A multi-view multi-label learning with incomplete data and self-adaptive correlations

Changming Zhu

cmzhu@shmtu.edu.cn

Shanghai Maritime University

Liju Han

Shanghai Maritime University

Research Article

Keywords: Self-adaptive, Incomplete data, Multi-view multi-label

Posted Date: September 23rd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4951118/v1

Additional Declarations: No competing interests reported.

A multi-view multi-label learning with incomplete data and self-adaptive correlations

Abstract

Multi-view multi-label (MVML) learning aims to process MVML data sets represented with multiple feature sets (i.e., views) and labeled with multiple class labels. While in current scenes, MVML data sets always encounter to two main phenomena. First, for MVML data, there exist some correlations among different features, instances, labels and these correlations usually have diverse representations including within-view, cross-view, and consensus-view representations. Due to there are usually some certain relationships between information exist, thus these correlations always be changed in a self-adaptive way. Second, for some unpredictable reasons, MVML data maybe incomplete and loss some information. To address these phenomena, we pay attention to the self-adaptive measurement of those correlations in different representations and the process of incomplete data, then a multi-view multi-label learning with incomplete data and self-adaptive correlations (MVML-IDSaC) is developed. Extensive experiments on 5 MVML data sets show the superiority of the developed algorithm and some conclusions are addressed. (1) MVML-IDSaC performs better than some related competitive algorithms in statistical over AUC and precision; (2) MVML-IDSaC can process incomplete MVML data much better; (3) considering comprehensive relationships about data and its inferring results with a feasible way, the performances of a multi-view multi-label algorithm is promoted further.

Keywords: Self-adaptive, Incomplete data, Multi-view multi-label

1 Introduction

1.1 Background

Multi-view multi-label (MVML) data sets are widely used in current applications and each MVML instance is represented with multiple feature sets (namely, views) and labeled with multiple class labels. In order to process these MVML data sets, many algorithms are developed [1, 2, 3, 4] and two kinds of algorithms are widely concerned. One is the correlation-based algorithm and the other is incomplete data-based algorithm. For the former one, algorithms pay more attention to the correlations among features, instances, and labels which can be demonstrated in multiple representations, namely, within-view, cross-view, and consensus-view representations (see Fig. 1) [5, 6, 7, 8]. Then for the latter one, algorithms aim to process incomplete MVML data with some strategies including indicator matrix, reconstruction error, etc. and incomplete data represent instances with missing labels, features, etc [9, 10, 11]. On the base of the above two kinds of algorithms, recent scholars make further promotes and develop some algorithms with multiple correlations and incomplete data considered [12, 13].

1.2 Main problem

Although the above mentioned algorithms can process MVML data sets, there still exist some common problems. First, some algorithms [1, 2, 3, 4, 5, 6, 7, 8] measure correlations among features, labels, or instances with low-rank preserving terms or self-adaptive constraints used. But these algorithms are hard to measure multiple correlations in within-view, cross-view, and consensus-view representations simultaneously. Moreover, these algorithms cannot reflect the self-adaptive change for multiple correlation information authentically. Second, some algorithms process incomplete data

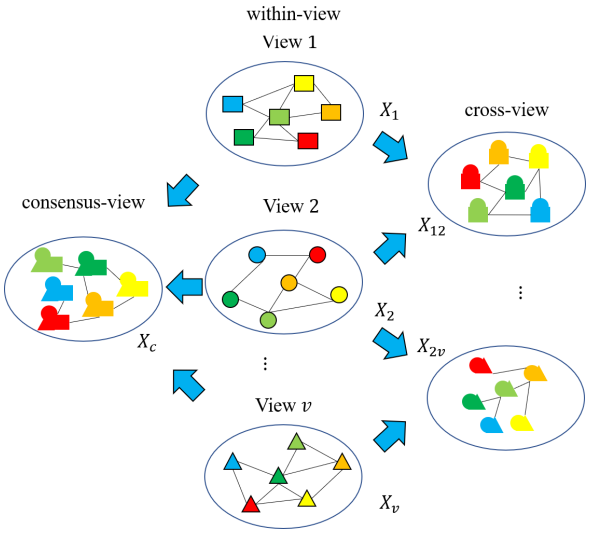

Fig. 1 Illustration of within-view, cross-view, and consensus-view representations. Three shapes (square, circle, triangle) and six colors (red, orange, yellow, green, dark green, blue) are used to distinguish different views and instances, respectively, and the instances of the same color belong to the same instance. Here, within-view representation demonstrates data information expressed in a view, cross-view representation is shared by two different views and its information can be treated as the source for the information of two views, consensus-view representation is shared by all views and its information describes the consensus representation and source of information from all different views.

without the consideration about multiple correlations in different representations [9, 10, 11] and the incomplete data cannot be inferred well. Third, the current algorithms with multiple correlations and incomplete data considered [12, 13] still measure the correlations with artificial setting and ignore the laws of self-adaptive change for multiple correlation information.

Indeed, how to explore the laws of self-adaptive change for multiple correlation information and apply the laws into the incomplete MVML data is an open problem.

1.3 Proposal and contributions

To solve the problems mentioned in subsection 1.2, we take the incomplete MVML data as the research objects and explore the laws of self-adaptive change for multiple correlation information in different representations including within-view, consensus-view, and cross-view ones according to the data characteristics. Then on the base of the explored laws, we design some feasible self-adaptive constraints and introduce a 'within-consensus-cross'-based individual- and mutual- balance model (WCC-IM) to infer the incomplete data better. Finally, on the base of self-adaptive constraints, WCC-IM, we propose a new algorithm named multi-view multi-label learning with incomplete data and self-adaptive correlations (MVML-IDSaC).

Compared with the previous algorithms, the major contributions of this study are (1) MVML-IDSaC can measure the correlations among features, labels, instances for a MVML data more authentically; (2) WCC-IM helps the MVML-IDSaC to infer incomplete data in a sound way; (3) the computational cost of MVML-IDSaC is also smaller than the traditional one, $O(n^3)$-level.

The main work of this study includes (1) we put forward a new design concept for models of multi-view multi-label learning and the infer of incomplete data. Then we elaborate the corresponding framework, optimization procedure, and computational cost; (2) we analyze influence of different correlations and laws of their self-adaptive change in the representations about within-view, consensus-view, and cross-view; (3) we report the significant performances of MVML-IDSaC compared with the baselines.

1.4 Framework

The study is organized as follows: Section 2 reviews the related works about multi-view learning, multi-label learning, and multi-view multi-label learning. Then for each kind of learning, the algorithms to process incomplete data are also reviewed. Section 3 shows the framework, optimization procedure, and computational cost of MVML-IDSaC. In section 4, we report on numerous experiments to evaluate the proposed algorithm. Finally, section 5 concludes this study and advises the future work.

2 Related work

In this section, we review some important work about multi-view learning, multi-label learning, and multi-view multi-label learning. The applications of them to process incomplete data are also reviewed.

2.1 Multi-view learning

Multi-view learning aims to process data sets which can be represented with multiple feature sets and many algorithms are developed. For example, Xue et. al [14] propose a multi-view learning framework with a linear computational cost which simultaneously analyze all features by learning an integrated projection matrix; Niu et. al [15] propose a multi-view representation learning with subspace transformation relationships used and the essential structural information of multi-view data are fully exploited; Lou et. al [16] propose two multi-view intuitionistic fuzzy support vector machines with insensitive pinball loss that can not only handle the general multi-view classification problems but also be robust to noisy data.

Besides for these algorithms, some algorithms are developed for the incomplete data. For example, He et. al [17] develop a multi-view clustering algorithm to discover the latent cluster structure and partition the incomplete multi-view data into different groups with an adaptive instance completion scheme; Sun et. al [18] develop a graph learning-based incomplete multi-view clustering algorithm and learn a common consensus graph from all incomplete views and obtain a clustering indicator matrix in a unified framework. Then a relaxed spectral clustering model is introduced to obtain a probability consensus representation with all positive elements that reflect the data clustering result; Mu et. al [19] propose a tensor-based incomplete multi-view clustering framework and the missing views are reconstructed based on non-negative matrix factorization and self-representation.

2.2 Multi-label learning

Multi-label learning pays more attention to the solution of data sets which instances can be labeled with class labels and many algorithms are developed. For example, Sun et. al [20] propose a co-evolutionary multilabel hypernetwork (Co-MLHN) as an attempt to exploit label correlations in an effective and efficient way and they can classify the multi-label data sets better; Che et. al [21] develop a multi-label algorithm which focuses on the label correlation adapted to overall data and the locally targeted information presented by some instances; Zhai et. al [22] develop a family of online multi-label classification algorithms, which can update the model instantly and efficiently, and make a timely online prediction when new data arrive.

