



# A Method of Optimizing Weight Allocation in Data Integration Based on Q-Learning for Drug-Target Interaction Prediction

Jiacheng Sun<sup>1,2,3</sup>, You Lu<sup>1,2,3\*</sup>, Linqian Cui<sup>1,2,3</sup>, Qiming Fu<sup>1,2,3</sup>, Hongjie Wu<sup>1</sup> and Jianping Chen<sup>2,4\*</sup>

<sup>1</sup>School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou, China, <sup>2</sup>Jiangsu Province Key Laboratory of Intelligent Building Energy Efficiency, Suzhou University of Science and Technology, Suzhou, China, <sup>3</sup>Suzhou Key Laboratory of Mobile Network Technology and Application, Suzhou University of Science and Technology, Suzhou, China, <sup>4</sup>School of Architecture and Urban Planning, Suzhou University of Science and Technology, Suzhou, China

## OPEN ACCESS

### Edited by:

Lei Deng,  
Central South University, China

### Reviewed by:

Wenzheng Bao,  
Xuzhou University of Technology,  
China  
Yi Xiong,  
Shanghai Jiao Tong University, China

### \*Correspondence:

You Lu  
luyou@usts.edu.cn  
Jianping Chen  
alanjpchen@allyun.com

### Specialty section:

This article was submitted to  
Molecular and Cellular Pathology,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

**Received:** 13 October 2021

**Accepted:** 14 February 2022

**Published:** 04 March 2022

### Citation:

Sun J, Lu Y, Cui L, Fu Q, Wu H and  
Chen J (2022) A Method of Optimizing  
Weight Allocation in Data Integration  
Based on Q-Learning for Drug-Target  
Interaction Prediction.  
Front. Cell Dev. Biol. 10:794413.  
doi: 10.3389/fcell.2022.794413

Calculating and predicting drug-target interactions (DTIs) is a crucial step in the field of novel drug discovery. Nowadays, many models have improved the prediction performance of DTIs by fusing heterogeneous information, such as drug chemical structure and target protein sequence and so on. However, in the process of fusion, how to allocate the weight of heterogeneous information reasonably is a huge challenge. In this paper, we propose a model based on Q-learning algorithm and Neighborhood Regularized Logistic Matrix Factorization (QLNRLMF) to predict DTIs. First, we obtain three different drug-drug similarity matrices and three different target-target similarity matrices by using different similarity calculation methods based on heterogeneous data, including drug chemical structure, target protein sequence and drug-target interactions. Then, we initialize a set of weights for the drug-drug similarity matrices and target-target similarity matrices respectively, and optimize them through Q-learning algorithm. When the optimal weights are obtained, a new drug-drug similarity matrix and a new drug-drug similarity matrix are obtained by linear combination. Finally, the drug target interaction matrix, the new drug-drug similarity matrices and the target-target similarity matrices are used as inputs to the Neighborhood Regularized Logistic Matrix Factorization (NRLMF) model for DTIs. Compared with the existing six methods of NetLapRLS, BLM-NII, WNN-GIP, KBMF2K, CMF, and NRLMF, our proposed method has achieved better effect in the four benchmark datasets, including enzymes(E), nuclear receptors (NR), ion channels (IC) and G protein coupled receptors (GPCR).

**Keywords:** drug-target interactions, heterogeneous information fusion, q-learning, weight distribution, drug similarity, target similarity

## INTRODUCTION

Diseases are usually caused by defective proteins in the body or the functional structure of viral proteins. Effective drugs can combine well with these proteins, removing the original function and achieving the therapeutic effect (Ding et al., 2021). Previous research and development of novel drugs mainly relied on biochemical experiments, which are risky, expensive and time-consuming. In

addition, protein-protein interactions play a key role in many biological processes (Guo et al., 2015), leading to the emergence of large-scale experimental data on genes and proteins, making drug discovery and repositioning in biomedical research more difficult (Ding et al., 2019a). The purpose of DTIs prediction is to identify potential novel drugs or novel targets for existing drugs, and provide a list of candidate drugs for drug discovery, thus greatly improving the efficiency of research and development and reducing the cost of experiments. In recent years, the research methods of DTIs prediction have been proposed successively. So far, a large number of associations known to date have been proved by previous experiments and stored in some public databases (Wang et al., 2021). Therefore, existing methods predict DTIs mainly based on a small number of experimentally validated interactions in existing databases, such as DrugBank (Wishart et al., 2008), KEGG DRUG (Kanehisa et al., 2012), and SuperTarget (Günther et al., 2008). Previous studies have shown that DTIs prediction based on experimental verification can effectively predict some novel interactions between drugs and targets, and the computational methods used to predict DTIs can significantly improve the efficiency of drug discovery.

In general, traditional computational methods proposed for DTIs prediction are mainly based on ligands and target-based (Liu et al., 2021). The ligand-based method usually compares the candidate ligand with the known interacting ligand of the target to determine the binding between them. However, the disadvantage is that the ligand-based method cannot be used for targets which have no or only a small amount of known binding ligands. The target-based (or docking simulation) method uses docking techniques to predict interactions between drug candidates and targets. Nevertheless, this method requires detailed structures. Not all proteins have structural information (Ding et al., 2020). For example, most targets of GPCRs (G protein coupled receptors) are unknown (Yan et al., 2019).

In order to solve the difficulties of traditional methods, chemical genomics methods have been successfully used for large-scale drug discovery and repositioning (Chen et al., 2018). The purpose of chemical genomics research is to integrate drug and target information into a unified framework, in order to identify potentially useful compounds (Ezzat et al., 2019). Chemical genomics methods are usually divided into ligand-based, target-based, and target-ligand, all of which are based on similarity between member proteins and targets. In fact, this significant similarity-based view of chemical genomics has made machine learning methods widely used in DTIs prediction tasks (Ding et al., 2017; Lo et al., 2018; Ding et al., 2019b; Ding et al., 2019c; Qu et al., 2019; Liu et al., 2020; Bagherian et al., 2021), where DTIs prediction is regarded as a binary classification problem, in which drug-target pairs are taken as examples, the internal chemical structure of the drug and the amino acid subsequence of the target are considered as features. Then, we can build a binary classification model through some classic classification methods, such as support vector machines (SVM), neural networks and nearest

neighbors. Yamanishi et al. (Yamanishi et al., 2010) proposed a framework for supervised bipartite graph reasoning, and this framework could predict unknown DTIs based on chemical, genomic and pharmacological data. Tabei et al. (Tabei et al., 2012) used complex protein pairs represented by eigenvectors to correspond to chemical substructures and protein domains, and used logistic regression and SVM to establish a prediction model. Mei et al. (Mei et al., 2013a) improved the bipartite local model by considering novel drug candidates through the interaction profile of neighbors. In short, DTIs prediction based on machine learning comes down to a process of data collection, feature representation, similarity calculation and machine learning modeling. Although machine learning has achieved good results in DTIs prediction, there are still bottlenecks. For example, most drug relocation methods based on machine learning only use a single metric to evaluate the similarity between diseases and between drugs. In fact, the similarity of drugs/diseases is not only noisy, but also multimodal, which can be measured from different aspects (Yang et al., 2021). The fusion of multiple similarity measures can avoid the noise and extract effective features in individual similarity calculation, thus effectively improving the accuracy of DTIs prediction.

