A tri-factor matrix factorization method for drug target interaction prediction.*

Dongliang Wu
Fudan University
Shanghai, China
xxx@xxx.edu.cn

Abstract

Abstract.

1 Introduction

In this paper, we proposed a novel framework to combine multiple similarity matrices in one model that improves prediction accuracy by taking advantages of both latent factor approaches and multi-similarities. Our contributions are summarized as follows:

- To incorporate the similarity of drugs(targets) for DTI prediction more effectively, we developed a tri-factor matrix factorization method to approximate similarity matrices.
- We provide a new way to combine multiple similarities with missing values.
- We empirically show that our model outperforms other state-of-the-art methods on real world datasets.

2 Method

2.1 Notation and Problem Setting

First we briefly formalize the problem of drug-target interaction(DTI) prediction. Given a drug target interaction network $\boldsymbol{G}$, a set of drugs $\boldsymbol{D} = \{d_1, d_2, \cdots, d_{N_d}\}$ and a set of targets $\boldsymbol{T} = \{t_1, t_2, \cdots, t_{N_t}\}$, we can construct a DTI matrix $\boldsymbol{Y}$. Here $\boldsymbol{Y}$ is an $N_t \times N_d$ binary matrix whose rows and columns represent targets and drugs respectively. If drug d_i and target t_j interact with each other, then $\boldsymbol{Y}_{ij}$ = 1, else $\boldsymbol{Y}_{ij}$ = 0.

Let $\{\boldsymbol{S}_d^1, \boldsymbol{S}_d^2, \cdots, \boldsymbol{S}_d^{M_d}\}$ be a set of drug similarity matrices, where each is a $N_d \times N_d$ matrix, and M_d is the number of drug similarity matrices. $\boldsymbol{S}_{\boldsymbol{d}}^{k_d}(i,j)$ is the similarity score between drug d_i and drug d_j in the k-th drug similarity matrix. Similarly let $\{\boldsymbol{S}_t^1, \boldsymbol{S}_t^2, \cdots, \boldsymbol{S}_t^{M_t}\}$ be a set of target similarity matrices, where each is a $N_t \times N_t$ matrix, and M_t is the number of drug similarity matrices. $\boldsymbol{S}_{\boldsymbol{t}}^{k_t}(i,j)$ is the similarity score between target t_i and target t_j in the k-th target similarity matrix.

Let $\hat{\boldsymbol{Y}}$ be a score matrix, where the (i,j)-element of $\hat{\boldsymbol{Y}}$, i.e. $\hat{\boldsymbol{Y}}(i,j)$, shows the score that drug d_i and target t_j interact with each other. The objective of this article is to estimate $\hat{\boldsymbol{Y}}$ so that $\hat{\boldsymbol{Y}}$ should be consistent with $\boldsymbol{Y}$.

2.2 Multiple Similarities Tri-factor Matrix Factorization (MSTMF)

We first introduce a state-of-the-art matrix factorization model as our basic. Given a drug-target interaction matrix $\boldsymbol{Y}_{N_t \times N_d}$, a general framework of matrix factorization is:

$$(\boldsymbol{P}, \boldsymbol{Q}) = \arg\min_{\boldsymbol{P},\boldsymbol{Q}} \|\boldsymbol{H} \odot (\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^T)\|_F^2 \qquad (1)$$

where $\boldsymbol{P} \in \boldsymbol{R}^{N_t \times k}$ and $\boldsymbol{Q} \in \boldsymbol{R}^{N_d \times k}$, corresponding to feature spaces of targets and drugs, respectively. $\boldsymbol{T}$ is a $N_t \times N_d$ matrix, in which $\boldsymbol{H}_{ij} = 1$ if $\boldsymbol{Y}_{ij}$ is a known drug-target pair(has interaction with each other or not); otherwise $\boldsymbol{H}_{ij} = 0$. $\boldsymbol{H} \odot \boldsymbol{Z}$ denotes the element-wise product of matrices $\boldsymbol{H}$ and $\boldsymbol{Z}$. By estimating $\boldsymbol{P}$ and $\boldsymbol{Q}$, we can reconstruct $\boldsymbol{Y}$ so that unknown drug-target pairs have prediction scores.

