Improved Multi-view Multi-label Learning with Incomplete Views and Labels

Changming Zhu
College of Information Engineering
Shanghai Maritime University, SMU
Shanghai, China
Email: cmzhu@shmtu.edu.cn

Duoqian Miao
Department of Computer Science and Technology
Tongji University, TJU
Shanghai, China
Email: dqmiao@tongji.edu.cn

Rigui Zhou
College of Information Engineering
Shanghai Maritime University, SMU
Shanghai, China
Email: rgzhou@shmtu.edu.cn

Lai Wei
College of Information Engineering
Shanghai Maritime University, SMU
Shanghai, China
Email: weilai@shmtu.edu.cn

***Abstract*—Multi-view multi-label learning has attracted the attention of many scholars and widely used in multiple fields. While in real-world applications, due to the lack of manpower and equipment failure, the data sets to be processed maybe loss some labels or views. Moreover, most multi-view multi-label learning methods neglect the global and local label correlations of both the whole data set and each view and the complementary information coming from different views sometimes. Furthermore, some methods ignore such a phenomenon that each class label might be determined by some specific features of its own. In order to improve the performance of such methods, in this paper, we develop an improved multi-view multi-label learning with incomplete views and labels (IMVL-IV). In framework of IMVL-IV, the usage of label-specific features makes the decision of label be determined by some specific features rather than all features so that we can pay more attention to portion specific features and save time; the introduction of label correlation matrix offsets the defect of missing labels; the adoption of low-rank assumption matrix restores missing views; global and local label correlations are taken into consideration with clustering technology; a consensus multi-view representation is put to use to encode the complementary information from different views. Different from traditional learning methods, this is the first attempt to design a multi-view multi-label learning method with incomplete views and labels by the learning of label-specific features, label correlation matrix, low-rank assumption matrix, global and local label correlations, and complementary information. Experimental results validate that IMVL-IV achieves a better performance and it is superior to the classical multi-view learning methods and multi-label learning methods.**

***Keywords*-incomplete views and labels; label-specific features; multi-view multi-label; label correlation.**

I. Introduction

A. What is multi-view multi-label data sets

In real-world applications, three kinds of data sets are widely encountered, i.e., multi-label, multi-view, and multi-view multi-label data sets [1].

Multi-label data sets consist of instances with multiple class labels. A classical example is that a scene image can be annotated with several tags [2]. Please see Fig. 1, a data set consists of two scene images (instances) and four tags (class labels). These four labels are nature, landscape, history, oil painting and '1' indicates yes, '0' represents 'no'. For the two instances, one is an oil painting which shows a natural landscape, i.e., a country in China, thus its label is $(1, 1, 0, 1)$, the other is an oil painting showing Napoleon Bonaparte who was a historical personage, thus its label is $(0, 0, 1, 1)$. Besides for this example, there are many other examples including a document may belong to multiple topics, and a piece of music may be associated with different genres.

In a multi-view data set, an instance (data point) is represented by multiple forms. Each form is a view. For example, as in Fig. 2, a web page data set consists of multiple web pages and each web page (instance) can be described by text, image, and video. Then, text, image, and video are regarded as three views. Moreover, each instance only has one class label. In Fig. 2, two of them belong to class science and the other two belong to class entertainment and arts.

Different from traditional multi-view data sets and multi-label data sets, multi-view multi-label data sets consists of instances which have multiple views and class labels. In a multi-view multi-label data set, each instance exhibits multiple views and in each view, the instance can be represented by multiple class labels. For example, as in Fig. 3, if there is publicity website about the Imperial Palace and people can appreciate and understand it from multiple aspects including the text introduction, image introduction, and video introduction. Now text, image, and video can be regarded as different views. Then from different views, the content of propaganda can be labeled differently. When it comes to any textual information, it can be treated as an introduction to history; according to image introduction, it can be treated as an oil painting which describes landscape and history; when people listen and watch the video, it can

2375-9259/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDMW.2019.00104

Figure 1. Example of a multi-label data set.

Figure 2. Example of a multi-view data set.

be treated as a stereoscopic introduction about landscape and history rather than an oil painting. Here, history, landscape, oil painting, etc. can be regarded as class labels and such a data set is a multi-view multi-label one.

B. Problems of traditional multi-view multi-label learning methods and previous solutions

Since multi-view multi-label data sets exist in real-world applications widely, thus some related tasks are put forward.

Figure 3. Example of a multi-view multi-label data set.

Among them, classification task is of a general nature.

For multi-view multi-label classification task, Zhu et. al [3] develop a classical solution named multi-view multi-sparsity kernel reconstruction (MMKR for short) model to process multi-class (multi-label) image classification. Given images (including test images and training images) representing with multiple visual features (multi-view), MMKR first maps them into a high-dimensional space, e.g., a reproducing kernel Hilbert space (RKHS), where test images are then linearly reconstructed by some representative training images, rather than all of them. Furthermore a classification rule is proposed to classify test images. Maeday et. al develop a multi-feature fusion based on supervised multi-view multi-label canonical correlation projection (sM2CP) [4]. The proposed method applies sM2CP-based feature fusion to multiple features obtained from various convolutional neural networks (CNNs) whose characteristics are different. Since new fused features with high representation ability can be obtained, performance improvement of multi-label classification is realized. Specifically, in order to tackle the multi-label problem, sM2CP introduces a label similarity information of label vectors into the objective function of supervised multi-view canonical correlation analysis. Thus, sM2CP can deal with complex label information such as multi-label annotation.

