Feature-Label Dual-Mapping for Missing-Label-Specific Features Learning

Lulu Zhang[1], Yusheng Cheng[1,2,*], Yibin Wang[1,2], Gensheng Pei[1]

[1](*School of Computer and Information, Anqing Normal University, Anqing* 246133)

[2](*Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Fujian Zhangzhou, 363000, China*)

***The e-mail address, and telephone number of the corresponding author: chengyshaq@163.com, +86-18055665796**

Abstract

Label-specific features learning can effectively exploit the unique features of each label, which alleviates the high dimensionality and improves the classification performance of multi-label. However, most existing label-specific features learning algorithms assume that label space is complete, ignoring the effect of missing labels on the classification accuracy. Some methods try to recover the missing labels first and then learn the mapping between the completed label matrix and the feature matrix. However, early intervention in the recovery of lost tags may affect the distribution of original labels to a certain extent. In this paper, Feature-Label Dual-Mapping for missing-label-specific features learning is proposed. According to the information that the label depends on the feature, the dual mapping weight of the complete feature space and the missing label space is jointly learned. Therefore, the proposed algorithm is to conduct latent missing labels recovery by feature-label dual-mapping to directly obtain target weight in this paper, avoiding the negative influence of early label recovery intervention. Compared with several state-of-the-art methods in 10 benchmark multi-label data sets, the results show that the proposed algorithm is reasonable and effective.

Keywords: Multi-label learning; Missing labels; Label-specific features; Feature-label dual-mapping

1. Introduction

With the rapid information development, machine learning has been successfully applied in medical [1], engineering [2,3], energy [4], and other fields. In machine learning, where traditional supervised learning is one of the most studied paradigms, each real-world object is represented by an instance associated with a label. However, in the real world, each instance can be assigned to multiple class labels simultaneously. For example, there may be multiple labels (e.g., *sea*, *boat*, *tree*, and *mountain*) in a landscape picture. Therefore, multi-label learning led researchers to take a keen interest in recent years.

Multi-label learning has received more and more attention, and various classification algorithms have been proposed. These algorithms are usually divided into two categories: problem transformation approaches and algorithm adaptation approaches. Problem transformation approaches transform a multi-label learning problem into either one or more independent single-label problems. For example, Binary Relevance (BR) [5] is a classic algorithm, which directly translates multi-label problems into multiple single-label learning problems without considering correlations among labels. Algorithm adaptation approaches meet the multi-label classification tasks by modifying the traditional single-label learning algorithms, and the multi-label k-nearest neighbor (ML-KNN) [6] is the representative algorithm. Furthermore, existing multi-label algorithms are generally divided into three categories according to the type of label correlation, namely first-order approaches, second-order approaches, and higher-order approaches. First-order approaches mainly transform multi-label learning problems into multiple independent single-label problems for solving, such as BR and ML-KNN, which ignore the correlation among labels. Second-order approaches transform multi-label learning problems into “label ranking” [7-9] problems and consider the correlation between label pairs. The high-order approaches tackle the multi-label learning problem by mining relationships among all the class labels or a subset of class labels, such as Ensembles of Classifier Chains (ECC) [10] and Group sensitive Classifier Chains (GCC) [11]. By investigating previous work on multi-label learning, it can be noticed that the existing multi-label classification methods mainly utilize the same features representation to distinguish all labels. At the same time, this popular strategy might be suboptimal. In multi-label learning, each class label might be determined by some specific features of its own. Some methods have been proposed [20-25] to learn label-specific features for multi-label learning. However, ineffective performance could be achieved by these methods when part of the class labels of the training data are missing.

The previous multi-label classification algorithms assumed that the label vector of each instance in the training data is correct and complete. However, with the rapid development of information, data volume is increasing significantly, which results in a higher cost in obtaining complete labels. Meanwhile, the lack of objective judgment and labeling experience also affects the labeling quality in manual label annotation. Therefore, the phenomenon of label missing is common in multi-label learning. For multi-label classification with missing labels, the simplest method is tantamount to transform the multi-label problem into multiple independent binary classification problems, such as BR [5]. However, this strategy ignores the correlation among the labels, resulting in an ineffective classification of the trained model. To address this problem, there have been attempts to recover the missing labels by exploiting label correlations [12-15]. For example, Zhu et al. [14] proposed the GLOCAL algorithm, which handles missing labels by learning a latent label representation and optimizing label manifolds, while exploring the correlation between global and local labels. He et al. [15] proposed MLMF, which joints multi-label classification and label correlations with missing labels and feature selection.

The above researches show that the multi-label learning with missing labels recovery algorithm has better classification performance. Still, most of them use the same features to predict all the labels. Moreover, the existing multi-label learning algorithms usually recover missing labels first, and then learn the mapping between the reconstructed label matrix and the feature matrix. Early intervention of missing labels recovery may affect the distribution of original labels to a certain extent. Thus, learning the accurate feature-label relationship and guiding the process of label-specific features remains a critical problem for multi-label learning with missing labels.

To address the problems mentioned above, it is proposed to jointly handle missing-labels recovery and label-specific features learning in a unified framework. The method of Feature-Label Dual-Mapping for missing-label-specific features learning (FLDM) is proposed. First, the dual mapping weight of the complete feature space and the missing label space is learned. Considering that the feature space of multi-label learning is complete, only label space is incomplete. Moreover, labels depend on the feature information of the instance [16-18], similar instances usually have similar labels, and similar labels will also have similar label-specific features. For example, if a picture is labeled “elephant” or “tiger”, then there must be a specific area (feature) associated with “elephant” or “tiger”. Therefore, we learn the dual mapping of the complete feature space and the missing label space to obtain the target weight. Then, we learn the label-specific features representation for each class label and L_1 norm is used to extract label-specific features.

The contributions of this paper can be summarized as follows.

- Unlike traditional multi-label learning with missing labels, this paper directly learns the target weight through the dual mapping between features and labels to carry out latent missing labels recovery learning. Therefore, FLDM avoids the negative influence of early intervention of label recovery to a certain extent.
- The traditional multi-label learning algorithms use the same feature sets to predict all the labels, ignoring the unique characteristics of them. Simultaneously, considering the unique features of the label, we combine with label-specific features learning and missing labels recovery together.
- Compared with several state-of-the-art methods in 10 benchmark multi-label data sets, the results show that the proposed algorithm is both reasonable and effective.

The rest of this paper is organized as follows. Section 2 reviews related works on multi-label learning with label-specific features and missing labels. Section 3 expresses the model construction and optimization of the proposed algorithm. The experimental results and comparative analysis are shown in Section 4. Finally, we conclude the paper in Section 5.

2. Related works

2.1 Label-specific features learning

Multi-label learning with label-specific features is to perform low-dimensional data representation learning for each class label. Unlike the traditional multi-label learning algorithm, it fully considers the unique features of the label and becomes a new research direction for multi-label learning [19]. Therefore, Numerous approaches have been proposed to learn label-specific features. For example, Zhang et al. [20, 21] first proposed the concept of multi-label learning with label-specific features (LIFT), which utilizes label-specific features to represent instances for predicting the corresponding class label. However, LIFT does not consider the correlation among labels. LLSF-DL [22] is a multi-label learning

algorithm for learning label-specific features and class-dependent labels, and it is extended to model high-order label correlations by learning class-dependent labels in a sparse stacking way. Weng et al. [23] proposed LF-LPLC, a multi-label learning algorithm based on label-specific features and local pairwise label correlation. MLFC [24] jointly learn label-specific features and label correlations, which assigns feature weight by designing an optimization model while constructing additional features to consider label correlation. However, the construction of label-specific features may encounter an increasing number of feature dimensionalities, and a large amount of redundant information exists in feature space. Xu et al. [25] proposed a multi-label algorithm named FRS-LIFT, which can implement label-specific feature reduction with the fuzzy rough set.

2.2 Multi-label learning with missing labels

In multi-label learning, an instance is often associated with multiple labels simultaneously. Moreover, it is not easy to obtain complete labels in the real world, especially when the label space is too large. The performance of multi-label classification is significantly influenced by missing labels. Thus, some scholars have proposed many multi-label learning algorithms with missing labels recovery to improve performance. According to the difference in learning strategy, these algorithms are usually divided into three categories. The first category of algorithms separately recovery the incomplete label and learn a multi-label classifier. Some scholars assume the missing labels recovery and the unknown label prediction independently, recover the incomplete label matrix of training data first, and then learn a multi-label classifier. For example, Xu et al. [26] assume that the target matrix and side information matrix share the same latent information and complete the label matrix with side information. MMLNECM [27] is a missing labels completion method to solve multi-label classification with missing labels. In MMLNECM, an unbalanced label completion matrix is first obtained by recovering the missing labels, and then using KELM for linear prediction. In the second category, some researchers assume that the missing labels of each instance are known, and construct the loss function without considering these missing labels. LEML [28] addresses the problem of missing labels with multi-label learning in a generic empirical risk minimization (ERM) framework based on the label data that are known. In WELL [29], a similarity matrix is first obtained from the feature space, and then it is used to learn a new label similarity matrix for each class label. Also, these similar matrices are used to guide the recovery of label matrices under the constraint that similar instances have similar class labels. Finally, the third category of algorithms jointly implement the completion of the label matrix and construct multi-label classifiers together. For example, FastTag [30] combines incomplete label reconstruction and classifier construction via co-regularization, and coerces them into an agreement. Huang et al. [31] argue that each class label may be associated with only some specific features and proposed improving multi-label classification with missing labels by learning label-specific features.

