Selective Matrix Factorization for Multi-Relational Data Fusion *

Yuehui Wang[1], Guoxian Yu[1,3]*, Carlotta Domeniconi[2], Jun Wang[1] and Xiangliang Zhang[3] and Maozu Guo[4]

[1] College of Comp. & Inf. Sci., Southwest University, Chongqing, China
[2] Department of Computer Science, George Mason University, Fairfax, USA
[3] King Abdullah University of Science and Technology, Thuwal, SA
[4] College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China
{yuehuiwang, gxyu, kingjun}@swu.edu.cn, carlotta@cs.gmu.edu, xiangliang.zhang@kaust.edu.sa, guomaozu@bucea.edu.cn *Corresponding author: gxyu@swu.edu.cn(Guoxian Yu)

Abstract. Matrix factorization based data fusion solutions can account for the intrinsic structures of multi-relational data sources, but most solutions equally treat these sources or prefer sparse ones, which may be irrelevant for the target task. In this paper, we introduce a Selective Matrix Factorization based Data Fusion approach (SelMFDF) to collaboratively factorize multiple inter-relational data matrices into low-rank representation matrices of respective object types and optimize the weights of them. To avoid preference to sparse data matrices, it additionally regularizes these low-rank matrices by approximating them to multiple intra-relational data matrices and also optimizes the weights of them. Both weights contribute to automatically integrate relevant data sources. Finally, it reconstructs the target relational data matrix using the optimized low-rank matrices. We applied SelMFDF for predicting inter-relations (lncRNA-miRNA interactions, functional annotations of proteins) and intra-relations (protein-protein interactions). SelMFDF achieves a higher AUROC (area under the receiver operating characteristics curve) by at least 5.88%, and larger AUPRC (area under the precision-recall curve) by at least 18.23% than other related and competitive approaches. The empirical study also confirms that SelMFDF can not only differentially integrate these relational data matrices, but also has no preference toward sparse ones.

Keywords: Matrix factorization · Data fusion · Multi-relational data · Association prediction.

1 Introduction

With the rapid growth of Internet and modern technologies, we can obtain various data sources that are directly related to the main task, and also other

* This work is supported by NSFC (61872300, 61741217 and 61873214).

data sources indirectly related to the task, which can still facilitate the completion of this task. For example, the accuracy of gene function prediction can be improved by integrating the gene-level data (gene expression, gene-gene interactions), and also by fusing transcript-level data (miRNA-gene interactions, miRNA-miRNA interactions) that convey complementary information about gene functions [7,26]. The ever-increasing heterogeneous data sources make data fusion approaches increasingly popular over the past decade, which aim to collectively explore interesting patterns from multiple data sources, and to reduce the impact of noisy or irrelevant ones [7, 15].

An intuitive solution to fuse multiple data sources is concatenating the feature vectors of the same object across different data sources into a longer feature vector, and then applying off-the-shelf learners on this long vector. But this concatenation ignores the intrinsic characteristics of these feature vectors and may (and often does) suffer from the issue of curse of dimensionality and of missing features. Another intuitive solution is to train a classifier on each feature view and then combine these classifiers for ensemble prediction [21], but this ensemble solution may be impacted by low-quality base classifiers independently trained on individual views, which can not ensure a base classifier with sufficient accuracy. Furthermore, the early fusion (feature concatenation) and late fusion (classifier ensemble) can not capture heterogeneous relations between different object types. For these reasons, many *inter-median* data fusion solutions have been proposed in recent years [6, 13, 23].

Inter-median data fusion methods can be generally divided into three categories: *multiple kernel(network) learning*-based (MKL), *Bayesian network*-based (BN) and *matrix factorization*-based (MF) [7]. MKL methods firstly transform multi-relational data matrices onto the homologous data matrices that are directly related with the target task, and then applies different techniques to combine these transformed data matrices for prediction [8,13,23]. These MKL-based methods can selectively integrate multiple homologous data matrices. However, they have to transform heterogeneous features or project multi-relational data into a common feature space before fusion. This hand-crafted transformation and projection may enshroud the intrinsic structure of multi-relational data, and thus does not make full usage of them [26]. BN-based approaches combine the concepts from probability and graph theory to represent and model causal relations between random variables [17]. BN was initially applied to gene function prediction [18] and also shows the potentiality in patient-specific data integration [25]. Although BN-based solution can capture conditional dependence between data sources and variables, it suffers from a heavy computational limitation and asks for sufficient training data with labels.

MF-based solutions generally factorize multiple data matrices into low-rank matrices to explore latent relationships between objects across different data sources. Solutions in this type do not need to project multi-relational data matrices into the common feature space and thus can account for the intrinsic structure of these information sources. To name a few, Ding *et al.* [5] extended the classical nonnegative matrix factorization (NMF) [14] to nonnegative

matrix tri-factorization (NMTF) to co-cluster heterogeneous data, but NMTF can only fuse inter-relational data matrices and ignore the intra-relational ones. Wang *et al.* [19] proposed the symmetric nonnegative matrix tri-factorization (SNMTF) to simultaneously cluster different types of objects, and incorporates the intra-relational ones through manifold regularization [1]. However, SNMTF has a heavy computational complexity and large runtime, because it performs matrix factorization on a big matrix, whose block matrices embody inter-relations between objects. Zitnik and Zupan [26] developed a penalized matrix tri-factorization based model (DFMF) to jointly factorize multiple relational data matrices for predicting gene functions and pharmacologic actions.

These aforementioned MF-based solutions show great potential in exploring the underlying relations between objects, but they ignore the different relevances of multi-relational data sources, since they implicitly assume each source having equal relevance toward the target prediction task, while they may not (and often does). To overcome this problem, Fu *et al.* [6] introduced a MF-based model (MFLDA) to predict lncRNA-disease associations by assigning different weights to multiple inter-relational data matrices for objects of different types and by jointly factorizing these matrices into low-rank ones. MFLDA then uses the optimized matrices to reconstruct the target matrix to predict new inter-relations between objects of different types. However, MFLDA does not account for the different relevances of intra-relational data matrices for objects of the same types, and thus its performance may be compromised by the low-quality or irrelevant data sources. To simultaneously account for the different relevances of multiple intra-relational data matrices, MFLDA was further extended to WMFLDA, which can selectively fuse multiple intra-relation matrices [24]. However, these extended solutions *prefer* to assigning larger weights to *sparse* data matrices, or have a priority toward sparser ones, which may be irrelevant (or even harmful) for the target task. In fact, this preference is also suffered by many MKL-based solutions [9, 20, 23].

To address these issues, we propose a *Sel*ective *M*atrix *F*actorization based *D*ata *F*usion (SelMFDF) solution for integrating multi-relational data. SelMFDF can avoid preferring the sparse relational data matrices during the fusing process. It performs collaborative matrix tri-factorization to optimize the low-rank representation matrices of respective object types and the weights of inter-relational data matrices. To selectively integrate multiple intra-relational data matrices, it further optimizes these low-rank matrices by approximating them to multiple intra-relational data matrices and the weights of these matrices. These two types of weights contribute to identify relevant data sources and remove irrelevant ones. After that, it approximates the target relational data matrix using the optimized low-rank matrices. The main contributions of this paper are summarized as follows:

(i) Our introduced SelMFDF can respect and explore the intrinsic structures of multi-relational data matrices to simultaneously predict inter(intra)-relation between objects of different (same) types, automatically discard irrelevant data sources and credit larger weights to the more relevant ones.