On the basis of above, we can plus a few more information into the framework. Suppose we have two similarity matrices, $\boldsymbol{S_t}$ for targets and $\boldsymbol{S_d}$ for drugs. A simple trying is leting the inner product of two drug feature vectors to approximate the corresponding similarity of two drugs ($\boldsymbol{S_d} \approx \boldsymbol{Q}\boldsymbol{Q}^T$) and this is also the case with target similarity ($\boldsymbol{S_t} \approx \boldsymbol{P}\boldsymbol{P}^T$).(Our previous work[Zheng *et al.*, 2013])

However, this is not a good enough method to capture the correlation between two feature vectors. Because the correlation between two vectors maybe much more complex. On the other hand, this method is too restrictive to $\boldsymbol{P}$ and $\boldsymbol{Q}$ so that $\hat{\boldsymbol{Y}} = \boldsymbol{P}\boldsymbol{Q}^T$ may give poor predictions. So we add extra factors $\boldsymbol{W_d}$ and $\boldsymbol{W_t}$ to absorb the different scales of $\boldsymbol{S_d}, \boldsymbol{S_t}, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{Y}$. I.e. let $\boldsymbol{S_d} \approx \boldsymbol{Q}\boldsymbol{W_d}\boldsymbol{Q}^T$, and $\boldsymbol{S_t} \approx \boldsymbol{P}\boldsymbol{W_t}\boldsymbol{P}^T$.[Tang *et al.*, 2013] So $\boldsymbol{W_d}(\boldsymbol{W_t})$ captures the drug feature(target feature) correlation. For multiple similarities, we can factorize them separately using different $\boldsymbol{W_d}$ or $\boldsymbol{W_t}$.

Thus put all above together, we can write out the entire loss function:

*some thanks.

$$\begin{aligned}\mathcal{L} &= \|\boldsymbol{H} \odot (\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^T)\|_F^2 \\ &+ \sum_{k_t=1}^{M_t} \|\boldsymbol{H}_{\boldsymbol{t}}^{k_t} \odot (\boldsymbol{S}_{\boldsymbol{t}}^{k_t} - \boldsymbol{P}\boldsymbol{W}_{\boldsymbol{t}}^{k_t}\boldsymbol{P}^T)\|_F^2 \\ &+ \sum_{k_d=1}^{M_d} \|\boldsymbol{H}_{\boldsymbol{d}}^{k_d} \odot (\boldsymbol{S}_{\boldsymbol{d}}^{k_d} - \boldsymbol{Q}\boldsymbol{W}_{\boldsymbol{d}}^{k_d}\boldsymbol{Q}^T)\|_F^2 \\ &+ \lambda(\|\boldsymbol{P}\|_F^2 + \|\boldsymbol{Q}\|_F^2 + \sum_{k_t=1}^{M_t} \|\boldsymbol{W}_{\boldsymbol{t}}^{k_t}\|_F^2 + \sum_{k_d=1}^{M_d} \|\boldsymbol{W}_{\boldsymbol{d}}^{k_d}\|_F^2)\end{aligned} \tag{2}$$

where $\boldsymbol{H}_{\boldsymbol{t}}^{(\cdot)}, \boldsymbol{H}_{\boldsymbol{d}}^{(\cdot)}$ are introduced to handle missing values of some similarity matrices, and $\lambda(\cdot)$ is the regularization term. The parameters can be learned by minimizing the above loss function by using gradient descent algorithm. We write out the gradient and update rule directly below. The detailed algorithm is shown in Algorithm 1.

$$\begin{aligned}\mathcal{L} &= \|\boldsymbol{H} \odot (\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^T)\|_F^2 \\ &+ \lambda_t \sum_{k_t=1}^{M_t} \|\boldsymbol{H}_{\boldsymbol{t}}^{k_t} \odot (\boldsymbol{S}_{\boldsymbol{t}}^{k_t} - \boldsymbol{P}\boldsymbol{W}_{\boldsymbol{t}}^{k_t}\boldsymbol{P}^T)\|_F^2 \\ &+ \lambda_d \sum_{k_d=1}^{M_d} \|\boldsymbol{H}_{\boldsymbol{d}}^{k_d} \odot (\boldsymbol{S}_{\boldsymbol{d}}^{k_d} - \boldsymbol{Q}\boldsymbol{W}_{\boldsymbol{d}}^{k_d}\boldsymbol{Q}^T)\|_F^2 \\ &+ \lambda_l(\|\boldsymbol{P}\|_F^2 + \|\boldsymbol{Q}\|_F^2 + \sum_{k_t=1}^{M_t} \|\boldsymbol{W}_{\boldsymbol{t}}^{k_t}\|_F^2 + \sum_{k_d=1}^{M_d} \|\boldsymbol{W}_{\boldsymbol{d}}^{k_d}\|_F^2)\end{aligned} \tag{3}$$

