International Journal of Engineering

Prediction of Drug-target Protein Interaction Based on the Minimization of Weighted Nuclear Norm and Similarity Graph between Drugs and Target Proteins

A. Ghanbari Sorkhi*[a], S. M. R. Hashemi[b], H. Yarmohammadi[c], M. Iranpour Mobarakeh[d]

[a] Faculty of Electrical and Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran
[b] Young Researchers and Elite Clu, Qazvin Branch, Islamic Azad University, Qazvin, Iran
[c] Faculty of Computer Engineering, Shahrood university of technology, Shahrood, Iran
[d] Computer engineering and It department, Payam Noor University, Tehran, Iran

P A P E R   I N F O

A B S T R A C T

Identification of drug-target protein interaction plays an important role in the drug discovery process. Given the fact that prediction experiments are time-consuming, tedious, and very costly, the computational prediction could be a proper solution for decreasing search space for evaluation of the interaction between drug and target. In this paper, a novel approach based on the known drug-target interactions based on similarity graphs is proposed. It was shown that use of this method was a low-ranking issue and WNNM (weighted nuclear norm minimization) method was applied to detect the drug-target interactions. In the proposed method, the interaction between the drug and the target is encoded by graphs. Also known drug-target interaction, drug-drug similarity, target-target and combination of similarities were used as input. The proposed method was performed on four benchmark datasets, including enzymes (Es), ion channels (IC), G protein-coupled receptors (GPCRs), and nuclear receptors (NRs) based on the AUC and AUPR criteria. Finally, the results showed the improved performance of the proposed method.

doi: 10.5829/ije.2021.34.07a.18

1. INTRODUCTION

The evaluation of the drug-target interactions (DTIs) has attracted the attention of researchers in the field of pharmaceutical science, recently [1]. Accordingly, extensive efforts have been dedicated to the assessment of drug repositioning to discover the interaction between new targets and the existing drugs. In fact, DTI is defined as detection factor between the target and drug interaction that leads to changes in the drug's behavior/use. On the other hand, the identification of these interactions will minimize the adverse side effects of drugs [2]. Wet-lab experiments to recognize these potential interactions are cost and time consuming. Therefore, computational prediction (CP) methods have been used in recent years [3]. In general, CP methods can be divided into three categories of ligand-based [4], docking approaches [5], and chemogenomic approaches [6].

However, many chemogenomic approaches have been attracted attention of many researchers lately. These methods can be extensively used on accessible biological data [7]. In fact, these methods use data that includes process information simultaneously to predict. Here, information about processes means the diagram of the chemical structure and genomic sequence for drugs and targets. This general technique is divided into two categories of feature-based and similarity-based methods. Supervised machine learning methods are exploited in the feature-based technique. In fact, the methods include feature vectors of sets of drug-target pairs along with class labels that show the presence of interaction (positive instances) and absence of interaction (negative instances) [8].

*Corresponding author email: ali.ghanbari@mazust.ac.ir (A. Ghanbari Sorkhi)

In similarity-based methods, two similarity matrices related to drugs and similarity targets along with the interaction matrix which represents the interaction between drug pairs and targets are used, respectively [9].

These similarities usually arise through the chemical structures for the drug as well as through the protein sequence alignment for the target. Similarity-based methods have many positive features [10].

Unlike feature-based methods, similarity-based methods do not require a feature extraction or feature selection, which is a difficult and complex process.

Computational similarity criteria have recently been developed and widely used, the similarity of the chemical structure of drugs as well as the similarity of genomic sequences of targets are examples of this.

Due to the direct relationship between similarity-based approaches and kernel methods, similarity-based methods have better performance in prediction.

Similarity matrices show relationships between drugs and genes through chemical space and genomic space, respectively.

These features represent the superiority of similarity-based approaches over other approaches.

In the present research, we used a method based on Low-Rank Matrix Approximation (LRMA) according to weighted nuclear norm minimization (WNNM). In addition, the graph of drug-drug similarity and target-target similarity, and drug-target interaction was used to improve the performance of the proposed method. The details of the proposed method and the steps of the algorithm are described in the following sections.

