SMFM4L: Multi-typed Objects Multi-view Multi-instance Multi-label Learning based on Selective Matrix Factorization*

Yuanlin Yang[1,2,*], Guangyang Han[1], Runming Wang[1], and Saowei Wei[1]

[1] College of Computer and Information Sciences, Southwest University, Chongqing, China
[2] Chonqing University, Chongqing, China
{ylyang,gyhan,rmwang,swwei}@email.swu.edu.cn

Abstract. Multi-typed Objects Multi-view Multi-instance Multi-label Learning (M4L) deals with interlinked multi-typed objects (or bags) that are made of a set of instances, represented with heterogeneous feature views and annotated with a set of non-exclusive multiple labels. M4L is more general and powerful than the typical Multi-view Multi-instance Multi-label Learning (M3L) which lacks the power of jointly modelling the naturally interlinked multi-typed objects in the physical world. However, the current M4L methods equally treat multi-type of objects or prefer sparse ones, which may be irrelevant to the target task. To combat with this more general but challenging learning task, we develop a Selective Matrix Factorization based solution (SMFM4L). Particularly, SMFM4L first collaboratively factorizes multiple inter-relational data matrices into low-rank representation matrices of their respective objects and optimizes their weights. To avoid the interference of sparse data, it then approximates multiple intra-relational data by regularizing these low-rank matrices and also optimize their weights so that both (inter and intra) weights can automatically integrate relevant objects. Next, SMFM4L cooperates an integration item to push the label of bags into the instances and aggregates the label of instances to their hosting bags. Finally, SMFM4L reconstructs the label relation of the bag or instance using the optimized low-rank representation matrix. Experimental results on benchmark datasets show that SMFM4L achieves significantly better results than state-of-the-art methods.

Keywords: Multi-instance learning · Multi-label learning · Multi-view learning · Matrix Factorization.

1 Introduction

With the advancement of technology and the development of the Internet, diverse and large amounts of data (such as images and text) are often tagged with multiple semantic labels and represent multiple views of information that

* Corresponding to Yuanlin Yang

describe the complex object from different perspectives. For example, the paper object in Figure 1 is connected with an author and a publishing conference, and the paper object contains multiple instances (paragraphs and regions) on the text view and image view. Besides, the complex multi-view multi-instance objects (paper) associated with multiple semantic labels (i.e., AI, NLP, and CV). Multi-view Multi-instance Multi-label Learning (M3L) is used to solve this complex problem [1, 2]. The aim of M3L is to learn the relation among bag, instance and label to construct bag-label or instance-label connection.

However, M3L only models the bag of homogeneous network or single typed object, ignoring the fact that the label of a bag is derived not only from the instance features it contains, but also from other interconnected objects [3]. As a result, M3L cannot directly handle multi-type object setting. For the heterogeneous multi-instance multi-lable network scenario in Figure 1, the label of paper is determined not only by the information it contains about the instance features, but also by the potentially interconnected type objects. As M3L cannot be applied to the M4 scenario directly, a possible approach can be used to address this problem, namely by projecting other type object information onto the target object (bag). The projection operation is often used in many applications, such as multi-view learning [4], multi-core learning [5], etc. However, the projection operation often destroys the intrinsic structure and attribute information of the data itself, which can lead to new problems - i.e. information loss. Many Matrix Factorization (MF) based methods are currently being applied to web data with good results. The matrix factorization approach not only preserves network intrinsic structure, but also makes effective use of attribute information [6, 7]. However, it accounts for objects as simple nodes and ignores instance level representation, wihch represents more fine-grained information, thus multi-instance objects cannot be modeled.

Recently, a new learning paradigm called Multi-type objects Multi-view Multi-instance Multi-label Learning (M4L) is used to model heterogeneous multi-instance multi-label networks data [3, 8]. M4L can not only learn network structure information of interconnected multi-type objects, but also extract fine-grained information. M4L-JMF [3] first performs matrix Factorization on the heterogeneous multi-instance multi-label network, then learns the low-rank representation of different type of objects, and reconstructs the relation among bags, instances and lables. Synchronously, the labels of bags are distributed to the included instances through the integration term, which then aggregates the labels of instances to their affiliated bags. However, M4L-JMF tends to select sparser relation matrices and give them large weight, even these sparse objects may be irrelevant (or even harmful) for the target task.

To address these issues, we propose a M4L method (SMFM4L) based on Selective Matrix Factorization to avoid more sparse data during the fusing process. It first learns the low-rank representation of objects through the matrix tri-factorization technology, and then uses optimization algorithms to avoid selecting sparse data. It needs to optimize both the weight matrix between different types of objects, and the weight matrix of the same type of objects. SMFM4L

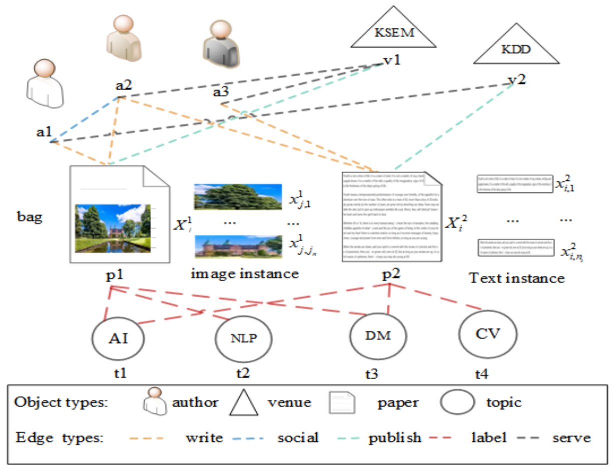

Fig. 1. An illustration of the Multi-typed objects Multi-view Multi-instance Multi-label (M4) data with three types of objects (author, venue and paper). The papers are encoded by the image and text view, and the paper objects (bags) in each view is further made of diverse instances (i.e., image patch x_1^1 in p1 and paragraph x_1^2 in p2), these objects are simultaneously tagged with multiple semantic labels (i.e., AI, NLP,DM and CV). These inter-connected multi-type objects naturally form a Heterogeneous Multi-Instance Network (HMIN).

