Label recovery and label correlation co-learning for multi-view multi-label classification with incomplete labels

Zhi-Fen He1,2 · Chun-Hua Zhang1,2 · Bin Liu1,2 · Bo Li1,2

Accepted: 27 June 2022 / Published online: 9 August 2022 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Multi-view multi-label learning (MVML) is an important paradigm in machine learning, where each instance is represented by several heterogeneous views and associated with a set of class labels. However, label incompleteness and the ignorance of both the relationships among views and the correlations among labels will cause performance degradation in MVML algorithms. Accordingly, a novel method, label recovery and label correlation co-learning for Multi-View Multi-Label classification with incomplete Labels (MV2ML), is proposed in this paper. First, a label correlation-guided binary classifier kernel-based is constructed for each label. Then, we adopt the multi-kernel fusion method to effectively fuse the multi-view data by utilizing the individual and complementary information among multiple views and distinguishing the contribution difference of each view. Finally, we propose a collaborative learning strategy that considers the exploitation of asymmetric label correlations, the fusion of multi-view data, the recovery of incomplete label matrix and the construction of the classification model simultaneously. In such a way, the recovery of incomplete label matrix and the learning of label correlations interact and boost each other to guide the training of classifiers. Extensive experimental results demonstrate that MV2ML achieves highly competitive classification performance against state-of-the-art approaches on various real-world multi-view multi-label datasets in terms of six evaluation criteria.

Keywords Multi-view multi-label classification · Incomplete labels · Label recovery · Label correlation · Multi-kernel fusion

1 Introduction
In many real-world applications, one instance may often have multiple heterogeneous feature representations and be associated with multiple class labels. For example, an image can be represented by heterogeneous features, such as color information (RGB), shape cues (SIFT) and global structure (GIST); at the same time, it can be annotated as sea, sunset and beach. A video usually contains diverse feature representations, such as audio, image and text, and can be labeled as Olympic Games, sports and stadium simultaneously. A piece of music can be characterized by rhythm and timbre attributes, and can be annotated with multiple labels, such as classical and sad. Multi-view multi-label learning (MVML) is a paradigm developed to deal with this type of data; it has received widespread attention in machine learning and has been successfully applied in numerous real-world application domains, such as image classification [1–4], video classification [5–7], emotion classification [8] and bioinformatics [9].

To solve an MVML problem, an intuitive strategy is to transform it into a multi-label learning (MLL) problem. On the one hand, by directly concatenating the features of each view into a single feature vector, an MVML problem
Label recovery and label correlation co-learning for multi-view multi-label classification with incomplete labels

The second strategy is to directly deal with multi-view multi-label data by constructing an MVML method. Although many MVML approaches have been proposed in recent years and have obtained excellent performance in different real-world applications, they share some limits. One limitation is that most methods focus on learning a shared latent subspace among different views to fuse the multi-view feature representations but do not explicitly consider both the individual information of each view and the complementary information among multiple views, which could degrade the performance of the MVML method. Another limitation is that most such methods are designed under the full-label assumption and fail to effectively handle multi-view multi-label data with incomplete labels. However, in real-world applications, it is rather expensive and difficult to obtain completely labeled data, while partially labeled data are easy to collect. For instance, in image annotation, markers with different backgrounds may only annotate an image with class labels that are familiar, especially when the label space is too large. In addition, in MVML, it is crucial to effectively discover and exploit label correlations to improve the performance of the multi-view multi-label classifier. Nevertheless, it is difficult to accurately learn or estimate the correlations among labels from the label matrix of the training instances, especially when part of the ground truth of labels is missing. Therefore, label incompleteness could cause performance degradation. To alleviate the performance degeneration caused by incomplete labels in MVML, some methods have been proposed and have achieved improved performance. However, determining how to learn high-order and asymmetric label correlations from training data with incomplete labels and how to exploit it to recover the incomplete label matrix and guide the construction of a multi-view multi-label classification model is still an important problem for MVML with incomplete labels.

To address the aforementioned problems, a novel method named MV2ML, i.e., label recovery and label correlation co-learning for Multi-View Multi-Label classification with incoMplete Labels, is proposed in this paper. Specifically, we construct a kernel-based binary classifier for each label by considering high-order and asymmetric label correlations among labels. Second, to utilize the individual and complementary information from different views and distinguish the contribution difference of each view, we employ the multi-kernel fusion method to effectively fuse multi-view features. Third, a unified learning framework for multi-view multi-label classification with incomplete labels is developed by jointly considering the utilization of both the individual and complementary information among views and the correlations among labels, the learning of different contribution weights of views, the recovery of the incomplete label matrix and the training of a multi-view multi-label classifier. Finally, we develop an alternative minimization method to iteratively optimize our proposed model. Experiments conducted on five multi-view multi-label benchmark datasets in terms of six evaluation criteria demonstrate that MV2ML achieves superior performance compared with state-of-the-art multi-label and multi-view multi-label classification algorithms.

In summary, the main contributions of this paper include the following:

1. Different from most existing methods in MVML with incomplete labels, MV2ML can explicitly discover and exploit label correlations, mine the individual and complementary information among different views, learn the contribution weight of each view and recover the label matrix with missing labels in a unified learning model.
2. The label correlations learned from the incomplete training data are employed to help recover the incomplete label matrix, while the recovered label matrix in turn affects the learning of the label correlations.
3. The label correlations and the different contribution coefficients of views are integrated into the classification model. Moreover, this model can be directly applied to predict the class labels of unseen instances.

The remainder of this paper is organized as follows: Section 2 briefly reviews some previous work on MLL and MVML. Section 3 presents our proposed method...
Most existing studies on MLL assume that all the labels of each instance in the training data are provided. However, in some real-world applications, it is difficult for such an assumption to hold, especially when the label space is very large. In addition, it is difficult to learn or obtain the accurate label correlations from the multi-label data with missing labels, and the performance of the MLL algorithm will also be affected by label incompleteness. In recent years, some MLL approaches with incomplete labels have been put forward [19–25]. Yu et al. [19] dealt with multi-label data with a large number of labels and missing labels by building a generic empirical risk minimization framework, but the label correlations were not exploited and obtained directly. Wang et al. [20] first modeled label correlations by utilizing a Bayesian network and then used it to guide the construction of a multi-label classifier. In addition, the proposed method would be further extended to handle missing-label data. Huang et al. [21] proposed an improved multi-label classification method with missing labels by learning label-specific features (LSML). LSML first learns high-order asymmetric label correlations and then uses them to guide the process of learning label-specific features to build a multi-label classifier. Bi et al. [22] developed a probabilistic model to effectively deal with missing-label data and automatically exploit pairwise label correlations represented by a covariance matrix. Zhu et al. [23] proposed an MLL approach based on global and local label correlations (GLOCAL). GLOCAL exploits global and local label correlations and can handle both full-label and missing-label data. Cheng et al [24] aimed to address the MLL task with label correlations and missing labels. However, before learning the multi-label classifier, the positive and negative label correlations are obtained from the missing-label data and used to help complete the missing-label data. He et al. [25] proposed a joint multi-label classification approach with label correlations, missing labels and feature selection (MLMF). MLMF can learn a multi-label classifier, explore and exploit high-order symmetric label correlations and select the most informative features in a unified learning model in a missing-label setting, but it drops the missing-label data in the learning process.

2.2 Multi-view multi-label learning

Previous MLL studies mainly focused on the assumption of a single-view setting and ignored the prevalent multi-view based multi-label scenario. However, in many real-world applications, multi-label data are often represented by multiple different feature views. For example, a news report can be described from multiple views, such as a written demonstration, picture demonstration and video
presentation, and can be annotated with pneumonia and COVID-19 simultaneously. To directly and effectively deal with data with multiple views and associated with multiple class labels, i.e., multi-view multi-label data, the MVML framework has been developed and numerous MVML approaches have been proposed in recent years, e.g., [2–9, 17, 26–38].

Luo et al. [2] introduced semi-supervised multi-view multi-label image classification based on matrix completion (MVMC) and proposed two variants of MVMC, i.e., MVMC-LS and MVMC-LP. Liu et al. [3] also presented a low-rank multi-view learning model for matrix completion-based multi-label image classification (lrMVL), which obtains a shared low-dimensional representation and then learns the combination weights to utilize complementary information among different views. Nevertheless, lrMVL obtains a low-rank common representation of multiple views as a preprocessing step before performing image classification. Zhang et al. [26] proposed a matrix factorization-based MVML framework (LAS-MML), which makes full use of complementary information among multiple views and latent semantic patterns to better obtain a consensus multi-view representation. Wang et al. [27] proposed a semi-supervised multi-view multi-label classification method based on nonnegative matrix factorization (NMF-SSMM), which fully explores the complementarity and consistency among views. However, these methods do not take into consideration the correlations among labels, which may lead to sub-optimal performance.

It is widely accepted that exploiting label correlations can further boost the performance of multi-view multi-label classifier [5, 8, 9, 28–30]. He et al. [28] aimed to improve the multi-view multi-label image annotation performance by exploring the consistency information among different views and label correlations across related labels. Zhang et al. [29] tried to handle the multi-label image classification problem by utilizing the metric learning approach to capture a joint representation of different views and the dependency among labels. Tan et al. [9] attempted to explicitly explore the individual and common information among multiple views and learn the correlations among multiple labels in a unified learning model, and an MVML framework based on individuality and commonality (ICM2L) was constructed. Zhao et al. [8] constructed a feedforward neural network for MVML, which can learn the consistency and diversity among multiple views and consider the label correlations and the contribution differences of each view in a joint learning model.

