

Collaborative Matrix Factorization with Multiple Similarities for Predicting Drug-Target Interactions

Xiaodong Zheng[†], Hao Ding[†], Hiroshi Mamitsuka[‡] and Shanfeng Zhu[†]

[†]Shanghai Key Lab of Intelligent Information Processing and School of Computer Science,
Fudan University, Shanghai 200433, China

[‡]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji 611-0011, Japan

ABSTRACT

We address the problem of predicting new drug-target interactions from three inputs: known interactions, similarities over drugs and those over targets. This setting has been considered by many methods, which however have a common problem of allowing to have only one similarity matrix over drugs and that over targets. The key idea of our approach is to use more than one similarity matrices over drugs as well as those over targets, where weights over the multiple similarity matrices are estimated from data to automatically select similarities, which are effective for improving the performance of predicting drug-target interactions. We propose a factor model, named Multiple Similarities Collaborative Matrix Factorization (MSCMF), which projects drugs and targets into a common low-rank feature space, which is further consistent with weighted similarity matrices over drugs and those over targets. These two low-rank matrices and weights over similarity matrices are estimated by an alternating least squares algorithm. Our approach allows to predict drug-target interactions by the two low-rank matrices collaboratively and to detect similarities which are important for predicting drug-target interactions. This approach is general and applicable to any binary relations with similarities over elements, being found in many applications, such as recommender systems. In fact, MSCMF is an extension of weighted low-rank approximation for one-class collaborative filtering. We extensively evaluated the performance of MSCMF by using both synthetic and real datasets. Experimental results showed nice properties of MSCMF on selecting similarities useful in improving the predictive performance and the performance advantage of MSCMF over six state-of-the-art methods for predicting drug-target interactions.

Categories and Subject Descriptors

G.1.2 [Numerical Analysis]: Approximation—*least squares approximation*; I.2.6 [Artificial Intelligence]: Learning—*knowledge acquisition, parameter learning*

General Terms

Algorithm, Experimentation, Performance

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
KDD'13, August 11–14, 2013, Chicago, Illinois, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
Copyright 2013 ACM 978-1-4503-2174-7/13/08 ...\$15.00.

Keywords

Cheminformatics; Drug-target interaction; Weighted low-rank approximation; Multiple types of similarities over drugs and targets

1. INTRODUCTION

Pharmaceutical sciences are an interdisciplinary research field of fundamental sciences, including biology, chemistry and physics, and a successfully developed engineering field, which has created a major industry of our society. The objective of pharmaceutical sciences is drug discovery, which starts with finding effective interactions between drugs and targets, where drugs are chemical compounds and targets are proteins (amino acid sequences). Known drug-target interactions are however limited to a small number [8], say only less than 7,000 compounds in fact having target protein information in PubChem [22], one of the largest chemical compound databases with currently around 35 million entries. Furthermore nowadays drug discovery, i.e. finding new drug-target interactions, needs much more cost and time, because relatively similar drugs (or targets) to those in known interactions have been already examined thoroughly. In this light efficient computational methods for predicting potential drug-target interactions are useful and long-awaited [11].

Two major computational approaches are docking simulation and data mining (or machine learning). Docking simulation is biologically well-accepted but has two serious problems: 1) simulation always needs three-dimensional (3D) structures of targets which are often unavailable [4] and 2) simulation is heavily time consuming. On the other hand, machine learning is much more efficient, by which a large number of candidates can be tested within a very short period of time. A straight-forward approach in machine learning is to set up a standard classification problem over a table of instances and their features, where instances are drug-target pairs, and features are chemical descriptors (for drugs) and amino acid subsequences (for targets). Any classification methods, such as support vector machine (SVM), can be applied to this table [18].

Drug-target interactions can be represented by a binary-labeled matrix Y of drugs and targets, where an element is 1 if the corresponding drug and target interact; and 0 if they do not interact. The problem of predicting drug-target interactions is to estimate labels of unknown elements from known elements in Y . For this problem, *similarities* between drugs and those between targets are helpful to predict drug-target interactions, assuming that similar drugs tend to share similar targets and vice versa [14]. A relatively simple idea of using the similarities is the pairwise kernel method (PKM) [13], which generates similarities (or kernels) between drug-target pairs from those between drugs and those between targets. PKM has to generate a huge matrix of all possible combinations of drug-target pairs, causing a serious drawback in computational efficiency. In-

stead, a typical procedure of similarity-based approaches is to use drug and target similarities to generate kernels over drugs and those over targets, respectively, from which drug-target interactions are estimated by kernel methods, such as SVM [2] and kernel regression [28, 26]. These approaches however have a common shortcoming that the kernel from drugs is generated independently from that from targets, meaning that predictions are done twice separately and the final result is obtained by averaging over the two predictions (See Section 3.2.3 for detail). This indicates that drug-target “interactions” are not captured well enough by the current similarity-based approaches. Furthermore so far similarity-based methods have used only one type of similarity for drugs and that for targets. In fact, chemical structure similarity and genomic sequence similarity are the most major metrics for drugs and targets, respectively. However, both drugs and targets have different types of similarity measures, and considering different types of similarities might enhance the predictive performance of drug-target interactions. We thus need to develop a method, which can incorporate multiple types of similarities from drugs and those from targets at once, to predict drug-target interactions.

Our proposed approach is to approximate the input drug-target interaction matrix \mathbf{Y} by two low-rank matrices \mathbf{A} and \mathbf{B} , which share the same feature space, so that \mathbf{A} and \mathbf{B} should be the space to be generated by the weighted similarity matrices of drugs and those of targets, respectively. In other words, \mathbf{Y} is collaboratively approximated by the inner products of the feature vectors of drugs, i.e. \mathbf{A} , and those of targets, i.e. \mathbf{B} , where the weighted drug (target) similarity matrices are also approximated by the inner products of drug (target) feature vectors themselves. We name this formulation *multiple similarities collaborative matrix factorization (MSCMF)*. We further propose an alternating least squares algorithm to estimate \mathbf{A} , \mathbf{B} and weights over drug and target similarities, by which MSCMF can select similarities which are the most consistent with the given drug-target interactions, resulting in the performance improvement for predicting drug-target interactions.

Low-rank approximation with respect to the Frobenius norm can be solved easily by singular value decomposition (SVD), if there are no constraints on factorized matrices. In data mining, the low-rank approximation is a starting point of many different disciplines. First, under no regularization terms, several variants of SVD, such as the generalized low-rank approximation model (GLRAM) [30] and the probabilistic matrix factorization (PMF) model [21], are proposed, being applied to information retrieval, particularly recommender systems [17]. Another series of related work are “non-negative” matrix factorization (NMF) [6, 7], in which factorized matrices must keep the elements nonnegative. On the other hand, MSCMF is a weighted low-rank approximation (WLRA) with regularization terms including a L2 (Tikhonov) regularization term over the low-rank matrices \mathbf{A} and \mathbf{B} , and there are no nonnegative constraints on \mathbf{A} and \mathbf{B} . This means that MSCMF is different from both simple low-rank approximation formulation and NMF.