While these classical methods always have no ability to process the data set with incomplete views and labels. Qian et. al develop a semi-supervised dimension reduction for multi-label and multi-view learning (SSDR-MML) to solve the challenges including incomplete labels [5]; Due to matrix completion (MC) has recently been introduced as a method for transductive (semi-supervised) multi-label classification, and has several distinct advantages, including robustness to missing data and background noise in both feature and label space, thus Luo et. al propose a multi-view MC (MVMC) framework [6] for multi-view cases. MVMC is robust to noise and can handle incomplete views and labels in multi-view multi-label learning.

While we know, among labels of instances may exist some correlation, for example, label 'nature' and label 'landscape' have a belonging relationship. If the label correlations are shared by all instances, the correlations are global label correlations. If the label correlations are shared only by a data subset, then the correlations are local label correlations. The above methods cannot take the label correlation into consideration. In 2014, Zhang et. al propose a high-order label correlation driven active learning (HoAL) approach that allows the iterative learning algorithm itself to select the informative insatnce-label pairs from which it learns so as to learn an accurate classifier with less annotation efforts. In the selection procedure, pair-wise label correlations and high-order label correlations are adopted [7]; Zhu et. al develop a multi-label learning with global and local label correlation (GLOCAL) model to deal with both the full-label and the missing-label cases, exploiting global and local label correlations simultaneously, through learning a latent label representation and optimizing label manifolds [2]; He et. al develop a multi-label classification approach joint with label correlations, missing labels and feature selection, named MLMF [8]. The proposed MLMF not only makes the joint learning of independent binary classifiers, but also allows the joint learning of multi-label classification and label correlations.

Moreover, we know each class label might be determined by some specific features of its own, and only a partial label set of each instance can be obtained for some real applications. Zhang et. al develop a multi-label learning with label specific features (LIFT). LIFT firstly constructs features specific to each label by conducting clustering analysis on its positive and negative instances, and then performs training and testing by querying the clustering result [9]; Someone develop a multi-label learning with label-specific features by resolving label correlations (MLFC). In MLFC, scholars propose to learn label-specific features using sparsity regularized optimization which cover the information of label correlations. Indeed, label correlations are represented by additional features generated in the optimization process, and a KNN-like method is designed to obtain label correlations-based features of test data [10]; Huang et. al develop a method to learn label-specific features for multi-label classification with missing labels, named LSML. First, a new supplementary label matrix is augmented from the incomplete label matrix by learning high-order label correlations. Then, a label-specific data representation for each class label is learned, and the multi-label classifier is constructed simultaneously based on it by incorporating the learned high-order label correlations [11].

Furthermore, since among multiple views, since views are always different, thus complementary information coming from different views is always important for the design of a learning machine. Zou et. al develop a multi-view multi-label (abbreviated by MVML) learning algorithm which take multiple features (multi-view) and ensemble learning into account simultaneously [12]. In MVML, they make full use of the complementarity among the views and the base learners of ensemble learning, leading to higher accuracy of image annotation. Wang et. al develop a semi-supervised multi-view multi-label classification learning method based on nonnegative matrix factorization (NMF-SSMM) [13]. In NMF-SSMM, it explores the complementary information by adopting multi-view NMF, regularizes the learned labels of each view towards a common consensus labeling, and obtains the labels of the unlabeled data guided by supervised information.

According to the above contents, we find that for multi-view multi-label learning, incomplete views and labels, global and local label correlation, label-specific features, and complementary information are four necessary issues should be solved.

C. Objectives

In order to solve the above issues, we develop an improved multi-view multi-label learning with incomplete views and labels (IMVL-IV). The framework of IMVL-IV consists of five parts. First, we use label-specific features so that each label is determined by some specific features rather than all features. Second, we introduce a label correlation matrix so as to solve the problem of incomplete labels. Third, we adopt low-rank assumption matrices to restore the missing information of views and solve the problem of incomplete views. Fourth, we use clustering technology to take the global and local label correlations of both the whole data set and each view into consideration. Fifth, we introduce a consensus multi-view representation into this new model so as to encode the complementary information coming from different views.

D. Novelty and contributions

Novelty: in the field of multi-view multi-label learning, it is the first attempt to process a multi-view multi-label data set with incomplete views and labels by the learning of label-specific features, label correlation matrix, low-rank assumption matrix, global and local label correlations, and complementary information. Different from the traditional learning methods, the proposed method can classify and process more complicated data sets well.

Contributions: (1) it has a better ability to process multi-view multi-label data sets; (2) it moves forward research of multi-view multi-label learning.