then must selectively assign weights to objects, which help to identify relevant objects and eliminate irrelevant objects in a heterogeneous multi-instance multi-label network. At last, the bag-label/instance-label relations are reconstructed through the optimized low-rank representation. The main contributions of this work are summarized below:
(i) We explicitly define a Heterogeneous Multi-Instance Network (HMIN) and naturally extend the M3L (Multi-view Multi-instance Multi-label Learning) into the M4L (Multi-typed Objects Multi-view Multi-instance Multi-label Learning), which is more effective and powerful than M3L (only focus on single-typed objects). This heterogeneity in HMIN includes not only multiple types of objects and views, but also multiple types of edges, so HMIN can unify M4 and M3 data more naturally.

(ii) We propose a novel Multi-typed Objects Multi-view Multi-instance Multi-label Learning solution based on Selective Matrix Factorization (SMFM4L), which can not only identify relevent objects and eliminate irrelevant data using automatically optimized weight matrices, but also credit larger weights to the more relevant ones for the prediction of bag/intance label relation.

(iii) Empirical results on multiple real-world datasets show that our proposed SMFM4L outperforms M3L solutions (M3Lcmf [1], M2IL [9] and ICM2L [10]), matrix factorization methods (DFMF [6], WMFLDA [11] and S-NMTF [12]) and M4L method (M4L-JMF [3]) on multiple prediction tasks.

The rest of the paper is organized as follows. Section 2 reviews the related work and Section 3 elaborates on the details of the proposed SMFM4L. The experimental results and analysis are presented in Section 4, and Section 5 concludes the paper.

2 Related work

Our work is closely related with Multi-view Multi-instance Multi-label Learning (M3L) and data fusion based on matrix factorization (MF) methods, and M3L is a more general or powerful framework than its degenerated version (such as Multi-instance Multi-label Learning [13–15], Multi-view Multi-instance Learning [16, 17] and Multi-view Multi-label Learning [18]).

Multi-view Multi-instance Multi-label Learning (M3L for short). These learning frameworks only consider a single type of object, and their degenerated versions ignore some association relation among bag, instance and label. Obviously, M3L is more powerful than other degenerated versions due to the use of the relationship between multiple views. In order to inherit the advantages of M3L for our proposed solution, we briefly introduce some representative M3L based methods. As far as the author knows, Multi-modal Multi-instance Multi-label Latent Dirichlet Allocation (M3LDA) [19] is the first M3L algorithm. This method uses Latent Dirichlet Allocation [20] to explore visual label topics and text label topics from image view and text view, and then let these two views predict bag label based on across view of consistent label.

In order to deal with the situation that some samples (bags) are missing in some views, Nguyen et al. [21] further proposed a Multi-view Multi-instance Multi-label learning method (MIMLmix) based on Hierarchical Bayesian networks, and derived an effective learning algorithm based on variational inference. The model assumes that in a continuous feature space, the label of the bag obeys Gaussian distribution; while in a discrete space, the label of the bag obeys a polynomial distribution, and these labels are distributed under different topics. Different from M3LDA and MIMLmix, Li et al. [16] proposed a Multi-view Multi-instance learning algorithm (M2IL) from the perspective of multi-view dictionary learning. M2IL first uses the sparse ϵ-graph model to construct multiple view information for the same sample, and uses different parameters to control the structural information for the same sample in different views. Then, M2IL integrates these different graphs into a unified framework, and uses sparse representation and a new multi-view dictionary learning to classify samples (bag). In order to learn large-scale Multi-view Multi-instance Multi-label (M3) data more efficiently, Yang et al. [22] proposed a Multi-view Multi-instance Multi-label learning based on Deep neural Networks (M3DN). This method performs deep network learning for each view separately, and requires that the prediction information of the same sample (bag) from different views is consistent. In addition, M3DN also uses the Optimal Transport theory [23] to capture the geometric information of the underlying label space and guide the bag label prediction. Xing et al. [1] proposed a Weakly Supervised Multi-view Multi-instance Multi-label Learning (WSM3L) algorithm based on multi-view dictionary learning to solve the situation where the correspondence between certain objects in different views is unknown or completely unknown. M3DNS [24] explored single-modal instance-level auto-encoders and improved bag-level based on Optimal Transport, then enhance consistency among modalities and predict the labels of bag using both instance-level and bag-level information. These M3L methods have achieved good performance in various domains, but they can only model a *single type* of object, and cannot be directly applied to the interconnected objects of multi-typed in real world.

Some data fusion methods based on matrix factorization can not only handle interconnected multi-type of objects, but also preserve the intrinsic structure among multi-type of objects. These approaches have been widely used in many fields [6, 7, 25]. To name a few, Symmetric Nonnegative Matrix Tri-Factorization (S-NMTF) [12] is based on matrix factorization by clustering multiple types at the same time. Since SNMTF performs matrix factorization on a large matrix, this fast matrix contains the internal relationships of objects and external relations, thus leading to a serious computational burden. Data Fusion Matrix Factorization (DFMF) [6] synergistically decomposes these fast matrices into low-rank matrices, and then reconstructs the target relationship matrix from them, so as to carry out the direct relationship between the target objects. Weighted Matrix Factorization on multi-relational data for LncRNA-Disease Association prediction (WMFLDA) [11] further considers the weights between different types of objects and selectively fuses relational data. However, although these schemes consider the network structure or multiple feature views between different types of objects, due to the lack of instanc-level information modeling capabilities, they cannot model multiple instances in complex objects and aggregate the instance feature onto the bag level. Multi-typed Objects Multi-View Multi-Instance Multi-Label Learning based on Joint Matrix Factorization (M4L-JMF) [3] combines multi-type object data with instance features (The M3L learning framework is extended for the first time to handle multi-type objects scenarios), and then M4L-JMF performs matrix factorization of these inter-association and intra-association data among different types of objects, then reconstructs label of bag or instance association relation based on low-rank matrix. However, because it cannot automatically identify sparse data, and assigns larger weights to sparse data, some noisy data cannot be eliminated well, causing a performance bottleneck.