An EM algorithm for estimating the factorized matrices \mathbf{A} and \mathbf{B} under WLRA is already presented [24]. WLRA with Tikhonov regularization over \mathbf{A} and \mathbf{B} is also equivalent to one formalization for recommender systems, called one-class collaborative filtering, where an alternating least squares algorithm was presented for estimating \mathbf{A} and \mathbf{B} [20, 19, 15]. The key difference of MSCMF from one-class collaborative filtering is that MSCMF further incorporates regularization terms to consider similarity matrices over drugs and those over targets, particularly multiple similarities of drugs and targets. In fact, similarity matrices over users and those over items are also considered in recommender systems [15, 10]. However, first, in [15], similarities are rather preprocessed and in-

corporated into the weight of WLRA, by which the formulation of one-class collaborative filtering is still used. Second, in [10], similarities are processed in a rather similar way to MSCMF in terms that the formulation contains a graph regularization term for similarity matrices, while the main factorization is weighted NMF, by which factorized matrices must be nonnegative. On the other hand, our formulation is WLRA with Tikhonov regularization and regularization terms over drug and target similarities, resulting in a new, original formulation, and we present an alternating least squares algorithm for estimating parameters in this formulation. In fact, alternating least squares as well as stochastic gradient descent are currently the two major and well-accepted approaches for computing matrix factorization.

We empirically evaluated the performance of MSCMF by using both synthetic and real datasets. We first examined MSCMF in terms of the performance improvement in adding similarity matrices and the selectivity of similarity matrices by using various types of synthetic data. We then evaluated the predictive performance of MSCMF under four benchmark datasets, comparing with six state-of-the-art similarity-based drug-target interaction prediction methods, in three settings of predicting 1) new (unknown) interactions (pair prediction), 2) new drugs (drug prediction) and 3) new targets (target prediction). Experimental results showed that MSCMF outperformed all competing methods, in terms of AUPR (Area Under the Precision Recall curve), for all three prediction settings and four datasets. In addition, we checked the performance difference by paired t -test, and the performance advantage of MSCMF was statistically significant in all 56 cases except only three cases at the significance level of 0.01 against p -values of the paired t -test. This result indicates a clear performance advantage over the competing methods.

2. METHOD

2.1 Notation and Problem Setting

Let $D = \{d_1, d_2, \dots, d_{N_d}\}$ be a given set of drugs and $T = \{t_1, t_2, \dots, t_{N_t}\}$ be a given set of targets, where N_d and N_t are the number of drugs and targets, respectively. Let $\{\mathbf{S}_d^1, \mathbf{S}_d^2, \dots, \mathbf{S}_d^{M_d}\}$ 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. We denote the (i, j) -element of \mathbf{S}_d^k by $s_d^k(d_i, d_j)$, which is equal to the similarity score between drugs d_i and d_j in the k -th drug similarity matrix \mathbf{S}_d^k . Similarly let $\{\mathbf{S}_t^1, \mathbf{S}_t^2, \dots, \mathbf{S}_t^{M_t}\}$ be a set of target similarity matrices, where each is a $N_t \times N_t$ matrix, m_t is the number of target similarity matrices, and the (i, j) -element of \mathbf{S}_t^k is denoted by $s_t^k(t_i, t_j)$, being equal to the similarity score between targets t_i and t_j in \mathbf{S}_t^k . Let \mathbf{Y} be a $N_d \times N_t$ binary matrix of true labels of drug-target interactions, where $\mathbf{Y}_{ij} = 1$ if drug d_i and target t_j interact with each other, and $\mathbf{Y}_{ij} = 0$ if they do not interact. The input of our method is the above two sets of similarity matrices and \mathbf{Y} . Let \mathbf{F} be a score function matrix, where the (i, j) -element of \mathbf{F} , i.e. \mathbf{F}_{ij} , shows the score that drug d_i and target t_j interact with each other. The problem is to estimate \mathbf{F} so that \mathbf{F} should be consistent with \mathbf{Y} .

2.2 Multiple Similarities Collaborative Matrix Factorization (MSCMF)

The main idea of MSCMF is to project drugs and targets into two low-rank matrices, corresponding to feature spaces of drugs and targets, respectively.

We thus factorize \mathbf{Y} into two low-rank feature matrices \mathbf{A} and \mathbf{B} so that one drug-target interaction should be approximated by the

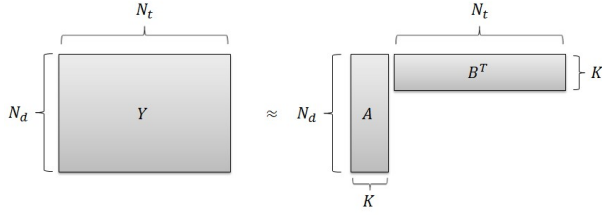


Figure 1: Schematic figure of matrix factorization.

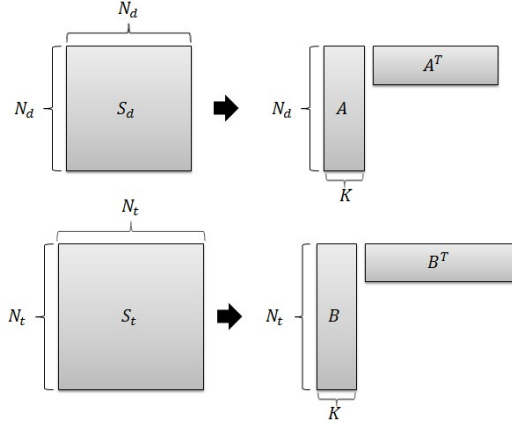


Figure 2: Schematic figure of similarity approximation.

inner product between the two feature vectors of the corresponding drug and target as follows:

$$\mathbf{Y} \approx \mathbf{A}\mathbf{B}^T, \quad (1)$$

where \mathbf{A} and \mathbf{B} are $N_d \times K$ and $N_t \times K$ feature matrices of drugs and targets, respectively, and K is the dimension of the feature spaces.

Estimating \mathbf{A} and \mathbf{B} leads us to reconstruct \mathbf{Y} , by which unknown interactions can have prediction scores. Figure 1 illustrates the process of Eq. (1). To estimate \mathbf{A} and \mathbf{B} in the matrix factorization, a reasonable approach is to minimize the squared error which can be our objective function:

$$\arg \min_{\mathbf{A}, \mathbf{B}} \|\mathbf{Y} - \mathbf{A}\mathbf{B}^T\|_F^2,$$

where $\|\cdot\|_F^2$ is the Frobenius norm.

In order to distinguish known drug-target pairs from unknown pairs, we consider one formulation, weighted low-rank approximation, which introduces a $N_d \times N_t$ weight matrix \mathbf{W} , in which $\mathbf{W}_{ij} = 1$ if \mathbf{Y}_{ij} is a known drug-target pair, i.e. an interacting or non-interacting pair; otherwise $\mathbf{W}_{ij} = 0$. \mathbf{W} is given as an input and used as follows:

$$\arg \min_{\mathbf{A}, \mathbf{B}} \|\mathbf{W} \cdot (\mathbf{Y} - \mathbf{A}\mathbf{B}^T)\|_F^2, \quad (2)$$

where $\mathbf{W} \cdot \mathbf{Z}$ denotes the element-wise product of matrices \mathbf{W} and \mathbf{Z} .

Then to avoid overfitting of \mathbf{A} and \mathbf{B} to training data, we apply L2 (Tikhonov) regularization to Eq. (2) by adding two terms regarding \mathbf{A} and \mathbf{B} .

$$\arg \min_{\mathbf{A}, \mathbf{B}} \|\mathbf{W} \cdot (\mathbf{Y} - \mathbf{A}\mathbf{B}^T)\|_F^2 + \lambda_l (\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2),$$

where λ_l is a regularization coefficient.