(ii) An alternative optimization procedure is developed to jointly optimize the low-rank matrix approximations and weights of multi-relational data matrices for the target prediction task.

(iii) Empirical study on predicting lncRNA-miRNA associations, gene functions and protein-protein interactions shows that SelMFDF significantly outperforms the related and competitive methods NMTF [5], SNMTF [19], DFMF [26], MFLDA [6], and WMFLDA [24].

2 Methodology

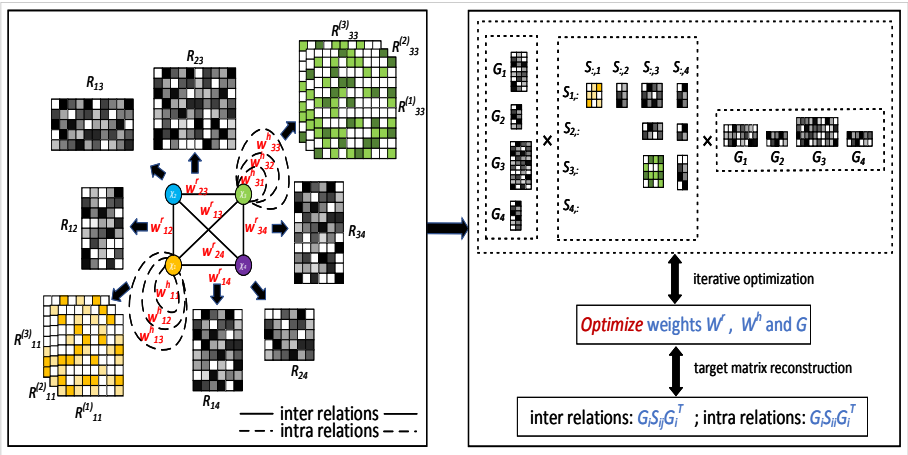


Fig. 1. The operating principle of SelMFDF. In the left figure, $\mathbf{R}_{ij}$ is the inter-relational data matrix between object type i and j, $\mathbf{R}_{ii}^{(v)}$ is the v-th intra-relational matrix of the i-th object type; in the right figure, $\mathbf{G}_i$ is the low-rank representation matrix of the i-th object type. $\mathbf{W}_{ij}^r$ and $\mathbf{W}_{iv}^h$ are the weights assigned to respective inter-relational and intra-relational data matrices.

The operating principle of SelMFDF is illustrated in Figure 1. SelMFDF presets weights for inter-relational and intra-relational data matrices, and performs collaborative low-rank matrix factorization. It then jointly optimizes the weights and the low-rank matrix approximations of these relational matrices. After that, it reconstructs the target relational data matrix based on the product of optimized low-rank matrices.

2.1 Matrix Factorization Model for Multiple Relational Data

The relationships between multi-type objects can be divided into inter-relations and intra-relations, both of which can be encoded by relational data matrices. To fuse these relational data sources, various solutions follow different principles to transform these data matrices toward the target relational matrix using the inter-relations between objects [7,22]. However, this transformation often overrides or even distorts the intrinsic structures of multi-relational data. To avoid this issue,

Zitnik and Zupan [26] introduced a penalized matrix factorization based data fusion framework (DFMF). The objective function of DFMF is:

$$\begin{aligned}\min_{G\geq 0}\mathcal{L}(\mathbf{G},\mathbf{S}) = &\sum_{\mathbf{R}_{ij}\in\mathcal{R}}\left\|\mathbf{R}_{ij}-\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right\|_F^2 \\ &+\sum_{i=1}^{m}\sum_{t=1}^{\max_i t_i} tr(\mathbf{G}^T\mathbf{\Theta}_i^{(t)}\mathbf{G})\end{aligned} \tag{1}$$

where $\|\cdot\|_F^2$ and $tr(\cdot)$ are the Frobenius norm of a matrix and the matrix trace operator. DFMF simultaneously considers m object types and fuses a collection of relational data sources ($\mathcal{R}$). The inter relations between n_i objects of type i and n_j objects of type j are stored in $\mathbf{R}_{ij} \in \mathbb{R}^{n_i\times n_j}$, $\mathbf{G}_i \in \mathbb{R}^{n_i\times k_i}$ is the low-rank representation of object type i, $\mathbf{S}_{ij} \in \mathbb{R}^{k_i\times k_j}$ encodes the latent relationship between $\mathbf{G}_i$ and $\mathbf{G}_j$, $k_i \ll n_i$ is the low-rank size of the respective object type, $\mathbf{G} = diag(\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_m)$. Without loss of generality, suppose the i-th object type has t_i intra relational data matrices and $\mathbf{\Theta}_i^{(t)}$ is the t-th one. $\mathbf{\Theta}^{(t)}$ collectively contains all the following block diagonal matrices: $\mathbf{\Theta}^{(t)} = diag(\mathbf{\Theta}_1^{(t)}, \mathbf{\Theta}_2^{(t)}, \cdots, \mathbf{\Theta}_m^{(t)})$, $t \in \{1, 2, \cdots, \max_i t_i\}$, and the i-th block matrix along the main diagonal of $\mathbf{\Theta}^{(t)}$ is zero if $t \geq t_i$.

Eq. (1) can respect and explore the intrinsic structure of multiple relational data matrices, since it does not project these matrices onto the same space for fusion. However, it *equally* treats all the relational matrices and ignores the different relevances of them toward the target task. As a result, its performance may be dragged down by noisy or irrelevant data sources. To address this issue, Fu *et al.* [6] extended DFMF by optimizing the weights assigned to inter-relational data matrices. However, it still can not differentiate noisy intra-relational matrices during the fusing process. Given that, Yu *et al.* [24] further specified weights to different intra-relational matrices. The theoretical analysis and experimental results show that these two extensions can indeed selectively fuse multiple relational data sources. However, they are inclined to select sparse ones with more zero elements, since the sparse data matrices generally have a smaller approximate loss ($\left\|\mathbf{R}_{ij}-\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right\|_F^2$) or smoothness loss ($tr(\mathbf{G}_i^T\mathbf{\Theta}_i^{(t)}\mathbf{G}_i)$). In practice, a too sparse data matrix often cannot encode sufficient information for the target task, and thus is irrelevant for the task.