After further analysis, it can be found that the above-mentioned MVML methods have no ability to process data with incomplete labels. In many real-world applications, it is rather expensive or difficult to obtain data with complete labels, while partially labeled data are more easily available. Therefore, accurately learn the label correlations and exploiting them to guide the construction of classifiers under a missing-label scenario is a major challenge in MVML. Zhu et al. [31–33] successively proposed three MVML methods based on global and local label correlations (IMVL, GLMVML and GLMVML-IVL). These methods handle data with heterogeneous views and missing labels and make use of global and local label correlations and complementarity between views. Tan et al. [34] developed a matrix completion-based multi-view weak-label learning model (McWL), which can jointly fuse multiple views and learn a matrix completion-based classifier in a unified learning framework. Zhao et al. [35] designed a two-step MVML algorithm with missing labels (TM3L), which first aims at a data representation of the common low-dimensional subspace of all views to solve the multi-view learning problem and then uses the label matrix completion method to solve the multi-label learning problem with missing labels. Tan et al. [36] proposed an incomplete multi-view weak-label learning method (iMVML), which simultaneously learns a shared subspace, local label correlations and a classifier from incomplete multi-view weak-label data. However, it is rather difficult to obtain accurate label correlations, especially when the label data are incompletely given. Li et al. [37] built a concise and effective model for MVML with non-aligned incomplete views and missing labels (NAIM3L). NAIM3L is a model with only one hyperparameter that simultaneously addresses three challenges: non-aligned views, incomplete views and missing labels. In this paper, we propose a novel multi-view multi-label classification method with incomplete labels, named MV2ML, to jointly optimize a multi-view multi-label classifier, learn a label correlation matrix from missing-label data, fuse data with multiple views and recover data with incomplete labels in a unified learning framework.

3 The proposed method

Let $D = \{(x_i, Y_i)\}_{i=1}^n$ be a multi-view multi-label dataset with $n$ instances and $V$ views, where $x_i = (x_i^1, x_i^2, ..., x_i^V) \in \mathbb{R}^d$ denotes the feature space of the $i$-th instance with $V$ views, $x_i^v \in \mathbb{R}^{d_v}$ represents the feature space of the $i$-th instance under the $v$-th view, $d = \sum_{v=1}^{V} d_v$ indicates the dimensionality of the feature space and $d_v$ denotes the dimensionality of features of the $v$-th view. Suppose that $x^v = (x_1^v, x_2^v, ..., x_n^v)^T \in \mathbb{R}^{n \times d_v}$ represents the feature matrix of the $v$-th view and $Y = (Y_1, Y_2, ..., Y_n)^T \in \{+1, 0, -1\}^{n \times L}$ is a label matrix with incomplete labels, where $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{iL})^T \in \{+1, 0, -1\}$ is the corresponding missing-label vector of $x_i$ and $L$ is the...
number of possible class labels, $Y_{il} = +1$ indicates that the $l$-th label is a relevant label for $x_i$, $Y_{il} = −1$ means that the $l$-th label is an irrelevant label for $x_i$ and $Y_{il} = 0$ represents that the $l$-th label for $x_i$ is missing.

### 3.1 Problem formulation

Inspired by the idea of the Representer Theorem [39], we define the prediction function of the $l$-th label as follows:

$$f_l(x) = \sum_{i=1}^{n} \alpha_{il} k(x, x_i) + b_l$$

where $k(\cdot, \cdot)$ is the kernel function and $\{\alpha_{il}\}_{i=1}^{n}$ and $b_l \in R$ denote the coefficients and bias of the $l$-th classifier, respectively.

Due to the large number of labels and relatively complex semantics in MVML, it is difficult to obtain numerous completely labeled instances, which leads to the label information of instances being partially or even completely missing. To avoid the influence of missing labels and improve the prediction performance and generalization ability of the MVML algorithm, it is crucial to effectively discover and exploit the correlations among labels in missing-label scenarios. In this paper, we assume that the actual prediction value is represented by the combination of missing-label scenarios. In this paper, we assume that the actual prediction value is represented by the combination of the prediction values of all labels [25]. Thus, we can obtain the final prediction function of the $l$-th label as follows:

$$g_l(x) = \sum_{i=1}^{n} \alpha_{il} k(x, x_i) + b_l + \sum_{q \neq l}^{n} s_{lq} \left( \sum_{i=1}^{n} \alpha_{il} k(x, x_i) + b_q \right)$$

$$= \left( \sum_{i=1}^{n} \alpha_{il} k(x, x_i) + b_l \right) s_{l}$$

where $\alpha_l = (\alpha_{il1}, \alpha_{il2}, \ldots, \alpha_{ilL})$, $b = (b_{l1}, b_{l2}, \ldots, b_{lL})$, $s_l = (s_{l1}, s_{l(l+1)}, \ldots, s_{lL})^T$ stands for the label correlation vector, and $s_{lq}$ denotes the relationship between the $l$-th label and the $q$-th label, which can be positive or negative. In addition, it is worth noting that $s_{lq}$ is not necessarily equal to $s_{ql}$.

In MVML, each instance is represented by several heterogeneous feature views and the feature dimensions of each view are usually different, so it is difficult to add multiple feature representations directly. A simple and intuitive method is to concatenate the features of each view into a long feature vector. However, such concatenation ignores the useful individuality of different views and may also lead to the “curse of dimensionality” or suboptimality problem. To make full use of the complementary and individual information from different views and distinguish the contribution differences of each view in the multi-view multi-label classification model, we employ a multi-kernel fusion method to effectively fuse different feature representations. Therefore, $k(x, x_i)$ in (2) is given as follows:

$$k(x, x_i) = \sum_{v=1}^{V} \beta^v k^v(x^v, x_{i}^v)$$

where $\beta^v$ and $k^v(\cdot, \cdot)$ are the contribution weight and the kernel function corresponding to the $v$-th view, respectively.

Finally, the objective optimization problem of MVML with incomplete labels by jointly learning label completion matrix and label correlations is formulated as follows:

$$\min_{A, b, \bar{Y}} \sum_{l=1}^{L} \sum_{i=1}^{n} \ell_l(g_l(x_i), \bar{Y}_{il}) + \lambda_A \Psi(A) + \lambda_S \Psi(S) + \lambda_\beta \Psi(\beta)$$

s.t. $s_{il} = 1, l = 1, \ldots, L$

$$\bar{P}_{\Omega_1}(\bar{Y}) = P_{\Omega_2}(\bar{Y})$$

$$\sum_{v=1}^{V} \beta^v = 1$$

$$\beta^v \geq 0, v = 1, 2, \ldots, V$$

where $Y = (Y_1, Y_2, \ldots, Y_n)^T$ is the incomplete label matrix, $Y_l = (Y_{l1}, Y_{l2}, \ldots, Y_{lL})^T \in \{+1, 0, -1\}^L$ is the label vector of $x_j$, $\bar{Y}$ is the label recovery matrix and $S = (s_{11}, s_{12}, \ldots, s_{LL}) \in R^{L \times L}$ denotes the label correlation matrix. $A = (\alpha_{1}, \alpha_{2}, \ldots, \alpha_{L})^T \in R^{n \times L}$ and $\beta = (\beta^1, \beta^2, \ldots, \beta^V)^T \in R^V$. $\Psi(A)$, $\Psi(S)$ and $\Psi(\beta)$ are the regularization terms for controlling the complexity of the model, and $\lambda_A$, $\lambda_S$ and $\lambda_\beta$ are the regularization parameters. In the experiments, $S$ is initialized to an $L$-order identity matrix and each value of $\beta$ is initially set to $1/V$. In addition, $S$ and $\beta$ are the trainable parameters in the optimization process.

(1) For the loss function $\ell_l(\cdot, \cdot)$ in (4), we choose the least square loss function in this paper. Thus, the first term of (4) can be written as:

$$\sum_{i=1}^{L} \sum_{l=1}^{n} \ell_l(g_l(x_i), \bar{Y}_{il}) = \sum_{i=1}^{L} \sum_{l=1}^{n} (g_l(x_i) - \bar{Y}_{il})^2$$

$$= \sum_{i=1}^{L} \sum_{l=1}^{n} (\sum_{j=1}^{n} \alpha_{il} k(x, x_j) + b_l s_{li} - \bar{Y}_{il})^2$$

$$= \|KAS + 1bS - \bar{Y}\|_F^2$$

where the kernel matrix

$$K = \sum_{v=1}^{V} \beta^v K^v,$$

$K^v \in R^{n \times n}$ denotes the kernel matrix of the $v$-th view, where each elements are $K_{ij}^v = k^v(x_i^v, x_j^v)$, $I$ is an $n$-dimensional column vector with all elements 1.
and $\| \cdot \|_F$ is the Frobenius norm. The complementary information among different views is captured by $K$ and the individual information of each view is captured by $(K^i)^T V_{i=1}^V$.

(2) For $\Psi(A), \Psi(S)$ and $\Psi(\beta)$ in (4), let $\Psi(A) = \text{tr}(A^T K A), \Psi(S) = \|S\|_F^2$ and $\Psi(\beta) = \|\beta\|_2^2$.

(3) For $P_{\Omega}(Y)$ and $P_{\Omega}(\bar{Y})$ in (4), we define

$$P_{\Omega}(Y_{ij}) = \begin{cases} Y_{ij}, & \text{if } (i, j) \in \Omega \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

and

$$P_{\Omega}(\bar{Y}_{ij}) = \begin{cases} \bar{Y}_{ij}, & \text{if } (i, j) \in \Omega \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $\Omega = \{(i, j)|Y_{ij} \neq 0, i = 1, 2, ..., n, j = 1, 2, ..., L\}$. Finally, (4) can be further rewritten as follows:

$$\min_{A,b,S,\beta,\bar{Y}} \|KAS + bS - \bar{Y}\|_F^2 + \lambda_A \text{tr}(A^T K A) + \lambda_S \|S\|_F^2 + \lambda_\beta \|\beta\|_2^2$$

s.t. $s_{ll} = 1, l = 1, ..., L$

$$P_{\Omega}(Y) = P_{\Omega}(\bar{Y})$$

$$\sum_{v=1}^V \beta^v = 1$$

$$\beta^v \geq 0, v = 1, ..., V$$

(9)

### 3.2 Optimization solution

The objective function in (9) is not jointly convex, so it is difficult to solve it directly. To solve the optimization problem (9), we adopt an alternating minimization strategy by alternatively optimizing one variable while fixing the other variables.

