0124	0	3.819	4.2419
A_9^p	4.6944	4.581	4.0836	4.2849	3.9766	5.2118	5.0255	3.819	0	4.4103
A_{10}^p	3.6442	3.8684	3.7849	3.904	4.1658	3.1552	4.8835	4.2419	4.4103	0

$\{2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4\}$ for λ and μ.

Tikhonov regularization parameter λ and manifold regularization parameter μ influence the values of latent feature vectors, while the dimension of drug latent feature vector k is utilized to find the low-dimensional space where drugs are projected. Therefore, the dimension of latent feature vector k is irrelevant to regularization parameters λ and μ. First, we fix $k = 55(\approx 10\% \times 548)$, and build MRMF models by using different values for λ and μ. As shown in Fig. 1(a), suitable values, which are not very great or very small, are very important for the parameters λ and μ. In the experiments, the prediction models achieve the best AUPR values when $\lambda = 2^2$ and $\mu = 2^2$. Then, we fix $\lambda = 2^2$ and $\mu = 2^2$, and test the influence of the parameter k. The MRMF models are constructed based on different values for k, and their AUPR values are shown in Fig. 1(b). Clearly, when $k < 44(\approx 8\% \times 548)$, the greater values of k lead to better performances of MRMF models. The MRMF model produces the best AUPR scores when $k = 44$. Based on above discussion, we set $k = 44$, $\lambda = 2^2$ and $\mu = 2^2$ for MRMF in the following experiments.

Since MRMF is initialized with two low-rank matrices X and Y, we use 10 different sets of initial matrices to train models and test the consensus of the prediction matrices constructed by MRMF. Here, we denote 10 constructed prediction matrices as $A_i^P, i = 1, \cdots, 10$, and then we measure the differences between prediction matrices by using Frobenius norm $||A_i^P - A_j^P||_F$. Table 2 demonstrates the pairwise differences of prediction matrices. Considering that there are 300,304 entries for the 548,548× prediction matrices, the differences between prediction matrices are negligible, and MRMF produces consensus prediction matrices by using different initial matrices.

3.2.2. Performances of MRMF models

To construct MRMF models, we introduce the manifold regularization term calculated in the feature space. Since we consider eight drug features and three similarity measures, we can obtain different manifold regularization terms. Here, we build different MRMF models by combining different features and different similarity measures to test the robustness of our method, and all models are evaluated by 20 runs of 5-fold cross validation (5-CV) under the same conditions.

First, we test influences of the three similarity measures, i.e., Jaccard similarity, cosine similarity and Gauss similarity. We obtain three drug substructure similarity-based manifold regularization terms by using three similarity measures, and build three MRMF models. As shown in Table 3, MRMF models that based on Jaccard similarity, cosine similarity and Gauss similarity produce the nearly same AUPR values of 0.7958, 0.7958 and 0.7957, respectively. The results demonstrate that MRMF models are robust to similarity measures, and can produce high-accuracy performances. Since different similarity measures produce similar performances, the Jaccard similarity is adopted in the following study.

Table 3
Performances of MRMF models based on substructures and different similarity measures.

Similarities	AUPR	AUC	Recall	Precision	Accuracy	F-measure
Jaccard	0.7958	0.9585	0.6812	0.7722	0.9546	0.7237
Cosine	0.7958	0.9584	0.6806	0.7729	0.9546	0.7237
Gauss	0.7957	0.9584	0.6808	0.7726	0.9546	0.7236

Further, we discuss the influences of eight features, i.e., substructures, targets, enzymes, transporters, pathways, indications, side effects and off side effects. Here, we adopt Jaccard similarity to calculate drug-drug similarities based on different features, and then build MRMF models. The results in Table 4 show that MRMF models using different feature-based manifold regularization terms have good performances. Moreover, we can observe that different features can lead to different performances. Substructures usually have the influence on drug functions, and side effects and off side effects are usually considered to be relevant [45]. Hence, substructures, side effects and off side effects can produce better performances than other features.

Therefore, MRMF models can produce satisfying results by using different features and different similarity measures, indicating the robustness of the proposed method.

