IET Systems Biology

Research Article

IET Journals
The Institution of Engineering and Technology

Drug repositioning via matrix completion with multi-view side information

ISSN 1751-8849
Received on 26th December 2018
Revised 14th June 2019
Accepted on 2nd July 2019
E-First on 30th August 2019
doi: 10.1049/iet-syb.2018.5129
www.ietdl.org

Yunda Hao[1], *Menglan Cai*[1], *Limin Li*[1] ✉

[1] *School of Mathematics and Statistics, Xi'an Jiaotong University, Xianning West 28, Xi'an, People's Republic of China*

✉ *E-mail: liminli@mail.xjtu.edu.cn*

Abstract: In the process of drug discovery and disease treatment, drug repositioning is broadly studied to identify biological targets for existing drugs. Many methods have been proposed for drug–target interaction prediction by taking into account different kinds of data sources. However, most of the existing methods only use one side information for drugs or targets to predict new targets for drugs. Some recent works have improved the prediction accuracy by jointly considering multiple representations of drugs and targets. In this work, the authors propose a drug–target prediction approach by matrix completion with multi-view side information (MCM) of drugs and proteins from both structural view and chemical view. Different from existing studies for drug–target prediction, they predict drug–target interaction by directly completing the interaction matrix between them. The experimental results show that the MCM method could obtain significantly higher accuracies than the comparison methods. They finally report new drug–target interactions for 26 FDA-approved drugs, and biologically discuss these targets using existing references.

Nomenclature

$\mathbf{P}$	$m_\mathrm{d} \times m_\mathrm{t}$ known drug–target interaction matrix
$\mathbf{A}^\mathrm{s}$	$k_\mathrm{d} \times k_\mathrm{t}$ complete low-rank matrix in the structural view
$\mathbf{A}^\mathrm{c}$	$k_\mathrm{d} \times k_\mathrm{t}$ complete low-rank matrix in the chemical view
$\mathbf{W}_\mathrm{d}^\mathrm{s}$	$m_\mathrm{d} \times m_\mathrm{d}$ drug–drug similarity matrix in the structural view
$\mathbf{W}_\mathrm{t}^\mathrm{s}$	$m_\mathrm{t} \times m_\mathrm{t}$ target–target similarity matrix in the structural view
$\mathbf{W}_\mathrm{d}^\mathrm{t}$	$m_\mathrm{d} \times m_\mathrm{d}$ drug–drug similarity matrix in the chemical view
$\mathbf{W}_\mathrm{t}^\mathrm{t}$	$m_\mathrm{t} \times m_\mathrm{t}$ target–target similarity matrix in the chemical view
$\mathbf{D}_\mathrm{s}$	$m_\mathrm{d} \times k_\mathrm{d}$ drugs feature matrix in the structural view
$\mathbf{G}_\mathrm{s}$	$m_\mathrm{t} \times k_\mathrm{t}$ protein targets feature matrix in the structural view
$\mathbf{D}_\mathrm{c}$	$m_\mathrm{d} \times k_\mathrm{d}$ drugs feature matrix in the chemical view
$\mathbf{G}_\mathrm{c}$	$m_\mathrm{t} \times k_\mathrm{t}$ protein targets feature matrix in the chemical view
$\mathbf{Q}$	$m_\mathrm{d} \times m_\mathrm{t}$ the common complete drug–target interaction matrix
$\mathbf{Z}$	$k_\mathrm{d} \times k_\mathrm{t}$ any given matrix
$\langle \cdot , \cdot \rangle$	inner product for matrices
∇	gradient operator
λ_1, λ_2	trade-off parameters

1 Introduction

Drugs take effects by acting on their corresponding targets, such as proteins. The identification of drug–target interactions becomes an important step in discovering new drugs. It helps the understanding of drug mechanism in treating diseases and provides inspirations for inventing new drugs. Although researchers can find some meaningful drug–target interactions through biological experiments, the high cost of carrying out those experiments forces people to develop computational methods to identify potential new targets for drugs.

Many methods have been proposed for identifying drug–target interactions. Among these researches, a diversity of data, including protein–protein interactions, gene expression data, chemical structure of drugs, metabolic network, protein sequence, drug response and drug side effects, are applied individually or jointly. For example, Liu *et al.* [1] apply neighbourhood regularised logistic matrix factorisation based on the protein sequences and drug structures to model how likely a drug interacts with a target. Yamanishi *et al.* [2] and Bleakley and Yamanishi [3] propose bipartite graph-based methods with the same dataset [1], by first defining a bipartite graph between drugs and proteins and then finding the latent common space for them. The drugs and targets closely situated are predicted as the interacted pairs. Mizutani *et al.* [4] make use of protein functions and drugs' side effects to identify novel targets for the already known anti-cancer drugs by sparse canonical correlation analysis. Chen and Zhang [5] propose a partial least square method with sparse network regularisation by integrating drug response data and gene expression to identify joint modular patterns. Li *et al.* [6] use the human metabolic network for the prediction of drug–target interactions by exploring drug-reaction interactions. Dorothea *et al.* [7] propose a network-based approach by combining a molecular interaction network and disease gene expression signatures. Ding *et al.* [8] and Zheng *et al.* [9] propose similarity-based methods to discover new drug targets. Li *et al.* [10] propose an efficient and effective multi-task machine learning approach for detecting potential drug targets, using both expression data and compound structure information.

Since drugs or proteins can be represented in different ways, the identification of drug targets by jointly considering their multi-view representations is a promising research field in the future due to the sufficient data varieties. For example, a drug can be described by its chemical response in different cells, or by its chemical structure. As for proteins, both their amino-acid sequences and their gene expression values in different cells can be regarded as their representations. We could consider the structure information of drugs and proteins as the structural view, while their chemical behaviour described by gene expression and drug response is regarded as the chemical view. In the field of machine learning, there are many multi-view methods which aim to do supervised or unsupervised learning by combining different representations of samples, such as [11–21] and so on. Unfortunately, the multi-view approaches could not be directly applied for multi-view drug–target prediction, where drugs and targets could construct a bipartite graph. Li [22] proposes a new graph-regularised-based single-view approach of single-view penalised graph (SPGraph) to identify drug targets by making use of the structural information or the chemical information individually, and extends it to a co-regularised multi-view method by fusing structural and chemical views of drugs and targets together. Li and Cai [23] develop a new multi-view low-rank embedding (MLRE) method by using a strategy of low-rank embedding. The results in [22, 23] suggest that the multi-view approaches perform significantly better than single-view approaches. Both [22, 23] take similar strategies by first obtaining new features for drugs and targets in a shared subspace and then doing clustering on all these representations by k-means. The proteins and drugs closely situated are predicted to have interactions. However, one might obtain different prediction results with different clustering methods or different initialisation at the clustering stage. Besides, the accuracy of prediction might be sensitive to the new representations of drugs and proteins in the common subspace. There is a high chance that drugs and proteins are wrongly clustered due to inappropriate representations. It is challenging to develop a new multi-view approach to identify drug targets.

In this work, our goal is to identify drug targets by directly completing the interaction matrix of drugs and proteins by using multi-view similarities among drugs or targets which we consider as their multi-view side information. Matrix completion is widely used in biology prediction problems, such as lncRNA-disease association [24], averse drug interaction [25], gene-disease associations [26] and miRNA-disease association [27]. For example, Chen *et al.* [27] propose an inductive matrix completion method with single-view side information. Although this work aims to predict miRNA-disease associations, it can also be applied for drug–target prediction. Unfortunately, this approach could only use one type of side information. Zhao *et al.* [21] propose to cluster n samples based on samples' multiple side information, by completing a 0-1 square clustering matrix whose entry represents whether the two samples are in the same cluster. However, the model proposed in multi-view matrix completion (MVMC) [21] can only be used in the case where the rows and the columns of completed matrix represent the same samples. In drug–target prediction problem, the rows and columns of the interaction matrix represent drugs and targets, respectively. Thus, MVMC could not be directly applied to predict drug–target interactions.