**Update $A$ and $b$ with fixed $S, \beta$ and $\bar{Y}$**

When $S, \beta$ and $\bar{Y}$ are fixed, the last two terms and the constraint term of (9) are constant and can be ignored accordingly. Therefore, the optimization problem in (9) reduces to

$$\min_{A,b} \|KAS + bS - \bar{Y}\|_F^2 + \lambda_A \text{tr}(A^T K A)$$

(10)

Let $\hat{A} = \begin{pmatrix} A \\ b \end{pmatrix} \in R^{(n+1) \times L}$ and $\hat{K} = (K, 1) \in R^{n \times (n+1)}$, (10) can be simplified as:

$$\min_{\hat{A}} \|\hat{K} \hat{A} S - \hat{Y}\|_F^2 + \lambda_A \text{tr}(\hat{A}^T \hat{K}^T \hat{A})$$

(11)

where $\hat{K} = \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix} \in R^{(n+1) \times (n+1)}$.

Therefore, $\hat{A}$ can be calculated by solving the following Sylvester equation:

$$\hat{A} S^T + \lambda_A (\hat{K}^T \hat{K})^{-1} \hat{K} \hat{A} = (\hat{K}^T \hat{K})^{-1} \hat{K}^T \hat{Y} S^T$$

(12)

Additionally, to avoid the irreversibility of the matrix $\hat{K}^T \hat{K}$, let it equal $\hat{K}^T \hat{K} + \xi I_{n+1}$, where $\xi$ is a very small positive number and $I_{n+1}$ is an $(n + 1)-$order identity matrix.

**Update $S$ with fixed $A, b, \beta$ and $\bar{Y}$**

With fixed $A, b, \beta$ and $\bar{Y}, S$ can be obtained by solving the following objective function:

$$\min_{S} \|KAS + bS - \bar{Y}\|_F^2 + \lambda_S \|S\|_F^2$$

s.t. $s_{ll} = 1, l = 1, ..., L$

(13)

The derivation of (13) w.r.t. $S$ is as follows:

$$S = ((KA + 1b)^T (KA + 1b) + \lambda_S I_L)^{-1}(KA + 1b)^T \bar{Y}$$

s.t. $s_{ll} = 1, l = 1, ..., L$

(14)

where $I_L$ is a $L-$order identity matrix.

**Update $\beta$ with fixed $A, b, S$ and $\bar{Y}$**

With $A, b, S$ and $\bar{Y}$ unchanged, we update $\beta$ by converting this problem into the following optimization problem:

$$\min_{\beta} \|KAS + bS - \bar{Y}\|_F^2 + \lambda_A \text{tr}(A^T K A) + \lambda_\beta \|\beta\|_2^2$$

s.t. $\sum_{v=1}^V \beta^v = 1$

$$\beta^v \geq 0, v = 1, ..., V$$

(15)

where $K = \sum_{v=1}^V \beta^v K^v$.

Equation 15 can be further expressed as:

$$\min_{\beta} \beta^T (M + \lambda_\beta I_V) \beta + \beta^T u$$

s.t. $\sum_{v=1}^V \beta^v = 1$

$$\beta^v \geq 0, v = 1, ..., V$$

(16)

where $M \in R^{V \times V}$ with each element $M_{ij} = \text{tr}(S^T A^T K^i K^j A S)$, $I_V$ is a $V-$order identity matrix, $u \in R^V$ with each element $u_v = \text{tr}(2S^T A^T K^v (bS - \bar{Y})) + \lambda_A A^T K^v A$ and $\text{tr}()$ represents the trace of matrix.

After calculating $\beta$, $K$ can be updated according to (6).

**Update $\bar{Y}$ with fixed $A, b, S$ and $\beta$**

After obtaining $A, b, S$ and $\beta$, the label prediction matrix $\hat{Y}$ can be calculated as $\hat{Y} = KAS + bS$ and the missing-label values in $\hat{Y}$ can be recovered according to the following formulation:

$$\hat{Y}_{ij} = \begin{cases} -1, & \text{if } \hat{Y}_{ij} \leq -1 \text{ and } (i, j) \in \hat{\Omega} \\ \hat{Y}_{ij}, & \text{otherwise} \end{cases}$$

(17)

$$\hat{Y}_{ij} = \begin{cases} +1, & \text{if } \hat{Y}_{ij} \geq +1 \text{ and } (i, j) \in \hat{\Omega} \\ \hat{Y}_{ij}, & \text{otherwise} \end{cases}$$

where $\hat{\Omega} = \{(i, j)|Y_{ij} = 0, i = 1, 2, ..., n, j = 1, 2, ..., L\}$.

Finally, the main optimization procedure of the proposed MV2ML approach is summarized in Algorithm 1.
3.3 Complexity analysis

There are four variables, A, b, S and β, in our proposed MV2ML approach. The time complexity of solving A and b in (12) is $O(n^3 + n^2L + nL^2 + L^3)$. The time complexities of updating S in (14) and β in (16) are $O(n^2L + nL^2 + L^3)$ and $O((n^2L + nL^2)V^2)$, respectively. The time complexity of matrix completion for $\bar{Y}$ is $O(n^2Q)$. Since $n \gg L$ and $n \gg V$, the total time complexity of MV2ML is $O(Tn^2(n + LV^2))$, where $T$ denotes the maximum number of iterations. In practice, $T$ does not exceed 50. Thus, the actual time cost of the above operations can be further reduced.

4 Experiments

4.1 Experimental settings

4.1.1 Datasets

The five multi-view multi-label datasets that we used in our experiments are all publicly available online. Their detailed descriptions are summarized in Table 1, including the number of instances (No. instances), number of views (No. views), number of features of each view (No. features of each view), number of class labels (No. labels), average number of labels per instance (LCard) and domain of each dataset (Domain). Specifically, Emotions is a music dataset with two views corresponding to rhythm (8 features) and timbre (64 features). Yeast is a biological dataset with two different views: one view is the genetic expression (79 features), the other is the phylogenetic profile of a gene (24 features). Human and Plant are biological datasets with three views, namely, amino acids (20 features), pseudo amino acids (20 features) and dipeptide compositions (400 features). Pascal VOC is an image dataset with six views: HUE (100 features), SIFT (1000 features), GIST (512 features), HSV (4096 features), RGB (4096 features) and LAB (4096 features).

4.1.2 Evaluation criteria

Six commonly-used multi-label evaluation criteria [18] are adopted for performance comparisons in our experiments, i.e., Average Precision, Coverage, Hamming Loss, OneError, Ranking Loss and Macro AUC. Let $T = \{(x_i, Y_i)\}_{i=1}^m \subset \mathbb{R}^d \times \{+1, -1\}^L$ be the test dataset, $Y_i$ be the set of relevant (irrelevant) labels of $x_i$. In addition, $rank_{g_l}(x)$ stands for a ranking function, where $rank_{g_l}(x) \leq rank_{g_q}(x)$ if $g_l(x) \geq g_q(x)$.

1. **Average Precision**

$$\text{Average Precision} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{|Y_i|} \sum_{l \in Y_i} \frac{|\{q \mid rank_{g_q} \leq rank_{g_l}(x), q \in Y_i\}|}{rank_{g_l}(x)}$$

The average precision evaluates the average fraction of a particular label $l \in Y_i$ that is ranked lower than the relevant labels.

2. **Coverage**

$$\text{Coverage} = \frac{1}{m} \sum_{i=1}^{m} \max_{l \in Y_i} |rank_{g_l}(x) - 1|$$

The coverage evaluates how many steps, on average, it is required to move down the ranked label list to cover all the relevant labels of the instance.

3. **Hamming Loss**

$$\text{Hamming Loss} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{L} |h(x_i) \Delta Y_i|$$

where $h(x_i)$ is the set of predicted labels of $x_i$, and $\Delta$ stands for the symmetric difference between two sets. The hamming loss evaluates the fraction of misclassified instance-label pairs.

4. **OneError**

$$\text{OneError} = \frac{1}{m} \sum_{i=1}^{m} \left\| \arg \max_{l \leq L} g_l(x_i) \notin Y_i \right\|$$

where $\| \pi \| = 1$ if $\pi$ holds, otherwise $\| \pi \| = 0$. The one-error metric evaluates the fraction of instances whose top-ranked label is not in the relevant label set.
Label recovery and label correlation co-learning for multi-view multi-label classification with incomplete labels

\[
\text{Ranking Loss} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{|\mathcal{Y}_i|} |\{(l, q) | \text{rank}_{g_l(x_i)} \geq \text{rank}_{g_q(x_i)}, (l, q) \in \mathcal{Y}_i \times \overline{\mathcal{Y}}_i\}| \quad (22)
\]

where \( \mathcal{Z}_i = \{x_i | Y_i = +1, 1 \leq i \leq m\} \) (\( \overline{\mathcal{Z}}_i = \{x_i | Y_i = -1, 1 \leq i \leq m\} \)) corresponds to the set of test instances with (without) the \( l \)-th label.

The first five evaluation criteria are example-based metrics, while the last is a label-based metric. For Average Precision and Macro AUC, the larger the evaluation value is, the better the algorithm’s performance, with an optimal value of 1. For the other four evaluation criteria, the smaller the evaluation value is, the better the algorithm’s performance, with an optimal value of 0.

4.1.3 Compared methods

We compare our proposed MV2ML method against three state-of-the-art single-view multi-label learning methods, GLOCAL [23], MLMF [25] and LSML [21], and three state-of-the-art MVML approaches, IC2ML [9], McWL [34] and NAIM3L [37].