3.3. Comparison with state-of-the-art methods

We consider existing DDI prediction methods to make comparison. Vilar utilized interaction information of the most similar drugs, and proposed the substructure similarity-based model [10] and interaction profile fingerprint (also known as common neighbors, CN) similarity-based model [11]. Zhang [14] adopted the label propagation algorithm to build substructure similarity-based model, side effect similarity-based model and off side effect similarity-based model. We name these models as Vilar's substructure-based model, Vilar's CN index-based model, substructure-based label propagation model, side effect-based label propagation model and off side effect-based label propagation model. These prediction models are implemented according to our

Table 4
Performances of MRMF models based on Jaccard similarity and different features.

Features	AUPR	AUC	Recall	Precision	Accuracy	F-measure
Substructures	0.7958	0.9585	0.6812	0.7722	0.9546	0.7237
Targets	0.7769	0.9553	0.6830	0.7381	0.9511	0.7092
Transporters	0.7393	0.9450	0.6641	0.6911	0.9447	0.6771
Enzymes	0.7548	0.9481	0.6662	0.7123	0.9473	0.6883
Pathways	0.7689	0.9513	0.6701	0.7275	0.9492	0.6974
Indications	0.7959	0.9604	0.6878	0.7726	0.9550	0.7275
Side effects	0.7960	0.9586	0.6803	0.7741	0.9547	0.7240
Off side effects	0.7963	0.9587	0.6809	0.7742	0.9548	0.7243

Table 5
Performances of different models evaluated by 20 runs of 5-CV.

Methods	Features	AUPR	AUC	Recall	Precision	Accuracy	F-measure
Neighbor recommender method	Substructures	0.7590	0.9360	0.6170	0.7650	0.9500	0.6830
	Targets	0.3650	0.8200	0.5480	0.3380	0.8670	0.4180
	Transporters	0.3290	0.7140	0.3890	0.2900	0.8620	0.3310
	Enzymes	0.3770	0.7560	0.3460	0.4710	0.9090	0.3990
	Pathways	0.5710	0.8120	0.4740	0.6570	0.9320	0.5500
	Indications	0.5990	0.9120	0.5910	0.5550	0.9230	0.5720
	Side effects	0.7540	0.9360	0.6180	0.7500	0.9490	0.6780
	Off side effects	0.7680	0.9400	0.6290	0.7650	0.9510	0.6910
Random walk method	Substructures	0.7580	0.9360	0.6160	0.7630	0.9500	0.6810
	Targets	0.5590	0.8520	0.5010	0.5960	0.9270	0.5440
	Transporters	0.3630	0.7130	0.3810	0.2970	0.8640	0.3290
	Enzymes	0.4700	0.7600	0.3440	0.6570	0.9270	0.4510
	Pathways	0.5940	0.8110	0.4790	0.7090	0.9370	0.5720
	Indications	0.7770	0.9410	0.6410	0.7680	0.9520	0.6990
	Side effects	0.7600	0.9360	0.6210	0.7640	0.9500	0.6850
	Off side effects	0.7630	0.9370	0.6270	0.7610	0.9500	0.6880
MP		0.7820	0.9480	0.6660	0.7550	0.9520	0.7070
Classic MF		0.6889	0.9388	0.6699	0.6274	0.9364	0.6477
IONMF		0.7667	0.9502	0.6639	0.7575	0.9520	0.7074
MRMF	Substructures	0.7958	0.9585	0.6812	0.7722	0.9546	0.7237
	Targets	0.7769	0.9553	0.6830	0.7381	0.9511	0.7092
	Transporters	0.7393	0.9450	0.6641	0.6911	0.9447	0.6771
	Enzymes	0.7548	0.9481	0.6662	0.7123	0.9473	0.6883
	Pathways	0.7689	0.9513	0.6701	0.7275	0.9492	0.6974
	Indications	0.7959	0.9604	0.6878	0.7726	0.9550	0.7275
	Side effects	0.7960	0.9586	0.6803	0.7741	0.9547	0.7240
	Off side effects	0.7963	0.9587	0.6809	0.7742	0.9548	0.7243

MP: matrix perturbation; Classic MF: classic matrix factorization; IONMF: integrative orthogonality-regularized nonnegative matrix factorization; MRMF: manifold regularized matrix factorization.

previous work [15].

Moreover, we consider several baseline methods. Since MRMF formulates the DDI prediction as a matrix completion task, we adopt the classic matrix factorization method (MF) proposed by Koren et al. [46] for comparison. We also adopt the integrative orthogonality-regularized nonnegative matrix factorization (iONMF) [47], which can integrate multiple features and report good performances [47]. By taking drugs and interactions as nodes and edges, we can reformulate the DDI prediction as a missing link prediction problem. A missing link prediction method named "matrix perturbation method" (MP) is proposed recently [48], and our studies [15] showed that MP produces best results among dozens of predictors in the DDI prediction.