In order to overcome the bottleneck encountered by machine learning in DTIs prediction, improving the prediction performance of DTIs by integrating heterogeneous information related to drugs and targets has been extensively studied by the academic community. Among them, the challenge is how to extract and integrate such information (Wang et al., 2011). Wang et al. (Wang et al., 2011) proposed a kernel function method based on SVM predictor to integrate heterogeneous information sources to improve the accuracy of DTIs prediction. Yan et al. (Yan et al., 2019) proposed a method based on multi-kernel learning and clustering to integrate heterogeneous information sources related to multiple drugs and targets to improve the accuracy of DTIs prediction, in which the weight of linear weighting was obtained through interior point optimization algorithm (Byrd et al., 1999). (Ding et al., 2020) used six different features to characterize protein sequences in the experimental process of predicting the subcellular localization of proteins. These features are constructed into corresponding kernels and combined by optimizing the weight of these kernels by Kernel Target Alignment-based Multiple Kernel Learning (KTA-MKL) (Ding et al., 2020). (Zhou et al., 2021) built a multi-kernel learning model with Hilbert-Schmidt independence criterion (HSIC) to obtain optimal weights for various features, and identified ncRNA subcellular localization by using the graph regularization k local hyperplane distance nearest neighbor model. Although the above methods have achieved good results, there are still some problems. For example, Wang et al. (Wang et al., 2011) found that the experimental effect of fusion of three information sources was inferior to that of fusion of two information sources in the process of fusion, and the reason may be that drug targets for the contributions to the similarity

**TABLE 1** | Information about the four datasets.

Dataset	E	IC	GPCR	NR
Drugs(n)	445	210	223	54
Targets(m)	664	204	95	26
Interactions	2926	1476	635	90

of DTIs prediction is different, and the integration of equal-weight information cannot reflect the difference of contribution ratio of heterogeneous information, DTIs prediction accuracy improved range is limited. To solve the above problems, we used Q-learning algorithm to reasonably allocate weights in the process of heterogeneous information integration, and realized DTIs prediction based on the NRLMF model proposed by Liu et al. (Liu et al., 2016). Finally, our method QLNRLMF achieved good results on four benchmark datasets.

The main contributions of this paper are summarized as follows:

- According to three heterogeneous data sources, including drug chemical structure, target protein sequence and drug-target interaction, we obtain three drug-drug similarity matrices and three target-target similarity matrices using different similarity calculation methods.
- We use Q-learning algorithm to reasonably allocate weight to the similarity matrices obtained above, and then perform linear weighted integration after obtaining the optimal weight.
- We perform DTIs prediction experiments on four benchmark datasets. Experimental results show that the DTIs prediction accuracy of QLNRLMF model is better to several advanced comparison models

## MATERIALS AND METHODS

### Problem Description

In this paper, the drug set is defined as  $D = \{d_i\}_{i=1}^m$ , and the target set is defined as  $T = \{t_j\}_{j=1}^n$ , where  $m$  and  $n$  respectively represent the number of drugs and the number of targets. The interaction between the drug and the target is defined as  $Y \in R^{m \times n}$ , which is composed of drug  $d_i$  ( $1 \leq i \leq m$ ) and target  $t_j$  ( $1 \leq j \leq n$ ), the matrix  $Y \in R^{m \times n}$  is defined as Eq. 1.

$$Y_{ij} = \begin{cases} 1 & \text{Drug } d_i \text{ is related to Target } t_j \\ 0 & \text{Drug } d_i \text{ is not related to Target } t_j \end{cases} \quad (1)$$

We take each drug-target pair in  $Y$  as a sample, and divide equally all drug-target pairs in  $Y$  into ten pieces by random seed shuffled. Then we select one sample as the test set in turn, and the remaining nine samples as the training set. At the same time, we obtain all the samples in the test set corresponding to the subscripts in  $Y$  and assign all the corresponding positions to 0, and a new drug and target interaction  $Y_{train} \in R^{m \times n}$  is obtained. The purpose of our experiment is to predict these unknown pairs of interactions labeled 0.

**TABLE 2** | Summary of similarity matrix of two feature spaces.

Space	Similarity matrix	Description
Drug	$SIM_{strdrug}$	Chemical structure
	$SIM_{cosdrug}$	Cosine similarity of drugs
	$SIM_{Jacdrug}$	Jaccard similarity of drugs
Target	$SIM_{strtar}$	Target sequence information
	$SIM_{costar}$	Cosine similarity of target
	$SIM_{Jactar}$	Jaccard similarity of the target

## Data Preparation

### Chemical Data

Yamanishi et al. (Yamanishi et al., 2008) obtained the chemical structure of the compound from the drugs and compounds section of the KEGG ligand database (Kanehisa et al., 2007). The chemical structure similarity between drugs is obtained by the SIMCOMP algorithm (Hattori et al., 2003), and the similarity score is calculated according to the number of common substructures between two compounds. In this paper, the chemical structure similarity between two compounds  $d_i$  and  $d_j$  can be calculated according to the Tanimoto coefficient as follows Eq. 2, and the chemical structure similarity matrix of the drug compound is expressed as  $SIM_{strdrug}$ .

$$SIM_{strdrug_{ij}} = \frac{|d_i \cap d_j|}{|d_i \cup d_j|} \quad (2)$$

### Genetic Data

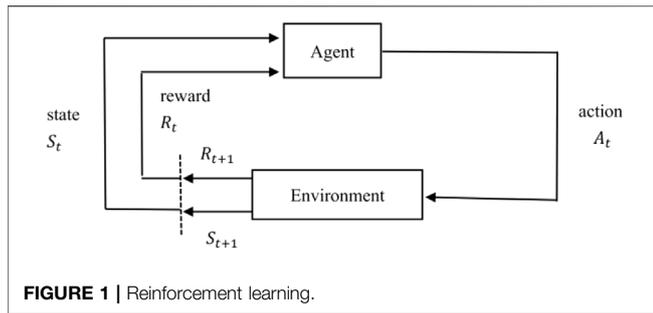
Due to the rapid development of sequencing technology, a large amount of data has been accumulated, we use amino acid sequence data to measure the similarity of proteins. In this paper, normalized Smith-Waterman score (Smith and Waterman, 1981) is used to calculate the sequence similarity between protein  $t_i$  and protein  $t_j$  as shown in Eq. 3, where  $SW(t_i, t_j)$  represents the Smith-Waterman score between the target protein  $t_i$  and  $t_j$ , and the sequence similarity matrix of the protein is expressed as  $SIM_{strtar}$ .

$$SIM_{strtar_{ij}} = \frac{SW(t_i, t_j)}{\sqrt{SW(t_i, t_i)SW(t_j, t_j)}} \quad (3)$$

### Drug-Target Interactions Data:

As for the performance evaluation of the DTIs prediction algorithm, we will verify it on four benchmark datasets, including E, IC, GPCR, and NR. Get the address of datasets: <http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/>. Table 1 counts the relevant information of the four datasets, including the number of drugs (n), the number of targets (m), the number of interactions.