1. GLOCAL [23]: A single-view multi-label learning algorithm that can simultaneously train the linear classifiers, exploit both global and local label correlations and recover the missing labels.
2. MLMF [25]: A single-view multi-label classification method that can learn binary classifiers, explore and exploit high-order asymmetric label correlations, reduce the dimensionality of the feature space and deal with missing-label data simultaneously.
3. LSML [21]: A single-view multi-label classification approach with missing labels and label-specific features. It trains classifiers, exploits label correlations and recovers the label matrix in a unified learning model.
4. IC2ML [9]: A transductive MVML framework that learns a common subspace of different views, the label correlation matrix and an ensemble classifier based on the individuality and commonality information of multiple views within a unified objective function.
5. McWL [34]: MVML with incomplete (weak) labels based on matrix completion. It optimizes multi-view data integration and matrix completion-based classification in a unified model.
6. NAIM3L [37]: A concise and effective multi-view multi-label classification approach for incomplete views and missing labels. It addresses incomplete and non-aligned views and missing-label problems with only one hyperparameter in a joint learning framework.
7. MV2ML: The method proposed in this paper, which learns and exploits high-order asymmetric label correlations, learns the linear classifiers, recovers missing-label data and fuses the data from different views in a unified learning framework.

In our experiments, all the approaches are implemented in MATLAB. The optimal parameters of the six competitive approaches are selected as suggested in the corresponding literature. For our proposed MV2ML, the regularization parameters \( \lambda_A, \lambda_S \) and \( \lambda_B \) are tuned by threefold cross-validation from the candidate set \( \{2^i\}_{i=7}^{10} \cup \{10^j\}_{j=3}^{3} \). For each kernel function \( k^v(\cdot, \cdot) (v = 1, 2, ..., V) \), the RBF (Radial Basis Function) kernel function is selected. In addition, for three MLL comparison methods, i.e., GLOCAL [23], MLMF [25] and LSML [21], we first transform the multi-view multi-label data into multi-label data by integrating multiple features into a long vector and then conduct the experiments. For fairness, NAIM3L is carried out with the incomplete-label setting.

4.2 Experimental results

For each multi-view multi-label dataset, we randomly sample 60% of it for training and use the remaining data for testing. The data partitioning is randomly repeated ten times. To generate training data with incomplete labels, we

---

1. Code: http://www.lamda.nju.edu.cn/code_Glocal.ashx
2. Code: https://jiunhwang.github.io/
3. Code: http://mlda.swu.edu.cn/codes.php?name=ICM2L
4. Code: http://mlda.swu.edu.cn/codes.php?name=McWL
use the same method as in [36]. More specifically, we set $p = \{0, 30\%, 50\%, 70\%\}$ as the ratio of incomplete labels and drop the assignment of the $l$-th label for $p$ randomly sampled positive and negative training instances of the $l$-th label. This reduces to the full-label case as $p = 0$. The mean value (mean) and the standard deviation (std) of MV2ML and the other compared approaches over each dataset in terms of the six evaluation metrics are recorded in Tables 2, 3, 4, 5 and 6. GLOCAL consumes considerable memory in training on the PASCAL VOC dataset. Thus, for GLOCAL, we first reduce the dimensions by utilizing principal component analysis (PCA) over the PASCAL VOC dataset, and the fraction of features remaining after PCA is set to 0.3 [11]. It is worth noting that $\circ/\circ$ indicates that MV2ML is statistically (through pairwise t-tests at the 5% significance level) superior/inferior to the compared methods; “↑” after the evaluation criterion indicates that the larger the value is, the better the classification performance, while “↓” indicates that the smaller the value is, the better the performance.

It can be seen from Tables 2, 3, 4, 5 and 6 that the performance of all the methods basically decreases as the incomplete-label ratio $p$ increases and our proposed approach MV2ML outperforms other compared approaches in most cases under different values of $p$. Both IC2ML and McWL are MVML approaches, but they are frequently inferior to the single-view multi-label learning methods MLMF and LSML. The main reason is twofold: (1) McWL is a matrix completion-based multi-view multi-label classifier, but it predicts labels based on the composition graph $W$ learned by integrating the weighted adjacency matrix of each view, which may cause its performance to degrade; (2) IC2ML is an MVML algorithm based on non-negative matrix factorization and makes predictions by combining the commonality and individuality information among multi-view data, but it ignores the incomplete labels scenarios.

To further statistically analyze whether there are performance differences among the comparison methods, we employ the popular Friedman test [40], which is a widely-used statistical comparison method for algorithms over multiple datasets. Table 7 summarizes the Friedman Statistics $F_F$ in terms of each evaluation criterion and the corresponding critical values at the 5% significance level. As seen from Table 7, for each evaluation criterion, the null-hypothesis that all comparison methods have equal performance is obviously rejected at the 5% significance level.

Therefore, we adopt the Nemenyi test as a post-hoc test at the 5% significance level to analyze whether the performance between any two comparison methods is significantly different. If the difference between the average ranks of two comparison approaches over all the datasets is greater than the critical difference (CD), then their
| p    | GLOCAL ± | MLMF ± | LSML ± | IC2ML ± | McWL ± | NAIM3L ± | MV2ML ± |
|------|----------|--------|--------|---------|--------|----------|---------|
| 0    | 0.783 ± 0.015| 0.803 ± 0.013| 0.792 ± 0.008| 0.615 ± 0.036| 0.770 ± 0.014| 0.774 ± 0.011| 0.820 ± 0.011|
| 30%  | 0.794 ± 0.015| 0.806 ± 0.018| 0.758 ± 0.019| 0.656 ± 0.040| 0.751 ± 0.025| 0.769 ± 0.015| 0.814 ± 0.007|
| 50%  | 0.766 ± 0.017| 0.789 ± 0.013| 0.755 ± 0.020| 0.674 ± 0.022| 0.731 ± 0.020| 0.749 ± 0.022| 0.813 ± 0.012|
| 70%  | 0.774 ± 0.015| 0.782 ± 0.014| 0.759 ± 0.016| 0.664 ± 0.036| 0.562 ± 0.032| 0.714 ± 0.039| 0.780 ± 0.014|
|      | 1.829 ± 0.097| 1.741 ± 0.056| 1.838 ± 0.059| 2.929 ± 0.176| 1.805 ± 0.047| 1.898 ± 0.074| 1.700 ± 0.088|
| 30%  | 1.818 ± 0.095| 1.768 ± 0.100| 2.030 ± 0.121| 2.621 ± 0.197| 1.910 ± 0.076| 1.940 ± 0.093| 1.747 ± 0.071|
| 50%  | 1.911 ± 0.145| 1.839 ± 0.062| 2.033 ± 0.096| 2.624 ± 0.215| 2.066 ± 0.091| 2.049 ± 0.130| 1.802 ± 0.082|
| 70%  | 1.910 ± 0.091| 1.902 ± 0.052| 2.038 ± 0.103| 2.671 ± 0.189| 3.179 ± 0.177| 2.223 ± 0.240| 1.944 ± 0.072|
|      | 0.311 ± 0.005| 0.203 ± 0.009| 0.209 ± 0.010| 0.361 ± 0.037| 0.230 ± 0.008| 0.219 ± 0.006| 0.266 ± 0.009|
| 30%  | 0.312 ± 0.007| 0.206 ± 0.007| 0.256 ± 0.022| 0.329 ± 0.030| 0.250 ± 0.016| 0.228 ± 0.006| 0.278 ± 0.011|
| 50%  | 0.312 ± 0.008| 0.215 ± 0.009| 0.257 ± 0.010| 0.319 ± 0.021| 0.273 ± 0.015| 0.240 ± 0.011| 0.295 ± 0.011|
| 70%  | 0.312 ± 0.003| 0.226 ± 0.015| 0.265 ± 0.017| 0.324 ± 0.031| 0.408 ± 0.015| 0.270 ± 0.024| 0.305 ± 0.007|
|      | 0.324 ± 0.028| 0.273 ± 0.026| 0.281 ± 0.016| 0.498 ± 0.048| 0.344 ± 0.029| 0.250 ± 0.096| 0.234 ± 0.022|
| 30%  | 0.301 ± 0.027| 0.265 ± 0.026| 0.332 ± 0.032| 0.465 ± 0.076| 0.370 ± 0.042| 0.283 ± 0.098| 0.243 ± 0.019|
| 50%  | 0.350 ± 0.032| 0.291 ± 0.019| 0.351 ± 0.045| 0.424 ± 0.031| 0.381 ± 0.032| 0.200 ± 0.113| 0.240 ± 0.026|
| 70%  | 0.324 ± 0.024| 0.295 ± 0.026| 0.324 ± 0.034| 0.439 ± 0.056| 0.565 ± 0.068| 0.250 ± 0.111| 0.292 ± 0.031|
|      | 0.176 ± 0.014| 0.158 ± 0.012| 0.171 ± 0.008| 0.377 ± 0.043| 0.181 ± 0.006| 0.188 ± 0.011| 0.147 ± 0.013|
| 30%  | 0.171 ± 0.013| 0.157 ± 0.018| 0.202 ± 0.019| 0.321 ± 0.042| 0.199 ± 0.018| 0.197 ± 0.015| 0.155 ± 0.007|
| 50%  | 0.189 ± 0.021| 0.172 ± 0.012| 0.205 ± 0.016| 0.307 ± 0.027| 0.226 ± 0.018| 0.214 ± 0.021| 0.157 ± 0.010|
| 70%  | 0.188 ± 0.016| 0.180 ± 0.010| 0.205 ± 0.015| 0.312 ± 0.038| 0.471 ± 0.046| 0.252 ± 0.045| 0.184 ± 0.011|
|      | 0.823 ± 0.013| 0.830 ± 0.008| 0.832 ± 0.007| 0.526 ± 0.090| 0.827 ± 0.007| 0.808 ± 0.007| 0.859 ± 0.010|
| 30%  | 0.828 ± 0.010| 0.824 ± 0.010| 0.803 ± 0.016| 0.631 ± 0.058| 0.805 ± 0.014| 0.799 ± 0.010| 0.851 ± 0.006|
| 50%  | 0.820 ± 0.012| 0.815 ± 0.010| 0.805 ± 0.018| 0.659 ± 0.046| 0.776 ± 0.013| 0.785 ± 0.013| 0.849 ± 0.006|
| 70%  | 0.813 ± 0.014| 0.802 ± 0.015| 0.800 ± 0.012| 0.679 ± 0.040| 0.552 ± 0.023| 0.747 ± 0.031| 0.817 ± 0.013|
The results (mean ± std) of MV2ML with compared methods on the Yeast dataset with different ratios of incomplete labels