2. Proposed Method

The drug-target interaction was shown with x-matrix, where the rows represent the drugs, and the columns represent the target. The matrix value is indicative of drug-target interactions. Since all interactions are not known, they are a relative matrix of observations which are expressed as follows:

$$Y = R.X \qquad (1)$$

In Equation (1), R is a subsampling operator. In this binary matrix, the value of 1 is indicative of known interaction and the value of zero shows unknown interaction or absence of interaction. A sampled DTI relative matrix is available. The goal of this equation is to estimate the x-matrix from known Ys and Rs. X is a low-rank matrix that needs to be retrieved. To this end, Equation (2) was applied:

$$\min_{X} \quad rank(X) \quad such\ that\ Y = R.X \qquad (2)$$

There are numerous methods for minimization of rank in various fields of vision and machine learning,

which have attracted the attention of many researchers in this field. One of the most important methods is nuclear norm minimization (NNM), which can guarantee the matrix rank exactly under some limited and theoretical conditions. Nonetheless, the NNM method is unable to make an exact approximation of the matrix rank for various real applications since it often tends to minimize the grade components too much. These methods are used to reconstruct the data by applying additional rank constraints to the estimated matrix. Given the fact that the direct minimization of the rank is an NP-hard problem, it is difficult to solve. In general, WNNM is used to minimize the matrix's rank. The nuclear norm of x-matrix, shown by $\|X\|_*$, is the sum of its singular values. For instance, in $\|X\|_* = \sum_i \sigma_i$ is the singular value of the x-matrix. The goal of NNM is retrieving the low-rank x-matrix from its degraded observation Y matrix by minimizing the $\|X\|_*$.

Recently, NNM-based methods are used in various areas, including removing noise from video, background extraction, and subspace clustering. Nonetheless, nuclear norm is often accepted as convex substitution of matrix rank. Although it has a theoretical guarantee, singular value thresholding (SVT) model reduces degree variables too much for NNM since it treats components of different degrees equally, and therefore, cannot estimate the matrix rank accurately. Numerous methods are proposed to improve the NNM performance. For intrinsic reconstruction, by solving an NNM problem, low-grade noise input can most likely be solved. In this method, nuclear norm proximal (NNP) can be defined as follows:

$$\hat{X} = prox_{\lambda\|.\|_*}(Y) = argmin_X \|Y - R.X\|_F^2 + \lambda\|X\|_* \qquad (3)$$

Equation (3) can be solved by applying a norm threshold action on singular values of the observation matrix in the form of Equation (4) [11, 12]:

$$\hat{X} = US_{\frac{\lambda}{2}}(\Sigma)V^T \qquad (4)$$

where $Y = U\Sigma V^T$ is a SVD of Y, and $S_{\frac{\lambda}{2}}(\Sigma)$ is the norm threshold in the Σ convex matrix with the $\frac{\lambda}{2}$ parameter. For each convex component, Σ_{ii} exists in Σ. The norm threshold function can be defined in the form of Equation (5):

$$S_{\frac{\lambda}{2}}(\Sigma)_{ii} = max\left(\Sigma_{ii} - \frac{\lambda}{2}, 0\right) \qquad (5)$$

While solving the equation is simple, the NNM has some limitations. The nuclear norm treats all singular values equally and ignores previous knowledge that often exists for matrix values. For instance, larger singular values of the data matrix are usually more important than smaller values in the most vision applications since they show the main components of the data. Different weights must be visually assigned to

different individual values so that the NNM flexibility is commensurate with the real scenarios. To correct the NNM's weakness, recent advancements have shown that the minimization of weighted nuclear norm can achieve a better matrix rank approximation, compared to NNM, which innovatively equates inverse weight with singular values. Researchers have proposed the WNNM method to improve NNM flexibility. The weighted imbalanced norm of the matrix is defined in the form of Equation (6):

$$\|X\|_{w,*} = \sum_i |w_i \sigma_i(X)| \tag{6}$$

where, $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_n(X)$, $w = [w_1, w_2, \ldots, w_n]$ and $w_i \geq 0$, the non-negative weight is allocated to $\sigma_i(X)$. The weight factor increases the ability to show the main nuclear norm. Logical weights determined based on prior knowledge and understanding of the problem use the model of the corresponding nuclear norm minimization of WNNM to make a better estimation of latent data from corrupted input. In this research, the WNNM method was applied for the DTI problem. As mentioned, the present research exploited the adjacent matrix, which shows the drug-target interaction matrix.