We evaluate all models by 20 runs of 5-fold cross validation (5-CV). As shown in Table 5, MRMF models based on different feature-based manifolds produce consistently satisfying results, achieving the AUPR value up to 0.7963. Among the eight features, substructures, side effects and off side effects can lead to better performances than other features. Based on the three features, MRMF models respectively achieve AUPR values of 0.7958, 0.7960 and 0.7963, better than neighbor recommender models (AUPR values of 0.7590, 0.7540 and 0.7680) and random walk models (AUPR values of 0.7580, 0.7600 and 0.7630). MRMF models also significantly improve the performance of the matrix factorization models: MF and IONMF, and produce better results than the missing link prediction method MP (AUPR value of 0.7820).

There are several reasons for good performances of MRMF. MRMF takes advantage of latent factor model that uncovers latent features of drugs, and introduces the manifold regularization to maintain the locally geometrical structures.

3.4. Case study

The primary goal of computational prediction is to screen the potential drug-drug interactions (DDIs) and guide the wet experiments to verify novel interactions. The benchmark dataset contains 48,584 interactions between 548 drugs from TWOSIDES. We train our prediction models by using all known interactions in TWOSIDES, and then predict novel interactions. TWOSIDES dataset was compiled in 2012, and DrugBank [31] is an up-to-date database for the drug-drug interactions. Therefore, we can check up on novel interactions that we predict in DrugBank database.

In the case study, we consider random walk method and matrix perturbation method for comparison, because they reported satisfying results. Here, we check up on top 10 prediction of each models in the latest online version of DrugBank, and count how many predicted interactions are confirmed, and results are demonstrated in Fig. 2. Since we have eight features for MRMF and random walk method, we can use

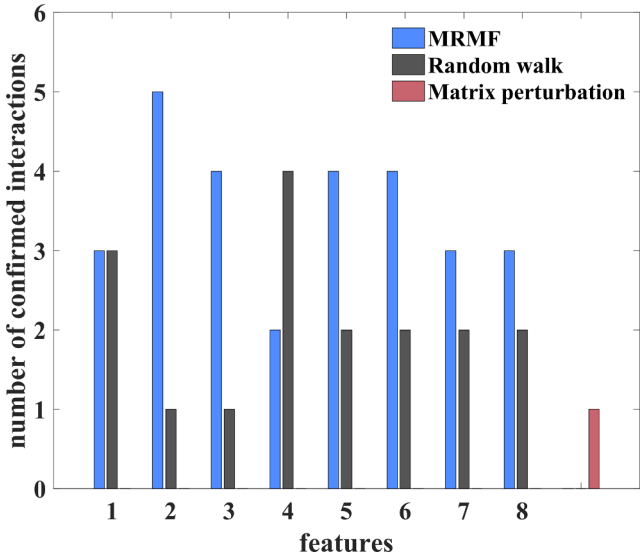

Fig. 2. Number of confirmed interactions in top 10 predictions of all prediction models. 1: substructures; 2: targets; 3: transporters; 4: enzymes; 5: pathways; 6: indications; 7: side effects; 8: off side effects.

Table 6
Top 10 interactions predicted by the target-based MRMF model.

Rank	Drug name	Drug name	Evidence in DrugBank
1	Norfloxacin	Lansoprazole	N.A.
2	Ciprofloxacin	Dorzolamide	N.A.
3	Captopril	Quinine	The metabolism of Captopril can be decreased when combined with Quinine
4	Clonazepam	Acetylcysteine	N.A.
5	Temozolomide	Alprazolam	N.A.
6	Acebutolol	Metoprolol	The serum concentration of Metoprolol can be increased when it is combined with Acebutolol
7	Rabeprazole	Zafirlukast	The metabolism of Zafirlukast can be decreased when combined with Rabeprazole.
8	Pantoprazole	Methadone	The metabolism of Methadone can be decreased when combined with Pantoprazole
9	Propranolol	Nadolol	Propranolol may increase the hypotensive activities of Nadolol
10	Nitroglycerin	Sirolimus	N.A.

N.A.: evidence is not available.

them to construct different MRMF models and random walk models. For seven out of eight features (excerpt enzymes), MRMF models identify more novel interactions than random walk models. MRMF also identifies more novel interactions than the matrix perturbation method.