Besides for these algorithms, some algorithms are also developed for the incomplete data. For example, Li et. al [23] propose a multi-label learning with missing labels (MLML) framework for facial action unit recognition under incomplete data. Different from most MLML works which usually use the same features for all classes, the proposed method discriminates each action unit based on the most related features. Selecting features for each action unit individually embeds the observation that occurrences of different action units produce feature changes of different face regions; Sun et. al [24] propose a weakly-supervised multi-label learning framework called WML-LSC, where the low-rank and sparse constrain schemes are jointly incorporated to capture the desired feature information. Then, the robustness of the learning model can be kept; Beigaite et. al [25] propose an algorithmic approach to reduce noise in the target labels and improve predictions and this algorithm allows us to preserve the existing information about the vegetation types in training data. At the same time, the algorithm potentially enhances this information by redistributing the fraction of human activity in urban areas and croplands across the training set;

2.3 Multi-view multi-label learning

Different from multi-view learning and multi-label learning, the multi-view multi-label learning is developed for MVML data. For example, Tan et. al [1] develop an approach named individuality- and commonality-based multi-view multi-label learning which improves the discriminant capacity of classifier toward rare labels and sufficiently considers the diverse characteristics of individual views; Zhu et. al [26] simultaneously conduct a hierarchical feature selection and a MVML learning for multi-view image classification, via embedding a proposed a new block-row regularizer into the MVML framework. Then this framework effectively conduct image classification by avoiding the adverse impact of both the redundant views and the noisy features; Ma et. al [27] propose a group-based model with local feature and label selection.

The proposed model can project instances into different groups by performing group-based feature selection with each view having its own importance for grouping, where each group has its own related labels. The proposed model can then predict the semantics of instances by performing group-based label selection with each group having its own weight for prediction. Besides, the inter-group correlation is also mined and introduced in the above group-based learning process to ensure effective multi-label classification.

In terms of incomplete problems, some corresponding algorithms are developed. For example, Wen et. al [13] propose an incomplete multi-view multi-label learning network which is composed of four major parts: view-specific deep feature extraction network, weighted representation fusion module, classification module, and view-specific deep decoder network. By, respectively, integrating the view missing information and label missing information into the weighted fusion module and classification module, the proposed algorithm can effectively reduce the negative influence caused by two such incomplete issues and sufficiently explore the available data and label information to obtain the most discriminative feature extractor and classifier; Qu et. al [12] propose an incomplete multi-view multi-label active learning (iMVMAL) algorithm to reduce the cost of querying MVML data with the usage of Autoencoder to learn the shared/individual representations from incomplete multi-view data, and leverages the commonality and individuality information, information from labels and that from sample features to query the most informative sample-label pairs; Li et. al [9] propose a concise yet effective model to simultaneously tackle the missing labels, incomplete views and non-aligned views challenges with only one hyper-parameter in the objective function.

3 Methodology

As what we mentioned before, although recent algorithms including the ones mentioned in section 2 can process those MVML data, multi-view data, multi-label data at some extents, they still forget some issues. First, multiple correlations in within-view, cross-view, and consensus-view representations cannot be self-adaptive measured simultaneously and the laws of self-adaptive change of the correlations are not considered. Second, they cannot infer the incomplete data well without the consideration about those multiple correlations.

In order to solve those problems, we construct the basic learning model and explore the laws of self-adaptive change for multiple correlation information. Then we design a model to infer the incomplete data better. Finally, we develop a multi-view multi-label learning with incomplete data and self-adaptive correlations, i.e., MVML-IDSaC.

3.1 Framework of MVML-IDSaC

Refer to our previous work [28], suppose there is an incomplete MVML data set $D = [X, Y]$ with V views and n instances which are information-complete or information-incomplete. Information-complete indicates that the instance has complete features and labels and the information-incomplete implies that there is some information missing among features or labels. Then for vth view, we let $X^v \in \mathbb{R}^{n\times d_v}$ and $Y^v \in \mathbb{R}^{n\times c_v}$ be the set of instances (including the information-complete instances and information-incomplete ones) and the corresponding labels. Meanwhile, $X^{Av} \in \mathbb{R}^{n_v\times d_{Av}}$ and $Y^{Av} \in \mathbb{R}^{n_v\times c_{Av}}$ are the set of information-complete instances and the corresponding labels. Here, d_v and c_v are feature dimension and number of labels for vth view respectively, d_{Av}, c_{Av} are the corresponding feature dimension, number of labels, respectively for n_v information-complete instances. Compared with X^{Av} and Y^{Av}, the incomplete parts in X^v and Y^v are set as 0.

Then according to X^{Av} and Y^{Av}, we set four pre-constructed matrices from the information-complete instances of vth view. They are $\bar{S}^v \in \mathbb{R}^{n_v\times n_v}$, $\bar{P}^v \in \mathbb{R}^{d_{Av}\times d_{Av}}$, $\bar{Q}^v \in \mathbb{R}^{c_{Av}\times c_{Av}}$, $\bar{W}^v \in \mathbb{R}^{c_{Av}\times d_{Av}}$ and each element of them is non-negative and represents the instance-instance correlation, feature-feature correlation, label-label correlation, feature-label correlation between two information-complete instances, respectively.

Since X^v and Y^v include some incomplete information, thus every pre-constructed matrix can not reveal the comprehensive relationships of all instances. Furthermore, $D = [X, Y]$ can be represented by within-view, cross-view, and consensus-view representations.

To this end and refer to [28], we should consider that how to infer the incomplete information with the complete information preserved firstly and how to reveal the comprehensive relationships between instances.

3.1.1 Preservation of complete information

Refer to [28], in order to infer the incomplete information from vth view, we set four corresponding referred (completed, comprehensive) matrices which store comprehensive relationships of all instances ultimately. Namely, $S^v \in \mathbb{R}^{n\times n}$, $P^v \in \mathbb{R}^{d_v\times d_v}$, $Q^v \in \mathbb{R}^{c_v\times c_v}$, $W^v \in \mathbb{R}^{c_v\times d_v}$ whose elements represent the instance-instance correlations, feature-feature correlations, label-label correlations, feature-label correlations between two instances, respectively. Then the correlation information of the information-complete instances in $\bar{S^v}$, $\bar{P^v}$, $\bar{Q^v}$, $\bar{W^v}$ should be preserved in the referred matrices S^v, P^v, Q^v, W^v, respectively. After that, in order to incorporate the detailed situation of missing information, we let $E^v \in \mathbb{R}^{d_v\times n}$ be the feature missing-index matrix and $F^v \in \mathbb{R}^{c_v\times n}$ be the label missing-index matrix for vth view, respectively. Here the element $E^v_{i_v,j}$ (or $F^v_{k_v,j}$) in E^v (or F^v) is set be 1 if i_vth feature (or k_vth label) of jth instance in vth view is available (or information-complete), and be 0, otherwise.

After that, we extend $\bar{S^v}$, $\bar{P^v}$, $\bar{Q^v}$, $\bar{W^v}$ to be $\widetilde{S^v} \in \mathbb{R}^{n\times n}$, $\widetilde{P^v} \in \mathbb{R}^{d_v\times d_v}$, $\widetilde{Q^v} \in \mathbb{R}^{c_v\times c_v}$, and $\widetilde{W^v} \in \mathbb{R}^{c_v\times d_v}$ (see work [28]) and build the following within-view preservation sub-models for vth view and $||\star||_F^2$ represents the Frobenius norm, $\odot$ indicates the element-wise based multiplication operation.

$$\min_{P^v}\sum_{v=1}^{V}||(\widetilde{P^v}-P^v)\odot(E^vE^{vT})||_F^2 \tag{1}$$

$$\min_{Q^v}\sum_{v=1}^{V}||(\widetilde{Q^v}-Q^v)\odot(F^vF^{vT})||_F^2 \tag{2}$$

$$\min_{W^v}\sum_{v=1}^{V}||(\widetilde{W^v}-W^v)\odot(F^vE^{vT})||_F^2 \tag{3}$$

Then since elements in S^v are related with instances, thus S^v-related within-view preservation sub-models are given as below and they are related with E^v and F^v, respectively.

$$\min_{S^v}\sum_{v=1}^{V}||(\widetilde{S^v}-S^v)\odot(E^{vT}E^v)||_F^2 \tag{4}$$

$$\min_{S^v}\sum_{v=1}^{V}||(\widetilde{S^v}-S^v)\odot(F^{vT}F^v)||_F^2 \tag{5}$$