Suppose that we have only one similarity matrix for drugs, \mathbf{S}_d and that for targets, \mathbf{S}_t . Our idea is that the generated low-rank ma-

Input:

true drug-target interaction matrix, \mathbf{Y} ;
drug similarity matrices, $\{\mathbf{S}_d^1, \mathbf{S}_d^2, \dots, \mathbf{S}_d^{M_d}\}$;
target similarity matrices, $\{\mathbf{S}_t^1, \mathbf{S}_t^2, \dots, \mathbf{S}_t^{M_t}\}$;
dimensionality of the feature space, K ;
weight matrix, \mathbf{W} ;
weight parameters, $\lambda_l, \lambda_d, \lambda_t$ and λ_ω ;

Output:

predicted interaction matrix, \mathbf{F} ;

- 1: Initialize \mathbf{A} , \mathbf{B} , ω_d and ω_t randomly;
- 2: **repeat**
- 3: Update each row vector of \mathbf{A} using Eq. (4);
- 4: Update each row vector of \mathbf{B} using Eq. (5);
- 5: Update weight vector ω_d using Eq. (6);
- 6: Update weight vector ω_t using Eq. (7);
- 7: Update \mathbf{F} using Eq. (8);
- 8: **until** Convergence;
- 9: Output \mathbf{F} ;

Figure 3: Pseudocode of Parameter Estimation in Multiple Similarities Collaborative Matrix Factorization

trices should be factorized matrices of the drug and target similarities. That is, the similarity between drugs should be approximated by the inner product of the corresponding two drug feature vectors, and this is also the case with the target similarity, as follows:

$$\mathbf{S}_d \approx \mathbf{A}\mathbf{A}^T, \quad \mathbf{S}_t \approx \mathbf{B}\mathbf{B}^T \quad (3)$$

Figure 2 shows a schematic picture of Eqs. (3).

Here we have a set of similarity matrices, instead of only one similarity matrix, for drugs and also for targets. We can then replace one similarity matrix with the one which combines multiple similarity matrices linearly as follows:

$$\mathbf{S}_d = \sum_{k=1}^{M_d} \omega_d^k \mathbf{S}_d^k, \quad \mathbf{S}_t = \sum_{k=1}^{M_t} \omega_t^k \mathbf{S}_t^k$$

$$s.t. \quad |\omega_d| = |\omega_t| = 1,$$

where ω_d^k and ω_t^k are weights over multiple similarity matrices for drugs and targets, respectively, $\omega_d = (\omega_d^1, \dots, \omega_d^{M_d})^T$ and $\omega_t = (\omega_t^1, \dots, \omega_t^{M_t})^T$.

We thus minimize the squared error between \mathbf{S}_d (\mathbf{S}_t) and $\mathbf{A}\mathbf{A}^T$ ($\mathbf{B}\mathbf{B}^T$), resulting in the two regularization terms. Thus the entire objective function (loss function \mathbf{L}) can be written as follows:

$$\arg \min_{\mathbf{A}, \mathbf{B}} \|\mathbf{W} \cdot (\mathbf{Y} - \mathbf{A}\mathbf{B}^T)\|_F^2$$

$$+ \lambda_l (\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2)$$

$$+ \lambda_d \left\| \sum_{k=1}^{M_d} \omega_d^k \mathbf{S}_d^k - \mathbf{A}\mathbf{A}^T \right\|_F^2$$

$$+ \lambda_t \left\| \sum_{k=1}^{M_t} \omega_t^k \mathbf{S}_t^k - \mathbf{B}\mathbf{B}^T \right\|_F^2$$

$$+ \lambda_\omega (\|\omega_t\|_F^2 + \|\omega_d\|_F^2)$$

$$s.t. \quad |\omega_t| = |\omega_d| = 1,$$

where λ_d, λ_t and λ_ω are regularization coefficients. Our regularization terms are all additive, and this manner of using regularization terms is typical (e.g. [27]).

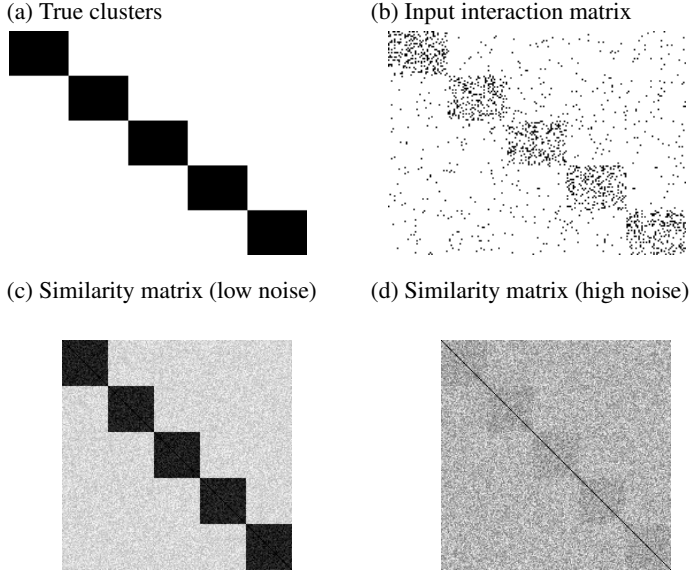


Figure 4: (a) True clusters, (b) the input interaction matrix, (c) a sample similarity matrix with low noise and (d) a sample similarity matrix with high noise, all for five balanced clusters

2.3 Alternating Least Squares Algorithm

We select alternating least squares to estimate \mathbf{A} , \mathbf{B} , ω_d and ω_t , which minimize \mathcal{L} . Here let \mathbf{a}_i and \mathbf{b}_j be the i -th and j -th row vectors of \mathbf{A} and \mathbf{B} , respectively. We take the partial derivative of \mathcal{L} , i.e. $\frac{\partial \mathcal{L}}{\partial \mathbf{a}_i}$ and obtain the updating rule of \mathbf{A} by setting $\frac{\partial \mathcal{L}}{\partial \mathbf{a}_i} = 0$, as follows:

$$\mathbf{a}_i = \left(\sum_{j=1}^{N_t} \mathbf{W}_{ij} \mathbf{Y}_{ij} \mathbf{b}_j + \lambda_d \sum_{p=1}^{N_d} \sum_{k=1}^{M_d} \omega_d^k s_d^k(d_i, d_p) \mathbf{b}_p \right) \left(\sum_{j=1}^{N_t} \mathbf{W}_{ij} \mathbf{b}_j^T \mathbf{b}_j + \lambda_l \mathbf{I}_K + \lambda_d \sum_{p=1}^{N_d} \mathbf{a}_p^T \mathbf{a}_p \right)^{-1}, \quad (4)$$

where \mathbf{I}_K is the $K \times K$ identity matrix.