Gradient:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{P}} &= \boldsymbol{H} \odot (\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^T)\boldsymbol{Q} \cdot (-1) \\ &+ \sum_{k_t=1}^{M_t} \Big[\Big(\boldsymbol{H}_{\boldsymbol{t}}^{k_t} \odot (\boldsymbol{S}_{\boldsymbol{t}}^{k_t} - \boldsymbol{P}\boldsymbol{W}_{\boldsymbol{t}}^{k_t}\boldsymbol{P}^T) \Big)^T \boldsymbol{P}\boldsymbol{W}_{\boldsymbol{t}}^{k_t} \cdot (-1) \\ &+ \boldsymbol{H}_{\boldsymbol{t}}^{k_t} \odot (\boldsymbol{S}_{\boldsymbol{t}}^{k_t} - \boldsymbol{P}\boldsymbol{W}_{\boldsymbol{t}}^{k_t}\boldsymbol{P}^T)\boldsymbol{P}(\boldsymbol{W}_{\boldsymbol{t}}^{k_t})^T \cdot (-1) \Big] \\ &+ \lambda \boldsymbol{P}\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{Q}} &= \boldsymbol{H} \odot (\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^T)\boldsymbol{P} \cdot (-1) \\ &+ \sum_{k_d=1}^{M_d} \Big[\Big(\boldsymbol{H}_{\boldsymbol{d}}^{k_d} \odot (\boldsymbol{S}_{\boldsymbol{d}}^{k_d} - \boldsymbol{Q}\boldsymbol{W}_{\boldsymbol{d}}^{k_d}\boldsymbol{Q}^T) \Big)^T \boldsymbol{Q}\boldsymbol{W}_{\boldsymbol{d}}^{k_d} \cdot (-1) \\ &+ \boldsymbol{H}_{\boldsymbol{d}}^{k_d} \odot (\boldsymbol{S}_{\boldsymbol{d}}^{k_d} - \boldsymbol{Q}\boldsymbol{W}_{\boldsymbol{d}}^{k_d}\boldsymbol{Q}^T)\boldsymbol{Q}(\boldsymbol{W}_{\boldsymbol{d}}^{k_d})^T \cdot (-1) \Big] \\ &+ \lambda \boldsymbol{Q}\end{aligned} \tag{5}$$

$$\begin{aligned}&\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}_{\boldsymbol{t}}^{k_t}} = \boldsymbol{P}^T(\boldsymbol{S}_{\boldsymbol{t}}^{k_t} - \boldsymbol{P}\boldsymbol{W}_{\boldsymbol{t}}^{k_t}\boldsymbol{P}^T)\boldsymbol{P} \cdot (-1) + \lambda \boldsymbol{W}_{\boldsymbol{t}}^{k_t} \\ &\quad (k_t = 1, \cdots, M_t)\end{aligned} \tag{6}$$

$$\begin{aligned}&\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}_{\boldsymbol{d}}^{k_d}} = \boldsymbol{Q}^T(\boldsymbol{S}_{\boldsymbol{d}}^{k_d} - \boldsymbol{Q}\boldsymbol{W}_{\boldsymbol{d}}^{k_d}\boldsymbol{Q}^T)\boldsymbol{Q} \cdot (-1) + \lambda \boldsymbol{W}_{\boldsymbol{d}}^{k_d} \\ &\quad (k_d = 1, \cdots, M_d)\end{aligned} \tag{7}$$