| p  | GLOCAL  | MLMF     | LSML     | IC2ML    | McWL     | NAIM\(^2\)L | MV2ML    |
|----|---------|----------|----------|----------|----------|-------------|----------|
| 0  | 0.616 ± 0.006\\ (<) & 0.758 ± 0.005 & 0.756 ± 0.005 & 0.699 ± 0.005 & 0.734 ± 0.007 & 0.738 ± 0.005 & 0.765 ± 0.004 |
| 30%| 0.610 ± 0.007 & 0.757 ± 0.008 & 0.739 ± 0.007 & 0.698 ± 0.011 & 0.752 ± 0.006 & 0.736 ± 0.005 & 0.764 ± 0.007 |
| 50%| 0.600 ± 0.007 & 0.752 ± 0.004 & 0.741 ± 0.009 & 0.691 ± 0.013 & 0.759 ± 0.005 & 0.732 ± 0.006 & 0.760 ± 0.008 |
| 70%| 0.596 ± 0.008 & 0.749 ± 0.007 & 0.738 ± 0.005 & 0.678 ± 0.016 & 0.756 ± 0.003 & 0.725 ± 0.006 & 0.753 ± 0.007 |

Coverage ↓

| 0  | 8.631 ± 0.102 & 6.389 ± 0.066 & 6.375 ± 0.073 & 7.027 ± 0.135 & 6.452 ± 0.078 & 6.631 ± 0.066 & 6.287 ± 0.078 |
| 30%| 8.728 ± 0.069 & 6.454 ± 0.098 & 6.855 ± 0.115 & 7.123 ± 0.229 & 6.335 ± 0.073 & 6.670 ± 0.079 & 6.349 ± 0.069 |
| 50%| 8.954 ± 0.093 & 6.466 ± 0.083 & 6.822 ± 0.093 & 7.206 ± 0.140 & 6.254 ± 0.071 & 6.764 ± 0.089 & 6.395 ± 0.090 |
| 70%| 9.187 ± 0.181 & 6.518 ± 0.083 & 6.861 ± 0.085 & 7.169 ± 0.268 & 6.304 ± 0.081 & 6.957 ± 0.086 & 6.512 ± 0.099 |

Hamming Loss ↓

| 0  | 0.302 ± 0.002 & 0.202 ± 0.003 & 0.216 ± 0.007 & 0.277 ± 0.007 & 0.238 ± 0.004 & 0.214 ± 0.003 & 0.237 ± 0.003 |
| 30%| 0.302 ± 0.002 & 0.203 ± 0.004 & 0.234 ± 0.006 & 0.278 ± 0.008 & 0.231 ± 0.003 & 0.215 ± 0.003 & 0.242 ± 0.005 |
| 50%| 0.302 ± 0.003 & 0.205 ± 0.002 & 0.232 ± 0.006 & 0.280 ± 0.007 & 0.227 ± 0.003 & 0.218 ± 0.003 & 0.246 ± 0.004 |
| 70%| 0.304 ± 0.003 & 0.206 ± 0.003 & 0.235 ± 0.007 & 0.283 ± 0.007 & 0.230 ± 0.003 & 0.222 ± 0.003 & 0.252 ± 0.003 |

OneError ↓

| 0  | 0.359 ± 0.013 & 0.234 ± 0.006 & 0.232 ± 0.009 & 0.261 ± 0.023 & 0.312 ± 0.011 & 0.454 ± 0.091 & 0.225 ± 0.007 |
| 30%| 0.373 ± 0.014 & 0.236 ± 0.009 & 0.242 ± 0.012 & 0.287 ± 0.048 & 0.261 ± 0.009 & 0.493 ± 0.102 & 0.231 ± 0.011 |
| 50%| 0.360 ± 0.014 & 0.226 ± 0.014 & 0.246 ± 0.013 & 0.262 ± 0.023 & 0.283 ± 0.010 & 0.532 ± 0.080 & 0.230 ± 0.012 |
| 70%| 0.379 ± 0.019 & 0.240 ± 0.010 & 0.246 ± 0.009 & 0.325 ± 0.069 & 0.243 ± 0.006 & 0.561 ± 0.079 & 0.238 ± 0.014 |

Ranking Loss ↓

| 0  | 0.345 ± 0.005 & 0.171 ± 0.003 & 0.172 ± 0.004 & 0.221 ± 0.006 & 0.189 ± 0.004 & 0.188 ± 0.004 & 0.168 ± 0.004 |
| 30%| 0.349 ± 0.005 & 0.173 ± 0.006 & 0.187 ± 0.006 & 0.220 ± 0.009 & 0.175 ± 0.005 & 0.190 ± 0.005 & 0.168 ± 0.005 |
| 50%| 0.359 ± 0.006 & 0.174 ± 0.003 & 0.185 ± 0.006 & 0.224 ± 0.007 & 0.169 ± 0.002 & 0.193 ± 0.005 & 0.169 ± 0.005 |
| 70%| 0.363 ± 0.011 & 0.176 ± 0.005 & 0.187 ± 0.004 & 0.226 ± 0.010 & 0.171 ± 0.004 & 0.201 ± 0.006 & 0.174 ± 0.005 |

MacroAUC ↑

| 0  | 0.667 ± 0.005 & 0.690 ± 0.005 & 0.687 ± 0.011 & 0.500 ± 0.010 & 0.701 ± 0.008 & 0.634 ± 0.006 & 0.702 ± 0.008 |
| 30%| 0.662 ± 0.007 & 0.681 ± 0.007 & 0.649 ± 0.006 & 0.508 ± 0.012 & 0.698 ± 0.010 & 0.628 ± 0.005 & 0.699 ± 0.006 |
| 50%| 0.657 ± 0.007 & 0.674 ± 0.009 & 0.647 ± 0.008 & 0.523 ± 0.012 & 0.695 ± 0.008 & 0.620 ± 0.009 & 0.692 ± 0.009 |
| 70%| 0.647 ± 0.011 & 0.663 ± 0.006 & 0.651 ± 0.007 & 0.524 ± 0.007 & 0.666 ± 0.009 & 0.610 ± 0.008 & 0.673 ± 0.009 |
Table 4: The results (mean ± std) of MV2ML with compared methods on the Human dataset with different ratios of incomplete labels

| p   | GLOCAL | MLMF     | LSML     | IC2ML   | McWL   | NAIM²L   | MV2ML   |
|-----|--------|----------|----------|---------|--------|----------|---------|
| 0   | 0.618 ± 0.010 | 0.624 ± 0.008 | 0.615 ± 0.010 | 0.505 ± 0.022 | 0.597 ± 0.011 | 0.568 ± 0.009 | 0.642 ± 0.006 |
| 30% | 0.612 ± 0.013 | 0.619 ± 0.008 | 0.565 ± 0.014 | 0.523 ± 0.023 | 0.593 ± 0.009 | 0.561 ± 0.009 | 0.637 ± 0.006 |
| 50% | 0.606 ± 0.011 | 0.610 ± 0.010 | 0.557 ± 0.013 | 0.529 ± 0.010 | 0.563 ± 0.008 | 0.551 ± 0.008 | 0.624 ± 0.010 |
| 70% | 0.603 ± 0.004 | 0.602 ± 0.007 | 0.561 ± 0.015 | 0.534 ± 0.008 | 0.518 ± 0.007 | 0.432 ± 0.005 | 0.617 ± 0.006 |

Coverage (↓)

| p   | GLOCAL | MLMF     | LSML     | IC2ML   | McWL   | NAIM²L   | MV2ML   |
|-----|--------|----------|----------|---------|--------|----------|---------|
| 0   | 2.263 ± 0.096 | 2.159 ± 0.056 | 2.215 ± 0.069 | 2.810 ± 0.113 | 2.321 ± 0.084 | 2.598 ± 0.085 | 2.103 ± 0.070 |
| 30% | 2.438 ± 0.148 | 2.189 ± 0.057 | 2.918 ± 0.119 | 2.696 ± 0.130 | 2.326 ± 0.063 | 2.726 ± 0.082 | 2.139 ± 0.050 |
| 50% | 2.461 ± 0.111 | 2.237 ± 0.083 | 2.934 ± 0.120 | 2.683 ± 0.068 | 2.526 ± 0.052 | 2.862 ± 0.069 | 2.193 ± 0.076 |
| 70% | 2.479 ± 0.106 | 2.309 ± 0.057 | 2.940 ± 0.146 | 2.640 ± 0.061 | 2.810 ± 0.052 | 3.989 ± 0.083 | 2.232 ± 0.063 |

Hamming Loss (↓)

| p   | GLOCAL | MLMF     | LSML     | IC2ML   | McWL   | NAIM²L   | MV2ML   |
|-----|--------|----------|----------|---------|--------|----------|---------|
| 0   | 0.085 ± 0.001 | 0.083 ± 0.001 | 0.106 ± 0.004 | 0.150 ± 0.006 | 0.128 ± 0.002 | 0.085 ± 0.001 | 0.084 ± 0.001 |
| 30% | 0.085 ± 0.001 | 0.084 ± 0.001 | 0.125 ± 0.008 | 0.142 ± 0.005 | 0.128 ± 0.002 | 0.086 ± 0.001 | 0.085 ± 0.001 |
| 50% | 0.084 ± 0.001 | 0.085 ± 0.001 | 0.128 ± 0.010 | 0.142 ± 0.004 | 0.135 ± 0.001 | 0.088 ± 0.001 | 0.084 ± 0.000 |
| 70% | 0.084 ± 0.000 | 0.086 ± 0.002 | 0.124 ± 0.005 | 0.139 ± 0.002 | 0.143 ± 0.002 | 0.210 ± 0.007 | 0.085 ± 0.001 |

