Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An incomplete multi-view multi-label learning with Universum

Changming Zhu

cmzhu@shmtu.edu.cn

Shanghai Maritime University

Lei Wang

Shanghai Maritime University

Research Article

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

Posted Date: September 18th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6496778/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An incomplete multi-view multi-label learning with Universum

Changming Zhu[1*] and Lei Wang[1]

[1*]College of Information Engineering, Shanghai Maritime University, No. 1550, Lingang Avenue, Pudong, 201306, Shanghai, China.

*Corresponding author(s). E-mail(s): cmzhu@shmtu.edu.cn;
Contributing authors: 202330310092@stu.shmtu.edu.cn;

Abstract

In current era, the incomplete multi-view multi-label data sets are widely encountered and they are hard to be processed due to the structure of each instance is complicated and some useful information maybe lost. To this end, this study uses the missing-information-index matrices and adopts Universum learning to promote the enhancements of useful information. Then we combine the self-adaptive measurements of within-view correlation, cross-view correlation, consensus-view correlation among features, labels, and instances in together so that the missing information can be recovered with a more feasible way. The developed learning machine is named as incomplete multi-view multi-label learning with Universum (IMVML-U). In order to demonstrate the superiority of IMVML-U, we adopt **5** MVML data sets for experiments and according to the experimental results, it is found that IMVML-U can process incomplete multi-view multi-label data sets much better, update the correlations among features, labels, and instances dynamically, and enhances the useful information with a less time complexity increased.

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

1 Introduction

1.1 Background

In current real-world, a data set is always represented by multiple feature sets (i.e., views) and marked with multiple class labels. Such a data set is named as multi-view multi-label (MVML) data and many learning machines are developed for processing this kind of data [1, 2, 3]. Now along with the time and the continue development in the science and technology, the forms of these data sets are undergoing changes and among those forms, the incomplete and insufficient form is widely encountered. As we know, due to some unpredictable reasons including the sampling equipment being damaged and the manpower is not sufficient, some instances maybe loss a certain amount of useful information and some instances maybe unlabeled. Such a form makes the useful information be insufficient and current traditional learning machines cannot predict the labels for data sets well.

Thus, to this end, many learning machines are developed to process incomplete multi-view multi-label data with some strategies [4, 18, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17] and the strategies can be summarized into three categories. One is missing-information-index matrices used, another is multiple representations used, and the third is artificial generation of new instances. But according to our depth analysis, it is found that these three strategies exist some shortcomings. First, for the strategy to use missing-information-index matrices, although many learning machines combine the data information matrices, correlation matrices among features, instances, labels with missing-information-index matrices, they cannot fully consider missing information in different representations simultaneously. Second, for the strategy to use multiple representations, many current learning machines design solutions with multiple representations about data considered including the within-view representations, cross-view representations, and consensus-view representations, and in their solutions, the correlations among features, labels, and instances in different representations are always used. While these correlations cannot be measured self-adaptive fully. Third, for the strategy to generate the new instances artificially, Universum learning is popular for generating more useful additional information. But this learning method always take the physical distances between instances as a measurement standard when to generate new information and the correlations between instances are always ignored.

1.2 Proposal

In this study, considering the disadvantages of the above mentioned strategies, we develop a solution named incomplete multi-view multi-label learning with Universum (IMVML-U) with the following steps.

(1) We make full use of the missing information in different representations simultaneously so as to lay the foundation for recovering missing information and updating correlations more feasible. In this step, the missing information

is demonstrated with the missing-information-index matrices used. (2) We consider the recovered quality and recovered quantity about the data and design an 'individual-mutual balance' model so that the details of the whole data and their labels can be represented well. (3) We consider and search some internal laws about the self-adaptive changes for different kinds of correlations so that we can measure the correlations among features, labels, and instances more feasible and accurate. (4) We combine the first three steps in together and design a temp objective function with the nuclear-norm constraint used. This function is called as incomplete multi-view multi-label model, namely, IMVML model. (5) Then we optimize the IMVML model and recover the current data with the correlations updated. (6) On the base of the recovered data and the updated correlations, we generate some new instances and initialize the corresponding labels. This procedure is named as Universum step ('U-step' for short) and the data set can be enhanced. (7) After the 'U-step', we update the corresponding correlations among features, labels, and instances in different representations again and the data set can be recovered more accurate.

1.3 Contributions

Compared with the current existing classical learning machines, the developed IMVML-U possesses the following contributions. First, IMVML-U can enhance the useful data information and as a result, it can recover the missing information in multiple representations more accurate since it has more useful information used. Second, IMVML-U can update the correlations among features, labels, and instances dynamically and more feasible.

1.4 Framework of this study

The framework of this study is given as below. Section 2 reviews the related work about incomplete multi-view multi-label learning and Universum learning. Section 3 describes the framework, optimization, computational cost of IMVML-U. Section 4 validates the performances of IMVML-U with extensive experiments on 5 MVML data sets. Section 5 concludes this study and point the future directions of our research.

2 Related work

In this section, we review the related work about incomplete multi-view multi-label learning and Universum learning as below.

2.1 Incomplete multi-view multi-label learning

Incomplete multi-view multi-label learning aims to process the data sets whose instances are represented by multiple feature sets, labeled by multiple class labels, and missing some useful information. Current incomplete multi-view multi-label learning are mainly developed on the base of matrix factorization

[18, 19, 20], kernel-based learning [4, 21, 22], deep neural networks [23, 24, 25], etc.

