Neurocomputing 557 (2023) 126729

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Within- cross- consensus-view representation-based multi-view multi-label learning with incomplete data

Changming Zhu [a,*], Yanchen Liu [a], Duoqian Miao [b], Yilin Dong [a], Witold Pedrycz [c,d]

[a] College of Information Engineering, Shanghai Maritime University, Shanghai, China
[b] Department of Computer Science and Technology, Tongji University, Shanghai, China
[c] Department of Electrical & Computer Engineering, Alberta University, Canada
[d] Istinye University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Sariyer/Istanbul, Turkiye

ARTICLE INFO

Communicated by C. Hou

Keywords:
Within-view representation
Cross-view representation
Consensus-view representation
Multi-view multi-label
Incomplete data

ABSTRACT

This article develops a multi-view multi-label learning for incomplete data which are ubiquitous with the usage of three kinds of representations including within-view representation, cross-view representation, and consensus-view representation. Different from the recent learning machines, the proposed learning machine takes the feature-oriented information, label-oriented information, and associated information between features and labels in multiple representations together and exploits the hidden useful information of available instances with the usage of instance–instance correlations, feature–feature correlations, label-label correlations, and feature–label correlations. The developed learning machine is named as within- cross- consensus-view representation-based multi-view multi-label learning with incomplete data (WCC-MVML-ID). Extensive experiments on multiple multi-view and multi-label data sets with incomplete data validate the effectiveness of WCC-MVML-ID and it can be concluded that (1) WCC-MVML-ID outperforms other compared learning machines and its performances are more stable even though the missing rates of features and labels being larger; (2) compared with within-view information and consensus-view information, cross-view information is more useful for the processing problem about incomplete data; (3) WCC-MVML-ID can converge within 45 iterations.

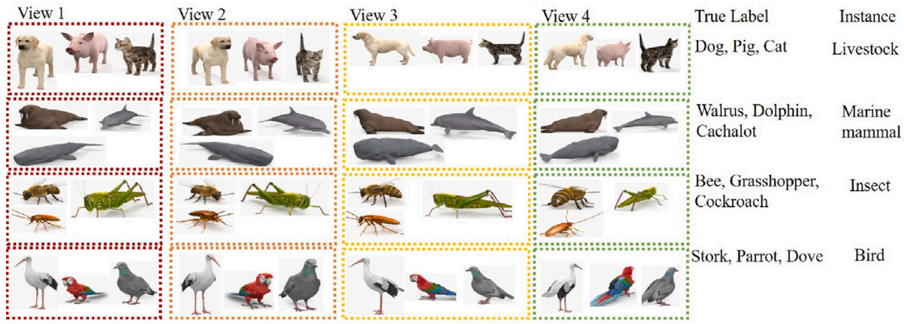
1. Introduction

Multi-view multi-label (MVML) data sets are ubiquitous in current real-world applications [1–4]. Different from traditional multi-view data sets [5–8] and multi-label data sets [9,10], a MVML instance can be represented by multiple features coming from different views and tagged by multiple class labels. Take Fig. 1-(a) as an example. There is an animal data set consisting of four instances, 'Livestock', 'Marine mammal', 'Insect', and 'Bird'. Each instance shows some animals and the instance can be labeled with multiple class labels. In addition, these animals can be demonstrated with four views where different views represent features from diversity shooting angles. In terms of such a MVML data set, traditional multi-view learning and multi-label learning maybe have no ability to process and a corresponding solution on the base of information about views and labels is subsequently proposed. This solution is named multi-view multi-label learning and it is more feasible for processing MVML data. We take the data set given in Fig. 1-(a) for experiment to validate that. In order to process this data set, we select three traditional solutions. One is multiple-view multiple-learner semi-supervised learning machine (MVMLSS) which is developed for multi-view data [11], another is multi-label classification machine with hierarchical embedding (MLCHE) which aims to process multi-label data [12], the third is multi-view based multi-label propagation (MVMLP) [13] which is a MVML solution. Suppose the selected three solutions have been fully trained, then according to Fig. 1-(b), it is found that for each instance, MVMLSS can only recognize one label and MLCHE cannot recognize all labels. In addition, sometimes MVMLSS and MLCHE give new labels unexpectedly (see the red words in Fig. 1-(b)). The main reason is that MVMLSS cannot train multiple labels of an instance simultaneously while MLCHE have to mix multiple features coming from different views in together and such an operation leads to the loss of some priori knowledge. That is why the MVML learning should be researched.

Moreover, for such a MVML data set, the instances, features, labels have three representations at least including the within-view representation, cross-view representation, and consensus-view representation. The within-view representation captures the information of data in

* Corresponding author.
E-mail addresses: cmzhu@shmtu.edu.cn (C. Zhu), 202230310107@stu.shmtu.edu.cn (Y. Liu), dqmiao@tongji.edu.cn (D. Miao), yldong@shmtu.edu.cn (Y. Dong), wpedrycz@ualberta.ca (W. Pedrycz).

https://doi.org/10.1016/j.neucom.2023.126729
Received 13 November 2022; Received in revised form 27 May 2023; Accepted 20 August 2023
Available online 25 August 2023



(a)

Instance	MVMLSS	MLCHE	MVMLP
Livestock	Dog	Pig, Pet	Dog, Pig, Cat
Marine mammal	Dolphin	Walrus, Cachalot	Walrus, Dolphin, Cachalot
Insect	Cockroach	Bee, Cockroach	Bee, Grasshopper, Cockroach
Bird	Hawk	Parrot	Stork, Parrot, Dove

(b)

Fig. 1. Representation of a MVML animal data set and a simple experiment to validate the effectiveness of a MVML learning.

each view, the cross-view representation demonstrates the shared information and information source of two different views, and the consensus-view representation contains the consensus information and information source of all different views.

In order to process those data sets, scholars develop many MVML solutions based on different representations. For example, Cheng et al. develop a MVML learning framework with view feature attention allocation and the within-view and consensus-view representations are obtained [14]; Zhao et al. learn view complementarity information and label consistency information in view-specific labels and propose a view-specific label learning method in MVML to solve the problem of non-aligned views [15]. They also develop a MVML feedforward neural network model which considers consensus-view representation and within-view representation information in MVML learning and use a similar ensemble learning method to construct the final model [16]; Zhao et al. propose a multi-view partial label machine based on maximum margin partial label learning so as to ensure the label consensus of multiple views and they employ a heuristic optimization framework to identify the ground-truth label and optimize the multi-class margin alternately [17].

Those solutions cannot process incomplete data. As is known, with the limitation of manpower and some other objective factors including equipment errors, data maybe miss some information about features and labels. We take Fig. 2 as an example. In this figure, an image instance with three animals is given and this image can be described with four views and labeled with three class labels. While with some objective factors, some information about features and labels is lost and we mark them with question marks. For such incomplete data, some learning methods are developed. For example, Qu et al. propose an incomplete MVML active learning approach to reduce the cost of querying MVML data with the learning of shared/individual representations of instances across/within incomplete views by an indicator matrix to indicate the missing instance of respective view [18]; Qian et al. develop a semi-supervised dimension reduction for multi-label and multi-view learning which performs optimization for dimension reduction and label inference in semi-supervised setting and address high dimensional with multiple (but possibly incomplete) labelings and views [19].

Indeed, for a MVML data set, besides the different representations, some correlations exist in the data set and also have influence on the performance of a MVML learning machine. For example, instance–instance correlation, feature–feature correlation, label–label correlation, and feature–label correlation. These correlations reflect the similarity information among instances (including features and labels)

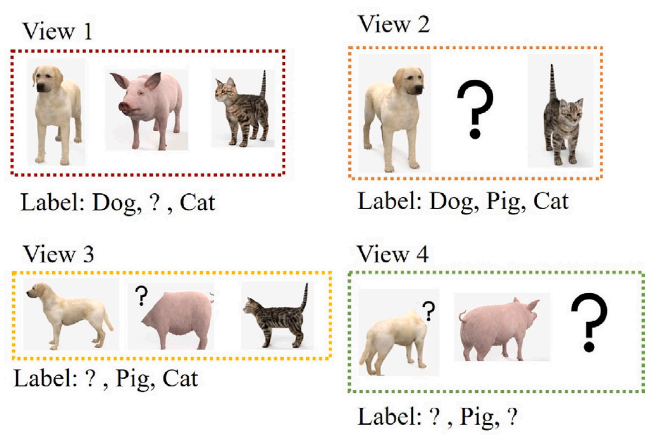


Fig. 2. Representation of a MVML data with incomplete information.

and they have not been considered in the above mentioned references. Even though some scholars develop solutions with correlations introduced [20–22], but how to process incomplete case with multiple representations and correlations is an open problem. Thus, in this article, we develop a within- cross- consensus-view representation-based multi-view multi-label learning with incomplete data (WCC-MVML-ID) by using multiple correlations including instance–instance one, feature–feature one, label–label one, and feature–label one.

Being different from the traditional MVML learning machines, the contributions of WCC-MVML-ID include:

(1) It has an ability to process incomplete case with the exploitation of hidden useful information from available instances by considering some correlations including instance–instance one, feature–feature one, label–label one, and feature–label one;

(2) It takes the feature-oriented information, label-oriented information, and associated information between features and labels from multiple representations in consideration which brings a better performance for a MVML learning machine.

The article is organized as follows: Section 2 demonstrates the working mechanism, optimization procedure, and covers time complexity of the WCC-MVML-ID. In Section 3, we report experiments to evaluate the proposed approach. Section 4 concludes this article and shows future studies.