3. The proposed method

In this part, we present how to learn label-specific features with missing labels based on the dependency between features and labels, and then introduce how to incorporate label correlation into the model. Finally, we give details of the proposed FLDM model and the optimization method.

In multi-label learning, the training data set can be expressed as $D = \{(\boldsymbol{x}_i, \boldsymbol{Y}_i)|i = 1,2,\cdots,n\}$ with n instances. The feature set is $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]^{\mathrm{T}} \in \mathbb{R}^{n\times d}$, and each instance has a d-dimensional feature vector, and the corresponding label set of each instance is $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n]^{\mathrm{T}} \in \{0,1\}^{n\times l}$, where l is the number of class labels, $y_{ij} = 1$ indicates the i-th instance has the j-th label, $y_{ij} = 0$ indicates the i-th instance does not have the j-th label or the label is missing. The task of multi-label learning is to learn a mapping: $f: \boldsymbol{X} \to \{0,1\}^l$. However, the mapping between features and labels is usually incorrect when the label is missing, resulting in an unsatisfactory classification of label predictions. In addition, the high dimensionality of multi-label data could cause more problems, such as slower solution speed, lower generalization performance, and higher computational cost. To address the above problems, this paper combines missing labels and label-specific features together and proposes a method of label-specific features learning with missing labels based on feature-label dual-mapping.

To show the performance of the proposed algorithm FLDM in more detail, we choose a toy data set to illustrate. The results are shown in Fig. 1. Where $\boldsymbol{X} \in \mathbb{R}^{6\times 5}$ is the input feature matrix of the toy data set, $\boldsymbol{Y} \in \mathbb{R}^{6\times 4}$ is the

corresponding label matrix, and the green box represents the missing labels. Firstly, the Lasso linear regression model is used to study the mapping of features to the missing-label space, and the $\widehat{\boldsymbol{Y}}_{\text{Lasso}} \in \mathbb{R}^{6\times4}$ is the prediction label. In Fig. 1, it can be found that only the mapping between the feature and the label space is learned, and the missing labels cannot be learned. Therefore, the mapping weight learned by the Lasso is not correct. According to the feature-label dependency, this paper adds the mapping learned from the label space to the feature in the basic model, and updates the weight coefficient $\boldsymbol{W} \in \mathbb{R}^{5\times4}$ through dual mapping learning to achieve label completion, where $\widehat{\boldsymbol{W}} \in \mathbb{R}^{5\times4}$ is the coefficient weight learned by FLDM. Using the updated $\widehat{\boldsymbol{W}}$ for label prediction, we find that the prediction label $\widehat{\boldsymbol{Y}}_{\text{FLDM}} \in \mathbb{R}^{6\times4}$ can effectively complete the missing label.

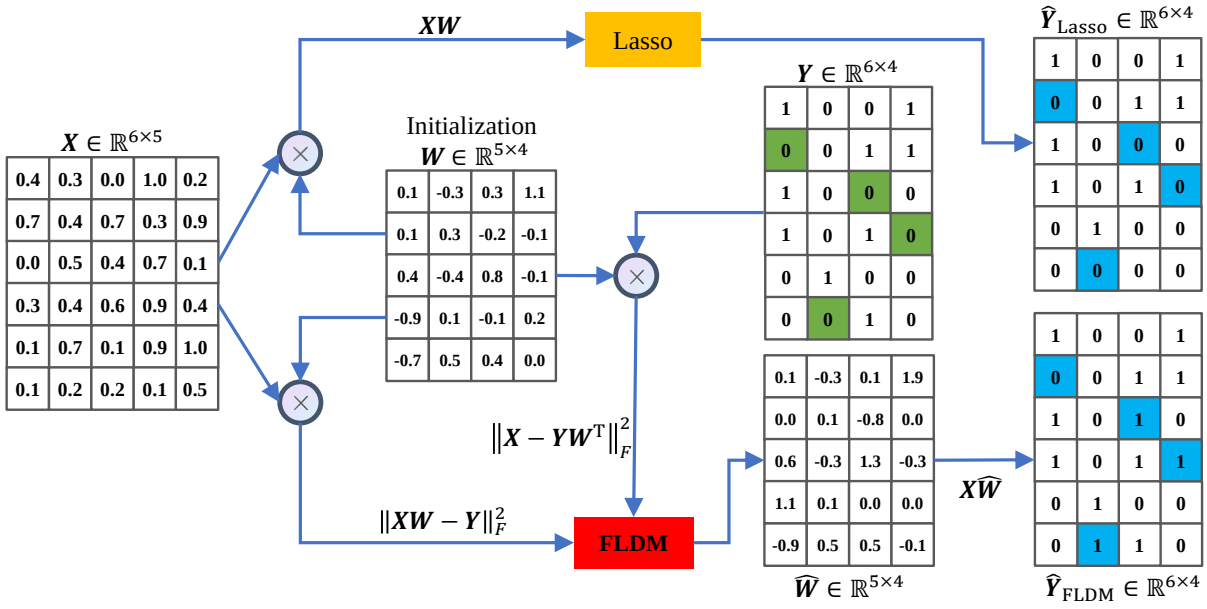

Fig. 1 The details of FLDM are shown in a toy data set. The green box indicates that the missing labels are represented by "0", and the blue box is the inferred labels at the location of missing labels.

3.1 Multi-label learning with label-specific features

Label-specific features learning is to find the most pertinent and discriminative to the corresponding class label, which effectively performs attribute reduction of the training features. Thus, contrasting to the original feature space, label-specific features for each class label are sparse. In this paper, we learn the label-specific features by Lasso linear regression and employ the L_1 regularized term sparse model to extract the label-specific features. Lasso is a compression estimation method with reducing dimensionality, and it compresses the coefficients of variables and makes some regression coefficients zero, thereby achieving variable selection. Therefore, the optimization problem is expressed as:

$$\min_{\boldsymbol{w}^i} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{W}_i - \boldsymbol{Y}_i\|_2^2 + \frac{\gamma}{2}\|\boldsymbol{w}^i\|_1 \tag{1}$$

Combining all the binary classifiers together, the optimization problem can be rewritten as,

$$\min_{\boldsymbol{W}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{W} - \boldsymbol{Y}\|_F^2 + \frac{\gamma}{2}\|\boldsymbol{W}\|_1 \tag{2}$$

where $\boldsymbol{W} = [\boldsymbol{W}_1, \boldsymbol{W}_2, \cdots, \boldsymbol{W}_l] \in \mathbb{R}^{d\times l}$ is the weight matrix, and γ is the parameter to control sparsity. The mapping from feature space to label space can be learned by Eq. (2), the non-zero elements in $\boldsymbol{W}_i$ indicate that the corresponding features are discriminative to label $\boldsymbol{Y}_i$, and these features can be regarded as the label-specific features of the label $\boldsymbol{Y}_i$. However, the missing label space will cause the target weights learned to be incorrect, which will affect the classification performance of multi-label learning. Therefore, it is crucial to learn the correct mapping weights from the feature space to the label space.

3.2 Missing-label-specific features learning via dual-mapping control strategy

In previous label-specific features learning, most algorithms assumed that the labels of the training data are all known. However, in the real world, a large amount of data and the inexperience of the annotator are sometimes difficult to obtain complete labels, which will affect the accuracy of classification. To address the problem, most methods usually perform missing labels recovery first, and then learn the mapping between the reconstructed label matrix and the feature matrix. However, early intervention of missing labels recovery may affect the distribution of original labels to a certain extent. Therefore, it is proposed feature-label dual-mapping for missing-label-specific features learning (FLDM), and it can alleviate the negative influence of early intervention of label recovery.

Till now, we have developed a mapping relation between the features and labels by Lasso. In this section, we try to impose another constraint to improve mapping matrix $\boldsymbol{W}$ via feature-label dual mapping. Since the missing of label leads to less information available in the label space, which reduces the classification accuracy. To enrich the label information, we changed the original mapping target to $\boldsymbol{YY}^{\mathrm{T}}$. In this way, when some labels are missing, the $n\times n$ matrix obtained by $\boldsymbol{YY}^{\mathrm{T}}$ will get more information, such as related information between instances. Furthermore, we can get valuable information about the position and the size of non-zero elements in the $n\times n$ matrix. The optimization function can be improved as follows:

$$\min_{\boldsymbol{W}} \frac{1}{2}\left\|\boldsymbol{XWY}^{\mathrm{T}}-\boldsymbol{YY}^{\mathrm{T}}\right\|_F^2+\frac{\gamma}{2}\|\boldsymbol{W}\|_1 \tag{3}$$