• In terms of the learning machines based on matrix factorization, Liu et. al [18] present a virtual-label guided matrix factorization (VLMF) to facilitate the incomplete clustering task. In VLMF, the graph regularization of each view is employed to mine the geometrical structure of data and the virtual label is used to guide matrix factorization and integrate the clustering process and the consensus latent representation learning of data into a joint framework; Wang et. al [19] present a cross-view multi-layer perceptron (CVMLP) to process the incomplete data processing task which integrates an auto-encoder module, cross-view classification loss, masked contrastive learning, and variance loss into a unified framework; Luong et. al [20] develop a deep non-negative matrix factorization-based framework for effective clustering task by uncovering both the non-linear relationships and the intrinsic components of the data. In their work, both the consensus and complementary information present in multiple views are sufficiently learned with the effective use of constraints such as normalized cut-type and orthogonal.

• In terms of the learning machines based on kernel-based learning, Zhang et. al [4] propose a transductive semi-supervised incomplete fuzzy system and they integrate the missing view imputation, pseudo label learning, and fuzzy system training to a single process to enable mutual learning with kernel technology used; Xiao et. al [21] develop a SVM-based multi-view multi-label learning method with privileged information learning and they implement both the consensus principle and complementarity principle in multi-view multi-label learning with kernel matrices used; Zhao et. al [22] develop a method that uses view-specific labels and label-feature dependence maximization. They first assume that each view and its corresponding label space have a smooth local structure. In this way, the view-specific label learning model is constructed, enhancing the consistency and complementarity of label space information. Then, multiple multi-label classifiers are constructed by maximizing label-feature dependence. Finally, the linear classification model is extended to the nonlinear with kernel-based learning used, and the prediction stage is combined with the contribution weight of each view.

• In terms of the learning machines based on deep neural networks, Li et. al [23] develop a multi-modal recommendation framework that integrates graph neural networks (GNNs) with deep mutual learning techniques. In their proposal, they use the mutual knowledge distillation technique to collaboratively train multiple uni-modal bipartite user-item graphs and then to generate uni-modal embeddings which allows the model to uncover and synchronize the latent semantic relationships among different modalities; Zhao et. al [24] develop a neural network multi-view multi-label learning framework to solve the problem of consistency and diversity among views and in their work, the label correlation factor is in addition to the classification model, and the view contribution factor is added to the prediction model; Sun et. al [25] develop a deep multi-view prototype-based disambiguation approach and specifically,

they innovatively employ the deep neural network for multi-view ambiguously-labeled image classification to enhance the representation ability, which makes use of the information fusion between multiple views.

2.2 Universum learning

Universum learning is a good method to enhance the data information and according to the introduction by [26, 27], Universum learning can generate a new data set and assist the classification of the target data set and in the traditional hand-written digits classification tasks. Many learning machines are developed on the base of Universum. For example, Liu et. al [28] develop a semi-supervised learning with Universum algorithm based on boosting technique, and focus on situations where only a few labeled instances are available. In this article, the developed method can improve text classification performance and the training error drops exponentially fast when each weak classifier is slightly better than random guessing; Richhariya et. al [29] develop an Universum least squares twin parametric-margin support vector machine (ULSTPMSVM) for classification problems which has a high generalization performance with lesser training time in comparison to existing algorithms; Han et. al [30] develop an Universum-inspired supervised contrastive learning (UniCon), which incorporates mixup strategy to generate mixup-induced Universum as Universum negatives and pushes them apart from anchor instances of the target classes. Their approach not only improves mixup with hard labels, but also innovates a novel measure to generate Universum data. With a linear classifier on the learned representations, UniCon shows state-of-the-art performance on various data sets.

3 The main framework of IMVML-U

This section shows the framework of proposed IMVML-U including its objective function, optimization, computational cost, etc.

3.1 Preliminary

Suppose $X = [X^1;\ X^2;\ ...;\ X^V]$ is a data set with V views and n instances. Then its corresponding labels are $Y = [Y^1;\ Y^2;\ ...;\ Y^V]$. Here, $X^v \in \mathbb{R}^{d_v \times n}$, $Y^v \in \mathbb{R}^{c_v \times n}$, d_v represents the number of features and c_v represents the one of labels for vth view. On the base of $[X;\ Y]$, we define four correlation matrices, they are instance-instance correlation matrix $S^v \in \mathbb{R}^{n \times n}$, feature-feature correlation matrix $P^v \in \mathbb{R}^{d_v \times d_v}$, label-label correlation matrix $Q^v \in \mathbb{R}^{c_v \times c_v}$, and label-feature correlation matrix $W^v \in \mathbb{R}^{c_v \times d_v}$. Each element in a matrix represents the corresponding correlation. Then we define two index matrices further, one is feature-missing index matrix $E^v \in \mathbb{R}^{d_v \times n}$, the other is label-missing index matrix $F^v \in \mathbb{R}^{c_v \times n}$. Similarly, each element in these matrices indicates the missing or not-missing case. For example, the element in pth row

and qth column of E^v, i.e., $E^v_{p,q} = 1$ indicates that in vth view, the pth feature of qth instance is not missing, otherwise, $E^v_{p,q} = 0$ indicates a missing case.