Update Rule:

$$\begin{aligned}\boldsymbol{P} &= \boldsymbol{P} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{P}} \\ \boldsymbol{Q} &= \boldsymbol{Q} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{Q}} \\ \boldsymbol{W}_{\boldsymbol{t}}^{k_t} &= \boldsymbol{W}_{\boldsymbol{t}}^{k_t} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}_{\boldsymbol{t}}^{k_t}} \\ &\quad (k_t = 1, \cdots, M_t) \\ \boldsymbol{W}_{\boldsymbol{d}}^{k_d} &= \boldsymbol{W}_{\boldsymbol{d}}^{k_d} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}_{\boldsymbol{d}}^{k_d}} \\ &\quad (k_d = 1, \cdots, M_d)\end{aligned} \tag{8}$$

Algorithm 1 the algorithm of MSTMF

Input:
drug-target interaction matrix $\boldsymbol{Y}$;
target similarity matrices, $\{\boldsymbol{S}_t^1, \boldsymbol{S}_t^2, \cdots, \boldsymbol{S}_t^{M_t}\}$;
drug similarity matrices, $\{\boldsymbol{S}_d^1, \boldsymbol{S}_d^2, \cdots, \boldsymbol{S}_d^{M_d}\}$;
the number of latent factors , K;
weight matrix, $\boldsymbol{H}$, $\boldsymbol{H}_{\boldsymbol{t}}^{k_t}$, $\boldsymbol{H}_{\boldsymbol{d}}^{k_d}$;
$(k_t = 1, \cdots, M_t, k_d = 1, \cdots, M_d)$
learning rate, η;

1: Initialize $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{W}_{\boldsymbol{t}}^{k_t}$, $\boldsymbol{W}_{\boldsymbol{d}}^{k_d}$ randomly;
$(k_t = 1, \cdots, M_t, k_d = 1, \cdots, M_d)$
2: **repeat**
3: Update $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{W}_{\boldsymbol{t}}^{k_t}$, $\boldsymbol{W}_{\boldsymbol{d}}^{k_d}$ using Eq.(8);
4: **until** Stopping criterion is met;

Output:
predicted interaction matrix $\hat{\boldsymbol{Y}} = \boldsymbol{P}\boldsymbol{Q}^T$;

2.3 Multiple Similarities Tri-factor Matrix Factorization With Weight (MSTMF-Weight)

Last section we introduced a method of combining multiple similarities to predict drug-target interaction. In this section, we propose another kind of method. As we can see, different similarity matrices provide different aspects of discriptions of similarities. And all similarity values are in the interval of [0,1]. So we combined the multiple similarity matrices linearly and acquired an unique similarity matrix as follows:

$$\begin{aligned}\boldsymbol{S}_{\boldsymbol{t}} &= \sum_{k_t=1}^{M_t} \alpha_{k_t} \boldsymbol{S}_{\boldsymbol{t}}^{k_t} \quad (\boldsymbol{S}_{\boldsymbol{d}} = \sum_{k_d=1}^{M_d} \beta_{k_d} \boldsymbol{S}_{\boldsymbol{d}}^{k_d}) \\ s.t. &\sum_{k_t=1}^{M_t} \alpha_{k_t} = 1 \quad (\sum_{k_d=1}^{M_d} \beta_{k_d} = 1)\end{aligned} \tag{9}$$

where α_{k_t} (β_{k_d}) is the weight of k_t-th (k_d-th) similarity matrix for targets (drugs). [Zheng *et al.*, 2013]

Thus the objective function can be written as follow:

$$\begin{aligned}\mathcal{L} &= \|\boldsymbol{H} \odot (\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^T)\|_F^2 \\ &+ \|(\sum_{k_t=1}^{M_t} \alpha_{k_t} \boldsymbol{S_t^{k_t}} - \boldsymbol{P}\boldsymbol{W_t}\boldsymbol{P}^T)\|_F^2 \\ &+ \|(\sum_{k_d=1}^{M_d} \beta_{k_d} \boldsymbol{S_d^{k_d}} - \boldsymbol{Q}\boldsymbol{W_d}\boldsymbol{Q}^T)\|_F^2 \\ &+ \lambda(\|\boldsymbol{P}\|_F^2 + \|\boldsymbol{Q}\|_F^2 + \|\boldsymbol{W_t}\|_F^2 + \|\boldsymbol{W_d}\|_F^2) \\ &+ \lambda_\alpha \|\boldsymbol{\alpha}\|_F^2 + \lambda_\beta \|\boldsymbol{\beta}\|_F^2 \\ s.t. \quad &\sum_{k_t=1}^{M_t} \alpha_{k_t} = 1, \quad \sum_{k_d=1}^{M_d} \beta_{k_d} = 1.\end{aligned} \tag{10}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_{M_t})^T, \boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_{M_d})^T$. And the *gradient*:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{P}} &= \boldsymbol{H} \odot (\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^T)\boldsymbol{Q} \cdot (-1) \\ &+ (\boldsymbol{S_t^T} - \boldsymbol{P}\boldsymbol{W_t^T}\boldsymbol{P}^T)\boldsymbol{P}\boldsymbol{W_t} \cdot (-1) \\ &+ (\boldsymbol{S_t} - \boldsymbol{P}\boldsymbol{W_t}\boldsymbol{P}^T)\boldsymbol{P}\boldsymbol{W_t^T} \cdot (-1) + \lambda \boldsymbol{P}\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{Q}} &= \boldsymbol{H} \odot (\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^T)\boldsymbol{Q} \cdot (-1) \\ &+ (\boldsymbol{S_d^T} - \boldsymbol{Q}\boldsymbol{W_d^T}\boldsymbol{Q}^T)\boldsymbol{Q}\boldsymbol{W_d} \cdot (-1) \\ &+ (\boldsymbol{S_d} - \boldsymbol{Q}\boldsymbol{W_d}\boldsymbol{Q}^T)\boldsymbol{Q}\boldsymbol{W_d^T} \cdot (-1) + \lambda \boldsymbol{Q}\end{aligned} \tag{12}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{W_t}} = \boldsymbol{P}^T(\boldsymbol{S_t} - \boldsymbol{P}\boldsymbol{W_t}\boldsymbol{P}^T)\boldsymbol{P} \cdot (-1) + \lambda \boldsymbol{W_t} \tag{13}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{W_d}} = \boldsymbol{Q}^T(\boldsymbol{S_d} - \boldsymbol{Q}\boldsymbol{W_d}\boldsymbol{Q}^T)\boldsymbol{Q} \cdot (-1) + \lambda \boldsymbol{W_d} \tag{14}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \alpha_{k_t}} = \Bigg\| \Big[&\sum_{k_t=1}^{M_t-1} \alpha_{k_t} \boldsymbol{S_t^{k_t}} + (1 - \sum_{k_t=1}^{M_t-1} \alpha_{k_t}) \boldsymbol{S_t^{M_t}} - \boldsymbol{P}\boldsymbol{W_t}\boldsymbol{P}^T \Big] \\ &\odot (\boldsymbol{S_t^{k_t}} - \boldsymbol{S_t^{M_t}}) \Bigg\|_F^2 + \lambda_\alpha \alpha_{k_t}, (k_t = 1, \cdots, M_t - 1)\end{aligned} \tag{15}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \beta_{k_d}} = \Bigg\| \Big[&\sum_{k_d=1}^{M_d-1} \beta_{k_d} \boldsymbol{S_d^{k_d}} + (1 - \sum_{k_d=1}^{M_d-1} \beta_{k_d}) \boldsymbol{S_d^{M_d}} - \boldsymbol{Q}\boldsymbol{W_d}\boldsymbol{Q}^T \Big] \\ &\odot (\boldsymbol{S_d^{k_d}} - \boldsymbol{S_d^{Md}}) \Bigg\|_F^2 + \lambda_\beta \beta_{k_d}, (k_d = 1, \cdots, M_d - 1)\end{aligned} \tag{16}$$

where

$$\boldsymbol{S_t} = \sum_{k_t=1}^{M_t} \alpha_{k_t} \boldsymbol{S_t^{k_t}} \quad , \quad \boldsymbol{S_d} = \sum_{k_d=1}^{M_d} \beta_{k_d} \boldsymbol{S_d^{k_d}}$$