According to the above  $Y_{train} \in R^{m \times n}$ , we use cosine similarity and Jaccard similarity coefficient to calculate the drug-drug similarity matrices  $SIM_{cosdrug}$ ,  $SIM_{Jacdrug}$ , target-target similarity matrices  $SIM_{costar}$ ,  $SIM_{Jactar}$ .



$$SIM_{cosdrug_{ij}} = \frac{|d_i \cap d_j|}{\sqrt{|d_i| |d_j|}} \quad (4)$$

$$SIM_{Jaddrug_{ij}} = \frac{|d_i \cap d_j|}{|d_i \cup d_j|} \quad (5)$$

$$SIM_{costar_{ij}} = \frac{|t_i \cap t_j|}{\sqrt{|t_i| |t_j|}} \quad (6)$$

$$SIM_{Jactar_{ij}} = \frac{|t_i \cap t_j|}{|t_i \cup t_j|} \quad (7)$$

Table 2 lists the above similarity matrices. There are three similarity matrices in the drug space and the target space respectively.

## NRLMF

NRLMF is the method proposed by Lin et al. (Liu et al., 2016), which combines Logistic matrix decomposition (LMF) and domain regularization to predict associations. Some of the equations that will be used in the model are as follows:

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} \sum_{i=1}^m \sum_{j=1}^n (1 + cy_{ij} - y_{ij}) \ln[1 + \exp(\mathbf{u}_i \mathbf{V}_j^T)] - cy_{ij} \mathbf{u}_i \mathbf{V}_j^T \\ + \frac{1}{2} \text{tr}[\mathbf{U}^T (\lambda_d \mathbf{I} + \alpha \mathbf{L}^d) \mathbf{U}] + \frac{1}{2} \text{tr}[\mathbf{V}^T (\lambda_t \mathbf{I} + \beta \mathbf{L}^t) \mathbf{V}] \end{aligned} \quad (8)$$

Denoting the objective function in Eq. 8 by L, the partial gradients with respect to U and V are as follows:

$$\frac{\partial L}{\partial \mathbf{U}} = \mathbf{P}\mathbf{V} + (c-1)(\mathbf{Y} \odot \mathbf{P})\mathbf{V} - c\mathbf{Y}\mathbf{V} + (\lambda_d \mathbf{I} + \alpha \mathbf{L}^d) \mathbf{U} \quad (9)$$

$$\frac{\partial L}{\partial \mathbf{V}} = \mathbf{P}^T \mathbf{U} + (c-1)(\mathbf{Y}^T \odot \mathbf{P}^T) \mathbf{U} - c\mathbf{Y}^T \mathbf{U} + (\lambda_t \mathbf{I} + \beta \mathbf{L}^t) \mathbf{V} \quad (10)$$

where  $\mathbf{P} \in R^{m \times n}$ , in which the  $(i, j)$  element is  $P_{ij}$ ,  $\odot$  denotes the Hadamard product of two matrices. View reference Lin et al. (Liu et al., 2016), where we can see more details about above equation.

## Q-Learning Algorithm

Reinforcement learning is an algorithm model that can make the best strategy through self-study in a particular situation. As shown in Figure 1, for real-world problems, reinforcement learning can be modeled by abstracting them as the interaction

process between agent and environment. For each time step, the agent receives the state of the environment and chooses to perform the corresponding action, and then in the next time step, the agent obtains a reward and a new state based on the feedback of the environment. In other words, reinforcement learning refers to continuously learning to adapt to the environment according to the rewards obtained, and the goal of the agent is to maximize the expected cumulative reward.

Q-learning is a value-based algorithm among reinforcement learning algorithms.  $Q(s, a)$  is the expectation that taking action  $a$  ( $a \in A$ ) can obtain benefits under state  $s$  ( $s \in S$ ) at a certain moment, and the environment will feedback corresponding return  $r$  according to the action taken by the agent. Therefore, the main idea of this algorithm is to construct a Q-Table (as shown in Table 3) between state S and action A to store Q value, and then select the action that can obtain the maximum benefit according to Q value.

The updated formula for Q value calculation in Q-Table is as follows:

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (11)$$

Where  $\alpha$  is learning rate and  $\gamma$  is discount rate.

## MDP Modeling State Space Design

Before the experiment begins, we initialize a set of weights for the drug similarity matrix and the target similarity matrix respectively, and then find a set of optimal weight assignments through the Q-learning algorithm. The state space at time step  $t$  is defined as  $S_t = (\alpha_t, \beta_t, \gamma_t, x_t, y_t, z_t)$ , where  $\alpha_t, \beta_t, \gamma_t, x_t, y_t$  and  $z_t$  represent respectively the weights in front of the three drug similarity matrices and the three target similarity matrices at the time step  $t$ , and  $\alpha_t + \beta_t + \gamma_t = 1, x_t + y_t + z_t = 1$ .

$$SIM_{drug} = \alpha SIM_{strdrug} + \beta SIM_{cosdrug} + \gamma SIM_{Jaddrug} \quad (12)$$

$$SIM_{tar} = x SIM_{strtar} + y SIM_{costar} + z SIM_{Jactar} \quad (13)$$

## Action Space Design

In order to optimize the weight of the drug similarity matrix and the target similarity matrix, we define the action space as  $A = (a, b, c, -(b+c))$ , where  $a \in \{0, 1\}$ , 0 means that the current action only acts on the weight of the drug similarity matrix, and 1 means that it only acts on the weight of the target similarity matrix;  $b, c \in \{-0.1, 0, 0.1\}$ , for example, the current state is (0.3, 0.3, 0.4, 0.3, 0.3, 0.4) and the executed action is (0, 0.1, 0.1, -0.2), so the next state is (0.4, 0.4, 0.2, 0.3, 0.3, 0.4).

TABLE 3 | Q-Table.