Moreover, we list the top 10 interactions predicted by the target-based MRMF model in Table 6, and provide the evidence for five interactions. DrugBank database collects drug-drug interactions records from manual search, FDA reports, PubChem literature or DailyMed. According to records in DrugBank[31], the metabolism of Captopril can be decreased when combined with Quinine; the serum concentration of Metoprolol can be increased when it is combined with Acebutolol; the metabolism of Zafirlukast can be decreased when combined with Rabeprazole; the metabolism of Methadone can be decreased when combined with Pantoprazole; Propranolol may increase the hypotensive activities of Nadolol. The confirmation of novel interactions depends on the available evidence which were reported in wet experiments. Although part of predictions are not confirmed, it only indicates no evidence for these predictions, and existences of these interactions are uncertain.

Therefore, MRMF can find out novel drug-drug interactions, which are not included in the benchmark dataset, and has great potential of predicting drug-drug interactions.

4. Conclusions

In this paper, we propose a novel computational method named "Manifold Regularized Matrix Factorization" (MRMF) to predict potential drug-drug interactions (DDIs). Specifically, MRMF projects drugs in the interaction space into a low-dimensional space, and introduces the manifold regularization on which the projected drugs are assumed to lie and preserve the locally geometrical structures of the drug feature-based manifold. The experimental results show that MRMF is robust to multiple drug features and different similarity measures, achieves high-accuracy performances on the benchmark dataset, and outperforms other state-of-the-art methods. However, MRMF has three parameters, and tuning parameters usually cost lots of time. Combining diverse features usually lead to the improved performances, and has many successfully applications in bioinformatics [49–60]. MRMF only uses individual feature-based manifold regulations, and how to combine diverse features in a manifold regulation is our future work.

Conflict of interest

We declare that we have no conflict of interest.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (61772381, 61572368), the Fundamental Research Funds for the Central Universities (2042017kf0219), and National Key Research and Development Program (2018YFC0407904). The fundings have no role in the design of the study and collection, analysis, and interpretation of data and writing the manuscript.

References

[1] B. Percha, R.B. Altman, Informatics confronts drug–drug interactions, Trends Pharmacol. Sci. 34 (3) (2013) 178–184.
[2] D.N. Juurlink, M. Mamdani, A. Kopp, A. Laupacis, D.A. Redelmeier, Drug-drug interactions among elderly patients hospitalized for drug toxicity, JAMA 289 (13) (2003) 1652–1658.
[3] R.W. van Leeuwen, E.L. Swart, F.A. Boom, M.S. Schuitenmaker, J.G. Hugtenburg, Potential drug interactions and duplicate prescriptions among ambulatory cancer patients: a prevalence study using an advanced screening method, BMC Cancer 10 (1) (2010) 679.
[4] N. Nagai, Drug interaction studies on new drug applications: current situations and regulatory views in Japan, Drug Metab. Pharmacokinet. 25 (1) (2010) 3–15.
[5] T. Prueksaritanont, X. Chu, C. Gibson, D. Cui, K.L. Yee, J. Ballard, T. Cabalu, J. Hochman, Drug–drug interaction studies: regulatory guidance and an industry perspective, AAPS J. 15 (3) (2013) 629–645.
[6] H. Kusuhara, How far should we go? Perspective of drug-drug interaction studies in drug development, Drug Metab. Pharmacokinet. 29 (3) (2014) 227–228.
[7] A. Cami, S. Manzi, A. Arnold, B.Y. Reis, Pharmacointeraction network models predict unknown drug-drug interactions, PLoS ONE 8 (4) (2013) e61468.
[8] F. Cheng, Z. Zhao, Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties, J. Am. Med. Inform. Assoc. 21 (e2) (2014) e278–e286.
[9] A. Gottlieb, G.Y. Stein, Y. Oron, E. Ruppin, R. Sharan, INDI: a computational framework for inferring drug interactions and their associated recommendations, Mol Syst Biol 8 (2012) 592.
[10] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, C. Friedman, Drug-drug interaction through molecular structure similarity analysis, J Am Med Inform Assoc 19 (6) (2012) 1066–1074.
[11] S. Vilar, E. Uriarte, L. Santana, N.P. Tatonetti, C. Friedman, Detection of drug-drug interactions by modeling interaction profile fingerprints, PLoS ONE 8 (3) (2013) e58321.
[12] P. Li, C. Huang, Y. Fu, J. Wang, Z. Wu, J. Ru, C. Zheng, Z. Guo, X. Chen, W. Zhou, et al., Large-scale exploration and analysis of drug combinations, Bioinformatics 31 (12) (2015) 2007–2016.
[13] K. Park, D. Kim, S. Ha, D. Lee, Predicting Pharmacodynamic Drug-Drug Interactions through Signaling Propagation Interference on Protein-Protein Interaction Networks, PLoS ONE 10 (10) (2015) e0140816.
[14] P. Zhang, F. Wang, J. Hu, R. Sorrentino, Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects, Sci Rep 5 (2015) 12339.
[15] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, X. Li, Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data, BMC Bioinf. 18 (1) (2017) 18.
[16] R. Ferdousi, R. Safdari, Y. Omidi, Computational prediction of drug-drug interactions based on drugs functional similarities, J. Biomed. Inform. 70 (2017) 54–64.
[17] W. Zhang, X. Liu, Y. Chen, W. Wu, W. Wang, X. Li, Feature-derived graph regularized matrix factorization for predicting drug side effects, Neurocomputing 287 (2018) 154–162.
[18] W. Zhang, X. Yue, W. Lin, W. Wu, R. Liu, F. Huang, F. Liu, Predicting drug-disease associations by using similarity constrained matrix factorization, BMC Bioinf. 19 (1) (2018).
[19] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Application of dimensionality reduction in recommender system-a case study, Minnesota Univ Minneapolis Dept of Computer Science, In., 2000.
[20] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
[21] A. Mnih, R.R. Salakhutdinov, Probabilistic matrix factorization, Advances in neural information processing systems, 2008, pp. 1257–1264.
[22] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[23] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319-+.