2.2 Objective Function of SelMFDF

Based on the above analysis, to reduce the impact of noisy data sources and to avoid inclined to sparse ones, we define the objective function of SelMFDF as follows:

$$\begin{aligned}\min_{\mathbf{G}\geq 0}\mathcal{L}(\mathbf{G},\mathbf{S},\mathbf{W}^r,\mathbf{W}^h) = &\sum_{R_{ij}\in\mathcal{R}}\mathbf{W}^r_{ij}\left\|\mathbf{R}_{ij}-\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right\|_F^2\\ &+\sum_{i=1}^{m}\sum_{t=1}^{\tau}\mathbf{W}^h_{it}\left\|\mathbf{R}^{(t)}_{ii}-\mathbf{G}_i\mathbf{S}_{ii}\mathbf{G}_i^T\right\|_F^2\\ s.t.\quad &\mathbf{W}^r\geq 0,\mathbf{W}^h\geq 0\end{aligned}\tag{2}$$

where $\mathbf{W}^r \in \mathbb{R}^{m\times m}$, $\mathbf{W}^h \in \mathbb{R}^{m\times\tau}$, $\tau = \max_i t_i$, $\mathbf{W}^r$ contains the weights assigned to different inter-relational data matrices, if $\mathbf{R}_{ij} \notin \mathcal{R}$, $\mathbf{W}^r_{ij} = 0$. $\mathbf{W}^h$ contains the weights assigned to different intra-relational data matrices. If $\mathbf{R}^{(t)}_{ii} \notin \mathcal{R}$ or $t > t_i$, $\mathbf{W}^h_{it} = 0$. Unlike Eq. (1), our objective function utilizes the shared low-rank matrices $\mathbf{G}_i$ and $\mathbf{S}_{ii} \in \mathbb{R}^{k_i\times k_i}$ across t_i intra-relational data matrices to approximate $\mathbf{R}^{(t)}_{ii}$. In this way, a data matrix inconsistent with other intra-relational data matrices of the same objects will be assigned with a lower weight. Particularly, for a sparse data matrix $\mathbf{R}^{(t)}_{ii}$, $\left\|\mathbf{R}^{(t)}_{ii}-\mathbf{G}_i\mathbf{S}_{ii}\mathbf{G}_i^T\right\|_F^2$ results in a large loss, because $\mathbf{R}^{(t)}_{ii}$ encodes much fewer relations between objects than its cousin matrices ($\{\mathbf{R}^{(t')}_{ii}\}_{t'=1}^{t_i}, t'\neq t\}$) and the loss is dominated by $tr(\mathbf{G}_i\mathbf{S}_{ii}\mathbf{G}_i^T)$. Similarly, for a dense matrix with noisy entries, $\left\|\mathbf{R}^{(t)}_{ii}-\mathbf{G}_i\mathbf{S}_{ii}\mathbf{G}_i^T\right\|_F^2$ also results in a big loss. To minimize the above objective function, a smaller weight will be automatically assigned to these two types of data matrices. Since $\mathbf{G}_i$ is also shared by the inter-relational data matrices, the first term in Eq. (2) can also avoid preferring to sparse ones. As a result, Eq. (2) can avoid the preference toward the sparse data matrices. We want to remark that low-rank matrix approximation can also reduce the impact of noises to some extent [4, 16].

However, Eq. (2) may only set $\mathbf{W}^r_{ij} = 1$ to $\mathbf{R}_{ij}$ if $\mathbf{R}_{ij}$ has the smallest approximation loss ($\|\mathbf{R}_{ij}-\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j\|_F^2$) among all the inter-relational matrices, and the other inter-relational ones will be discarded. Eq. (2) may also assign $\mathbf{W}^h_{it} = 1$ to $\mathbf{R}^{(t)}_{ii}$, if $\mathbf{R}^{(t)}_{ii}$ has the smallest approximation loss among all the intra-relational matrices. As a result, the contribution of other intra-relational ones will be disregarded. To remedy this issue, we add two l_2-norm based regularizations on $\mathbf{W}^r$ and $\mathbf{W}^h$, and update the objective function as follows:

$$\begin{aligned}\min_{G\geq 0}\mathcal{L}(\mathbf{G},\mathbf{S},\mathbf{W}^r,\mathbf{W}^h) = &\sum_{R_{ij}\in\mathbb{R}}\mathbf{W}^r_{ij}\left\|\mathbf{R}_{ij}-\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right\|_F^2\\ &+\sum_{i=1}^{m}\sum_{t=1}^{\tau}\mathbf{W}^h_{it}\left\|\mathbf{R}^{(t)}_{ii}-\mathbf{G}_i\mathbf{S}_{ii}\mathbf{G}_i^T\right\|_F^2\\ &+\alpha\left\|vec(\mathbf{W}^r)\right\|_F^2+\beta\left\|vec(\mathbf{W}^h)\right\|_F^2\\ s.t.\quad &\mathbf{W}^r\geq 0,\mathbf{W}^h\geq 0,\sum vec(\mathbf{W}^r)=1,\sum vec(\mathbf{W}^h)=1\end{aligned}\tag{3}$$

where $vec(\mathbf{W}^r)$ and $vec(\mathbf{W}^h)$ are the vectorization operator that stacks the rows of $\mathbf{W}^r$ and $\mathbf{W}^h$, $\alpha > 0$ and $\beta > 0$ are used to control the complexity of $vec(\mathbf{W}^r)$ and $vec(\mathbf{W}^h)$. By adding these two regularization terms, SelMFDF can selectively integrate several relevant data matrices, and automatically remove irrelevant ones. Our following optimization procedure for $\mathbf{W}^h$ and $\mathbf{W}^r$ will theoretically confirm this advantage.

$\tilde{\mathbf{G}}$ can be viewed as the optimized low-rank matrices of these object types, we can approximate the target inter-relational data matrix between object type i and j as Eq. (4). Similarly, we can also approximate the intra-relational data matrix as Eq. (5).

$$\widehat{\mathbf{R}}_{ij} = \tilde{\mathbf{G}}_i \tilde{\mathbf{S}}_{ij} \tilde{\mathbf{G}}_j^T \tag{4}$$

$$\widehat{\mathbf{R}}_{ii} = \tilde{\mathbf{G}}_i \tilde{\mathbf{S}}_{ii} \tilde{\mathbf{G}}_i^T \tag{5}$$

In this way, SelMFDF can not only predict the inter-relations between different types of objects, but also the intra-relations between objects of the same type.

2.3 Optimization of SelMFDF

The optimization problem in Eq. (3) is non-convex with respect to $\mathbf{G}$, $\mathbf{S}$, $\mathbf{W}^r$ and $\mathbf{W}^h$ simultaneously. It is difficult to seek the global optimal solutions for all the variables at the same time. Here, we follow the idea of alternating direction method of multipliers (ADMM) [2] and DFMF [26] to alternatively optimize one variable by fixing other three of these four variables in an iterative way.

To account for $\mathbf{G}_i \geq 0$, we import the Lagrangian multipliers $\{\lambda_i\}_{i=1}^m$ and reformulate Eq. (3) as follows:

$$\begin{aligned}
&\min_{\mathbf{G}\geq 0} \tilde{\mathcal{L}}(\mathbf{G}, \mathbf{S}, \mathbf{W}^r, \mathbf{W}^h, \lambda) = \\
&\sum_{R_{ij}\in\mathcal{R}} \mathbf{W}_{ij}^r tr(\mathbf{R}_{ij}^T \mathbf{R}_{ij} - 2\mathbf{R}_{ij}^T \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T + \mathbf{G}_j^T \mathbf{G}_j \mathbf{S}_{ij}^T \mathbf{G}_i^T \mathbf{G}_i \mathbf{S}_{ij}) \\
&+ \sum_{i=1}^{m} \sum_{t=1}^{\tau} \mathbf{W}_{it}^h tr({\mathbf{R}_{ii}^{(t)}}^T \mathbf{R}_{ii}^{(t)} - 2{\mathbf{R}_{ii}^{(t)}}^T \mathbf{G}_i \mathbf{S}_{ii} \mathbf{G}_i^T + \mathbf{G}_i^T \mathbf{G}_i \mathbf{S}_{ii}^T \mathbf{G}_i^T \mathbf{G}_i \mathbf{S}_{ii}) \\
&+ \alpha \left\| vec(\mathbf{W}^r) \right\|_F^2 + \beta \left\| vec(\mathbf{W}^h) \right\|_F^2 - \sum_{i=1}^{m} tr(\lambda_i \mathbf{G}_i^T) \\
&s.t. \quad \mathbf{W}^r \geq 0, \mathbf{W}^h \geq 0, \sum vec(\mathbf{W}^r) = 1, \sum vec(\mathbf{W}^h) = 1
\end{aligned} \tag{6}$$

Next, we goto the alternative optimization procedure.