Moreover, since each instance can be represented by three representations at least, so the above items can also have the below corresponding representations. The corresponding cross-view representations are $X^{ij} \in \mathbb{R}^{d_{ij}\times n}$, $Y^{ij} \in \mathbb{R}^{c_{ij}\times n}$, $S^{ij} \in \mathbb{R}^{n\times n}$, $P^{ij} \in \mathbb{R}^{d_{ij}\times d_{ij}}$, $Q^{ij} \in \mathbb{R}^{c_{ij}\times c_{ij}}$, $W^{ij} \in \mathbb{R}^{c_{ij}\times d_{ij}}$, $E^{ij} \in \mathbb{R}^{d_{ij}\times n}$, $F^{ij} \in \mathbb{R}^{c_{ij}\times n}$ and the corresponding consensus-view representations are $X^c \in \mathbb{R}^{k\times n}$, $Y^c \in \mathbb{R}^{l\times n}$, $S^c \in \mathbb{R}^{n\times n}$, $P^c \in \mathbb{R}^{k\times k}$, $Q^c \in \mathbb{R}^{l\times l}$, $W^c \in \mathbb{R}^{l\times k}$, $E^c \in \mathbb{R}^{k\times n}$, $F^c \in \mathbb{R}^{l\times n}$. Here, d_{ij} and c_{ij} are the number of features and labels for cross-view representation about ith and jth views. k and l are the number of features and labels for the consensus-view representation about all views.

Then according to the introduction about subsection 1.2, the framework of IMVML-U consists of the below subsections.

3.2 Construction of preservation model

As is known to all, for any data, their correlation matrices can store the comprehensive relationships of all instances. Thus no matter the correlation matrices be complete or incomplete, the available information should be preserved. To this end, take the vth view as example, we let $\widetilde{Q^v}$, $\widetilde{W^v}$, $\widetilde{S^v}$ be the complete correlation matrices corresponding to Q^v, W^v, S^v, respectively and then use the below Eq. (1) to reflect the preservation case. Moreover, since data has three representations at least, so the Eq. (1) can also be extended to the cross-view version (Eq. (2)) and consensus-view version (Eq. (3)).

$$\begin{aligned}\sum_{v=1}^{V}\Big[&||(\widetilde{P^v}-P^v)\odot(E^vE^{vT})||_F^2+||(\widetilde{Q^v}-Q^v)\odot(F^vF^{vT})||_F^2\\&+||(\widetilde{W^v}-W^v)\odot(F^vE^{vT})||_F^2+||(\widetilde{S^v}-S^v)\odot(E^{vT}E^v)||_F^2\\&+||(\widetilde{S^v}-S^v)\odot(F^{vT}F^v)||_F^2\Big]\end{aligned}\tag{1}$$

$$\begin{aligned}\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\Big[&||(\widetilde{P^{ij}}-P^{ij})\odot(E^{ij}E^{ij^T})||_F^2\\&+||(\widetilde{Q^{ij}}-Q^{ij})\odot(F^{ij}F^{ij^T})||_F^2+||(\widetilde{W^{ij}}-W^{ij})\odot(F^{ij}E^{ij^T})||_F^2\\&+||(\widetilde{S^{ij}}-S^{ij})\odot(E^{ij^T}E^{ij})||_F^2+||(\widetilde{S^{ij}}-S^{ij})\odot(F^{ij^T}F^{ij})||_F^2\Big]\end{aligned}\tag{2}$$

$$\begin{aligned}&||(\widetilde{P^c}-P^c)\odot(E^cE^{cT})||_F^2+||(\widetilde{Q^c}-Q^c)\odot(F^cF^{cT})||_F^2\\&+||(\widetilde{W^c}-W^c)\odot(F^cE^{cT})||_F^2+||(\widetilde{S^c}-S^c)\odot(E^{cT}E^c)||_F^2\\&+||(\widetilde{S^c}-S^c)\odot(F^{cT}F^c)||_F^2]\end{aligned}\tag{3}$$

3.3 Construction of 'individual-mutual balance' model

Since current data always including some incomplete instances, thus we should recover them during the design of a learning machine. But in order to pursue a better recover result, in this study, we suppose that the recovered information should keep a 'individual-mutual balance'. Here, 'individual' indicates that we can recover more missing information as many as possible, namely, pursue the recovering quantity. 'mutual' indicates that we pursue the recovering quality, namely, the recovered information has a higher quality and the recovered feature information can predict the labels as accurate as possible. 'balance' indicates that during the process of recovering the missing information, we cannot blindly pursue the recovering quantity or recovering quality and consider the recovering balance. Since the data has three representations at least, so the 'individual-mutual balance'-model should consist of the below three sub-models.

At the vth view, the individual, mutual, and balance parts are given in Eqs. (4)$\sim$(6), respectively.

$$\sum_{v=1}^{V}\Big[||X^v - P^v X^v||_F^2 + ||X^v - X^v S^v||_F^2 + ||Y^v - Y^v S^v||_F^2 + ||Y^v - Q^v Y^v||_F^2\Big] \tag{4}$$

$$\sum_{v=1}^{V}\Big[||X^v - W^{v^T} Y^v||_F^2 + ||Y^v - W^v X^v||_F^2\Big] \tag{5}$$

$$\frac{Eq.(4)}{Eq.(5)} \tag{6}$$

Then for the cross-view form, the individual, mutual, and balance parts are given in Eqs. (7)$\sim$(9), respectively.

$$\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\Big[||X^{ij} - P^{ij} X^{ij}||_F^2 + ||X^{ij} - X^{ij} S^{ij}||_F^2 + ||Y^{ij} - Y^{ij} S^{ij}||_F^2 + ||Y^{ij} - Q^{ij} Y^{ij}||_F^2\Big] \tag{7}$$

$$\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\Big[||X^{ij} - W^{ij^T} Y^{ij}||_F^2 + ||Y^{ij} - W^{ij} X^{ij}||_F^2\Big] \tag{8}$$

$$\frac{Eq.(7)}{Eq.(8)} \tag{9}$$

Finally, corresponding consensus-forms are given in Eqs. (10)∼(12), respectively.

$$||X^c - P^c X^c||_F^2 + ||X^c - X^c S^c||_F^2 + ||Y^c - Y^c S^c||_F^2 + ||Y^c - Q^c Y^c||_F^2 \tag{10}$$

$$||X^c - W^{cT} Y^c||_F^2 + ||Y^c - W^c X^c||_F^2 \tag{11}$$

$$\frac{Eq.(10)}{Eq.(11)} \tag{12}$$

3.4 Construction of self-adaptive model

As we know, the correlation matrices can store comprehensive relationships of instances. In order to get more accurate correlation matrices, we make full use of 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.