Furthermore, as we said before, the MVML data has three representations including within-view, cross-view, and consensus-view ones. Thus, we let $X^{ij} \in \mathbb{R}^{n\times d_{ij}}$ and $Y^{ij} \in \mathbb{R}^{n\times c_{ij}}$ be the cross-view representations for the sets of instances (including the information-complete instances and information-incomplete ones) and the corresponding labels between ith view and jth view. d_{ij} and c_{ij} are the corresponding feature dimension and number of labels. Now according to the same definitions, $S^{ij} \in \mathbb{R}^{n\times n}$, $P^{ij} \in \mathbb{R}^{d_{ij}\times d_{ij}}$, $Q^{ij} \in \mathbb{R}^{c_{ij}\times c_{ij}}$, $W^{ij} \in \mathbb{R}^{c_{ij}\times d_{ij}}$ are four corresponding cross-view referred matrices and the related extended matrices are $\widetilde{S^{ij}} \in \mathbb{R}^{n\times n}$, $\widetilde{P^{ij}} \in \mathbb{R}^{d_{ij}\times d_{ij}}$, $\widetilde{Q^{ij}} \in \mathbb{R}^{c_{ij}\times c_{ij}}$, and $\widetilde{W^{ij}} \in \mathbb{R}^{c_{ij}\times d_{ij}}$. Similarly, $E^{ij} \in \mathbb{R}^{d_{ij}\times n}$ is the cross-view feature missing-index matrix and $F^v \in \mathbb{R}^{c_{ij}\times n}$ is the cross-view label missing-index matrix between ith view and jth view. Then refer to the Eq. (1)$\sim$Eq. (5), the corresponding cross-view preservation sub-models are given as below.

$$\min_{P^{ij}}\sum_{i=1}^{V}\sum_{j=1}^{V}||(\widetilde{P^{ij}}-P^{ij})\odot(E^{ij}{E^{ij}}^T)||_F^2 \tag{6}$$

$$\min_{Q^{ij}}\sum_{i=1}^{V}\sum_{j=1}^{V}||(\widetilde{Q^{ij}}-Q^{ij})\odot(F^{ij}{F^{ij}}^T)||_F^2 \tag{7}$$

$$\min_{W^{ij}}\sum_{i=1}^{V}\sum_{j=1}^{V}||(\widetilde{W^{ij}}-W^{ij})\odot(F^{ij}{E^{ij}}^T)||_F^2 \tag{8}$$

$$\min_{S^{ij}} \sum_{i=1}^{V} \sum_{j=1}^{V} ||(\widetilde{S^{ij}} - S^{ij}) \odot ({E^{ij}}^T E^{ij})||_F^2 \quad (9)$$

$$\min_{S^{ij}} \sum_{i=1}^{V} \sum_{j=1}^{V} ||(\widetilde{S^{ij}} - S^{ij}) \odot ({F^{ij}}^T F^{ij})||_F^2 \quad (10)$$

Similar, $X^c \in \mathbb{R}^{n \times l}$ and $Y^{ij} \in \mathbb{R}^{n \times k}$ are the consensus-view representations for instances and labels of all views where l is the feature dimension, k is the number of labels in consensus-view representation. Their corresponding referred matrices and extended matrices are $S^c \in \mathbb{R}^{n \times n}$, $P^c \in \mathbb{R}^{l \times l}$, $Q^c \in \mathbb{R}^{k \times k}$, $W^c \in \mathbb{R}^{k \times l}$, $\widetilde{S^c} \in \mathbb{R}^{n \times n}$, $\widetilde{P^c} \in \mathbb{R}^{l \times l}$, $\widetilde{Q^c} \in \mathbb{R}^{k \times k}$, and $\widetilde{W^c} \in \mathbb{R}^{k \times l}$. Now the corresponding consensus-view preservation sub-models are given as below.

$$\min_{P^c} ||(\widetilde{P^c} - P^c) \odot (E^c {E^c}^T)||_F^2 \quad (11)$$

$$\min_{Q^c} ||(\widetilde{Q^c} - Q^c) \odot (F^c {F^c}^T)||_F^2 \quad (12)$$

$$\min_{W^c} ||(\widetilde{W^c} - W^c) \odot (F^c {E^c}^T)||_F^2 \quad (13)$$

$$\min_{S^c} ||(\widetilde{S^c} - S^c) \odot ({E^c}^T E^c)||_F^2 \quad (14)$$

$$\min_{S^c} ||(\widetilde{S^c} - S^c) \odot ({F^c}^T F^c)||_F^2 \quad (15)$$

Finally, we combine all the above 15 sub-models and build a total preservation model as below where $\Theta = \{S^v, P^v, Q^v, W^v, S^{ij}, P^{ij}, Q^{ij}, W^{ij}, S^c, P^c, Q^c, W^c\}$ and η_x is the penalty parameter for Eq. (x) to balance the importance of the corresponding constraints.

$$\min_{\Theta} (\eta_1 Eq.(1) + ... + \eta_{15} Eq.(15)) \quad (16)$$

3.1.2 'within-consensus-cross'-based individual- and mutual- balance model (WCC-IM)

According to subsection 3.1.1, the referred matrices can store comprehensive relationships of all instances ultimately, thus we try to recover the incomplete information with them. In order to recover the incomplete information better, we should consider both the quantity ('individual') and quality ('mutual') of recovering simultaneously. The quantity of recovering means that how many incomplete information has been recovered and the quality of recovering implies that how many improvements can the recovered information brings to the performances.

First, we take X^v and Y^v as example. In order to measure the quantity of recovering, we consider the difference between the recovered versions and the corresponding view-specific data, i.e., X^v and Y^v be small enough so that the available information can be preserved in the recovered versions as far as possible. Then we design the following 'individual'-measured sub-models where $S^v X^v$, $X^v P^v$ and $S^v Y^v$, $Y^v Q^v$ can be regarded as the recovered versions of X^v and Y^v, respectively.

$$\min_{S^v} \sum_{v=1}^{V} ||X^v - S^v X^v||_F^2 \quad (17)$$

$$\min_{P^v} \sum_{v=1}^{V} ||X^v - X^v P^v||_F^2 \quad (18)$$

$$\min_{S^v} \sum_{v=1}^{V} ||Y^v - S^v Y^v||_F^2 \quad (19)$$

$$\min_{Q^v} \sum_{v=1}^{V} ||Y^v - Y^v Q^v||_F^2 \quad (20)$$

Similarly, in order to measure the quality of recovering, we consider the mapping relationship (i.e., W^v) between X^v and Y^v and W^v is always used to map the X^v into Y^v. Then we design the following 'mutual'-measured sub-models where $Y^v W^v$ and $X^v {W^v}^T$ can be regarded as the recovered versions of X^v and Y^v, respectively.

$$\min_{W^v} \sum_{v=1}^{V} ||X^v - Y^v W^v||_F^2 \quad (21)$$

$$\min_{W^v} \sum_{v=1}^{V} ||Y^v - X^v W^{vT}||_F^2 \tag{22}$$

Second, in order to make the recovered information be more feasible, namely, neither too much emphasis on quantity of recovering nor too much emphasis on quality of recovering, we introduce a balance factor $\alpha^v = (\frac{(17)+(18)+(19)+(20)}{(21)+(22)})$ which is related to the above 'individual'-measured sub-models and 'mutual'-measured sub-models. In general, if $\alpha^v = 0.5$, the recovered results are most feasible. Then we can build a 'within'-based individual- and mutual- balance model as below.

$$\min_{S^v,P^v,Q^v,W^v} (\eta_{17} Eq.(17) + ... + \eta_{22} Eq.(22) + \eta_v \alpha^v) \tag{23}$$

Third, similar with subsection 3.1.1, the above 'within'-based individual- and mutual- balance model has its cross-view version (see Eq. (24)).

$$\min_{S^{ij},P^{ij},Q^{ij},W^{ij}} (\eta_{25} Eq.(25) + ... + \eta_{30} Eq.(30) + \eta_{ij} \alpha^{ij}) \tag{24}$$

where

$$\min_{S^{ij}} \sum_{i=1}^{V} \sum_{j=1}^{V} ||X^{ij} - S^{ij} X^{ij}||_F^2 \tag{25}$$

$$\min_{P^{ij}} \sum_{i=1}^{V} \sum_{j=1}^{V} ||X^{ij} - X^{ij} P^{ij}||_F^2 \tag{26}$$

$$\min_{S^{ij}} \sum_{i=1}^{V} \sum_{j=1}^{V} ||Y^{ij} - S^{ij} Y^{ij}||_F^2 \tag{27}$$

$$\min_{Q^{ij}} \sum_{i=1}^{V} \sum_{j=1}^{V} ||Y^{ij} - Y^{ij} Q^{ij}||_F^2 \tag{28}$$

$$\min_{W^{ij}} \sum_{i=1}^{V} \sum_{j=1}^{V} ||X^{ij} - Y^{ij} W^{ij}||_F^2 \tag{29}$$

$$\min_{W^{ij}} \sum_{i=1}^{V} \sum_{j=1}^{V} ||Y^{ij} - X^{ij} W^{ij^T}||_F^2 \tag{30}$$

are 6 sub-models including the 'individual'-measured sub-models and 'mutual'-measured sub-models and $\alpha^{ij} = (\frac{(25)+(26)+(27)+(28)}{(29)+(30)})$.