The contributions of this work are twofold. On one hand, we propose a novel inductive matrix completion with multi-view side information (MCM) for drug target prediction. We complete the association matrix directly with drugs similarity and targets similarity rather than clustering on the new representation of drugs and targets to predict the latent drug targets. The common completed matrix and two single-view completed matrices are alternately optimised by our MCM algorithm. The method can be considered as a general MCM and be applied to other scenarios. On the other hand, we compare our method MCM with other comparison partners in two experimental settings on real datasets, and the experimental results show that our method performs significantly better than other methods. We also report new and reliable drug–target interactions for 26 FDA-approved drugs. Most of the prediction results can be supported by existing references, which shows the effectiveness of our proposed method MCM.

2 Materials and method

In this section, we first describe materials used to obtain the drug similarities and protein targets similarities of two sides in Section 2.1. Then we formulate the multi-view problem for predicting drug targets in Section 2.2. In Section 2.3, a single-view approach is introduced by inductive matrix completion. Finally in Section 2.4, we propose our multi-view drug–target prediction method MCM.

2.1 Materials

The data of drug structures and protein sequences are downloaded from KEGG database [28]. Drug structure similarities are computed by SIMCOMP [29], a software program for structural global alignment using the shared substructures of the two compounds' structures. The similarities between protein sequences are calculated by Smith-Waterman algorithm [30].

The NCI60 human tumour cell line screen method, which is developed by National Cancer Institute (NCI), aims to screen a substances of cytotoxic activity in 60 cell lines for various cancer types. Specifically, the growth inhibition is measured by the sulforhodamine B assay for a cellular protein after a cell line was exposed to a drug for two days. 50% growth inhibition (GI50) is qualified the concentration of compound. The Developmental Therapeutics Program (DTP) human tumour cell line screening data is obtained from the DTP database https://dtp.cancer.gov/, and the gene expression data (mRNA:Affy-U133B, GCRMA-normalised) in NCI-60 cell lines conducted in [31] are downloaded from NCI website [32]. Using drug response data, the drug similarities are computed by the Gaussian kernel, for which the parameter σ is chosen as the median distance of pairwise distances among all drugs. We construct protein chemical similarities from gene expression data in the same way as the drug response similarities.

We obtain 326 common drugs from the drug response data and the drug structure data. Meanwhile, 608 overlapping proteins are also selected from the gene expression data and the protein sequence data. On the Drug Bank Database [33], the known drug–target associations are downloaded. We then obtain 114 known associations among the selected 326 drugs and 608 protein targets.

For either the drugs or the protein targets in our dataset, there are two types of representations: structural and chemical views. The protein sequence similarities and the drug structural similarities are used to construct the structural view representations. On the other hand, we construct chemical view by drug response data and proteins gene expression data in NCI60 cell lines.

2.2 Problem formulation

Suppose we have structural similarities and chemical similarities for $m_{\rm d}$ drugs and $m_{\rm t}$ proteins targets, respectively. Denote the drug–drug similarities and target–target similarities in the structural view as $\boldsymbol{W}_{\rm d}^{\rm s} \in R^{m_{\rm d} \times m_{\rm d}}$ and $\boldsymbol{W}_{\rm t}^{\rm s} \in R^{m_{\rm t} \times m_{\rm t}}$, and denote the drug–drug similarities and target–target similarities in the chemical view as $\boldsymbol{W}_{\rm d}^{\rm c} \in R^{m_{\rm d} \times m_{\rm d}}$ and $\boldsymbol{W}_{\rm t}^{\rm c} \in R^{m_{\rm t} \times m_{\rm t}}$, respectively. Among these drugs and protein targets, the known drug–target associations are denoted as the interaction matrix $\boldsymbol{P} \in R^{m_{\rm d} \times m_{\rm t}}$, which is defined as

$$P_{ij} = \begin{cases} 1, & \text{there is known interaction between the } i\text{th drug} \\ & \text{and the } j\text{th protein,} \\ 0, & \text{otherwise}\,. \end{cases}$$

We also denote $\Omega = \{(i, j) \big| P_{ij} = 1\}$ to be all the drug–target pairs which are known to be interacted. Our goal is to predict new drug–target associations by completing the matrix $\boldsymbol{P}$ based on all the given information. Nomenclature section summarises the notations used in this paper.

2.3 Drug–target prediction by matrix completion with single view side information

The inductive matrix completion is proposed in [34] to recover a latent matrix based on limited information. SIMCLDA [24] method applies inductive matrix completion to predict new associations between lncRNA and diseases. The model is based on the assumption that associations between lncRNAs and disease are dependent on the feature vectors extracted from some side information, such as RNA-RNA similarities and disease-disease similarities. It first extracts features for lncRNAs and diseases from their similarity matrices, respectively, and then applies the inductive matrix completion model with single view side information (MCS) to recover the unknown interactions between lncRNAs and diseases.

Although the method is developed for a different problem, it could be directly used for drug–target prediction. Similarly to lncRNAs or diseases, the feature vectors of drugs or protein targets could be obtained by eigenvalue decomposition of the similarity matrices in the problem of drug repositioning. In detail, we construct drug feature matrix $\boldsymbol{D} \in R^{m_{\rm d} \times k_{\rm d}}$ by the eigenvectors of the drug similarity matrix $\boldsymbol{W}_{\rm d}$ corresponding to its $k_{\rm d}$ largest

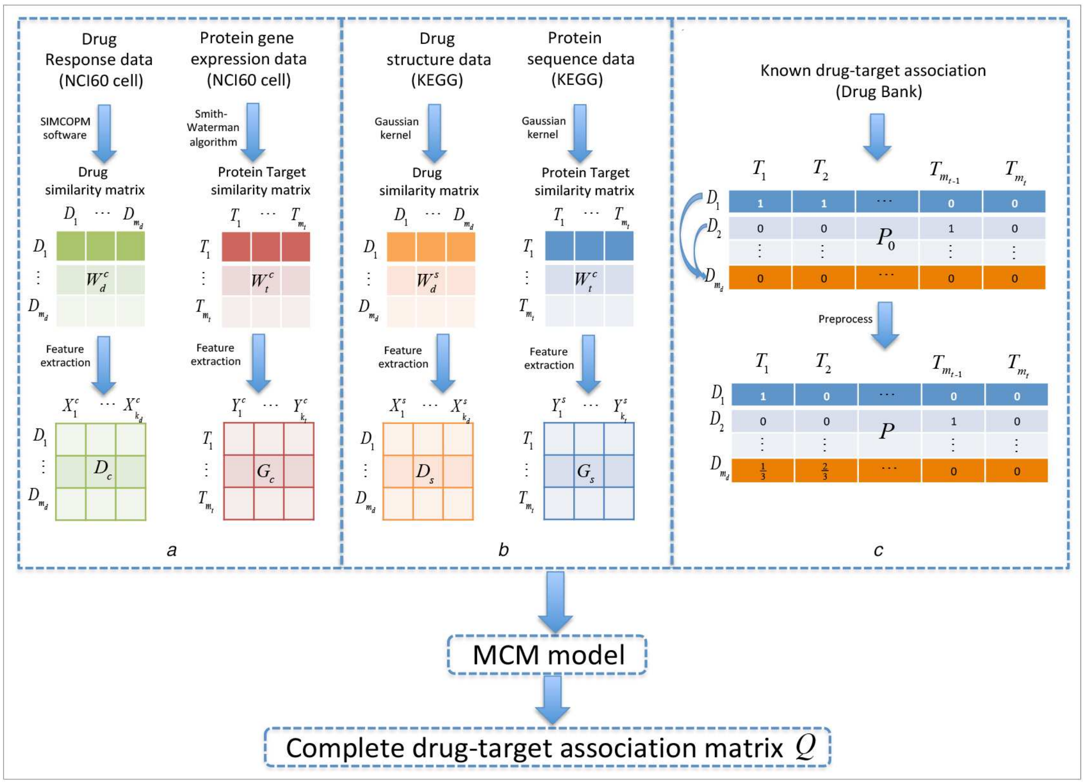
Drug Response data (NCI60 cell)
Protein gene expression data (NCI60 cell)
SIMCOPM software
Smith-Waterman algorithm
Drug similarity matrix
Protein Target similarity matrix
W_d^c
W_t^c
Feature extraction
D_c
G_c
a
Drug structure data (KEGG)
Protein sequence data (KEGG)
Gaussian kernel
Gaussian kernel
Drug similarity matrix
Protein Target similarity matrix
W_d^s
W_t^c
Feature extraction
D_s
G_s
b
Known drug-target association (Drug Bank)
P_0
Preprocess
P
c
MCM model
Complete drug-target association matrix Q