II. Improved Multi-view Multi-label Learning with Incomplete Views and Labels (IMVL-IV)

A. Data preparation

Suppose there is a multi-view multi-label data set $X = \{x_1, ..., x_i, ..., x_n\} \in \mathbb{R}^{d\times n}$ with v views and n instances. Dimension $d = \sum_{j=1}^{v} d_j$ where d_j is the number of features or dimension in jth view. If x_{ib}^j is the bth feature of ith instance in its jth view, then $x_i^j = \{x_{i1}^j, ..., x_{ib}^j, ..., x_{id_j}^j\}^T \in \mathbb{R}^{d_j\times 1}$ is the representation of ith instance in its jth view. Under this definition, $x_i = \{x_i^1, ..., x_i^j, ..., x_i^v\} \in \mathbb{R}^{d\times 1}$ represents ith instance and $X^j = \{x_1^j, ..., x_i^j, ..., x_n^j\} \in \mathbb{R}^{d_j\times n}$ represents jth view. Here, $b \in [1, d_j]$, $i \in [1, n]$ and $j \in [1, v]$.

Furthermore, in different views, an instance always possesses different labels, thus suppose $y_i^j \in \mathbb{R}^{l_j\times 1}$ is a label vector of ith instance in the jth view and each component of y_i^j indicates the label of x_i^j for the corresponding class. l_j represents that at jth view, instances have l_j classes. If the k_jth component of y_i^j, namely, $y_{ik_j}^j = 1$, it means x_i^j belongs to k_jth class. If $y_{ik_j}^j = -1$, this indicates that x_i^j does not belong to k_jth class. Then $Y^j = \{y_1^j, ..., y_i^j, ..., y_n^j\} \in \mathbb{R}^{l_j\times n}$

represents the label matrix of jth view. Under this definition, $y^{j-k_j} \in \mathbb{R}^{1\times n}$ indicates the label vector for k_jth label and each component of y^{j-k_j} ($y_i^{j-k_j}$) represents whether x_i^j belongs to k_jth label or not. Here, $k_j \in [1, l_j]$.

B. Multi-view multi-label learning with label-specific features

Suppose $w^{j-p_j} \in \mathbb{R}^{d_j\times 1}$ is the p_jth label vector in jth view and it indicates which features are related to some labels. Different from y_i^j and y^{j-k_j}, w^{j-p_j} represents the relevances between p_jth label and d_js features in jth view, and each component is a related relevance. Here, $p_j \in [1, l_j]$. Then the optimization problem of a multi-view multi-label learning with label-specific features for one label is given below. Here, λ_3^j is the tradeoff parameter for jth view.

$$\min_{w^{j-p_j}} \sum_{j=1}^{v} \left(\frac{1}{2} \left|\left| {w^{j-p_j}}^T X^j - y^{j-k_j} \right|\right|_2^2 + \lambda_3^j \left|\left| w^{j-p_j} \right|\right|_1 \right) \quad (1)$$

Then we consider all labels and Eq. (1) is rewritten as below.

$$\min_{W^j} \sum_{j=1}^{v} \left(\frac{1}{2} \left|\left| {W^j}^T X^j - Y^j \right|\right|_2^2 + \lambda_3^j \left|\left| W^j \right|\right|_1 \right) \quad (2)$$

where $W^j = [w^{j-1}, ..., w^{j-p_j}, ..., w^{j-l_j}] \in \mathbb{R}^{d_j\times l_j}$.

C. Multi-view multi-label learning with missing labels

Suppose any missing labels can be reconstructed by the values of other labels by the correlations between them and let $C^j \in \mathbb{R}^{l_j\times l_j}$ represents the label correlation matrix for l_j labels in jth view. In this matrix, p_j^1th row and p_j^2th column element $c_{p_j^1 p_j^2}$ indicates the degree of correlation that label $y^{j-p_j^1}$ is correlated with $y^{j-p_j^2}$ and in most cases, $c_{p_j^1 p_j^2} = c_{p_j^2 p_j^1}$. Note that one class label may be correlated with only a subset of class labels, thus we add the ℓ_1-norm regularizer over C^j to learn sparse label dependencies. The objective function can be written as,

$$\begin{aligned} \min_{W^j, C^j} \sum_{j=1}^{v} \Bigg(& \frac{1}{2} \left|\left| {W^j}^T X^j - C^j Y^j \right|\right|_F^2 + \\ & \frac{\lambda_1^j}{2} \left|\left| C^j Y^j - Y^j \right|\right|_F^2 + \lambda_2^j \left|\left| C^j \right|\right|_1 + \lambda_3^j \left|\left| W^j \right|\right|_1 + \\ & \lambda_4^j \sum_{p_j^1, p_j^2 = 1}^{l_j} c_{p_j^1 p_j^2} \left|\left| w^{j-p_j^1} - w^{j-p_j^2} \right|\right| \Bigg) \\ s.t. \quad & C^j \succeq 0 \end{aligned} \quad (3)$$

and this objective function combines label-specific features.