Fourth, the consensus-view version of above 'within'-based individual- and mutual- balance model is given in Eq. (31).

$$\min_{S^c,P^c,Q^c,W^c} (\eta_{32} Eq.(32) + ... + \eta_{37} Eq.(37) + \eta_c \alpha^c) \tag{31}$$

where

$$\min_{S^c} ||X^c - S^c X^c||_F^2 \tag{32}$$

$$\min_{P^c} ||X^c - X^c P^c||_F^2 \tag{33}$$

$$\min_{S^c} ||Y^c - S^c Y^c||_F^2 \tag{34}$$

$$\min_{Q^c} ||Y^c - Y^c Q^c||_F^2 \tag{35}$$

$$\min_{W^c} ||X^c - Y^c W^c||_F^2 \tag{36}$$

$$\min_{W^c} ||Y^c - X^c W^{cT}||_F^2 \tag{37}$$

are 6 sub-models including the 'individual'-measured sub-models and 'mutual'-measured sub-models and $\alpha^c = (\frac{(32)+(33)+(34)+(35)}{(36)+(37)})$.

Finally, we combine the Eq. (23), Eq. (24), and Eq. (31) in together to build the 'within-consensus-cross'-based individual- and mutual-balance model, i.e., WCC-IM.

$$\min_{\Theta} (Eq.(23) + Eq.(24) + Eq.(31)) \tag{38}$$

3.1.3 Revealing of comprehensive relationships

As what we said before, the referred (completed, comprehensive) matrices can store comprehensive

relationships of all instances ultimately. In order to get the better referred matrices, we can explore the laws of self-adaptive change for multiple correlation information in different representations including within-view, consensus-view, and cross-view ones according to the data characteristics.

We consider the within-view representation firstly. As the summarized of existing work including [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], it is found that the correlations can be described in self-adaptive ways and they change followed by some laws. For example, (i) if two instances are strongly correlated, their corresponding features and the predictive labels might be more similar; (ii) if two labels are strongly correlated, their corresponding outputs might be more similar; (iii) if two features are strongly correlated, their corresponding information of features might be more similar.

Thus in order to realize the first law, we take vth view as example and utilize two regularizer terms. One is related with X^v which includes the information about features, and another is related with $Y^v = X^v W^{vT}$ which indicates the information about predictive labels. In terms of the former regularizer term, if the ath instance of X^v (x_a^v) and the bth instance of X^v (x_b^v) are strongly correlated, the similarity between x_a^v and x_b^v will be large. Similarly, for the latter regularizer term, if x_a^v and x_b^v are strongly correlated, their corresponding predictive labels, i.e., y_a^v and y_b^v will be large as well. Then we can define the following two regularizer terms.

$$\sum_{a=1}^{n}\sum_{b=1}^{n} S_{ab}^v ||x_a^v - x_b^v|| = tr(X^{vT} L_{S^v} X^v) \quad (39)$$

$$\begin{aligned}\sum_{a=1}^{n}\sum_{b=1}^{n} S_{ab}^v ||y_a^v - y_b^v|| &= tr(Y^{vT} L_{S^v} Y^v) \quad (40)\\ &= tr(W^v X^{vT} L_{S^v} X^v W^{vT})\end{aligned}$$

where S_{ab}^v describes the instance-instance correlation between instance x_a^v and instance x_b^v and L_{S^v} is the Laplacian matrix for S^v.

In order to realize the second law, take vth view as an example, we also utilize a regularizer term. Since $Y^v = X^v W^{vT}$ and its pth (or qth) column y_{vp} (or y_{vq}) describes the output of pth (or qth) label. Then we design a corresponding regularizer term as below.

$$\begin{aligned}\sum_{p=1}^{c_v}\sum_{q=1}^{c_v} Q_{pq}^v ||y_{vp} - y_{vq}|| &= tr(Y^v L_{Q^v} Y^{vT}) \quad (41)\\ &= tr(X^v W^{vT} L_{Q^v} W^v X^{vT})\end{aligned}$$

where Q_{pq}^v describes the label-label correlation between pth and qth labels. Similar with L_{S^v}, L_{Q^v} is the Laplacian matrix for Q^v.

In order to realize the third law, we still utilize a regularizer term for vth view. Since rth (or sth) column x_{vr} (or x_{vs}) describes the information of rth (or sth) feature. Then we design a corresponding regularizer term as below.

$$\sum_{r=1}^{d_v}\sum_{s=1}^{d_v} P_{rs}^v ||x_{vr} - x_{vs}|| = tr(X^v L_{P^v} X^{vT}) \quad (42)$$

where P_{rs}^v describes the feature-feature correlation between rth and sth features. Similar with L_{S^v}, L_{P^v} is the Laplacian matrix for P^v.

Then referring to subsection 3.1.1, these regularizer terms can also be migrated to cross-view form (see Eq. (43)$\sim$ Eq. (46)) and consensus-view form (see Eq. (47)$\sim$ Eq. (50)) where $\Rightarrow$ describes the migration operation and $L_{S^{ij}}$, $L_{Q^{ij}}$, $L_{P^{ij}}$, L_{S^c}, L_{Q^c}, L_{P^c} are the Laplacian matrices for the corresponding matrices.

$$tr(X^{vT} L_{S^v} X^v) \Rightarrow tr(X^{ijT} L_{S^{ij}} X^{ij}) \quad (43)$$

$$\begin{aligned}&tr(W^v X^{vT} L_{S^v} X^v W^{vT}) \Rightarrow \quad (44)\\ &tr(W^{ij} X^{ijT} L_{S^{ij}} X^{ij} W^{ijT})\end{aligned}$$

$$\begin{aligned}&tr(X^v W^{vT} L_{Q^v} W^v X^{vT}) \Rightarrow \quad (45)\\ &tr(X^{ij} W^{ijT} L_{Q^{ij}} W^{ij} X^{ijT})\end{aligned}$$

$$tr(X^v L_{P^v} X^{vT}) \Rightarrow tr(X^{ij} L_{P^{ij}} X^{ijT}) \quad (46)$$

$$tr(X^{vT} L_{S^v} X^v) \Rightarrow tr(X^{cT} L_{S^c} X^c) \quad (47)$$

$$\begin{aligned}&tr(W^v X^{vT} L_{S^v} X^v W^{vT}) \Rightarrow \quad (48)\\ &tr(W^c X^{cT} L_{S^c} X^c W^{cT})\end{aligned}$$

$$tr(X^v W^{vT} L_{Q^v} W^v X^{vT}) \Rightarrow tr(X^c W^{cT} L_{Q^c} W^c X^{cT}) \quad (49)$$

$$tr(X^v L_{P^v} X^{vT}) \Rightarrow tr(X^c L_{P^c} X^{cT}) \quad (50)$$

Followed by above laws, the correlations can be expressed self-adaptively and we can get a self-adaptive model as below.

$$\min_{\Theta}(\eta_{39} Eq.(39) + ... + \eta_{50} Eq.(50)) \quad (51)$$

3.1.4 Correlations imposed with nuclear-norm

Since low-rank characteristic of representation can be enforced by matrix nuclear norm [29, 30], so we represent each correlation with such norm and build three nuclear-norm-based sub-models, namely, $||\theta^v||_*^2$, $||\theta^c||_*^2$, $||\theta^{ij}||_*^2$ where

$$||\theta^v||_*^2 = \sum_{C \in S,P,Q,W} \eta_{C^v} ||C^v||_*^2 \quad (52)$$

$$||\theta^{ij}||_*^2 = \sum_{C \in S,P,Q,W} \eta_{C^{ij}} ||C^{ij}||_*^2 \quad (53)$$

$$||\theta^c||_*^2 = \sum_{C \in S,P,Q,W} \eta_{C^c} ||C^c||_*^2 \quad (54)$$

Then the final nuclear-norm-based model is given as below.

$$\min_{\Theta}(Eq.(52) + Eq.(53) + Eq.(54)) \quad (55)$$

3.1.5 Final model of MVML-IDSaC

According to the above mentioned, the final model of MVML-IDSaC is the combination of four models, i.e., Eq. (16), Eq. (38), Eq. (51), and Eq. (55) which is given as below.

$$\min_{\Theta}(Eq.(16) + Eq.(38) + Eq.(51) + Eq.(55)) \quad (56)$$

3.2 Optimization

For the optimization of Eq. (56), we also refer to [28] and gradient descent way is adopted. Namely, we update a correlation and leave the others fixed,

Table 1 Computation of $\frac{\partial \mathcal{L}}{\partial \mathcal{C}}$ in Eq. (56) in 9 forms and $\mathcal{A}$, $\mathcal{B}$, $\mathcal{D}$ describe different terms.