Q-Table	$a_1$	$a_2$
$S_1$	$q(S_1, a_1)$	$q(S_1, a_2)$
$S_2$	$q(S_2, a_1)$	$q(S_2, a_2)$
$S_3$	$q(S_3, a_1)$	$q(S_3, a_2)$

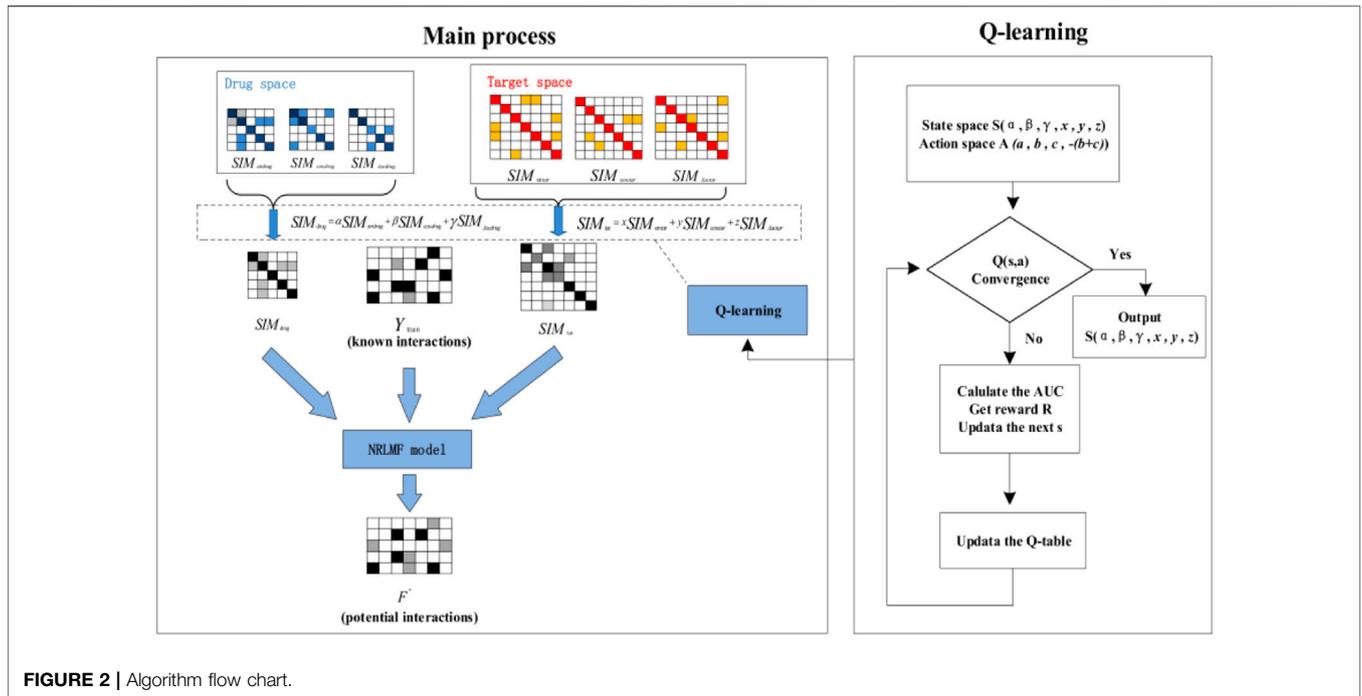


FIGURE 2 | Algorithm flow chart.

### Design of Reward Function

After each action is performed, environment will respond with an instant reward  $R$  in return. In this paper, we will use AUC as a reward. When we perform an action  $a$  at each time step, we will get a new state  $s$  that is a new weight. Then, data integration is carried out according to the new weights, and the integration results and drug-target interactions data are taken as the input of the NRLMF model, and the output result AUC is used as the instant reward after performing the corresponding action. If the state does not match after the action is taken, that is, there is a situation that is less than 0 or greater than 1 in the state, we will feedback 0 as an instant reward.

$$R(S_t) = \begin{cases} AUC & \text{If all weights are between 0 and 1} \\ 0 & \text{Otherwise} \end{cases} \quad (14)$$

### QLNRLMF

In our experiment, we divide the dataset into training set and test set in a ratio of 9:1, and validate the performance with ten-fold-cross-validation. The main ideas of our method are as follows: First, we use different similarity calculation methods to calculate heterogeneous data sources and obtain drug Space and target Space. Then, we initialize weights for the matrices in Drug space and target space respectively and optimize them through Q-learning algorithm. After obtaining the optimal weight, linear weighted integration is carried out to obtain  $SIM_{drug}$  and  $SIM_{tar}$ . Finally, we take  $SIM_{drug}$ ,  $SIM_{tar}$  and  $Y_{train}$  as the inputs of NRLMF model for DTIs prediction.

The pseudocode of QLNRLMF algorithm is shown in Algorithm 1. The overall flow chart for QLNRLM is shown in Figure 2.

### Algorithm 1. Qlnrlmf

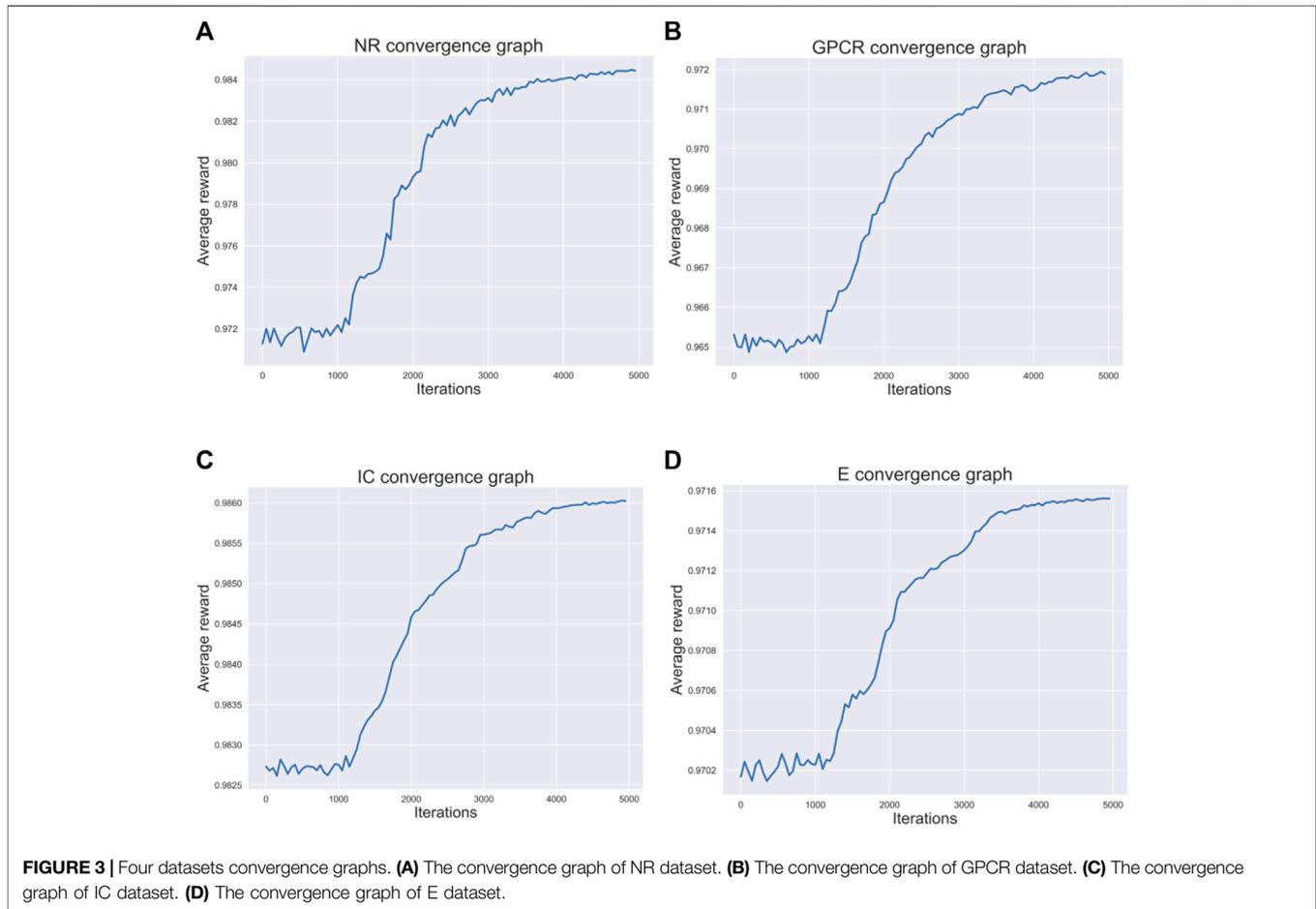
```

Require: drugs set  $D$ , targets set  $T$ , drug-target interactions  $Y$ , state space  $S$ , action space  $A$ , reward  $R$ , learning_rate  $\alpha$ , discount rate  $\gamma$ 
Ensure: The prediction of  $S(\alpha, \beta, \gamma, x, y, z)$ 
1. Divide the dataset and get the  $Y_{train}$ 
2. Calculate the similarity matrix  $SIM_{strdrug}$ ,  $SIM_{cosdrug}$ ,  $SIM_{jadrug}$ ,  $SIM_{sttar}$ ,  $SIM_{costar}$ ,  $SIM_{jactar}$ 
3. Initialize  $Q(s, a)$ 
4. repeat (for each episode):
5.   Initialize  $s$ 
6.   repeat (for each step of episode):
7.     choose  $a$  from  $s$  using policy derived from  $Q$  (e.g.,  $e\_greedy$ )
8.     take action  $a$ 
9.      $s' \leftarrow s, a$ 
10.     $SIM_{drug} = \alpha SIM_{strdrug} + \beta SIM_{cosdrug} + \gamma SIM_{jadrug}$ 
11.     $SIM_{tar} = x SIM_{sttar} + y SIM_{costar} + z SIM_{jactar}$ 
12.     $r \leftarrow NRLMF(SIM_{drug}, SIM_{tar}, Y_{train})$ 
13.     $Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$ 
14.     $s \leftarrow s'$ 
15.   until  $s$  is terminal
16. until  $Q(s, a)$  convergence
    
```

