Predicting Drug-Target Interactions Using Neihborhood Regularied Neural Network Matrix Factorization

Qingguo Zeng

Department of Information and Computing Science, South China Normal University, Guangzhou, 510631

1 METHODS

In fact, drug-target interactions (DTI) prediction can be served as a recommendation task which aims to predict the preference of each user (drug) on the each item (gene), so the recommendation technique can be applied to the DTI prediction. Matrix factorization technique is widely used for the recommendation. Matrix factorization approximates the entries of the matrix by acting inner product on the latant vectors for the corresponding row and column, which is a fixed function, resulting in potentially limiting the performance of the model. To solve this problem, we replace inner product with a neural network which is regarded as an effective model to approximates arbitrary functions, leading to improve the performance of model. This method named Neural Network Matrix Factorization (NNMF).

Although NNMF can estimate the global structure of DTI matrix, it ignores the relationship between the similar genes or similar drugs, that is, the similarity of drugs or genes in the original data space may not be effectively maintained in the latent space. In other words, the corresponding latent vector does not well represent the characteristics of the drug or gene and thus, we can not get a satisfactory accuracy of DTI prediction. In order to overcome it, we impose neiborhood regularization constraints [1] on the model so that latent vectors corresponding to the similar drugs or genes are similar, then the original features is very likely to be inherited in the latent space, thereby enhancing the DTI prediction accuracy. We call our proposal Regularized Neural Network Matrix Factorization (NRNNMF). In the following, I will introduce the model in detail from NNMF and Neighborhood Regularization constraints.

1.1 Neural Network Matrix Factorization

In this work, NNMF is the core component. The method has been tried to be used in the recommendation system[2; 3], but so far no one has used it for DTI prediction. It models the interaction probability p_{ij} of a drug-gene pair(d_i, g_j) as followed:

$$p_{ij}(\theta) = f_\theta(u_i, v_j, u'_i \bullet v'_j) \tag{1}$$

where f_θ is a neural network with weights θ, $\{u_i, u'_i\}$ is the latent variable set of drug d_i, $\{v_j, v'_j\}$ is the latent variable set of gene g_j, $\bullet$ denotes the Hadamard product of vector or matrix. For each drug or each gene, NNMF learns two latent vectors, where the dimension of u_i and v_j is D, and the dimension of u'_i and v'_j is D' [3]. Then, the input layer of the neural network has $2D + D'$ neurons. The first D neurons are drug-specific features; the next D are gene-specific features; and the last D' are the Hadamard product between latent vectors of drug and gene. For simplicity, we denote the latent vectors of drugs and genes in different dimensions, ie $U = \{u_1^\top, u_2^\top, ..., u_m^\top\}^\top$, $V = \{v_1^\top, v_2^\top, ..., v_n^\top\}^\top$, $U' = \{u'^\top_1, u'^\top_2, ..., u'^\top_m\}^\top$ and $V' = \{v'^\top_1, v'^\top_2, ..., v'^\top_n\}^\top$. In addition, write $Z = \{U, V, U', V', \theta\}$ for all the parameters of the model.

From another point of view, NNMF transforms the DTI prediction into classification. Due to the sparseness of DTI matrix, we suffer from a serious class-imbalance problem. Meanwhile, from the background, the confidence in the a known drug-gene interaction pair and an unknown pair are not the same. The former one have been experimentally verified, thus they are more credible[1]. In this paper, we use the importance weighting strategy to deal with the difference importance between positive and negative sample, that is, when training the model, each known pair is treated as $c(c \geqslant 1)$ positive training samples, and each unknown pair is treated as a negative sample, where the constant c controls the importance of the observed interacting drug-gene pair.