Form	Computational results
$\frac{\partial ||(\mathcal{A}-\mathcal{C})\odot\mathcal{B}||_F^2}{\partial \mathcal{C}}$	$2(\mathcal{C}\odot\mathcal{B}\odot\mathcal{B}^T - \mathcal{A}\odot\mathcal{B}\odot\mathcal{B}^T)$
$\frac{\partial ||\mathcal{A}-\mathcal{B}\mathcal{C}||_F^2}{\partial \mathcal{C}}$	$2(\mathcal{B}^T\mathcal{B}\mathcal{C} - \mathcal{B}^T\mathcal{A})$
$\frac{\partial ||\mathcal{A}-\mathcal{B}\mathcal{C}^T||_F^2}{\partial \mathcal{C}}$	$2(\mathcal{C}\mathcal{B}^T\mathcal{B} - \mathcal{A}^T\mathcal{B})$
$\frac{\partial ||\mathcal{A}-\mathcal{B}\mathcal{C}\mathcal{D}||_F^2}{\partial \mathcal{C}}$	$2(\mathcal{B}^T\mathcal{B}\mathcal{C}\mathcal{D}\mathcal{D}^T - \mathcal{B}^T\mathcal{A}\mathcal{D}^T)$
$\frac{\partial ||\mathcal{A}-\mathcal{B}\mathcal{C}\mathcal{C}^T\mathcal{D}||_F^2}{\partial \mathcal{C}}$	$-2(\mathcal{D}\mathcal{A}^T\mathcal{B}\mathcal{C} + \mathcal{B}^T\mathcal{A}\mathcal{D}^T\mathcal{C}) + 4\mathcal{B}^T\mathcal{B}\mathcal{C}\mathcal{C}^T\mathcal{D}\mathcal{D}^T\mathcal{C}$
$\frac{\partial ||\mathcal{C}||_*^2}{\partial \mathcal{C}}$	$\mathcal{C}_U\mathcal{C}_V^T$
$\frac{\partial tr(\mathcal{A}L_{\mathcal{C}}\mathcal{A}^T)}{\partial \mathcal{C}}$	$\frac{1}{2}[\mathcal{E}]_{\mathcal{C}}$
$\frac{\partial tr(\mathcal{C}\mathcal{A}\mathcal{C}^T)}{\partial \mathcal{C}}$	$\mathcal{C}\mathcal{A}^T + \mathcal{C}\mathcal{A}$
$\frac{\partial tr(\mathcal{A}\mathcal{C}^T\mathcal{B}\mathcal{C}\mathcal{A}^T)}{\partial \mathcal{C}}$	$\mathcal{B}^T\mathcal{C}\mathcal{A}^T\mathcal{A} + \mathcal{B}\mathcal{C}\mathcal{A}\mathcal{A}^T$

i.e., $\mathcal{C}(t) \leftarrow \mathcal{C}(t-1) - \frac{\partial \mathcal{L}(t-1)}{\partial \mathcal{C}(t-1)}$ where $\mathcal{C}$ describes a correlation matrix and t represents the iteration index. According to Eq. (56), the $\frac{\partial \mathcal{L}}{\partial \mathcal{C}}$ has multiple forms (see Table 1) and singular value decomposition of $\mathcal{C} \in \mathbb{R}^{n_1 \times n_2}$ of rank r is $\mathcal{C} = \mathcal{C}_U \mathcal{C}_\Sigma \mathcal{C}_V^*$ where $\mathcal{C}_\Sigma = diag(\{\sigma_k\}_{k \le i \le r})$, $\mathcal{C}_U$ and $\mathcal{C}_V$ are respectively $n_1 \times r$ and $n_2 \times r$ matrices with orthonormal columns, and the singular values σ_k are positive. Furthermore, $[\mathcal{E}]_{\mathcal{C}}$ has the same dimensionality of $\mathcal{C}$ and its pth row and qth column element is $||\mathcal{A}_{p,:}^T - \mathcal{A}_{q,:}^T||^2$ where $\mathcal{A}_{p,:}$ and $\mathcal{A}_{q,:}$ stand for the pth and qth rows of $\mathcal{A}$, respectively. The optimization procedure will be terminate until the changes about the normalized value of $\mathcal{L}$ is lesser than some threshold values.

3.3 Computational cost

According to the above contents, the computations of the correlations determine the computational cost of MVML-IDSaC. Then with the computation of these correlations on the basis of Table 1, it is found the maximum computational cost for a correlation is $DO(Cn^2)$ where C and D are two constants and this causes the total computational cost of MVML-IDSaC is less than $NDO(Cn^2)$ where N is the number of correlations. Compared with some existing algorithms [31, 32, 33, 34] whose computational costs are $O(n^3)$-levels, the computational cost of MVML-IDSaC is much smaller.

4 Experiments

In order to validate the effectiveness of the developed MVML-IDSaC, we adopt 5 real-world

MVML data sets for experiments and answer the following questions.

1) How effective is MVML-IDSaC compared with other related algorithms in process MVML data sets?

2) What is the impact of those multiple correlations and related terms on MVML-IDSaC?

4.1 Experimental setup

Data setting: we refer to [28] and also adopt 5 MVML data sets for experiments (see Table 2). Their detailed information can be found in [28]. Moreover, where to download them are also given.

Compared algorithms: to study the performance of MVML-IDSaC, we compare it with 3 state-of-the-art MVML algorithms including NAIM^{3}L [9], iMVMAL [12], DD-IMvMLC-net [13].

Parameter setting: for the compared algorithms, the selection of corresponding optimal parameter values can be referred to the original papers. For MVML-IDSaC, the elements in any variable $A \in \Theta$ where $\Theta = \{S^v, P^v, Q^v, W^v, S^{ij}, P^{ij}, Q^{ij}, W^{ij}, S^c, P^c, Q^c, W^c\}$ are initialized in equipartition and updated according to subsection 3.2. For penalty parameters ηs, their optimal parameter values can be selected from the set $\{0.1, 0.2, ..., 0.8, 0.9\}$. For the incomplete setting, the missing rates of features and labels are 50% and we remove some information so as to obtain this rate.

Evaluation: we use AUC and precision as the main metrics for experimental comparisons and some further comparison including convergence and training time are also given.

Selection of optimal parameters: there are some tuning parameters in MVML-IDSaC. In order to select the optimal parameter values, avoid over-fitting in model selection, and ensure the authenticity of experimental results, we adopt the same way given in [28]. In simple speaking, for each data set, we select 70% instances for training and validation and the rest for testing. Then the optimal parameter values can be selected with ten-fold cross-validation (Here, AUC is adopted as the metric to select the optimal parameters). The operations to select optimal parameters are repeated for five times independently and the average performances and the corresponding standard deviations of algorithms on the used data sets are also reported.

Experimental environment: the operation system is RedHat Linux Enterprise 9.0, the processor is Intel Core i7-12700 (12 CPUs), and the coding environment is MATLAB 2022b.

4.2 Performance and efficiency study

Here, we use Fig. 2 to demonstrate the AUC, precision, and corresponding standard deviations of all algorithms on all data sets. Then Fig. 3 demonstrates the training time comparison results. Moreover, Fig. 4 demonstrates the convergence of MVML-IDSaC on all used data sets. In this figure, left sub-figure shows the convergence curve and the right sub-figure reports the iteration index[1].

According to these figures, it can be found that in terms of AUC and precision, although the performance of MVML-IDSaC is not the best stable, it still outperforms the other algorithms in most cases with a relatively low training time kept. Moreover, although there are many terms are considered in the model of MVML-IDSaC, it can get a convergence within 30 iterations in our experiments.

4.3 Statistical analysis

To validate the effectiveness of the MVML-IDSaC further, we adopt Friedman-Nemenyi statistical test [35] for statistical analysis and check if the differences between MVML-IDSaC and other compared algorithms are significant or not. Here, for convenience, we use AUC and precision for statement.

(1) According to the AUC results, we demonstrate the average ranks of all used algorithms, rank differences between MVML-IDSaC and others, and corresponding statistical values (see Fig. 5-(a) and Fig. 5-(b)).

Then refer to [35], (i) we carry out Friedman test at first. Since we adopt 5 data sets and 4 algorithms (i.e., $N = 5$ and $k = 4$) for

[1]Iteration index stands for the index when the changes of normalized objective value is smaller than 0.01.