Optimizing $\mathbf{S}_{ij}$: Suppose $\mathbf{G}$, $\mathbf{W}^r$ and $\mathbf{W}^h$ are known and fixed, and let the partial derivative of Eq. (6) with respect to $\mathbf{S}_{ij}$ and $\mathbf{S}_{ii}$ equal to 0, we can obtain the explicit solution of $\mathbf{S}_{ij}$ and $\mathbf{S}_{ii}$ as follows:

$$\mathbf{S}_{ij} = (\mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{R}_{ij} \mathbf{G}_j (\mathbf{G}_j^T \mathbf{G}_j)^{-1} \tag{7}$$

$$\mathbf{S}_{ii} = (\mathbf{G}_i^T \mathbf{G}_i)^{-1} \frac{\sum_{t=1}^{\tau} \mathbf{W}_{it}^h (\mathbf{G}_i^T \mathbf{R}_{ii}^{(t)} \mathbf{G}_i)}{\sum_{t=1}^{\tau} \mathbf{W}_{it}^h} (\mathbf{G}_i^T \mathbf{G}_i)^{-1} \tag{8}$$

Optimizing $\mathbf{G}_i$: Similar as the optimization of $\mathbf{S}$, we also take the partial derivative of Eq. (6) with respect to $\mathbf{G}_i$ with known $\mathbf{S}$, $\mathbf{W}^r$ and $\mathbf{W}^h$:

$$\begin{aligned}\frac{\partial \tilde{\mathcal{L}}}{\mathbf{G}_i} = &\sum_{j:\mathbf{R}_{ij}\in\mathcal{R}} \mathbf{W}^r_{ij}(-2\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}^T_{ij} + 2\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}^T_j\mathbf{G}_j\mathbf{S}^T_{ij}) \\ &+ \sum_{j:\mathbf{R}_{ji}\in\mathcal{R}} \mathbf{W}^r_{ji}(-2\mathbf{R}^T_{ji}\mathbf{G}_j\mathbf{S}_{ji} + 2\mathbf{G}_i\mathbf{S}^T_{ji}\mathbf{G}^T_j\mathbf{G}_j\mathbf{S}_{ji}) \\ &+ \sum_{t=1}^{\tau} \mathbf{W}^h_{it}2\mathbf{R}^{(t)}_{ii}\mathbf{G}_i - \lambda_i\end{aligned} \tag{9}$$

Multipliers λ_i can be obtained from Eq. (9) by letting $\frac{\partial \tilde{\mathcal{L}}}{\mathbf{G}_i} = 0$ and the KKT (Karush-Kuhn-Tucker) complementary condition [2] for nonnegativity of $\mathbf{G}_i$ as:

$$0 = \lambda_i \circ \mathbf{G}_i \tag{10}$$

where $\circ$ denotes the Hadamard product. Eq. (10) is a fixed point equation and the solution must satisfy it at convergence. Thus, we can obtain:

For $\mathbf{R}_{ij} \in \mathcal{R}$:

$$\begin{aligned}\mathbf{G}^{(e)}_i &+= \mathbf{W}^r_{ij}(\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}^T_{ij})^+ + \mathbf{W}^r_{ij}\mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}^T_j\mathbf{G}_j\mathbf{S}^T_{ij})^- \\ \mathbf{G}^{(d)}_i &+= \mathbf{W}^r_{ij}(\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}^T_{ij})^- + \mathbf{W}^r_{ij}\mathbf{G}_i(\mathbf{S}_{ij}\mathbf{G}^T_j\mathbf{G}_j\mathbf{S}^T_{ij})^+ \\ \mathbf{G}^{(e)}_j &+= \mathbf{W}^r_{ij}(\mathbf{R}^T_{ij}\mathbf{G}_i\mathbf{S}_{ij})^+ + \mathbf{W}^r_{ij}\mathbf{G}_j(\mathbf{S}^T_{ij}\mathbf{G}^T_i\mathbf{G}_i\mathbf{S}_{ij})^- \\ \mathbf{G}^{(d)}_j &+= \mathbf{W}^r_{ij}(\mathbf{R}^T_{ij}\mathbf{G}_i\mathbf{S}_{ij})^- + \mathbf{W}^r_{ij}\mathbf{G}_j(\mathbf{S}^T_{ij}\mathbf{G}^T_i\mathbf{G}_i\mathbf{S}_{ij})^+\end{aligned} \tag{11}$$

For $t = 1, 2, \ldots, \tau$:

$$\begin{aligned}\mathbf{G}^{(e)}_i &+= 2\mathbf{W}^h_{it}(\mathbf{R}^{(t)}_{ii}\mathbf{G}_i\mathbf{S}^T_{ii})^+ + 2\mathbf{W}^h_{it}(\mathbf{G}_i\mathbf{S}_{ii}\mathbf{G}^T_i\mathbf{G}_i\mathbf{S}^T_{ii})^- \\ \mathbf{G}^{(d)}_i &+= 2\mathbf{W}^h_{it}(\mathbf{R}^{(t)}_{ii}\mathbf{G}_i\mathbf{S}^T_{ii})^- + 2\mathbf{W}^h_{it}(\mathbf{G}_i\mathbf{S}_{ii}\mathbf{G}^T_i\mathbf{G}_i\mathbf{S}^T_{ii})^+\end{aligned} \tag{12}$$

where the matrices with positive and negative symbols are defined as $\mathbf{A}^+ = \frac{|\mathbf{A}|+\mathbf{A}}{2}$ and $\mathbf{A}^- = \frac{|\mathbf{A}|-\mathbf{A}}{2}$, respectively. Then we can construct $\mathbf{G}$ as:

$$\mathbf{G} \leftarrow \mathbf{G} \circ diag(\sqrt{\frac{\mathbf{G}^{(e)}_1}{\mathbf{G}^{(d)}_1}}, \sqrt{\frac{\mathbf{G}^{(e)}_2}{\mathbf{G}^{(d)}_2}}, \ldots, \sqrt{\frac{\mathbf{G}^{(e)}_m}{\mathbf{G}^{(d)}_m}}) \tag{13}$$

Optimizing $\mathbf{W}^r$: After updating $\mathbf{S}$ and $\mathbf{G}$, we view them as known and take the partial derivative of Eq. (6) with respect to $\mathbf{W}^r$. In this case, the second, fourth and fifth terms on the right of Eq. (6) are irrelevant to $\mathbf{W}^r$. Then we have:

$$\begin{aligned}\widetilde{\mathcal{L}}(\mathbf{G}, \mathbf{S}, \mathbf{W}^r) = &\sum_{\mathbf{R}_{ij}\in\mathcal{R}} \mathbf{W}^r_{ij} \left\|\mathbf{R}_{ij} - \mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}^T_j\right\|^2_F \\ &+ \alpha \left\|vec(\mathbf{W}^r)\right\|^2_F \\ s.t. \quad &\mathbf{W}^r_{ij} \geq 0, \sum vec(\mathbf{W}^r) = 1\end{aligned} \tag{14}$$

Let $\mathbf{L}_{ij} = \left\|\mathbf{R}_{ij} - \mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right\|_F^2$ be the reconstruction loss for $\mathbf{R}_{ij}$, then Eq. (14) can be updated as:

$$\begin{aligned}\widetilde{\mathcal{L}}(\mathbf{L}, \mathbf{W}^r, \delta, \gamma) = vec(\mathbf{W}^r)^T vec(\mathbf{L}) + \alpha vec(\mathbf{W}^r)^T vec(\mathbf{W}^r) \\ - \sum_{i,j=1}^{m} \delta_{ij}\mathbf{W}_{ij}^r - \gamma(\sum_{i,j=1}^{m} \mathbf{W}_{ij}^r - 1)\end{aligned} \tag{15}$$

Eq. (15) is a quadratic optimization problem with respect to $vec(\mathbf{W}^r)$ and the Lagrangian multipliers (δ and γ) are the two constraints of $\mathbf{W}^r$.