Update Rule:

$$\begin{aligned}\boldsymbol{P} &= \boldsymbol{P} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{P}} \\ \boldsymbol{Q} &= \boldsymbol{Q} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{Q}} \\ \boldsymbol{W_t} &= \boldsymbol{W_t} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{W_t}} \\ \boldsymbol{W_d} &= \boldsymbol{W_d} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{W_d}} \\ \alpha_{k_t} &= \alpha_{k_t} - \eta_\alpha \cdot \frac{\partial \mathcal{L}}{\partial \alpha_{k_t}} \\ &(k_t = 1, \cdots, M_t - 1) \\ \alpha_{Mt} &= 1 - \sum_{k_t=1}^{Mt-1} \alpha_{k_t} \\ \beta_{k_d} &= \beta_{k_d} - \eta_\beta \cdot \frac{\partial \mathcal{L}}{\partial \beta_{k_d}} \\ &(k_d = 1, \cdots, M_d - 1) \\ \beta_{Md} &= 1 - \sum_{k_d=1}^{Md-1} \beta_{k_d}\end{aligned} \tag{17}$$

The detailed algorithm is shown in Algorithm 2.

Algorithm 2 the algorithm of MSTMF-weight

Input:
drug-target interaction matrix $\boldsymbol{Y}$;
target similarity matrices, $\{\boldsymbol{S}_t^1, \boldsymbol{S}_t^2, \cdots, \boldsymbol{S}_t^{M_t}\}$;
drug similarity matrices, $\{\boldsymbol{S}_d^1, \boldsymbol{S}_d^2, \cdots, \boldsymbol{S}_d^{M_d}\}$;
the number of latent factors , K;
weight matrix, $\boldsymbol{H}$;
learning rate, $\eta, \eta_\alpha, \eta_\beta$;
1: Initialize $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W_t}, \boldsymbol{W_d}$ randomly;
2: **repeat**
3: Update $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W_t}, \boldsymbol{W_d}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ using Eq.(17);
4: **until** Stopping criterion is met;
Output:
predicted interaction matrix $\hat{\boldsymbol{Y}} = \boldsymbol{P}\boldsymbol{Q}^T$;

3 Experiment

3.1 Datasets

Drug-Target Interaction Data

We used 4 benchmark datasets *GPCR*, *Enzyme*, *IC*, *NR* , which were provided by [Ashburner *et al.*, 2000]. They were all collected from four general databases and frequently used in predicting drug-target interactions.[Gönen, 2012] The statistics of these four datasets are shown in table 1.

Similarities over Drugs and Targets

We use the following two types of similarities for drugs and targets, respectively.
Targets:

	GPCR	*Enzyme*	*IC*	*NR*
# of interactions	635	2926	1476	90
# of drugs	223	445	210	54
# of targets	95	664	204	26

Table 1: The statistics of datasets

- **GS(Genomic sequence) similarity** is computed by a normalized Smith-Waterman score[Ashburner *et al.*, 2000] between two target sequences.
- **GO(Gene Ontology) similarity** is the overlap of GO annotations [Yamanishi *et al.*, 2008] of two targets. We considered only one options of GO—molecular functions (MF), because it is more effective than biological processes(BP) and cellular component (CC).

Drugs:

- **CS(Chemical structure) similarity** is computed by the number of shared substructures in chemical structures between two drugs.
- **ATC(Anatomical Therapeutic Chemical) similarity**[Skrbo *et al.*, 2004] is computed by using a hierarchical drug classification system. We used a general method in [Lin, 1998] to compute the similarity between two nodes (drugs) in this classification tree.

