Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 1279253, 8 pages
https://doi.org/10.1155/2020/1279253

Research Article

Multilabel Classification Using Low-Rank Decomposition

Bo Yang,[1,2] **Kunkun Tong**,[1] **Xueqing Zhao**,[1] **Shanmin Pang**,[3] **and Jinguang Chen**[1]

[1]*Shaanxi Key Laboratory of Clothing Intelligence, National and Local Joint Engineering Research Center for Advanced Networking & Intelligent Information Services, School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China*
[2]*School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China*
[3]*School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China*

Correspondence should be addressed to Jinguang Chen; chenjinguang@xpu.edu.cn

Received 19 December 2019; Accepted 21 February 2020; Published 7 April 2020

Guest Editor: Longzhuang Li

In the multilabel learning framework, each instance is no longer associated with a single semantic, but rather with concept ambiguity. Specifically, the ambiguity of an instance in the input space means that there are multiple corresponding labels in the output space. In most of the existing multilabel classification methods, a binary annotation vector is used to denote the multiple semantic concepts. That is, +1 denotes that the instance has a relevant label, while −1 means the opposite. However, the label representation contains too little semantic information to truly express the differences among multiple different labels. Therefore, we propose a new approach to transform binary label into a real-valued label. We adopt the low-rank decomposition to get latent label information and then incorporate the information and original features to generate new features. Then, using the sparse representation to reconstruct the new instance, the reconstruction error can also be applied in the label space. In this way, we finally achieve the purpose of label conversion. Extensive experiments validate that the proposed method can achieve comparable to or even better results than other state-of-the-art algorithms.

1. Introduction

Classification is a high-frequency vocabulary in machine learning. We often say that classification generally refers to single-label classification, that is, an object is given a category. In multilabel learning, the meaning of classification is multilabel classification. Specifically, an instance is associated with more than one class label simultaneously. Multilabel learning has many application fields, such as web mining [1–3], text categorization [4–6], multimedia contents annotation [7–11], and bioinformatics [12–14].

In recent years, the field of multilabel learning has gradually attracted significant attention. A variety of algorithms have been proposed, which can be basically divided into two categories [15]: algorithm adaptation and problem transformation. The core idea of the former is to transform the previous supervised learning algorithm so that it can be used to solve multilabel learning problems, such as ML-kNN [16], while the latter is to convert the multilabel learning problem into other known problems to solve, such as BR [17]. Some multilabel algorithms solve the multilabel learning problem without using the correlation among different labels, such as LIFT [18]. The main idea of the LIFT is to obtain the identifying characteristics of each label and build a new feature space. It first obtains the positive and negative examples corresponding to each label and then performs cluster analysis on the corresponding set of examples to obtain the cluster centers and finally uses the cluster centers to construct the label-specific features. In the process of solving the multilabel learning problem, LIFT does not consider label correlations; hence, it can be regarded as a new feature conversion method. Some algorithms consider the label correlation [19–25] for solving the multilabel learning problem. For example, the basic idea in [20] is to model the correlation among labels based on the Bayesian network and to achieve efficient learning by using the approximate strategy. Indeed, the rational use of the correlation among labels can effectively boost the

performance of multilabel classification. For example, if an image has labels "football" and "rainforest," it is likely to be labeled "Brazil". It has a low probability of being labeled "river" if a document is annotated with "desert". Therefore, how to effectively explore and make full use of label correlations is a crucial problem for multilabel learning.

In fact, for an object with multiple labels, the importance of the related labels is still different. Although the importance of each label is not given directly, we can judge the importance of each label through external observation. Generally speaking, the larger the proportion in the original object, the more important the corresponding label. Accordingly, how to accurately express the importance of the label is also a challenge.

The method in [26] decomposes the original output space in order to obtain potential label semantic information, which can effectively increase the ability of the subsequent feature selection. Motivated by the decomposition of the label space in [26], in the paper, we propose a method named label low-rank decomposition (LLRD) for multilabel classification. The LLRD algorithm first performs low-rank decomposition on the label matrix, then combines the decomposed results with the original features to form new features, and mines the structural information of the feature through sparse reconstruction. Third, it transforms the binary label into the real-valued and finally converts the classification problem into a regression problem.

The contribution of this paper is as follows:

(1) Utilize low-rank decomposition to reveal the global label correlations and achieve good classification results

(2) Combine the low-rank decomposition results with the original features reducing the information loss in the subsequent label transformation process

(3) Carry out extensive experiments on different field datasets to verify the effectiveness of different algorithms

2. Materials and Methods

2.1. Datasets. In this experiment, a total of 13 datasets were used covering four fields: audio, text, image, and biology. All these data resources can be collected from Mulan (http://mulan.sourceforge.net/datasets.html) and Meka (http://meka.sourceforge.net/#datasetsru). Table 1 gives the specific details of the datasets. The number of instances, label space, and the dimension of features are denoted by |S|, L(S), and D(S), respectively. LDen (S) is the density of label, which is the result of the normalization of label cardinality LCard(S).