In our knowledge, the correlations can be measured and changed with the following laws. (*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.

In order to realize the first law, we take the information from vth view as the example and utilize two regularizer terms. One is related with X^v which includes the information about features, and another is related with $Y^v = W^v X^v$ 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.

$$\sum_{a=1}^{n}\sum_{b=1}^{n} S_{ab}^v ||x_a^v - x_b^v|| = tr(X^v L_{S^v} X^{vT}) \tag{13}$$

$$\begin{aligned}\sum_{a=1}^{n}\sum_{b=1}^{n} S_{ab}^v ||y_a^v - y_b^v|| &= tr(Y^v L_{S^v} Y^{vT}) \\ &= tr(W^v X^v L_{S^v} X^{vT} W^{vT})\end{aligned} \tag{14}$$

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.

Then for V views, the total model for within-view representations according to the first law, is given in Eq. (15) and its cross-view representation and the consensus-view one are given in Eqs. (16) and (17), respectively.

$$\sum_{v=1}^{V}\left[tr(X^v L_{S^v} X^{vT}) + tr(W^v X^v L_{S^v} X^{vT} W^{vT})\right] \tag{15}$$

$$\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\left[tr(X^{ij} L_{S^{ij}} X^{ij^T}) + tr(W^{ij} X^{ij} L_{S^{ij}} X^{ij^T} W^{ij^T})\right] \tag{16}$$

$$tr(X^c L_{S^c} X^{cT}) + tr(W^c X^c L_{S^c} X^{cT} W^{cT}) \tag{17}$$

In order to realize the second law, we also take vth view as an example, we utilize a regularizer term. Since $Y^v = W^v X^v$ 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.

$$\begin{aligned}\sum_{p=1}^{c_v}\sum_{q=1}^{c_v} Q^v_{pq}||y_{vp} - y_{vq}|| &= tr(Y^{vT} L_{Q^v} Y^v) \\ &= tr(X^{vT} W^{vT} L_{Q^v} W^v X^v)\end{aligned} \tag{18}$$

where Q^v_{pq} 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.

Similar, for V views, for the second law, the total model for within-view representations is given in Eq. (19) and its cross-view representation and the consensus-view one are given in Eqs. (20) and (21), respectively.

$$\sum_{v=1}^{V} tr(X^{vT} W^{vT} L_{Q^v} W^v X^v) \tag{19}$$

$$\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V} tr(X^{ij^T} W^{ij^T} L_{Q^{ij}} W^{ij} X^{ij}) \tag{20}$$

$$tr(X^{cT} W^{cT} L_{Q^c} W^c X^c) \tag{21}$$

In order to realize the third law, we still utilize a regularizer term for vth view. Since rth (or sth) column x_{cr} (or x_{cs}) describes the information of rth (or sth) feature for the X^c. Then we design a corresponding regularizer term

as below.

$$\sum_{r=1}^{k}\sum_{s=1}^{k} S_{rs}^{v}||x_{cr} - x_{cs}|| = tr(X^{c}L_{S^{v}}X^{cT}) \tag{22}$$

where S_{rs}^{v} describes the feature-feature correlation between rth and sth features. Then similarly, for information about labels, we can get

$$tr(Y^{c}L_{S^{v}}Y^{cT}) \tag{23}$$

Then, for the third law, the total models are given in Eqs. (24), (25) and (26), respectively.

$$\sum_{v=1}^{V} tr(X^{c}L_{S^{v}}X^{cT}) \tag{24}$$

$$\sum_{v=1}^{V} tr(Y^{c}L_{S^{v}}Y^{cT}) \tag{25}$$

$$\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\left[tr(X^{ij}L_{S^{i}}X^{ijT}) + tr(Y^{ij}L_{S^{i}}Y^{ijT})\right] \tag{26}$$

3.5 Construction of nuclear-norm model

Since low-rank characteristic of representation can be enforced by matrix nuclear norm [31, 32], 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

$$||\theta^{v}||_{*}^{2} = \sum_{C\in S,P,Q,W} \eta_{C^{v}}||C^{v}||_{*}^{2} \tag{27}$$

$$||\theta^{ij}||_{*}^{2} = \sum_{C\in S,P,Q,W} \eta_{C^{ij}}||C^{ij}||_{*}^{2} \tag{28}$$

$$||\theta^{c}||_{*}^{2} = \sum_{C\in S,P,Q,W} \eta_{C^{c}}||C^{c}||_{*}^{2} \tag{29}$$

and η_{x} is the penalty parameter for different matrices to balance the importance of the corresponding constraints.

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

$$\sum_{v=1}^{V}||\theta^{v}||_{\star}^{2} + \sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}||\theta^{ij}||_{\star}^{2} + ||\theta^{c}||_{\star}^{2} \tag{30}$$

Table 1 Computation of $\frac{\partial \mathcal{L}}{\partial \mathcal{C}}$ in Eq. (31) in 9 forms and $\mathcal{A}$, $\mathcal{B}$, $\mathcal{D}$ describe different terms.