In order to comprehensively capture the fine-grained multi-instance information and fuse network structure information in the Multi-type Multi-view Multi-instance Multi-label data (M4 data) or Heterogeneous Multi-Instance Network (HMIN), and filter out related objects and eliminate irrelevant objects, we introduce a method based on Selection Matrix Factorization called SMFM4L to model interconnection complex multi-type objects. The next section will introduce this method in detail.

Table 1. The used main notations

Notation	Explanation
$\mathbf{R}_{ij}$	Inter-relational matrices of different types of objects
$\mathbf{\Theta}_i^{(t)}$	Intra-association matrices of the i-th typed objects with the t-th view
$\mathbf{R}_{bi}$	Bag-Instance relational matrix
$\mathbf{R}_{bl}$	Bag-Label relational matrix
$\mathbf{R}_{il}$	Instance-Label relational matrix
$\mathbf{G}_i$	Low-rank representation of objects of the i-th type
$\mathcal{S}_{ij}$	The latent inter-relations between objects of the i-th type and the j-th type
$\mathcal{R}$	The set of inter-relational matrices $\mathbf{R}_{ij}$
$\mathbf{\Theta}$	A set of all the intra-relational matrices $\mathbf{\Theta}_i^{(t)}$
$\mathcal{G}$	A Heterogeneous Multi-instance Network (HMIN)
$\mathbf{W}_{ij}^r$	Weight of the inter-relational matrix of the i-th and i-th typed objects
$\mathbf{W}_{pt}^h$	Weight of the t-th intra-relational matrix of the p-th object type
α	Regularization weight for $\mathbf{W}_{ij}^r$
β	Regularization weight of $\mathbf{W}_{pt}^h$

3 Method

3.1 Problem Statement

A Heterogeneous Multi-instance Network (HMIN) $\mathcal{G} = \{\mathcal{V}, \mathcal{R}, \mathbf{\Theta}\}$ is a special Heterogeneous Information Network (HIN), where $\mathcal{V}$ contains multi-typed objects, $\mathcal{R}$ and $\mathbf{\Theta}$ contain inter-association relation among multi-type of objects and intra-association relation between same type of objects (also as multi-view information), respectively. The difference of HMIN and HIN lies in two aspects: The first is in the prediction task, HMIN focuses on bag/instance-label prediction, while HIN focuses on the node-level tasks (such as node classification and recommendation [26]) or network level tasks (such as clustering and community discovery [27]). The second is that HMIN *must* include the *complex objects (bags)* which are composed of multiple instances, while HIN does not. Suppose there are m types of objects that are directly or indirectly related to each other, and they are coded into the relation of different types of objects matrix $\mathbf{R}_{ij} \in \mathbb{R}^{n_i \times n_j}$, the same type objects relation matrix $\mathbf{\Theta}_i^{(t)} \in \mathbb{R}^{n_i \times n_i}$. The relation data matrix $\mathbf{R}_{ij}$ means that the inter-relation between n_i objects of the i-th type and n_j objects of the j-th type, and the relation data matrix $\mathbf{\Theta}_i^{(t)}$ encodes the intra-relation for the i-th type. The goal of SMFM4L is to learn the mapping function $f(\mathcal{V}, \mathcal{R}, \mathbf{\Theta}) \in \{0, 1\}^q$ to label the target object with respect to q different but related labels. Then we can get the bag-label relation matrix $\mathbf{R}_{bl}$ for N_b objects and N_l labels or instance-label relation matrix $\mathbf{R}_{il}$ for N_i objects and N_l tags. Here $\mathcal{R}$ stores the entire inter-relational data matrices $\mathbf{R}_{ij}$, while $\mathbf{\Theta}$ holds the intra-relational matrix $\mathbf{\Theta}_i$.

To achieve the goal of predicting the correlation label of bags or instances, we first pre-set the weights of the intra-association and inter-association relation matrices, and then synergistically decomposes the intra-association and inter-association relation matrices in $\mathbf{R}$ and $\mathbf{\Theta}$ to obtain the low-rank representation of these types of objects. Then, collaborative optimization is carried out on the weight matrices and these low-rank matrices. After that, We fuse the multi-type object information through these object latent representation matrices, and reconstruct the association matrix $\mathbf{R}_{bl}$ (or $\mathbf{R}_{il}$) between bags (or isntances) and labels from the obtained low-rank matrix representation.

3.2 Selected Matrix Factorization

Fusing other types of object information can improve the representation ability of bags or instances, and achieve the purpose of further improving the prediction of bag-label (or instance-label) relations. Other types of objects can be used to project toward the target type (bag) to form a composite bag-bag or instance-instance relational data matrices, and then known label information can be used to predict label information of unknown bags or instances. Although these projection methods can integrate information of interconnected objects (such as multiple kernel learning, classifier ensemble based data fusion solutions) [5, 28]. However, projection operation may overrides or even distorts the intrinsic structure information of multi-type objects, resulting in information loss and compromise the performance [29, 30].