Similarly, according to $\frac{\partial \mathcal{L}}{\partial \mathbf{b}_j} = 0$, we obtain the updating rule of \mathbf{B} as follows:

$$\mathbf{b}_j = \left(\sum_{i=1}^{N_d} \mathbf{W}_{ij} \mathbf{Y}_{ij} \mathbf{a}_i + \lambda_t \sum_{q=1}^{N_t} \sum_{k=1}^{M_t} \omega_t^k s_t^k(t_j, t_q) \mathbf{a}_q \right) \left(\sum_{i=1}^{N_d} \mathbf{W}_{ij} \mathbf{a}_i^T \mathbf{a}_i + \lambda_l \mathbf{I}_K + \lambda_t \sum_{q=1}^{N_t} \mathbf{b}_q^T \mathbf{b}_q \right)^{-1} \quad (5)$$

Again, updating rules of ω_d and ω_t are given as follows:

$$\omega_d = (\Phi_d + \lambda_\omega \mathbf{I}_{M_d})^{-1} \left(\zeta_d - \frac{(\mathbf{1}_{M_d}^T (\Phi_d + \lambda_\omega \mathbf{I}_{M_d})^{-1} \zeta_d - 1)}{(\mathbf{1}_{M_d}^T (\Phi_d + \lambda_\omega \mathbf{I}_{M_d})^{-1} \mathbf{1}_{M_d})} \mathbf{1}_{M_d} \right) \quad (6)$$

$$\omega_t = (\Phi_t + \lambda_\omega \mathbf{I}_{M_t})^{-1} \left(\zeta_t - \frac{(\mathbf{1}_{M_t}^T (\Phi_t + \lambda_\omega \mathbf{I}_{M_t})^{-1} \zeta_t - 1)}{(\mathbf{1}_{M_t}^T (\Phi_t + \lambda_\omega \mathbf{I}_{M_t})^{-1} \mathbf{1}_{M_t})} \mathbf{1}_{M_t} \right), \quad (7)$$

where $\mathbf{1}_K$ is the vector, in which all K elements are 1, letting $\Phi_d(i, j)$ and $\Phi_t(i, j)$ be the (i, j) -elements of Φ_d and Φ_t , respec-

tively, Φ_d and Φ_t are given as follows:

$$\Phi_d(i, j) = \text{tr}(\mathbf{S}_d^i \mathbf{S}_d^{jT}), \quad \Phi_t(i, j) = \text{tr}(\mathbf{S}_t^i \mathbf{S}_t^{jT})$$

Furthermore letting $\zeta_d(k)$ and $\zeta_t(k)$ be the k -th elements of vectors ζ_d and ζ_t , ζ_d and ζ_t are given as follows:

$$\zeta_d(k) = \text{tr}(\mathbf{B}^T \mathbf{S}_d^k \mathbf{B}), \quad \zeta_t(k) = \text{tr}(\mathbf{A}^T \mathbf{S}_t^k \mathbf{A})$$

Fig. 3 shows a pseudocode of the alternating least squares algorithm for estimating \mathbf{A} , \mathbf{B} , ω_d and ω_t . In this algorithm, we first initialize \mathbf{A} , \mathbf{B} , ω_d and ω_t randomly and repeat updating \mathbf{A} , \mathbf{B} , ω_d and ω_t , according to Eqs. (4), (5), (6) and (7), respectively, until convergence. Finally, the matrix of predicted drug-target interactions \mathbf{F} is given as follows:

$$\mathbf{F} = \mathbf{A} \mathbf{B}^T \quad (8)$$

F_{ij} is the predicted interaction score of drug d_i and target t_j , and the drug and target in a highly ranked pair in terms of the scores are predicted to interact with each other.

3. EXPERIMENTS

3.1 Synthetic Clustering Data

3.1.1 Experimental Settings

Drug-target interactions as well as user-item collaborations would have latent clusters (or factors) from which interactions (or collaborations) can be generated. Thus we first embedded true clusters into the interaction matrix and then simulated a real-world situation by adding a certain amount of noise to the interaction matrix so that the true clusters should not be easily retrieved from the interaction matrix only. For simplicity, we focused on disjoint clusters (by which it is easy to generate synthetic similarity matrices). Figure 4 shows an example of true clusters and the input interaction matrix generated by using the true clusters with noise, for five balanced clusters. We further generated drug similarity matrices and target similarity matrices by first incorporating the cluster information of drug-target interaction matrix (so clusters are on the diagonal of the similarity matrices) and then adding a certain amount of noise, so that multiple similarity matrices have a different amount of noise. Figure 4 shows two samples of similarity matrices (with low and high noise), again for five balanced clusters. We then checked how well the true clusters can be estimated from the interaction matrix and how much this performance can be improved by adding similarity matrices to the interaction matrix. Note that we added similarity matrices keeping the diversity in terms of noise in the similarity matrices. At the same time we examined what type of similarity matrices will be selected, where we expect that less noisy matrices will be selected and more noisy matrices will be discarded.

We here show more detailed experimental settings below: All interaction matrices we used have 200 drugs and 150 targets. We first embedded several true clusters in the interaction matrix by which the (i, j) -element of this matrix is 1 if d_i and t_j are in the same cluster; otherwise this element value is 0. We then completed the input interaction matrix by considering two types of noise: 1) we simply replaced 80% of 1 of the interaction matrix with 0 randomly, and 2) we randomly flipped 2% of all 0 to 1. Then similarity matrices are generated in the following manner: We first generated a true similarity matrix \mathbf{S}_{true} so that the (i, j) -element of \mathbf{S}_{true} is 1 if d_i and d_j (or t_i and t_j) belong to the same true cluster (note that this is easy because true clusters are disjoint); otherwise this value is 0. We then generated a random matrix $\mathbf{S}_{\text{random}}$ where each element of this matrix randomly takes a value between

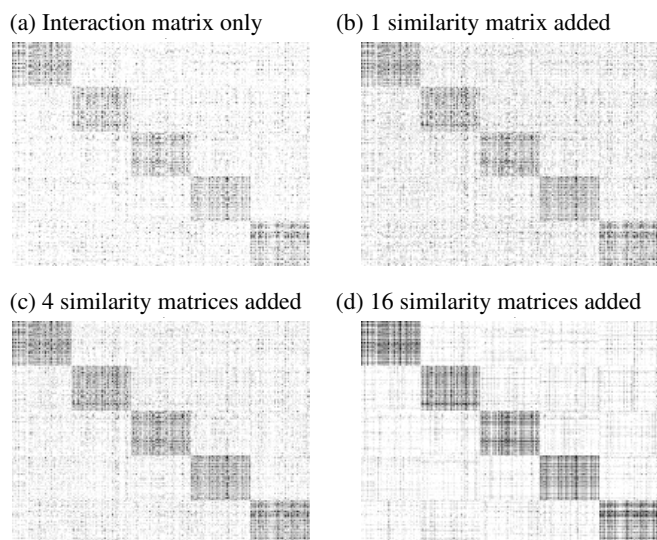


Figure 5: Estimated interactions (a) without any similarity matrices, with (b) one similarity matrix, (c) four similarity matrices and (d) 16 similarity matrices, when five true clusters are balanced.

0 and 1 and the diagonal elements are forced to be 1. We finally generated similarity matrices (for both drugs and targets) by $S_{\text{true}} - (\text{noise_level}) \times S_{\text{random}}$, where *noise_level* is a value being changed from 0.15 to 0.9 by the interval of 0.05, resulting in 16 similarity matrices, all having different noise levels totally.

3.1.2 Performance Results

We started our experiments for the five balanced clusters shown in Figure 4, in which each cluster has 40 drugs and 30 targets. Figure 5 shows the estimated drug-target interactions (clusters) from the input interaction matrix and similarity matrices. From this figure, we can see that clusters (interactions) was not clearly predicted if the input was the interaction matrix only, while these clusters were made clearer by adding a larger number of similarity matrices. This indicates that clusters (interactions) were more clearly predicted by adding a larger number of similarity matrices, and MSCMF works well for this addition of similarity matrices to the input interaction matrix. We then made the size of clusters unbalanced, keeping the number of clusters at the same, where the sizes of five true clusters were (70, 10), (55, 20), (40, 30), (25, 40), (10, 50) for drugs and targets, respectively. Figure 6 shows the estimated drug-target interactions (clusters) from the input interaction matrix and similarity matrices. This figure also clearly shows the advantage of adding similarity matrices and the efficiency of MSCMF for selecting the similarity matrices useful to improve the performance.