Fig. 1 *Flowchart of our MCM method. We construct similarity matrices in chemical and structural views for drugs and protein targets and extract features from these similarity matrices. Meanwhile, we preprocess the known drug–target association matrix P_0. Finally, a complete drug–target association matrix Q is obtained by MCM model with D_c, G_c, D_s, G_s and P as inputs*

(a) Chemical view construction, ***(b)*** Structural view construction, ***(c)*** Association matrix preprocessing

eigenvalues. Similarly, we could obtain protein feature matrix $\boldsymbol{G} \in R^{m_t \times k_t}$ by k_t eigenvectors of protein similarity matrix $\boldsymbol{W}_t$.

For either the chemical view or the structural view, the interaction matrix $\boldsymbol{P}$ can be recovered by matrix completion with single view side information (MCS) by SIMCLDA proposed in [24]

$$\min_{A \in R^{k_d \times k_t}} \quad \| \boldsymbol{A} \|_* \qquad \text{s.t.} \quad \mathscr{R}_\Omega(\boldsymbol{DAG}^T) = \mathscr{R}_\Omega(P) \tag{1}$$

where $\| \cdot \|_*$ is the nuclear norm, and $\mathscr{R}_\Omega(\cdot): R^{m_d \times m_t} \to R^{m_d \times m_t}$ is defined as follows:

$$\mathscr{R}_\Omega(M)_{ij} = \begin{cases} M_{ij}, & \text{if } (i,j) \in \Omega \\ 0, & \text{otherwise} \end{cases}$$

for any matrix $\boldsymbol{M} \in R^{m_d \times m_t}$ and Ω is a collection of observed indicators of interacting drugs and protein targets. After solving the optimisation problem (1) for the optimal $\boldsymbol{A}$, $\boldsymbol{DAG}^T$ could be used as the completed matrix of $\boldsymbol{P}$. The entry with larger value in the matrix $\boldsymbol{DAG}^T$ implies that the corresponding drugs and protein targets have higher probability to be interacted.

However, SIMCLDA does not consider multiple similarities between lncRNAs or diseases from different fields or views, thus it could not be applied for the case when the multi-view side information is available. In next section, we will propose a matrix completion method for drug–target prediction with multi-view side information.

2.4 MCM for drug–target prediction

Note that for the structural view, we could first compute $\boldsymbol{D}_s$ and $\boldsymbol{G}_s$ from $\boldsymbol{W}_d^s$ and $\boldsymbol{W}_t^s$, respectively, and then apply the MCS model to obtain a corresponding completed matrix $\boldsymbol{D}_s\boldsymbol{A}^s\boldsymbol{G}_s^T$. We call this MCS-S (S is the short for structural). Similarly, as for the chemical view, we could also obtain a completed matrix $\boldsymbol{D}_c\boldsymbol{A}^c\boldsymbol{G}_c^T$, which we call MCS-C (C is the short for chemical). In this section, we extend the above single-view model MCS to the multi-view case. We hope that the two completed matrices obtained from the structural and chemical views are as consistent as enough, and thus propose a MCM as follows:

$$\min_{\boldsymbol{Q} \in R^{m_d \times m_t}, \boldsymbol{A}^s, \boldsymbol{A}^c \in R^{k_d \times k_t}} \quad F_1 + F_2 + F_3 \qquad \text{s.t.} \quad 0 \le Q_{ij} \le 1, \quad i = 1, \dots, m_d, j = 1, \dots, m_t. \tag{2}$$

where

$$F_1 = \frac{1}{2}\left(\| \mathscr{R}_\Omega(\boldsymbol{D}_s\boldsymbol{A}^s\boldsymbol{G}_s^T - \boldsymbol{P}) \|_F^2 + \| \mathscr{R}_\Omega(\boldsymbol{D}_c\boldsymbol{A}^c\boldsymbol{G}_c^T - \boldsymbol{P}) \|_F^2 \right),$$

$$F_2 = \lambda_1\left(\| \boldsymbol{A}^s \|_* + \| \boldsymbol{A}^c \|_* \right),$$

$$F_3 = \lambda_2\left(\| (\boldsymbol{D}_s\boldsymbol{A}^s\boldsymbol{G}_s^T - \boldsymbol{Q} \|_F^2 + \| \boldsymbol{D}_c\boldsymbol{A}^c\boldsymbol{G}_c^T - \boldsymbol{Q} \|_F^2 \right),$$

and $\lambda_1 \ge 0$ and $\lambda_2 \ge 0$ are trade-off parameters. Note that by minimising the first item F_1, the known entries in the completed matrices can be preserved well. Minimising the second item F_2 is to force the low rank of the two matrices $\boldsymbol{A}^s$ and $\boldsymbol{A}^c$ closer. The third term aims to make the two completed interaction matrices be as similar as possible by introducing a common completed matrix $\boldsymbol{Q}$. The details of our method are shown in Fig. 1. We also note that the MCM model could be easily extended for the case when more than two views are available.

2.5 Algorithm

In order to solve the optimisation problem (2), we develop an algorithm by updating $\boldsymbol{A}^s$, $\boldsymbol{A}^c$ and $\boldsymbol{Q}$ alternately. First, we fix $\boldsymbol{A}^s$ and $\boldsymbol{A}^c$ to solve $\boldsymbol{Q}$ and get the following sub-problem:

```
Inputs. $W_d^s$,$W_t^s$,$W_d^c$,$W_t^c$,$P$
        $\lambda_1, \lambda_2, k$
Outputs. the completed matrix $Q$
1. Extract the feature vectors $D_s$ and $G_s$ from $W_d^s$ and $W_t^s$,
   $D_c$ and $G_c$ from $W_d^c$ and $W_t^c$ by eigenvalue decomposition
   for structural and chemical view, respectively.
2. Set initials for $A^s$ and $A^c$. Repeat
     Fix $A^s$ and $A^c$ to obtain $Q$ by (4);
     Fix $A^s$ and $Q$ to solve $A^c$ by (8) - (10) iteratively;
     Fix $A^c$ and $Q$ to solve $A^s$ by (8) - (10) iteratively;
   until the value of the objective function in (2) does not change.
```

Fig. 2 *MCM algorithm for drug–target prediction*

$$\min_{0 \le Q_{ij} \le 1} \ \| \boldsymbol{D}_{\rm s}\boldsymbol{A}^{\rm s}\boldsymbol{G}_{\rm s}^{\rm T} - \boldsymbol{Q} \|_{\rm F}^2 + \| \boldsymbol{D}_{\rm c}\boldsymbol{A}^{\rm c}\boldsymbol{G}_{\rm c}^{\rm T} - \boldsymbol{Q} \|_{\rm F}^2 . \tag{3}$$

The optimal $\boldsymbol{Q}$ for this problem is

$$\hat{\boldsymbol{Q}} = \boldsymbol{Proj}\left(\frac{1}{2}\boldsymbol{D}_{\rm s}\boldsymbol{A}^{\rm s}\boldsymbol{G}_{\rm s}^{\rm T} + \frac{1}{2}\boldsymbol{D}_{\rm c}\boldsymbol{A}^{\rm c}\boldsymbol{G}_{\rm c}^{\rm T}\right), \tag{4}$$

where

$$[\boldsymbol{Proj}(M)]_{ij} = \begin{cases} 0, & \text{if } M_{ij} < 0 \\ 1, & \text{if } M_{ij} > 1 \\ M_{ij}, & \text{otherwise} . \end{cases}$$