OneError (↓)

| p   | GLOCAL | MLMF     | LSML     | IC2ML   | McWL   | NAIM²L   | MV2ML   |
|-----|--------|----------|----------|---------|--------|----------|---------|
| 0   | 0.553 ± 0.015 | 0.545 ± 0.011 | 0.561 ± 0.013 | 0.706 ± 0.031 | 0.593 ± 0.016 | 0.762 ± 0.055 | 0.524 ± 0.008 |
| 30% | 0.555 ± 0.015 | 0.549 ± 0.010 | 0.606 ± 0.019 | 0.681 ± 0.033 | 0.594 ± 0.012 | 0.774 ± 0.054 | 0.530 ± 0.011 |
| 50% | 0.566 ± 0.016 | 0.563 ± 0.011 | 0.621 ± 0.019 | 0.669 ± 0.016 | 0.621 ± 0.012 | 0.800 ± 0.039 | 0.550 ± 0.015 |
| 70% | 0.567 ± 0.009 | 0.574 ± 0.011 | 0.613 ± 0.019 | 0.664 ± 0.013 | 0.670 ± 0.012 | 0.860 ± 0.051 | 0.559 ± 0.010 |

Ranking Loss (↓)

| p   | GLOCAL | MLMF     | LSML     | IC2ML   | McWL   | NAIM²L   | MV2ML   |
|-----|--------|----------|----------|---------|--------|----------|---------|
| 0   | 0.148 ± 0.005 | 0.141 ± 0.003 | 0.145 ± 0.005 | 0.191 ± 0.009 | 0.154 ± 0.006 | 0.174 ± 0.005 | 0.136 ± 0.005 |
| 30% | 0.150 ± 0.010 | 0.144 ± 0.004 | 0.194 ± 0.009 | 0.183 ± 0.009 | 0.155 ± 0.005 | 0.183 ± 0.005 | 0.138 ± 0.003 |
| 50% | 0.162 ± 0.007 | 0.147 ± 0.006 | 0.196 ± 0.008 | 0.182 ± 0.005 | 0.169 ± 0.004 | 0.193 ± 0.005 | 0.142 ± 0.005 |
| 70% | 0.163 ± 0.007 | 0.152 ± 0.005 | 0.195 ± 0.010 | 0.179 ± 0.005 | 0.193 ± 0.003 | 0.284 ± 0.005 | 0.146 ± 0.005 |

MacroAUC (↑)

| p   | GLOCAL | MLMF     | LSML     | IC2ML   | McWL   | NAIM²L   | MV2ML   |
|-----|--------|----------|----------|---------|--------|----------|---------|
| 0   | 0.709 ± 0.009 | 0.722 ± 0.013 | 0.728 ± 0.006 | 0.593 ± 0.023 | 0.701 ± 0.012 | 0.670 ± 0.008 | 0.742 ± 0.013 |
| 30% | 0.672 ± 0.015 | 0.717 ± 0.006 | 0.632 ± 0.010 | 0.602 ± 0.028 | 0.670 ± 0.009 | 0.661 ± 0.011 | 0.734 ± 0.014 |
| 50% | 0.659 ± 0.020 | 0.704 ± 0.015 | 0.635 ± 0.015 | 0.615 ± 0.016 | 0.625 ± 0.013 | 0.643 ± 0.009 | 0.719 ± 0.013 |
| 70% | 0.653 ± 0.017 | 0.689 ± 0.008 | 0.632 ± 0.020 | 0.624 ± 0.027 | 0.519 ± 0.012 | 0.592 ± 0.011 | 0.701 ± 0.016 |
Table 5  The results (mean ± std) of MV2ML with compared methods on the Plant dataset with different ratios of incomplete labels

| p     | GLOCAL  | MLMF    | LSML    | IC2ML   | McWL    | NAIM^3L  | MV2ML   |
|-------|---------|---------|---------|---------|---------|----------|---------|
|       | 0       | 0.570 ± 0.016• | 0.581 ± 0.011• | 0.577 ± 0.015• | 0.520 ± 0.016• | 0.539 ± 0.015• | 0.514 ± 0.014• | 0.609 ± 0.013• |
| 30%   | 0.557 ± 0.019• | 0.574 ± 0.015• | 0.509 ± 0.016• | 0.523 ± 0.019• | 0.499 ± 0.009• | 0.428 ± 0.006• | 0.604 ± 0.008• |
| 50%   | 0.549 ± 0.015• | 0.570 ± 0.016• | 0.518 ± 0.024• | 0.513 ± 0.020• | 0.494 ± 0.013• | 0.347 ± 0.025• | 0.584 ± 0.010• |
| 70%   | 0.538 ± 0.018• | 0.559 ± 0.015• | 0.507 ± 0.024• | 0.529 ± 0.019• | 0.488 ± 0.007• | 0.354 ± 0.028• | 0.577 ± 0.015• |
|       | 0       | 2.246 ± 0.158• | 2.028 ± 0.072  | 2.201 ± 0.136• | 2.660 ± 0.152• | 2.474 ± 0.130• | 2.890 ± 0.089• | 1.959 ± 0.140• |
| 30%   | 2.489 ± 0.142• | 2.105 ± 0.142• | 2.862 ± 0.180• | 2.619 ± 0.168• | 2.842 ± 0.115• | 3.697 ± 0.058• | 1.993 ± 0.072• |
| 50%   | 2.556 ± 0.191• | 2.230 ± 0.083• | 2.806 ± 0.197• | 2.674 ± 0.138• | 2.864 ± 0.108• | 4.641 ± 0.353• | 2.125 ± 0.097• |
| 70%   | 2.653 ± 0.231• | 2.302 ± 0.129• | 2.767 ± 0.187• | 2.548 ± 0.078• | 2.884 ± 0.066• | 4.541 ± 0.271• | 2.159 ± 0.092• |
|       | 0       | 0.090 ± 0.001  | 0.093 ± 0.002• | 0.122 ± 0.011• | 0.169 ± 0.004• | 0.164 ± 0.003• | 0.102 ± 0.002• | 0.089 ± 0.001• |
| 30%   | 0.090 ± 0.001  | 0.094 ± 0.002• | 0.146 ± 0.015• | 0.169 ± 0.005• | 0.180 ± 0.003• | 0.224 ± 0.007• | 0.090 ± 0.001• |
| 50%   | 0.089 ± 0.001  | 0.094 ± 0.002• | 0.147 ± 0.016• | 0.171 ± 0.007• | 0.178 ± 0.002• | 0.234 ± 0.009• | 0.090 ± 0.001• |
| 70%   | 0.090 ± 0.001  | 0.096 ± 0.003• | 0.145 ± 0.015• | 0.168 ± 0.005• | 0.182 ± 0.003• | 0.240 ± 0.035• | 0.090 ± 0.001• |
|       | 0       | 0.623 ± 0.022• | 0.614 ± 0.016• | 0.605 ± 0.024• | 0.679 ± 0.024• | 0.653 ± 0.024• | 0.750 ± 0.064• | 0.571 ± 0.018• |
| 30%   | 0.629 ± 0.027• | 0.621 ± 0.017• | 0.688 ± 0.021• | 0.673 ± 0.023• | 0.697 ± 0.013• | 0.822 ± 0.067• | 0.577 ± 0.012• |
| 50%   | 0.638 ± 0.015• | 0.619 ± 0.029• | 0.674 ± 0.034• | 0.689 ± 0.026• | 0.705 ± 0.021• | 0.844 ± 0.057• | 0.606 ± 0.014• |
| 70%   | 0.646 ± 0.021• | 0.632 ± 0.022• | 0.695 ± 0.029• | 0.671 ± 0.030• | 0.715 ± 0.016• | 0.892 ± 0.058• | 0.615 ± 0.030• |
|       | 0       | 0.191 ± 0.014• | 0.171 ± 0.006• | 0.187 ± 0.011• | 0.228 ± 0.012• | 0.214 ± 0.011• | 0.249 ± 0.008• | 0.166 ± 0.011• |
| 30%   | 0.212 ± 0.014• | 0.177 ± 0.011• | 0.247 ± 0.019• | 0.226 ± 0.016• | 0.248 ± 0.010• | 0.323 ± 0.005• | 0.168 ± 0.006• |
| 50%   | 0.220 ± 0.017• | 0.189 ± 0.008• | 0.241 ± 0.018• | 0.232 ± 0.012• | 0.250 ± 0.009• | 0.408 ± 0.032• | 0.180 ± 0.009• |
| 70%   | 0.227 ± 0.020• | 0.195 ± 0.011• | 0.239 ± 0.015• | 0.220 ± 0.007• | 0.251 ± 0.006• | 0.398 ± 0.024• | 0.183 ± 0.007• |
|       | 0       | 0.740 ± 0.018• | 0.759 ± 0.009• | 0.755 ± 0.016• | 0.629 ± 0.030• | 0.627 ± 0.041• | 0.661 ± 0.008• | 0.783 ± 0.011• |
| 30%   | 0.691 ± 0.022• | 0.747 ± 0.018• | 0.630 ± 0.038• | 0.625 ± 0.021• | 0.500 ± 0.014• | 0.594 ± 0.009• | 0.774 ± 0.009• |
| 50%   | 0.666 ± 0.026• | 0.725 ± 0.014• | 0.642 ± 0.022• | 0.650 ± 0.038• | 0.492 ± 0.018• | 0.566 ± 0.019• | 0.752 ± 0.016• |
| 70%   | 0.665 ± 0.032• | 0.704 ± 0.016• | 0.629 ± 0.027• | 0.644 ± 0.024• | 0.501 ± 0.008• | 0.547 ± 0.015• | 0.739 ± 0.014• |
Table 6 The results (mean ± std) of MV2ML with compared methods on the Pascal VOC dataset with different ratios of incomplete labels