To analyze the WNNP problem, a lemma is presented [13] which following special Lemma 1 is derived from this lemma [14]:

Lemma 1. For any $m \times n$ matrices A and B, $tr(A^T B) \leq \sum_i \sigma_i(A)\sigma_i(B)$, where $\sigma_1(A) \geq \sigma_2(A) \ldots \geq 0$ and $\sigma_1(B) \geq \sigma_2(B) \ldots \geq 0$ are the descending singular values of A and B, respectively. Equality occurs if it is only possible to find units U and V which concurrently singular value analyze A and B because
$A = U\Sigma_A V^T$, and $B = U\Sigma_B V^T$,

where the ordered eigenvalue matrices are showed by Σ_A and Σ_B with singular value $\sigma(A)$ and $\sigma(B)$ along the diagonal with the same order, respectively.

The following main theorem is concluded based on the result of Lemma 1 [15].

Theorem 1 Given $Y \in \Re^{m \times n}$, without loss of generality, it is assumed that $m \geq n$, and let $Y = U\Sigma V^T$ be the SVD of Y, where $\Sigma = \begin{pmatrix} diag(\sigma_1, \sigma_2, \ldots, \sigma_n) \\ 0 \end{pmatrix} \in \Re^{m \times n}$. $\widehat{X} = U\widehat{D}V^T$ is expressed as the universal optimal WNNP problem in (3), where $D = \begin{pmatrix} diag(d_1, d_2, \ldots, d_n) \\ 0 \end{pmatrix}$ is a diagonal non-negative matrix and the solution $(d_1, d_2, \ldots, d_n)$ is for the following convex optimization problem:

$$\min_{d_1, d_2, \ldots, d_n} \sum_{i=1}^n (\sigma_i - d_i)^2 + w_i d_i$$
$$\text{s.t. } d_1 \geq d_2 \geq \cdots \geq d_n \tag{7}$$

According to theorem 1, the WNNP problem is a new quadratic optimization problem with linear constraints whose global optimization is easily calculated by off-the-shelf convex optimization solvers. Therefore, for the non-convex WNNP problem, a global solution can be obtained through (7). The next results show that when the weights are arranged in non-descending order, the global solution (7) can be obtained in closed-form [15].

Result 1 If $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ and the weights convince

$0 \leq w_1 \leq w_2 \leq \cdots \leq w_n$, then the global optimization of (7) is $d = \max\left(\sigma - \frac{w_1}{2}, 0\right)$

The conclusion in result 1 is very useful considering that the singular values of a matrix are arranged in non-ascending order and the larger singular values usually correlate with the subspaces of the most important components of the data matrix.

Larger singular values have shrinked less to preserve original and valid information of the underneath data. Therefore, through result 1, there is an optimal closed-form solution to the WNNP problem using the weighted singular value soft-thresholding operation [15]:

$$prox_{\lambda\|.\|_*}(Y) = US_{\frac{w}{2}}(\Sigma)V^T \tag{3}$$

where $Y = U\Sigma V^T$ is the SVD of Y, and $S_{\frac{w}{2}}(\Sigma)$ is the generalized soft-thresholding operator with weight vector w

$$S_{\frac{w}{2}}(\Sigma)_{ii} = \max\left(\Sigma_{ii} - \frac{w_i}{2}, 0\right) \tag{3}$$

Also the above WNNP solver exactly decadents to the NNP solver for the traditional NNM problem when all the weights w_i are set the same.

In this matrix, the value is 1 in case of the presence of a known interaction between the drug (d_t) and target (t_j); otherwise, the value is zero. In the present study, the drug similarity matrix (S_d) and target similarity matrix (S_s) were applied in addition to the interaction matrix .

In addition, we applied the SIMCOMP similarity method [16] based on the number of common substructures in chemical structure. In fact, S_d shows the similarity of the chemical structure of drug pairs. Also S_t shows the degree of similarity between the two proteins, estimated according to the genome sequence similarity based on the amino acid sequence of target protein.

Notably, the normalized Smith-Waterman method [17] was applied for estimating this case.

in addition to the application of the introduced similarity matrix there are four other similarity matrices, including cosine (S_{cos}), correlation (S_{cor}), hamming (S_{ham}), and jaccard (S_{jac}) which were used for DTI prediction [5].