Next, we fix $\boldsymbol{A}^{\rm s}$ and $\boldsymbol{Q}$ and solve $\boldsymbol{A}^{\rm c}$ by the following sub-problem:

$$\begin{aligned} \min_{A^{\rm c} \in R^{k_{\rm d} \times k_{\rm t}}} \ & \frac{1}{2} \| \mathscr{R}_\Omega(\boldsymbol{D}_{\rm c}\boldsymbol{A}^{\rm c}\boldsymbol{G}_{\rm c}^{\rm T} - \boldsymbol{P}) \|_{\rm F}^2 + \lambda_1 \| \boldsymbol{A}^{\rm c} \|_* \\ & + \lambda_2 \| \boldsymbol{D}_{\rm c}\boldsymbol{A}^{\rm c}\boldsymbol{G}_{\rm c}^{\rm T} - \boldsymbol{Q} \|_{\rm F}^2 . \end{aligned} \tag{5}$$

Let

$$h(\boldsymbol{A}^{\rm c}) = \frac{1}{2} \| \mathscr{R}_\Omega(\boldsymbol{D}_{\rm c}\boldsymbol{A}^{\rm c}\boldsymbol{G}_{\rm c}^{\rm T} - \boldsymbol{PP}) \|_{\rm F}^2 + \lambda_2 \| \boldsymbol{D}_{\rm c}\boldsymbol{A}^{\rm c}\boldsymbol{G}_{\rm c}^{\rm T} - \boldsymbol{Q} \|_{\rm F}^2 .$$

For any given $\boldsymbol{Z} \in R^{k_{\rm d} \times k_{\rm t}}$, one can approximate $h(\boldsymbol{A}^{\rm c})$ by the following quadratic approximation:

$$\begin{aligned} h(\boldsymbol{A}^c) \simeq \tilde{h}(\boldsymbol{A}^{\rm c}, \boldsymbol{Z}) &= h(\boldsymbol{Z}) + < \nabla h(\boldsymbol{Z}), \boldsymbol{A}^{\rm c} - \boldsymbol{Z} > + \frac{t}{2} \| \boldsymbol{A}^{\rm c} - \boldsymbol{Z} \|_{\rm F}^2 \\ &= \frac{t}{2} \left\| \boldsymbol{A}^{\rm c} - \left(\boldsymbol{Z} - \frac{1}{t}\nabla h(\boldsymbol{Z})\right) \right\|_{\rm F}^2 + h(\boldsymbol{Z}) - \frac{1}{2t} \| \nabla h(\boldsymbol{Z}) \|_{\rm F}^2 , \end{aligned} \tag{6}$$

where

$$\nabla h(\boldsymbol{Z}) = \boldsymbol{D}_{\rm c}^{\rm T}\{\mathscr{R}_\Omega(\boldsymbol{D}_{\rm c}\boldsymbol{Z}\boldsymbol{G}_{\rm c}^{\rm T} - \boldsymbol{P}) + 2\lambda_2(\boldsymbol{D}_{\rm c}\boldsymbol{Z}\boldsymbol{G}_{\rm c}^{\rm T} - \boldsymbol{Q})\}\boldsymbol{G}_{\rm c},$$

$\langle \cdot , \cdot \rangle$ denotes the inner product for matrices, and the proximal parameter t determines the estimation of the second-order gradient $\nabla^2 h(Z)$. Thus, (5) can be rewritten as

$$\min_{A^{\rm c} \in R^{k_{\rm d} \times k_{\rm t}}} \lambda_1 \| \boldsymbol{A}^{\rm c} \|_* + \frac{t}{2} \left\| \boldsymbol{A}^{\rm c} - \left(\boldsymbol{Z} - \frac{1}{t}\nabla h(\boldsymbol{Z})\right) \right\|_{\rm F}^2 . \tag{7}$$

We then apply accelerated gradient descent (APG) [35] to obtain optimal solution of (7) by the following iterative procedure

$$step\ 1\text{: let } \boldsymbol{Z}_l = A_l^c + \gamma_l(\gamma_{l-1}^{-1} - 1)(A_l^c - A_{l-1}^c), \tag{8}$$

$$step\ 2\text{: } solve$$

$$\boldsymbol{A}_{l+1}^c = \min_{A^{\rm c} \in R^{k_{\rm d} \times k_{\rm t}}} \lambda_1 \| \boldsymbol{A}^{\rm c} \|_* + \frac{t}{2} \left\| \boldsymbol{A}^{\rm c} - \left(\boldsymbol{Z}_l - \frac{1}{t}\nabla h(\boldsymbol{Z}_l)\right) \right\|_{\rm F}^2 , \tag{9}$$

$$step3\text{:}compute\, \gamma_{l+1} = \frac{1}{2}\left(\sqrt{\gamma_l^4 + 4\gamma_l^2} - \gamma_l^2\right), \tag{10}$$

where $\gamma_l, \gamma_{l+1} \in (0, 1]$.

For step 2, we solve the optimisation problem by applying the following singular value thresholding algorithm [36]. Let $\boldsymbol{B}_l = \boldsymbol{Z}_l - (1/t)\nabla h(\boldsymbol{Z}_l)$. Suppose the singular value decomposition (SVD) of $\boldsymbol{B}_l$ is

$$B_l = \boldsymbol{V}_1 \Sigma \boldsymbol{V}_2^{\rm T}, \quad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_q),$$

where $\boldsymbol{V}_1 = [v_1^1, v_1^2, \ldots, v_1^q] \in \mathscr{R}^{k_{\rm d} \times q}$ and $\boldsymbol{V}_2 = [v_2^1, v_2^2, \ldots, v_2^q] \in \mathscr{R}^{k_{\rm t} \times q}$ are unitary matrices and $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_q > 0$ are singular values. The solution of (9) is then given by

$$\boldsymbol{A}_{l+1}^c \leftarrow F_t(B_l) = \sum_i^{\sigma_i \ge (\lambda_1/t)} \left(\sigma_i - \frac{\lambda_1}{t}\right)\boldsymbol{v}_1^i(\boldsymbol{v}_2^i)^{\rm T}, \tag{11}$$

where $\boldsymbol{v}_1^i$ and $\boldsymbol{v}_2^i$ are the left and right singular vectors of B_l corresponding to σ_i, respectively.

Finally, we fix $\boldsymbol{A}^{\rm c}$ and $\boldsymbol{Q}$ and solve $\boldsymbol{A}^{\rm s}$ by the similar way that we use to solve $\boldsymbol{A}^{\rm c}$ in the previous step. The iterations stop until the change of the value of the objective function in (2) are less than a small number. We thus obtain the recovered matrix $\tilde{\boldsymbol{P}}$ by $\boldsymbol{Q}$. We show a summary for the procedure to solve the optimisation problem (2) in algorithm box MCM.

2.6 Computation complexity analysis

There are two stages in our algorithm MCM. In the first stage, the eigenvalue decomposition is adopted to extract features for drugs and targets in each view, and a computation cost of $O(m_d^3)$ or $O(m_t^3)$ is required. At each iteration of the second stage, $\boldsymbol{Q}$, $\boldsymbol{A}^{\rm s}$ and $\boldsymbol{A}^{\rm c}$ are updated in three steps, respectively. For the first step, $\boldsymbol{Q}$ is updated by the mean of recovered drug–target association matrices from both two views, which requires a computation cost of $O(m_{\rm d}k_{\rm d}k_{\rm t} + m_{\rm d}k_{\rm t}m_{\rm t})$. For the second step, $\boldsymbol{A}^{\rm c}$ is updated by the SVD of $B = Z - (1/t)\nabla h(Z)$, where $h(Z)$ is the quadratic approximation of $\boldsymbol{A}^{\rm c}$ with any given $Z \in R^{k_{\rm d} \times k_{\rm t}}$. A computation cost of $O(k_{\rm d}^2 k_{\rm t} + k_{\rm t}^3)$ is required for the second step. For the third step, the same computation cost is required as the second step. Overall, the MCM algorithm takes computation time of $O(m_{\rm d}^3)$ or $O(m_t^3)$ (see Fig. 2).