Then, by making the assumption that all the pair is independent, the likelihood function is as follows:

$$\begin{aligned}
&p(Y|Z)\\
&=\left(\prod_{1\leqslant i\leqslant m,1\leqslant j\leqslant n,y_{ij}=1}\left[p_{ij}(\theta)^{y_{ij}}(1-p_{ij}(\theta))^{(1-y_{ij})}\right]^{c}\right)\left(\prod_{1\leqslant i\leqslant m,1\leqslant j\leqslant n,y_{ij}=0}p_{ij}(\theta)^{y_{ij}}(1-p_{ij}(\theta))^{(1-y_{ij})}\right)\\
&=\left(\prod_{1\leqslant i\leqslant m,1\leqslant j\leqslant n,y_{ij}=1}p_{ij}(\theta)^{cy_{ij}}(1-p_{ij}(\theta))^{(1-y_{ij})}\right)\left(\prod_{1\leqslant i\leqslant m,1\leqslant j\leqslant n,y_{ij}=0}p_{ij}(\theta)^{cy_{ij}}(1-p_{ij}(\theta))^{(1-y_{ij})}\right)\\
&=\prod_{i=1}^{m}\prod_{j=1}^{n}p_{ij}(\theta)^{cy_{ij}}(1-p_{ij}(\theta))^{(1-y_{ij})},
\end{aligned}\tag{2}$$

where y_{ij} is the element of the i^{th} row and j^{th} column of the DTI matrix Y.

In addition, we place zero-mean spherical Gaussian priors on the latent vectors of drugs and genes to regularize the model, and thus avoiding over-fitting.

$$\begin{aligned}
p(U|\sigma_U^2)=\prod_{i=1}^{m}\mathcal{N}(u_i|0,\sigma_U^2\mathbf{I}), \qquad p(U|\sigma_V^2)=\prod_{i=1}^{n}\mathcal{N}(v_i|0,\sigma_V^2\mathbf{I}),\\
p(U'|\sigma_{U'}^2)=\prod_{i=1}^{m}\mathcal{N}(u_i'|0,\sigma_{U'}^2\mathbf{I}), \qquad p(V'|\sigma_{V'}^2)=\prod_{i=1}^{m}\mathcal{N}(v_i'|0,\sigma_{V'}^2\mathbf{I})
\end{aligned}\tag{3}$$

Where $\sigma_U,\sigma_V,\sigma_U'$ and σ_V' are scalar controling the variances of Gaussian distribution, $\mathbf{I}$ denotes the identity matrix. According to Baye's rule, we have

$$p(Z|Y,\sigma_U^2,\sigma_V^2,\sigma_{U'}^2,\sigma_{V'}^2)\propto p(Y|Z)p(U|\sigma_U^2)p(V|\sigma_V^2)p(U'|\sigma_{U'}^2)p(V'|\sigma_{V'}^2).\tag{4}$$

For simplicity, we take log of the posterior distribution, that is

$$\begin{aligned}
\log p(Z|Y,\sigma_U^2,\sigma_V^2,\sigma_{U'}^2,\sigma_{V'}^2)=&\left(\sum_{i=1}^{m}\sum_{j=1}^{n}cy_{ij}p_{ij}(\theta)+(1-y_{ij})(1-p_{ij}(\theta))\right)-\frac{1}{2\sigma_U^2}\sum_{i=1}^{m}||u_i||_2^2\\
&-\frac{1}{2\sigma_V^2}\sum_{i=1}^{n}||v_i||_2^2-\frac{1}{2\sigma_{U'}^2}\sum_{i=1}^{m}||u_i'||_2^2-\frac{1}{2\sigma_{V'}^2}\sum_{i=1}^{n}||v_i'||_2^2+const,
\end{aligned}\tag{5}$$

Where $const$ is a constant term independent of the model parameter Z and is not need to be considered when solving the parameters. The parameters of the model can be learned by maximizing the posterior probability, which is equivalent to minimizing the following objective functions,

$$\min_{Z}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-1)(1-p_{ij}(\theta))-cy_{ij}p_{ij}(\theta)\right)+\frac{1}{2\sigma_U}||U||_F^2+\frac{1}{2\sigma_V}||V||_F^2+\frac{1}{2\sigma_{U'}}||U'||_F^2+\frac{1}{2\sigma_{V'}}||V'||_F^2\tag{6}$$

Where $||\cdot||_F$ denote the Frobenius norm of a matrix.