2. Methodology

As we said in the previous section, besides being unable to handle incomplete case, useful information from multiple representations and hidden in available instances are also less considered in traditional MVML learning. In this context, we exploit useful information with the usage of multiple correlations in diverse representations and propose a novel multi-view multi-label learning named WCC-MVML-ID.

2.1. Framework of WCC-MVML-ID

For any incomplete MVML data set $D = [X, Y]$ with n instances and V views, if in a view, an instance has complete features and labels, we treat this instance be information-complete, otherwise, the instance is information-incomplete. Then we let $X^v \in \mathbb{R}^{d_v \times n}$ and $Y^v \in \mathbb{R}^{c_v \times n}$ be the set of instances (including the information-complete instances and information-incomplete ones) and the corresponding labels from vth view where d_v and c_v are feature dimension and number of labels for vth view respectively. Here, $j, m \in [1, n]$, $i_v, r_v \in [1, d_v]$, $k_v, s_v \in [1, c_v]$ are indexes which indicate jth (or mth) instance, i_vth (or r_vth) feature, and k_vth (or s_v) label. In addition, we let $X^{Av} \in \mathbb{R}^{d_{Av} \times n_v}$ and $Y^{Av} \in \mathbb{R}^{c_{Av} \times n_v}$ be the set of information-complete instances and the corresponding labels from vth view, where n_v is the number of information-complete instances and d_{Av}, c_{Av} are the corresponding feature dimension, number of labels, respectively. 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 let (1) $\bar{S^v} \in \mathbb{R}^{n_v \times n_v}$, $\bar{P^v} \in \mathbb{R}^{d_{Av} \times d_{Av}}$, $\bar{Q^v} \in \mathbb{R}^{c_{Av} \times c_{Av}}$, $\bar{W^v} \in \mathbb{R}^{c_{Av} \times d_{Av}}$ denote the four matrices pre-constructed from the information-complete instances of vth view, where all elements of them are non-negative; (2) each element in $\bar{S^v}$, $\bar{P^v}$, $\bar{Q^v}$, $\bar{W^v}$ represents the instance–instance correlation, feature–feature correlation, label–label correlation, feature–label correlation between two information-complete instances, respectively.

Due to the data missing, every $\bar{S^v}$, $\bar{P^v}$, $\bar{Q^v}$, $\bar{W^v}$ cannot reveal the comprehensive relationships of all instances. Indeed, if we can get the comprehensive relationships, the missing information about features and labels can be inferred and the incomplete MVML data set D can be recovered.

For this purpose, we take the matrices pre-constructed as basis to explore complementary information of different views, analyze the shortcomings of the existing work, and summarize that a feasible model to solve an incomplete MVML data set should compose of feature-oriented part, label-oriented part, and associated part in within-view representations, cross-view representations, and consensus-view representations at least. The parts related with within-view representations are mainly used to preserve the available information and recover the view-specific data, the parts related with cross-view representations are used to infer missing data with different views, and the parts related with consensus-view representations aim to learn the consensus representation of different views.

As a summary, WCC-MVML-ID is mainly composed of the following components: within-view preservation and recovering, cross-view inferring, and consensus representation learning.

2.1.1. Within-view preservation

As we said before, a main objective of WCC-MVML-ID is to reveal the comprehensive relationships of all instances, thus we let $S^v \in \mathbb{R}^{n \times n}$, $P^v \in \mathbb{R}^{d_v \times d_v}$, $Q^v \in \mathbb{R}^{c_v \times c_v}$, $W^v \in \mathbb{R}^{c_v \times d_v}$ be the four referred (completed, comprehensive) matrices of the vth view, each element in S^v, P^v, Q^v, W^v represents the instance–instance correlation, feature–feature correlation, label–label correlation, feature–label correlation between two instances, respectively. ***Due to these four referred matrices store comprehensive relationships of all instances ultimately, thus for vth view, 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***. Thus, to this end, we design the following four within-view preservation sub-models where $\|\star\|_F^2$ represents the Frobenius norm.

$$\min_{S^v} \sum_{v=1}^{V} \left\|(S^v)_A - \bar{S^v}\right\|_F^2 \tag{1}$$

$$\min_{P^v} \sum_{v=1}^{V} \left\|(P^v)_A - \bar{P^v}\right\|_F^2 \tag{2}$$

$$\min_{Q^v} \sum_{v=1}^{V} \left\|(Q^v)_A - \bar{Q^v}\right\|_F^2 \tag{3}$$

$$\min_{W^v} \sum_{v=1}^{V} \left\|(W^v)_A - \bar{W^v}\right\|_F^2 \tag{4}$$

where $(S^v)_A \in \mathbb{R}^{n_v \times n_v}$, $(P^v)_A \in \mathbb{R}^{d_{Av} \times d_{Av}}$, $(Q^v)_A \in \mathbb{R}^{c_{Av} \times c_{Av}}$, $(W^v)_A \in \mathbb{R}^{c_{Av} \times d_{Av}}$ denote the sub-matrices of S^v, P^v, Q^v, W^v, respectively. Each element of $(S^v)_A$, $(P^v)_A$, $(Q^v)_A$, $(W^v)_A$ denotes the correlation of the corresponding two information-complete instances as that in $\bar{S^v}$, $\bar{P^v}$, $\bar{Q^v}$, $\bar{W^v}$, respectively.

Then in order to incorporate the detailed situation of missing information into the previous four within-view preservation sub-models, 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. In these two matrices, 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. Based on E^v and F^v, the Eqs. (2)∼(4) can be transformed as below where $\odot$ indicates the element-wise based multiplication operation.

$$\min_{P^v} \sum_{v=1}^{V} \left\|(\widetilde{P^v} - P^v) \odot (E^v E^{vT})\right\|_F^2 \tag{5}$$

$$\min_{Q^v} \sum_{v=1}^{V} \left\|(\widetilde{Q^v} - Q^v) \odot (F^v F^{vT})\right\|_F^2 \tag{6}$$

$$\min_{W^v} \sum_{v=1}^{V} \left\|(\widetilde{W^v} - W^v) \odot (F^v E^{vT})\right\|_F^2 \tag{7}$$

Then since elements in S^v are related with instances, thus Eq. (1) can be transformed with the below two equations which related with E^v and F^v, respectively.

$$\min_{S^v} \sum_{v=1}^{V} \left\|(\widetilde{S^v} - S^v) \odot (E^{vT} E^v)\right\|_F^2 \tag{8}$$

$$\min_{S^v} \sum_{v=1}^{V} \left\|(\widetilde{S^v} - S^v) \odot (F^{vT} F^v)\right\|_F^2 \tag{9}$$

In these transformed equations, if we define $M^{v1} = E^v E^{vT} \in \mathbb{R}^{d_v \times d_v}$, $M^{v2} = F^v F^{vT} \in \mathbb{R}^{c_v \times c_v}$, $M^{v3} = F^v E^{vT} \in \mathbb{R}^{c_v \times d_v}$, $M^{v4} = E^{vT} E^v \in \mathbb{R}^{n \times n}$, $M^{v5} = F^{vT} F^v \in \mathbb{R}^{n \times n}$, then element $M^{v1}_{i_v,r_v}$ of M^{v1} means that how many instances are information-complete on both i_vth feature and r_vth feature, element $M^{v2}_{k_v,s_v}$ of M^{v2} means that how many instances are information-complete on both k_vth label and s_vth label, element $M^{v3}_{k_v,i_v}$ of M^{v3} means that how many instances are information-complete on both k_vth label and i_vth feature, element $M^{v4}_{j,m}$ of M^{v4} means that how many features are information-complete simultaneously for both jth instance and mth instance, element $M^{v5}_{j,m}$ of M^{v5} means that how many labels are information-complete simultaneously for both jth instance and mth instance.

Moreover, $\widetilde{S^v} \in \mathbb{R}^{n \times n}$, $\widetilde{P^v} \in \mathbb{R}^{d_v \times d_v}$, $\widetilde{Q^v} \in \mathbb{R}^{c_v \times c_v}$, $\widetilde{W^v} \in \mathbb{R}^{c_v \times d_v}$ in these transformed equations are four extended matrices which are filled by $\bar{S^v}$, $\bar{P^v}$, $\bar{Q^v}$, $\bar{W^v}$, respectively. In these extended matrices, the elements related to the missing instances are set as 0. Through mathematical calculations, it can be inferred that there are some connections between extended matrices and pre-constructed matrices, i.e.,

$$\widetilde{S^v} = \Gamma^v_S \bar{S^v} \Gamma^{vT}_S \tag{10}$$

$$\widetilde{P^v} = \Gamma_P^v \bar{P}^v \Gamma_P^{vT} \tag{11}$$

$$\widetilde{Q^v} = \Gamma_Q^v \bar{Q}^v \Gamma_Q^{vT} \tag{12}$$

$$\widetilde{W^v} = \Gamma_Q^v \bar{W}^v \Gamma_P^{vT} \tag{13}$$

where $\Gamma_S^v \in \mathbb{R}^{n\times n_v}$, $\Gamma_P^v \in \mathbb{R}^{d_v\times d_{Av}}$, $\Gamma_Q^v \in \mathbb{R}^{c_v\times c_{Av}}$ are three missing-index matrices. The definitions of them are given as below.

In the vth view, the jth row and bth ($b \in [1, n_v]$) column element in Γ_S^v is 1 if the index of bth information-complete instance in all instances is j, and 0, otherwise. Similarly, the i_vth row and cth ($c \in [1, d_{Av}]$) column element in Γ_P^v is 1 if the index of cth information-complete feature in all features is i_v, and 0, otherwise; the k_vth row and dth ($d \in [1, c_{Av}]$) column element in Γ_Q^v is 1 if the index of dth information-complete label in all labels is k_v, and 0, otherwise.