Table 2 Summarized information of used data sets where ALPI indicates Avg. label per instance .

data sets	No. instances	No. labels	ALPI	URL
Corel5k	4999	260	3.396	http://archive.ics.uci.edu/ml/datasets/Corel +Image+Features
Espgame	20770	268	4.686	https://www.kaggle.com/datasets/parhamsalar/espgame
IAPRTC12	19627	291	5.719	https://www.imageclef.org/photodata
Mirflickr	25000	38	4.716	https://press.liacs.nl/mirflickr/
Pascal07	9963	20	1.465	http://host.robots.ox.ac.uk/pascal/VOC/

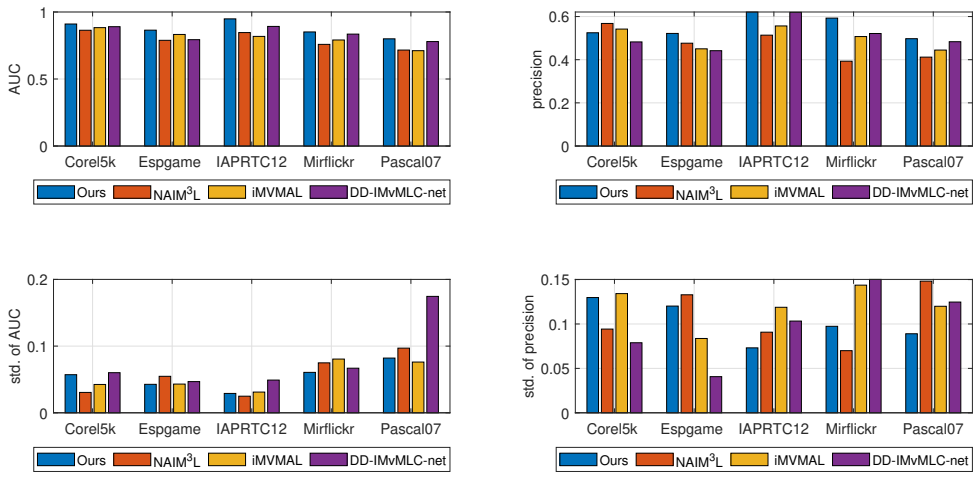

Fig. 2 AUC, precision, and corresponding standard deviation (std.) comparisons.

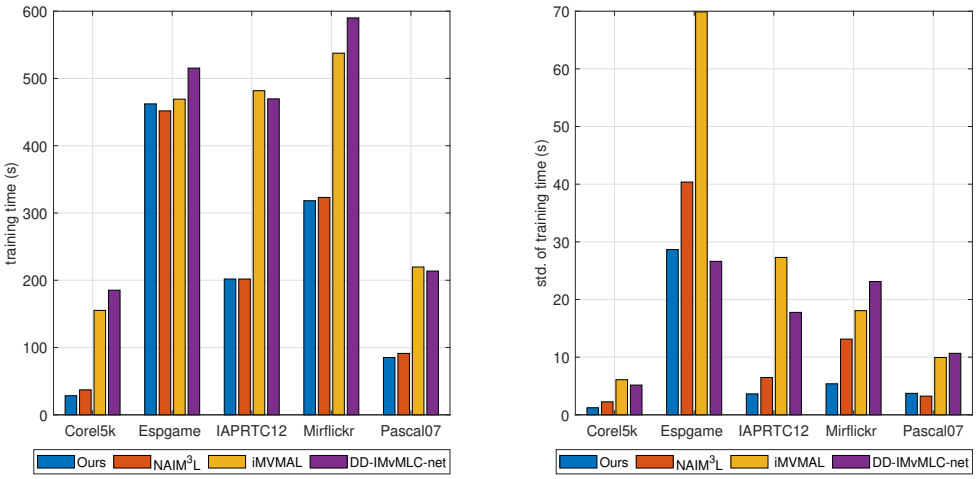

Fig. 3 Training time and the corresponding standard deviation (std.) comparisons.

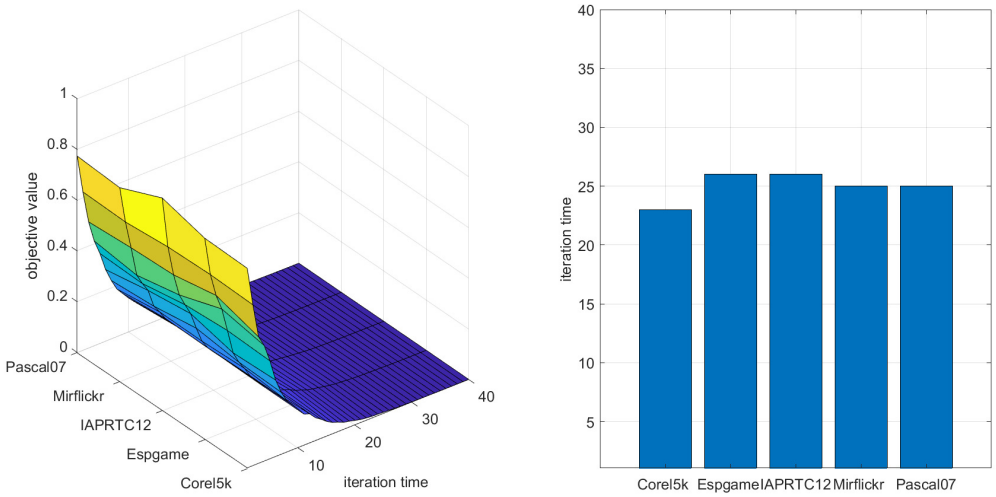

Fig. 4 Convergence of MVML-IDSaC on the used data sets and the objective value has been normalized.

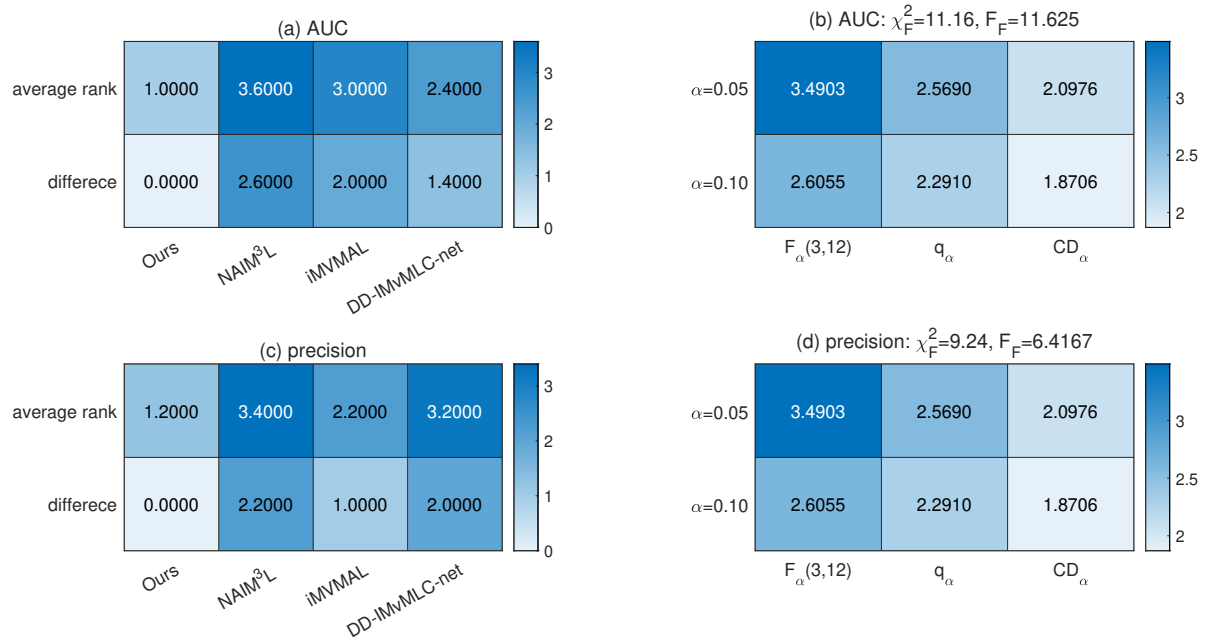

Fig. 5 Statistical analysis for MVML-IDSaC in terms of AUC and precision.

experiments, then we can get Friedman statistic as follows. $\chi_F^2 = \frac{12\times N}{k(k+1)}[1.0000^2 + 3.6000^2 + 3.0000^2 + 2.4000^2 - \frac{k(k+1)^2}{4}] = 11.16$, $F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} = 11.625$, $F_{0.05}(k-1,(k-1)(N-1)) = F_{0.05}(3,12) = 3.4903$, and $F_{0.10}(k-1,(k-1)(N-1)) = F_{0.10}(3,12) = 2.6055$. Since $F_F > F_{0.05}(3,12)$ and $F_F > F_{0.10}(3,12)$, so we reject the null-hypothesis and draw a conclusion that the differences between all compared algorithms on multiple data sets are significant. (ii) Then we carry out Nemenyi test for pairwise comparisons. Since $N = 5$ and $k = 4$, thus critical value at $q_{0.05}$ is 2.5690 and corresponding critical difference (CD) is $CD_{0.05} = q_{0.05}\sqrt{\frac{k\cdot(k+1)}{6\cdot N}} = 2.0976$ while the one at $q_{0.10}$ is 2.2910 and corresponding CD

is $CD_{0.10} = q_{0.10}\sqrt{\frac{k\cdot(k+1)}{6\cdot N}} = 1.8706$. Since under the case of $CD_{0.05}$ and $CD_{0.10}$, rank differences between MVML-IDSaC and NAIM^{3}L are larger than $CD_{0.05}$ and $CD_{0.10}$, so we say on this case, the performance of MVML-IDSaC is significant better than $NAIM^3L$. Similar, MVML-IDSaC is significant better than iMVMAL when $\alpha = 0.10$.