If label $y^{j-p_j^1}$ and label $y^{j-p_j^2}$ are strongly correlated, namely, $c_{p_j^1 p_j^2}$ or $c_{p_j^2 p_j^1}$ is large, they will have similar label-specific features. Thus, the corresponding model coefficients $w^{j-p_j^1}$ and $w^{j-p_j^2}$ will be quite similar, and the Euclidean distance between them will be small. Otherwise, $w^{j-p_j^1}$ and $w^{j-p_j^2}$ will be dissimilar, and the Euclidean distance between them will be large. After some mathematical operations, the optimization problem can be rewritten as,

$$\begin{aligned} \min_{W^j, C^j} \sum_{j=1}^{v} \Bigg(& \frac{1}{2} \left|\left| {W^j}^T X^j - C^j Y^j \right|\right|_F^2 + \\ & \frac{\lambda_1^j}{2} \left|\left| C^j Y^j - Y^j \right|\right|_F^2 + \lambda_2^j \left|\left| C^j \right|\right|_1 + \lambda_3^j \left|\left| W^j \right|\right|_1 + \\ & \lambda_4^j tr(W^j L^j {W^j}^T) \Bigg) \\ s.t. \quad & C^j \succeq 0 \end{aligned} \quad (4)$$

where $L^j \in \mathbb{R}^{l_j\times l_j}$ is the graph Laplacian matrix of C^j and $tr(A)$ represents the trace of A..

D. Multi-view multi-label learning with incomplete views

In order to restore the incomplete views, we suppose $Z^j \in \mathbb{R}^{d_j\times n}$ is the low-rank assumption matrix of X^j and it can be decomposed into the form $Z^j = U^j V^j$ where $U^j \in \mathbb{R}^{d_j\times r_j}$ and $V^j \in \mathbb{R}^{r_j\times n}$ and $rank(Z^j) = r_j$.

Then the corresponding optimization problem can be rewritten as,

$$\begin{aligned} \min_{U^j, V^j} \frac{1}{2} \sum_{j=1}^{v} \Bigg(& \lambda_5^j \left|\left| U^j V^j - Z^j \right|\right|_F^2 + \\ & \lambda_6^j \left|\left| f(U^j, V^j) - Y^j \right|\right|_2^2 \Bigg) \end{aligned} \quad (5)$$

where $f(U^j, V^j) - Y^j = (P^j U^j V^j + V_0^j) \circ Y^j - I^j - B^j \in \mathbb{R}^{l_j\times n}$ and $V_0^j \in \mathbb{R}^{l_j\times n}$ is a bias matrix, $B^j \in \mathbb{R}^{l_j\times n}$ is a loose variable matrix, $I^j \in \mathbb{R}^{l_j\times n}$ is an all-one matrix, and $P^j \in \mathbb{R}^{l_j\times d_j}$ is a weight matrix and it is different from W^j. P^j is used to classify the restored X^j, i.e., $U^j V^j$ while W^j is a label matrix which is applied to original X^j. Moreover, the operation of $\circ$ is given as follows. $(P^j U^j V^j + V_0^j) \circ Y^j = \begin{pmatrix} (P^j U^j V^j + V_0^j)_1 diag((Y^j)_1) \\ (P^j U^j V^j + V_0^j)_2 diag((Y^j)_2) \\ \vdots \\ (P^j U^j V^j + V_0^j)_{l_j} diag((Y^j)_{l_j}) \end{pmatrix}$ where $diag(A)$ indicates the diagonalization operation and $(A)_i$ represents the ith row of A.

Thus, Eq. (5) can be rewritten as Eq. (6).

$$\begin{aligned} \min_{U^j, V^j} \frac{1}{2} \sum_{j=1}^{v} \Bigg(& \lambda_5^j \left|\left| U^j V^j - Z^j \right|\right|_F^2 + \\ & \lambda_6^j \left|\left| (P^j U^j V^j + V_0^j) \circ Y^j - I^j - B^j \right|\right|_2^2 \Bigg) \end{aligned} \quad (6)$$

E. Multi-view multi-label learning with global and local label correlations

Under j-th view, suppose X^j is divided into g^j groups by some clustering methods, i.e., $X^j = \{X_1^j, ..., X_m^j, ..., X_{g^j}^j\}$

and m-th group of X^j is $X_m^j \in \mathbb{R}^{d_j \times n_m^j}$ where n_m^j is the number of instances in X_m^j and $n_1^j+n_2^j+...+n_m^j+...+n_{g^j}^j = n$. In our method, we adopt active three-way clustering (ATC) [14] for clustering. Details of ATC can be referred to [14]. Then let $F_0^j = P^j X^j \in \mathbb{R}^{l_j \times n}$ represent the classifier output matrix of X^j and $F_m^j = P^j X_m^j \in \mathbb{R}^{l_j \times n_m^j}$ represent the one of X_m^j.