2.1.2. Within-view recovering

According to Section 2.1.1, the four referred (completed, comprehensive) matrices of the vth view, namely, S^v, P^v, Q^v, W^v, can store comprehensive relationships of all instances ultimately, thus we try to recover the view-specific data X^v and Y^v with them.

In simple speaking, the X^vS^v, P^vX^v and Y^vS^v, Q^vY^v can be regarded as the recovered versions of X^v and Y^v, respectively. In addition, for most data sets, there exist a mapping relationship between X^v and Y^v and with this relationship, the X^v can be mapped into Y^v. Thus, according to the definition of W^v, we can also treat the $W^{vT}Y^v$ and W^vX^v as another recovered versions of X^v and Y^v, respectively. ***Moreover, the difference between the recovered versions and the corresponding view-specific data should be small enough so that the available information can be preserved in the recovered versions as far as possible.***

To this end, we design the following six within-view recovering sub-models.

$$\min_{S^v} \sum_{v=1}^{V} \|X^v - X^vS^v\|_F^2 \tag{14}$$

$$\min_{P^v} \sum_{v=1}^{V} \|X^v - P^vX^v\|_F^2 \tag{15}$$

$$\min_{S^v} \sum_{v=1}^{V} \|Y^v - Y^vS^v\|_F^2 \tag{16}$$

$$\min_{Q^v} \sum_{v=1}^{V} \|Y^v - Q^vY^v\|_F^2 \tag{17}$$

$$\min_{W^v} \sum_{v=1}^{V} \left\|X^v - W^{vT}Y^v\right\|_F^2 \tag{18}$$

$$\min_{W^v} \sum_{v=1}^{V} \|Y^v - W^vX^v\|_F^2 \tag{19}$$

2.1.3. Cross-view inferring

For the incomplete MVML data, it is hard to obtain the complete information about data by only exploring the within-view information since the lack of similarity information (namely, correlations) of the information-incomplete instances and information-complete instances. Fortunately, since many MVML data contain some complementary information among views and there exist some connected relationships among instances in at least one view for incomplete MVML data, thus it is possible to infer the missing data by accommodating the similarity information (namely, correlations) from other views [23]. ***Take S^v as an example, if we regard the $B^S \in \mathbb{R}^{V\times V}$ as a similarity matrix or self-representation matrix for V views and its uth row and vth column element $B_{u,v}^S$ is the similarity between uth view and vth view. Then $\sum_{u=1,u\neq v}^{V} S^uB_{u,v}^S$ can be treated as the total cross-view similarity information for vth view. In order to adaptively select the most reliable information about other views for matrix completion such that more feasible correlation-based matrices can be achieved, we let the difference between S^v and $\sum_{u=1,u\neq v}^{V} S^uB_{u,v}^S$ be minimize. This can be used for optimizing S^v and inferring the incomplete MVML data with X^vS^v and Y^vS^v.***

Thus, we design the following four cross-view inferring sub-models to recover the information-incomplete instances and these sub-models are also sparse representation based.

$$\min_{S^v,B^S} \sum_{v=1}^{V} \left\|S^v - \sum_{u=1,u\neq v}^{V} S^uB_{u,v}^S\right\|_F^2 \tag{20}$$
$$s.t. \begin{cases} 0 \le S^v \le 1, S^{vT}I_v^S = I_v^S, & S_{j,j}^v = 0 \\ 0 \le B_{u,v}^S \le 1, \sum_{u=1,u\neq v}^{V} B_{u,v}^S = 1, & B_{v,v}^S = 0 \end{cases}$$

$$\min_{P^v,B^P} \sum_{v=1}^{V} \left\|P^v - \sum_{u=1,u\neq v}^{V} P^uB_{u,v}^P\right\|_F^2 \tag{21}$$
$$s.t. \begin{cases} 0 \le P^v \le 1, P^{vT}I_v^P = I_v^P, & P_{i_v,i_v}^v = 0 \\ 0 \le B_{u,v}^P \le 1, \sum_{u=1,u\neq v}^{V} B_{u,v}^P = 1, & B_{v,v}^P = 0 \end{cases}$$

$$\min_{Q^v,B^Q} \sum_{v=1}^{V} \left\|Q^v - \sum_{u=1,u\neq v}^{V} Q^uB_{u,v}^Q\right\|_F^2 \tag{22}$$
$$s.t. \begin{cases} 0 \le Q^v \le 1, Q^{vT}I_v^Q = I_v^Q, & Q_{k_v,k_v}^v = 0 \\ 0 \le B_{u,v}^Q \le 1, \sum_{u=1,u\neq v}^{V} B_{u,v}^Q = 1, & B_{v,v}^Q = 0 \end{cases}$$

$$\min_{W^v,B^W} \sum_{v=1}^{V} \left\|W^v - \sum_{u=1,u\neq v}^{V} W^uB_{u,v}^W\right\|_F^2 \tag{23}$$
$$s.t. \begin{cases} 0 \le W^v \le 1, W^{vT}I_v^{W1} = I_v^{W2} & , W^vI_v^{W2} = I_v^{W1}, W_{k_v,i_v}^v = 0 \\ 0 \le B_{u,v}^W \le 1, \sum_{u=1,u\neq v}^{V} B_{u,v}^W = 1, & B_{v,v}^W = 0 \end{cases}$$

where B^S, B^P, B^Q, B^W are four $V\times V$ self-representation matrices and $I_v^S \in \mathbb{R}^{n\times 1}$, $I_v^P \in \mathbb{R}^{d_v\times 1}$, $I_v^Q \in \mathbb{R}^{c_v\times 1}$, $I_v^{W1} \in \mathbb{R}^{c_v\times 1}$, $I_v^{W2} \in \mathbb{R}^{d_v\times 1}$ are five column vectors with all elements as one.

2.1.4. Consensus representation learning

As we mentioned, for the incomplete MVML data, consensus-view representation contains the consensus information and information source of all different views and all views always share a same consensus representation. If we can learn the consensus representation, the incomplete parts of X^v and Y^v can also be inferred. Concretely speaking, we let $X^c \in \mathbb{R}^{d\times n}$ and $Y^c \in \mathbb{R}^{l\times n}$ be the consensus representation of X^v and Y^v, respectively. d and l are the numbers of features and labels for the consensus-view representation. In addition, x_j^c (or x_m^c) and y_j^c (or y_m^c) are the consensus representations of jth (or mth) instance and its corresponding label.

Then, we refer to many related work including [24] and suppose that ***if the similarity information (namely, correlation) between jth instance and mth instance is larger, their corresponding consensus representations are more similar. In order to make the optimization of similarity information be adaptive***, we design the following consensus-view sub-models.

$$\min_{X^c,S^v} \frac{1}{2} \sum_{j=1}^{n} \sum_{m=1}^{n} \left\|x_j^c - x_m^c\right\|_2^2 s_{jm} \tag{24}$$

$$\min_{Y^c,S^v} \frac{1}{2} \sum_{j=1}^{n} \sum_{m=1}^{n} \left\|y_j^c - y_m^c\right\|_2^2 s_{jm} \tag{25}$$

If we can optimize S^v with the solution of these two consensus-view sub-models, then the incomplete parts of X^v and Y^v can also be inferred with X^vS^v and Y^vS^v.

Moreover, suppose the affinity matrix for S^v is $W_{S^v} = \frac{(S^v + S^{vT})}{2}$ and its Laplacian matrix is L_{S^v} where $L_{S^v} = D_{S^v} - W_{S^v}$ and D_{S^v} is the diagonal degree matrix whose elements are column (or row, since W_{S^v} is symmetric) sums of W_{S^v} [5]. Then, the above consensus-view sub-models can be transformed as the following ones where tr is the trace of a matrix, $X^{cT}X^c = I_n$ and $Y^{cT}Y^c = I_n$ are two constraints, and $I_n \in \mathbb{R}^{n\times n}$ is an identity matrix.

$$\min_{X^c} \sum_{v=1}^{V} tr(X^c L_{S^v} X^{cT}) \quad s.t. X^{cT}X^c = I_n \tag{26}$$

$$\min_{Y^c} \sum_{v=1}^{V} tr(Y^c L_{S^v} Y^{cT}) \quad s.t. Y^{cT}Y^c = I_n \tag{27}$$

2.2. The optimization problem

According to Section 2.1, we combine the Eqs. (5)∼(9), (14)∼(23), (26)∼(27) in together. In addition, we refer to [24] and also consider that different views may contain different degrees of useful information, so we also introduce adaptively weighted learning and the optimization problem of WCC-MVML-ID is given as below where $B \in \{B^S, B^P, B^Q, B^W\}$, α^v is the weight of vth view which denotes the importance of the view, $r > 1$ is the smoothing parameter to control the distribution of weights for different views, λs are the penalty parameters to balance the importance of the corresponding constraints.