3 Experiments results

3.1 Evaluation of our method

We evaluate the performance of our methods MCS-S, MCS-C and MCM by comparing their prediction accuracies with some other existing methods including single view methods including support vector machine (SVM), bipartite graph learning (BGL) [2], SPGraph [22] and single-view rank embedding (SLRE) (2017) [23] and multi-view methods including multi-view SVM, multi-view penalised graph (MPGraph) [22] and multi-view rank embedding (MLRE) (2017) [23]. Among our methods, MCS-S and MCS-C are single-view methods for structural view and chemical view, respectively, while MCM is the multi-view method. We first describe the experimental settings in detail, then introduce the comparison methods, and finally show results for all experiments.

3.1.1 Experimental setting: We collect a smaller dataset from the whole dataset by removing drugs with no known targets and targets with no known drugs. About 65 drugs and 80 targets are remained, and there are 114 known pairs among them in total. For the smaller dataset, we design two experimental settings called NT (new coming target) and NDNT (new coming drug and new coming target) to compare different methods with our methods. For the NT setting, our goal is to find the drugs that are associated with the test

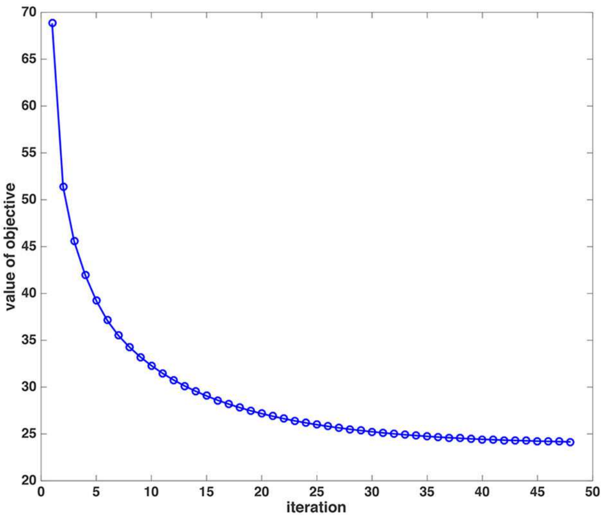

Fig. 3 *Convergence of our MCM algorithm*

targets. In the setting of NDNT, we aim to obtain the interactions between test drugs and test targets.

For the NT setting, we divide all the 80 targets in the small dataset into five folds. Each fold of targets is chosen as test data in turn while the remaining four folds of targets were considered as training data. We use the associations between the training targets and all the drugs to recover the interaction matrix by methods of MCS-S, MCS-C and MCM, respectively. When the interaction matrix is computed, the probabilities of interactions between test targets and all the drugs are obtained. For each test target, the drugs with the k highest association values are considered to be interacted with it. By changing the threshold k, we can obtain a receiver operating characteristic curve and the corresponding area under the curve (AUC) value. In our multi-view method, we calculate the AUC values in $\boldsymbol{D}_s\boldsymbol{A}^s\boldsymbol{G}_s^T$, $\boldsymbol{D}_c\boldsymbol{A}^c\boldsymbol{G}_c^T$ and $\boldsymbol{Q}$, and report the maximum value among these three AUC values as the final AUC value in all of our multi-view experiments. We take the same way to calculate AUC values in other compared multi-view methods. In NDNT, we divide all drugs and all targets into five folds, respectively, and select drugs and targets in each fold as test data for each time while the other remaining drugs and targets as training data. With the known associations between training drugs and training targets, one can recover the potential interaction matrix and compute AUCs with the same way in NT setting. We repeat the procedure for 50 times in each setting and report the average AUCs and standard errors.

In all three methods: MCS-S, MCS-C and MCM, the parameters λ, λ_1 and λ_2 are chosen from the set $\{0.001, 0.01, 0.1, 1\}$. We fixed $k = 40$ and reported the best results when parameters are chosen from the above set. To make a fair comparison, the same parameter range of λ and k are used to compute the final results for SPGraph, MPGraph, SLRE and MLRE approaches.

3.1.2 Comparison methods:

(a) Single-view and multi-view SVMs: On training datasets, SVMs can learn a classifier which can classify pairs of drug–target into categories 'having interaction' or 'not having interaction'. The Kronecker product $K = \boldsymbol{W}_d \otimes \boldsymbol{W}_t$ of drug similarity matrix $\boldsymbol{W}_d$ and protein similarity matrix $\boldsymbol{W}_t$ represents the kernel between drug–protein pairs. For each specific view, SVM with the corresponding Kronecker kernel is applied to solve drug–target prediction problem. For the multi-view SVM method, we simply apply the SVM approach with multiple kernels from the two views.
(b) BGL [2]: For either structural view or chemical view, BGL can be used to predict drug–target associations as a single-view approach.
(c) SPGraph and MPGraph [22]: SPGraph is a single-view method to predict drug–target associations, and it can be used for either view. MPGraph is the extended multi-view method, in which both two views can be integrated for drug–target prediction.
(d) SLRE and MLRE [23]: SLRE is a low-rank embedding based single-view method, which can be used for either view. MLRE is a multi-view method which uses both structural and chemical views for identifying drug targets.

3.1.3 Results: We first checked the convergence property of our MCM algorithm with $\lambda_1 = 0.1$, $\lambda_2 = 0.1$ and $k = 40$ on the smaller dataset. The results are shown in Fig. 3, where the x-axis represents the times of iteration, and the y-axis represents the values of the optimisation objective function. From the figure, we can see that the algorithm converges quite fast.

The results for our methods and the comparison methods with $k = 40$ are shown in Table 1, where '—' denotes that the corresponding single-view method does not have multi-view version. Note that single-view methods with structural view obtained higher AUC values than those with chemical view in most cases. For both of the two views, the single-view method MCS performed the best in both the NT and the NDNT settings. We can see from the table that, in both settings, graph-based multi-view method (MPGraph) and multi-view method through low rank embedding (MLRE) performed better than their corresponding single-view methods (SPGraph and SLRE), and our matrix completion based multi-view method (MCM) worked better than the corresponding single-view method (MCS). The results imply that applying multi-view information of drugs and targets could strengthen the prediction accuracy. Besides, our method MCM performed the best among the multi-view methods for the settings of both NT and NDNT. This shows that our methods are effective in discovering the potential associations between drugs and targets.

Table 1 Average AUCs for all nine methods and *t*-test *p*-values of significant difference in results between our methods (bold) and the second best methods (italic)

	SVM	BLG	SPGraph	SLRE	MCS	*P*-value
Structure view						
NT	0.492	0.443	*0.509*	0.498	**0.598**	1.969×10^{-15}
NDNT	0.523	0.479	0.527	*0.591*	**0.660**	1.738×10^{-07}
Chemical view						
NT	0.493	0.497	*0.541*	0.513	**0.543**	5.534×10^{-01}
NDNT	0.472	*0.497*	0.477	0.431	**0.575**	4.035×10^{-07}

	MKL-SVM	BLG	MPGraph	MLRE	MCM	P-value
Multi-view						
NT	0.536	—	*0.565*	0.517	**0.616**	2.173×10^{-26}
NDNT	0.520	—	0.599	*0.629*	**0.719**	1.791×10^{-11}