Then, on the base of X^j, X_m^j and their corresponding label matrices, we compute the label correlation matrices. Take X^j as instance, $S_0^j = \{[S_0^j]_{p_j^1 p_j^2}\}$ denotes global label correlation matrix under jth view and $[S_0^j]_{p_j^1 p_j^2} = \frac{y^{j-p_j^1} {y^{j-p_j^2}}^T}{\left|\left|y^{j-p_j^1}\right|\right|\left|\left|y^{j-p_j^2}\right|\right|}$ represents the global label correlation of p_j^2-th label with respect to p_j^2-th label and $y^{j-p_j^1}$ is the p_j^1th row of Y^j. Then we let L_0^j be the Laplacian matrix of S_0^j. Similarly, for X_m^j, $S_m^j = \{[S_m^j]_{p_j^1 p_j^2}\}$ is the corresponding local label correlation matrix and L_m^j is its Laplacian matrix. Dimensions of S_0^j, L_0^j, S_m^j, L_m^j are both $l_j \times l_j$ and $p_j^1, p_j^2 \in [1, l_j]$.

With the above definitions, since we want the classifier outputs can be closer if two labels are more positively correlated, thus the corresponding problem is given as follows.

$$\min_{P^j} \sum_{j=1}^{v} \left(\lambda_7^j tr({F_0^j}^T L_0^j F_0^j) + \lambda_8^j \sum_{m=1}^{g^j} tr({F_m^j}^T L_m^j F_m^j) \right) \tag{7}$$

F. Multi-view multi-label learning with complementary information originating from different views

Let $T^j \in \mathbb{R}^{d_j \times r_j}$ be the basic matrix of X^j and $Q \in \mathbb{R}^{r_j \times n}$ be a latent representation matrix. Here, T^j has a similar function with U^j and Q has a similar function V^j. r_j is the rank of X^j. Then the corresponding optimization problem is given as,

$$\min_{T^j, Q} \lambda_9 \sum_{j=1}^{v} \left|\left|X^j - T^j Q\right|\right|_F^2 + \lambda_{10} \sum_{j \neq t} IND(T^j, T^t) \tag{8}$$

here, $\sum_{j=1}^{v} \left|\left|X^j - T^j Q\right|\right|_F^2$ searches a comprehensive multi-view representation and $\sum_{j \neq t} IND(T^j, T^t)$ is used to measure the independence between different views where $IND(T^j, T^t) = -HSIC(T^j, T^t)$ and $HSIC$ is a Hilbert-Schmidt independence criterion estimator [15].

G. Finally objective function of IMVL-IV

The final objective function of IMVL-IV is given below and all terms are important.

$$\begin{aligned}
\min_{\Omega} & \sum_{j=1}^{v} \Bigg(\frac{1}{2} \left|\left|{W^j}^T X^j - C^j Y^j\right|\right|_F^2 + \\
& \frac{\lambda_1^j}{2} \left|\left|C^j Y^j - Y^j\right|\right|_F^2 + \lambda_2^j \left|\left|C^j\right|\right|_1 + \lambda_3^j \left|\left|W^j\right|\right|_1 + \\
& \lambda_4^j tr(W^j L^j {W^j}^T) \Bigg) + \frac{1}{2} \sum_{j=1}^{v} \Bigg(\lambda_5^j \left|\left|U^j V^j - Z^j\right|\right|_F^2 + \\
& \lambda_6^j \left|\left|(P^j U^j V^j + V_0^j) \circ Y^j - I^j - B^j\right|\right|_2^2 \Bigg) + \\
& \sum_{j=1}^{v} \left(\lambda_7^j tr({F_0^j}^T L_0^j F_0^j) + \lambda_8^j \sum_{m=1}^{g^j} tr({F_m^j}^T L_m^j F_m^j) \right) + \\
& \lambda_9 \sum_{j=1}^{v} \left|\left|X^j - T^j Q\right|\right|_F^2 + \lambda_{10} \sum_{j \neq t} IND(T^j, T^t) \\
s.t. \quad & C^j \succeq 0
\end{aligned} \tag{9}$$

where $\Omega = \{W^j, C^j, U^j, V^j, Z^j, P^j, V_0^j, B^j, T^j, Q\}$.

H. Realization

In order to solve Eq. (9), alternating optimization is adopted here and in each iteration, we update one of the variables in $\{W^j, C^j, U^j, V^j, Z^j, P^j, V_0^j, B^j, T^j, Q\}$ with gradient descent and leave the others fixed. After we get the ∇_A where $A \in \{W^j, C^j, U^j, V^j, Z^j, P^j, V_0^j, B^j, T^j, Q\}$, we can use $A := A - \eta \nabla_A$ to update A where η is the step size. Finally, when we get the optimal matrices, ${W^j}^T U^j V^j$ can be used to compute the classifier outputs for X^j.

I. Computational complexity

In order to solve the Eq. (9) and optimize the IMVL-IV, in each iteration, we update one of the variables in $\{W^j, C^j, U^j, V^j, Z^j, P^j, V_0^j, B^j, T^j, Q\}$ with gradient descent and leave the others fixed. Thus, the computational complexity of IMVL-IV is depended on the ones of the update of these parameters. What's more, since the computational complexity of matrix multiplication is much larger than matrix subtraction, thus the computational complexity of the update for a variable is mainly depended on the computation of ∇_A rather than the computation of $A := A - \eta \nabla_A$. So, we can say that the computational complexity of IMVL-IV is finally depended on the computation of ∇_As. By detailed computation, the total computational complexity of IMVL-IV is $O(Cn^2)$ where $C = \sum_{j=1}^{v} (r_j + 6l_j + d_j)$ is a constant.