| p     | GLOCAL | MLMF  | LSML  | IC2ML    | McWL  | NAIM$	extsuperscript{3}$L | MV2ML |
|-------|--------|-------|-------|----------|-------|-----------------------------|-------|
| 0     | 0.390 ± 0.004 ± | 0.489 ± 0.004 ± | 0.490 ± 0.004 ± | 0.186 ± 0.003 ± | 0.560 ± 0.006 ± | 0.499 ± 0.003 ± | 0.596 ± 0.007 ± |
| 30%   | 0.364 ± 0.008 ± | 0.480 ± 0.005 ± | 0.478 ± 0.003 ± | 0.186 ± 0.002 ± | 0.565 ± 0.004 ± | 0.497 ± 0.003 ± | 0.580 ± 0.008 ± |
| 50%   | 0.343 ± 0.006 ± | 0.482 ± 0.003 ± | 0.480 ± 0.003 ± | 0.186 ± 0.002 ± | 0.555 ± 0.007 ± | 0.495 ± 0.003 ± | 0.570 ± 0.005 ± |
| 70%   | 0.341 ± 0.006 ± | 0.478 ± 0.004 ± | 0.479 ± 0.005 ± | 0.187 ± 0.002 ± | 0.511 ± 0.005 ± | 0.492 ± 0.004 ± | 0.550 ± 0.007 ± |
|       | Coverage (↓)    |       |       |          |       |                             |       |
| 0     | 8.460 ± 0.065 ± | 5.329 ± 0.052 ± | 5.429 ± 0.051 ± | 10.313 ± 0.056 ± | 4.230 ± 0.082 ± | 5.149 ± 0.036 ± | 3.632 ± 0.099 ± |
| 30%   | 8.879 ± 0.145 ± | 5.557 ± 0.047 ± | 5.721 ± 0.056 ± | 10.332 ± 0.043 ± | 4.114 ± 0.078 ± | 5.190 ± 0.038 ± | 3.897 ± 0.133 ± |
| 50%   | 9.254 ± 0.101 ± | 5.561 ± 0.066 ± | 5.701 ± 0.040 ± | 10.302 ± 0.075 ± | 4.199 ± 0.079 ± | 5.261 ± 0.034 ± | 3.953 ± 0.078 ± |
| 70%   | 9.338 ± 0.099 ± | 5.669 ± 0.063 ± | 5.723 ± 0.056 ± | 10.267 ± 0.056 ± | 4.861 ± 0.110 ± | 5.374 ± 0.057 ± | 4.321 ± 0.096 ± |
|       | Hamming Loss (↓) |       |       |          |       |                             |       |
| 0     | 0.074 ± 0.001 ± | 0.073 ± 0.001 ± | 0.104 ± 0.005 ± | 0.164 ± 0.000 ± | 0.103 ± 0.001 ± | 0.071 ± 0.001 ± | 0.070 ± 0.000 ± |
| 30%   | 0.074 ± 0.001 ± | 0.073 ± 0.000 ± | 0.102 ± 0.003 ± | 0.164 ± 0.001 ± | 0.101 ± 0.001 ± | 0.071 ± 0.001 ± | 0.071 ± 0.000 ± |
| 50%   | 0.074 ± 0.001 ± | 0.073 ± 0.000 ± | 0.102 ± 0.003 ± | 0.164 ± 0.001 ± | 0.102 ± 0.001 ± | 0.071 ± 0.001 ± | 0.071 ± 0.001 ± |
| 70%   | 0.074 ± 0.000 ± | 0.073 ± 0.000 ± | 0.102 ± 0.004 ± | 0.164 ± 0.001 ± | 0.107 ± 0.001 ± | 0.072 ± 0.001 ± | 0.072 ± 0.001 ± |
|       | OneError (↓)    |       |       |          |       |                             |       |
| 0     | 0.701 ± 0.006 ± | 0.584 ± 0.006 ± | 0.579 ± 0.005 ± | 1.000 ± 0.000 ± | 0.554 ± 0.008 ± | 0.683 ± 0.048 ± | 0.496 ± 0.007 ± |
| 30%   | 0.730 ± 0.009 ± | 0.585 ± 0.006 ± | 0.611 ± 0.003 ± | 1.000 ± 0.000 ± | 0.544 ± 0.006 ± | 0.683 ± 0.045 ± | 0.509 ± 0.009 ± |
| 50%   | 0.757 ± 0.007 ± | 0.581 ± 0.005 ± | 0.610 ± 0.004 ± | 1.000 ± 0.000 ± | 0.550 ± 0.011 ± | 0.677 ± 0.041 ± | 0.519 ± 0.006 ± |
| 70%   | 0.757 ± 0.007 ± | 0.586 ± 0.006 ± | 0.608 ± 0.010 ± | 1.000 ± 0.000 ± | 0.573 ± 0.005 ± | 0.712 ± 0.027 ± | 0.535 ± 0.011 ± |
|       | Ranking Loss (↓) |       |       |          |       |                             |       |
| 0     | 0.361 ± 0.003 ± | 0.213 ± 0.002 ± | 0.216 ± 0.003 ± | 1.000 ± 0.000 ± | 0.166 ± 0.003 ± | 0.203 ± 0.002 ± | 0.135 ± 0.004 ± |
| 30%   | 0.382 ± 0.007 ± | 0.222 ± 0.003 ± | 0.225 ± 0.002 ± | 1.000 ± 0.000 ± | 0.161 ± 0.003 ± | 0.205 ± 0.002 ± | 0.146 ± 0.006 ± |
| 50%   | 0.400 ± 0.005 ± | 0.222 ± 0.002 ± | 0.224 ± 0.002 ± | 1.000 ± 0.000 ± | 0.165 ± 0.004 ± | 0.208 ± 0.002 ± | 0.150 ± 0.004 ± |
| 70%   | 0.403 ± 0.005 ± | 0.226 ± 0.004 ± | 0.225 ± 0.002 ± | 1.000 ± 0.000 ± | 0.194 ± 0.005 ± | 0.213 ± 0.003 ± | 0.165 ± 0.005 ± |
|       | MacroAUC (↑)    |       |       |          |       |                             |       |
| 0     | 0.549 ± 0.004 ± | 0.693 ± 0.008 ± | 0.720 ± 0.003 ± | 0.500 ± 0.000 ± | 0.806 ± 0.003 ± | 0.737 ± 0.003 ± | 0.845 ± 0.004 ± |
| 30%   | 0.561 ± 0.002 ± | 0.666 ± 0.005 ± | 0.700 ± 0.004 ± | 0.500 ± 0.000 ± | 0.805 ± 0.004 ± | 0.734 ± 0.003 ± | 0.829 ± 0.008 ± |
| 50%   | 0.556 ± 0.007 ± | 0.664 ± 0.006 ± | 0.700 ± 0.003 ± | 0.500 ± 0.000 ± | 0.791 ± 0.004 ± | 0.728 ± 0.002 ± | 0.823 ± 0.008 ± |
| 70%   | 0.564 ± 0.006 ± | 0.661 ± 0.005 ± | 0.699 ± 0.003 ± | 0.500 ± 0.000 ± | 0.743 ± 0.005 ± | 0.719 ± 0.004 ± | 0.802 ± 0.008 ± |
performances are significantly different. Figure 1 plots the comparison results of all approaches against each other in terms of each evaluation metric. In each subfigure, the CD (CD=2.949×√(7×(7+1))/(6×20)≈2.015) is plotted above the axis, the average ranks of all methods are marked along the axis with higher ranks to the left. The groups of compared approaches that are not significantly different, i.e., that have a difference between their average ranks of less than one CD, are connected with a bold solid line.

In Fig. 1(a), MV2ML is connected with MLMF and not linked with the rest of the comparison algorithms. This means that MV2ML achieves comparable performance to that of MLMF and performs better than the rest of the compared approaches on Average Precision. On the other five evaluation criteria, MV2ML achieves a comparable performance with those of MLMF (Fig. 1(b)), MLMF, GLOCAL and NAIM^3L (Fig. 1(c)), MLMF (Fig. 1(d)), MLMF (Fig. 1(e)) and MLMF (Fig. 1(f)). In summary, for each approach, there are 36 comparisons (6 compared algorithms and 6 evaluation criteria). MV2ML obtains statistically superior performance in 80.6% of the cases and achieves statistically comparable performance in only 19.4% of the cases. Therefore, MV2ML achieves better performance than the other compared approaches.

### 4.3 Sensitivity analysis of parameters

To analyze the sensitivity of the parameters of MV2ML w.r.t. $\lambda_A$, $\lambda_S$ and $\lambda_\beta$, we conduct an experiment on the Human dataset, changing the value of one parameter and fixing the values of the other parameters at optimal values. Figs. 2, 3 and 4 report the Average Precision, Coverage, Hamming Loss, OneError, Ranking Loss and Macro_AUC results of MV2ML for different values of $\lambda_A$, $\lambda_S$ and $\lambda_\beta$ on the Human dataset. In addition,
Fig. 2 Sensitivity analysis of parameter $\lambda_A$

Fig. 3 Sensitivity analysis of parameter $\lambda_S$
similar results can be obtained for the other datasets. As shown in Fig. 2, the performance of MV2ML improves as $\lambda_A$ increases and then decreases as $\lambda_A$ increases. The parameter $\lambda_S$ mainly controls the importance of the label correlations. From Fig. 3, we can see that the performance of MV2ML increases as $\lambda_S$ increases and then achieves stable performance when $\lambda_S$ is larger than 32. In MV2ML, the importance degree of each view is controlled by $\lambda_{\beta}$. As seen from Fig. 4, when $\lambda_{\beta}$ is small, the performance of MV2ML has no obvious influence, possibly because $\lambda_{\beta}$ is already heavily constrained to be positive and sum to one.