In Eq. (3), we can learn the mapping of the feature space to the label space. Due to the missing labels in the trained multi-label data, it could cause a certain lack of the mapping relationship between the feature and the label space. Simultaneously, considering the dependencies between feature and label, some information is shared between them. We add the mapping of label space to feature space, and the $\boldsymbol{W}$ can be updated by minimizing the square loss $\left\|\boldsymbol{X}-\boldsymbol{YW}^{\mathrm{T}}\right\|_F^2$. Combining with Eq. (3), we learn the dual mapping between feature space and label space to obtain the correct mapping weight $\boldsymbol{W}$. To show the effectiveness of the FLDM in the missing-labels learning more intuitively, we show the prediction results for the feature-label dual mapping weight in Fig. 1. From the experimental results, it can be found that FLDM can correct the prediction errors caused by the missing-label in the dual-weight mapping learning and effectively perform the potential learning of the missing-labels. The optimization problem is formulated as:

$$\min_{\boldsymbol{W}} \frac{1}{2}\left\|\boldsymbol{XWY}^{\mathrm{T}}-\boldsymbol{YY}^{\mathrm{T}}\right\|_F^2+\frac{\alpha}{2}\left\|\boldsymbol{X}-\boldsymbol{YW}^{\mathrm{T}}\right\|_F^2+\frac{\gamma}{2}\|\boldsymbol{W}\|_1 \tag{4}$$

where α is the penalty factor. In multi-label learning, each class labels often have correlations with each other. If there is a strong connection between two class labels, then the discriminative features of one class labels will also be discriminatory to another. On the contrary, the discriminative features of one class label may not be discriminative to another if two class labels are uncorrelated or weakly correlated. In the model of this paper, label-specific features are determined by the non-zero value of the coefficient matrix $\boldsymbol{W}_i$. If the label y_i is strongly correlated with y_j, the similarity between $\boldsymbol{W}_i$ and $\boldsymbol{W}_j$ is large, otherwise the similarity is small. Therefore, the label correlation plays a vital role in dealing with multi-label problems. To further optimize the objective function, the label correlation is incorporated in this paper. Thus, the optimization function is reconstructed as:

$$\min_{\boldsymbol{W}} \frac{1}{2}\left\|\boldsymbol{XWY}^{\mathrm{T}}-\boldsymbol{YY}^{\mathrm{T}}\right\|_F^2+\frac{\alpha}{2}\left\|\boldsymbol{X}-\boldsymbol{YW}^{\mathrm{T}}\right\|_F^2+\frac{\beta}{2}\sum_{j=1}^{l} R_{ij}\boldsymbol{W}_i^T\boldsymbol{W}_j+\frac{\gamma}{2}\|\boldsymbol{W}\|_1 \tag{5}$$

where $R_{ij}=1-C_{ij}$, and $\boldsymbol{R}$ is a symmetric matrix. C_{ij} is the correlation coefficient between labels y_i and y_j, $\boldsymbol{C}\in\mathbb{R}^{l\times l}$ is the label correlation matrix, and β is the parameter that controls label correlation. In this paper, we use cosine similarity to calculate correlations among the labels. The final optimization formulation can be written as:

$$\min_{\boldsymbol{W}} \frac{1}{2}\left\|\boldsymbol{XWY}^{\mathrm{T}}-\boldsymbol{YY}^{\mathrm{T}}\right\|_F^2+\frac{\alpha}{2}\left\|\boldsymbol{X}-\boldsymbol{YW}^{\mathrm{T}}\right\|_F^2+\frac{\beta}{2}tr(\boldsymbol{RW}^{\mathrm{T}}\boldsymbol{W})+\frac{\gamma}{2}\|\boldsymbol{W}\|_1 \tag{6}$$

3.3 Optimization

The objective function is a convex optimization problem, but it is not smooth due to the L_1 norm. Therefore, this

paper uses the accelerated proximal gradient method for solving, considering the original problem as an optimization function with the L_1 norm. It can be re-expressed as:

$$\min_{\boldsymbol{W}\in\mathcal{K}} F(\boldsymbol{W}) = f(\boldsymbol{W}) + g(\boldsymbol{W}) \tag{7}$$

where $\mathcal{H}$ is a real Hilbert space. Both $f(\boldsymbol{W})$ and $g(\boldsymbol{W})$ are convex and $f(\boldsymbol{W})$ is further Lipschitz continuous $\|\nabla f(\boldsymbol{W}_1) - \nabla f(\boldsymbol{W}_2)\| \leq L_f\|\Delta \boldsymbol{W}\|$, where L_f is the *L*ipschitz constant. Therefore, the function $F(\boldsymbol{W})$ in Eq. (7) is divided into two subproblems of $f(\boldsymbol{W})$ and $g(\boldsymbol{W})$, which are expressed as:

$$f(\boldsymbol{W}) = \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}} - \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\right\|_F^2 + \frac{\alpha}{2}\left\|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{W}^{\mathrm{T}}\right\|_F^2 + \frac{\beta}{2}tr(\boldsymbol{R}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}) \tag{8}$$

$$g(\boldsymbol{W}) = \frac{\gamma}{2}\|\boldsymbol{W}\|_1 \tag{9}$$

Since $g(\boldsymbol{W})$ is convex but not smooth, the problem can be solved by Accelerated Proximal Gradient Descent (APGD) [32]. APGD belongs to the greedy minimization iterative strategy algorithm, which can quickly solve the minimization problem with the L_1 norm. Instead of directly minimizing $F(\boldsymbol{W})$, APGD minimizes a sequence of separable quadratic approximations to $F(\boldsymbol{W})$, denoted as $\Phi(\boldsymbol{W}, \boldsymbol{W}^{(t)})$.

$$\Phi(\boldsymbol{W}, \boldsymbol{W}^{(t)}) = f(\boldsymbol{W}^{(t)}) + \left\langle \nabla f(\boldsymbol{W}^{(t)}), \boldsymbol{W} - \boldsymbol{W}^{(t)} \right\rangle + \frac{L_f}{2}\left\|\boldsymbol{W} - \boldsymbol{W}^{(t)}\right\|_F^2 + g(\boldsymbol{W}) \tag{10}$$

Let $\boldsymbol{z}_t = \boldsymbol{W}^{(t)} - \frac{1}{L_f}\nabla f(\boldsymbol{W}^{(t)})$, and the iterative formula of Eq. (10) can be obtained by

$$\begin{aligned}
\boldsymbol{W}_{t+1} &= S_\varepsilon\left[z^{(t)}\right] = \arg\min_{\boldsymbol{W}} \Phi\left(\boldsymbol{W}, \boldsymbol{W}^{(t)}\right) \\
&= \arg\min_{\boldsymbol{W}} \frac{L_f}{2}\left\|\boldsymbol{W} - \boldsymbol{z}^{(t)}\right\|_F^2 + g(\boldsymbol{W}) \\
&= \arg\min_{\boldsymbol{W}} \frac{L_f}{2}\left\|\boldsymbol{W} - \boldsymbol{z}^{(t)}\right\|_F^2 + \frac{\gamma}{L_f}\|\boldsymbol{W}\|_1
\end{aligned} \tag{11}$$

In [32], the work has shown that setting $\boldsymbol{W}^{(t)} = \boldsymbol{W}_t + \frac{b_{t-1}-1}{b_t}(\boldsymbol{W}_t - \boldsymbol{W}_{t-1})$, b_t satisfying $b_{t+1}^2 - b_{t+1} \leq b_t^2$, and $\boldsymbol{W}_t$ is the results of $\boldsymbol{W}$ at *t*-th iteration. Where $S_\varepsilon[\cdot]$ indicates soft threshold, for arbitrary w_{ij}>0, ε>0, soft thresholds can be defined as:

$$S_\varepsilon[w_{ij}] = \begin{cases} w_{ij} - \varepsilon & \text{if } w_{ij} > \varepsilon \\ w_{ij} + \varepsilon & \text{if } w_{ij} < -\varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

According to Eq. (8), $\nabla f(\boldsymbol{W})$ can be calculated as:

$$\nabla f(\boldsymbol{W}) = \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{W}\boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{Y} - \alpha(\boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{Y}^TY) + \beta\boldsymbol{W}\boldsymbol{R} \tag{13}$$

Since solving the coefficient weight $\boldsymbol{W}$ in Eq. (11) requires the Lipschitz constant. According to Eq. (13), given any $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, it can be inferred that:

$$\begin{aligned}
\|\nabla f(\boldsymbol{W}_1) - \nabla f(\boldsymbol{W}_2)\|_F^2 &= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\Delta\boldsymbol{W}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - \alpha\Delta\boldsymbol{W}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} + \beta\Delta\boldsymbol{W}\boldsymbol{R} \\
&\leq 2\left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\Delta\boldsymbol{W}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right\|_F^2 + 2\left\|\alpha\Delta\boldsymbol{W}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right\|_F^2 + 2\|\beta\Delta\boldsymbol{W}\boldsymbol{R}\|_F^2 \\
&\leq 2\left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right\|_2^2\left\|\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right\|_2^2\|\Delta\boldsymbol{W}\|_F^2 + 2\left\|\alpha\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right\|_2^2\|\Delta\boldsymbol{W}\|_F^2 + 2\|\beta\boldsymbol{R}\|_2^2\|\Delta\boldsymbol{W}\|_F^2 \\
&= (2\left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right\|_2^2\left\|\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right\|_2^2 + 2\left\|\alpha\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right\|_2^2 + 2\|\beta\boldsymbol{R}\|_2^2)\|\Delta\boldsymbol{W}\|_F^2 \\
&= (2\sigma_{\max}^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\sigma_{\max}^2(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}) + 2\sigma_{\max}^2(\alpha\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}) + 2\sigma_{\max}^2(\beta\boldsymbol{R}))\|\Delta\boldsymbol{W}\|_F^2
\end{aligned} \tag{14}$$

where $\Delta\boldsymbol{W} = \boldsymbol{W}_1 - \boldsymbol{W}_2$, and σ_{max} is the maximum singular value of the matrix. Therefore, the *L*ipschitz constant L_f can be calculated as:

$$L_f = \sqrt{2\sigma_{\max}^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\sigma_{\max}^2(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}) + 2\sigma_{\max}^2(\alpha\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}) + 2\sigma_{\max}^2(\beta\boldsymbol{R})} \tag{15}$$