Base on the KKT conditions, the optional $\mathbf{W}^r$ should satisfy the following four conditions:

(i) Stationary condition: $\frac{\partial \widetilde{\mathcal{L}}}{\partial \mathbf{W}^r} = \mathbf{L} + 2\alpha\mathbf{W}^r - \delta - \gamma = 0$
(ii) Feasible condition: $\mathbf{W}_{ij}^r \geq 0, \sum_{i,j=1}^{m} \mathbf{W}_{ij}^r - 1 = 0$
(iii) Dual feasibility: $\delta_{ij} \geq 0, \forall \mathbf{R}_{ij} \in \mathcal{R}$
(iv) Complementary slackness: $\delta_{ij}\mathbf{W}_{ij}^r = 0, \forall \mathbf{R}_{ij} \in \mathcal{R}$

From the stationary condition, $\mathbf{W}_{ij}^r$ can be computed as follows:

$$\mathbf{W}_{ij}^r = \frac{\delta_{ij} + \gamma - \mathbf{L}_{ij}}{2\alpha} \tag{16}$$

We can find that $\mathbf{W}_{ij}^r$ depends on the specification of δ_{ij} and γ, and the specification of δ_{ij} and γ can be analyzed in the following three cases:

(i) If $\gamma > \mathbf{L}_{ij}$, then $\mathbf{W}_{ij}^r > 0$, because of the complementary slackness $\delta_{ij}\mathbf{W}_{ij}^r = 0, \delta_{ij} = 0$ and $\mathbf{W}_{ij}^r = \frac{\gamma - \mathbf{L}_{ij}}{2\alpha}$
(ii) If $\gamma = \mathbf{L}_{ij}$, because of $\delta_{ij}\mathbf{W}_{ij}^r = 0$ and $\mathbf{W}_{ij}^r = \frac{\delta_{ij}}{2\alpha}$, then $\delta_{ij} = 0$ and $\mathbf{W}_{ij}^r = 0$
(iii) If $\gamma < \mathbf{L}_{ij}$, since $\mathbf{W}_{ij}^r \geq 0$, it requires $\delta_{ij} > 0$; because $\delta_{ij}\mathbf{W}_{ij}^r = 0$, then $\mathbf{W}_{ij}^r = 0$

From the above analysis, we can set $\mathbf{W}_{ij}^r$ as:

$$\mathbf{W}_{ij}^r = \begin{cases} \frac{\gamma - \mathbf{L}_{ij}}{2\alpha} & \text{if } \gamma > \mathbf{L}_{ij} \text{ and } \mathbf{R}_{ij} \in \mathcal{R} \\ 0 & \text{if } \gamma \leq \mathbf{L}_{ij} \text{ or } \mathbf{R}_{ij} \notin \mathcal{R} \end{cases}, \tag{17}$$

Let $\mathbf{v}_L \in \mathbb{R}^{|\mathcal{R}|}$ store the entries of vector $vec(\mathbf{L})$ in ascending order with entries corresponding to $\mathbf{R}_{ij} \notin \mathcal{R}$ removed. Accordingly, $\mathbf{v}^r \in \mathbb{R}^{|\mathcal{R}|}$ stores the corresponding entries of $vec(\mathbf{W}^r)$. For a not too big predefined α, there exists $p \in \{1, 2, \ldots, |\mathcal{R}|\}$ with $\mathbf{v}_L(p) < \gamma$ and $\mathbf{v}_L(p+1) \geq \gamma$, satisfying $\sum \mathbf{v}_L = \sum_{\mathbf{v}_L(p) < \gamma} \frac{\gamma - \mathbf{v}_L(p)}{2\alpha} = 1$. Then $\mathbf{v}^r(p')$ has the following explicit solution:

$$\mathbf{v}^r(p') = \begin{cases} \frac{\gamma - \mathbf{v}_L(p')}{2\alpha} & \text{if } p' \leq p \\ 0 & \text{if } p' > p \end{cases}, \tag{18}$$

From $\sum_{p'=1}^{|\mathcal{R}|} \mathbf{v}^r(p') = \sum_{p'=1}^{p} \frac{\gamma - \mathbf{v}_L(p')}{2\alpha} = 1$, we can get the value for γ as:

$$\gamma = \frac{2\alpha + \sum_{p'=1}^{p} \mathbf{v}_L(p')}{p} \tag{19}$$

To search the optimal p, we initialize $p = |\mathcal{R}|$ and decrease it step by step. In each step, we repeatedly refer to Eqs. (18-19) and stop the search once a feasible p is obtained. From Eq. (19), we can observe that for a nonnegative γ, at least one inter-relational data matrix can be selected.

Optimizing $\mathbf{W}^h$: When $\mathbf{G}$, $\mathbf{S}$ and $\mathbf{W}^r$ are fixed, the first, the third and fifth terms on the right of Eq. 3 are irrelevant to $\mathbf{W}^h$, and can be ignored. Then we can follow the similar procedure as that of $\mathbf{W}^r$ to obtain the explicit solution of $\mathbf{W}^h$:

$$\mathbf{W}^h_{it} = \begin{cases} \frac{\mu - \mathbf{O}_i^{(t)}}{2\beta} & \text{if } \mu > \mathbf{O}_i^{(t)} \text{ and } t \leq \max_i t_i \\ 0 & \text{if } \mu \leq \mathbf{O}_i^{(t)} \text{ and } t > \max_i t_i \end{cases}, \tag{20}$$

where $\mathbf{O}_i^{(t)} = \left\| \mathbf{R}_{ii}^{(t)} - \mathbf{G}_i \mathbf{S}_{ii} \mathbf{G}_i^T \right\|_F^2$, $\mu = \frac{2\beta + \sum_{h'=1}^{h} \mathbf{v}_O(h')}{h}$, $\mathbf{v}_O$ stores the entries of vector $vec(\mathbf{O})$ in ascending order with entries corresponding to $\{\mathbf{R}_{ii}^{(t)}\}_{t=1}^{t_i}$ $(i = 1, 2, \cdots, m)$, and h can also be sought in the similar way as p in Eq. (18). We can see from Eq. (20) that if $\mathbf{O}_i^{(t)}$ is larger, $\mathbf{W}^h_{it}$ will be smaller. Once $\mathbf{G}_i \mathbf{S}_{ii} \mathbf{G}_i^T$ is a fixed appropriation, a sparser (or denser) $\mathbf{R}_{ii}^{(t)}$ causes a larger reconstruction loss ($\mathbf{O}_i^{(t)}$). As a result, the explicit solution of $\mathbf{W}^h$ can also avoid the preference toward the 'sparse' data matrices.