Instead of checking the obtained interaction matrices directly, we then checked normalized mutual information (NMI) between the true clusters and the obtained interactions for both the balanced and unbalanced cases (Note that NMI is a standard measure to check the performance of clustering methods). Figure 7 shows the NMI when we changed the number of added similarity matrices. This figure reveals that NMI was clearly bigger by adding similarity matrices for the both cases, while NMI was saturated when around six similarity matrices were added. This would be because low noise similarity matrix was already included when we added around six or so similarity matrices, keeping the diversity of noise level in the

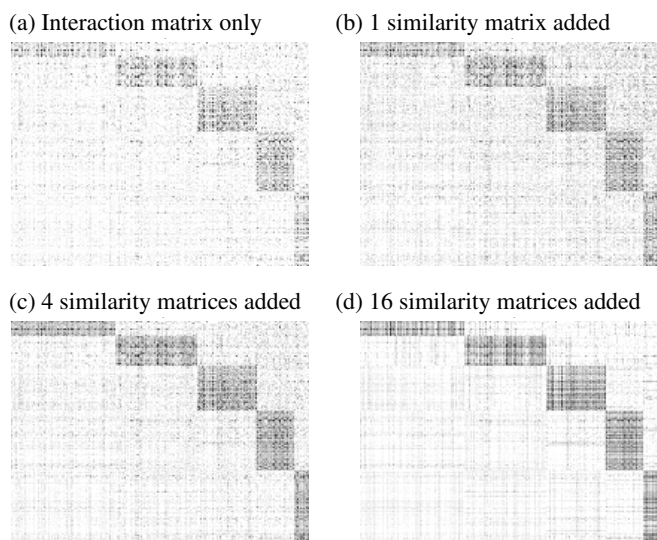


Figure 6: Estimated interactions (a) without any similarity matrices, with (b) one similarity matrix, (c) four similarity matrices and (d) 16 similarity matrices, when five true clusters are unbalanced.

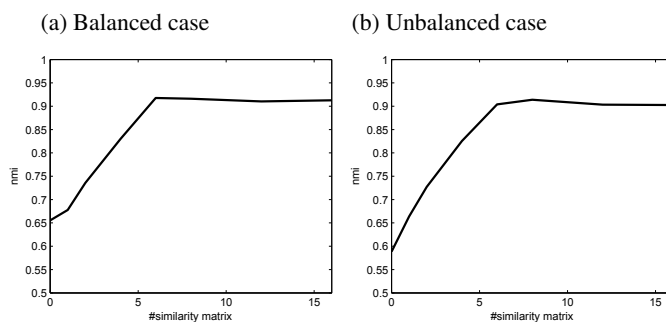


Figure 7: NMI for the (a) balanced 5 clusters and (b) unbalanced 5 clusters.

matrices. We thus checked the change of weights over similarity matrices during the iteration of our alternating least squares algorithm. Figure 8 shows the resultant weight change during the iteration, when we used five similarity matrices with different noise levels: 0.15, 0.3, 0.5, 0.7 and 0.9. From this figure, we can see that the initial weights were 0.2 (we used a uniform distribution for this case), while in both cases, only two similarity matrices had high weights (around 0.5) finally. These selected similarity matrices were with the noise level of 0.15 and 0.3, indicating that low-noise similarity matrices were selected by MSCMF automatically and high-noise similarity matrices were not, implying the efficiency of MSCMF for selecting better similarity matrices.

Finally we changed the number of clusters from 3 to 7, keeping the cluster unbalancedness. In fact, the sizes of clusters we tested were, for drugs and targets, (110, 20), (60, 50) and (30, 80) for 3 clusters, (80, 10), (60, 20), (40, 40) and (20, 80) for 4 clusters, (65, 5), (50, 10), (40, 20), (25, 30), (15, 40) and (5, 45) for 6 clusters, and (70, 5), (50, 10), (30, 15), (20, 20), (15, 25), (10, 35) and (5, 40) for 7 clusters. Figures 9 and 10 show the NMI when we changed the number of added similarity matrices and the weights over five similarity matrices (with the same noise levels as the case of five clusters) during iterations of our algorithm, respectively. The results of these figures were totally consistent with those obtained

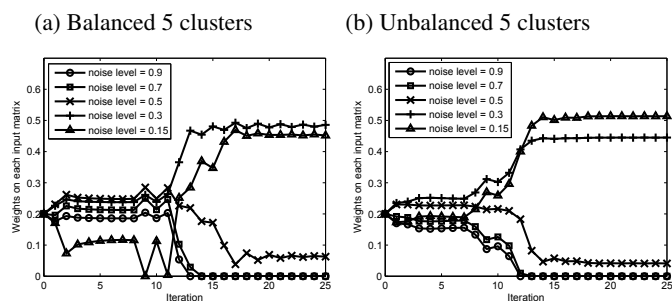


Figure 8: Change of weights over similarity matrices during iteration of our algorithm for the (a) balanced 5 clusters and (b) unbalanced 5 clusters.

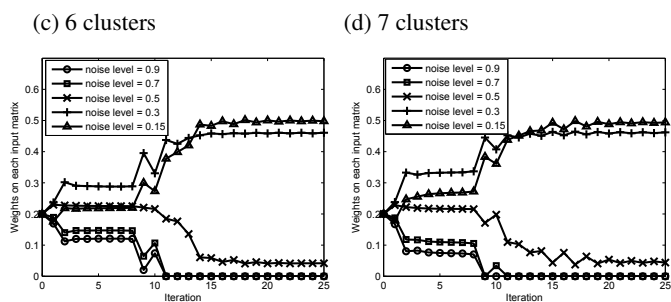
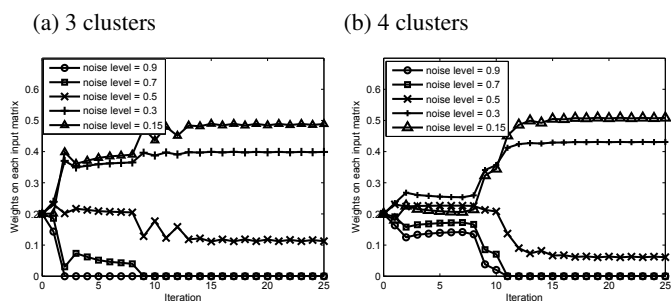


Figure 10: Variation of weights during the algorithm iteration for (a) 3 clusters, (b) 4 clusters, (c) 6 clusters and (d) 7 clusters, all being unbalanced clusters.