The specific process of solving the output weight $\boldsymbol{W}$ by APGD is shown in Algorithm 1:

Algorithm 1: Optimization of FLDM

Input: Training data set $D = \{\boldsymbol{x}_i, \boldsymbol{Y}_i\}_{i=1}^{n}$, test data set $D^* = \{\boldsymbol{x}_j^*\}_{j=1}^{n^*}$, model parameters: α, β, γ, λ;
Output: Coefficient weight: $\boldsymbol{W}^*$
1: **Initialization**: b_0, $b_1 \leftarrow 1$, $t \leftarrow 1$ $\boldsymbol{W}_0, \boldsymbol{W}_1 \leftarrow (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$;
2: compute matrix $\boldsymbol{R}$;
3: **Repeat**
4: calculate *Lipschitz* constant L_f according to Eq. (15);
5: $\boldsymbol{W}^{(t)} \leftarrow \boldsymbol{W}^{(t)} + \frac{b_{t-1}-1}{b_t}(\boldsymbol{W}_t - \boldsymbol{W}_{t-1})$;
6: $\boldsymbol{z}_t = \boldsymbol{W}^{(t)} - \frac{1}{L_f}\nabla f(\boldsymbol{W}^{(t)})$;
7: $\boldsymbol{W}_{t+1} = \arg\min_{\boldsymbol{W}} \frac{L_f}{2}\left\|\boldsymbol{W} - \boldsymbol{z}^{(t)}\right\|_F^2 + \frac{\gamma}{L_f}\|\boldsymbol{W}\|_1$;
8: $\boldsymbol{W}^{(t+1)} \leftarrow \boldsymbol{W}^{(t)}$;
9: $t \leftarrow t + 1$
10: $b_{t+1} \leftarrow \frac{1+\sqrt{4b_t^2+1}}{2}$;
11: **until** *converge*;
12: $\boldsymbol{W}^* \leftarrow \boldsymbol{W}_t$

3.4 Complexity analysis

In our proposed method, $\boldsymbol{X} \in \mathbb{R}^{n\times d}$, $\boldsymbol{Y} \in \{0,1\}^{n\times l}$, and $\boldsymbol{W} \in \mathbb{R}^{d\times l}$, where n is the number of instances, d is the number of features, and l is the number of labels. The time complexity of FLDM mainly includes three parts in Algorithm 1; the most time-consuming steps are 1, 4, and 6. In step 1, the time complexity of initializing $\boldsymbol{W}$ is $O(dnl + d^2n + d^2l)$; In step 4, the time complexity is caused by calculating the Lipschitz constant, which is $O(d^3 + l^3)$; In step 6, the gradient of $F(\boldsymbol{W})$ should be calculated, which produces the complexity of $O(dnl + dl^2 + nl^2 + d^2l)$. Therefore, the total time complexity of FLDM is: $O(dnl + (n + l)d^2 + (n + d)l^2 + d^3 + l^3)$.

4. Experimental results and analysis

4.1 Data sets

To verify the performance of the proposed algorithm, ten public benchmark multi-label data sets were selected in the experiment. Moreover, the data set includes multiple fields, such as music, text, and images. Cal500 is the music data, the Corel6k1 belongs to image data, and the rest of the data sets are all text data sets. The number of instances varies from 502 to 16766, the number of features varies from 68 to 47236, and the number of class labels varies from 22 to 174. The details of each data set are shown in Table 1, where "Cardinality" means label cardinality of the data set. All of these data sets can be obtained from Mulan and PALM.

Table 1 Multi-label data sets

Data set	Instances	Features	Labels	Cardinality	Domain
Cal500	502	68	174	26.044	music
Medical	978	1449	45	1.245	text
Arts	5000	462	26	1.636	text
Education	5000	550	33	1.461	text
Recreation	5000	606	22	1.432	text
Science	5000	743	40	1.451	text
Social	5000	1047	39	1.283	text
Society	5000	636	27	1.692	text
Rcv1s2	6000	47236	101	2.634	text
Corel16k1	16766	500	153	2.859	image

Mulan http://mulan.sourceforge.net/datasets-mlc.html

PALM http://cse.seu.edu.cn/PersonalPage/zhangml/files/Image.rar

4.2 Comparing algorithms

To verify the effectiveness of FLDM, seven state-of-the-art multi-label algorithms were selected for comparative experiments, and five-fold cross-validation was used. In the experiment, the parameters of each comparison algorithm are set according to the parameter range of the original paper. The specific algorithm introduction and parameter setting are summarized as follows:

FastTag [30]: Fast image tagging. It combines both incomplete label reconstruction and classifier construction via co-regularization, and coerces them into an agreement. Parameter γ of it is tuned in $\{10^0, 10^1, ..., 10^4\}$.

FaIE [33]: Feature-aware Implicit label space Encoding. FaIE uses feature-aware implicit label space coding to reduce the label space dimension and complete the label. Parameter α of it is selected from $\{10^{-1}, 10^0, ..., 10^4\}$, τ is chosen from {0.1, 0.2, ..., 1.0}, and the predefined sparsity level in CS is selected from {1, 2, ..., M}, with M being the maximal number of labels in an instance.

Maxide [26]: Speedup Matrix Completion with Side Information. It is a problem transform method for incomplete multi-label learning. The regularization parameter λ is selected from $\{2^{-10}, 2^{-9}, ..., 2^{10}\}$. Parameters γ and ϵ are set to be 2 and 10^{-5}, respectively.

Glocal [14]: It can simultaneously recover the missing labels and train the linear classifiers, explore both global and local label correlations. Parameter $\lambda = 1$, λ_1 to λ_5 are searched in $\{10^{-5}, 10^{-4}, ..., 10^1\}$, k is tuned in $\{0.1l, 0.2l, ..., 0.6l\}$, and g is tuned in {5, 10, 15, 20}.

ESMC [34]: Efficient Semi-supervised Multi-label Classifier. The algorithm is an embedding-based method. It uses a random method to nonlinearly embed label vectors, which can predict tail labels more accurately. It has a good mechanism for dealing with missing labels, large-scale data sets, and unlabeled data. Parameters λ, ρ, and σ_z are set from {10, 100, 1000, 10000, 100000}.

LSML [31]**:** Learning Label-Specific features for multi-label classification with Missing Labels. It combines label-specific features learning and label matrix recovery for multi-label classification. Parameters λ_1, λ_2, λ_3, λ_4 are searched in $\{10^{-5}, 10^{-4}, ..., 10^3\}$.

MNECM [27]: Missing Multi-label learning with Non-Equilibrium Based on Classification Margin algorithm. It performs missing multi-labels learning with non-equilibrium by expanding the label classification margin. The kernel parameter γ and the regularization parameter C are both set to 1.

FLDM: The proposed algorithm is a latent missing labels recovery method that obtains the correct target weight by learning the dual mapping of complete features. Moreover, it combines missing labels and label-specific features for multi-label learning. Parameters α, β, and γ are tuned in $\{2^{-4}, 2^{-2}, ..., 2^{13}\}$, and parameter λ is 2^{-1}.

4.3 Evaluation metrics

To comprehensively evaluate the performance of each comparison algorithm, this paper selects six common evaluation metrics to verify the performance of each algorithm in multi-label classification, which are Hamming Loss, Exact Match, Accuracy, F_1, Macro F_1 and Micro F_1, respectively [10,35,36]. Given a multi-label data set $D = \{(\boldsymbol{x}_i, \boldsymbol{Y}_i)|1 \le i \le n\}$, $\boldsymbol{y}_i \in \{0,1\}^l$ represents the true label of the i-th instance, the predicted label is $\hat{y}$, and the specific definition of each evaluation index is:

Hamming Loss $= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{l}|\boldsymbol{y}_i \Delta \hat{\boldsymbol{y}}_i|$, Δrepresents the symmetric difference between the two sets. Hamming Loss examines the misclassification of instances in a single label. The smaller the indicator, the better the performance of the algorithm.

Exact Match $= \frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{y}_i = \hat{\boldsymbol{y}}_i]$, where $[x]$ is the indication function. Exact Match evaluates how many times the ground truth labels and the predicted labels are exactly matched.

Accuracy $= \frac{1}{n}\sum_{i=1}^{n}\frac{|\boldsymbol{y}_i \wedge \boldsymbol{y}_i|}{|\boldsymbol{y}_i \vee \boldsymbol{y}_i|}$, which evaluates Jaccard similarity between the ground truth labels and the predicted labels.

$F_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{2p_i r_i}{p_i + r_i}$, where p_i and r_i are the precision and recall for the i-th instance, and F_1 is the integrated version

of precision and recall for each instance.

Macro $F_1 = \frac{1}{l}\sum_{i=1}^{l}\frac{2p_i r_i}{p_i+r_i}$ is the integrated version of precision and recall for each label, where p_i and r_i are the precision and recall for the i-th label.

Micro $F_1 = \frac{2\sum_{j=1}^{l}\sum_{i=1}^{n} y_{ij}\hat{y}_{ij}}{\sum_{j=1}^{l}\sum_{i=1}^{n} y_{ij}+\sum_{j=1}^{l}\sum_{i=1}^{n}\hat{y}_{ij}}$ is an extended version of the single label F_1 measure for multi-label classification, and it treats every entry of the label vector as an individual instance regardless of label distinction.