1.2 Neighborhood Regularization Constraint

Although NNMF can predict DTI, it does not take the local structure of drugs and genes into account, which potentially limits the performance of the model. To overcome it, we add the neighborhood regularization constraint to NNMF model to ensure the similarity between the drug or gene can be maintained in the latent space, that is, minimizing the distance between the drug or gene and its neighbors from the perspective of optimization. Note that we only consider the nearest K_1 neighbors of each gene or drug to prevent from large noise caused by defective similarity measurement or other faulty operation. As a result, we further improve the accuracy of DTI prediction. Next, we show the derivation[1].

For a drug d_i, its K_1 nearest neighbors are denoted as a set $N(d_i)$. In this paper, We use an adjacency matrix $A = \{a_{ik}\}$ to describe the neighborhood information of the drug d_i,

$$a_{ik} = \begin{cases} s^d_{iu}, & if\ d_k \in N(d_i) \\ 0, & otherwise, \end{cases} \tag{7}$$

Where s^d_{iu} is the entry in i^{th} row and u^{th} column of matrix S^d which describe the drug similarity scores. Likewise, write $N(g_j)$ for a set including K_1 nearest neighbors of gene g_j, s^g_{jl} for the entry in j^{th} row and l^{th} column of matrix S^g and $B = \{b_{jl}\}$ for describing the neighborhood information of the gene g_j, then

$$b_{il} = \begin{cases} s^g_{jl}, & if\ d_l \in N(g_j) \\ 0, & otherwise, \end{cases} \tag{8}$$

Then, the objective function for the neighborhood regularization constraint for the drug in the $D-$dimensional space latent space is shown below,

$$\begin{aligned} &\sum_{i=1}^{m}\sum_{k=1}^{m} a_{ik}||u_i - u_k||_2^2 \\ &= \left[\sum_{i=1}^{m}(\sum_{k=1}^{m} a_{ik})u_i u_i^\top + \sum_{k=1}^{m}(\sum_{i=1}^{m} a_{ik})u_k u_k^\top\right] - tr(U^\top AU) - tr(U^\top A^\top U) \\ &= tr(U^\top L^d U) \end{aligned} \tag{9}$$

Where $tr(\cdot)$ is the trace of a matrix, $L^d = (D^d + \tilde{D}^d) - (A + A^\top)$, both D^d and $\tilde{D}^d$ are two diagonal matrices whose diagonal elements are $D^d_{ii} = \sum_{k=1}^{m} a_{ik}$ and $\tilde{D}^d_{kk} = \sum_{i=1}^{m} a_{ik}$, respectively. Likewise, we can get other three neighborhood regularization constraint terms as follows:

$$\begin{aligned} &\sum_{i=1}^{m}\sum_{k=1}^{m} a_{ik}||u'_i - u'_k||_2^2 = tr(U'^\top L^d U'), \\ &\sum_{j=1}^{n}\sum_{l=1}^{n} b_{jl}||v_j - v_l||_2^2 = tr(V^\top L^g V), \\ &\sum_{j=1}^{n}\sum_{l=1}^{n} b_{jl}||v'_j - v'_l||_2^2 = tr(V'^\top L^g V'), \end{aligned} \tag{10}$$

Where $L^g = (D^g + \tilde{D}^g) - (B + B^\top)$, D^g and $\tilde{D}^g$ are also two diagonal matrices whose diagonal elements are $D^g_{jj} = \sum_{l=1}^{n} b_{jl}$ and $\tilde{D}^d_{ll} = \sum_{j=1}^{n} b_{jl}$, respectively.