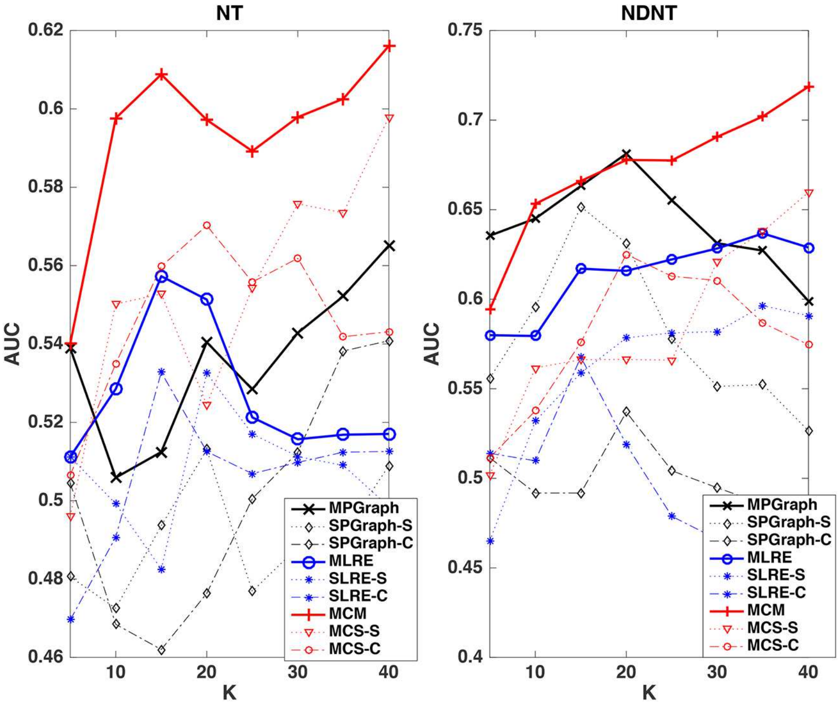

Fig. 4 *Average AUC results computed by nine approaches in two settings of NT and NDNT with different values of the parameter k*

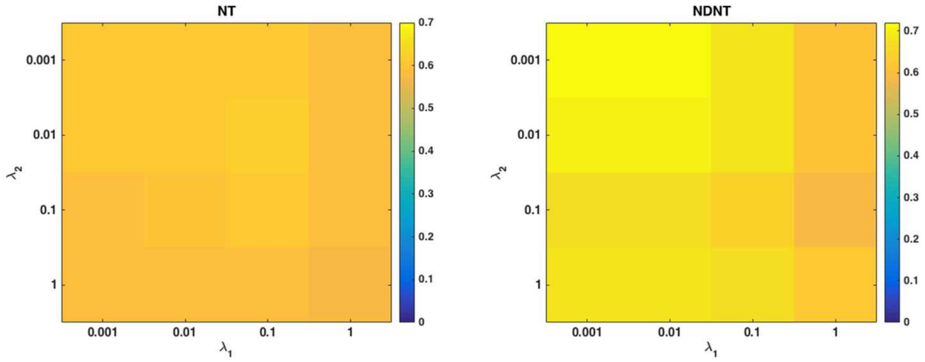

Fig. 5 *Average AUCs on NT and NDNT settings with different values of parameters λ_1 and λ_2*

To show whether the MCM method outperforms significantly the other methods, we also calculated the *t*-test *p*-values by comparing the 50 AUCs between our MCM method and the second best method. In Table 1, we reported the *p*-values for all the cases. It shows that our method could obtain significantly better results than the compared methods.

To show the robustness of our approaches with respect to the parameter *k*, we took *k* from the set {5:5:40} and reported the results of SPGraph-S, SPGraph-C, MPGraph, SLRE-S, SLRE-C, MLRE, MCS-S, MCS-C and MCM for the two settings NT and NDNT in Fig. 4. In the NT setting, we can see that the graph-based methods and the low-rank embedding based methods sometimes performed even worse than the matrix completion based single-view method MCS. In the NDNT setting, all methods obtained higher AUC values and performed stably. We also note that generally the multi-view methods performed better than the single-view methods for any *k* in the parameter set, and our multi-view method of MCM performed the best for each case.

To show the robustness of our method with respect to the parameters λ_1 and λ_2, we reported the results of the average AUC values on both NT and NDNT settings with different values of these two parameters varying from the set of {0.001, 0.01, 0.1, 1} in Fig. 5. We can see that our method could obtain better results on NDNT setting than NT setting. The results on each setting changed a little when the parameters vary. This shows that our method MCM performed robustly for the given set of parameters λ_1 and λ_2.

3.2 Prediction of new drug–target associations in the whole dataset

We applied our proposed MCM method on the whole dataset to predict new drug–target interactions by completing the association matrix $\boldsymbol{P}$. The parameter is set as $\lambda_1 = 0.1$, $\lambda_2 = 0.1$ and $k = 40$. In the proposed MCM method, when the latent matrix is recovered from $\boldsymbol{P}$, the probabilities of the associations between all drugs and targets are obtained. For target *i*, we selected the top *t* percentage of drugs based on the values in the *i*th column of the completed matrix and predicted them as the drugs that can interact with the target.

We evaluated the prediction results of our method of MCM in the following steps. We first randomly removed *l* known interactions from the association matrix $\boldsymbol{P}$, where *l* is a number chosen from the set {5,10,15,20}, and solved the MCM model to recover the interaction matrix $\boldsymbol{P}$. We then selected the associated drug–target pairs in the complete $\boldsymbol{P}$ by varying the threshold *t* in the set {10,20,30,40,50,60}, and finally computed the percentage of the recalled drug–target pairs. Fig. 6 shows the percentage of the recalled pairs with different rank thresholds *t* and different number of removed known interactions *l*. We can see that the percentage of recalled pairs increases along with the increase of *t* at each fixed *l*. In most cases, over 50% interactions that were removed in the first step could be recovered by our method of MCM. This implies that the prediction results recovered from our MCM method are highly credible. Furthermore, for the new drug *d* that we are interested in, we conducted prediction experiment between 66 drugs (65 drugs

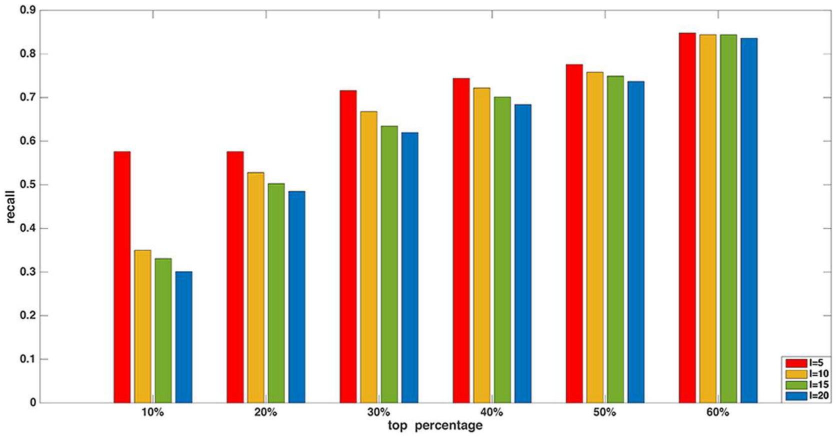

Fig. 6 *Percentage of the recalled pairs with different rank thresholds t and different number of removed known interactions l*

from the smaller dataset and the new drug *d*) and 608 proteins with 114 known associations to find its corresponding target proteins. Note that there are no known interactions between the drug *d* and the 608 proteins. The same parameter settings as the previous experiments are used. Table 2 shows the new identified targets for 26 Food and Drug Administration (FDA)-approved drugs with the top 0.5% of recovered probabilities in each column of the recovered latent matrix when the parameter *k* is chosen to be 40. We found that some of the predicted targets in Table 2 can be validated by some existing research results, which are discussed in the discussion part.