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

3.6 objective function

The objective function consists of the above four parts and it is given as below.

$$\begin{aligned}\min \Big[& Eq.(1)+Eq.(2)+Eq.(3)+Eq.(4)+Eq.(5)+Eq.(6)+Eq.(7) \\ & +Eq.(8)+Eq.(9)+Eq.(10)+Eq.(11)+Eq.(12)+Eq.(15)+Eq.(16) \\ & +Eq.(17)+Eq.(19)+Eq.(20)+Eq.(21)+Eq.(24)+Eq.(25)+ \\ & Eq.(26)+Eq.(30)\Big]\end{aligned} \tag{31}$$

3.7 Optimization

For the optimization of Eq. (31), we also refer to [33] and gradient descent way is adopted. Namely, we update a correlation and leave the others fixed, 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. (31), 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.

Then the optimal results are $\tau^{\omega-o1}$ where $\tau \in \{P, Q, W, S\}$ and $\omega \in \{v, ij, c\}$. Once we get the current optimal results, we can recover the data initial with the terms like $X^{v-o1} = P^{v-o1}X^v$ and $Y^{v-o1} = Q^{v-o1}Y^v$.

3.8 The enhancement for data set (U part of IMVML-U)

We let $\frac{1}{n-1}$ be the threshold. The at the vth view, if the corresponding instance-instance correlation value between pth instance and qth instance, i.e., $S^v_{p,q} > \frac{1}{n-1}$, we think these two instances have a higher similarity and we can use $\frac{x^v_p + x^v_q}{2}$ to generate a new instance and N new instances are generated and compose of X^v_N. Here, x^v_p and x^v_q are the pth column and qth column of X^{v-o1}. Then for cross-view representation and consensus-view representation, the results are similar. Then we have $X^{v'} = \left[X^{v-o1};\ X^v_N\right] \in \mathbb{R}^{d_v \times (n+N)}$, $X^{ij'} = \left[X^{ij-o1};\ X^{ij}_N\right] \in \mathbb{R}^{d_{ij} \times (n+N)}$, $X^{c'} = \left[X^{c-o1};\ X^c_N\right] \in \mathbb{R}^{d_c \times (n+N)}$.

Then we let $Y^v_N = W^{v-o1} X^v_N$, $Y^{ij}_N = W^{ij-o1} X^{ij}_N$, $Y^c_N = W^{c-o1} X^c_N$ and we have $Y^{v'} = \left[Y^{v-o1};\ Y^v_N\right] \in \mathbb{R}^{c_v \times (n+N)}$, $Y^{ij'} = \left[Y^{ij-o1};\ Y^{ij}_N\right] \in \mathbb{R}^{c_{ij} \times (n+N)}$, $Y^{c'} = \left[Y^{c-o1};\ Y^c_N\right] \in \mathbb{R}^{c_c \times (n+N)}$.

Now the data set has been enhanced.

3.9 The second update of the data set

Once we enhanced the data set, its correlations and information should be updated again.

First, we update the S^{v-o1}, S^{ij-o1}, S^{c-o1}. The dimension of these three matrices are $(n+N) \times (n+N)$ and each element in these matrices are decided by the Eular distance between two instances and the matrices are normalized.

Second, we update the matrices by the above section 3.7 and we can get optimal results which are represented by $\tau^{\omega-o2}$ where $\tau \in \{X, Y, P, Q, W, S\}$ and $\omega \in \{v, ij, c\}$.

After that, we can get the newest recovered and enhanced data sets are below. $X^{v''} = X^{v'} S^{vo-2}$, $Y^{v''} = Y^{v'} S^{vo-2}$. For the cross-view and consensus-view representations, the results are similar.

3.10 The computational cost of IMVML-U

According to the above contents, the computations of the correlations determine the computational cost of IMVML-U. 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 IMVML-U is less than $NDO(Cn^2)$ where N is the number of correlations. Compared with some existing algorithms [34, 35, 36, 37] whose computational costs are $O(n^3)$-levels, the computational cost of IMVML-U is much smaller.

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

data sets	instances-labels-views-ALPI
Corel5k	4999-260-6-3.396
Espgame	20770-268-6-4.686
IAPRTC12	19627-291-6-5.719
Mirflickr	25000-38-6-4.716
Pascal07	9963-20-6-1.465

4 Experiments

4.1 Experimental settings

Data setting: we refer to [7] and adopt 5 MVML data sets for experiments (see Table 2). Here, Corel5k can be downloaded from http://archive.ics.uci.edu/ml/datasets/Corel+Image+Features, Espgame can be downloaded from https://www.kaggle.com/datasets/parhamsalar/espgame, IAPRTC12 can be downloaded from https://www.imageclef.org/photodata, Mirflickr can be downloaded from https://press.liacs.nl/mirflickr/, Pascal07 can be downloaded from http://host.robots.ox.ac.uk/pascal/VOC/.

Compared algorithms: in order to validate the performance of IMVML-U, we adopt 3 classical MVML algorithms for comparisons. They are NAIM^{3}L [8], iMVMAL [14], and DD-IMvMLC-net [7].

Parameter setting: for NAIM^{3}L, iMVMAL, and DD-IMvMLC-net, their optimal parameter values can be selected according to the corresponding references. For IMVML-U, 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 3.7. For penalty parameters ηs, their optimal parameter values can be selected from the set $\{0.1, 0.2, ..., 0.8, 0.9\}$. Moreover, in order to get the incomplete setting, we remove 50% features and 50% labels in random.

Evaluation: for convenience, we adopt training time, AUC (adapted area under curve) and AP (average precision) as the main metrics for experimental comparisons.