## RESULTS

### Evaluation Measurements

In order to verify the performance of our proposed method, we evaluate it from the feasibility, efficiency and accuracy of the algorithm. First of all, we prove the feasibility of our algorithm through the convergence graph of the average reward for each episode. Secondly, we compare the Q-learning algorithm with the brute force method in terms of the time required to find the optimal weights, to prove the effectiveness of our algorithm.



**TABLE 4 |** Time contrast between Q-learning algorithm and Brute Force algorithm.

Dataset	1	2	3	4	5	6	7	8	9	10	Avg	Avg/ S <sub>num</sub>
NR	1023	1001	1032	1027	1029	987	1008	1016	1035	1005	1016	0.784
GPCR	1033	1069	1001	1050	1015	1055	1050	1015	1050	1024	1036	0.799
IC	997	1021	1030	1034	1045	1034	1027	1012	1006	1017	1022	0.789
E	1045	994	1036	1039	1031	1001	1032	1064	982	1026	1025	0.791

Finally, we compare our algorithm with other algorithms on AUC and AUPR to prove the accuracy of our algorithm.

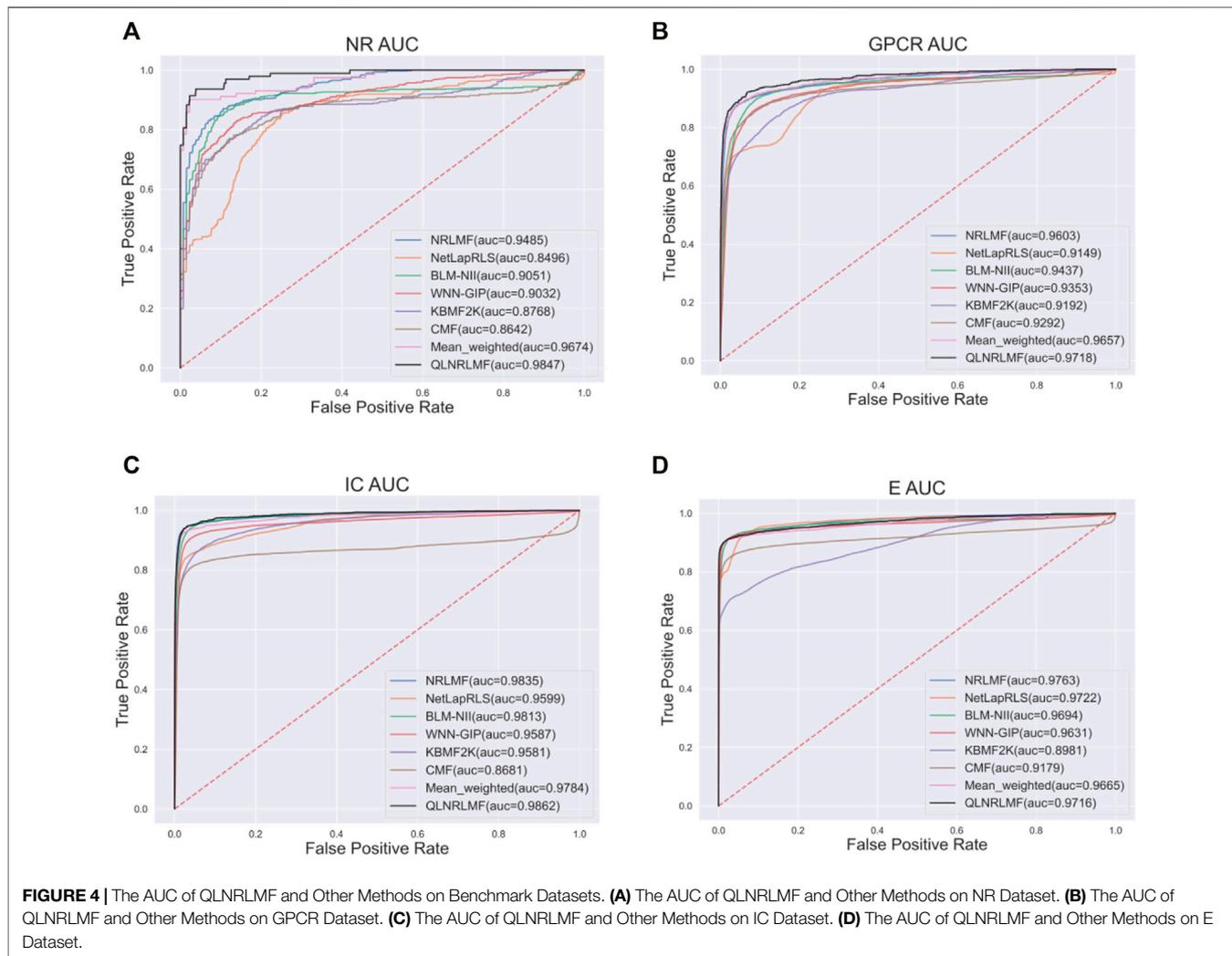
### Average Reward Convergence Graph

Figure 3 respectively describes the average reward convergence graph of four benchmark datasets under q-Learning algorithm, where the abscissa represents the number of iterations and the ordinate represents the average reward. As shown in the Figure 3, the average reward does not show an obvious upward trend in the first 1000 iterations, because the agent is constantly exploring trial and error at the beginning. After that, with the continuous accumulation of experience, the strategies learned by agents are getting better and better, and the average reward also

presents an upward trend. When the number of iterations reaches 4000, the average reward basically tends to be stable, indicating that agent has learned an optimal strategy.