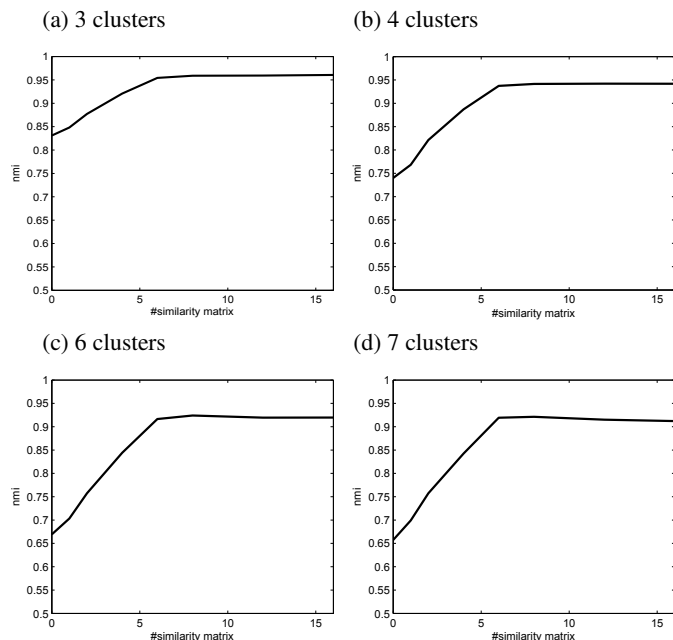


Figure 9: Variation of NMI by adding similarity matrices for (a) 3 clusters, (b) 4 clusters, (c) 6 clusters and (d) 7 clusters, all being unbalanced clusters.

when we used five clusters, and for example, only two similarity matrices were finally selected regardless of the number of clusters in Figure 10. One additional item of note is that when we used random initial weights, the final weights were almost the same as those of uniform distributions which are shown in Figures 8 and 10. This indicates that our optimization algorithm is stable against the change of initial values. Overall these results indicate that the efficiency of MSCMF is robust against the cluster size as well as cluster unbalancedness.

3.2 Real Drug-Target Interaction Data

3.2.1 Drug-Target Interaction Data

We used four real benchmark datasets, called *Nuclear receptor*, *GPCR*, *Ion channel* and *Enzyme*, which were originally provided by [29]¹. These datasets were collected from four general databases and frequently used in predicting drug-target interactions [2, 28, 26, 9]. Table 1 shows the statistics of these four datasets.

¹All datasets are downloadable from <http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/>

	#interactions	#drugs	#targets
Nuclear receptor	90	54	26
GPCR	635	223	95
Ion channel	1476	210	204
Enzyme	2926	445	664

3.2.2 Similarity Matrices over Drugs and Targets

We used the following two and four types of similarities for drugs and targets, respectively, by considering the effectiveness of predicting drug-target interactions [12].

Drugs:

Chemical structure similarity (CS) is computed by the number of shared substructures in chemical structures between two drugs.

ATC similarity (ATC) is computed by using a hierarchical drug classification system, called ATC (Anatomical Therapeutic Chemical) [23]. We used a general method in [16] to compute the similarity between two nodes (drugs) in this classification tree.

Targets:

Genomic sequence similarity (GS) is computed by a normalized Smith-Waterman score [29] between two target sequences.

Gene Ontology (GO) similarity is the overlap of GO annotations [1] of two targets for which we simply used GOSemSim [31]. We considered two options of GO: molecular functions (MF) and biological processes (BP).

Protein-protein interaction (PPI) network similarity (PPI) is computed from the shortest distance between two targets in a human protein-protein interaction (PPI) network [25].

We note that CS and GS are the most standard similarities (which have been used for predicting drug-target interactions) and so we just downloaded the data of CS and GS along with the drug-target interaction data, while we computed the other similarities by using

the procedure described above. We further note that these similarities are diverse. For example, GS is derived from static (sequence) information, while PPI is more dynamic and noisy, and GO is rather in between these two types of similarities.

3.2.3 Competing Methods

We here briefly review six state-of-the-art similarity-based methods for predicting drug-target interactions, all being compared with our method in this experiment.

Pairwise Kernel Method (PKM) [13] generates similarities (kernels) over drug-target pairs, which can be the input instances of SVM. The similarity between two drug-target pairs, say (d, t) and (d', t') , is computed from given drug and target similarities, $s_d(d, d')$ and $s_t(t, t')$, as follows:

$$\mathbf{K}((d, t), (d', t')) = s_d(d, d')s_t(t, t') \quad (9)$$

Bipartite local model (BLM) [2] also uses SVM. To predict a score of drug-target pair (d, t) , first drug d is fixed, and a SVM is trained by using known interactions of drug d as instances where the kernel over instances is the target similarity matrix. Second target t is fixed, and a SVM is trained by using interactions of target t . In [2], the prediction is obtained by the maximum of the results of the two trained SVM, while in our experiments we used the average over them, since this is standard in other methods, such as [28, 26].

Net Laplacian regularized least squares (NetLapRLS) [28] minimizes the least squared error between \mathbf{Y} and \mathbf{F} . Note that \mathbf{F} is obtained twice, i.e. drugs and targets, separately. For drugs, we first compute a $N_d \times N_d$ matrix \mathbf{O}_d , showing how many targets are overlapped between two drugs. We then have matrix \mathbf{V}_d by a linear combination of \mathbf{O}_d with \mathbf{S}_d : $\mathbf{V}_d = t\mathbf{S}_d + (1-t)\mathbf{O}_d$. \mathbf{F}_d is then given by $\mathbf{F}_d = \mathbf{V}_d\boldsymbol{\alpha}_d$, where $\boldsymbol{\alpha}_d$ is the parameter matrix to be estimated. Note that NetLapRLS uses only one similarity matrix over drugs. The entire formulation is given as follows:

$$\min_{\boldsymbol{\alpha}_d} \{ \|\mathbf{Y} - \mathbf{V}_d\boldsymbol{\alpha}_d\|^2 + \lambda_n \text{Tr}(\boldsymbol{\alpha}_d^T \mathbf{V}_d \mathbf{L}_d \mathbf{V}_d \boldsymbol{\alpha}_d) \},$$

where λ_n is a weight and \mathbf{L}_d is a normalized graph Laplacian of \mathbf{S}_d . This is a convex optimization problem, which gives the following direct solution:

$$\hat{\mathbf{F}}_d = \mathbf{V}_d(\mathbf{V}_d + \lambda_n \mathbf{L}_d \mathbf{V}_d)^{-1} \mathbf{Y}$$

The same operation is done for the target side to have $\hat{\mathbf{F}}_t$, and the final result is obtained by the average: $\hat{\mathbf{F}} = \frac{\hat{\mathbf{F}}_d + \hat{\mathbf{F}}_t}{2}$.

Regularized Least Squares with Gaussian Interaction Profiles (RLS-GIP) [26] is similar to but simpler than NetLapRLS in terms of regularization. For drugs, we can first compute a $N_d \times N_d$ matrix \mathbf{Q}_d , in which the (i, j) -element is $\exp(-\gamma \|\mathbf{Y}_i - \mathbf{Y}_j\|^2)$, where γ is a parameter and \mathbf{Y}_i is the i -th row vector of \mathbf{Y} . We then have matrix \mathbf{V}_d by a linear combination with \mathbf{S}_d : $\mathbf{V}_d = t\mathbf{S}_d + (1-t)\mathbf{Q}_d$. A least square classifier with simpler regularization leads to a direct and simpler solution for estimating \mathbf{F} :

$$\hat{\mathbf{F}}_d = \mathbf{V}_d(\mathbf{V}_d + \lambda_g \mathbf{I}_{N_d})^{-1} \mathbf{Y}$$

Again the final result is obtained by averaging over $\hat{\mathbf{F}}_d$ and $\hat{\mathbf{F}}_t$.

RLS-GIP with Kronecker product kernel (RLS-GIP-K) [26] uses the regularized least square but incorporates the idea of PKM, i.e. a kernel over drug-target pairs, which is, in [26], a (Kronecker) product kernel from \mathbf{V}_d and \mathbf{V}_t , as follows:

$$\mathbf{K}((d, t), (d', t')) = \mathbf{V}_d(d, d')\mathbf{V}_t(t, t')$$

We then use the regularized least square, being the same as RLS-GIP, to estimate the final \mathbf{F} as follows:

$$\hat{\mathcal{F}} = \mathbf{K}(\mathbf{K} + \lambda_h \mathbf{I}_{N_d \times N_t})^{-1} \mathcal{Y},$$

where \mathcal{Y} and \mathcal{F} are vectors (with $N_d N_t$ elements), corresponding to \mathbf{Y} and \mathbf{F} , respectively.