These relational data types can be further divided into inter-relational matrix and intra-relational matrix. To fuse these multi-type object relations, Data Fusion technology based on Matrix Factorization (DFMF) has been proposed to solve this problem without destroying the intrinsic structure of multi-relation data [6]. The basic framework can be formalized as follows:

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

where $\mathbf{R}_{ij}$ is inter relations between n_i objects of type i and n_j objects of type j, and $\mathbf{G} = diag(\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_m)$, $\mathbf{G}_i \in \mathbb{R}^{n_i\times k_i}$ is the low-rank representation of objects of the i-th type. $\mathbf{S}$ is made of $\mathbf{S}_{ij} \in \mathbb{R}^{k_i\times k_j}$, which encodes the latent relationship between $\mathbf{G}_i$ and $\mathbf{G}_j$, and $k_i \ll n_i, k_j \ll n_j$ are the low-rank size of the respective object type. $\mathbf{\Theta}^{(t)}$ collectively contains all the following block diagonal matrices: $\mathbf{\Theta}^{(t)} = diag(\mathbf{\Theta}_1^{(t)}, \mathbf{\Theta}_2^{(t)}, \cdots, \mathbf{\Theta}_m^{(t)})$ ($t \in \{1, 2, \cdots, max_i t_i\}$), where the i-th block matrix along the main diagonal of $\Theta^{(t)}$ is zero if $t > t_i$, and $\tau = max_i t_i$, $tr(\cdot)$ is the matrix trace operator, $\|\cdot\|_F^2$ is the Frobenius norm of a matrix.

Without loss of generality, the above function aims to reconstruct the incomplete $\mathbf{R}_{bl}$ (or $\mathbf{R}_{il}$) to replenish the relation among bags, instances and labels. Entries in intra-association data matrices within an association are negative for similar objects and positive for dissimilar objects. The negative entry is treated as a *must-link* constraint, which forces object pairs to be close together in the presentation space; The positive entries entries can be seen as a *cannot-link* constraint [31, 32], it forces the different objects in low rank representation space away from each other.

However, Equation (1) not only ignores the important *bag-instance* association, but also fails to distinguish the dependencies of multi-typed objects. A real-world object (bag) may further be composed of several distinct sub-objects (instances), and the label of the bag is determined by the label of its instances [13, 33]. In many practical fields (i.e., medical image analysis and biology), instance-level precise labels are more important and interesting than bag-level labels, which carry more specific knowledge about the regions (i.e. local patches and functional sites) of bag (i.e. images and molecules) [34]. Unfortunately, the labels of the instances are often unknown, while the labels of the bags can be collected easily [35].

Based on above analysis, to leverage the bag-instance association between two types of objects, we define the objective function of SMFM4L as follows:

$$
\begin{aligned}
\min_{\mathbf{G}\geq 0} \Omega(\mathbf{G},\mathbf{S}) = \sum_{i,j=1}^{m} \|\mathbf{R}_{ij} - \mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\|_F^2 + \|\mathbf{R}_{bl} - \mathbf{R}_{bi}\mathbf{G}_i\mathbf{S}_{il}\mathbf{G}_l^T\|_F^2 \\
+ \sum_{k=1}^{m}\sum_{t=1}^{\tau} tr(\mathbf{G}^T\mathbf{\Theta}_i^{(t)}\mathbf{G})
\end{aligned}
\tag{2}
$$

where a dispatch and aggregation term ($\|\mathbf{R}_{bl} - \mathbf{R}_{bi}\mathbf{G}_i\mathbf{S}_{il}\mathbf{G}_l^T\|_F^2$) has been introduced to push the labels of bags into their affiliated instances and reverse aggregate the labels of instances into their hosting bags in Equation (2), and $\mathbf{G}_i\mathbf{S}_{il}\mathbf{G}_l^T$ integrates the labels of instances towards their hosting bags. In this way, Equation (2) can not only fuse multiple types of objects, but also consider complex objects composed of multiple sub-objects (instances) to predict the labels of bags and instances in a consistent manner.

Equation (2) can not only respect the intrinsic structure of multi-type objects, but also model the fine-grained information for bag-instance association. Because this function does not project these multi-type object association information onto the same space for fusion, and introduce a dispatch and aggregation term to predict the labels of bags and instances in a coherent way. However, it equally treats inter-association relation and intra-association relation and ignores the related relevance of them to the target task. As a result, its performance may be degraded due to noisy or irrelevant data. To solve this problem, Fu *et al.* [25] extended DFMF by optimizing the weight matrix of different types of objects. However, it still could not avoid the noise interference in the intra-association relation data during the fusion phase. Yang *et al.* [3] further extended DFMF by considering special weights for intra-association relation data and by considering important bag-instance correlations into M4L task. The theoretical analysis and experimental results of the two extended methods demonstrate that selective fusion of different intra- and inter-association correlations can further improve the representation of objects during fusion of complex objects. However, they prefer to select sparse objects, which own relation matrix with more zero element, and sparse data usually have a smaller smoothing loss ($tr(\mathbf{G}^T\mathbf{\Theta}_i^{(t)}\mathbf{G})$) or approximation loss ($\|\mathbf{R}_{ij} - \mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\|_F^2$). In practice, a very sparse data matrix usually does not encode sufficient information for the target task, so sparse data may be irrelevant to the target task.

3.3 Unified Objective Function

Multiple inter-(intra-) relational data matrices contain complementary information for different types of objects, but they may also contain some noisy or irrelevant data matrices [3, 36]. Although the low-rank matrix factorization can reduce the inner noises of individual data matrices to some extent [37, 38], it is still necessary to selectively integration and fusion of different relevance relation between data matrix. Specifically, We support adaptive weights for the set of inter (intra)-relational data matrices, which explicitly remove the noisy data matrices. In addition, in order to reduce the preference for sparse data or the influence of noisy data, we define the objective function of SMFM4L as follows:

$$
\begin{aligned}
\min_{G\geq 0} \Omega(\boldsymbol{G},\boldsymbol{S},\boldsymbol{W}^r,\boldsymbol{W}^h) = &\sum_{i,j=1}^{m} \boldsymbol{W}_{ij}^r \|\boldsymbol{R}_{ij} - \boldsymbol{G}_i\boldsymbol{S}_{ij}\boldsymbol{G}_j^T\|_F^2 + \|\mathbf{R}_{bl} - \mathbf{R}_{bi}\mathbf{G}_i\mathbf{S}_{il}\mathbf{G}_l^T\|_F^2 \\
&+ \sum_{p=1}^{m}\sum_{t=1}^{\tau} \boldsymbol{W}_{pt}^h \|\boldsymbol{R}_{pp}^t - \boldsymbol{G}_p\boldsymbol{S}_{pp}\boldsymbol{G}_p^T\|_F^2 + \alpha\|vec(\boldsymbol{W}^r)\|_F^2 + \beta\|vec(\boldsymbol{W}^h)\|_F^2 \\
&s.t. \boldsymbol{W}^r \geq 0, \boldsymbol{W}^h \geq 0, \sum vec(\boldsymbol{W}_i^r) = 1, \sum vec(\boldsymbol{W}_i^h) = 1
\end{aligned}
\tag{3}
$$

where $\tau = max_i t_i$, $\boldsymbol{S}_{ij} \in \mathbb{R}^{k_i \times k_j}$, $\mathbf{W}^r \in \mathbb{R}^{m\times m}$ and $\mathbf{W}^h \in \mathbb{R}^{m\times\tau}$ are the weight matrices assigned to different inter-relational data matrices and different intra-relational data matrices, respectively. If $\boldsymbol{R}_{ij} \notin \mathcal{R}$, $\boldsymbol{W}_{ij}^r = 0$. $\mathbf{W}_{pt}^h$ is the weight of the t-th intra-relational matrix of the p-th object type. If $\boldsymbol{R}_{pp}^t \notin \mathcal{R}$ or $t > t_i$, $\boldsymbol{W}_{pt}^r = 0$. Unlike Equation (2), our objective function not only selectively integrates inter-relation and intra-relation data matrices using weight matrices, but also approximates $\boldsymbol{R}_{pp}^t$ to fuse it by using shared low-rank matrices $\boldsymbol{G}_p$ and $\boldsymbol{S}_{pp}$ in the process of crossing t_i intra-relation data. In this way, a data matrix inconsistent with other data matrices of the same objects will be assigned with a lower weight. So, data matrices that are inconsistent with other data matrices within the same object's intra-relational matrices will be given lower weights. Especially for dense data matrices $\boldsymbol{R}_{pp}^t$, $\|\boldsymbol{R}_{pp}^t - \boldsymbol{G}_p\boldsymbol{S}_{pp}\boldsymbol{G}_p^T\|_F^2$ has a larger loss because $\boldsymbol{R}_{pp}^t$ encodes intra-relation more than its cousin matrix ($\{\boldsymbol{R}_{pp}^t\}_t^{t_p}, t_p \neq t$) is much less and the losses are mainly occupied by $tr(\mathbf{G}^T\mathbf{\Theta}_i^{(t)}\mathbf{G})$. In the same way, for sparse matrices with noisy elements, $\|\boldsymbol{R}_{pp}^t - \boldsymbol{G}_p\boldsymbol{S}_{pp}\boldsymbol{G}_p^T\|_F^2$ leads to a large loss. $vec(\mathbf{W}_i^r)$ and $vec(\mathbf{W}_i^h)$ are the vectorisation operator that stacks the

i-th row of $\mathbf{W}^r$ and $\mathbf{W}^h$, α and β are the regularization weights for these two weight matrices, they are used to control the complexity of $vec(\mathbf{W}^)$ and $vec(\mathbf{W}^h)$. Therefore, SMFM4L can not only selectively fuse and integrate several related data matrices, but also automatically remove unrelated data matrices using these two regularization items.

To minimize the above objective function, smaller inter association weight and intra association weight are automatically assigned to different types of objects and same types of objects, respectively. For the first term in objective function, due to $\boldsymbol{G}_i$ is also shared by an interconnected objects, sparse selection of this term can also be avoided. Therefore, Equation (3) avoids the preference for sparse association relation. Overall, Equation (3) can not only explore the contribution of different intra-relational data matrices and inter-relational matrices by assigning weights to them, but also avoid favoring sparse objects. We would like to point out that this objective function provides a more general and comprehensive framework for effectively modeling complex objects and effectively avoiding interference from sparse or noisy data.

The objective function of our SMFM4L is non-convex in $\mathbf{G}$, $\mathbf{S}$, $\mathbf{W}^r$ and $\mathbf{W}^h$ altogether. We can use the idea of auxiliary functions which is frequently used in the convergence proof of approximate matrix factorization algorithms to optimize $\mathbf{G}$ and $\mathbf{S}$ in (3) [39, 40], and the idea of convex optimization and Karush-Kuhn-Tucker (KKT) multipliers [41] to optimize $\mathbf{W}^r$ and $\mathbf{W}^h$. Due to the page limit, the optimization process and convergence proof are ignored.

Table 2. Statistics of datasets used for the experiments. *avgBL* is the average number of labels per bag and *avgBI* is the average number of instances per bag.

Dataset	Bags	Instances	Labels	AvgBL	AvgBI	View	Node Type	Link Type
Letter Frost	144	565	26	3.6	3.9	1		
Letter Carroll	166	717	26	3.9	4.3	1		
MSRC v2	591	1,758	23	2.5	1.0	1		
Birds	548	10,232	13	2.1	18.7	1		
Isoform	795	6,457	704	6.7	5.2	2	5	5

4 Empirical Evaluation

4.1 Experimental Setup

To study the performance of our proposed algorithm, we designed three experiments to test the performance of SMFM4L. In the first experiment, we used real-world M4 data (Isoform dataset) [3] to study the performance of SMFM4L and compare it with data fusion methods, M4L and its degenerated versions. In the second experiment, we compared SMFM4L against M4L and its degenerated versions (Multi-view Multi-lable learning, Multi-view Multi-instance learning and Multi-instance Multi-label learning) on benchmark multi-instance multi-label datasets [1, 42–44]. We have listed the statistics for these datasets in Table

Table 3. Results on Isoform dataset of M4L, M3L-based methods and matrix factorization (MF)-based methods by 5-fold cross validation. •/◦ indicates whether SMFM4L is statistically (according to pairwise t-test at 95% significance level) superior/inferior to the other method.