(2) Similarly, Fig. 5-(c) and Fig. 5-(d) demonstrates the statistical analysis about MVML-IDSaC on the precision. According to this figure, in terms of precision, our MVML-IDSaC is significant better than NAIM^{3}L as well when $\alpha = 0.05$ and $\alpha = 0.10$ and is significant better DD-IMvMLC-net when $\alpha = 0.10$.

In general, our MVML-IDSaC performs best as demonstrated by statistical tests, especially compared with NAIM^{3}L.

4.4 Parameter study

According to the objective function of MVML-IDSaC (i.e., Eq. (56)), there are 60 parameters can be adjusted and different parameter values lead to diverse average AUC, precision, training time, and convergence. In order to validate the influence of parameters to the average performances vary, we take data set Corel5k for statement (for other data sets, the conclusions are similar) and the influence can be found in Fig. 6.

From this figure, it is found that if we want to get a better AUC or precision, (*i*) for terms about balance factors which are related to the 'individual'-measured sub-models and 'mutual'-measured sub-models, we can set the corresponding penalty parameters be 0.8 or 0.9. This indicates that the balance factors are more important to balance the difference between 'individual'-measured sub-models and 'mutual'-measured sub-models; (*ii*) for terms about nuclear-norm-based sub-models, they have little effects on the performance and we can set their corresponding penalty parameters be in random; (*iii*) for other terms about within-view parts, penalty parameters can be set as 0.4 or 0.5; for other terms about consensus-view parts, penalty parameters can be set as 0.5 or 0.6; for other terms about cross-view parts, penalty parameters can be set as $0.6 \sim 0.8$. This conclusion indicates that compared with within-view information and consensus-view information, the cross-view information is more important.

To this end, in the future experiments, we can set the parameters referring to the above conclusions.

4.5 Ablation study

As we know, the model of MVML-IDSaC consists of many terms and we want to see which term is essential. To this end, we carry out ablation study. In simple speaking, we set a penalty parameter be 0 which equals to removing the corresponding term and see the average vary of performances about AUC and precision on all used data sets. In this subsection, we adopt data set Mirflickr for statement and for other data sets, the conclusions are similar. Results are given in Fig. 7 and in this figure, curve 'best'/'worst' stands for the best/worst performances when we adjust the penalty parameter values while curve 'ablation' stands for the performances when we set the penalty parameter values be 0, namely, removing the corresponding terms.

According to this figure, it can be seen that removing terms about balance factors which are related to the 'individual'-measured sub-models and 'mutual'-measured sub-models brings a greater reduction to the performances and this validates the essential of WCC-IM, especially for the within-view one and the cross-view one. Moreover, removing terms about revealing of the comprehensive relationships also bring some reduction to the performances to a certain extent. This also means that considering the comprehensive relationships about MVML data can improve the ability of algorithms to process MVML data sets effectively.

5 Conclusion and future work

To process multi-view multi-label data sets more effectively, this study develops a multi-view multi-label learning with incomplete data and self-adaptive correlations (MVML-IDSaC). Different from state-of-the-art algorithms, the MVML-IDSaC considers correlations among different instances, features, labels in consensus-view, within-view, and cross-view representations and explores the laws of self-adaptive change for these correlations. Moreover, the developed algorithm can process incomplete MVML data well with a 'within-consensus-cross'-based individual-

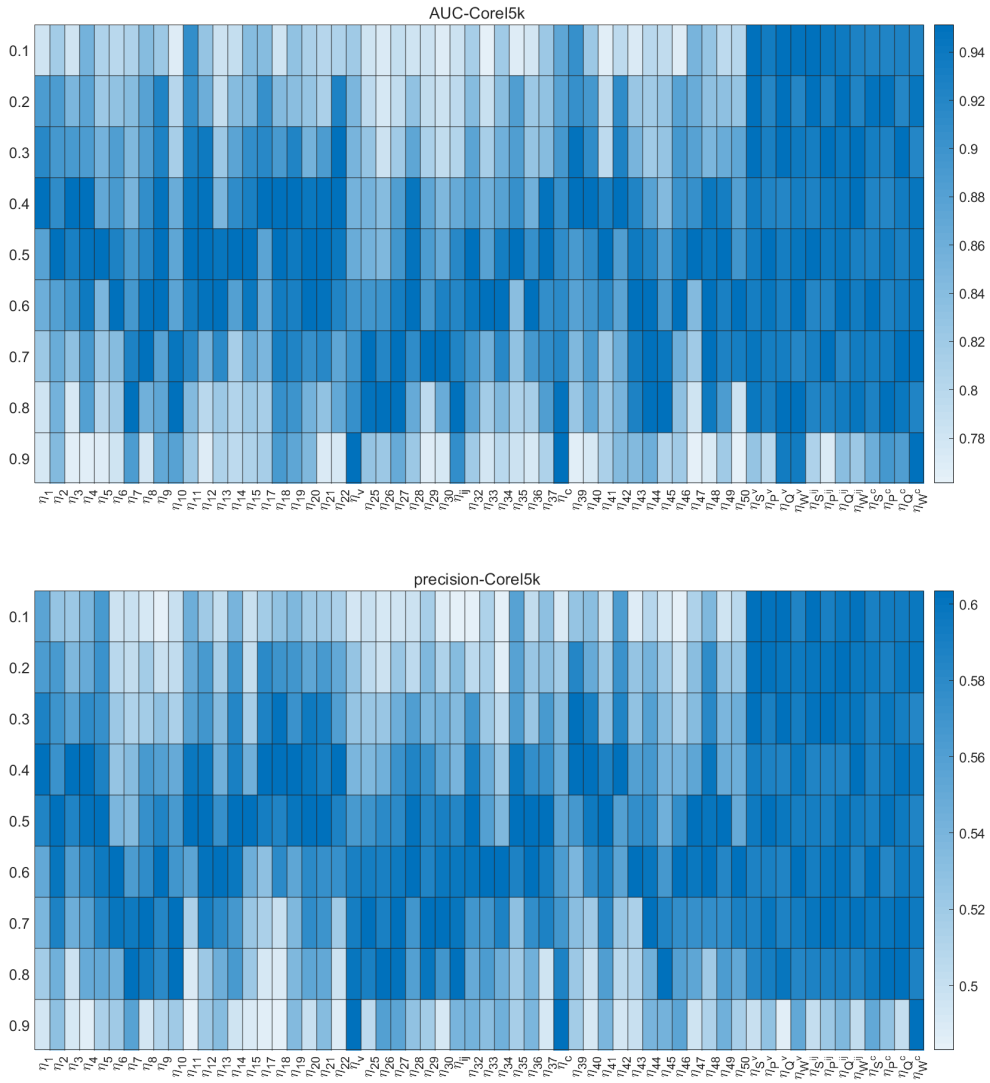

Fig. 6 Average influence of parameter values for MVML-IDSaC on the used data set Corel5k of AUC and precision.

and mutual- balance model introduced. Experiments completed on several benchmark data sets demonstrate the superiority of MVML-IDSaC over related competitive ones.

With the complexity of tasks and the arrival of big data era, traditional sampling equipments have no ability to capture data information in real time and this leads to an online study problem which also affect the performances of MVML-IDSaC. Thus, in the future, we plan to design a new algorithm on the basis of MVML-IDSaC to solve these online study problems.

Acknowledgments and funding

This work is supported by National Natural Science Foundation of China (CN) under grant numbers 62276164, 'Science and technology innovation action plan' Natural Science Foundation of Shanghai under grant number 22ZR1427000. Furthermore, this work is also sponsored by Shanghai Oriental Talent Program-Youth Program. The authors would like to thank their supports.