1.3 Neighborhood Regularized Neural Network Matrix Factorization

By summarizing the last two subsection, we can get the final objective function of the model by adding Eq(9) and Eq(10) to Eq(6).

$$\begin{aligned} \min_Z &\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - 1)(1 - p_{ij}(\theta)) - cy_{ij}p_{ij}(\theta)\right) + \frac{1}{2}tr[U^\top(\lambda_d + \alpha L^d)U] \\ &+ \frac{1}{2}tr[V^\top(\lambda_g + \beta L^g)V] + \frac{1}{2}tr[U'^\top(\lambda'_d + \alpha' L^d)U'] + \frac{1}{2}tr[V'^\top(\lambda'_g + \beta' L^g)V] \end{aligned} \tag{11}$$

Where $\lambda_d = \frac{1}{\sigma^2_U}, \lambda_g = \frac{1}{\sigma^2_V}, \lambda'_d = \frac{1}{\sigma^2_{U'}}, \lambda'_g = \frac{1}{\sigma^2_{V'}}, \alpha, \beta, \alpha', \beta'$ are all the regularization coefficients. The solution of this optimization problem shown in Eq(11) is described in Algorithm 1.

So far, the construction and solution of the model has been completely described, but there is still a problem——"cold start". The new drugs or genes do not have interaction information, and it is likely to incorrectly make the prediction by directly using NRNNMF model. Therefore, we replace the original latent vector with linear combination of latent vectors between K_2 nearest neighbors for the new drugs or genes. Then, the interaction probability p_{ij} of a drug-gene pair (d_i, g_j) is changed into:

$$\hat{p}_{ij}(\theta) = f_\theta(\hat{u}_i, \hat{v}_j, \hat{u'}_i \bullet \hat{v'}_j) \tag{12}$$

Where,

Algorithm 1 NRNNMF

Input: $Y, S^d, S^g, c, D, D', K_1, K_2, \lambda_d, \lambda_g, \lambda'_d, \lambda'_g, \alpha, \beta, \alpha', \beta', max\_iter, early\_stop\_max\_iters$
Output: $Z = \{U, U', V, V', \theta\}$

Initialize Z randomly;
Construct the adjacency matrices A and B according to Eq (7) and Eq (8), respectively;
Compute the neiborhood regularization matrices L^d and L^t according to Eq (9) and Eq(10)
Set $RMSE^0_{valid} = \inf$
for $t = 1, ..., max\_iter$ **do**
    Sampling a batch of training data randomly;
    Compute the partial derivative of Z about the function Eq (11) and using gradient descent algorithm to update the parameter Z;
    Compute the RMSE for the validation data, and write it for $RMSE^t_{valid}$, respectively;
    if $RMSE^t_{valid} < RMSE^{t-1}_{valid}$ **then**
        Set $RMSE^{t-1}_{valid} = RMSE^t_{valid}$
    else if early_stop_iters == early_stop_max_iters **then**
        break
    else if $RMSE^t_{valid} > RMSE^{t-1}_{valid}$ **then**
        early_stop_iters $+= 1$
    end if
end for

$$\hat{u}_i = \begin{cases} u_i, & if\ d_i \in D^+ \\ \frac{1}{\sum_{k \in N^+(d_i)}} s^d_{ik} u_k, & if\ d_i \in D^-, \end{cases} \tag{13}$$

$$\hat{v}_j = \begin{cases} v_j, & if\ g_j \in T^+ \\ \frac{1}{\sum_{l \in N^+(g_j)}} s^g_{jl} v_l, & if\ g_j \in T^-, \end{cases} \tag{14}$$

$$\hat{u'}_i = \begin{cases} u'_i, & if\ d_i \in D^+ \\ \frac{1}{\sum_{k \in N^+(d_i)}} s^d_{ik} u'_k, & if\ d_i \in D^-, \end{cases} \tag{15}$$

$$\hat{v'}_j = \begin{cases} v'_j, & if\ g_j \in T^+ \\ \frac{1}{\sum_{l \in N^+(g_j)}} s^g_{jl} v_l, & if\ g_j \in T^-, \end{cases} \tag{16}$$

Where $D^+ = \{d_i | \sum_{j=1}^n y_{ij} > 0, \forall 1 \leqslant i \leqslant m\}$ and $T^+ = \{g_j | \sum_{i=1}^n y_{ij} > 0, \forall 1 \leqslant j \leqslant n\}$ are defined as set of positive drugs and targets, respectively. Then, the set of negative drugs and negative genes are denotes as $D^- = D / D^+$ and $T^- = T / T^+$, respectively.