4.4 Result analysis

We randomly selected 80% of each data set as the training part and 20% as the test part to find the best experimental results. The missing rate of class labels for each data set is set to 20% and 60%. According to the missing preset rate, the observed labels of the training data are discarded. To avoid empty classes or instances without positive labels, we choose the largest positive label as the output label to ensure that each instance at least contains one positive label. The comparison results of the experiment in ten multi-label data sets are shown in Tables 2-7, which corresponds to the experimental results of Hamming Loss, Exact Match, Accuracy, F_1, Macro F_1, and Micro F_1, respectively. The symbol "↑" indicates that an evaluation metric equates to a better result. While the converse is true for the symbol "↓", the experimental results include the average and standard deviation of each indicator. Moreover, the best experimental results in each data set are indicated in bold font, and win / tie/loss is given at the end of each table, where "win" indicates that FLDM is superior to the comparison algorithm, "tie" indicates no significant difference, "loss" indicates that FLDM is worse than this comparison algorithm.

Table 2 shows the results of each comparison algorithm in Hamming Loss. With the same missing rate, the present algorithm differs less from other comparison algorithms. In the Medical and Rcv1s2 data sets, the FLDM is superior to other comparison algorithms and slightly better than FastTag and Maxide under the same default rate on all data sets. The results of FaIE have the best performance on all data, followed by Glocal and MNECM, and there is no significant difference between the comparison algorithms on the Corel16k1 data set.

The results of each comparison algorithm under the Exact Match metric are shown in Table 3. The results show that FLDM significantly outperforms the seven comparison algorithms in all the nine data sets except Cal500. Because the Cal500 data set has many labels, each algorithm is difficult to achieve complete matching, so each algorithm result is 0. According to the experimental results, it is can found that the significant performance of FLDM for latent missing labels recovery.

Table 4 shows the experimental results of each comparison algorithm in Accuracy, and it can be found that the results of FLDM in all experimental data sets are better than other comparison algorithms. Moreover, under the evaluation metrics of F_1 and Micro F_1, the proposed algorithm FLDM is also significantly better than other algorithms in all data sets. The specific results are shown in Table 5-6. Table 7 shows the comparison results of each comparison algorithm under the Micro F1, and the FLDM is dominant in most data sets. When the missing rate is 20%, there is no significant difference between the algorithm of FLDM and LSML on the Science data set, but FLDM is slightly worse than LSML on the Rcv1s2 data set.

Table 1 Results of each comparing algorithm in terms of Hamming Loss

Data set	Missing Rate	FastTag	FaIE	Maxide	Glocal	ESMC	LSML	MNECM	FLDM
		Hamming Loss ↓							
Cal500	0.2	0.143±0.002	**0.137**±0.003	0.141±0.002	**0.137**±0.003	0.142±0.003	**0.137**±0.004	0.140±0.002	0.141±0.003
	0.6	0.144±0.002	0.138±0.003	0.142±0.001	**0.137**±0.001	0.146±0.002	0.139±0.002	0.138±0.003	0.144±0.003
Medical	0.2	0.018±0.004	0.012±0.001	0.019±0.006	0.020±0.005	0.019±0.004	0.011±0.002	0.011±0.001	**0.010**±0.002
	0.6	0.019±0.003	0.015±0.005	0.020±0.003	0.018±0.002	0.020±0.004	0.015±0.001	0.015±0.001	**0.011**±0.002
Arts	0.2	0.057±0.001	0.054±0.001	0.061±0.003	0.060±0.002	0.056±0.002	0.055±0.001	**0.053**±0.001	0.059±0.001
	0.6	0.058±0.001	**0.055**±0.001	0.061±0.003	0.060±0.002	0.057±0.002	0.058±0.001	0.057±0.001	0.059±0.002

Education	0.2	0.040±0.001	0.038±0.001	0.039±0.001	0.039±0.001	0.038±0.001	0.038±0.001	**0.036**±0.001	0.042±0.001
	0.6	0.040±0.001	**0.039**±0.000	**0.039**±0.001	**0.039**±0.001	**0.039**±0.001	0.041±0.001	0.040±0.001	0.042±0.000
Recreation	0.2	0.064±0.002	0.056±0.002	0.067±0.001	0.063±0.003	0.060±0.002	0.056±0.002	**0.055**±0.002	0.061±0.002
	0.6	0.064±0.002	**0.057**±0.002	0.067±0.001	0.063±0.003	0.059±0.003	0.058±0.001	0.058±0.001	0.061±0.002
Science	0.2	0.035±0.001	**0.032**±0.001	0.035±0.002	0.033±0.001	0.033±0.002	0.033±0.002	0.033±0.001	0.035±0.001
	0.6	0.036±0.001	**0.032**±0.001	0.036±0.001	0.033±0.001	0.034±0.001	0.034±0.001	0.033±0.001	0.036±0.001
Social	0.2	0.023±0.002	**0.021**±0.001	0.022±0.001	**0.021**±0.001	0.022±0.001	**0.021**±0.001	**0.021**±0.001	0.022±0.001
	0.6	0.022±0.001	**0.021**±0.001	0.022±0.000	**0.021**±0.001	0.023±0.001	0.023±0.000	0.022±0.001	0.022±0.001
Society	0.2	0.056±0.001	**0.052**±0.001	0.056±0.002	0.055±0.002	0.054±0.002	0.053±0.002	0.053±0.001	0.054±0.001
	0.6	0.056±0.002	**0.054**±0.001	0.057±0.001	0.055±0.002	0.055±0.001	**0.054**±0.001	0.055±0.001	0.055±0.001
Rcv1s2	0.2	0.026±0.001	0.024±0.000	0.026±0.000	0.025±0.001	0.025±0.001	0.024±0.001	0.026±0.001	**0.023**±0.001
	0.6	0.026±0.001	0.025±0.001	0.026±0.000	0.025±0.001	0.025±0.001	0.025±0.001	0.026±0.001	**0.024**±0.000
Corel16k1	0.2	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	0.021±0.000
	0.6	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	**0.019**±0.000	0.021±0.000
win/tie/loss	0.2	6/1/3	2/0/8	5/3/2	5/0/5	3/2/5	2/0/8	2/0/8	/
	0.6	4/3/3	2/0/8	5/2/3	4/1/5	4/1/5	3/0/7	2/2/6	

Table 2 Results of each comparing algorithm in terms of Exact Match

Data set	Missing Rate	FastTag	FaIE	Maxide	Glocal	ESMC	LSML	MNECM	FLDM
		Exact Match ↑							
Cal500	0.2	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
	0.6	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Medical	0.2	0.548±0.035	0.577±0.021	0.558±0.033	0.578±0.032	0.584±0.028	0.641±0.039	0.620±0.029	**0.719**±0.027
	0.6	0.511±0.040	0.489±0.027	0.533±0.029	0.580±0.049	0.558±0.031	0.525±0.032	0.482±0.032	**0.686**±0.034
Arts	0.2	0.200±0.008	0.192±0.009	0.203±0.007	0.217±0.010	0.215±0.011	0.277±0.009	0.235±0.010	**0.343**±0.010
	0.6	0.188±0.010	0.172±0.007	0.197±0.009	0.216±0.007	0.207±0.009	0.236±0.009	0.203±0.007	**0.341**±0.011
Education	0.2	0.223±0.014	0.218±0.008	0.235±0.015	0.235±0.013	0.251±0.017	0.274±0.011	0.249±0.015	**0.350**±0.011
	0.6	0.190±0.018	0.193±0.005	0.231±0.017	0.231±0.018	0.230±0.011	0.222±0.010	0.227±0.012	**0.338**±0.008
Recreation	0.2	0.232±0.011	0.227±0.015	0.230±0.014	0.241±0.015	0.253±0.012	0.318±0.011	0.272±0.013	**0.396**±0.011
	0.6	0.218±0.016	0.220±0.016	0.221±0.016	0.241±0.017	0.237±0.016	0.276±0.010	0.231±0.008	**0.395**±0.016
Science	0.2	0.204±0.008	0.208±0.007	0.210±0.009	0.214±0.007	0.241±0.008	0.272±0.016	0.242±0.008	**0.358**±0.013
	0.6	0.190±0.008	0.195±0.007	0.201±0.007	0.212±0.007	0.220±0.006	0.234±0.004	0.209±0.010	**0.352**±0.015
Social	0.2	0.495±0.007	0.482±0.008	0.511±0.005	0.504±0.006	0.512±0.006	0.536±0.002	0.509±0.003	**0.591**±0.011
	0.6	0.475±0.009	0.470±0.011	0.486±0.011	0.504±0.010	0.500±0.009	0.479±0.005	0.487±0.004	**0.586**±0.009
Society	0.2	0.262±0.013	0.269±0.006	0.272±0.020	0.282±0.023	0.298±0.018	0.309±0.009	0.293±0.010	**0.365**±0.012
	0.6	0.241±0.016	0.247±0.005	0.257±0.010	0.278±0.016	0.280±0.012	0.291±0.017	0.274±0.015	**0.359**±0.010
Rcv1s2	0.2	0.144±0.007	0.143±0.009	0.144±0.004	0.152±0.006	0.149±0.009	0.153±0.010	0.147±0.011	**0.203**±0.010
	0.6	0.099±0.009	0.083±0.008	0.101±0.010	0.152±0.008	0.087±0.007	0.092±0.009	0.088±0.009	**0.179**±0.010
Corel16k1	0.2	0.004±0.002	0.005±0.001	0.002±0.001	0.003±0.000	0.007±0.001	0.010±0.001	0.006±0.001	**0.019**±0.001
	0.6	0.003±0.002	0.003±0.001	0.001±0.001	0.004±0.001	0.004±0.001	0.006±0.001	0.004±0.001	**0.018**±0.002
win/tie/loss	0.2	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	/
	0.6	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	