$$\begin{aligned}\min_{\substack{S^v,P^v,Q^v,W^v\\ B,X^c,Y^c,\alpha^v}} \mathcal{L} = \sum_{v=1}^{V}(\alpha^v)^r[&\lambda_1 \left\|(\widetilde{S^v} - S^v)\odot(E^{vT}E^v)\right\|_F^2 + \\ &\lambda_2 \left\|(\widetilde{S^v} - S^v)\odot(F^{vT}F^v)\right\|_F^2 + \lambda_3 \left\|(\widetilde{P^v} - P^v)\odot(E^v E^{vT})\right\|_F^2 + \\ &\lambda_4 \left\|(\widetilde{Q^v} - Q^v)\odot(F^v F^{vT})\right\|_F^2 + \lambda_5 \left\|(\widetilde{W^v} - W^v)\odot(F^v E^{vT})\right\|_F^2 + \\ &\lambda_6 \|X^v - X^v S^v\|_F^2 + \lambda_7 \|X^v - P^v X^v\|_F^2 + \lambda_8 \|Y^v - Y^v S^v\|_F^2 + \\ &\lambda_9 \|Y^v - Q^v Y^v\|_F^2 + \lambda_{10} \left\|X^v - W^{vT}Y^v\right\|_F^2 + \lambda_{11} \|Y^v - W^v X^v\|_F^2 + \\ &\lambda_{12} \left\|S^v - \sum_{u=1,u\neq v}^{V} S^u B^S_{u,v}\right\|_F^2 + \lambda_{13} \left\|P^v - \sum_{u=1,u\neq v}^{V} P^u B^P_{u,v}\right\|_F^2 + \\ &\lambda_{14} \left\|Q^v - \sum_{u=1,u\neq v}^{V} Q^u B^Q_{u,v}\right\|_F^2 + \lambda_{15} \left\|W^v - \sum_{u=1,u\neq v}^{V} W^u B^W_{u,v}\right\|_F^2 + \\ &\lambda_{16} tr(X^c L_{S^v} X^{cT}) + \lambda_{17} tr(Y^c L_{S^v} Y^{cT})]\end{aligned} \tag{28}$$

$$s.t.\begin{cases} 0 \le S^v, P^v, Q^v, W^v, B^S_{u,v}, B^P_{u,v}, B^Q_{u,v}, B^W_{u,v} \le 1 \\ S^{vT} I^S_v = I^S_v, P^{vT} I^P_v = I^P_v, Q^{vT} I^Q_v = I^Q_v, \\ \quad W^{vT} I^{W1}_v = I^{W2}_v, W^v I^{W2}_v = I^{W1}_v \\ S^v_{j,j} = P^v_{i_v,i_v} = Q^v_{k_v,k_v} = W^v_{k_v,i_v} = B^S_{v,v} = B^P_{v,v} = B^Q_{v,v} = B^W_{v,v} = 0 \\ \sum_{u=1,u\neq v}^{V} B^S_{u,v} = \sum_{u=1,u\neq v}^{V} B^P_{u,v} = \sum_{u=1,u\neq v}^{V} B^Q_{u,v} = \sum_{u=1,u\neq v}^{V} B^W_{u,v} = 1 \\ X^{cT}X^c = I_n, Y^{cT}Y^c = I_n \end{cases}$$

2.3. Optimization

It is hard for us to optimize Eq. (28) since there are $5V+6$[1] variables to optimize. But refer to [24–29], we find that the alternative iterative optimization approach they adopted is a relatively reasonable method to find the local optimal solution and in this article, we adopt the same method. Detailed optimization steps are given as follows:

We let $A \in \{S^v, P^v, Q^v, W^v, B^S, B^P, B^Q, B^W, X^c, Y^c, \alpha^v\}$ firstly and fix other variables when to optimize an variable in A so as to obtain the corresponding optimization problem of Eq. (28) w.r.t. A (see Eqs. (29)∼(39)). Here, in Eq. (39), e^v indicates the remaining part of Eq. (28) except for $\sum_{v=1}^{V}(\alpha^v)^r$.

$$\begin{aligned}\min_{\substack{0\le S^v\le 1\\ S^{vT}I^S_v=I^S_v\\ S^v_{j,j}=0}} \mathcal{L}(S^v) = \sum_{v=1}^{V}(\alpha^v)^r[&\lambda_1 \left\|(\widetilde{S^v} - S^v)\odot(E^{vT}E^v)\right\|_F^2 + \\ &\lambda_2 \left\|(\widetilde{S^v} - S^v)\odot(F^{vT}F^v)\right\|_F^2 + \lambda_6 \|X^v - X^v S^v\|_F^2 + \lambda_8 \|Y^v - Y^v S^v\|_F^2 + \\ &\lambda_{12} \left\|S^v - \sum_{u=1,u\neq v}^{V} S^u B^S_{u,v}\right\|_F^2 + \lambda_{16} tr(X^c L_{S^v} X^{cT}) + \lambda_{17} tr(Y^c L_{S^v} Y^{cT})]\end{aligned} \tag{29}$$

$$\begin{aligned}\min_{\substack{0\le P^v\le 1\\ P^{vT}I^P_v=I^P_v\\ P^v_{i_v,i_v}=0}} \mathcal{L}(P^v) = \sum_{v=1}^{V}(\alpha^v)^r[&\lambda_3 \left\|(\widetilde{P^v} - P^v)\odot(E^v E^{vT})\right\|_F^2 + \\ &\lambda_7 \|X^v - P^v X^v\|_F^2 + \lambda_{13} \left\|P^v - \sum_{u=1,u\neq v}^{V} P^u B^P_{u,v}\right\|_F^2]\end{aligned} \tag{30}$$

$$\begin{aligned}\min_{\substack{0\le Q^v\le 1\\ Q^{vT}I^Q_v=I^Q_v\\ Q^v_{k_v,k_v}=0}} \mathcal{L}(Q^v) = \sum_{v=1}^{V}(\alpha^v)^r[&\lambda_4 \left\|(\widetilde{Q^v} - Q^v)\odot(F^v F^{vT})\right\|_F^2 + \\ &\lambda_9 \|Y^v - Q^v Y^v\|_F^2 + \lambda_{14} \left\|Q^v - \sum_{u=1,u\neq v}^{V} Q^u B^Q_{u,v}\right\|_F^2]\end{aligned} \tag{31}$$

$$\begin{aligned}\min_{\substack{0\le W^v\le 1\\ W^{vT}I^{W1}_v=I^{W2}_v\\ W^v I^{W2}_v=I^{W1}_v\\ W^v_{k_v,i_v}=0}} \mathcal{L}(W^v) = \sum_{v=1}^{V}(\alpha^v)^r[&\lambda_5 \left\|(\widetilde{W^v} - W^v)\odot(F^v E^{vT})\right\|_F^2 + \\ &\lambda_{10} \left\|X^v - W^{vT}Y^v\right\|_F^2 + \lambda_{11} \|Y^v - W^v X^v\|_F^2 + \\ &\lambda_{15} \left\|W^v - \sum_{u=1,u\neq v}^{V} W^u B^W_{u,v}\right\|_F^2]\end{aligned} \tag{32}$$

$$\min_{\substack{0\le B^S_{u,v}\le 1, B^S_{v,v}=0\\ \sum_{u=1,u\neq v}^{V} B^S_{u,v}=1}} \mathcal{L}(B^S) = \sum_{v=1}^{V}(\alpha^v)^r[\lambda_{12} \left\|S^v - \sum_{u=1,u\neq v}^{V} S^u B^S_{u,v}\right\|_F^2] \tag{33}$$

$$\min_{\substack{0\le B^P_{u,v}\le 1, B^P_{v,v}=0\\ \sum_{u=1,u\neq v}^{V} B^P_{u,v}=1}} \mathcal{L}(B^P) = \sum_{v=1}^{V}(\alpha^v)^r[\lambda_{13} \left\|P^v - \sum_{u=1,u\neq v}^{V} P^u B^P_{u,v}\right\|_F^2] \tag{34}$$

$$\min_{\substack{0\le B^Q_{u,v}\le 1, B^Q_{v,v}=0\\ \sum_{u=1,u\neq v}^{V} B^Q_{u,v}=1}} \mathcal{L}(B^Q) = \sum_{v=1}^{V}(\alpha^v)^r[\lambda_{14} \left\|Q^v - \sum_{u=1,u\neq v}^{V} Q^u B^Q_{u,v}\right\|_F^2] \tag{35}$$

$$\min_{\substack{0\le B^W_{u,v}\le 1, B^W_{v,v}=0\\ \sum_{u=1,u\neq v}^{V} B^W_{u,v}=1}} \mathcal{L}(B^W) = \sum_{v=1}^{V}(\alpha^v)^r[\lambda_{15} \left\|W^v - \sum_{u=1,u\neq v}^{V} W^u B^W_{u,v}\right\|_F^2] \tag{36}$$

$$\min_{X^{cT}X^c=I_n} \mathcal{L}(X^c) = \sum_{v=1}^{V}(\alpha^v)^r[\lambda_{16} tr(X^c L_{S^v} X^{cT})] \tag{37}$$

$$\min_{Y^{cT}Y^c=I_n} \mathcal{L}(Y^c) = \sum_{v=1}^{V}(\alpha^v)^r[\lambda_{17} tr(Y^c L_{S^v} Y^{cT})] \tag{38}$$

$$\min_{0\le\alpha^v\le 1, \sum_{v=1}^{V}\alpha^v=1} \mathcal{L}(\alpha^v) = \sum_{v=1}^{V}(\alpha^v)^r[e^v] \tag{39}$$

Then the optimization of Eqs. (29)∼(36) accounts for iteration over these variables in the form below where t indicates the iteration

[1] 5 V corresponds to $S^v, P^v, Q^v, W^v, \alpha^v$, 6 corresponds to $B^S, B^P, B^Q, B^W, X^c, Y^c$.

Table 1
Algorithm: WCC-MVML-ID.