Kernelized Bayesian matrix factorization (KBMF2K) [9] has a similar idea to our method in the sense that drug and target similarities are both projected onto low-dimensional spaces with the same dimension so that they can reconstruct true drug-target interactions. More concretely, \mathbf{S}_d and \mathbf{S}_t are reduced into low-dimensional matrices \mathbf{G}_d and \mathbf{G}_t , respectively, so that $\mathbf{Y} \approx \mathbf{G}_d \mathbf{G}_t^T$. The entire scheme is a graphical model that makes \mathbf{G}_d and \mathbf{G}_t latent variables, which are estimated by a variational estimation algorithm. Graphical models are likely to have unavoidable constraints. So drug and target similarity matrices cannot be weighted in KBMF2K.

3.2.4 Experimental Settings

We compared the performance of MSCMF with six latest competing methods, BLM, PKM, NetLapRLS, RLS-GIP, RLS-GIP-K and KBMF2K. In addition, as a reference, we checked the performance of two downgraded variations of MSCMF: 1) OCCF: MSCMF without any similarities, meaning weighted low-rank approximation with Tikhonov regularization only, being equivalent to the formulation of one-class collaborative filtering [20] and 2) CMF: MSCMF with only one type of similarity, i.e. chemical structure similarity for drugs and genomic sequence similarity for targets. The evaluation was done by 5×10 -fold cross-validation (CV). That is, we repeated the following one CV five times: the entire dataset was randomly divided into ten folds, from which we repeated training by nine folds and testing for the rest, ten times, changing the test fold. The results were averaged over the total 50 ($= 5 \times 10$) runs. We considered three different types of prediction by randomly dividing 1) all drug-target interactions (*pair prediction*), 2) all drugs (*drug prediction*) and 3) all targets (*target prediction*). Note that RLS-GIP, RLS-GIP-K and OCCF cannot be applied to drug and target prediction, by which we compared four methods for drug and target prediction. We evaluated the performance by AUPR (Area Under the Precision-Recall curve) instead of a more standard measure, AUC (Area Under the ROC Curve), because AUPR punishes highly ranked false positives much more than AUC [5], this point being important practically 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.

MSCMF has five parameters, K , λ_l , λ_d , λ_t and λ_ω . For each pair of training and test datasets in cross-validation, we selected parameter values, by using an usual manner of (10-fold) cross-validation: only a part (nine folds) of the training dataset for estimating parameters of MSCMF and the rest (one fold) for evaluation. In this parameter value selection, we considered all combinations of the following values: $\{50, 100\}$ for K , $\{2^{-2}, \dots, 2^1\}$ for λ_l , $\{2^{-3}, 2^{-2}, \dots, 2^5\}$ for λ_d and λ_t , $\{2^1, 2^2, \dots, 2^{10}\}$ for λ_ω .

We implemented PKM and BLM by using LIBSVM [3], where the regularization parameter of SVM was set at 1, according to [13] and [2]. To train PKM, the number of negative examples (interactions) was set at the same number of positive interactions, due to limitations of the main memory size. We implemented NetLapRLS exactly according to [28] and set its parameter values as specified in [28]. RLS-GIP and RLS-GIP-K were run by using the software originally developed in [26], where the parameter setting we used was also the same as those in [26]. Similarly KBMF2K was run by the software in [9] with the same parameter setting as that in [9].

3.2.5 Performance Results

Table 2 shows resultant AUPR (with p -values of paired t -test between each method and the best method in the same column) which

Table 2: AUPR values obtained by 5×10-fold cross validation. The highest AUPR value for each column is highlighted in boldface. The *p*-values between the best method and each method are in parentheses.

(a) Pair prediction

Methods	Nuclear receptor		GPCR	Ion channel		Enzyme
BLM	0.204 (2.84×10^{-25})	0.464 (1.23×10^{-36})	0.592 (6.68×10^{-42})	0.496 (2.07×10^{-54})		
PKM	0.514 (4.27×10^{-12})	0.474 (2.19×10^{-37})	0.663 (1.20×10^{-42})	0.627 (2.86×10^{-45})		
NetLapRLS	0.563 (1.39×10^{-7})	0.708 (1.17×10^{-17})	0.900 (2.25×10^{-20})	0.874 (1.46×10^{-18})		
RLS-GIP	0.599 (1.10×10^{-5})	0.733 (3.07×10^{-11})	0.904 (1.21×10^{-18})	0.880 (7.43×10^{-15})		
RLS-GIP-K	0.604 (1.93×10^{-5})	0.727 (7.91×10^{-13})	0.898 (2.00×10^{-20})	0.884 (1.20×10^{-10})		
KBMF2K	0.508 (5.16×10^{-11})	0.686 (5.23×10^{-18})	0.876 (3.00×10^{-22})	0.796 (1.34×10^{-41})		
OCCF	0.387 (4.41×10^{-16})	0.657 (2.47×10^{-22})	0.883 (5.03×10^{-25})	0.775 (8.96×10^{-42})		
CMF	0.643 (3.08×10^{-2})	0.746 (1.31×10^{-7})	0.937 (4.76×10^{-1})	0.887 (2.51×10^{-13})		
MSCMF	0.673	0.773	0.937	0.894		

(b) Drug prediction

Methods	Nuclear receptor		GPCR	Ion channel		Enzyme
BLM	0.194 (1.07×10^{-19})	0.210 (3.26×10^{-32})	0.167 (1.82×10^{-25})	0.092 (1.12×10^{-27})		
PKM	0.484 (2.47×10^{-5})	0.323 (1.14×10^{-19})	0.328 (3.64×10^{-8})	0.254 (3.18×10^{-17})		
NetLapRLS	0.481 (1.78×10^{-5})	0.397 (6.64×10^{-16})	0.343 (2.50×10^{-13})	0.298 (7.51×10^{-16})		
KBMF2K	0.450 (1.39×10^{-5})	0.357 (2.39×10^{-17})	0.296 (6.19×10^{-9})	0.253 (4.52×10^{-17})		
CMF	0.497 (2.87×10^{-4})	0.398 (6.14×10^{-15})	0.342 (8.68×10^{-10})	0.326 (8.00×10^{-12})		
MSCMF	0.572	0.474	0.419	0.432		

(c) Target prediction

Methods	Nuclear receptor		GPCR	Ion channel		Enzyme
BLM	0.325 (5.10×10^{-6})	0.367 (8.85×10^{-18})	0.641 (2.51×10^{-25})	0.611 (1.03×10^{-23})		
PKM	0.413 (4.08×10^{-2})	0.400 (7.02×10^{-15})	0.659 (1.30×10^{-21})	0.587 (9.33×10^{-29})		
NetLapRLS	0.433 (7.95×10^{-2})	0.503 (3.92×10^{-9})	0.762 (2.67×10^{-9})	0.787 (4.90×10^{-5})		
KBMF2K	0.404 (4.62×10^{-2})	0.412 (2.00×10^{-11})	0.725 (2.16×10^{-12})	0.607 (2.90×10^{-27})		
CMF	0.435	0.556	0.798	0.796		
MSCMF	0.431 (4.04×10^{-1})	0.505 (8.04×10^{-8})	0.785 (5.20×10^{-3})	0.795 (4.21×10^{-1})		

Table 3: A typical case of resultant similarity weights under pair prediction.