Table 2 Predicted targets for 26 FDA-approved drugs by our MCM method

KEGG ID	Drug name	Gene name
D05905	sparsomycin	UROD, JARID1D, KIF1A
D00372	thiabendazole	SLC1A4
D00433	silver sulfadiazine	SDS, SCNN1A, RARRES1, TSTA3, NPPB, SST, SULT2B1, GSTA2, CPB1
D03936	econazole	FCER1A, NDUFS8, SCNN1A, ALOX5, IFNAR2, RARA, CMA1, GSTM5
D00413	zidovudine	ALDH2
D00237	auranofin	COL1A1, TYR, TTPA, PLCL1, KLK1, APOE, MTAP, CP, S100P, EEA1, JARID1D, P4HB, CRYBB1
D01334	cyclacillin	ALDH2, CLPP
D01364	ciclopirox	VCAM1, JARID1D
D04115	1,8-cineole	JARID1D
D00214	dactinomycin	PYGL, COL1A1, SLC1A4, NDUFS1, HMOX1, TGM2, ACADM, CFD, JARID1D, POR
D06265	uracil mustard	JARID1D
D00188	cholecalciferol	GRIK1, GRIA1, GRIK2, GRIA2, GRIA4, GRIK3
D00297	digitoxin	NOS1, SLC1A4
D06067	temozolomide	SDS, CALM1, ACVR1B
D00254	carmustine	CALM1, DNMT1, MCM6, PAICS
D00478	procarbazine	ALDH2
D00343	ifosfamide	ALDH3B2
D00966	tamoxifen	SCNN1A
D00153	testolactone	ALDH2, GRIK1, GRIA1
D00399	valproic acid	GAMT, OXTR, CAST, CDC2, UCK2, NR1H2, ITPKA, HAGH, SCN4A, CAPN1
D01068	vinblastine	SLC1A4, NDUFS1, HMOX1, CFD
D01211	leucovorin	GRM1, GRM4, GRM8, MGST2, GRM3
D00275	cisplatin	SDS
D00266	chlorambucil	JARID1D
D01363	carboplatin	JARID1D, CLPP
D01747	idarubicin	SLC1A4, NDUFS1, HMOX1, CFD

4 Discussion

In this section, we discuss the biological meaning of the predicted drug–target interactions by our method MCM.

Carmustine is usually referred as an antineoplastic agent used in the treatment of brain tumours. Hagelkrüys [37] reported that the absence of DNMT1 in the brain leads to a severe neurological phenotype, a dramatically disorganised brain architecture and death. This supports our predicted interaction between the target DNMT1 and the drug Carmustine. The work in [38] shows that IDNMT1 overexpression is correlated with a reduction of MGMT protein expression in high-grade astrocytic tumour. It is reported in [39] that astrocytic tumours form the most common histologic group among childhood brain tumours. This further validates DNMT1 plays an important role in brain tumour and DNMT1 most likely is a key target for drug Carmustine. Besides, it has been known that Tamoxifen can be used for the treatment and prevention of estrogen receptor positive breast cancer. Varley *et al.* [40] reported two fusion transcripts that were identified in breast cancer cell lines, confirmed across breast cancer primary tumours, and were not detected in normal tissues (SCNN1A-TNFRSF1A and CTSD-IFITM10). This strongly validates our predicted drug–target interaction of Tamoxifen and SCNN1A, which is predicted by our methods. Another drug Testolactone is an antineoplastic agent that is used to treat advanced breast cancer. Choi *et al.* [41] found that alcohol and genetic polymorphisms of cyp2e1 and aldh2 play an important role in breast cancer development. This supports the predicted interaction of Testolactone and ALDH2 in our results.

Leucovorin sometimes can be used in combination with 5-fluorouracil to prolong survival in the palliative treatment of patients with advanced colorectal cancer. Yi *et al.* [42] demonstrated that expression of GRM3 is significantly upregulated in majority of human colonic adenocarcinomas tested and colon cancer cell lines. GRM3 and Leucovorin are all related to colon cancer or colorectal cancer so they probably have some interaction, so our finding of interactions between them is reasonable. Valproic acid is a histone deacetylase inhibitor and is under investigation for the treatment of HIV and various cancers. In our prediction results, we found that OXTR, UCK2 and ITPKA may be the targets for valproic acid. We also found evidence which indicates that all these targets play important roles in various types of cancer. Zhong *et al.* [43] showed that OXT receptor (OXTR) is the primary target of OXT in androgen-independent prostate cancer cell lines (DU145 and PC3). UCK2 is of particular scientific interest due to its overexpression in tumour cell lines [44], which makes it a target in anti-cancer treatments [45]. Wang *et al.* [46] showed that ITPKA expression is up-regulated in many types of cancer including lung and breast cancers, and overexpressed ITPKA contributes to tumourigenesis. These results suggest that valproic acid may interact with targets OXTR, UCK2 and ITPKA to function in different types of cancer, which supports our results.

Both carboplatin and chlorambucil can possess antineoplastic activity or be used as antineoplastic agent for the treatment of various malignant and non-malignant diseases. The results in [47] demonstrated that JARID1D levels were highly down-regulated in metastatic prostate tumours compared with normal prostate tissues and primary prostate tumours. This indicates that JARID1D might be the target for carboplatin and chlorambucil in the treatment of prostate cancer. Idarubicin is a kind of anthracycline antineoplastics. The results in [48] showed that the panel with NDUFS1 and NDUFS8 reflecting tumour metabolism status is a novel prognostic predictor for lung cancer. This indicates that NDUFS1 would be the target for idarubicin in the treatment of lung cancer.

5 Conclusion

Many research results have already shown the effectiveness of multi-view methods for the applications when multiple information of an object are available. In this work, we propose a MVMC method for prediction of the interactions between two types of samples, say drugs and targets. We apply a single-view approach MCS to identify drug targets by integrating the structural information from drug structures and protein sequences, or integrating the chemical information from both drug response and gene expression. We then extend the single-view MCS method to the corresponding multi-view approach MCM, which jointly considers both the structural and chemical information of the drugs and proteins. Our experimental results demonstrate that our approaches work significantly the best in most cases. Although in this work we only consider two types of information for drugs and proteins, our proposed MCM method can be applied for the case when more than two views are available. Extending MCM to three views is an interesting topic, which could strengthen the learning ability. We will do more research on this in the future.

6 Acknowledgments

The work was supported by the NSFC projects (no. 11471256 and no. 11631012) and the Fundamental Research Funds for the Central Universities.

7 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), p. e1004760

[2] Yamanishi, Y., Araki, M., Gutteridge, A., *et al.*: 'Prediction of drug–target interaction networks from the integration of chemical and genomic spaces', *Bioinformatics*, 2008, **24**, (13), pp. i232–i240

[3] Bleakley, K., Yamanishi, Y.: 'Supervised prediction of drug–target interactions using bipartite local models', *Bioinformatics*, 2009, **25**, (18), pp. 2397–2403

[4] Mizutani, S., Pauwels, E., Stoven, V., *et al.*: 'Relating drugprotein interaction network with drug side effects', *Bioinformatics*, 2012, **28**, (18), p. i522

[5] Chen, J., Zhang, S.: 'Integrative analysis for identifying joint modular patterns of gene-expression and drug-response data', *Bioinformatics*, 2016, **32**, (11), pp. 1724–1732

[6] Li, L., Zhou, X., Ching, W., *et al.*: 'Predicting enzyme targets for cancer drugs by profiling human metabolic reactions in NCI-60 cell lines', *BMC Bioinformatics*, 2010, **11**, (1), p. 501

[7] Dorothea, E., Alexander, I., Olga, P., *et al.*: 'Drug target prediction and repositioning using an integrated network-based approach', *Plos One*, 2013, **8**, (4), pp. e60618–e60618

[8] Ding, H., Takigawa, I., Mamitsuka, H., *et al.*: 'Similarity-based machine learning methods for predicting drug–target interactions: a brief review', *Brief. Bioinform.*, 2014, **15**, (5), pp. 734–747

[9] Zheng, X., Ding, H., Mamitsuka, H., *et al.*: 'Collaborative matrix factorization with multiple similarities for predicting drug–target interactions'. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Chicago, 2013, pp. 1025–1033

[10] Li, L., He, X., Borgwardt, K.: 'Multi-target drug repositioning by bipartite blockwise sparse multi-task learning', *BMC Syst. Biol.*, 2018, **12**, (4), p. 55