3.2 Competing Methods

We compare the proposed framework **MSTMF** and **MSTMF-weight** with the following representative methods:
BLM: This method turns the problem of predicting edges in a bipartite graph(drug target interaction graph here) into a binary supervised problem.[Gönen, 2012]
PKM: This is a straightforward SVM-based method and uses drug-target interactions as instances, implying that the similarity between drug-target pairs needs to be computed. PKM then uses the similarity matrix (kernel) of drug-target pairs with known labels to train an SVM classifier, which can then predict the scores of arbitrary drug-target pairs.[Jacob and Vert, 2008]
LapRLS: This method attempts to directly estimate interaction score matrix $\boldsymbol{F}$ for drugs and targets, separately, which we denote by $\boldsymbol{F_d}$ and $\boldsymbol{F_t}$, respectively. For drugs, LapRLS minimizes the squared loss between $\boldsymbol{Y}$ and $\boldsymbol{F_d}$ with a regularized term of $\boldsymbol{S_d}$ and $\boldsymbol{F_d}$. The same procedure can be performed for targets. Finally, $\boldsymbol{F}$ is obtained by averaging over $\boldsymbol{F_d}$ and $\boldsymbol{F_t}$. [Xia *et al.*, 2010]
NetLapRLS: This method is a modification of LapRLS to consider drug-target interactions more directly.
GIP: This method generates a Gaussian kernel from the interaction profiles and liearly combines the Gaussian kernel with the drug similarity matrix into the kernel. Then it incorporates the idea of PKM, which computes a pairwise kernel over drugtarget pairs. Finally, the resultant kernel matrix is used in the straight-forward framework of regularized least squares, which minimize the square loss between the score function and the true labels with a regularizer.[van Laarhoven *et al.*, 2011]
KBMF2K: The idea behind KBMF2K is first to project the drug and target spaces into two low-dimensional spaces through kernels (similarity matrices) and then to estimate drugtarget interactions under the lowdimensional spaces. [Gönen, 2012]

3.3 Experimental Settings

Evaluation The evaluation was done by 5×10-fold cross-validation(CV). That is we randomly divided the dataset into 10 folds for 5 times. Each time we implemented a 10-fold cross-validation on the datasets. The results were averaged over the total 50(= 5×10) runs. For each round of 5-fold cross-validation, we selected parameter values by doing 9-fold cross-validation on the training dataset, and the rest for evaluation. We used the metric AUPR (Area Under the Precision-Recall curve) to evaluated the performance of our mothod. Instead of AUC (Area Under the ROC Curve), AUPR punishes highly ranked false positives much more than AUC. This point is important since only highly ranked drug-target pairs in prediction will be biologically or chemically tested later in an usual drug discovery process, meaning that highly ranked false positives should be avoided. [Zheng *et al.*, 2013]

We consider only one type of prediction problems:[Ding *et al.*, 2013]

- Pair prediction—predicting a interaction for the new pair of a drug and a target that already has one or more interactions.

3.4 Experimental Results

Table 2 shows the performance comparison of the proposed methods TMF, MSTMF, MSTMF-weight and other methods introduced before. Note that TMF is the MSTMF with only one type of similarities, and CMF is the MSCMF with only one type of similarities. (chemical structure similarity for drugs and genomic sequence similarity for targets). While MSCMF, MSTMF, MSTMF-weight use all kinds of the obtained similarities: CS, GS, ATC, and GOMF.

From the results, we can observe that: For single similarity setting, the results of TMF significantly better than any other methods in all four benchmark datasets. And for multiple similarities setting, MSTMF and MSTMF-weight outperform MSCMF as well. While the disparities between MSTMF and MSTMF-weight are small .

The above observations demonstrate the effectiveness of our approachs.

3.5 Parameter selection

Regularization parameter For simplicity, our model has only one regularization parameter λ. We vary the value of λ as $\{2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}, 2^{3}\}$. Figure 1 shows how the performance of MSTMF varies with the parameter λ. As we can see, the MSTMF is very stable with respect to the parameter λ.
Dimension of latent factors We investigate the effect of dimensionality of latent factors. As is shown in figure 2. We

Table 2: AUPR values obtained by 5×10-fold cross validation. The highest AUPR value for each column is highlighted in boldface.
Pair prediction

Methods	GPCR	Enzyme	IC	NR
BLM	0.666	0.830	0.847	0.509
PKM	0.401	0.633	0.617	0.499
LapRLS	0.639	0.826	0.804	0.539
NetLapRLS	0.660	0.840	0.848	0.545
GIP	0.710	0.869	0.889	0.596
KBMF2K	0.686	0.796	0.876	0.508
CMF	0.739	0.896	0.938	0.568
TMF	**0.794**	**0.913**	**0.945**	**0.611**
MSCMF	0.743	0.913	0.938	0.562
MSTMF	0.787	**0.927**	0.949	**0.658**
MSTMF-weight	**0.796**	0.925	**0.951**	0.645

can see that, the result converges until the number of dimensions was larger than 256.