The current research also exploited five similarity matrices estimated by the drug-target interaction matrix. In fact, similarity matrices are used for DTI, as shown in Equation (7):

$$\min_{Z} \quad \|Y - W(Z))\|_F^2 + \lambda\|Z\|_* \tag{7}$$
$$+\alpha_1 Tr\left(Z^T \sum_{i=1}^{nsim} L_d^i X\right) + \alpha_2 Tr\left(Z^T \sum_{i=1}^{nsim} L_t^i X^T\right)$$

In Equation (7), $\alpha_1 > 0$ and $\alpha_2 > 0$ are balancing parameters, $Tr(.)$ is the operator of the matrix's transposes, nsim shows the number of similarity matrixes. Here, five similarity matrixes were considered. Moreover, L_d and L_t are Laplacian graph for S_d and S_t, estimated in the form of $L_d = D_d - S_d$ and $L_t = D_t - S_t$, respectively.

In this regard, D_d and D_t are degree matrices for drugs and targets, computed by $D_d^{ii} = \sum_j S_d^{ij}$ and $D_t^{ii} = \sum_j S_t^{ij}$.

In this section, the WNNM method shown in algorithm (1) is used to solve Equation (7).

Algorithm 1. multi graph regularized nuclear norm minimization [5] method combined with proposed WNNM method

---

Procedure Alg(M,A,S_d^{com}, S_t^{com})
      Sparsify: S_d^{com}, S_t^{com}
      Initialize: $\lambda, \alpha_1, \alpha_2, v_1, v_2, L_t^{com}, L_d^{com}, Y = M, Z = M^T$

$$AA \leftarrow \begin{pmatrix} A \\ \sqrt{v_1}I \\ \sqrt{v_2}I \end{pmatrix}$$

      For loop, iterate (k)

$$YY_k \leftarrow \begin{pmatrix} M \\ \sqrt{v_1}Z^T \\ \sqrt{v_2}Y \end{pmatrix}$$

$X_k \leftarrow WNNM(YY_k, AA, \lambda)$
      $Y_k \leftarrow solve - sylvester(v_1 I, \alpha_1 L_d^{com}, v_1 X_k')$
      $Z_k \leftarrow solve - sylvester(v_2 I, \alpha_2 L_t^{com}, v_2 X_k)$
      End Loop

In Algorithm 1, $S_d^{com} = S_d + S_d^{cos} + S_d^{cor} + S_d^{ham} + S_d^{jac} = \sum_{i=1}^{nsim} S_d^i$ and $S_t^{com} = S_t + S_t^{cos} + S_t^{cor} + S_t^{ham} + S_t^{jac} = \sum_{i=1}^{nsim} S_t^i$ show the combined similarity for drug and target, $D_d^{com} = diag(\sum_j S_d^{Com})$ and $D_t^{com} = diag(\sum_j S_t^{Com})$ show the combined degree matrix for the drug and target also $L_d^{com} = D_d^{com} - D_d^{com}$ and $L_t^{com} = D_t^{com} - D_t^{com}$ show the combined Laplacian matrix for the drug and target, respectively. This equation is solved using the method presented in [5]. Please refer to the mentioned article for more details.

3. Experiments and Analysis of Results

In this section, the experiments and results of the proposed method are analyzed separately.

3. 1. Dataset and Evaluation Criteria The information related to the interactions between drugs and target proteins for public databases of KEGG BRITE, RENDA, SuperTarget and DrugBank have been assessed by Yamanishi et al. [7]. Similar to Yamanishi et al. study, we applied four benchmark datasets from four different classes of target protein. In fact, these criteria are simulated from public databases. The following is a description of these datasets:
• Enzymes (Es): 445 drugs, 664 targets, and 2926 interactions were extracted in this dataset .
•Ion channels (IC): 201 drugs, 204 targets, and 1476 interactions are extracted in this dataset .
• G protein-coupled receptors (GPCRs): 223 drugs, 95 targets, and 635 interactions are extracted in this dataset .
• Nuclear receptors (NRs): 54 drugs, 26 targets, and 90 interactions are extracted in this dataset .

It is notable that the foregoing datasets were simulated from public databases, which are available with the address of http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget publicly. In the present research, cross-validation settings of leave-one-out (LOO) were used for data segmentation. Three modes of the dataset were considered in the results section. In addition, CVS for drug prediction, CVS for target prediction, and interaction prediction was introduced with titles of CVS1, CVS2, and CVS3, respectively. This segmentation was based on Mongia et al. study [5], as presented below:
• CVS1/drug prediction: All drug profiles are set aside to be used as the experiment set, which tests the algorithm's ability to predict the interactions of new drugs, that is, drugs for which no cross-information is available.
• CVS2/target prediction: The entire target profiles are set aside to be used as the experiment set to assess the algorithm's ability to predict interactions of new targets .
• CVS3/pair prediction: Random drug-target pairs are set aside as the experiment set for prediction. This a normal adjustment for validation and evaluation.