3 Experiments

3.1 Experimental setup

To investigate the effectiveness of SelMFDF, we apply it for inter-relation and intra-relation prediction tasks. The inter-relation prediction tasks include lncRNA-miRNA associations and Gene Ontology (GO) annotations of genes, where the target relational matrix is a binary matrix, representing associations between lncRNAs and miRNAs or between genes and GO terms (labels). The intra-relation prediction task is to predict protein-protein interactions by reconstructing the target adjacent matrix of proteins. We collect five object types: lncRNA, genes, miRNA, diseases and Gene Ontology, and adopt eight inter-relational data sources and twelve intra-relational data sources between these objects for experiments. The details of these sources are provided in Table 1.

To comparatively study the performance of SelMFDF, we compare it against five matrix factorization based data fusion methods, including NMTF [5], S-NMTF [19], DFMF [26], MFLDA [6] and WMFLDA [24]. The first three comparing methods equally treat inter-relational matrices or intra-relational matrices

Table 1. Details on the collected inter-relations and intra-relations from different data sources

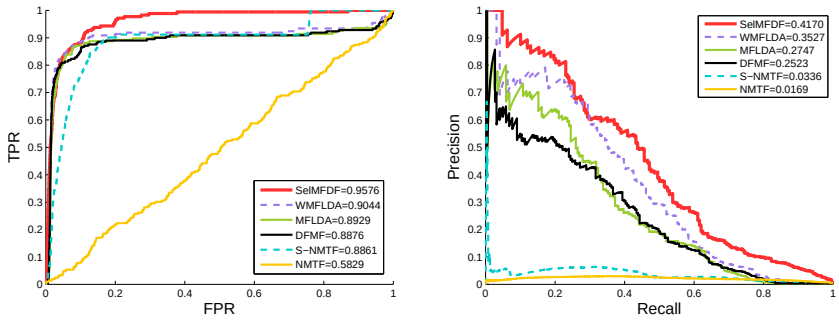
Datasets	Size	#Associations		Sources
LncRNA-Gene	240 × 15527	6186	$\mathbf{R}_{12}$	http://www.lncrna2target.org/
LncRNA-miRNA	240 × 495	1002	$\mathbf{R}_{13}$	http://starbase.sysu.edu.cn/mirLncRNA.php/
LncRNA-Disease	240 × 412	2697	$\mathbf{R}_{14}$	http://www.cuilab.cn/lncrnadisease/
LncRNA-GO	240 × 6428	3094	$\mathbf{R}_{15}$	ftp://ftp.ncbi.nih.gov/gene/GeneRIF/
Gene-Disease	15527 × 412	115317	$\mathbf{R}_{24}$	http://www.disgenet.org/
Gene-GO	15527 × 6428	1191503	$\mathbf{R}_{25}$	http://geneontology.org/
miRNA-Gene	495 × 15527	135852	$\mathbf{R}_{32}$	http://mirtarbase.mbc.nctu.edu.tw/
miRNA-Disease	495 × 412	13562	$\mathbf{R}_{34}$	http://www.cuilab.cn/hmdd/
Gene-Gene	2719 × 2719	4551	$\mathbf{R}_{22}^{(1)}$	http://dip.doe-mbi.ucla.edu/dip/Main.cgi
	7898 × 7898	32097	$\mathbf{R}_{22}^{(2)}$	http://hprd.org/index_html
	13106 × 13106	283306	$\mathbf{R}_{22}^{(3)}$	http://ophid.utoronto.ca/ophidv2.204/index.jsp
	11778 × 11778	113973	$\mathbf{R}_{22}^{(4)}$	http://www.ebi.ac.uk/intact/
	7898 × 7898	32097	$\mathbf{R}_{22}^{(5)}$	http://mint.bio.uniroma2.it/
	13086 × 13086	223546	$R_{22}^{(6)}$	http://thebiogrid.org/
miRNA-miRNA	239 × 239	57121	$\mathbf{R}_{33}^{(1)}$	http://doi.org/10.1186/1471-2164-8-166
	443 × 443	196249	$\mathbf{R}_{33}^{(2)}$	http://doi.org/10.1093/bioinformatics/btx019
	495 × 495	225645	$\mathbf{R}_{33}^{(3)}$	http://www.cuilab.cn/hmdd/
	495 × 495	202833	$\mathbf{R}_{33}^{(4)}$	http://mirtarbase.mbc.nctu.edu.tw/
	495 × 495	42723	$\mathbf{R}_{33}^{(5)}$	http://starbase.sysu.edu.cn/mirLncRNA.php/
	22 × 22	32	$\mathbf{R}_{33}^{(6)}$	http://doi.org/10.1016/j.gene.2012.09.066

during the fusion process. MFLDA optimizes weights to different inter-relational ones and WMFLDA further assigns weights to different intra-relational ones. The input parameters of these methods are set as specified by the authors in the code, or optimized in the suggested ranges. We use the area under the receiver operating characteristic curve (AUROC) and the area under the precision recall curve (AUPRC) to quantify the overall performance. We run five fold cross validation for ten independent rounds, and report the average results.

3.2 Results of inter-relation prediction tasks

For this investigation, we randomly divide the original lncRNA-miRNA associations ($\mathbf{R}_{13}$) into five folds for cross validation. Next, we plot the ROC curves of the comparing methods and report their corresponding AUROCs in Figure 2(a). We can see that SelMFDF always has the highest TPRs (true positive rates) under the same FPRs (false positive rates), and achieves the highest AUROC among these methods. Figure 2(b) plots the PR curves and reports the AUPRCs, we can also observe that SelMFDF consistently outperforms these comparing methods.

SelMFDF performs significantly better than WMFLDA and MFLDA, although the latter two also account for the different relevances of multiple relational data matrices. This is because they both use the manifold regularization and approximation loss to determine the relevance of these matrices. As such, they prefer sparse data matrices during the fusion process. However, those sparse matrices may be irrelevant for the target task. SelMFDF does not have such preference, and thus it obtains better results than WMFLDA and MFLDA. The other comparing methods equally treat all the data matrices. As expected, they have much lower AUROC and AUPRC than those of WMFLDA and



(a) ROC curves and AUROCs (b) PR curves and AUPRCs

Fig. 2. Results of lncRNA-miRNA association prediction. (a) ROC curve and AUROCs. (b) PR curve and AUPRCs.

MFLDA, and say nothing of SelMFDF. In practice, S-NMTF costs the largest runtime costs and memory, since it performs matrix factorization on a big adjacency matrix of all objects. NMTF only fuses inter relational data matrix and thus loses to all the comparing methods.