### Comparisons with Brute Force Algorithm

In this experiment, the number  $S_{num}$  of the state space is 1296. If the Brute Force algorithm is adopted, the optimal state can be found by traversing 1296 states, that is, the optimal weight allocation can be obtained. If we use the Q-learning algorithm, we separately record the number of different states visited when each fold finds the optimal weight. As shown in Table 4, we can conclude that the ratio of the time required by the four datasets to the time required by the Brute Force algorithm are: 78.4, 79.9,



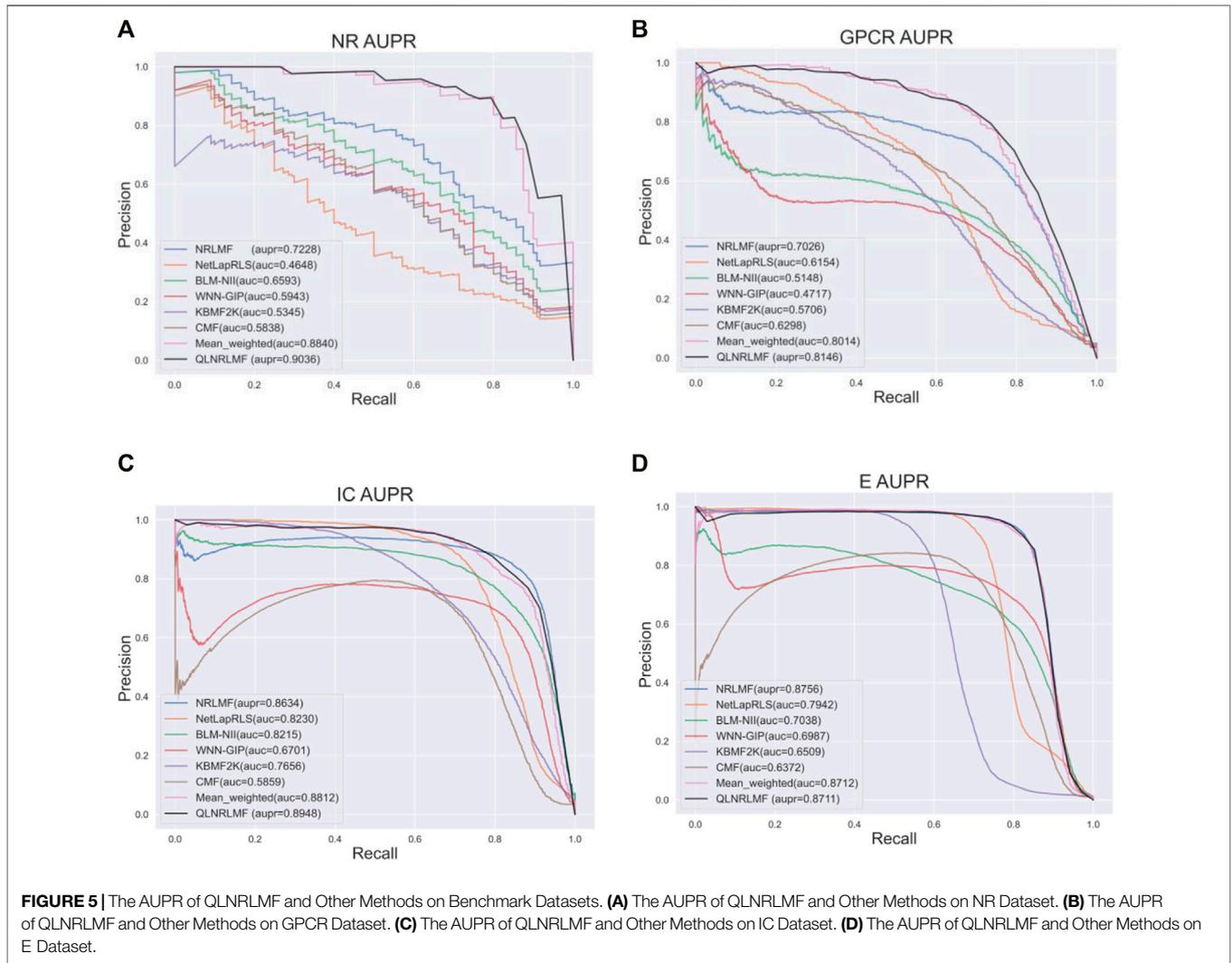
78.9, 79.1%, and the average time ratio is 79.1%. All in all, the running time of Q-learning algorithm has been reduced by nearly 20.9% on average.

## Comparisons with Other Methods

There are many evaluation indexes for DTIs prediction methods. Among them, AUC and AUPR are widely used. AUC is the area under the receiver operating characteristic (ROC) curve, which can also be understood as the probability that a positive sample and a negative sample are randomly given, and the prediction probability of the positive sample is greater than the probability of the prediction probability of the negative sample. The value of AUC can directly evaluate the performance of the DTIs prediction method, and the greater the value, the better. AUPR is the area under the curve of precision and recall rate, and is a quantitative measurement method that can determine the average separation degree between the predicted fraction of real interactions and the predicted fraction of real non-interactions (Peng et al., 2017). Relatively few interactions are known for DTIs

prediction. Therefore, AUPR is a more effective evaluation indicator than AUC, because AUPR takes some measures to reduce the impact of predicted fake DTIs data on the highest ranking scores (Davis and Goadrich, 2006). Therefore, we use these two indicators to evaluate the performance of our proposed method.