When at least one DTI is known for d_i and t_j respectively in the training data the CVS1 predicts the unknown pair (d_i, t_j). To prevent using the pairs, CV used the pairs between the drugs having $>= 2$ targets and the targets interacting with $>= 2$ drugs, which should be used in three other scenarios. Some of these pairs are selected by random for testing in each round of CV and the union of the rest of them and other entries are used for training.

However, when there are no DTIs for observation of new drugs and new targets in the training data, CVS2 and CVS3 predict new drugs and new targets respectively.

Performance of CV on drugs in CSV2, where the rows corresponding to drugs are randomly blinded for

testing and the resting rows are used for training. Also performance of CV on targets in CSV3 where the columns (accounting for targets) are randomly blinded for testing and the resting columns are used for training as well.

We have made various tasks of CV under 3 scenarios showed in Figure 1 respectively.
In addition, area under the ROC curve (AUC) and area under the Precision-Recall (AUPR) were applied to assess the performance of the proposed method according to Mongia et al. study [5].

3. 2. Analysis of Experiments' Results   This section includes a comparison of the proposed method with previous works in recent years. In the current research, we applied the techniques presented in others study for comparison [18, 19]. Notably, all methods are performed from the same data set and the same CVS. In addition, the results of other works were extracted from the articles. The results are shown in the tables below based on the AUC and AUPR criteria .

Table 1 presents a comparison of the methods based on the AUC criterion in four benchmark datasets and various CVSs. According to the results, the proposed method had acceptable performance in the evaluation of DTI in similar datasets, compared to other techniques. A very important point in these tables is related to the prediction of the target-drug pair. In this regard, the rows related to CVS3 are shown in Table 1.

Table 2 compares the methods based on the AUPR criteria and the results obtained from the techniques in four benchmark datasets in three different CVSs. According to the results, the proposed method had acceptable performance in DTI evaluation (CVS3) in all four benchmark datasets, compared to other methods.

This paper presents a new approach based on the known drug-target interactions based on similarity graphs. The weighted nuclear norm minimization method was used to identify the drug-target interactions. Our proposed method encodes the adjacency between the drug and the target by graphs. Also, known drug-
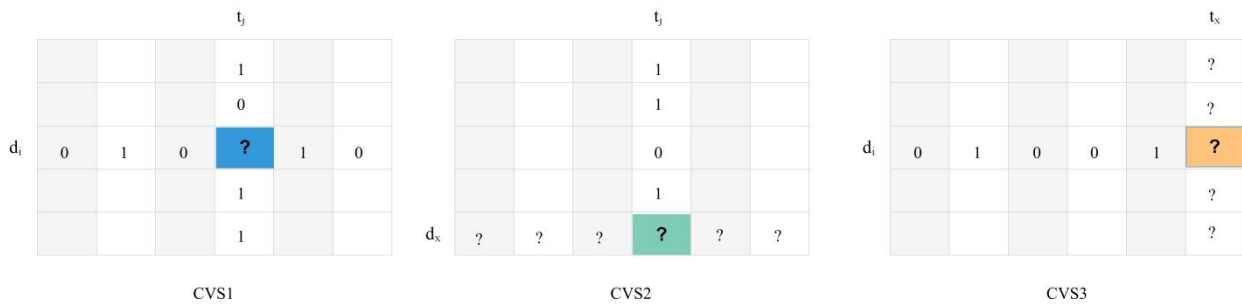

Figure 1. Presentation of cross-validation schemes for three scenarios. Each column represents a scenario. Row includes the DTI matrices, in which the entries marked with "?" are the pairs of interest to be tested

TABLE 1. Comparison of the proposed method with other techniques based on AUC criteria in four datasets in various CVSs