[24] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Adv. Neural Inform. Process. Syst. Vols 1 and 2 (14) (2002) 585–591.
[25] N.P. Tatonetti, P.P. Ye, R. Daneshjou, RB. Altman, Data-driven prediction of drug effects and interactions, Sci. Transl. Med. 4 (125) (2012).
[26] Y. Wang, J. Xiao, T.O. Suzek, J. Zhang, J. Wang, S.H. Bryant, PubChem: a public information system for analyzing bioactivities of small molecules, Nucleic Acids Res. 37 (2009).
[27] Q. Li, T. Cheng, Y. Wang, S.H. Bryant, PubChem as a public resource for drug discovery, Drug Discovery Today 15 (23) (2010) 1052–1057.
[28] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Res. (2006) 34(90001).
[29] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets, Nucleic Acids Res. 36 (2008) 901–906 (Database issue).
[30] C. Knox, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs, Nucleic Acids Res. (2011) 39.
[31] C. Knox, Y. Djoumbou, T. Jewison, A.C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A.Y. Tang, DrugBank 4.0: shedding new light on drug metabolism, Nucleic Acids Res. (2014) 42.
[32] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, KEGG for representation and analysis of molecular networks involving diseases and drugs, Nucleic Acids Res. (2010) 38.
[33] M. Kuhn, M. Campillos, I. Letunic, L.J. Jensen, P. Bork, A side effect resource to capture phenotypic effects of drugs, Mol. Syst. Biol. 6 (1) (2010).
[34] F. Wang, C. Zhang, Label Propagation through Linear Neighborhoods, IEEE Trans. Knowl. Data Eng. 20 (1) (2008) 55–67.
[35] W. Zhang, X. Yue, Y.L. Chen, W.R. Lin, B.L. Li, F. Liu, X.H. Li, Predicting drug-disease associations based on the known association bipartite network, Ieee Int. Conf. Bioinform. Biomed. (Bibm) 2017 (2017) 503–509.
[36] W. Zhang, Y.L. Chen, S.K. Tu, F. Liu, QL. Qu, Drug side effect prediction through linear neighborhoods and multiple data source integration, Ieee Int. C Bioinform. (2016) 427–434.
[37] W. Zhang, Y. Chen, D. Li, Drug-target interaction prediction through label propagation with linear neighborhood information, Molecules 22 (12) (2017) 2056.
[38] W. Zhang, X. Yue, Y. Chen, W. Lin, B. Li, F. Liu, X. Li, Predicting drug-disease associations based on the known association bipartite network, IEEE International Conference on Bioinformatics and Biomedicine, 2017, pp. 503–509.
[39] W. Zhang, X. Yue, F. Liu, Y. Chen, S. Tu, X. Zhang, A unified frame of predicting side effects of drugs by using linear neighborhood similarity, BMC Syst. Biol. 11 (6) (2017) 101.
[40] W. Zhang, X. Yue, F. Huang, R. Liu, Y. Chen, C. Ruan, Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network, Methods (2018).
[41] W. Zhang, X. Yue, W. Lin, W. Wu, R. Liu, F. Huang, F. Liu, Predicting drug-disease associations by using similarity constrained matrix factorization, BMC Bioinf. 19 (1) (2018) 233.
[42] T. van Laarhoven, S.B. Nabuurs, E. Marchiori, Gaussian interaction profile kernels for predicting drug–target interaction, Bioinformatics 27 (21) (2011) 3036–3043.