2 RESULTS

In the previous section, we have described the principles and solutions of NRNNMF model in detail. Next, we will evaluate the performance of the model and compare it with some state-of-the-arts. In the field of biological information, Receiver Operator Characteristic (ROC) and Precision-Recall (PR) curves is commonly used for evaluating the performance of the model. For highly skewed data, ROC curve tend to give an overoptimistic picture of a models performance, while Precision-Recall(PR) curves give a more informative one. There are usually few positive DTIs. Thus, we mainly use AUPR(Area under Precision-Recall Curve) as the evaluation metric, meanwhile AUC (Area under Receiver Operator Characteristic Curve) is only for reference. Due to randomness involved in the optimization(such as data partitioning, parameter initialization, etc), the performance of models are evaluated through 5 trials of 10 fold cross-validation and calculate the average AUC and AUPR.

2.1 Comparisons with the State-of-the-Arts

To validate the model, we compare the proposed NRNNMF method with five state-of-the-arts methods for DTI prediction, including NetLapPLS[5], BLM-NII[6], WNN-GIP[7], CMF[8] and LMF[9]. Tab 1 show the result of all the methods.

In summary, NRNNMF achieved the best performance with AUC and AUPR on the given dataset, especially the AUPR. The AUPR obtained by our model is 0.915930, which is 6.55% higher than that attained by the second best model LMF. In the following, we have a detailed look at why NRNNMF outperforms the other methods from the structure of the model.

Table 1. Results on DTI prediction

Method	AUC	AUPR
NetLapPLS	0.960554 ± 0.002423	0.825708 ± 0.007145
BLM-NII	0.980521 ± 0.002222	0.824273 ± 0.009971
WNN-GIP	0.957403 ± 0.003900	0.665780 ± 0.022617
CMF	0.860660 ± 0.014519	0.590802 ± 0.034268
LMF	0.963534 ± 0.002745	0.850424 ± 0.006054
NRNNMF	$\mathbf{0.982524 \pm 0.002589}$	$\mathbf{0.915930 \pm 0.007159}$

2.2 The learning Ability of Neural Network Matrix Factorization

In addition to NRNNMF, there are still two methods based on matrix factorization mentioned above, CMF and LMF. In order to research the different performance of these matrix factorization techniques, neighborhood regularization constraint in NRNNMF is removed and NRNNMF degenerates into NNMF. Then, we obtain AUC and AUPR of NNMF model up to 0.971009 and 0.873134, respectively, which are higher than those of CMF and LMF. This result shows that NNMF has a strong learning ability inheriting from neural network and is able to learn the features of drugs and genes. Secondly, the NNMF does not use a fixed function for matrix factorization, leading to make the model more flexible and extract more information.

2.3 The Influence of Neighborhood Regularization Constraint

To exploit the local structure of the interaction data, we add neighborhood regularization constraint. Thus, the feature of data in original data space can be effectively maintained in the latent space. We examine how the neighborhood regularization constraint affects the accuracy of DTI prediction and whether different number of neighbors used for regularization have different impact on DTI prediction accuracy.