How to select the optimal parameters: In order to select the optimal parameter values for IMVML-U, avoid over-fitting in model selection, and ensure the authenticity of experimental results, we refer to the strategy given in [33] and for each data set, 70% instances are selected for training and validation and the rest are selected for testing. Then according to the results about AUC, the optimal parameter values can be selected with ten-fold cross-validation. 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: in our study, we select RedHat Linux Enterprise 9.0 as the operation system and the coding environment is MATLAB 2022b.

4.2 Performance and efficiency study

Here, we use Fig. 1 to demonstrate the AUC, precision, and corresponding standard deviations of all algorithms on all data sets. Then Fig. 2 demonstrates the training time comparison results. Moreover, Fig. 3 demonstrates the convergence of IMVML-U on all used data sets. In this figure, top sub-figures show the convergence curve and the bottom sub-figures report the iteration index[1].

According to these figures, it can be found that in terms of AUC and precision, although the performance of IMVML-U 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 IMVML-U, it can get a convergence within 30 iterations in our experiments.

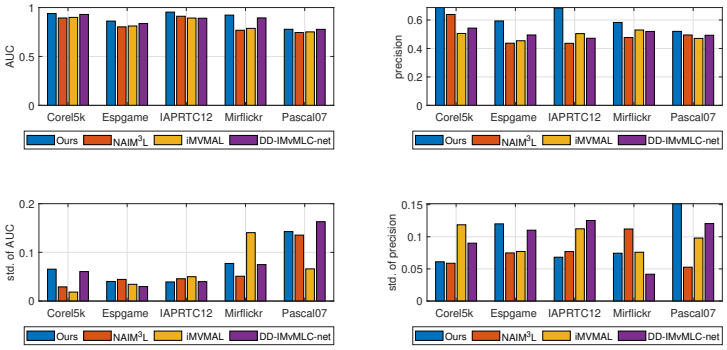


Fig. 1 AUC, precision, and corresponding standard deviation (std.) comparisons.

4.3 Statistical analysis

To validate the effectiveness of the IMVML-U further, we adopt Friedman-Nemenyi statistical test [38] for statistical analysis and check if the differences between IMVML-U 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 IMVML-U and others, and corresponding statistical values (see Fig. 4-(a) and Fig. 4-(b)).

Then refer to [38], (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 experiments, then we can get Friedman statistic as follows. $\chi_F^2 = \frac{12\times N}{k(k+1)}[1.0000^2 + 3.6000^2 + 3.0000^2 + 2.4000^2 - \frac{k(k+1)^2}{4}] = 11.16$, $F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} = 11.625$,

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

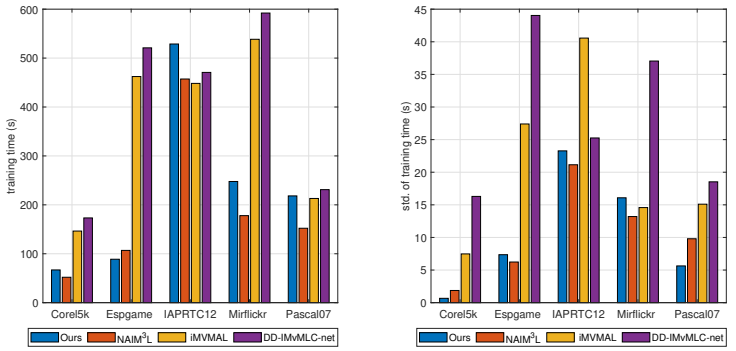


Fig. 2 Training time and the corresponding standard deviation (std.) comparisons.

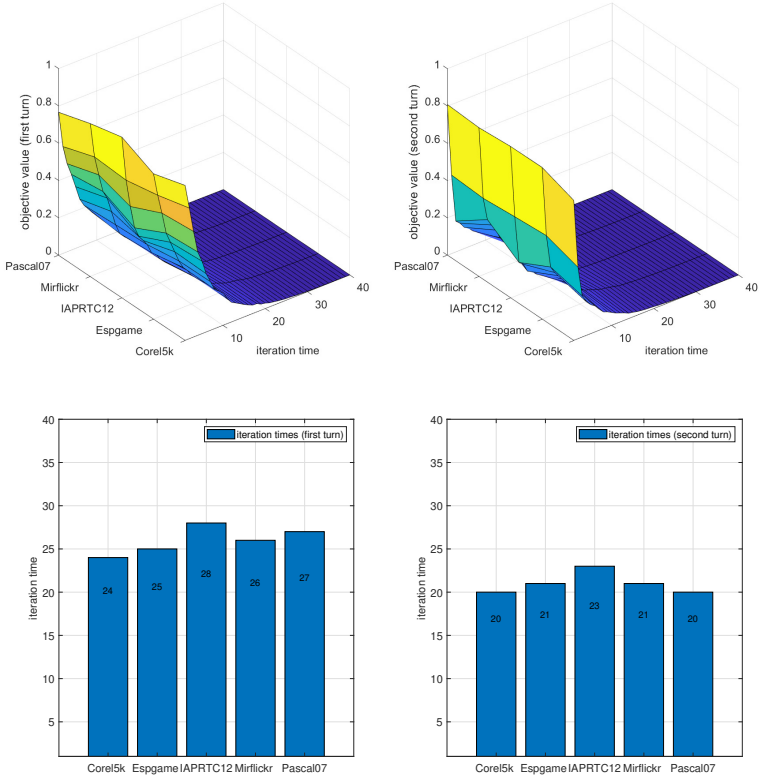


Fig. 3 Convergence of IMVML-U on the used data sets and the objective value has been normalized.