CVS	Dataset	[18]	[19]	[20]	[21]	Proposed method
CSV1	Es	0.9272	0.9067	0.96	0.97	**0.9721**
	IC	0.9368	0.9286	0.97	**0.98**	0.9526
	GPCRs	0.8966	0.8694	0.94	**0.96**	0.9024
	NRs	0.8373	0.8124	0.88	0.92	**0.9421**
CSV2	Es	0.7755	0.7952	0.78	0.84	**0.8512**
	IC	0.7669	0.7576	0.79	**0.94**	0.8013
	GPCRs	0.8800	0.8067	0.88	0.91	**0.9186**
	NRs	0.8615	0.8124	0.86	0.90	**0.9015**
CSV3	Es	0.9705	0.9635	0.93	0.92	**0.9512**
	IC	0.9832	0.9786	0.94	0.97	**0.9969**
	GPCRs	0.9493	0.9458	0.88	0.93	**0.9902**
	NRs	0.8679	0.9329	0.79	0.88	**0.9339**

TABLE 2. Comparison of the proposed method with other techniques based on the AUPR criteria in four databases in different CVSs

CVS	Dataset	[18]	[19]	[20]	[21]	Proposed method
CSV1	Es	0.7808	0.5465	0.87	**0.92**	0.8532
	IC	0.7786	0.7437	**0.92**	0.92	0.8011
	GPCRs	0.5989	0.5397	0.73	0.79	**0.7944**
	NRs	0.4774	0.4907	0.60	0.83	**0.7720**
CSV2	Es	0.3848	0.2409	0.40	0.73	**0.7322**
	IC	0.3538	0.3090	0.36	0.69	**0.6921**
	GPCRs	0.4059	0.3463	0.42	**0.63**	0.5812
	NRs	0.5203	0.5373	0.56	0.71	**0.7366**
CSV3	Es	0.8837	0.8093	0.80	0.82	**0.9055**
	IC	0.9373	0.8459	0.81	0.80	**0.9411**
	GPCRs	0.7543	0.6933	0.60	0.61	**0.7601**
	NRs	0.6383	0.7072	0.46	0.64	**0.7888**

target interaction, drug-drug similarity, target-target and combination of similarities have been used as input. The proposed method was performed on four benchmark based on the AUC and AUPR criteria. Eventually, the results showed an improvement in the performance of the proposed method

4. CONCLUSION

The present research proposed a novel approach to identify drug-target interactions, which applied the drug-drug, target-target, and target-drug interaction similarity graph method. In the current research, the proposed method's performance was improved by using the WNNM in order to eliminate NNM limitations in the DTI use. In addition, the proposed technique was assessed in four benchmark datasets based on the AUC and AUPR criteria. The final results were indicative of the improved performance of the proposed method compared to previous approaches in the field.

5. REFERENCES

1. Sachdev, K., Gupta, M.K., "A comprehensive review of feature based methods for drug target interaction prediction", *Journal of Biomedical Informatics*, Vol. 93, (2019), 1031-59, doi: 10.1016/j.jbi.2019.103159

2. Sachdev, K., Gupta, M.K., "A hybrid ensemble-based technique for predicting drug–target interactions", *Chemical Biology & Drug Design*, Vol. 6, (2020), 1447-1455. doi: 10.1111/cbdd.13753

3. Ezzat, A., Wu, M., Li, X.L., Kwoh, C.K., "Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey", *Briefings in Bioinformatics*, Vol. 20, (2019), 1337-1357. doi: 10.1093/bib/bby002

4. Keiser, M.J., Roth, B.L., Armbruster, B. N., Ernsberger, P., Irwin, J.J., Shoichet, B.K., "Relating protein pharmacology by ligand chemistry", *Nature Biotechnology*, Vol. 25, (2007), 197-206, doi: 10.1038/nbt1284

5. Mongia, A., Majumdar, A., "Drug-target interaction prediction using Multi Graph Regularized Nuclear Norm Minimization", *Plos one*, Vol. 15, (2020), 226-484, doi: 10.1371/journal.pone.0226484

6. Yıldırım, M.A., Goh, K.I., Cusick, M. E., Barabási, A.L., Vidal, M., "Drug—target network," *Nature Biotechnology*, Vol. 25, (2007), 1119-1126, doi: 10.1038/nbt1338

7. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., Kanehisa, M., "Prediction of drug–target interaction networks from the integration of chemical and genomic spaces", *Bioinformatics*, Vol. 24, (2008), 232-240, doi: 10.1093/bioinformatics/btn162

8. Ezzat, A., Wu, M., Li, X.L., Kwoh, C.K., "Drug-target interaction prediction via class imbalance-aware ensemble learning", *BMC Bioinformatics*, Vol. 17, (2061), 267-276, doi: 10.1186/s12859-016-1377-y