(a) Similarities over drugs

Similarities	Nuclear receptor	GPCR	Ion channel	Enzyme
CS	0.6042	0.68	0.5804	0.5626
ATC	0.3958	0.32	0.4196	0.4374

(b) Similarities over targets

Similarities	Nuclear receptor	GPCR	Ion channel	Enzyme
GS	0	0.5297	0	0
GO (MF)	0.4409	0.1286	0.5262	0.3827
GO (BP)	0.5591	0	0.4738	0.3652
PPI	0	0.3417	0	0.2521

were all obtained by 5 × 10-fold CV. For pair prediction, MSCMF outperformed all six competing methods, being statistically significant in all cases. This directly indicates the clear performance advantage of our approach over existing state-of-the-art methods for predicting drug-target interactions. In addition, MSCMF completely outperformed OCCF for all four datasets, being statistically significant, and achieved higher values than CMF, being statistically significant for two datasets (GPCR and Enzyme). This result implies that adding similarity matrices is generally useful but sometimes insignificant. For drug prediction, again MSCMF outperformed all four competing methods, all being statistically sig-

nificant in all datasets. In addition, this case, MSCMF clearly outperformed CMF for all four datasets. This result shows that the scheme of MSCMF worked for prediction, and using more than one similarity matrices is useful for drug prediction. For target prediction, MSCMF outperformed all four competing methods, being statistically significant except three cases in the Nuclear Receptor dataset at the significant level of 0.01. On the other hand, CMF outperformed MSCMF in all four datasets, two cases being statistically significant. Thus we can say that the framework of CMF or MSCMF works for target prediction, while incorporating more similarity matrices might not be necessarily useful for improving the performance in this case.

Finally we checked resultant weights over similarity matrices of drugs and targets. Table 3 shows a typical set of resultant weights under pair prediction. In this table, for drugs, chemical structure similarity (CS) always had larger weights than ATC code similarity (ATC) for all four datasets, being consistent with the fact that chemical structure similarity is the most well-used similarity. On the other hand, for targets, interestingly, the most popular genomic sequence similarity (GS) had weights of zero for three datasets (Nuclear receptor, Ion channel and Enzyme), implying that this similarity might not work well for prediction. Instead, two GO-based similarities both achieved high values for the three datasets, implying that GO-based similarities were very useful and should be used more than genomic sequence similarity. However, genomic sequence similarity achieved the highest weight value for GPCR, which is in reality the most major target in drug discovery (more

than 50% of all targets are GPCR), which might be the reason why genomic sequence similarity has been used so far.

4. CONCLUDING REMARKS

We have presented a new formulation based on weighted low-rank approximation for predicting drug-target interactions. The key feature of our approach is to use multiple types of similarity matrices for both drugs and targets. In particular we stress that multiple similarity matrices are explicitly incorporated into our optimization formulation as regularization terms, by which our method can select similarity matrices, which are the most useful for predicting drug-target interactions, by which the predictive performance can be improved. We have demonstrated the advantage of our proposed method by using both synthetic and real datasets. Synthetic data experiments have revealed the favorable selectivity on similarity matrices, and real data experiments have shown the high predictive performance over the six current state-of-the-art methods for predicting drug-target interactions.

5. ACKNOWLEDGMENTS

This work has been partially supported by MEXT KAKENHI (24300054), ICR-KU International Short-term Exchange Program for Young Researchers, SRF for ROCS, SEM and National Natural Science Foundation of China (61170097).

6. REFERENCES

- [1] M. Ashburner et al. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.*, 25(1):25–29, 2000.
- [2] K. Bleakley and Y. Yamanishi. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics*, 25(18):2397–2403, 2009.
- [3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
- [4] A. C. Cheng, R. G. Coleman, K. T. Smyth, Q. Cao, P. Souillard, D. R. Caffrey, A. C. Salzberg, and E. S. Huang. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.*, 25(1):71–75, 2007.
- [5] J. Davis and M. Goadrich. The relationship between Precision-Recall and ROC curves. In *ICML*, pages 233–240, 2006.
- [6] I. S. Dhillon and S. Sra. Generalized nonnegative matrix approximations with bregman divergences. In *NIPS*, pages 283–290, 2005.
- [7] C. H. Q. Ding, T. Li, and M. I. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1):45–55, Jan. 2010.
- [8] C. M. Dobson. Chemical space and biology. *Nature*, 432:824–828, 2004.
- [9] M. Gönen. Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization. *Bioinformatics*, 28(18):2304–2310, 2012.
- [10] Q. Gu, J. Zhou, and C. H. Q. Ding. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs. In *SDM*, pages 199–210. SIAM, 2010.
- [11] A. L. Hopkins. Drug discovery: predicting promiscuity. *Nature*, 462:167–168, 2009.
- [12] M. Iskar, G. Zeller, X. M. Zhao, V. van Noort, and P. Bork. Drug discovery in the age of systems biology: the rise of computational approaches for data integration. *Curr. Opin. Biotechnol.*, 23(4):609–616, Aug 2012.
- [13] L. Jacob and J. P. Vert. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics*, 24(19):2149–2156, 2008.
- [14] T. Klabunde. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br. J. Pharmacol.*, 152(1):5–7, 2007.
- [15] Y. Li, J. Hu, C. Zhai, and Y. Chen. Improving one-class collaborative filtering by incorporating rich user information. In *CIKM*, pages 959–968, New York, NY, USA, 2010. ACM.
- [16] D. Lin. An information-theoretic definition of similarity. In *ICML*, pages 296–304, 1998.
- [17] H. Ma, H. Yang, M. R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In *CIKM*, pages 931–940, 2008.
- [18] N. Nagamine and Y. Sakakibara. Statistical prediction of protein chemical interactions based on chemical structure and mass spectrometry data. *Bioinformatics*, 23(15):2004–2012, 2007.
- [19] R. Pan and M. Scholz. Mind the gaps: weighting the unknown in large-scale one-class collaborative filtering. In *KDD*, pages 667–676, New York, NY, USA, 2009. ACM.
- [20] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. Lukose, M. Scholz, and Q. Yang. One-class collaborative filtering. In *ICDM*, pages 502–511, 2008.
- [21] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *NIPS*, 2007.
- [22] E. W. Sayers et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, 40(Database issue):D13–D25, 2012.
- [23] A. Skrbo, B. Begovic, and S. Skrbo. Classification of drugs using the ATC system (anatomic, therapeutic, chemical classification) and the latest changes. *Med. Arh.*, 58(1 Suppl 2):138–141, 2004.
- [24] N. Srebro and T. Jaakkola. Weighted low-rank approximations. In *ICML*, pages 720–727, 2003.
- [25] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, 34(Database issue):D535–D539, 2006.
- [26] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics*, 27(21):3036–3043, 2011.
- [27] F. Wang, X. Wang, and T. Li. Semi-supervised multi-task learning with task regularizations. In *ICDM*, pages 562–568, 2009.
- [28] Z. Xia, L. Y. Wu, X. Zhou, and S. T. Wong. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst. Biol.*, 4(Suppl 2):S6, 2010.
- [29] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 24(13):i232–i240, 2008.
- [30] J. Ye. Generalized low rank approximations of matrices. *Machine Learning*, 61(1-3):167–191, 2005.
- [31] G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, and S. Wang. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, 26(7):976–978, 2010.