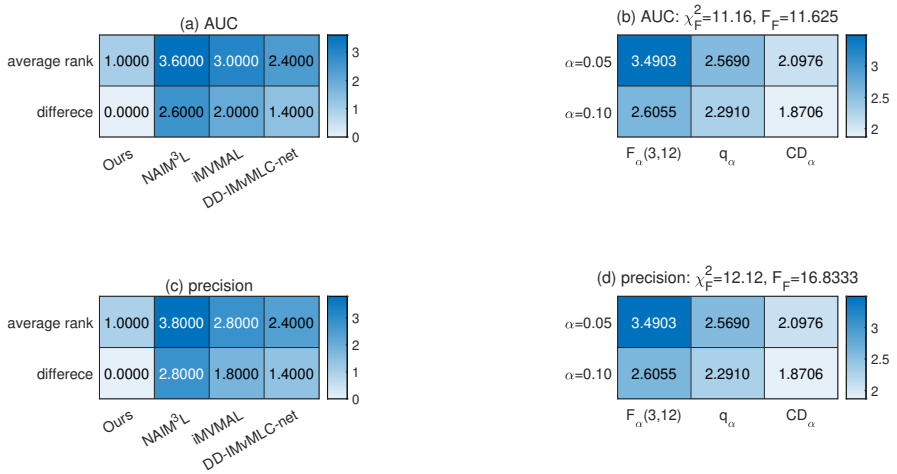


Fig. 4 Statistical analysis for IMVML-U in terms of AUC and precision.

$F_{0.05}(k-1,(k-1)(N-1)) = F_{0.05}(3,12) = 3.4903$, 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 IMVML-U and NAIM^{3}L are larger than $CD_{0.05}$ and $CD_{0.10}$, so we say on this case, the performance of IMVML-U is significant better than $NAIM^3L$. Similar, IMVML-U is significant better than iMVMAL when $\alpha = 0.10$.

(2) Similarly, Fig. 4-(c) and Fig. 4-(d) demonstrates the statistical analysis about IMVML-U on the precision. According to this figure, in terms of precision, our IMVML-U is significant better than NAIM^{3}L.

5 Conclusions and future work

To process multi-view multi-label data sets more effectively, this study develops an incomplete multi-view multi-label learning with Universum (IMVML-U). Different from state-of-the-art algorithms, the IMVML-U 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- and mutual- balance model introduced. Experiments completed on several benchmark data sets demonstrate the superiority of IMVML-U 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 IMVML-U. Thus, in the future, we plan to design a new algorithm on the basis of IMVML-U to solve these online study problems.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62276164, in part by the 'Science and technology innovation action plan' Natural Science Foundation of Shanghai under Grant 22ZR1427000, in part by Shanghai Oriental Talent Program-Youth Program. (Corresponding author: Changming Zhu)

References

[1] J.H. Ma, H.Y. Sun, and T. Zhu, "Multiview multilabel classification with group-based feature and label selection," *IEEE Transactions on Consumer Electronics*, 2023, DOI: 10.1109/TCE.2023.3278457.

[2] Q.Y. Tan, G.X. Yu, J. Wang, C. Domeniconi, and X.L. Zhang, "Individuality- and commonality-based multiview multilabel learning," *IEEE Transactions on Cybernetics*, vol. 51, no. 3, pp. 1716-1727, 2021.

[3] G.X. Yu, Y.Y. Xing, J. Wang, C. Domeniconi, and X.L. Zhang, "Multi-view multi-instance multilabel active learning, " *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 9, pp. 4311-4321, 2022.

[4] W. Zhang, Z.H. Deng, Q.D. Lou, T. Zhang, K.S. Choi, and S.T. Wang, "Takagi-Sugeno-Kang fuzzy system towards label-scarce incomplete multi-view data classification," *Information Sciences*, vol. 647, 119466, 2023.

[5] X.Y. Liu and P. Song, "Incomplete multi-view clustering via virtual-label guided matrix factorization," *Expert Systems with Applications*, vol. 210, 118408, 2022.

[6] Z. Wang, H. Zhou, P. Zhong, and H. Zou, "Cross-view multi-layer perceptron for incomplete multi-view learning," *Applied Soft Computing*, vol. 157, 111510, 2024.

[7] J. Wen, C.L. Liu, S.J. Deng, Y.C. Liu, L.K. Fei, K. Yan, and Y. Xu, "Deep Double Incomplete Multi-View Multi-Label Learning With Incomplete Labels and Missing Views," *IEEE Transactions on Neural Networks and Learning Systems*, 2023, DOI: 10.1109/TNNLS.2023.3260349.

[8] X. Li and S.C. Chen, "A concise yet effective model for non-aligned incomplete multi-view and missing multi-label learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 5918-5932, 2022.

[9] Y.S. Zhang, J. Wu, Z.H. Cai, and P.S. Yu, "Multi-view multi-label learning with sparse feature selection for image annotation," *IEEE Transactions on Multimedia*, vol. 22, no. 11, pp. 2844-2857, 2020.

[10] B.Y. Qian, X. Wang, J.P. Ye, and I. Davidson, "A reconstruction error based framework for multi-label and multi-view learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 3, pp. 594-607, 2015.

[11] Y.Q. Li, B.Y. Wu, B. Ghanem, Y.P. Zhao, H.X. Yao, and Q. Ji, "Facial action unit recognition under incomplete data based on multi-label learning with missing labels," *Pattern Recognition*, vol. 60, pp. 890-900, 2016.

[12] C.L. Liu, J. Wen, X.L. Luo, and Y. Xu, "Incomplete multi-view multi-label learning via label-guided masked view- and category-aware transformers," *in 2023 Proceedings of the AAAI conference on artificial intelligence (AAAI)*, Washington DC, United States, 2023, pp. 8816-8824.