[43] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: a geometric framework for learning from labeled and unlabeled examples, J. Machine Learning Res. 7 (1) (2006) 2399–2434.
[44] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press, 2004.
[45] M. Duran-Frigola, P. Aloy, Recycling side-effects into clinical markers for drug repositioning, Genome Med. 4 (1) (2012) 3.
[46] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (8) (2009) 30–37.
[47] M. Strazar, M. Zitnik, B. Zupan, J. Ule, T. Curk, Orthogonal matrix factorization enables integrative analysis of multiple RNA binding proteins, Bioinformatics 32 (10) (2016) 1527–1535.
[48] L. Lu, L. Pan, T. Zhou, Y.C. Zhang, H.E. Stanley, Toward link predictability of complex networks, Proc. Natl. Acad. Sci. U S A 112 (8) (2015) 2325–2330.
[49] W. Zhang, Y. Niu, Y. Xiong, M. Zhao, R. Yu, J. Liu, Computational prediction of conformational B-cell epitopes from antigen primary structures by ensemble learning, PLoS ONE 7 (8) (2012) e43575.
[50] W. Zhang, J. Liu, Y. Xiong, M. Ke, K. Zhang, Predicting immunogenic T-cell epitopes by combining various sequence-derived features, December 18-21, IEEE International Conference on Bioinformatics and Biomedicine, IEEE Computer Society, Shanghai, 2013, pp. 4–9.
[51] W. Zhang, F. Liu, L. Luo, J. Zhang, Predicting drug side effects by multi-label learning and ensemble learning, BMC Bioinf. 16 (2015) 365.
[52] W. Zhang, Y. Niu, H. Zou, L. Luo, Q. Liu, W. Wu, Accurate prediction of immunogenic T-cell epitopes from epitope sequences using the genetic algorithm-based ensemble learning, PLoS ONE 10 (5) (2015) e0128194.
[53] D. Li, L. Luo, W. Zhang, F. Liu, F. Luo, A genetic algorithm-based weighted ensemble method for predicting transposon-derived piRNAs, BMC Bioinf. 17 (1) (2016) 329.
[54] L. Luo, D. Li, W. Zhang, S. Tu, X. Zhu, G. Tian, Accurate prediction of transposon-derived piRNAs by integrating various sequential and physicochemical features, *PLoS ONE* 11 (4) (2016).
[55] W. Zhang, Y. Chen, S. Tu, F. Liu, Q. Qu, Drug side effect prediction through linear neighborhoods and multiple data source integration, IEEE Int. Conf. Bioinform. Biomed. (BIBM) 2016 (2016) 427–434.
[56] W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, W. Xiao, Predicting potential side effects of drugs by recommender methods and ensemble learning, Neurocomputing 173 (2016) 979–987.
[57] W. Zhang, J. Shi, G. Tang, W. Wu, X. Yue, D. Li, Predicting small RNAs in bacteria via sequence learning ensemble method, IEEE International Conference on Bioinformatics and Biomedicine, 2017, pp. 643–647.
[58] W. Zhang, X. Zhu, Y. Fu, J. Tsuji, Z. Weng, Predicting human splicing branchpoints by combining sequence-derived features and multi-label learning methods, BMC Bioinf. 18 (Suppl 13) (2017) 464.
[59] W. Zhang, Q. Qu, Y. Zhang, W. Wang, The linear neighborhood propagation method for predicting long non-coding RNA–protein interactions, Neurocomputing 273 (2018) 526–534.
[60] W. Zhang, X. Zhu, Y. Fu, J. Tsuji, Z. Weng, The prediction of human splicing branchpoints by multi-label learning, *IEEE International Conference on Bioinformatics and Biomedicine*, 2016, pp. 254–259.