	Method	**AvgF1**	**AUROC**	**AUPRC**
M3L	M3Lcmf	0.152±0.004•	0.663±0.018•	0.154±0.013•
	ICM2L	0.074±0.001•	0.533±0.001•	0.041±0.022•
	M2IL	0.025±0.004•	0.544±0.009•	0.032±0.013•
MF	DFMF	0.051±0.001•	0.943±0.009•	0.637±0.054•
	S-NMTF	0.021±0.001•	0.790±0.012•	0.015±0.002•
	WMFLDA	0.032±0.001•	0.949±0.004•	0.612±0.007•
M4L	M4L-JMF	0.055±0.002•	0.967±0.004•	0.674±0.026•
	SMFM4L	0.256±0.001	0.981±0.002	0.713±0.012

(2). In particular, we compare the performance of SMFM4L against some related methods, including M3L solutions (M3Lcmf [1], ICM2L [45] and M2IL [9]), data fusion solutions based on matrix factorization (S-NMTF [12], DFMF [6] and WMFLDA [36]), M4L solution (M4L-JMF [3]). Input parameters of the comparative method are based on the authors specified in the code or the advice of the paper (or optimized). We will explore the sensitivity of these parameters later. For each comparison method, we performed a five-fold cross-validation in ten independent experiments and reported the average results. All experiments are performed on a server configured with CentOS 7.3, 256GB RAM and Intel Exon E5-2678 v3.

To effectively evaluate the performance of SMFM4L, three frequently used evaluation metrics (*AUPRC*, *AUROC* and *AvgF1*) in multi-label learning and bioinformatics are adopted to quantify the overall performance: The average area under the precision-recall curve (*AUPRC*), which calculates the value of the area under the Precision-Recall (PR) curve for each label and then takes the average values for all labels; The average area under the receiver operating curve (*AUROC*), which computes the area under the ROC curve for each label, and then takes the average for all labels; and the average F1-score (*AvgF1*) across all classes by computing the harmonic precision value and recall value for each label, and then takes the average of all labels, the formal definitions of these metrics are omitted due to the page limit, and can be found in [46]. The larger values are an indication of a better performance for all metrics.

4.2 Results on M4 data

To effectively evaluate the performance of SMFM4L, we applied the SMFM4L algorithm to predict the association between isoforms and GO (Gene Ontology) terms (as the labels of isoform objects) on the Isoform dataset [3]. Since M3L methods cannot handle M4 data directly, we first project other types of objects to Gene to form M3 data, and then use these M3L methods to target task. As for other data fusion based on matrix factorization methods, we use this classical setup [3], we ignore this particular bag-instance relation, and then use the multi-type object information to predict the association between labels and genes. We use the top K labels corresponding to the largest entry in each row of $\mathbf{R}_{bl}$ ($\mathbf{R}_{il}$) as the relevant labels for the bag (or instance), here K denotes the average number of labels per bag or instance.

In Table 3 we report the results of the first experiment. From Table 3, we have three important and interesting observations: The first is that feature representation of bags or instances can be significantly improved by fusing multi-type object information. Since M3L methods cannot process M4 data directly, they can only process M3 data formed after projection. The results in Table 3 show that the M4L methods are more powerful than the M3L methods in terms of performance as shown by the AUROC and AUPRC values in Table 3. This is because the projection operation causes information loss and therefore hurts the modeling capability of M3L. It is further shown that our proposed SMFM4L method can naturally handle multi-type objects, while the M3L method only considers a single type of object and is not directly adaptable to multi-type object setting;

The second observation is that bag-instance association information is important in M4L learning tasks, and selective fusion of data can improve performance. This finding comes from comparing the matrix factorization based method with M4L solution, which has higher AUROC and AUPRC values, because M4L-JMF or SMFM4L account for finer-grained multi-instance information. Overall, this comparison confirms the importance of bag-instance association in the M4L learning task, and selective fusion of multi-type object information can improve classifier performance;

The final interesting observation is that avoiding sparse data preferences can better fuse multiple types of objects. This observation comes from the comparison between M4L-JMF and SMFM4L, where SMFM4L performs significantly better than M4L-JMF although M4L-JMF also accounts for the relation between different types of objects. This is because the M4L-JMF method uses the manifold regularization and approximate losses to determine the relevance of these objects. Therefore, the M4L-JMF method prefers sparse data and gives them greater weights in the fusion process. However, these sparse data are irrelevant to the target task, so the performance of M4L-JMF is lower than that of SMFM4L, especially in *AvgF1* value. In contrast, SMFM4L does not have this preference, so SMFM4L gives the best results among all the compared methods.

In conclusion, the experimental results on real-world M4 datasets suggest that SMFM4L can not only adequately model M4 data without destroying the intrinsic structure between multiple types of objects and ignoring bag-instance association relation, but also avoid the interference of sparse data. It is due to these advantages that it beats all comparison methods and obtains the best results.

Table 4. Results on four datasets with bag-level labels of compared methods by 5-fold cross validation. •/◦ indicates whether SMFM4L is statistically (according to pairwise t-test at 95% significance level) superior/inferior to the other method.