Table 4 Results of each comparing algorithm in terms of Accuracy

Data set	Missing Rate	FastTag	FaIE	Maxide	Glocal	ESMC	LSML	MNECM	FLDM
		Accuracy ↑							

Cal500	0.2	0.176±0.009	0.187±0.006	0.181±0.016	0.196±0.006	0.182±0.017	0.188±0.012	0.181±0.004	**0.239**±0.013
	0.6	0.156±0.006	0.161±0.003	0.163±0.013	0.196±0.006	0.198±0.012	0.196±0.006	0.158±0.002	**0.219**±0.016
Medical	0.2	0.632±0.023	0.643±0.025	0.652±0.020	0.682±0.033	0.684±0.027	0.761±0.029	0.693±0.028	**0.802**±0.023
	0.6	0.579±0.022	0.551±0.028	0.598±0.021	0.689±0.032	0.677±0.020	0.663±0.020	0.544±0.033	**0.777**±0.028
Arts	0.2	0.240±0.006	0.237±0.009	0.266±0.007	0.275±0.006	0.298±0.008	0.354±0.021	0.300±0.009	**0.433**±0.010
	0.6	0.204±0.009	0.207±0.006	0.217±0.005	0.274±0.008	0.279±0.010	0.286±0.015	0.289±0.008	**0.429**±0.011
Education	0.2	0.273±0.018	0.267±0.008	0.281±0.010	0.291±0.015	0.315±0.017	0.355±0.016	0.298±0.013	**0.434**±0.012
	0.6	0.244±0.014	0.234±0.006	0.257±0.013	0.286±0.012	0.299±0.012	0.275±0.009	0.243±0.011	**0.426**±0.006
Recreation	0.2	0.264±0.014	0.261±0.015	0.277±0.013	0.284±0.012	0.328±0.011	0.389±0.012	0.308±0.015	**0.462**±0.012
	0.6	0.245±0.016	0.248±0.016	0.258±0.018	0.284±0.019	0.299±0.013	0.325±0.014	0.291±0.011	**0.458**±0.017
Science	0.2	0.232±0.006	0.248±0.008	0.256±0.007	0.262±0.005	0.312±0.010	0.360±0.022	0.282±0.010	**0.429**±0.009
	0.6	0.223±0.008	0.229±0.008	0.233±0.009	0.261±0.008	0.267±0.009	0.288±0.013	0.248±0.009	**0.419**±0.012
Social	0.2	0.523±0.008	0.527±0.010	0.534±0.007	0.550±0.005	0.587±0.008	0.600±0.007	0.579±0.008	**0.651**±0.010
	0.6	0.491±0.012	0.508±0.013	0.511±0.009	0.551±0.010	0.543±0.010	0.534±0.005	0.537±0.014	**0.644**±0.011
Society	0.2	0.361±0.004	0.340±0.006	0.354±0.014	0.363±0.024	0.398±0.013	0.418±0.004	0.401±0.006	**0.470**±0.011
	0.6	0.327±0.008	0.308±0.004	0.319±0.013	0.359±0.013	0.369±0.011	0.384±0.022	0.363±0.006	**0.462**±0.010
Rcv1s2	0.2	0.211±0.009	0.226±0.010	0.258±0.005	0.281±0.003	0.311±0.008	0.333±0.006	0.291±0.009	**0.354**±0.008
	0.6	0.189±0.010	0.193±0.007	0.218±0.012	0.280±0.013	0.263±0.011	0.276±0.011	0.266±0.010	**0.333**±0.008
Corel16k1	0.2	0.026±0.003	0.028±0.002	0.031±0.002	0.024±0.001	0.043±0.005	0.060±0.004	0.040±0.006	**0.137**±0.003
	0.6	0.023±0.002	0.020±0.002	0.026±0.002	0.032±0.003	0.033±0.002	0.039±0.004	0.032±0.007	**0.134**±0.002
win/tie/loss	0.2	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	/
	0.6	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	

Table 5 Results of each comparing algorithm in terms of F_1

Data set	Missing Rate	FastTag	FaIE	Maxide	Glocal	ESMC	LSML	MNECM	FLDM
		$F_1 \uparrow$							
Cal500	0.2	0.303±0.010	0.310±0.009	0.312±0.011	0.322±0.008	0.306±0.014	0.311±0.017	0.314±0.007	**0.381**±0.017
	0.6	0.287±0.011	0.273±0.004	0.293±0.013	0.323±0.009	0.301±0.010	0.323±0.008	0.268±0.003	**0.355**±0.022
Medical	0.2	0.657±0.027	0.666±0.026	0.697±0.023	0.719±0.033	0.782±0.025	0.802±0.027	0.718±0.028	**0.831**±0.022
	0.6	0.561±0.018	0.572±0.029	0.601±0.019	0.726±0.027	0.717±0.019	0.712±0.019	0.565±0.034	**0.808**±0.027
Arts	0.2	0.247±0.007	0.254±0.009	0.269±0.008	0.297±0.005	0.327±0.017	0.383±0.024	0.331±0.010	**0.466**±0.010
	0.6	0.215±0.012	0.221±0.006	0.242±0.012	0.295±0.008	0.311±0.015	0.305±0.017	0.309±0.008	**0.462**±0.011
Education	0.2	0.278±0.015	0.285±0.009	0.293±0.013	0.311±0.016	0.331±0.013	0.385±0.018	0.316±0.013	**0.466**±0.012
	0.6	0.231±0.012	0.249±0.006	0.278±0.009	0.306±0.010	0.302±0.008	0.294±0.008	0.297±0.007	**0.459**±0.006
Recreation	0.2	0.267±0.014	0.274±0.015	0.289±0.012	0.300±0.011	0.357±0.018	0.414±0.012	0.321±0.015	**0.486**±0.012
	0.6	0.245±0.016	0.258±0.017	0.268±0.017	0.300±0.020	0.321±0.014	0.343±0.015	0.298±0.017	**0.481**±0.018
Science	0.2	0.259±0.005	0.263±0.009	0.269±0.006	0.279±0.004	0.339±0.017	0.392±0.025	0.296±0.010	**0.456**±0.008
	0.6	0.238±0.012	0.241±0.008	0.248±0.009	0.278±0.010	0.289±0.013	0.307±0.017	0.270±0.009	**0.445**±0.012
Social	0.2	0.540±0.009	0.543±0.010	0.555±0.008	0.567±0.005	0.597±0.010	0.623±0.009	0.601±0.010	**0.673**±0.010
	0.6	0.508±0.012	0.521±0.013	0.538±0.009	0.568±0.010	0.541±0.008	0.553±0.006	0.544±0.013	**0.665**±0.011
Society	0.2	0.346±0.006	0.366±0.006	0.357±0.016	0.394±0.024	0.421±0.009	0.460±0.008	0.430±0.007	**0.511**±0.010
	0.6	0.310±0.005	0.332±0.004	0.324±0.011	0.390±0.013	0.400±0.013	0.420±0.026	0.407±0.009	**0.502**±0.011
Rcv1s2	0.2	0.261±0.013	0.259±0.010	0.284±0.008	0.334±0.004	0.358±0.011	0.407±0.016	0.349±0.010	**0.416**±0.009
	0.6	0.220±0.016	0.218±0.007	0.223±0.012	0.333±0.014	0.311±0.017	0.349±0.018	0.318±0.012	**0.398**±0.008
Corel16k1	0.2	0.043±0.006	0.038±0.003	0.031±0.003	0.033±0.002	0.056±0.007	0.083±0.006	0.060±0.007	**0.191**±0.003
	0.6	0.035±0.008	0.027±0.002	0.025±0.007	0.045±0.004	0.049±0.005	0.054±0.006	0.044±0.008	**0.188**±0.003

win/tie/loss	0.2	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	/
	0.6	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	