III. Experiments

A. Experimental setting

1) Data set: In our experiments, we adopt three kinds of data sets for experiments. First kind is 6 multi-view

Table I
DETAILED INFORMATION OF MULTI-VIEW DATA SETS.

Order	Data set	No. instances	No. labels	No. views
1	Mfeat	2000	10	6
2	Reuters	111740	6	5
3	Corel	1000	10	4
4	VOC	9963	20	2
5	MIR	23691	38	2
6	3Source	169	6	3

Table II
DETAILED INFORMATION OF MULTI-LABEL DATA SETS.

Order	Data set	No. instances	No. features	No. labels	label/ instance
7	Arts	5000	462	26	1.64
8	Business	5000	438	30	1.59
9	Computers	5000	681	33	1.51
10	Education	5000	550	33	1.46
11	Entertainment	5000	640	21	1.42
12	Health	5000	612	32	1.66

data sets, namely, Mfeat[1], Reuters[2], Corel[3], Pascal VOC 2007 (VOC)[4], MIR-Flickr (MIR)[5], 3Source[6] (see Table I). The second kind is 6 multi-label data sets which are also used in [2], [10], [11]. Table II shows information of them and $label/instance$ represents the average number of labels possessed by each instance. The third kind is a multi-view multi-label data set, namely, NUS-WIDE (let its order be 13). NUS-WIDE consists of 810 images (instances) and 81 labels. Each instance is related with some of the 81 labels and has 6 views including color histogram, color correlogram, edge direction histogram, wavelet texture, block-wise color moments extracted over 5×5 fixed grid partitions, and bag of words based on SIFT descriptions [16], [17].

Furthermore, since our proposed IMVL-IV can process incomplete views and labels, thus for each used multi-view data set, suppose it has n instances and we randomly select $|\Omega|$ instances from these n instances and randomly remove one view from each instance. Then in our experiments, $|\Omega|/n$ falls in $[0.1, 1]$. Indeed, $|\Omega|/n$ can be regarded the rate of missing views. Furthermore, for each available multi-label data set, we randomly sample ρ percent of the elements in the label matrix as missing, and the rest as observed. The ρ falls in $[10, 90]$. For NUS-WIDE, both $|\Omega|/n$ and ρ are used, but the range should include value 0 so that for NUS-WIDE, it should cover three cases, namely, incomplete views, incomplete labels, and incomplete views and labels. For convenience, we use $order - x$ to represents these different missing cases. The meanings of x are given in Table III. For example, $1 - a$ represents Mfeat with 10% views missing. In order to describe the cases for NUS-WIDE, we can use the formation $order - x - x$. For example, $36 - a - k$ represents NUS-WIDE with 10% views and 20% labels missing.

Table III
DETAILED MEANING OF X.

x	meaning	x	meaning
a	missing 10% views	j	missing 10% labels
b	missing 20% views	k	missing 20% labels
c	missing 30% views	l	missing 30% labels
d	missing 40% views	m	missing 40% labels
e	missing 50% views	n	missing 50% labels
f	missing 60% views	o	missing 60% labels
g	missing 70% views	p	missing 70% labels
h	missing 80% views	q	missing 80% labels
i	missing 90% views	r	missing 90% labels

Table IV
CHARACTERISTICS OF COMPARED METHODS.

	1	2	3	4	5	6	7
LMSC [18]	✓						
MLDL [19]	✓						
MLCHE [20]		✓					
MMKR [3]			✓				
sM2CP [4]			✓				
SSDR-MML [5]			✓	✓			
MVMC [6]			✓	✓			
GLOCAL [2]		✓		✓	✓		
MLMF [8]		✓		✓	✓		
LF-LPLC [21]		✓			✓	✓	
LIFT [9]		✓				✓	
MLFC [10]		✓				✓	
LSML [11]		✓		✓	✓	✓	
MVML [12]			✓				✓
NMF-SSMM [13]			✓				✓

2) Compared method: Since three kinds of data sets are adopted in our experiments, thus for the fair comparison, we also adopt three kinds of learning methods for comparisons. They are multi-view learning methods, multi-label ones, and multi-view multi-label ones. Table IV shows the characteristics of the used compared methods. In this table, $1 \sim 7$ indicate 7 characteristics, namely, 1-multi-view, 2-multi-label, 3-multi-view multi-label, 4-incomplete data, 5-label correlations, 6-label-specific features, 7-complementary information.

3) Parameter setting and how to get the experimental results: For the compared methods, the parameter settings of them can be found in the respective references. Then for the proposed IMVL-IV, the settings are given below.