**Figure 4; Figure 5** respectively describe the comparison of AUC and AUPR between our method and other seven methods under four benchmark datasets. As shown in the figures, the comparison between our QLNRLMF method and NRLMF method shows that in four datasets, our method improves 3.62, 1.15, 0.27% and -0.47% respectively in AUC, and 18.08, 11.23, 3.25% and -0.48% respectively in AUPR. In general, our method is superior to NRLMF method. In order to prove the advantages of Q-learning algorithm, we also perform the integration experiment of linear equal-weight strategy. The experimental results show that the linear integration strategy based on Q-learning algorithm is obviously better than the linear equal-weight integration strategy. In addition, we also compare



**TABLE 5 |** Comparison with the other seven methods.

AUC								
Dataset	NetLapRLS	BLM-NII	WNN-GIP	KBMF2K	CMF	Mean-weighted	NRLMF	QLNRLMF
NR	0.849	0.905	0.903	0.876	0.864	0.967	0.948	<b>0.984</b>
GPCR	0.914	0.943	0.935	0.919	0.929	0.965	0.960	<b>0.971</b>
IC	0.959	0.981	0.958	0.958	0.868	0.978	0.983	<b>0.986</b>
E	0.972	0.969	0.963	0.898	0.917	0.966	<b>0.976</b>	0.971
Avg	0.923	0.949	0.939	0.912	0.894	0.969	0.966	<b>0.978</b>
AUPR								
NR	0.464	0.659	0.594	0.534	0.583	0.884	0.722	<b>0.903</b>
GPCR	0.615	0.514	0.471	0.570	0.629	0.801	0.702	<b>0.814</b>
IC	0.823	0.821	0.670	0.765	0.585	0.881	0.863	<b>0.894</b>
E	0.794	0.703	0.698	0.650	0.637	0.871	<b>0.875</b>	0.871
Avg	0.674	0.674	0.608	0.629	0.608	0.859	0.790	<b>0.870</b>

The values in bold mean the best result for each line.

our algorithm with other advanced five methods, which contain NetLapRLS (Xia et al., 2010), BLM-NII (Mei et al., 2013b), WNN-GIP (van Laarhoven and Marchiori, 2013), KBMF2K (Gönen,

2012) and CMF (Zheng et al., 2013), we can find that our proposed method has certain advantages from the figures. Finally, we summarize the experimental results and data of all

the above methods into **Table 5**, from which we find that our QLNRLMF method achieves better results in DTIs prediction.

## CONCLUSION AND DISCUSSION

In this paper, we propose a model for optimizing weight allocation of heterogeneous data based on Q-learning algorithm to improve the accuracy of DTIs prediction. We obtain multiple drug-drug similarity matrices and target-target similarity matrices from heterogeneous data through different similarity calculation methods, and then optimize the linear weighted weights of these similarity matrices based on Q-learning algorithm. Finally, we perform the experiment of DTIs prediction based on NRLMF model. To evaluate the performance and advantages of our proposed QLNRLMF method, we perform a series of experiments on four benchmark datasets to demonstrate the feasibility, efficiency, and accuracy of our proposed method. There are two main advantages to our approach. In our study, we use AUC and AUPR as evaluation indicators to evaluate the performance of our proposed method. On the one hand, it can be seen from the experimental results that our method achieves better results on the four benchmark datasets compared with other methods. On the other hand, we add reinforcement learning method, which speeds up the acquisition of optimal weight and enables us to predict DTIs more effectively.

Through this experiment, we can find that the integration of multiple information can improve the prediction accuracy of DTIs to some extent, and the rational allocation of the weight of these information also plays a key role in the prediction

performance of DTIs. One of our future work directions is to further propose better optimization algorithms for weight distribution. For example, we can change the action space from discrete to continuous to make the weight distribution more accurate, so as to further improve the accuracy of DTIs prediction.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

JS, YL, and HW: conception. JS: experiment and analysis of data. JS, LC and QF: preparation of the manuscript. JC: supervision. All authors contributed to the article and approved the submitted version.

## FUNDING

The research was funded by the National Key Research and Development Program of China (2020YFC2006602), the National Natural Science Foundation of China (62176175, 62073231, 62072324, 61902272, 61902271, 61876217, 61876121, 61772357); The Key Research and Development Program of Jiangsu Province (BE2020026).

## REFERENCES

- Bagherian, M., Sabeti, E., Wang, K., Sartor, M. A., Nikolovska-Coleska, Z., and Najarian, K. (2021). Machine Learning Approaches and Databases for Prediction of Drug-Target Interaction: a Survey Paper. *Brief. Bioinform.* 22, 247–269. doi:10.1093/bib/bbz157
- Byrd, R. H., Hribar, M. E., and Nocedal, J. (1999). An Interior Point Algorithm for Large-Scale Nonlinear Programming. *SIAM J. Optim.* 9 (4), 877–900. doi:10.1137/S1052623497325107
- Chen, R., Liu, X., Jin, S., Lin, J. W., and Liu, J. (2018). Machine Learning for Drug-Target Interaction Prediction. *Mol. : A J. Synth. Chem. Nat. Product. Chem.* 23, n. pag. doi:10.3390/molecules23092208
- Davis, J., and Goadrich, M. H. (2006). “The Relationship between Precision-Recall and ROC Curves,” in Proceedings of the 23rd international conference on Machine learning, n. pag. doi:10.1145/1143844.1143874
- Ding, Y., Tang, J., and Guo, F. (2021). Identification of Drug-Target Interactions via Multi-View Graph Regularized Link Propagation Model. *Neurocomputing* 2021, n.p.-ag. doi:10.1016/j.neucom.2021.05.100
- Ding, Y., Tang, J., and Guo, F. (2019). Identification of Drug-Target Interactions via Fuzzy Bipartite Local Model. *Neural Comput. Appl.* 32, 10303–10319. doi:10.1007/s00521-019-04569-z
- Ding, Y., Tang, J., and Guo, F. (2019). Identification of Drug-Side Effect Association via Multiple Information Integration with Centered Kernel Alignment. *Neurocomputing* 325, 211–224. doi:10.1016/j.neucom.2018.10.028
- Ding, Y., Tang, J., and Guo, F. (2019). Identification of Drug-Side Effect Association via Semisupervised Model and Multiple Kernel Learning. *IEEE J. Biomed. Health Inform.* 23, 2619–2632. doi:10.1109/jbhi.2018.2883834
- Ding, Y., Tang, J., and Guo, F. (2020). Identification of Drug-Target Interactions via Dual Laplacian Regularized Least Squares with Multiple Kernel Fusion. *Knowledge-Based Syst.* 204, 106254. doi:10.1016/j.knsys.2020.106254
- Ding, Y., Tang, J., and Guo, F. (2020). Human Protein Subcellular Localization Identification via Fuzzy Model on Kernelized Neighborhood Representation. *Appl. Soft Comput.* doi:10.1016/j.asoc.2020.106596
- Ding, Y., Tang, J., and Guo, F. (2017). Identification of Drug-Target Interactions via Multiple Information Integration. *Inf. Sci.* 418–419, 546–560. doi:10.1016/j.ins.2017.08.045
- Ezzat, A., Wu, M., Li, X., and Kwok, C. (2019). Computational Prediction of Drug-Target Interactions Using Chemogenomic Approaches: an Empirical Survey. *Brief. Bioinformatics* 2019, n. pag. doi:10.1093/bib/bby002
- Gönen, M. (2012). Predicting Drug-Target Interactions from Chemical and Genomic Kernels Using Bayesian Matrix Factorization. *Bioinformatics* 28, 2304–2310. doi:10.1093/bioinformatics/bts360
- Günther, S., Kuhn, M., Dunkel, M., Campillos, M., Sen-ger, C., Petsalaki, E., et al. (2008). “SuperTarget and Matador: Resources for Exploring Drug-Target Relationships. *Nucleic Acids Res.* 36, D919–D922. doi:10.1093/nar/gkm863
- Guo, F., Ding, Y., Li, Z., and Tang, J. (2015). Identification of Protein-Protein Interactions by Detecting Correlated Mutation at the Interface. *J. Chem. Inf. Model.* 55 (9), 2042–2049. doi:10.1021/acs.jcim.5b00320
- Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M. (2003). Development of a Chemical Structure Comparison Method for Integrated Analysis of Chemical and Genomic Information in the Metabolic Pathways. *J. Am. Chem. Soc.* 125, 11853–11865. doi:10.1021/JA036030U
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2007). KEGG for Linking Genomes to Life and the Environment. *Nucleic Acids Res.* 36, D480–D484. doi:10.1093/nar/gkm882

- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets. *Nucleic Acids Res.* 40, D109–D114. doi:10.1093/nar/gkr988
- Liu, B., Pliakos, K., Vens, C., and Tsoumakas, G. (2021). Drug-Target Interaction Prediction via an Ensemble of Weighted Nearest Neighbors with Interaction Recovery. *ArXiv abs/2012.12325*, n. pag. doi:10.1007/S10489-021-02495-Z
- Liu, H., Ren, G., Chen, H., Liu, Q., Yang, Y., and Zhao, Q. (2020). “Predicting lncRNA-miRNA Interactions Based on Logistic Matrix Factorization with Neighborhood Regularized. *Knowl. Based Syst.* 191, 105261. doi:10.1016/j.knsys.2019.105261
- Liu, Y., Wu, M., Miao, C., Zhao, P., and Li, X. (2016). Neighborhood Regularized Logistic Matrix Factorization for Drug-Target Interaction Prediction. *PLoS Comput. Biol.* 12, n. pag. doi:10.1371/journal.pcbi.1004760
- Lo, Y.-C., Rensi, S. E., Torng, W., and Altman, R. B. (2018). Machine Learning in Chemoinformatics and Drug Discovery. *Drug Discov. Today* 23, 1538–1546. doi:10.1016/j.drudis.2018.05.010
- Mei, J.-P., Kwoh, C.-K., Yang, P., Li, X.-L., and Zheng, J. (2013). Drug-target Interaction Prediction by Learning from Local Information and Neighbors, 29): 238–245. doi:10.1093/bioinformatics/bts670
- Mei, J.-P., Kwoh, C.-K., Yang, P., Li, X.-L., and Zheng, J. (2013). Drug-target Interaction Prediction by Learning from Local Information and Neighbors. *Bioinformatics* 29, 238–245. doi:10.1093/bioinformatics/bts670
- Peng, L., Liao, B., Zhu, W., Li, Z., and Li, K. (2017). Predicting Drug-Target Interactions with Multi-Information Fusion. *IEEE J. Biomed. Health Inform.* 21, 561–572. doi:10.1109/JBHI.2015.2513200
- Qu, J., Chen, X., Yin, J., Zhao, Y., and Li, Z. (2019). Prediction of Potential miRNA-Disease Associations Using Matrix Decomposition and Label Propagation. *Knowl. Based Syst.* 186, n. pag. doi:10.1016/J.KNOSYS.2019.104963
- Smith, T. F., and Waterman, M. S. (1981). Identification of Common Molecular Subsequences. *J. Mol. Biol.* 147, 195–197. doi:10.1016/0022-2836(81)90087-5
- Tabei, Y., Pauwels, E., Stoven, V., Takemoto, K., and Yamanishi, Y. (2012). Identification of Chemogenomic Features from Drug-Target Interaction Networks Using Interpretable Classifiers. *Bioinformatics* 28, i487–i494. doi:10.1093/bioinformatics/bts412
- van Laarhoven, T., and Marchiori, E. (2013). Predicting Drug-Target Interactions for New Drug Compounds Using a Weighted Nearest Neighbor Profile. *PLoS ONE* 8, n. pag. doi:10.1371/journal.pone.0066952
- Wang, Y.-C., Zhang, C.-H., Deng, N.-Y., and Wang, Y. (2011). Kernel-based Data Fusion Improves the Drug-Protein Interaction Prediction. *Comput. Biol. Chem.* 35, 353–362. doi:10.1016/j.compbiolchem.2011.10.003
- Wang, H., Tang, J., Ding, Y., and Guo, F. (2021). Exploring Associations of Non-coding RNAs in Human Diseases via Three-Matrix Factorization with Hypergraph-Regular Terms on center Kernel Alignment. *Brief. Bioinform.* doi:10.1093/bib/bbaa409
- Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., et al. (2008). DrugBank: a Knowledgebase for Drugs, Drug Actions and Drug Targets: a Knowledgebase for Drugs, Drug Actions and Drug Targets. *Nucleic Acids Res.* 36, D901–D906. doi:10.1093/nar/gkm958
- Xia, Z., Wu, L.-Y., Zhou, X., and Wong, S. T. (2010). Semi-supervised Drug-Protein Interaction Prediction from Heterogeneous Biological Spaces. *BMC Syst. Biol.* 4, S6. doi:10.1186/1752-0509-4-S2-S6
- Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of Drug-Target Interaction Networks from the Integration of Chemical and Genomic Spaces. *Bioinformatics* 24, i232–i240. doi:10.1093/bioinformatics/btn162
- Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S. (2010). Drug-target Interaction Prediction from Chemical, Genomic and Pharmacological Data in an Integrated Framework. *Bioinformatics* 26, i246–i254. doi:10.1093/bioinformatics/btq176
- Yan, X.-Y., Zhang, S.-W., and He, C.-R. (2019). Prediction of Drug-Target Interaction by Integrating Diverse Heterogeneous Information Source with Multiple Kernel Learning and Clustering Methods. *Comput. Biol. Chem.* 78, 460–467. doi:10.1016/j.compbiolchem.2018.11.028
- Yang, M., Wu, G., Zhao, Q., Li, Y., and Wang, J. (2021). Computational Drug Repositioning Based on Multi-Similarities Bilinear Matrix Factorization. *Brief. Bioinformatics* 2021, n. pag. doi:10.1093/bib/bbaa267
- Zheng, X., Ding, H., Mamitsuka, H., and Zhu, S. (2013). “Collaborative Matrix Factorization with Multiple Similarities for Predicting Drug-Target Interactions,” in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, n. pag. doi:10.1145/2487575.2487670
- Zhou, H., Wang, H., Tang, J., Ding, Y., and Guo, F. (2021). Identify ncRNA Subcellular Localization via Graph Regularized Klocal Hyperplane Distance Nearest Neighbor Model on Multikernel Learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi:10.1109/TCBB.2021.3107621

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sun, Lu, Cui, Fu, Wu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.