Table 6 Results of each comparing algorithm in terms of Macro F_1

Data set	Missing Rate	FastTag	FaIE	Maxide	Glocal	ESMC	LSML	MNECM	FLDM
		Macro F_1 ↑							
Cal500	0.2	0.036±0.003	0.037±0.001	0.050±0.002	0.047±0.004	0.046±0.004	0.038±0.003	0.040±0.002	**0.055**±0.004
	0.6	0.031±0.003	0.032±0.001	0.040±0.003	0.047±0.003	0.045±0.005	0.042±0.002	0.034±0.001	**0.051**±0.004
Medical	0.2	0.247±0.023	0.255±0.020	0.266±0.022	0.284±0.022	0.236±0.026	0.315±0.021	0.318±0.032	**0.359**±0.033
	0.6	0.213±0.021	0.228±0.020	0.243±0.019	0.289±0.010	0.223±0.016	0.306±0.016	0.263±0.008	**0.348**±0.027
Arts	0.2	0.137±0.013	0.124±0.006	0.176±0.009	0.161±0.006	0.186±0.011	0.193±0.014	0.181±0.012	**0.220**±0.012
	0.6	0.123±0.010	0.107±0.004	0.121±0.007	0.159±0.005	0.143±0.009	0.138±0.006	0.133±0.010	**0.210**±0.011
Education	0.2	0.099±0.008	0.105±0.013	0.100±0.009	0.108±0.008	0.120±0.010	0.142±0.008	0.132±0.014	**0.165**±0.011
	0.6	0.089±0.007	0.096±0.012	0.098±0.006	0.102±0.005	0.088±0.009	0.096±0.009	0.091±0.012	**0.155**±0.018
Recreation	0.2	0.188±0.007	0.172±0.010	0.213±0.009	0.215±0.005	0.233±0.007	0.258±0.006	0.240±0.013	**0.274**±0.005
	0.6	0.173±0.008	0.165±0.008	0.189±0.010	0.212±0.012	0.200±0.008	0.199±0.015	0.195±0.017	**0.259**±0.011
Science	0.2	0.104±0.011	0.119±0.006	0.121±0.008	0.124±0.003	0.134±0.010	0.168±0.009	0.156±0.013	**0.174**±0.013
	0.6	0.088±0.008	0.100±0.006	0.118±0.007	0.125±0.010	0.108±0.008	0.111±0.005	0.115±0.008	**0.167**±0.010
Social	0.2	0.098±0.010	0.127±0.009	0.130±0.007	0.133±0.013	0.118±0.011	0.122±0.004	0.120±0.005	**0.172**±0.008
	0.6	0.059±0.008	0.090±0.008	0.107±0.005	0.134±0.013	0.077±0.009	0.085±0.008	0.079±0.006	**0.158**±0.014
Society	0.2	0.103±0.009	0.115±0.005	0.121±0.012	0.135±0.010	0.143±0.010	0.156±0.011	0.131±0.008	**0.167**±0.002
	0.6	0.089±0.007	0.102±0.003	0.109±0.010	0.135±0.011	0.119±0.008	0.126±0.009	0.103±0.010	**0.156**±0.008
Rcv1s2	0.2	0.055±0.007	0.064±0.003	0.078±0.014	0.086±0.003	0.097±0.011	0.102±0.007	0.090±0.009	**0.109**±0.004
	0.6	0.045±0.009	0.051±0.002	0.065±0.012	0.086±0.006	0.080±0.008	0.074±0.007	0.076±0.008	**0.098**±0.004
Corel16k1	0.2	0.011±0.001	0.012±0.001	0.009±0.002	0.006±0.001	0.019±0.001	0.020±0.002	0.014±0.001	**0.032**±0.002
	0.6	0.008±0.001	0.006±0.001	0.005±0.002	0.009±0.002	0.014±0.002	0.015±0.003	0.010±0.002	**0.032**±0.002
win/tie/loss	0.2	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	/
	0.6	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	

Table 7 Results of each comparing algorithm in terms of Micro F_1

Data set	Missing Rate	FastTag	FaIE	Maxide	Glocal	ESMC	LSML	MNECM	FLDM
		Micro F_1 ↑							
Cal500	0.2	0.310±0.013	0.305±0.010	0.301±0.011	0.318±0.009	0.305±0.011	0.306±0.017	0.297±0.005	**0.378**±0.017
	0.6	0.228±0.010	0.268±0.004	0.288±0.009	0.319±0.009	0.301±0.009	0.318±0.008	0.264±0.003	**0.352**±0.022
Medical	0.2	0.708±0.024	0.741±0.020	0.680±0.038	0.676±0.059	0.698±0.027	0.808±0.028	0.772±0.020	**0.818**±0.019
	0.6	0.671±0.017	0.651±0.011	0.643±0.021	0.696±0.027	0.678±0.015	0.721±0.012	0.653±0.016	**0.789**±0.029
Arts	0.2	0.328±0.011	0.312±0.010	0.318±0.012	0.353±0.007	0.379±0.011	0.408±0.018	0.387±0.008	**0.431**±0.009
	0.6	0.267±0.012	0.274±0.008	0.285±0.009	0.349±0.009	0.355±0.008	0.322±0.011	0.366±0.010	**0.419**±0.011
Education	0.2	0.345±0.012	0.365±0.010	0.387±0.013	0.391±0.016	0.401±0.012	0.439±0.015	0.397±0.013	**0.453**±0.011
	0.6	0.312±0.007	0.320±0.006	0.336±0.006	0.389±0.012	0.347±0.011	0.337±0.008	0.340±0.009	**0.443**±0.017
Recreation	0.2	0.353±0.011	0.338±0.019	0.370±0.017	0.368±0.010	0.399±0.010	0.441±0.011	0.389±0.018	**0.452**±0.011
	0.6	0.330±0.014	0.319±0.020	0.347±0.022	0.365±0.023	0.356±0.013	0.380±0.011	0.361±0.013	**0.450**±0.017
Science	0.2	0.337±0.014	0.329±0.011	0.334±0.017	0.351±0.005	0.383±0.021	**0.430**±0.027	0.371±0.016	**0.430**±0.010
	0.6	0.310±0.018	0.303±0.013	0.299±0.025	0.351±0.011	0.348±0.017	0.343±0.015	0.340±0.011	**0.414**±0.012
Social	0.2	0.582±0.006	0.589±0.009	0.613±0.008	0.604±0.007	0.600±0.009	0.627±0.002	0.611±0.010	**0.635**±0.012
	0.6	0.551±0.006	0.571±0.012	0.582±0.006	0.605±0.009	0.579±0.006	0.573±0.004	0.575±0.009	**0.625**±0.013

Society	0.2	0.361±0.007	0.386±0.005	0.411±0.010	0.407±0.022	0.437±0.005	0.451±0.006	0.445±0.008	**0.463**±0.009
	0.6	0.349±0.008	0.358±0.003	0.374±0.013	0.404±0.014	0.401±0.003	0.417±0.023	0.397±0.012	**0.452**±0.009
Rcv1s2	0.2	0.303±0.014	0.268±0.008	0.351±0.009	0.357±0.003	0.381±0.009	**0.404**±0.012	0.391±0.011	0.384±0.009
	0.6	0.264±0.012	0.213±0.004	0.325±0.015	0.353±0.012	0.341±0.013	0.348±0.017	0.334±0.010	**0.354**±0.005
Corel16k1	0.2	0.060±0.006	0.053±0.003	0.041±0.005	0.045±0.004	0.087±0.008	0.106±0.007	0.091±0.008	**0.189**±0.004
	0.6	0.039±0.008	0.036±0.002	0.034±0.004	0.060±0.006	0.069±0.007	0.071±0.008	0.068±0.009	**0.185**±0.003
win/tie/loss	0.2	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	8/1/1	10/0/0	/
	0.6	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	

The above analysis shows that the algorithm in this paper is superior to most of the other five evaluation metrics except Hamming Loss. The rationality and effectiveness of FLDM are further illustrated by statistical hypothesis testing and stability analysis. To analyze the performance differences between different algorithms in different data sets, we use the non-parametric Friedman test [37] for performance analysis. In addition, there are eight algorithms and 20 data sets (the sum of the data sets for the missing rate of 20% and 60%). Table six summarizes the F_F value, and the critical value of the Friedman statistic for each evaluation metric at the significance level is $\alpha = 0.05$, and the degree of freedom is $F(7,133)$. Moreover, it can also be clearly seen from Table 8 that there are significant differences between the algorithms.

Table 8 Summary of the Friedman statistics F_F on each evaluation metric and the critical value of $F(7, 133)$ for $\alpha = 0.05$

Evaluation metrics	F_F	Critical value ($\alpha = 0.05$)
Hamming Loss	9.4733	7.0000
Exact Match	28.7666	
Accuracy	85.3308	
F_1	85.0926	
Macro F_1	37.3112	
Micro F_1	41.1243	

Then, we perform the Nemenyi [20,37] test with a significance level of 5% on all algorithms to further analyze the relative performance among all algorithms. If the average ranking between the two algorithms is greater than the critical difference (CD), there is a clear difference between them. Otherwise, there is no significant difference. In Fig. 2, the results show each comparison algorithm in six evaluation metrics, with the CD value being 2.3478. The performance of each algorithm increases from left to right, and a solid black line connects the algorithms with no significant difference. By observing the Nemenyi test results among compared algorithms under each evaluation metric, FLDM is all on the rightmost side of subplots (b)-(f) in Figure 2. Therefore, the proposed algorithm in this paper is overall superior to other algorithms.

For FLDM, in Fig. 2(a), there is no significant difference among Maxide, FastTag, ESMC, Glocal, LSML, and MNECM. FaIE is superior to FLDM. In Fig. 2(b)-(f), there is no significant difference among the FLDM and LSML algorithms in this paper, and it is superior to Maxide, FastTag, ESMC, Glocal, and MNECM. In Fig.2 (c), there is no significant difference between FLDM and ESMC algorithms under the Accuracy metric, and it is clearly superior to Maxide, FastTag, Glocal, and MNECM. Through statistics, it is found that FLDM is superior to other algorithms at a level of 69%, and it has no significant difference from other algorithms at a level of 28.6%.

For LSML, it is superior to other algorithms at a level of 38.1%, and there is no significant difference from other algorithms at a level of 61.9%.

For ESMC, it can be found that the algorithm is superior to the other algorithms at a level of 26.2%, and there is no significant difference from other algorithms at a level of 64.3%.

For Glocal, it is superior to other comparison algorithms at a level of 23.8% and has no significant difference from other algorithms at a level of 61.9%.

Through the above analysis, it can be found that FLDM has the best performance, which is dominant at a level of 69%, followed by the LSML algorithm, which is superior to other algorithms at a level of 38.1%. Moreover, the ESMC

algorithm ranks third, and it is better than other comparison algorithms at a level of 23.8%. Based on the above experimental results analysis, it is further verified that the proposed algorithm FLDM has good performance.