For IMVL-IV, we divide the data set into several groups with ATC and the setting of ATC can refer to [14]. $W^j, C^j, U^j, V^j, Z^j, P^j, V_0^j, B^j, T^j, Q$ can be initialized according to the X^j and their corresponding groups. For λ_i^js, they are selected from $\{2^{-5}, 2^{-4}, ..., 2^0\}$. Furthermore, the maximum number of iterations is set to be 1000.

In order to get the optimal results and according to the compared methods' demands, for each data set, we randomly

[1] http://archive.ics.uci.edu/ml/datasets/Multiple+Features
[2] http://archive.ics.uci.edu/ml/datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection
[3] http://archive.ics.uci.edu/ml/datasets/Corel+Image+Features
[4] http://host.robots.ox.ac.uk/pascal/VOC/
[5] http://press.liacs.nl/mirflickr/
[6] http://mlg.ucd.ie/datasets/3sources.html

select $\{10\%, 20\%, ..., 60\%\}$ for training and the rest for test. Then we repeat the experiments with each parameter combination for ten times and get the average results and the corresponding standard deviation. The best parameters are the ones whose average precision is the best. Then, the other performance indexes including the AUC (i.e., the area under the receiver operating characteristic (ROC) curve), running time, convergence, etc. are given with the optimal parameters. Here, we should notice that for each data set, different methods should process same data.

4) Experimental environment: The experimental environment is given below. All computations are performed on a node of compute cluster with 32 CPUs (Intel Core Due 3.0GHz) running RedHat Linux Enterprise 5 with 48GB main memory. The coding environment is python 3.0.

B. Experimental results

In order to validate the effectiveness of the IMVL-IV, we conduct experiments from multiple aspects including AUC, precision, running time, hamming loss, ranking loss, macro-F1, or micro-F1. **For the limitation of the length for this manuscript, we only show the results on Mfeat, Arts, and NUS-WIDE on AUC, precision, running time. But for other data sets and evaluation indexes, the results are similar.** Moreover, for NUS-WIDE, only results about AUC are given. Results can be found in Fig. (4) and Fig. (5). According to the results, we can see that our IMVL-IV performs best in terms of AUC and precision. Although it should cost more time, but the better classification performances offset such a disadvantage. Furthermore, missing more views or labels brings a worse performance, but our IMVL-IV still performs best in average. Combining the results about the standard deviation, we can find the performances about AUC, precision, and running time are statable in generally.

IV. Conclusions

Multi-view multi-label data sets with incomplete data is widely used in real-world application and most traditional multi-view multi-label learning methods cannot process that. This manuscript develops an improved multi-view multi-label learning with incomplete views and labels (IMVL-IV) for this issue. In framework of IMVL-IV, five important factors which enhance the processing ability to process multi-view multi-label data sets with incomplete data are introduced, i.e., label-specific features, label correlation matrix, low-rank assumption matrix, global and local label correlations, and consensus multi-view representation. Different from traditional learning methods, this is the first attempt to design a multi-view multi-label learning method with incomplete views and labels by the learning of these five factors. Experimental results validate that IMVL-IV achieves a better performance and it is superior to the classical multi-view learning methods and multi-label learning methods.

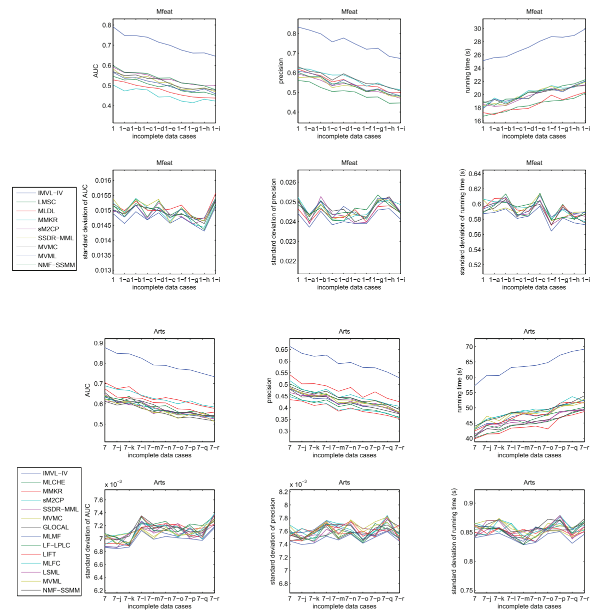

Figure 4. Experimental results on data sets Mfeat and Arts.

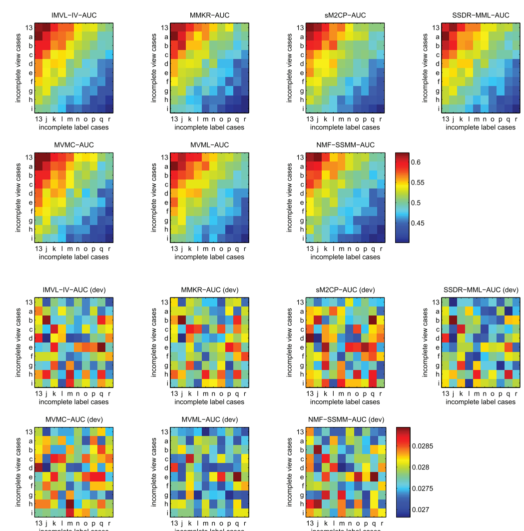

Figure 5. Experimental results about AUC on data set NUS-WIDE.