4.4 Convergence analysis

Figure 5 plots the convergence curve of MV2ML on the Emotions and Yeast datasets with full labels. For both datasets, the objective function value decreases dramatically and then declines gradually as the number of iterations
increases. A similar convergence trend can be observed on the other datasets.

5 Conclusions

In this paper, we address the problems of how to discover label correlations, recover the incomplete label matrix and fuse multi-view data in a collaborative way for multi-view multi-label classification with incomplete labels. A novel approach called label recovery and label correlation co-learning for multi-view multi-label classification with incomplete labels (MV2ML) is proposed. In MV2ML, the asymmetric label correlation matrix is automatically learned from the incomplete training data and used to recover the incomplete label matrix and guide the building of multi-view multi-label classification model by effectively fusing multi-view data iteratively. In this way, the recovery of incomplete label matrix and the learning of label correlations guide each other to boost the prediction performance of the classification model. The experimental results on various multi-view multi-label benchmark datasets verify the effectiveness of our proposed method in solving the problem of MVML with incomplete labels. In future work, we will further extend our proposed model architecture to a deep learning model.

Acknowledgements

This work is supported by National Natural Science Foundation of China under Grants 62041604, 62172198, 61762064, 62063029, Natural Science Foundation of Jiangxi Province under Grant 20222BAB121005, Jiangxi Science Fund for Distinguished Young Scholars under Grant 20192BCBL23001, Science and Technology Research Project of Jiangxi Provincial Education Department under Grant GJJ191152 and Scientific Startup Foundation for Doctors under Grant EA202107235. The authors would like to thank the anonymous referees and the editors for their helpful comments and suggestions.

References

1. Luo Y, Tao DC, Xu C, Liu H, Wen YG (2013) Multiview vector-valued manifold regularization for multilabel image classification. IEEE Trans Neural Netw Learn Syst 24(5):709–722
2. Luo Y, Liu TL, Tao DC, Xu C (2015) Multiview matrix completion for multi-label image classification. IEEE Trans Image Process 24(8):2355–2368
3. Liu M, Luo Y et al (2015) Low-rank multi-view learning in matrix completion for multi-label image classification. 29th AAAI Conference on Artificial Intelligence, pp 2778–2784
4. Zhang YS, Wu J, Cai Z, Yu PS (2020) Multi-View Multi-Label Learning With Sparse Feature Selection for Image Annotation. IEEE Trans Multimed 22(11):2844–2857
5. Wu X, Chen QG et al (2019) Multi-view multi-label learning with view-specific information extraction. 28th International Joint Conference on Artificial Intelligence, pp 3884–3890
6. Chen ZS, Wu X, Chen QG, Hu Y, Zhang ML (2020) Multi-view partial multi-label learning with graph-based disambiguation. In: Proceedings of the 34th AAAI Conference on artificial intelligence (AAAI’20) New York, NY, pp 3553–3560
7. Wu JH, Wu X, Chen QG, Hu Y, Zhang ML (2020) Feature-induced manifold disambiguation for multi-view partial multi-label learning. In: Proceedings of the 26th ACM SIGKDD Conference on knowledge discovery and data mining (KDD’20), Virtual Event, pp 557–565
8. Zhao DW, Gao QW et al (2021) Consistency and Diversity neural network multi-view multi-label learning. Knowl Based Syst 218:106841
9. Tan QY, Yu GX, Wang J et al (2021) Individuality- and commonality-based multiview multilabel learning. IEEE Trans Cybern 51(3):1716–1727
10. Zhang ML, Zhou ZH (2007) ML-KNN: A lazy learning approach to multi-label learning. Pattern Recog 40(7):2038–2048
11. Zhang ML, Peña JM, Robles V (2009) Feature selection for multi-label naive Bayes classification. Inf Sci 179(19):3218–3229
12. Fürnkranz J, Hüllermeier E, Mancia EL, Brinker K (2008) Multilabel classification via calibrated label ranking. Mach Learn 73(2):133–153
13. Zhang Y, Yeung DY (2013) Multilabel relationship learning. ACM Trans Knowl Discov Data 7(2):1–30
14. Huang SJ, Yu Y, Zhou ZH (2012) Multi-label hypothesis reuse. In: Proceedings of the 18th ACM SIGKDD International conference on knowledge discovery and data mining, Beijing, China, pp 525–533
15. Guo YH, Xue W (2013) Probabilistic multi-label classification with sparse feature learning. In: Proceedings of the 23rd International joint conference on artificial intelligence, pp 1373–1379
16. He ZF, Yang M (2019) Sparse and low-rank representation for multi-label classification. Appl Intell 49:1708–1723
17. Ren WIY, Zhang L et al (2017) Robust mapping learning for multi-view multi-label classification with missing labels. International conference on knowledge science, engineering and management, Springer, pp 543–551
18. Zhang ML, Zhou ZH (2014) A review on multi-label learning algorithms. IEEE Trans Knowl Data Eng 26(8):1819–1837
19. Yu HF, Jain P, Kar P, Dhillon IS (2014) Large-scale multi-label learning with missing labels. In: Proceedings of the 31st international conference on machine learning, pp 392–401
20. Wang SF, Wang J, Wang ZY, Ji Q (2014) Enhancing multi-label classification by modeling dependencies among labels. Pattern Recognit 47(10):3405–3413
21. Huang J, Qin F, Zheng X et al (2019) Improving multi-label classification with missing labels by learning label-specific features. Inf Sci 492:124–146
22. Bi W, Kwok JT (2014) Multilabel classification with label correlations and missing labels. In: Proceedings of the 28th AAAI Conference on artificial intelligence, pp 1680–1686
23. Zhu Y, Kwok JT, Zhou ZH (2018) Multi-Label Learning with Global and Local Label Correlation. IEEE Trans Knowl Data Eng 30(6):1081–1094
24. Cheng ZW, Zeng ZW (2020) Joint label-specific features and label correlation for multi-label learning with missing label. Appl Intell 50:4029–4049
25. He ZF, Yang M, Gao Y, Liu HD, Yin YL (2019) Joint multi-label classification and label correlations with missing labels and feature selection. Knowl Based Syst 163:145–158
26. Zhang CQ, Yu ZW et al (2018) Latent semantic aware multi-view multi-label classification. 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, pp 4414–4421
27. Wang GX, Zhang CQ, Zhu PF, Hu QH (2017) Semi-supervised multi-view multi-label classification based on nonnegative matrix
factorization. International Conference on Artificial Neural Networks, pp 340–348.

28. He ZY, Chen C, Bu JJ, Li P, Cai D (2015) Multi-view based multi-label propagation for image annotation. Neurocomputing 168:853–860

29. Zhang MY, Li CS, Wang ZF (2019) Multi-view metric learning for multi-label image classification. IEEE International Conference on Image Processing, pp 2134–2138

30. Sun SL, Zong DM (2021) LCBM: A Multi-view probabilistic model for multi-label classification. IEEE Trans Pattern Anal Mach Intell 43(8):2682–2696

31. Zhu CM, Miao DQ et al (2019) Improved multi-view multi-label learning with incomplete views and labels. International conference on data mining workshops (ICDMW), Beijing, China, pp 689–696

32. Zhu CM, Miao DQ et al (2020) Global and local multi-view multi-label learning. Neurocomputing 371:67–77

33. Zhu CM, Wang PH, Ma L et al (2020) Global and local multi-view multi-label learning with incomplete views and labels. Neural Comput Appl 32:15007–15028

34. Tan QY, Yu GX, Domeniconi C, Wang J (2018) Multi-view weak-label learning based on matrix completion. In: Proceedings of the 2018 SIAM International conference on data mining, pp 450–458

35. Zhao DW, Gao QW, Lu YX, Sun D (2021) Two-step multi-view and multi-label learning with missing label via subspace learning. Appl Soft Comput 102(2):107–120

36. Tan QY, Yu GX, Domeniconi C, Wang J, Zhang ZL (2018) Incomplete multi-view weak-label learning. 27th International Joint Conference on Artificial Intelligence, pp 2703–2709

37. Li X, Chen SC (2020) A concise yet effective model for non-aligned incomplete multi-view and missing multi-label learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, in press

38. Liu XY, Sun LJ, Feng SH (2021) Incomplete multi-view partial multi-label learning. Applied Intelligence. in press

39. Schölkopf B, Herbrich R, Smola AJ (2001) A generalized representer theorem. In: Helmbold, D, Williamson, B, eds Lecture notes in artificial intelligence 2111. Berlin: Springer-Verlag, pp 416–426

40. Demsar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Zhi-Fen He received the Ph.D. degree in the School of Mathematical Sciences from Nanjing Normal University, Nanjing, China, in 2015. She is currently a Lecturer with the School of Mathematics and Information Science, Nanchang Hangkong University, Nanchang, China. Her current research interests include machine learning and computer vision.

Chun-Hua Zhang received the Ph.D. degree in the University of Macau, Macao, China, in 2020. He is currently a Lecturer with the School of Mathematics and Information Science, Nanchang Hangkong University, Nanchang, China. His current research interests include machine learning and computational mathematics.

Bin Liu received the Ph.D. degree in computational mathematics from Dalian University of Technology, Dalian, China, in 2021. He is currently a Lecturer with the School of Mathematics and Information Science, Nanchang Hangkong University, Nanchang, China. His current research interests include geometric processing and deep learning.

Bo Li received the Ph.D. degree in 2008 in computational mathematics, Dalian University of Technology (DUT), Dalian, China. Now he is the professor in the School of Mathematics and Information Science of Nanchang Hangkong University. His current research interests include the areas of image processing and computer graphics.