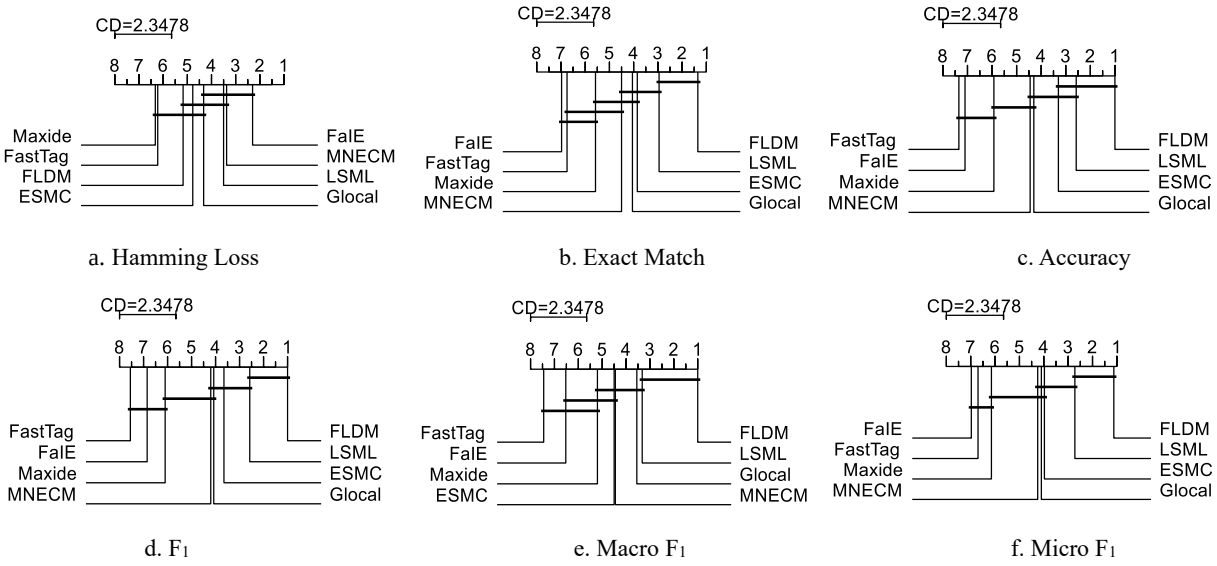

Fig. 2 The performance comparison of algorithms

5. Conclusion

Different from many multi-label learning algorithms with missing labels, which reconstructs the missing label space first, then learning the mapping of features to the reconstructed label space. However, early intervention in the recovery of lost tags may affect the distribution of original labels. In this paper, it is proposed to learn the target weight directly by feature-label dual-mapping. Combined with label-specific features learning, we propose feature-label dual-mapping for missing-label-specific features learning (FLDM), which avoids the negative influence of early intervention of label recovery to a certain extent. Compared with several state-of-the-art methods, the results show that the proposed algorithm is both reasonable and effective. However, the algorithm assumes that the feature space is complete and the label space is incomplete. If the feature space is also incomplete, the algorithm cannot handle it well. Therefore, how to perform multi-label learning with the features and labels are all missing is the focus of future considerations.

Compliance with ethical standards

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61702012 and Key Laboratory of Data Science and Intelligence Application, Fujian Province University (NO. D202005) and Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education (Anhui University) (No.2020A003).

Conflict of interest: The authors declared that they have no conflicts of interest in this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest is connected with the work submitted.

Ethical approval: This article does not contain any studies with human participants or animals performed by any authors.

References:

[1] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in Medicine, 2001, 23(1), 89-109.

[2] Trung N T, Shahgoli A F, Zandi Y, Shariati M, Wakil K, Safa M, Khorami M. Moment-rotation prediction of precast beam-to-column connections using extreme learning machine. Structural Engineering and Mechanics, 2019, 70(5): 639-647.

[3] Shariati M, Trung N T, Wakil K, Mehrabi P, Safa M, Khorami M. Estimation of moment and rotation of steel rack connections using extreme learning machine. Steel and Composite Structures, 2019, 31(5): 427-435.

[4] Petković D, Nikolić V, Mitić V V, & Kocić L. Estimation of fractal representation of wind speed fluctuation by artificial neural network with different training algorothms. Flow Measurement and Instrumentation, 2017, 54: 172-176.

[5] Boutell M R, Luo J, Shen X, Brown C M. Learning multi-label scene classification. Pattern Recognition, 2004, 37(9): 1757-1771.

[6] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition, 2007, 40(7): 2038-2048.

[7] Hüllermeier E, Fürnkranz J, Cheng W, Brinker K. Label ranking by learning pairwise preferences. Artificial Intelligence, 2008, 172(16-17): 1897-1916.

[8] Fürnkranz J, Hüllermeier E, Mencía E L, Brinker K. Multilabel classification via calibrated label ranking. Machine Learning, 2008, 73(2): 133-153.

[9] Zhang M L, Zhou Z H. Multilabel neural networks with applications to functional genomics and text categorization. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(10): 1338-1351.

[10] Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. Machine Learning, 2011, 85(3): 333-359.

[11] Huang J, Li G R, Wang S H, Zhang W G, Huang Q M. Group sensitive classifier chains for multi-label classification, *In Proceedings of IEEE International Conference Multimedia Expo*, 2015: 1–6.

[12] Bi W, Kwok J T. Multilabel classification with label correlations and missing labels. *In Proceedings of 28th AAAI Conference on Artificial Intelligence*, 2014: 1680–1686.

[13] Xu L, Wang Z, Shen Z, Wang Y, Chen E. Learning low-rank label correlations for multi-label classification with missing labels. *In Proceedings of the 14th IEEE International Conference on Data Mining*, 2014: 1067–1072.

[14] Zhu Y, Kwok J T, Zhou Z H. Multi-label learning with global and local label correlation. IEEE Transactions on Knowledge and Data Engineering, 2017, 30(6): 1081-1094.

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

[16] Yu G, Domeniconi C, Rangwala H, Zhang G. Protein function prediction using dependence maximization. *In Proceedings of Conference on Machine Learning and Knowledge Discovery in Databases*. 2013: 574-589.

[17] Zhang Y, Zhou Z H. Multilabel dimensionality reduction via dependence maximization. ACM Transactions on Knowledge Discovery from Data (TKDD), 2010, 4(3): 1-21.

[18] Song L, Smola A, Gretton A, Bedo J, Borgwardt K. Feature selection via dependence maximization. Journal of Machine Learning Research, 2012, 13: 1393-1434.

[19] Zhang M L, Li Y K, Liu X Y, Geng X. Binary relevance for multi-label learning: an overview. Frontiers of Computer Science, 2018, 12(2): 191-202.

[20] Zhang M L, Wu L. Lift: Multi-label learning with label-specific features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 107-120.

[21] Wu L, Zhang M L. Research of label-specific features on multi-label learning algorithm. Journal of Software, 2014, 25(9): 1992-2001 (in Chinese).

[22] Huang J, Li G, Huang Q, Wu X. Learning label-specific features and class-dependent labels for multi-label classification. IEEE Transactions on Knowledge and Data Engineering. 2016, 28(12): 3309-3323.

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

[24] Zhang J, Li C, Cao D, Lin Y, Su S, Dai L, Li S. Multi-label learning with label-specific features by resolving label correlations. Knowledge-Based Systems, 2018, 159: 148-157.
[25] Xu S, Yang X, Yu H, Yu D J, Yang J, Tsang E C. Multi-label learning with label-specific feature reduction. Knowledge-Based Systems, 2016, 104: 52-61.
[26] Xu M, Jin R, Zhou Z H. Speedup matrix completion with side information: Application to multi-label learning. *In Proceedings of 26th International Conference on Neural Information Processing Systems*, 2013: 2301-2309.
[27] Cheng Y, Qian K, Wang Y, Zhao D. Missing multi-label learning with non-equilibrium based on classification margin. Applied Soft Computing, 2020, 86: 105924.
[28] Yu H F, Jain P, Kar P, Dhillon I. Large-scale multi-label learning with missing labels. *International Conference on Machine Learning*. 2014: 593-601.
[29] Sun Y Y, Zhang Y, Zhou Z H. Multi-label learning with weak label. *Proceedings of 24th AAAI Conference on Artificial Intelligence*. 2010: 593-598.
[30] Chen M, Zheng A, Weinberger K. Fast image tagging. *Proceedings of International Conference on Machine Learning*. 2013: 1274-1282.
[31] Huang J, Qin F, Zheng X, Cheng Z, Yuan Z, Zhang W, Huang Q. Improving multi-label classification with missing labels by learning label-specific features. Information Sciences, 2019, 492: 124-146.
[32] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.
[33] Lin Z, Ding G, Hu M, Wang J. Multi-label classification via feature-aware implicit label space encoding. *Proceedings of International Conference on Machine Learning*. 2014: 325-333.
[34] Akbarnejad A H, Baghshah M S. An efficient semi-supervised multi-label classifier capable of handling missing labels. IEEE Transactions on Knowledge and Data Engineering, 2018, 31(2): 229-242.
[35] Wang X, Sukthankar G. Multi-label relational neighbor classification using social context features. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2013: 464-472.
[36] Zhang M L, Zhou Z H. A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 2013, 26(8): 1819-1837.
[37] Demšar J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 2006, 7: 1-30.