Metric	ICM2L	M3Lcmf	S-NMTF	DFMF	WMFLDA	M4L-JMF	SMFM4L
				Birds			
AvgF1	0.372±0.009•	0.485±0.012◦	0.208±0.064•	0.252±0.012•	0.256±0.011•	0.268±0.014•	0.395±0.011
AUROC	0.720±0.008•	0.604±0.032•	0.738±0.027•	0.902±0.009•	0.911±0.006•	0.944±0.002•	0.955±0.003
AUPRC	0.255±0.019•	0.425±0.012•	0.101±0.025•	0.886±0.045•	0.897±0.009•	0.963±0.008	0.984±0.005
				Letter Carroll			
AvgF1	0.317±0.001•	0.543±0.044◦	0.095±0.011•	0.247±0.014•	0.269±0.012•	0.288±0.014•	0.483±0.0107
AUROC	0.589±0.006•	0.649±0.022•	0.705±0.003•	0.906±0.008•	0.916±0.003•	0.924±0.002•	0.951±0.003
AUPRC	0.101±0.002•	0.421±0.012•	0.084±0.016•	0.909±0.056•	0.916±0.013•	0.948±0.008•	0.971±0.006
				Letter Frost			
AvgF1	0.252±0.001•	0.538±0.050◦	0.077±0.017•	0.242±0.018•	0.248±0.031•	0.250±0.009•	0.392±0.006
AUROC	0.638±0.010•	0.665±0.002•	0.645±0.033•	0.907±0.003•	0.912±0.003•	0.924±0.002•	0.951±0.003
AUPRC	0.211±0.024•	0.495±0.083•	0.128±0.007•	0.894±0.049•	0.912±0.025•	0.951±0.008•	0.973±0.012
				MSRC v2			
AvgF1	0.219±0.003•	0.426±0.028	0.054±0.022•	0.198±0.008•	0.211±0.007•	0.215±0.004•	0.412±0.006
AUROC	0.668±0.063•	0.698±0.018•	0.695±0.041•	0.937±0.002•	0.941±0.003•	0.958±0.001•	0.972±0.004
AUPRC	0.253±0.026•	0.444±0.083•	0.106±0.046•	0.880±0.021•	0.917±0.009•	0.933±0.003•	0.958±0.007

4.3 Results on M3 data

We further investigate the performance of SMFM4L on four benchmark canonical single-view multi-instance multi-label datasets . For the experiments of four datasets in Table 2, we randomly split the data view into two views, each containing half the features of the original views. Other pre-treatment are the same as those in the previous subsection. Table 4 reported the experimental results of four multi-instance multi-label datasets (listed in Table 2).

The experimental results in Table 4 demonstrate that our proposed SMFM4L works better than the other comparison methods and has relatively high $AUPRC$ and $AUROC$ values. Compared to the second-best baseline, the $AUPRC$ and $AUROC$ values are at least 2% and 1% higher, respectively. The increase in $AUPRC$ values is even more significant, this observation implies that those four multiple-instance multi-label datasets are class imbalanced datasets. Thus, in terms of overall performance, our proposed method performs better.

5 Conclusion

In this paper, we introduced a Multi-typed objects Multi-view Multi-instance Multi-label Learning solution based on Selective Matrix Factorization (SMFM4L) for naturally interlinked multiple type of objects. Unlike existing M4L based on matrix factorization approaches or degraded versions, SMFM4L is more powerful and effective than them due to it can not only selectively integrate multi-relational matrix, but also avoid preferring to sparse association data. Experimental results on real-word and benchmark datasets validated that SMFM4L can more comprehensively fuse multiple type of objects by selective matrix factorization and mine the complex relation among bags, instances and labels, and achieves a much better performance than the state-of-the-art methods in predicting bag/instance label relation.

For future works, we plan to consider more network structure and semantic information such as graph structure information in heterogeneous multi-instance biological networks for cooperative driver pathway discovery, and study the impact of different semantic information on prediction tasks (bag and instance level).

References

1. Yuying Xing, Guoxian Yu, Carlotta Domeniconi, Jun Wang, Zili Zhang, and Maozu Guo. Multi-view multi-instance multi-label learning based on collaborative matrix factorization. In *AAAI Conference on Artificial Intelligence*, pages 5508–5515, 2019.
2. Cam-Tu Nguyen, Xiaoliang Wang, Jing Liu, and Zhi-Hua Zhou. Labeling complicated objects: Multi-view multi-instance multi-label learning. In *AAAI Conference on Artificial Intelligence*, pages 2013–2019, 2014.
3. Yuanlin Yang, Guoxian Yu, Jun Wang, Carlotta Domeniconi, and Xiangliang Zhang. Multi-type objects multi-view multi-instance multi-label learning. *IEEE International Conference on Data Mining (ICDM)*, pages 1370–1375, 2020.
4. Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017.
5. Mehmet Gönen and Ethem Alpaydın. Multiple kernel learning algorithms. *JMLR*, 12:2211–2268, 2011.
6. Marinka Žitnik and Blaž Zupan. Data fusion by matrix factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):41–53, 2015.
7. Yuehui Wang, Guoxian Yu, Carlotta Domeniconi, Jun Wang, Xiangliang Zhang, and Maozu Guo. Selective matrix factorization for multi-relational data fusion. In *DASFAA*, pages 313–329, 2019.
8. Yuanlin Yang, Guoxian Yu, Carlotta Domeniconi, and Xiangliang Zhang. Deep multi-type objects multi-view multi-instance multi-label learning. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, pages 486–494. SIAM, 2021.
9. Bing Li, Chunfeng Yuan, Weihua Xiong, Weiming Hu, Houwen Peng, Xinmiao Ding, and Steve Maybank. Multi-view multi-instance learning based on joint sparse representation and multi-view dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2554–2560, 2017.
10. Qiaoyu Tan, Guoxian Yu, Jun Wang, Carlotta Domeniconi, and Xiangliang Zhang. Individuality-and commonality-based multiview multilabel learning. *IEEE T. on CYB*, 99(1):1–13, 2020.
11. Guoxian Yu, Yuehui Wang, Jun Wang, Guangyuan Fu, Maozu Guo, and Carlotta Domeniconi. Weighted matrix factorization based data fusion for predicting lncrna-disease associations. In *BIBM*, pages 572–577, 2018.
12. Hua Wang, Heng Huang, and Chris Ding. Simultaneous clustering of multi-type relational data via symmetric nonnegative matrix tri-factorization. In *the Conference on Information and Knowledge Management*, pages 279–284, 2011.