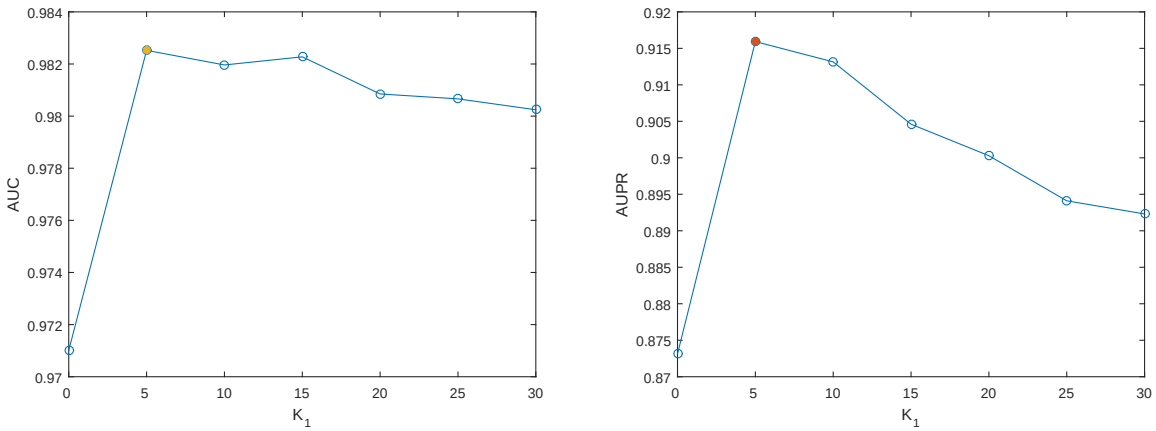

Fig. 1. How the number of neighbors K_1 affects the DTI prediction results, where the solid point is the optimal point.

Fig 1 shows how the number of neighbors (K_1) affects the AUC and AUPR obtain by our model. We find that neighborhood regularization constraint is able to effectively improve the performance of the model. As shown in Fig 1, the optimal values of K_1 is 5. However, the more neighbor information involved in training, the smaller AUC and AUPR attained by the model on DTI prediction, which is mainly caused by two reasons: 1) The technique of calculating the similarity of drug or gene may be defective, which potentially leads to the introduction of noise when using neighbors information. With the increase of K_1, the noise in the data increases, and the prediction accuracy of the model decreases; 2) Neighbors information plays the role of regularization in the model. In machine learning, the regularization technique is mainly used to limit the complexity of the model and prevent the model from over-fitting. However, if the regularization term is too large, it will result in underfitting. Therefore, utilizing excessive neighbors information give rise to the large regularization, reducing the performance of the model.

3 DISCUSSION

This paper proposes a novel machine-learning approach to predict drug-gene interaction, namely Neihborhood Regularied Neural Network Matrix Factorization. Our approach use neural network matrix factorization to factorize highly sparse DTI matrix into four low-rank matrix, two of which effectively describe the feature of drug and the others well represents the characteristic of gene. Then, the importance weighting strategy is used to highlight the importance of positive observations so as to alleviate the problem of category imbalance. Morevoer, neihborhood regularization constraint is utilized to ensures that the data in the latent space retains the characteristics of drugs or genes in the original space, leading to further improve the accuracy of DTI prediction. In the experiment, our method outperforms several state-of-the-arts.

In addition, our approach can be easily extended to solve the multi-type DTI interaction by modifying the structure of output layer in neural network[4].

However, there are some unresolved problems in our model. The time for training our model is nearly 20 times that of other models. In addiction, There are many hyperparameters in the model, which costs us a lot of time to determine them for acquiring an optimal model. In the future, we will use Bayesian optimization based on Gaussian Process to achieve automatic determination on hyperparameters[10].

REFERENCES

[1]Liu Y, Wu M, Miao C, et al. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. PLoS Comput Biol, 2016, 12(2): e1004760.
[2]Dziugaite G K, Roy D M. Neural network matrix factorization. arXiv preprint arXiv:1511.06443, 2015.
[3]Stolee J, Patterson N. Matrix Factorization with Neural Networks and Stochastic Variational Inference.
[4]Wang Y, Zeng J. Predicting drug-target interactions using restricted Boltzmann machines. Bioinformatics, 2013, 29(13): i126-i134.
[5]Xia Z, Wu L Y, Zhou X, et al. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. BMC systems biology. BioMed Central Ltd, 2010, 4(Suppl 2): S6.
[6]Mei J P, Kwoh C K, Yang P, et al. Drugtarget interaction prediction by learning from local information and neighbors. Bioinformatics, 2013, 29(2): 238-245.
[7]van Laarhoven T, Marchiori E. Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile[J]. PloS one, 2013, 8(6): e66952.
[8]Zheng X, Ding H, Mamitsuka H, et al. Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013: 1025-1033.
[9]Johnson C C. Logistic matrix factorization for implicit feedback data. Advances in Neural Information Processing Systems, 2014, 27.
[10]Snoek J, Larochelle H, Adams R P. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems. 2012: 2951-2959.