Input: an incomplete MVML data set **Output**: $A \in \{S^v, P^v, Q^v, W^v, B^S, B^P, B^Q, B^W, X^c, Y^c, \alpha^v\}$ where $v = 1, 2, \dots, V$
1. initialize variables in A, penalty parameters λs, and smoothing parameter r; 2. **repeat** 3. update each variable in $\{B^S, B^P, B^Q, B^W, X^c, Y^c\}$ according to Section 2.3 when other variables are fixed; 4. for $v = 1, 2, \dots, V$ 5. update each variable in $\{S^v, P^v, Q^v, W^v, \alpha^v\}$ according to Section 2.3 when other variables are fixed; 6. end for 7. **until** normalized value of $\mathcal{L}$ is small than 0.01.

number.[2]

$$A(t+1) = A(t) - \frac{\partial \mathcal{L}(A)}{\partial A} \tag{40}$$

The optimization of Eqs. (37) and (38) turns out to be a typical eigenvalue decomposition problem. Thus, suppose $x_1^c, x_2^c, \dots, x_d^c$ are the eigenvectors corresponding to the first d minimum eigenvalues of matrix $\sum_{v=1}^{V}(\alpha^v)^r L_{S^v}$ and then the optimal solution to problem Eq. (37) is expressed as $X^c = [x_1^c; x_2^c; \dots; x_d^c] \in d \times n$. Similarly, the optimal solution to problem Eq. (38) is expressed as $Y^c = [y_1^c; y_2^c; \dots; y_l^c] \in l \times n$ where $y_1^c; y_2^c; \dots; y_l^c$ are the eigenvectors corresponding to the first l minimum eigenvalues of matrix $\sum_{v=1}^{V}(\alpha^v)^r L_{S^v}$.

The optimization of Eq. (39) can be referred to [24] and its optimal solution is given below.

$$\alpha^v = (e^v / \sum_{v=1}^{V} e^v)^{1/(r-1)} \tag{41}$$

According to the above contents, the Eq. (28) can be optimized step by step until the changes about the normalized value of $\mathcal{L}$ is small than 0.01. In addition, the procedure of WCC-MVML-ID can be summarized in the Table 1.

2.4. Computational complexity analysis

Refer to [24–29] and after careful calculation, the computational complexity of WCC-MVML-ID should consist of three parts. For S^v, P^v, $Q^v, W^v, B^S, B^P, B^Q, B^W$, their total computational overhead is $(4V + 2)O(n^3)$. For the X^c and Y^c, the most computational overhead comes by the eigenvalue decomposition. In this article, we adopt 'eigs' functions [30] to speed up the computational efficiency and the overheads are $O(Cn^2)$ for both X^c and Y^c where C is a constant. For α^v, it is easy to get the optimal results via numerical division operation and its computational complexity can be ignored. Thus, according to the above analysis, the computational complexity of the optimization problem for Eq. (28) is about $O((4V+2)tn^3)$.

3. Experiments

3.1. Experimental setup

Data setting: we refer to [31] and use 5 MVML data sets for experiments (see Table 2). All data sets belong to the image domain [32]. Here, Corel5k was first used in [33] and since then, it has become an important benchmark for keyword based image retrieval and image annotation. Espgame was obtained from an online game where two players, that cannot communicate outside the game, gain points by agreeing on words describing the image [34]. IAPRTC12 was initially published for cross-lingual retrieval and it can be transformed into a format comparable to the other sets by extracting common nouns using natural language processing techniques [35]. Mirflickr comprises 25 000 images from the Flickr website which are redistributable for research purposes and represent a real community of users both in the image content and image tags [36]. Pascal07 consists of annotated consumer photographs collected from the Flickr photo-sharing website [37]. The detailed information and setting of these data sets can be found in [31,38], and website 'http://lear.inrialpes.fr/people/guillaumin/data.php'. Each data set involves six views which can be found in many other Refs. [39–42]: HUE (the color portion of the model, expressed as a number from 0 to 360 degrees), SIFT (scale invariant feature transform), GIST (an abstract representation of the scene that spontaneously activates memory representations of scene categories), HSV (hue, saturation, value), RGB (red, green, blue), and LAB (lightness, red/green value, blue/yellow value).

Table 2
Detailed information of used data sets.

Data sets	No. instances	No. labels	Avg. label per instance
Corel5k	4999	260	3.396
Espgame	20 770	268	4.686
IAPRTC12	19 627	291	5.719
Mirflickr	25 000	38	4.716
Pascal07	9963	20	1.465

Baseline learning machines: to study the performance of WCC-MVML-ID, we compare it with 4 state-of-the-art MVML learning machines including MVMLP [13], SSDR-MML [43], LSA-MML [44], ICM2L [45].

Parameter setting: optimal parameter values of compared learning machines are selected from the set as suggested in the original papers. For WCC-MVML-ID, variables $A \in \{S^v, P^v, Q^v, W^v, B^S, B^P, B^Q, B^W, X^c, Y^c, \alpha^v\}$ are initialized with an average way, namely, elements are initialized in equipartition and they have been updated according to Section 2.3. For penalty parameters λs, the optimal parameter values can be selected from the set $\{0.1, 0.2, \dots, 0.8, 0.9\}$ and the smoothing parameter r is selected from the set $\{2, 3, \dots, 19, 20\}$. For the incomplete setting, the missing rates of features and labels can be selected from the set $\{10\%, 20\%, \dots, 80\%, 90\%\}$.

Evaluation: for the comparison about the used learning machines on the experimental data sets, we adopt AUC as the main metric for performance comparisons. AUC is the abbreviation of 'the area under the ROC (receiver operating characteristic) curve'. As we know, ROC curve is used to select the optimal threshold and AUC is a standard used to measure the quality of the classification model [46]. Indeed, there are some other metrics including precision, recall, etc., but with the limitation of the article's length, we only focus on the performances on AUC. On the base of AUC, some experimental results including the different influence of missing rates for features and labels are given. Moreover, the convergence of WCC-MVML-ID is also given.

Selection of optimal parameters: there are some tuning parameters in WCC-MVML-ID. In order to select the optimal parameter values, avoid over-fitting in model selection, and ensure the authenticity of experimental results, we should ensure that a data set D consists of the independent test set D_{te} and the set D_{tr_va} which is used for training and validating a final learning machine. Thus, 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 the algorithm in Table 3 and the corresponding performances of learning machines are obtained. **In simple speaking, we evenly divide D_{tr_va} into ten parts firstly and ten-fold cross-validation is adopted on the same partitions so as to search optimal parameter values. After that, the optimal parameter values are used to test the performances of learning machines on the test set.** In addition, the above operations are repeated for five times independently and we report average performances and the corresponding standard deviations of learning machines on the used data sets.

[2] Limited by the length of this article, details of $\frac{\partial F(A)}{\partial A}$ are not given here, but their computations can refer to various algebra books.

Table 3
Algorithm: selection of optimal parameter values and the acquirement of performances for learning machines.

Input: a MVML data set $D = \{D_{te}, D_{tr\_va}\}$ and all combinations of parameters for a learning machine $P(q)$ where $q \in [1, Q]$ and Q is the total number of combinations;
Output: optimal parameter values $P^{\star}$ and corresponding performances $T(P^{\star})$ including AUC and convergence for learning machines;

1. split D_{tr_va} into 10 non-overlapping parts in average, i.e., $D_{tr\_va} = \{D_{tr\_va1,\ldots,D_{tr\_va10}}\}$;
2. for $q = 1, \ldots, Q$
3. select one combination of parameters, i.e., $P(q)$;
4. for $k = 1, \ldots, 10$
5. choose D_{tr_vak} as the validation set and the rest 9 parts as the training set;
6. adopt the training set to train a temporary learning machine $LM(k)$;
7. adopt D_{tr_vak} to validate the effectiveness of $LM(k)$ and get AUC on D_{tr_vak}, i.e., $A(P(q) - k)$;
8. end for
9. get the average AUC on D_{tr_va} with $P(q)$, i.e., $A(P(q)) = \sum_{r=1}^{10} A(P(q) - r)/10$;
10. end for
11. get the optimal parameter values $P^{\star} = argmax(A(P(q)))$ and the corresponding optimal learning machine $LM(P^{\star})$ is also gotten;
12. test D_{te} with $LM(P^{\star})$ and get $T(P^{\star})$.

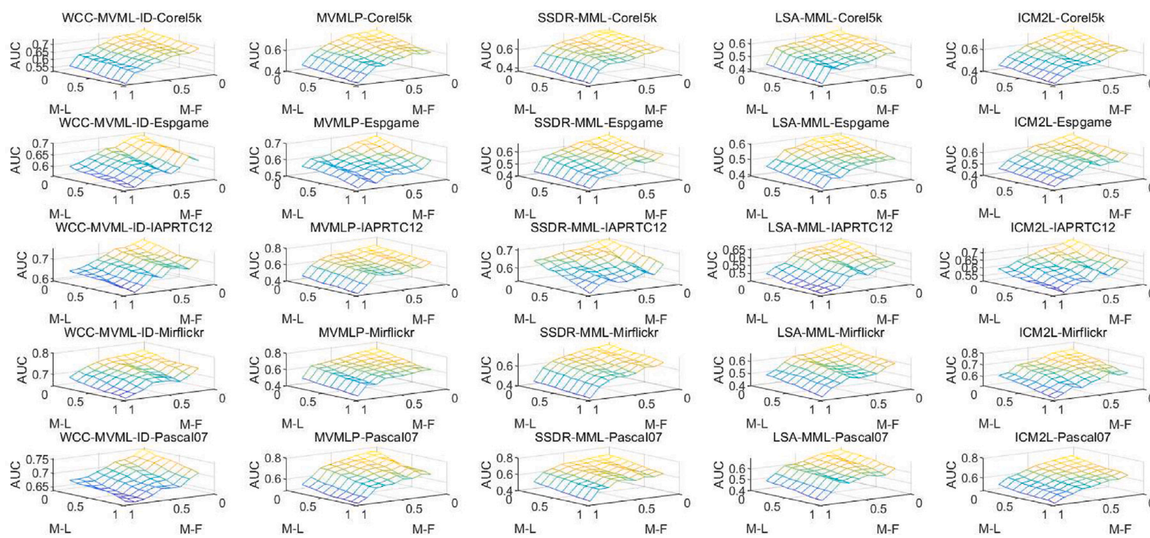


Fig. 3. AUC comparisons for different learning machines on the used data sets with different missing rates of features (M-F) and missing rates of labels (M-L).