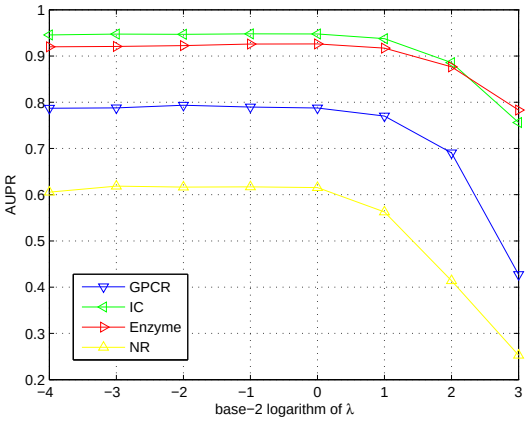

Figure 1: The effect of parameter λ

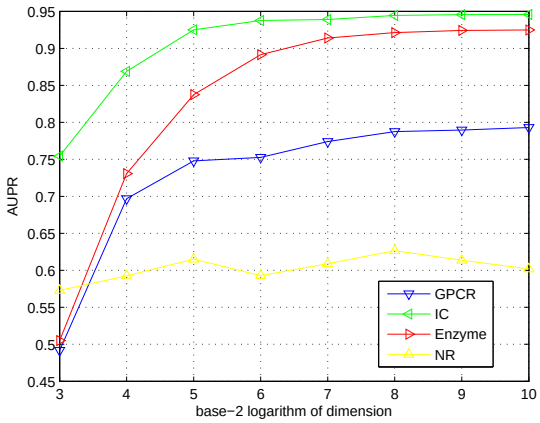

Figure 2: The effect of dimensions.
The dimensionality is increased from 8 to 1024.

4 Conclusion

In this paper, we proposed two new methods (MSTMF and MSTMF-weight) to model multiple similarities for drug target interaction prediction. And we experimentally improved that our methods achieved significantly better performance than any other methods using AUPR as a criteria. In addition, our proposed method MSTMF has only one parameter λ, and the prediction result is quite stable with respect to its parameter. This is a big advantage which MSCMF doesn't have.

5 Acknowledgments

This is Acknowledgments.

References

[Ashburner *et al.*, 2000] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.

[Ding *et al.*, 2013] Hao Ding, Ichigaku Takigawa, Hiroshi Mamitsuka, and Shanfeng Zhu. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Briefings in bioinformatics*, page bbt056, 2013.

[Gönen, 2012] Mehmet Gönen. Predicting drug–target interactions from chemical and genomic kernels using bayesian matrix factorization. *Bioinformatics*, 28(18):2304–2310, 2012.

[Jacob and Vert, 2008] Laurent Jacob and Jean-Philippe Vert. Protein-ligand interaction prediction: an improved

chemogenomics approach. *Bioinformatics*, 24(19):2149–2156, 2008.

[Lin, 1998] Dekang Lin. An information-theoretic definition of similarity. In *ICML*, volume 98, pages 296–304, 1998.

[Skrbo *et al.*, 2004] Armin Skrbo, Begler Begovi, and Selma Skrbo. [classification of drugs using the atc system (anatomic, therapeutic, chemical classification) and the latest changes]. *Medicinski arhiv*, 58(1 Suppl 2):138141, 2004.

[Tang *et al.*, 2013] Jiliang Tang, Xia Hu, Huiji Gao, and Huan Liu. Exploiting local and global social context for recommendation. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2712–2718. AAAI Press, 2013.

[van Laarhoven *et al.*, 2011] Twan van Laarhoven, Sander B. Nabuurs, and Elena Marchiori. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics*, 27(21):3036–3043, November 2011.

[Xia *et al.*, 2010] Zheng Xia, Ling-Yun Wu, Xiaobo Zhou, and Stephen TC Wong. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC systems biology*, 4(Suppl 2):S6, 2010.

[Yamanishi *et al.*, 2008] Yoshihiro Yamanishi, Michihiro Araki, Alex Gutteridge, Wataru Honda, and Minoru Kanehisa. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 24(13):i232–i240, 2008.

[Zheng *et al.*, 2013] Xiaodong Zheng, Hao Ding, Hiroshi Mamitsuka, and Shanfeng Zhu. 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*, pages 1025–1033. ACM, 2013.