[13] K. Maeda, S. Takahashi, T. Ogawa, and M. Haseyama, "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, 2020, pp. 46-50.

[14] C.W. Qu, K.M. Wang, H. Zhang, G.X. Yu, and C. Domeniconi, "Incomplete multi-view multi-label active learning," *in 2021 IEEE International Conference on Data Mining (ICDM)*, Auckland, New Zealand, 2021, pp. 1294-1299.

[15] L.J. Sun, P. Ye, G.Y. Lyu, S.H. Feng, G.J. Dai, and H. Zhang, "Weakly-supervised multi-label learning with noisy features and incomplete labels," *Neurocomputing*, vol. 413, pp. 61-71, 2020.

[16] B. Richhariya and M. Tanveer, "A reduced Universum twin support vector machine for class imbalance learning," *Pattern Recognition*, vol. 102, 107150, 2020.

[17] C.L. Liu, W.H. Hsaio, C.H. Lee, T.H. Chang, and T.H. Kuo, "Semi-supervised text classification with Universum learning," *IEEE Transactions on Cybernetics*, vol. 46, no. 2, pp. 462-473, 2016.

[18] X.Y. Liu and P. Song, "Incomplete multi-view clustering via virtual-label guided matrix factorization," *Expert Systems with Applications*, vol. 210, 118408, 2022.

[19] Z. Wang, H. Zhou, P. Zhong, and H. Zou, "Cross-view multi-layer perceptron for incomplete multi-view learning," *Applied Soft Computing*, vol. 157, 111510, 2024.

[20] K. Luong, R. Nayak, T. Balasubramaniam, and M.A. Bashar, "Multi-layer manifold learning for deep non-negative matrix factorization-based multi-view clustering," *Pattern Recognition*, vol. 131, 108815, 2022.

[21] Y.S. Xiao, J.F. Chen, B. Liu, L. Zhao, X.J. Kong, Z.F. Hao, "A new multi-view multi-label model with privileged information learning," *Information Sciences*, vol. 656, 119911, 2024.

[22] D.W. Zhao, Q.W. Gao, Y.X. Lu, and D. Sun, "Learning view-specific labels and label-feature dependence maximization for multi-view multi-label classification", *Applied Soft Computing*, vol. 124, 109071, 2022.

[23] J.N. Li, C.Q. Yang, G.H. Ye, and Q.V.H. Nguyen, "Graph neural networks with deep mutual learning for designing multi-modal recommendation systems", *Information Sciences*, vol. 654, 119815, 2024.

[24] D.W. Zhao, Q.W. Gao, Y.X. Lu, D. Sun, and Y.S. Cheng, "Consistency and diversity neural network multi-view multi-label learning", *Knowledge-Based Systems*, vol. 218, 106841, 2021.

[25] S.D. Sun, X.T. Yu, and Y.J. Tian, "Multi-view prototype-based disambiguation for partial label learning", *Pattern Recognition*, vol. 141, 109625, 2023.

[26] J. Weston, R. Collobert, F. Sinz, L. Bottou, and V. Vapnik, "Inference with the Universum", *in 23rd Proceedings of International Conference on Machine Learning (ICML)*, pp. 1009-1016, 2006.

[27] O. Chapelle, A. Agarwal, F. Sinz, and B. Schölkopf, "An analysis of inference with the Universum", *in Advances in Neural Information Processing Systems (NIPS)*, vol. 20, pp. 1369-1376, 2007.

[28] C.L. Liu, W.H. Hsaio, C.H. Lee, T.H. Chang, and T.H. Kuo, "Semi-supervised text classification with Universum learning," *IEEE Transactions on Cybernetics*, vol. 46, no. 2, pp. 462-473, 2016.

[29] B. Richhariya and M. Tanveer, "Universum least squares twin parametric-margin support vector machine," *in 2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, DOI: 10.1109/IJCNN48605.2020.9206865.

[30] A.Y. Han, C.X. Geng, and S.C. Chen, "Universum-Inspired Supervised Contrastive Learning," *IEEE Transactions on Image Processing*, vol. 32, pp. 4275-4286, 2023.

[31] T. Zhou, C.Q. Zhang, X. Peng, H. Bhaskar, J. Yang, "Dual shared-specific multiview subspace clustering," *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3517–3530, 2020.

[32] X. Peng, C.Y. Lu, Z. Yi, and H.J. Tang, "Connections between nuclear-norm and frobenius-norm-based representations," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 218–224, 2018.

[33] C.M. Zhu, Y.C. Liu, D.Q. Miao, Y.L. Dong, and W. Pedrycz, "Within-cross- consensus-view representation-based multi-view multi-label learning with incomplete data," *Neurocomputing*, vol. 557, 126729, 2023.

[34] S.E. Hajjar, F. Dornaika, and F. Abdallah, "One-step multi-view spectral clustering with cluster label correlation graph," *Information Sciences*, vol. 592, pp. 97–111, 2022.

[35] G.S. Cui and Y. Li, "Nonredundancy regularization based nonnegative matrix factorization with manifold learning for multiview data representation," *Information Fusion*, vol. 82, pp. 86–98, 2022.

[36] Y. Xie, J.Y. Liu, Y.Y. Qu, D.C. Tao, W.S. Zhang, L.Q. Dai, and L.Z. Ma, "Robust kernelized multiview self-representation for subspace clustering,"

IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 868–881, 2021.

[37] C.Q. Zhang, H.Z. Fu, Q.H. Hu, X.C. Cao, Y. Xie, D.C. Tao, and D. Xu, "Generalized latent multi-view subspace clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 1, pp. 86–99, 2020.

[38] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 1–30, 2006.