Experimental environment: all computations are performed on a node of compute cluster with 12 CPUs (Intel Core i7-12700) running RedHat Linux Enterprise 9.0. The coding environment is MATLAB 2020a.

3.2. Performances comparison with different missing rates of features and labels

As we said before, with the limitation of manpower and some other objective factors, a MVML data maybe miss some information about features and labels. Thus, we discuss the influence of different missing rates of features and labels on AUC. In this article, the selection set of missing rates of features and labels is $\{10\%, 20\%, \ldots, 80\%, 90\%\}$. Fig. 3 shows the influence of different values of missing rates for features and labels on AUC and Fig. 4 gives the corresponding standard deviations (std.). Furthermore, in order to validate the effectiveness of WCC-MVML-ID statistically, we adopt the p-value for comparison (see Fig. 5) and in Fig. 5, each sub-figure represents the p-value of WCC-MVML-ID compared with a compared learning machine on a data set with different missing rates of features and labels.

According to these figures, it is found that (1) with the missing rates of features and labels being larger, the AUCs of these learning machines on different data sets have downward trends; (2) even under the case of the same missing rate of features and labels, WCC-MVML-ID still outperforms other compared learning machines on these data sets; (3) the standard deviations of WCC-MVML-ID are smaller than others which indicates the performances of WCC-MVML-ID are more stable; (4) under most cases, the p-values of WCC-MVML-ID compared with other learning machines are smaller than 0.05 which validates the effectiveness of WCC-MVML-ID in generally.

3.3. Influence of λs

In WCC-MVML-ID, there are 17 λs should be tuned and different values of them will bring diverse performances. We discuss the influence of different values of λs and the results are given in Fig. 6. In this figure, each datum represents the best AUC of a λ under a given value on an used data set. According to this figure, it is found that (1) for λ_{12}, λ_{13}, λ_{14}, λ_{15} which correspond to the cross-view expressions, the best AUC achieves when the values of these λs are set to be 0.9; (2) for other λs, the AUC achieves best when the values of these λs be 0.4, 0.5, or 0.6; (3) according to the former conclusions, we can see that compared with other parts, the cross-view information plays a more important role and this indicates that information between different views is more useful for the processing problem about incomplete data; (4) for other parts including the within-view information and consensus-view information, their roles to the process of incomplete MVML data are similar.

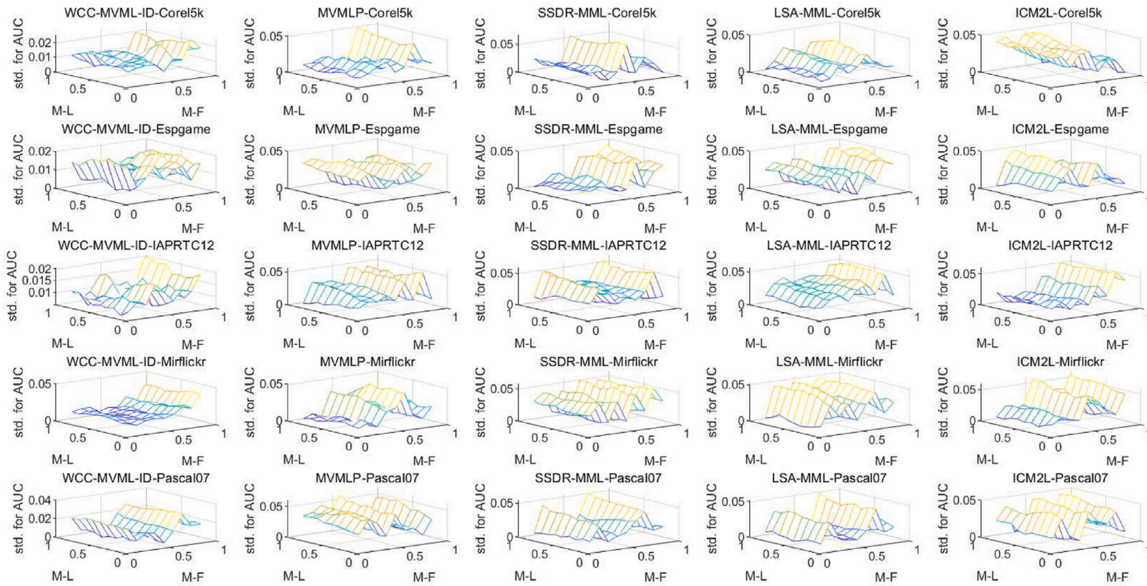


Fig. 4. The std. of AUC comparisons for different learning machines on the used data sets with different missing rates of features (M-F) and missing rates of labels (M-L).

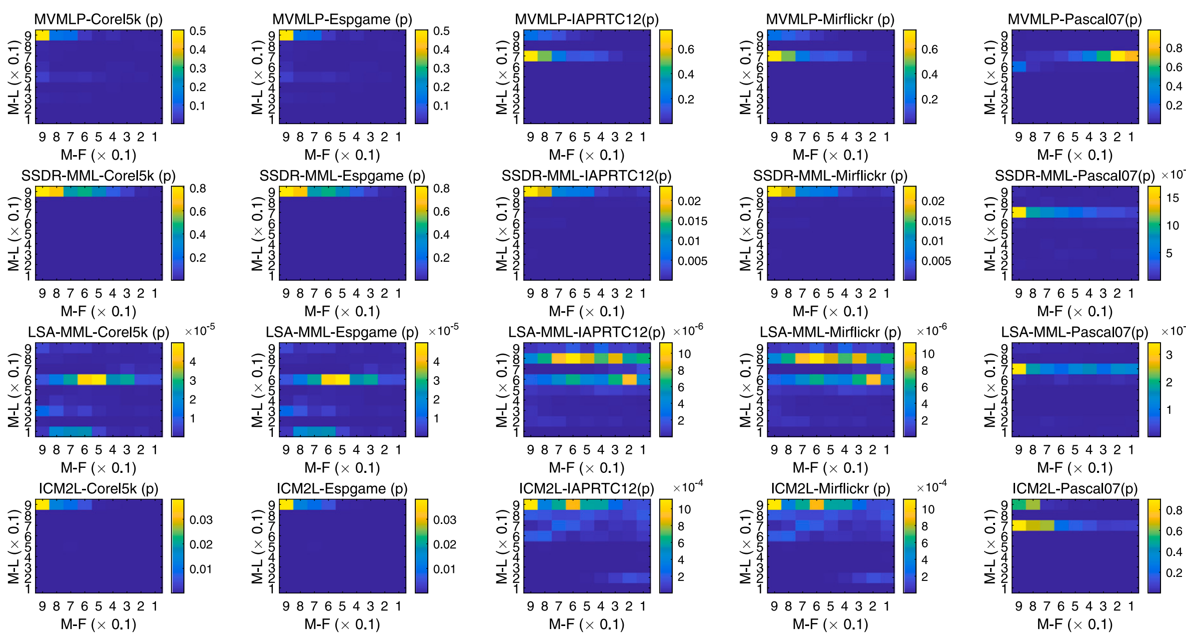


Fig. 5. p-value of AUC comparisons for different learning machines on the used data sets with different missing rates of features (M-F) and missing rates of labels (M-L).

3.4. Influence of r

r in WCC-MVML-ID is a smoothing parameter to control the distribution of weights for different views and its values can be selected in $\{2, 3, \dots, 19, 20\}$. Here, we also show the influence of r on the performances of WCC-MVML-ID with Fig. 7. According to this figure, it is found that the developed WCC-MVML-ID is insensitive to the values of parameter r on data sets Corel5k, Espgame, Pascal07 to some extent (the performance change about AUC is about 2%) and it can achieve a relative good AUC on IAPRTC12 and Mirflickr when r is selected from the set of $[2, 9]$. According to the experimental results in Fig. 7, we can simply select parameter r from the range $[2, 9]$.

3.5. Convergence of WCC-MVML-ID

In this subsection, we show the convergence of WCC-MVML-ID with Fig. 8. In this figure, under any missing rate of features and labels, the iteration numbers are given and according to this figure, it is found that (1) on each used data set, WCC-MVML-ID can converge within 45 iterations; (2) with the missing rates being larger, WCC-MVML-ID should converge with higher iteration numbers.