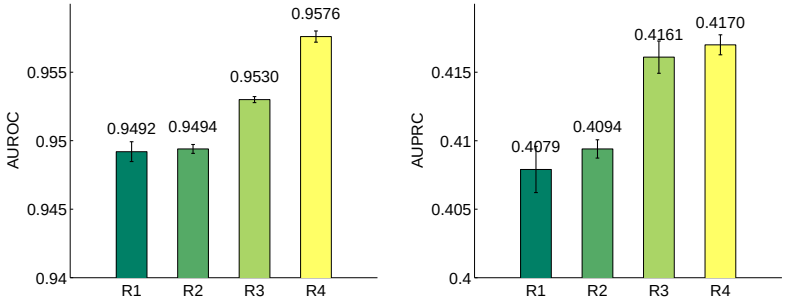


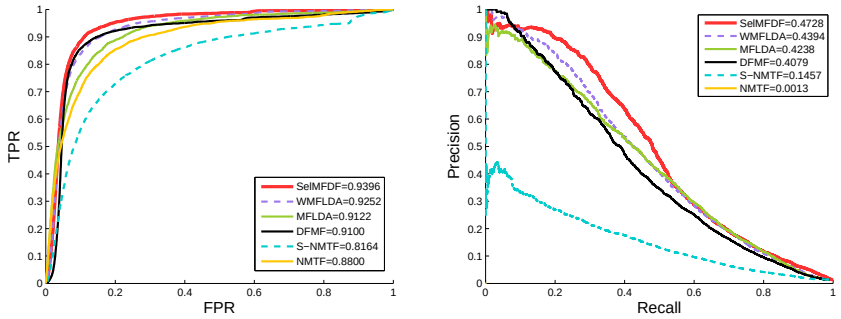
Fig. 3. The AUROCs and AUPRCs of SelMFDF with different collections of intra-relational data matrices. $\mathcal{R}1 = \{\mathbf{R}_{22}^{(1)}, \mathbf{R}_{22}^{(2)}, \mathbf{R}_{22}^{(3)}, \mathbf{R}_{22}^{(4)}, \mathbf{R}_{22}^{(5)}, \mathbf{R}_{22}^{(6)}\}$, $\mathcal{R}2 = \mathcal{R}1 - \mathbf{R}_{22}^{1}$, $\mathcal{R}3 = \mathcal{R}1 - \mathbf{R}_{22}^{3}$, $\mathcal{R}4 = \mathcal{R}1 - \mathbf{R}_{22}^{1} - -\mathbf{R}_{22}^{3}$.

To investigate whether SelMFDF has the capability to identify relevant data matrices and avoid too sparse ones, we report the weights assigned to different intra-relational matrices. The weights assigned to $\mathbf{R}_{22}^{(i)}, (i = 1, \ldots, 6)$ are (0, 0.0946, 0, 0.0537, 0.0946, 0.1084) and the weights assigned to $\mathbf{R}_{33}^{(i)}, (i = 1, \ldots, 6)$ are (0.1033, 0.1572, 0.1133, 0.1092, 0.1568, 0.0089). We can see SelMFDF assigns a zero weight to the sparsest $\mathbf{R}_{22}^{(1)}$ and $\mathbf{R}_{33}^{(6)}$, and it also assigns a zero weight to the densest $\mathbf{R}_{22}^{(3)}$. The sparsity of these data matrices is included in Table 1 (column '#Associations'). These two assignments are expected from Eq. (20) that SelMFDF can avoid preferring to too sparse and too dense data matrices

by crediting lower weights to them. In contrast, these comparing methods either equally integrate them or prefer the sparse ones.

To prove these discarded matrices are indeed irrelevant, we further report the results of SelMFDF by discarding $\mathbf{R}_{22}^{(1)}$ and $\mathbf{R}_{22}^{(3)}$, in Figure 3. SelMFDF obtains the highest AUROC and AUPRC when $\mathbf{R}_{22}^{(1)}$ and $\mathbf{R}_{22}^{(3)}$ are excluded. We also see that $\mathbf{R}_{22}^{(1)}$ has little contribution. This observation confirms the sparse data matrix has a tiny impact on the target prediction task, since it is too sparse to encode sufficient information for the target task. In addition, SelMFDF has an increased performance when $\mathbf{R}_{22}^{(3)}$ is discarded. That is possible because $\mathbf{R}_{22}^{(3)}$ is a dense matrix with many noisy entries.

We further apply these comparing methods to predict GO annotations of genes (the target relational matrix is $\mathbf{R}_{25}$) in five-fold cross validation. The AUROCs and AUPRCs of these comparing methods are revealed in Figure 4. We can clearly see that SelMFDF again performs consistently better than the other five approaches and the results give the similar conclusions as those on predicting lncRNA-miRNA associations.



(a) ROC curves and AUROCs (b) PR curves and AUPRCs

Fig. 4. Results of predicting GO annotations of proteins. (a) ROC curve and AUROCs. (b) PR curve and AUPRCs.

3.3 Results of intra-relation prediction task

To further explore the usage of SelDFMF in predicting intra-relations between the same type of objects, we apply SelMFDF to predict protein-protein interactions. For this study, we pick out the protein-protein interaction matrix collected from BioGrid [3] from the collection of intra relational matrices $\{\mathbf{R}_{22}^{(t)}\}_{t=1}^{6}$, and then use $\mathbf{G}_2\mathbf{S}_{22}\mathbf{G}_2^T$ to approximate the target intra-relational data matrix. Next, we select the top K predicted interact-pairs and check them by referring to available interactions in BioGrid [3]. The number of confirmed interactions under different K is reported in Table 2.

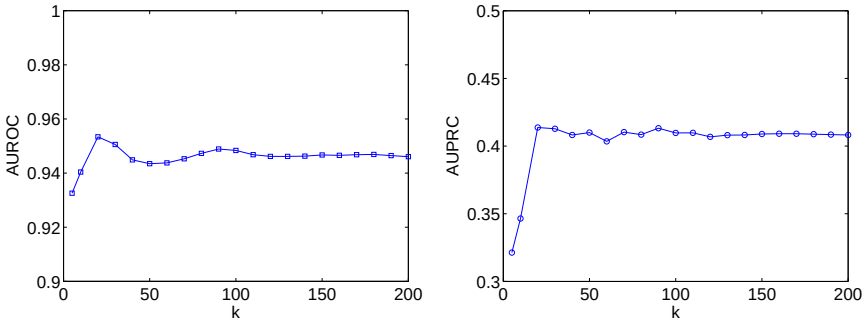
From Table 2, we can clearly see that SelMFDF always more accurately predicts protein-protein interactions than other methods. In addition, from the remaining 15 interactions (not recorded in the BioGrid) in top 20 predicted by

Table 2. Number of confirmed PPIs (from BioGrid) predicted by comparing methods.

Methods	Confirmed Interactions					
	K=20	K=50	K=100	K=500	K=1000	K=10000
SelMFDF	5	9	17	56	118	879
WMFLDA	2	5	10	31	69	521
MFLDA	2	4	13	26	52	511
DFMF	2	4	10	23	43	482
S-NMTF	2	4	4	9	24	140
NMTF	0	0	0	1	1	9

SelMFDF, we further find 6 interactions confirmed by HRPD [11], IntAct [10] and I2D [12] databases. These results indicate SelMFDF can be more reliably applied for the intra-relations prediction.

3.4 Parameter Analysis

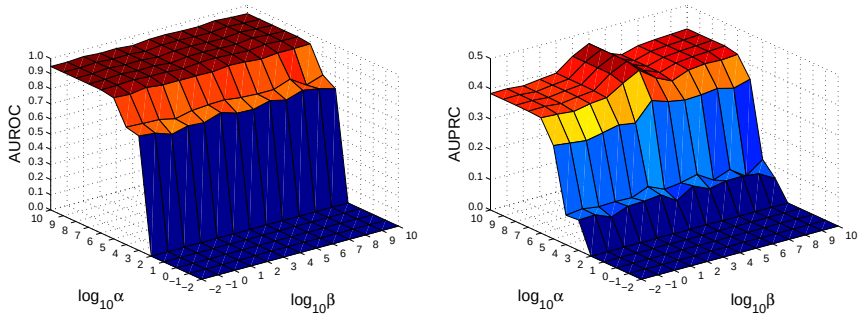


(a) AUROCs (b) AUPRCs

Fig. 5. The AUROC and AUPRC of SelMFDF under different low-rank sizes k.