9. Thafar, M.A., Olayan, R.S., Ashoor, H., Albaradei, S., Bajic, V.B., Gao, X., "DTiGEMS :+drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques", *Journal of Cheminformatics*, Vol. 12, (2020), 1-17, doi: 10.1186/s13321-020-00447-2

10. Ding, H ,.et al., Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Briefings in Bioinformatics*, Vol. 15, No. 5, (2014), 734-747, doi: 10.1093/bib/bbt056

11. Sun, X., Y. Wang, and X. Zhang. Multi-View Subspace Clustering via Non-Convex Tensor Rank Minimization. in 2020 IEEE International Conference on Multimedia and Expo (ICME). 2020. IEEE, doi: 10.1137/20M1318006

12. Yang, Y., et al., Foreground-Background Separation via Generalized Nuclear Norm and Structured Sparse Norm Based Low-Rank and Sparse Decomposition. *IEEE Access*, (2020), Vol. 8, 84217-84229, doi: 10.1109/ACCESS.2020.2992132

13. Gu, S., et al. Weighted nuclear norm minimization with application to image denoising. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, doi: 10.1109/CVPR.2014.366

14. D. Rhea, "The case of equality in the von Neumann trace inequality," preprint, 2011, doi: 10.1016/j.laa.2015.05.015

15. Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X., & Zhang, L. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision*, Vol. 121, No. 2, (2017), 183-208, doi: 10.1007/s11263-016-0930-5

16. Hattori, M., Tanaka, N., Kanehisa, M., Goto, S., "SIMCOMP/SUBCOMP: chemical structure search servers for network analyses", *Nucleic Acids Research*, Vol. 38, (2010), 652-656, doi: 10.1093/nar/gkq367

17. AMoZ, J., "Identification of common molecular subsequences", *Journal of Molecular Biology*, Vol. 147, (1981), 195-197. doi: 10.1016/0022-2836(81)90087-5

18. Zheng, X., Ding, H., Mamitsuka, H., Zhu, S., "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 2013, 1025-1033, (2013). doi : 10.1145/2487575.2487670

19. Van, T., Laarhoven and Marchiori, E., "Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile", *Plos one*, Vol. 8, (2013), 52-566. doi: 10.1371/journal.pone.0066952

20. Ezzat, A., et al., Drug-target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 14, No. 3, (2016), 646-656. doi: 10.1109/TCBB.2016.2530062

21. Olayan, R.S., H. Ashoor, and V.B. Bajic, DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. *Bioinformatics*, Vol 34, No. 7, (2018), 1164-1173. doi: 10.1093/bioinformatics/btx731

Persian Abstract

چکیده

شناسایی تعامل بین دارو و پروتئین‌های هدف، نقش بسیار مهمی در فرآیند کشف داور دارد. از آنجایی که انجام آزمایشات پیش‌بینی این فرآیند زمانبر، بسیار پرهزینه و خسته کننده می‌باشد. پیش‌بینی محاسباتی می‌تواند یک راهکار مناسبی برای کاهش فضای جستجو برای بررسی تعامل بین دارو و هدف بجای استفاده از آزمایشات پر هزینه باشد. در این مقاله یک راهکار نوین بر اساس، تعامل‌های شناخته شده در بین دارو-هدف برپایه گراف شباهت ارائه شده است. در این مقاله نشان داده شده است که این کاربرد جز مسائل با مرتبه پایین می‌باشند که در این راستا از روش حداقل رساندن نرم هسته‌ای وزنی برای تشخیص تعامل بین دارو و هدف استفاده شده است. همچنین در این مقاله برای نمایش مجاورت بین دارو و هدف از گراف مجاورت استفاده شد که تعامل دارویی-هدف، شباهت دارو-دارو، هدف-هدف و ترکیبی از شباهت‌ها به عنوان ورودی استفاده شد. روش پیشنهادی بر روی چهار مجموعه داده معیار شامل آنزیم‌ها (Es)، کانال‌های یونی(IC)، گیرنده‌های همراه پروتئین G (GPCRs) و گیرنده‌های هسته‌ای (NRs)، بر اساس معیارهای AUC، AUPR مورد ارزیابی قرار گرفته است. نتایج بدست آمده نشان دهنده بهبود عملکرد روش پیشنهادی می‌باشد.