4. Conclusions and future studies

Incomplete multi-view multi-label (MVML) data sets are ubiquitous in current real-world applications and they are hard to be solved by traditional MVML learning machines. To this end, we develop a within- cross- consensus-view representation-based multi-view multi-label learning with incomplete data (WCC-MVML-ID) to process incomplete MVML data sets better. Different from traditional solutions, WCC-MVML-ID takes feature-oriented information, label-oriented information, and associated information between features and labels from within-view representations, cross-view representations, and consensus

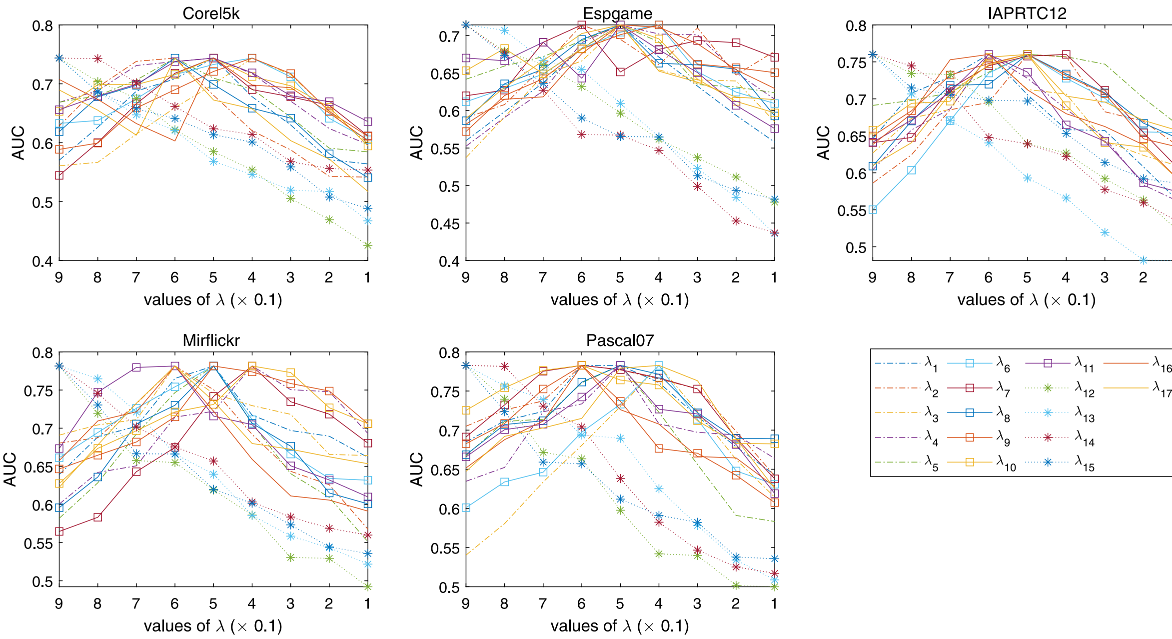


Fig. 6. Influence of different values for different λs on AUC of WCC-MVML-ID.

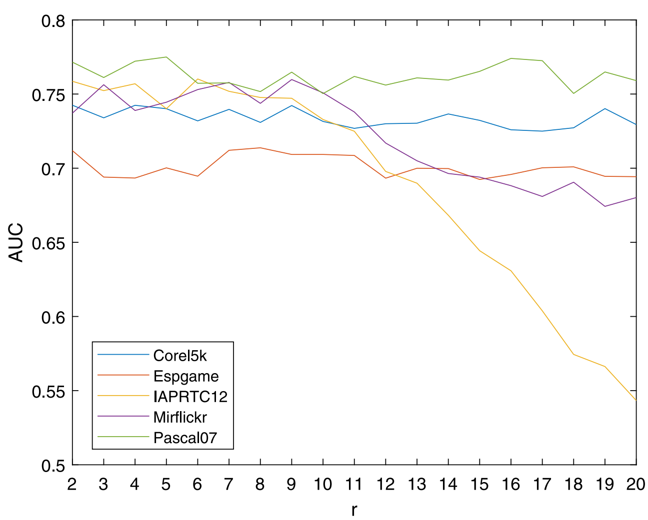


Fig. 7. Influence of different values for r on AUC of WCC-MVML-ID.

-view representations in consideration. In addition, the correlations among instances, features, and labels are also considered to exploit the hidden useful information from available instances. On the base of these considerations, WCC-MVML-ID is mainly composed of three components, i.e., within-view preservation and recovering, cross-view inferring, and consensus representation learning and it has a better ability to process incomplete MVML data sets with the optimization of correlations and inferring of information-incomplete parts.

Experiments on several benchmark data sets demonstrate the superiority of the proposed WCC-MVML-ID over related competitive learning machines. In simple speaking, (1) for MVML data sets, with the missing rates of features and labels being larger, the performances of competitive MVML learning machines have downward trends while WCC-MVML-ID still outperforms other compared ones and its performances are more stable; (2) compared with within-view information and consensus-view information, cross-view information plays a more important role to process incomplete MVML data; (3) WCC-MVML-ID can converge within 45 iterations on used data sets and it always costs higher iteration numbers with the missing rates being larger.

Although WCC-MVML-ID has some advantages to process incomplete MVML data sets, some further studies also attract our attentions and can be considered in the future. First, in current applications, more and more data sets arrive in real-time and this phenomenon will affect the performances of WCC-MVML-ID. Second, affected by some labor costs, some data sets arrive without any priori knowledge and this leads to the useless of current many MVML learning machines. To this end, how to solve real-time MVML data sets and how to carry out zero-shot learning are two open problems in the future.

CRediT authorship contribution statement

Changming Zhu: Conceptualization, Methodology, Funding acquisition, Writing – original draft. **Yanchen Liu:** Validation, Investigation. **Duoqian Miao:** Review & editing. **Yilin Dong:** Formal analysis, Data curation. **Witold Pedrycz:** Review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by National Natural Science Foundation of China (CN) [grant numbers 62276164, 61602296, and 62076182], 'Science and technology innovation action plan' Natural Science Foundation of Shanghai, China [grant number 22ZR1427000]. Furthermore, this work is also sponsored by 'Chenguang Program' supported by Shanghai Education Development Foundation, China and Shanghai Municipal Education Commission, China [grant number 18CG54]. The authors would like to thank their support.

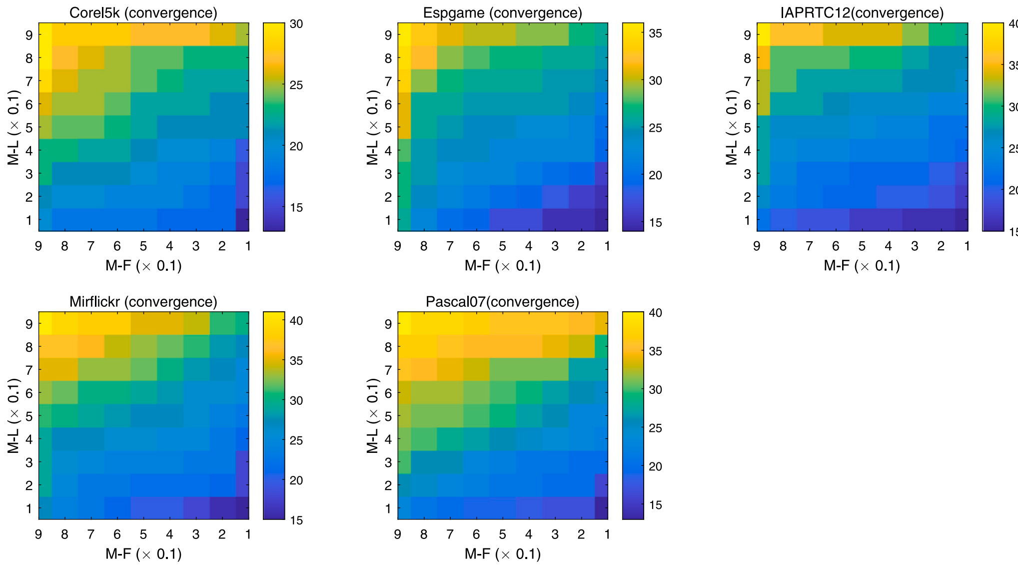


Fig. 8. Convergence of WCC-MVML-ID on different data sets with different missing rates of features and labels.