The low-rank size k_i is an important parameter for low-rank matrix approximation based solutions. To study the sensitivity of k_i, we fix all $k_i = k$ across these five types of objects for simplicity, and then increase k from 10 to 200. Fig. 5 reports the AUROC and AURPC under different input values of k_i in predicting lncRNA-miRNA associations in five-fold cross validation. Both the AUROC value and AUPRC value increase as the increase of k and reach to a highest when $k \approx 20$. Then the AUROC value has a slight decrease and keeps stable after $k > 100$. The AURPC value nearly keeps stable when $k \geq 20$. Given these observations, we adopt $k = 20$ for experiments.

From Eq. (18) and Eq. (20), we can find that once the input value of α or β is specified, the weights $\mathbf{W}^r$ and $\mathbf{W}^h$ assigned to the relational data matrices are also determined. Thus, we further conduct five-fold validation to evaluate the performance of SelMFDF under different combinations of α and β. We vary α and β in $\{10^{-2}, 10^{-1}, \cdots, 10^{10}\}$ and report the average AUROC and AUPRC in Fig. 6. We can clearly see that when $\alpha = 10^7$ and $\beta = 10^4$, SelMFDF achieves the highest AUPRC. The input value of α significantly affects the performance; the AUROC value and AUPRC value increase as α increase, and reach a plateau when $\alpha > 10^7$. The input value of β also affects the performance; the AUPRC value increases as β get larger, and then it slightly decreases when $\beta > 10^4$.



(a) ROC curves and AUROCs (b) PR curves and AUPRCs

Fig. 6. AUROC and AUPRC of SelMFDF under different input values of α and β. (a) AUROCs. (b) AUPRCs.

From these results, we can conclude that SelMFDF can automatically identify irrelevant relational data matrices, and achieve a more prominent performance on predicting the inter- and intra-relations between multiple object types. In addition, it is effective in a wide combination of α and β values, and low-rank sizes.

4 Conclusion

We introduced a selective matrix factorization based solution (SelMFDF) to fuse multi-relational data matrices. Unlike existing matrix factorization based data fusion approaches, SelMFDF can not only selectively integrate multi-relational data matrices, but also avoid preferring to sparse ones and dense ones. Extensive experimental results show that SelMFDF achieves a much better performance than the state-of-the-art solutions in predicting inter-relations and intra-relations between objects. In our future work, we will extend SelMFDF for large scale heterogeneous data fusion. The code and datasets are available at http://mlda.swu.edu.cn/codes.php?name=SelMFDF.

References

1. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. JMLR **7**(11), 2399–2434 (2006)
2. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge Univ. Press (2004)
3. Chatr-Aryamontri, A., Oughtred, R., et al.: The biogrid interaction database: 2017 update. Nucleic Acids Research **45**(D1), D369–D379 (2017)
4. Chen, X., Yu, G., Domeniconi, C., Wang, J., Zhang, Z.: Matrix factorization for identifying noisy labels of multi-label instances. In: PRICAI. pp. 508–517 (2018)
5. Ding, C., Li, T., Peng, W., Park, H.: Orthogonal nonnegative matrix t-factorizations for clustering. In: KDD. pp. 126–135 (2006)
6. Fu, G., Wang, J., Domeniconi, C., Yu, G.: Matrix factorization-based data fusion for the prediction of lncrna–disease associations. Bioinformatics **34**(9), 1529–1537 (2018)

7. Gligorijević, V., Pržulj, N.: Methods for biological data integration: perspectives and challenges. Journal of the Royal Society Interface **12**(112), 20150571 (2015)
8. Gönen, M., Alpaydın, E.: Multiple kernel learning algorithms. JMLR **12**(7), 2211–2268 (2011)
9. Karasuyama, M., Mamitsuka, H.: Multiple graph label propagation by sparse integration. TNNLS **24**(12), 1999–2012 (2013)
10. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., et al.: The intact molecular interaction database in 2012. Nucleic Acids Research **40**(D1), D841–D846 (2011)
11. Keshava Prasad, T., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al.: Human protein reference database2009 update. Nucleic Acids Research **37**(S1), D767–D772 (2008)
12. Kotlyar, M., Pastrello, C., Sheahan, N., Jurisica, I.: Integrated interactions database: tissue-specific view of the human and model organism interactomes. Nucleic Acids Research **44**(D1), D536–D541 (2015)
13. Lanckriet, G.R., De Bie, T., Cristianini, N., Jordan, M.I., Noble, W.S.: A statistical framework for genomic data fusion. Bioinformatics **20**(16), 2626–2635 (2004)
14. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS. pp. 556–562 (2001)
15. Li, Y., Wu, F.X., Ngom, A.: A review on machine learning principles for multi-view biological data integration. Briefings in Bioinformatics **19**(2), 325–340 (2016)
16. Meng, D., De La Torre, F.: Robust matrix factorization with unknown noise. In: ICCV. pp. 1337–1344 (2013)
17. Nielsen, T.D., Jensen, F.V.: Bayesian networks and decision graphs. Springer Science & Business Media (2009)
18. Troyanskaya, O.G., Dolinski, K., Owen, A.B., Altman, R.B., Botstein, D.: A bayesian framework for combining heterogeneous data sources for gene function prediction (in saccharomyces cerevisiae). PNAS **100**(14), 8348–8353 (2003)
19. Wang, H., Huang, H., Ding, C.: Simultaneous clustering of multi-type relational data via symmetric nonnegative matrix tri-factorization. In: CIKM. pp. 279–284 (2011)
20. Wang, M., Hua, X.S., Hong, R., Tang, J., Qi, G.J., Song, Y.: Unified video annotation via multigraph learning. TCSVT **19**(5), 733–746 (2009)
21. Yu, G., Domeniconi, C., Rangwala, H., Zhang, G., Yu, Z.: Transductive multi-label ensemble classification for protein function prediction. In: KDD. pp. 1077–1085 (2012)
22. Yu, G., Fu, G., Lu, C., Ren, Y., Wang, J.: Brwlda: bi-random walks for predicting lncrna-disease associations. Oncotarget **8**(36), 60429 (2017)
23. Yu, G., Rangwala, H., Domeniconi, C., Zhang, G., Zhang, Z.: Predicting protein function using multiple kernels. TCBB **12**(1), 219–233 (2015)
24. Yu, G., Wang, Y., Wang, J., Fu, G., Guo, M., Domeniconi, C.: Weighted matrix factorization based data fusion for predicting lncrna-disease associations. In: BIBM. pp. 1–6 (2018)
25. Yuan, Y., Savage, R.S., Markowetz, F.: Patient-specific data fusion defines prognostic cancer subtypes. PLoS Computational Biology **7**(10), e1002227 (2011)
26. Žitnik, M., Zupan, B.: Data fusion by matrix factorization. TPAMI **37**(1), 41–53 (2015)