Acknowledgment

This work is supported by National Natural Science Foundation of China (CN) under grant number 61602296, Natural Science Foundation of Shanghai under grant number 16ZR1414500, Project funded by China Postdoctoral Science Foundation under grant number 2019M651576. Furthermore, this work is also sponsored by 'Chenguang Program' supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission under grant number 18CG54. The authors would like to

thank their supports.

References

[1] X.F. Zhu, X.L. Li, and S.C. Zhang, *Block-Row Sparse Multiview Multilabel Learning for Image Classification*. IEEE Transactions on Cybernetics, 46(2):450-461, 2016.

[2] Y. Zhu, J.T. Kwok, and Z.H. Zhou, *Multi-Label Learning with Global and Local Label Correlation*. IEEE Transactions on Knowledge and Data Engineering, 99:1-24, 2017.

[3] X.F. Zhu, Q. Xie, Y.H. Zhu, X.Y. Liu, and S.C. Zhang, *Multi-view multi-sparsity kernel reconstruction for multi-class image classification*. Neurocomputing, 169:43-49, 2015.

[4] K. Maeday, S. Takahashi, T. Ogaway, and M. Haseyama, *Multi-feature fusion based on supervised multi-view multi-label canonical correlation projection*. 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2019, doi: 10.1109/ICASSP.2019.8682060.

[5] 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, 27(3):594-607, 2015.

[6] Y. Luo, T.L. Liu, D.C. Tao, and C. Xu, *Multiview matrix completion for multilabel image classification*. IEEE Transactions on Image Processing, 24(8): 2355-2368, 2015.

[7] B. Zhang, Y. Wang, and F. Chen, *Multilabel image classification via high-order label correlation driven active learning*. IEEE Transactions on Image Processing, 23(3):1430-1441, 2014.

[8] Z.F. He, M. Yang, Y. Gao, H.D. Liu, and Y.L. Yin, *Joint multi-label classification and label correlations with missing labels and feature selection*. Knowledge-Based Systems, 163:145-158, 2019.

[9] M.L. Zhang and L. Wu, *Lift: Multi-Label Learning with Label-Specific Features*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(1): 107-120, 2015.

[10] J. Zhang, C.D. Li, D.L. Cao, Y.J. Lin, S.Z. Su, L. Dai, and S.Z. Li, *Multi-label learning with label-specific features by resolving label correlations*. Knowledge-Based Systems, 159:148-157, 2018.

[11] J. Huang, F. Qin, X. Zheng, Z.K. Cheng, Z.X. Yuan, W.G. Zhang, and Q.M. Huang, *Improving multi-label classification with missing labels by learning label-specific features*. Information Sciences, 492:124-146, 2019.

[12] F.H. Zou, Y. Liu, H. Wang, J.K. Song, J. Shao, K. Zhou, and S. Zheng, *Multi-view multi-label learning for image annotation*. Multimedia Tools and Applications, 75(20):12627-12644, 2016.

[13] G.X. Wang, C.Q. Zhang, P.F. Zhu, and Q.H. Hu, *Semi-supervised multi-view multi-label classification based on nonnegative matrix factorization*. International conference on artificial neural networks, artificial neural networks and machine learning, pp. 340-348, 2017.

[14] H. Yu, X.C. Wang, G.Y. Wang, X.H. Zeng, *An active three-way clustering method via low-rank matrices for multi-view data*. Information Sciences, 2018, doi: 10.1016/j.ins.2018.03.009.

[15] C.Q. Zhang, Z.W. Yu, Q.H. Hu, P.F. Zhu, X.W. Liu, and X.B. Wang, *Latent Semantic Aware Multi-View Multi-Label Classification*. Thirty-Second AAAI Conference on Artificial Intelligence, pp. 4414-4421, 2018.

[16] T.S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, *Nus-wide: a real-world web image database from national university of Singapore*. in: Proceedings of the ACM International Conference on Image and Video Retrieval, pp 48, 2009.

[17] Z.Y He, C. Chen, J.J. Bu, P. Li, and D. Cai, *Multi-view based multi-label propagation for image annotation*. Neurocomputing, 168:853-860, 2015.

[18] 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, 2018, doi: 10.1109/TPAMI.2018.2877660.

[19] F. Wu, X.Y. Jing, X.G. You, D. Yue, R.M. Hu, and J.Y Yang, *Multi-view low-rank dictionary learning for image classification*. Pattern Recognition, 50:143-154, 2016.

[20] V. Kumar, A.K. Pujari, V. Padmanabhan, S.K. Sahu, and V.R. Kagita, *Multi-label classification using hierarchical embedding*. Expert Systems with Applications, 91:263-269, 2018.

[21] W. Weng, Y.J. Lin, S.X Wu, Y.W. Li, and Y. Kang, *Multi-label learning based on label-specific features and local pairwise label correlation*. Neurocomputing, 273:385-394, 2018.