References

[1] X.Y. Liu, P. Song, Incomplete multi-view clustering via virtual-label guided matrix factorization, Expert Syst. Appl. 210 (2022) 118408.
[2] B. Liu, W.B. Li, Y.S. Xiao, X.D. Chen, L.W. Liu, C.D. Liu, K. Wang, P. Sun, Multi-view multi-label learning with high-order label correlation, Inform. Sci. 624 (2023) 165–184.
[3] D.W. Zhao, Q.W. Gao, Y.X. Lu, D. Sun, Learning view-specific labels and label-feature dependence maximization for multi-view multi-label classification, Appl. Soft Comput. 124 (2022) 109071.
[4] S.D. Sun, X.T. Yu, Y.J. Tian, Multi-view prototype-based disambiguation for partial label learning, Pattern Recognit. 141 (2023) 109625.
[5] J.L. Liu, S.H. Teng, L.K. Fei, W. Zhang, X.Z. Fang, Z.X. Zhang, N.Q. Wu, A novel consensus learning approach to incomplete multi-view clustering, Pattern Recognit. 115 (2021) 107890.
[6] H. Tao, C.P. Hou, D.Y. Yi, J.B. Zhu, D.W. Hu, Joint embedding learning and low-rank approximation: a framework for incomplete multiview learning, IEEE Trans. Cybern. 51 (3) (2021) 1690–1703.
[7] W.Z. Zhuge, H. Tao, T.J. Luo, L.L. Zeng, C.P. Hou, D.Y. Yi, Joint representation learning and clustering: a framework for grouping partial multiview data, IEEE Trans. Knowl. Data Eng. 34 (8) (2022) 3826–3840.
[8] Z.M. Xia, Y. Chen, C. Xu, Multiview PCA: a methodology of feature extraction and dimension reduction for high-order data, IEEE Trans. Cybern. 52 (10) (2022) 11068–11080.
[9] K.W. Sun, C.H. Lee, J. Wang, Multilabel classification via co-evolutionary multilabel hypernetwork, IEEE Trans. Knowl. Data Eng. 28 (9) (2016) 2438–2451.
[10] Y. Dai, W.W. Song, Z. Gao, L.Y. Fang, Global-guided weakly-supervised learning for multi-label image classification, J. Vis. Commun. Image Represent. 93 (2023) 103823.
[11] S.L. Sun, Q.J. Zhang, Multiple-view multiple-learner semi-supervised learning, Neural Process. Lett. 34 (2011) 229–240.
[12] V. Kumar, A.K. Pujari, V. Padmanabhan, S.K. Sahu, V.R. Kagita, Multi-label classification using hierarchical embedding, Expert Syst. Appl. 91 (2018) 263–269.
[13] Z.Y. He, C. Chen, J.J. Bu, P. Li, D. Cai, Multi-view based multi-label propagation for image annotation, Neurocomputing 168 (2015) 853–860.
[14] Y.S. Cheng, Q.Y. Li, Y.B. Wang, W.J. Zheng, Multi-view multi-label learning with view feature attention allocation, Neurocomputing 501 (2022) 857–874.
[15] D.Q. Zhao, Q.W. Gao, Y.X. Lu, D. Sun, Learning view-specific labels and label-feature dependence maximization for multi-view multi-label classification, Appl. Soft Comput. 124 (2022) 109071.
[16] D.Q. Zhao, Q.W. Gao, Y.X. Lu, D. Sun, Y.S. Cheng, Consistency and diversity neural network multi-view multi-label learning, Knowl.-Based Syst. 218 (2021) 106841.
[17] L. Zhao, Y.S. Xiao, B. Liu, Z.F. Hao, Multi-view partial label machine, Inform. Sci. 586 (2022) 310–325.
[18] C.W. Qu, K.M. Wang, H. Zhang, G.X. Yu, C. Domeniconi, Incomplete multi-view multi-label active learning, in: 2021 IEEE International Conference on Data Mining, ICDM, 2022, pp. 1294–1299.
[19] B.Y. Qian, X. Wang, J.P. Ye, I. Davidson, A reconstruction error based framework for multi-label and multi-view learning, IEEE Trans. Knowl. Data Eng. 27 (3) (2015) 594–607.
[20] Y.S. Zhang, J. Wu, Z.H. Cai, P.S. Yu, Multi-view multi-label learning with sparse feature selection for image annotation, IEEE Trans. Multimed. 22 (11) (2020) 2844–2857.
[21] Y. Zhang, J.D. Shen, C. Yu, C.J. Wang, Views meet labels: personalized relation refinement network for multiview multilabel learning, IEEE MultiMedia 29 (2) (2022) 104–113.
[22] S.E. Hajjar, F. Dornaika, F. Abdallah, One-step multi-view spectral clustering with cluster label correlation graph, Inform. Sci. 592 (2022) 97–111.
[23] S. Bhadra, S. Kaski, J. Rousu, Multi-view kernel completion, Mach. Learn. 106 (5) (2017) 713–739.
[24] J. Wen, K. Yan, Z. Zhang, Y. Xu, J.Q. Wang, L.K. Fei, B. Zhang, Adaptive graph completion based incomplete multi-view clustering, IEEE Trans. Multimed. (2021) 2493–2504.
[25] S.Y. Yi, Y.Y. Liang, Z.Y. He, Y. Li, Y.M. Cheung, Dual pursuit for subspace learning, IEEE Trans. Multimed. 21 (6) (2019) 1399–1411.
[26] J. Wen, Z.F. Zhong, Z. Zhang, L.K. Fei, Z.H. Lai, R.Z. Chen, Adaptive locality preserving regression, IEEE Trans. Circuits Syst. Video Technol. 30 (1) (2020) 75–88.
[27] Z. Zhang, Z.H. Lai, Z. Huang, W.K. Wong, G.S. Xie, L. Liu, L. Shao, Scalable supervised asymmetric hashing with semantic and latent factor embedding, IEEE Trans. Image Process. 28 (10) (2019) 4803–4818.
[28] M.C. Shang, C. Liang, J.W. Luo, H.X. Zhang, Incomplete multi-view clustering by simultaneously learning robust representations and optimal graph structures, Inform. Sci. 640 (2023) 119038.
[29] L. Yao, G.F. Lu, J.B. Zhao, B. Cai, Multi-view clustering based on a multimetric matrix fusion method, Expert Syst. Appl. 228 (2023) 120272.
[30] T.G. Wright, L.N. Trefethen, Large-scale computation of pseudospectra using ARPACK and eigs, SIAM J. Sci. Comput. 23 (2) (2001) 591–605.
[31] X. Li, S.C. Chen, A concise yet effective model for non-aligned incomplete multi-view and missing multi-label learning, IEEE Trans. Pattern Anal. Mach. Intell. 44 (10) (2022) 5918–5932.
[32] M. Guillaumin, T. Mensink, J. Verbeek, C. Schmid, TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 309–316.
[33] P. Duygulu, K. Barnard, N. De Freitas, D. Forsyth, Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary, in: 7th European Conference on Computer Vision, 2002, pp. 97–112.
[34] L.V. Ahn, L. Dabbish, Labeling images with a computer game, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2004, pp. 319–326.
[35] M. Grubinger, Analysis and Evaluation of Visual Information Systems Performance (Ph.D. thesis), Victoria University, 2007.
[36] M.J. Huiskes, M.S. Lew, The mir flickr retrieval evaluation, in: Proceedings of the ACM International Conference on Multimedia Information Retrieval, 2008, pp. 39–43.
[37] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[38] Q.Y. Tan, G.X. Yu, C. Domeniconi, J. Wang, Z.L. Zhang, Incomplete multi-view weak-label learning, in: 2018 International Joint Conference on Artificial Intelligence, 2018, pp. 2703–2709.

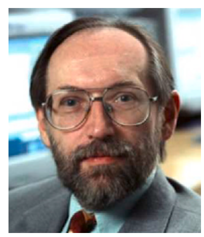
[39] A. Friedman, Framing pictures: The role of knowledge in automatized encoding and memory for gist, J. Exp. Psychol. [Gen.] 108 (1979) 316–335.
[40] R.D. Fan, T.J. Luo, W.Z. Zhuge, S. Qiang, C.P. Hou, Multi-view subspace learning via bidirectional sparsity, Pattern Recognit. 108 (2020) 107524.
[41] X.Y. Liu, L.J. Sun, S.H. Feng, Incomplete multi-view partial multi-label learning, Appl. Intell. 52 (2022) 3289–3302.
[42] M. Guillaumin, J. Verbeek, C. Schmid, Multimodal semi-supervised learning for image classification, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 902–909.
[43] B.Y. Qian, X. Wang, J.P. Ye, I. Davidson, A reconstruction error based framework for multi-label and multi-view learning, IEEE Trans. Knowl. Data Eng. 27 (3) (2015) 594–607.
[44] C.Q. Zhang, Z.W. Yu, Q.H. Hu, P.F. Zhu, X.W. Liu, X.B. Wang, Latent semantic aware multi-view multi-label classification, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018, pp. 4414–4421.
[45] Q.Y. Tan, G.X. Yu, J. Wang, C. Domeniconi, X.L. Zhang, Individuality- and commonality-based multiview multilabel learning, IEEE Trans. Cybern. 51 (3) (2021) 1716–1727.
[46] L.L. Xia, J.P. An, C. Ma, H.J. Hou, Y.P. Hou, L.Y. Cui, X.H. Jiang, W.Q. Li, Z.K. Gao, Neural network model based on global and local features for multi-view mammogram classification, Neurocomputing 536 (2023) 21–29.

Changming Zhu received the Ph.D. degree in computer science and technology from the School of information science and Engineering, East China University of Science and Technology, Shanghai, China, in 2015. From 2018 to 2020, he was a Postdoctoral Fellow with Tongji University, Shanghai, China. He is currently a Professor with the College of Information Engineering, Shanghai Maritime University, Shanghai, China. His current research interesting include image process, and multi-view learning.

Yanchen Liu received the bachelor's degree in computer science and technology from the College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, China, in 2021. He is currently a postgraduate student with the College of Information Engineering, Shanghai Maritime University, Shanghai, China. His current research interesting include multi-view learning.

Duoqian Miao received the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1997. He is currently a Professor with the School of Electronics and Information Engineering and the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, China. His main research interests include soft computing, rough sets, pattern recognition, data mining, and machine learning.

Yilin Dong received the Ph.D. degree in control science and engineering from the School of Automation, Southeast University, Nanjing, China, in 2020. He is currently with the College of Information Engineering, Shanghai Maritime University, Shanghai, China. His research interests include belief function theory, information fusion, and pattern recognition.



Witold Pedrycz (Fellow, IEEE) received the M.Sc., Ph.D., and D.Sc. degrees from the Silesian University of Technology, Gliwice, Poland. He is a Professor with Istinye University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Sariyer/Istanbul, Turkiye and the Canada Research Chair (Computational Intelligence) with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. His main research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering. He has published numerous papers in the above areas. Prof. Pedrycz is extensively involved in editorial activities. He is an Editor-in-Chief of Information Sciences and the IEEE Transactions on Systems, Man, and Cybernetics: Systems. He currently serves as an Associate Editor for the IEEE Transactions on Fuzzy Systems. He is a member of a number of editorial boards of other international journals.