13. Zhi-Hua Zhou, Min-Ling Zhang, Sheng-Jun Huang, and Yu-Feng Li. Multi-instance multi-label learning. *Artificial Intelligence*, 176(1):2291–2320, 2012.
14. Sheng-Jun Huang, Wei Gao, and Zhi-Hua Zhou. Fast multi-instance multi-label learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(11):2614–2627, 2018.
15. Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D Manning. Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pages 455–465, 2012.
16. Bing Li, Chunfeng Yuan, Weihua Xiong, Weiming Hu, Houwen Peng, Xinmiao Ding, and Steve Maybank. Multi-view multi-instance learning based on joint sparse representation and multi-view dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2554–2560, 2017.
17. Bin Wu, Erheng Zhong, Andrew Horner, and Qiang Yang. Music emotion recognition by multi-label multi-layer multi-instance multi-view learning. In *ACM MM*, page 117–126, 2014.
18. Meng Liu, Yong Luo, Dacheng Tao, Chao Xu, and Yonggang Wen. Low-rank multi-view learning in matrix completion for multi-label image classification. In *AAAI Conference on Artificial Intelligence*, pages 2778–2784, 2015.
19. Cam-Tu Nguyen, De-Chuan Zhan, and Zhi-Hua Zhou. Multi-modal image annotation with multi-instance multi-label lda. In *IJCAI*, pages 1558–1564, 2013.
20. David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
21. Cam-Tu Nguyen, Xiaoliang Wang, Jing Liu, and Zhi-Hua Zhou. Labeling complicated objects: Multi-view multi-instance multi-label learning. In *AAAI Conference on Artificial Intelligence*, 2014.
22. Yang Yang, Yi-Feng Wu, De-Chuan Zhan, Zhi-Bin Liu, and Yuan Jiang. Complex object classification: A multi-modal multi-instance multi-label deep network with optimal transport. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2594–2603, 2018.
23. Cédric Villani. Optimal transport: old and new. *Bull. Amer. Math. Soc*, 47:723–727, 2010.
24. Yang Yang, Zhao-Yang Fu, De-Chuan Zhan, Zhi-Bin Liu, and Yuan Jiang. Semi-supervised multi-modal multi-instance multi-label deep network with optimal transport. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 99(1):1–14, 2021.
25. Guangyuan Fu, Jun Wang, Carlotta Domeniconi, and Guoxian Yu. Matrix factorization-based data fusion for the prediction of lncrna–disease associations. *Bioinformatics*, 34(9):1529–1537, 2018.
26. Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 701–710, 2014.
27. Xiaokai Wei, Linchuan Xu, Bokai Cao, and Philip S. Yu. Cross view link prediction by learning noise-resilient representation consensus. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 1611–1619, 2017.
28. Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017.
29. Vladimir Gligorijević and Nataša Pržulj. Methods for biological data integration: perspectives and challenges. *Journal of the Royal Society Interface*, 12(112):20150571, 2015.

30. Marinka Žitnik, Francis Nguyen, Bo Wang, Jure Leskovec, Anna Goldenberg, and Michael Hoffman. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 50:71–91, 2019.
31. Sugato Basu, Arindam Banerjee, and Raymond J Mooney. Active semi-supervision for pairwise constrained clustering. In *SIAM International Conference on Data Mining*, pages 333–344, 2004.
32. Mikhail Bilenko, Sugato Basu, and Raymond J Mooney. Integrating constraints and metric learning in semi-supervised clustering. In *International Conference on Machine Learning*, pages 59–68, 2004.
33. Thomas Dietterich, Richard Lathrop, and Tomas Lozano-Perez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2):31–71, 1997.
34. Peng Jiang, Haibin Ling, Jingyi Yu, and Jingliang Peng. Salient region detection by ufo: Uniqueness, focusness and objectness. In *CVPR*, pages 1976–1983, 2013.
35. Guoxian Yu, Keyao Wang, Carlotta Domeniconi, Maozu Guo, and Jun Wang. Isoform function prediction based on bi-random walks on a heterogeneous network. *Bioinformatics*, 36(1):303–310, 2020.
36. Yuehui Wang, Guoxian Yu, Jun Wang, Guangyuan Fu, Maozu Guo, and Carlotta Domeniconi. Weighted matrix factorization on multi-relational data for lncrna-disease association prediction. *Methods*, 173:32–43, 2020.
37. Deyu Meng and Fernando De La Torre. Robust matrix factorization with unknown noise. In *IEEE International Conference on Computer Vision*, pages 1337–1344, 2013.
38. Xia Chen, Guoxian Yu, Carlotta Domeniconi, Jun Wang, Zhao Li, and Zili Zhang. Cost effective multi-label active learning via querying subexamples. In *IEEE International Conference on Data Mining (ICDM)*, pages 905–910, 2018.
39. Chris HQ Ding, Tao Li, and Michael I Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Trans on Pattern Analysis and Machine*, 32(1):45–55, 2008.
40. Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Neural Information Processing Systems*, pages 556–562, 2001.
41. Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.
42. Forrest Briggs, Xiaoli Z Fern, and Raviv Raich. Rank-loss support instance machines for miml instance annotation. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 534–542, 2012.
43. Guoxian Yu, Xia Chen, Carlotta Domeniconi, Jun Wang, Zhao Li, Zili Zhang, and Xiangliang Zhang. Cmal: Cost-effective multi-label active learning by querying subexamples. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 99(1):1–14, 2021.
44. Sheng-Jun Huang, Nengneng Gao, and Songcan Chen. Multi-instance multi-label active learning. In *IJCAI*, pages 1886–1892, 2017.
45. Qiaoyu Tan, Guoxian Yu, Jun Wang, Carlotta Domeniconi, and Xiangliang Zhang. Individuality- and commonality-based multiview multilabel learning. *IEEE TCYB*, 99(1):1–12, 2021.
46. Minling Zhang and Zhihua Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 26(8):1819–1837, 2014.