[11] Shen, R., Mo, Q., Schultz, N., *et al.*: 'Integrative subtype discovery in glioblastoma using iCluster', *Plos One*, 2012, **7**, (4), p. e35236

[12] Mo, Q., Wang, S., Seshan, V., *et al.*: 'Pattern discovery and cancer gene identification in integrated cancer genomic data', *Proc. Natl. Acad. Sci. USA*, 2013, **110**, (11), pp. 4245–4250

[13] Gönen, M., Margolin, A.: 'Localized data fusion for kernel k-means clustering with application to cancer biology', *Adv. Neural. Inf. Process. Syst.*, 2014, pp. 1305–1313

[14] Lanckriet, G., Cristianini, N., Bartlett, P., *et al.*: 'Learning the kernel matrix with semidefinite programming', *J. Mach. Learn. Res.*, 2004, **5**, (1), pp. 27–72

[15] Yu, S., Tranchevent, L., Liu, X., *et al.*: 'Optimized data fusion for kernel k-means clustering', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (5), pp. 1031–1039

[16] Tilman, L., Joachim, M.: 'Fusion of similarity data in clustering'. Proc. of Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2006

[17] Tang, W., Lu, Z., Dhillon, I.: 'Clustering with multiple graphs'. IEEE Int. Conf. on Data Mining, Sparks, NV, United States, 2009, pp. 1016–1021

[18] Song Chen, C., Chuang, Y., Huang, H.: 'Affinity aggregation for spectral clustering'. IEEE Conf. on Computer Vision and Pattern Recognition, Providence, Rhode Island, 2012, pp. 773–780

[19] Kumar, A., Rai, P.: 'Co-regularized multi-view spectral clustering'. Int. Conf. on Neural Information Processing Systems, Granada Spain, 2011, pp. 1413–1421

[20] Wang, B., Mezlini, A., Demir, F., *et al.*: 'Similarity network fusion for aggregating data types on a genomic scale', *Nat. Methods*, 2014, **11**, (3), pp. 333–337

[21] Zhao, P., Jiang, Y., Zhou, Z.: 'Multi-view matrix completion for clustering with side information'. Pacific-Asia Conf. on Knowledge Discovery and Data Mining, Halifax, Nova Scotia - Canada, 2017, pp. 403–415

[22] Li, L.: 'MPGraph: multi-view penalised graph clustering for predicting drug–target interactions', *IET Syst. Biol.*, 2014, **8**, (2), pp. 67–73

[23] Li, L., Cai, M.: 'Drug target prediction by multi-view low rank embedding', *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2017, **PP**, (99), pp. 1–1

[24] Lu, C., Yang, M., Luo, F., *et al.*: 'Prediction of lncRNA-disease associations based on inductive matrix completion', *Bioinformatics*, 2018, **34**, pp. 3357–3364

[25] Natarajan, N., Dhillon, I.: 'Inductive matrix completion for predicting gene–disease associations', *Bioinformatics*, 2014, **30**, (12), pp. i60–i68

[26] Li, R., Dong, Y., Kuang, Q., *et al.*: 'Inductive matrix completion for predicting adverse drug reactions (ADRs) integrating drug–target interactions', *Chemometr. Intell. Lab. Syst.*, 2015, **144**, pp. 71–79

[27] Chen, X., Wang, L., Qu, J., *et al.*: 'Predicting mirna-disease association based on inductive matrix completion', *Bioinformatics*, 2018, **34**, (24), pp. 4256–4265

[28] Kanehisa, M., Goto, S., Hattori, M., *et al.*: 'From genomics to chemical genomics: new developments in KEGG', *Nucleic Acids Res.*, 2006, **34**, (Database issue), p. D354

[29] Hattori, M., Okuno, Y., Goto, S., *et al.*: 'Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways', *J. Am. Chem. Soc.*, 2003, **125**, (39), pp. 11853–11865

[30] Smith, T.F., Waterman, M.S.: 'Identification of common molecular subsequences', *J. Mol. Biol.*, 1981, **147**, (1), pp. 195–197

[31] Shankavaram, T., Reinhold, C., Nishizuka, S., *et al.*: 'Transcript and protein expression profiles of the nci-60 cancer cell panel: an integromic microarray study', *Mol. Cancer Ther.*, 2007, **6**, (3), pp. 820–832

[32] Reinhold, C., Sunshine, M., Liu, H., *et al.*: 'Cellminer: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set', *Eur. J. Cancer*, 2012, **48**, (14), pp. 82–82

[33] Knox, C., Law, V., Jewison, T., *et al.*: 'Drugbank 3.0: a comprehensive resource for 'omics' research on drugs', *Nucleic Acids Res.*, 2010, **39**, pp. D1035–D1041

[34] Jain, P., Dhillon, I.S.: 'Provable inductive matrix completion', CoRR, abs/1306.0626, 2013

[35] Toh, K., Yun, S.: 'An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems', *Pac. J. Optim.*, 2010, **6**, (3), pp. 615–640

[36] Cai, J., Candès, J., Shen, Z.: 'A singular value thresholding algorithm for matrix completion', *SIAM J. Optim.*, 2008, **20**, (4), pp. 1956–1982

[37] Hagelkrüys, A.: 'The role of hdac1 and dnmt1 in erythropoiesis and brain development', *Nat. Methods*, 2009, **11**, (3), pp. 333–337

[38] Wan, F., Rahman, K., Nafi, S., *et al.*: 'Overexpression of DNA methyltransferase 1 (DNMT1) protein in astrocytic tumour and its correlation with O6-methylguanine-DNA methyltransferase (MGMT) expression', *Int. J. Clin. Exp. Pathol.*, 2015, **8**, (6), p. 6095

[39] Burzynski, S.R.: 'Treatments for astrocytic tumors in children', *Pediatr. Drugs*, 2006, **8**, (3), pp. 167–178

[40] Varley, E., Gertz, J., Roberts, S., *et al.*: 'Recurrent read-through fusion transcripts in breast cancer', *Breast Cancer Res. Treat.*, 2014, **146**, (2), pp. 287–297

[41] Choi, J.Y., Abel, J., Neuhaus, T., *et al.*: 'Role of alcohol and genetic polymorphisms of CYP2E1 and ALDH2 in breast cancer development', *Pharmacogenetics*, 2003, **13**, (2), pp. 67–72

[42] Yi, H., Geng, L., Black, A., *et al.*: 'The miR-487b-3p/GRM3/TGFβ signaling axis is an important regulator of colon cancer tumorigenesis', *Oncogene*, 2017, **36**, (24), pp. 3477–3489

[43] Zhong, M., Clarke, S., Khan, S.: 'Abstract 474: the essential role of giα2 in prostate cancer progression', *Cancer Res.*, 2012, **72**, (8 Supplement), pp. 474–474

[44] Rompay, A., Norda, A., Lindén, K., *et al.*: 'Phosphorylation of uridine and cytidine nucleoside analogs by two human uridine-cytidine kinases', *Mol. Pharmacol.*, 2001, **59**, (5), pp. 1181–1186

[45] Schumacher, F., Wang, Z., Skotheim, R., *et al.*: 'Testicular germ cell tumor susceptibility associated with the UCK2 locus on chromosome 1q23', *Hum. Mol. Genet.*, 2013, **22**, (13), pp. 2748–2753

[46] Wang, Y., Ma, X., Zhang, Y., *et al.*: 'ITPKA gene body methylation regulates gene expression and serves as an early diagnostic marker in lung and other cancers', *J. Thorac. Oncol.*, 2016, **11**, (9), pp. 1469–1481

[47] Li, N., Dhar, S., Chen, T., *et al.*: 'JARID1D is a suppressor and prognostic marker of prostate cancer invasion and metastasis', *Cancer Res.*, 2016, **76**, (4), p. 831

[48] Su, C., Chang, Y., Yang, C., *et al.*: 'The opposite prognostic effect of NDUFS1 and NDUFS8 in lung cancer reflects the oncojanus role of mitochondrial complex I', *Sci. Rep.*, 2016, **6**, p.31357