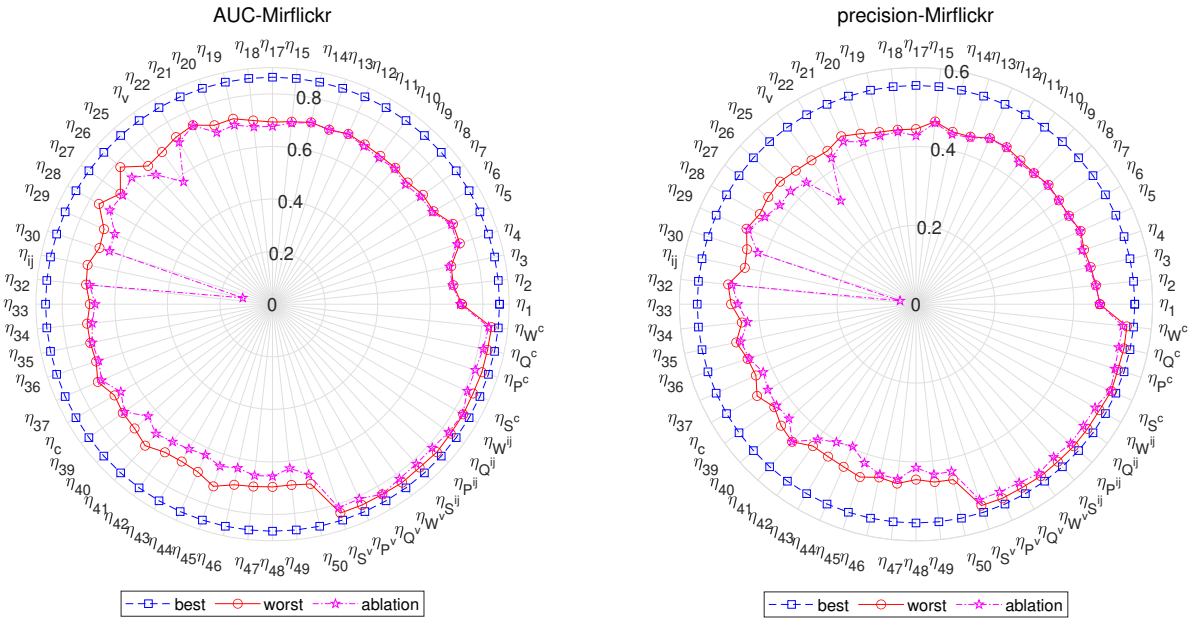

Fig. 7 Ablation study for MVML-IDSaC on data set Mirflickr in terms of AUC and precision where only one term removed is considered here.

References

[1] Tan QY, Yu GX, Wang J, Domeniconi C, Zhang XL (2021) Individuality- and commonality-based multiview multilabel learning. IEEE Transactions on Cybernetics 51(3): 1716-1727.

[2] Luo Y, Liu TL, Tao DC, Xu C (2015) Multi-view matrix completion for multilabel image classification. IEEE Transactions on Image Processing 24(8): 2355-2368.

[3] Zhang YB, Lian HR, Yang G, Zhao SY, Ni P, Chen H, Li CP (2023) Inaccurate-supervised learning with generative adversarial nets. IEEE Transactions on Cybernetics 53(3): 1522-1536.

[4] Zhang YS, Wu J, Cai ZH, Yu PS (2020) Multi-view multi-label learning with sparse feature selection for image annotation. IEEE Transactions on Multimedia 22(11): 2844-2857.

[5] Yu GX, Xing YY, Wang J, Domeniconi C, Zhang XL (2022) Multiview multi-instance multilabel active learning. IEEE Transactions on Neural Networks and Learning Systems 33(9): 4311-4321.

[6] Lai Q, Vong CM, Zhou JH, Zhou YM, Chen CLP (2023) Fast broad multiview multi-instance multilabel learning (FBM3L) with viewwise intercorrelation. IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2023.3286876.

[7] Zhang Y, Shen JD, Yu C, Wang CJ (2022) Views meet labels: personalized relation refinement network for multiview multilabel learning. IEEE MultiMedia 29(2): 104-113.

[8] Zhang LH, Wang L, Bai ZJ, Li RC (2022) A Self-consistent-field iteration for orthogonal canonical correlation analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(2): 890-904.

[9] Li X, Chen SC (2022) A concise yet effective model for non-aligned incomplete multi-view and missing multi-label learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(10: 5918-5932.

[10] Maeda K, Takahashi S, Ogawa T, Haseyama M (2020) Feature integration via geometrical supervised multi-view multi-label canonical correlation for incomplete label assignment. in 2020 IEEE International Conference on Image Processing(ICIP), Abu Dhabi, United Arab Emirates, pp 46-50.

[11] Qian BY, Wang X, Ye JP, Davidson I (2015) A reconstruction error based framework for multi-label and multi-view learning. IEEE

Transactions on Knowledge and Data Engineering 27(3): 594-607.

[12] Qu CW, Wang KM, Zhang H, Yu GX, Domeniconi C (2021) Incomplete Multi-view Multi-label Active Learning. in 2021 IEEE International Conference on Data Mining(ICDM), Auckland, New Zealand, pp 1294-1299.

[13] Wen J, Liu CL, Deng SJ, Liu YC, Fei LK, Yan K, Xu Y (2023) Deep double incomplete multi-view multi-label learning with incomplete labels and missing views. IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2023.3260349.

[14] Xue XW, Nie FP, Li ZH, Wang S, Li X, Yao M (2018) A multiview learning framework with a linear computational cost. IEEE Transactions on Cybernetics 48(8): 2416-2425.

[15] Niu XY, Zhang CW, Ma Y, Hu LH, Zhang JF (2023) A multi-view subspace representation learning approach powered by subspace transformation relationship. Knowledge-Based Systems 277, 110816.

[16] Lou CL, Xie XJ (2023) Multi-view intuitionistic fuzzy support vector machines with insensitive pinball loss for classification of noisy data. Neurocomputing 549, 126458.

[17] He WJ, Zhang Z, Wei YH (2023) Scalable incomplete multi-view clustering with adaptive data completion. Information Sciences, 119562. https://doi.org/10.1016/j.ins.2023.119562.

[18] Sun LL, Wen J, Liu CL, Fei LK, Li LS (2023) Balance guided incomplete multi-view spectral clustering. Neural Networks 166: 260-272.

[19] Mu JS, Song P, Yu YW, Zheng WM (2023) Tensor-based consensus learning for incomplete multi-view clustering. Expert Systems with Applications 234, 121013.

[20] Sun KW, Lee CH, Wang J (2016) Multilabel classification via co-evolutionary multilabel hypernetwork. IEEE Transactions on Knowledge and Data Engineering 28(9): 2438-2451.

[21] Che XY, Chen DG, Mi JS (2023) Learning instance-level label correlation distribution for multilabel classification with fuzzy rough sets. IEEE Transactions on Fuzzy Systems 31(8): 2871-2884.

[22] Zhai TT, Wang H (2022) Online passive-aggressive multilabel classification algorithms. IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2022.3164906.

[23] Li YQ, Wu BY, Ghanem B, Zhao YP, Yao HX, Ji Q (2016) Facial action unit recognition under incomplete data based on multi-label learning with missing labels. Pattern Recognition 60: 890-900.

[24] Sun LJ, Ye P, Lyu GY, Feng SH, Dai GJ, Zhang H (2020) Weakly-supervised multi-label learning with noisy features and incomplete labels. Neurocomputing 413: 61-71.

[25] Beigaite R, Read J, Zliobaite I (2022) Multi-output regression with structurally incomplete target labels: A case study of modelling global vegetation cover. Ecological Informatics 72, 101849.

[26] Zhu XF, Li XL, Zhang SC (2016) Block-row sparse multiview multilabel learning for image classification. IEEE Transactions on Cybernetics 46(2): 450-461.

[27] Ma JH, Sun HY, Zhu T (2023) Multiview multilabel classification with group-based feature and label selection. IEEE Transactions on Consumer Electronics. https://doi.org/10.1109/TCE.2023.3278457.

[28] Zhu CM, Liu YC, Miao DQ, Dong YL, Pedrycz W (2023) Within- cross- consensus-view representation-based multi-view multi-label learning with incomplete data. Neurocomputing 557, 126729.

[29] Zhou T, Zhang CQ, Peng X, Bhaskar H, Yang J (2020) Dual shared-specific multiview subspace clustering. IEEE Transactions on Cybernetics 50(8): 3517-3530.

[30] Peng X, Lu CY, Yi Z, Tang HJ (2018) Connections between nuclear-norm and frobenius-norm-based representations. IEEE Transactions on Neural Networks and Learning Systems 29(1): 218-224.

[31] Hajjar SE, Dornaika F, Abdallah F (2022) One-step multi-view spectral clustering with cluster label correlation graph. Information Sciences 592: 97-111.

[32] Cui GS, Li Y (2022) Nonredundancy regularization based nonnegative matrix factorization with manifold learning for multiview data representation. Information Fusion 82: 86-98.

[33] Xie Y, Liu JY, Qu YY, Tao DC, Zhang WS, Dai LQ, Ma LZ (2021) Robust kernelized multiview self-representation for subspace clustering. IEEE Transactions on Neural Networks and Learning Systems 32(2): 868-881.

[34] Zhang CQ, Fu HZ, Hu QH, Cao XC, Xie Y, Tao DC, Xu D (2020) Generalized latent multi-view subspace clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 42(1): 86-99.

[35] Demsar J (2006) Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7(1): 1-30.