2.2. Notations. Formally, suppose $\mathcal{X} = \mathbb{R}^d$ be the d-dimensional input space and $\mathcal{Y} = \{l_1, l_2, \ldots, l_q\}$ denote the output domain of q class labels. Let $\mathcal{D} = \{(x_i, y_i) | 1 \le i \le p\}$ be the multilabel training dataset with p examples, where $x_i \in \mathcal{X}$ is a d-dimensional instance vector and $y_i \subseteq \mathcal{Y}$ is the label vector corresponding to x_i. Let

Table 1: Properties of the experimental datasets.

Datasets	|S|	D(S)	L(S)	LCard(S)	LDen(S)	Domain
cal500	502	68	174	26.044	0.150	Audio
Emotions	593	72	6	1.868	0.311	Audio
Medical	978	1449	45	1.245	0.028	Text
Llog	1460	1004	75	1.180	0.016	Text
Image	2000	294	5	1.236	0.247	Image
Scene	2407	294	5	1.074	0.179	Image
Yeast	2417	103	14	4.237	0.303	Biology
Slashdot	3782	1079	22	1.181	0.054	Text
rcv1subset1	6000	500	101	2.880	0.029	Text
rcv1subset2	6000	500	101	2.634	0.026	Text
rcv1subset3	6000	500	101	2.614	0.026	Text
rcv1subset4	6000	500	101	2.484	0.025	Text
rcv1subset5	6000	500	101	2.642	0.026	Text

$X = [x_1, x_2, \ldots, x_p] \in \mathbb{R}^{d \times p}$ represent the input data matrix, and $X_i = [x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p]$ denote the matrix from which x_i is removed from X. Let $Y = [y_1, y_2, \cdots, y_p] \in \{-1, 1\}^{q \times p}$ is a matrix composed of label vector.

2.3. The Process of LLRD. First, LLRD decomposes the label matrix with low-rank method. In the framework of multilabel learning, label matrix is often considered to be low rank [27, 28] due to the existence of label correlations. Low-rank structure is also a way to explore the global relationship between labels. Therefore, we can perform low-rank decomposition on the label matrix. Assuming that the rank of Y is $r < q$, Y can be written as follows:

$$Y \simeq AB, \tag{1}$$

where $A \in \mathbb{R}^{q \times r}$ represents the dependency of $B \in \mathbb{R}^{r \times p}$ on the original label space and B is a mapping of the original label and also contains label correlation information.

Second, we combine B with X to form a new feature space $N = [X; B][n_1, n_2, \ldots, n_p] = \in \mathbb{R}^{(r+d) \times p}$. In order to reveal the inner structure of the feature space, we use sparse reconstruction [29] method to model the relationship between the training instances. Specifically, we use $W[s_{ij}]_{=p \times p}$ to represent the training object relationship matrix, where s_{ij} is a measure of the relationship between n_i and n_j. Let $S_i = [s_{1i}, \ldots, s_{i-1,i}, s_{i+1,i}, \ldots, s_{pi}]^{\mathrm{T}}$ denote the corresponding sparse reconstruction coefficient related to n_i. According to the sparse representation theory, S_i can be calculated as follows:

$$\min_{S_i} \left\| N_i S_i - n_i \right\|_2^2 + \eta \left\| S_i \right\|_1, \tag{2}$$

where $N_i = [n_1, n_2, \ldots, n_{i-1}, n_{i+1}, \ldots, n_p]$ represent a combination of all training instances except n_i. We can solve the above problem using alternating direction method of multiplier [30].

Third, we transform the original binary label set $y_i = (l_{i1}, l_{i2}, \ldots, l_{iq})^{\mathrm{T}}$ associated with any x_i in the training set into a real-valued label vector $c_i = (c_{i1}, c_{i2}, \ldots, c_{iq})^{\mathrm{T}}$, where $l_{ij} \in \{-1, 1\}$ and $c_{ij} \in \mathbb{R}$. Because the real value contains more information, and through the size of the value, we can also infer the importance of the label. Since the input space

and the label space are often interrelated, it is assumed that the relationship between n_i and n_j in the input space also exists between c_i and c_j in the label space. Accordingly, the representation errors of different elements in the label space can be written as follows:

$$\min_{C} \sum_{i=1}^{p} \left\| c_i - \sum_{j=1}^{p} s_{ij} c_j \right\|_2^2 \quad s.t.\ k_1 \le l_{ij} c_{ij} \le k_2 \quad (1 \le i \le p, 1 \le j \le q), \tag{3}$$

where $c = [c_1, c_2, \ldots, c_p]$. The above quadratic programming problem can be solved by mature tools related to quadratic programming. The original multilabel classification problem can be transferred into a multioutput regression problem. There are many solutions [31] to solve it. The learning of LLRD method contains three phases: low-rank decomposition, sparse reconstruction, and multioutput regression. The time complexity of low-rank decomposition and sparse reconstruction is $O(d^2 p + d^3)$. If we choose multioutput support vector regression to realize the classification, the time complexity is $O(qp^3)$. Thus, the total complexity of LLRD is $O(d^2 p + d^3 + qp^3)$.

3. Results and Discussion

3.1. Experiment Setup. In this subsection, we investigate comparisons between our LLRD and other six multilabel learning methods on six multilabel evaluation criteria, which include two categories: example-based and label-based metrics [32]. The example-based metric is to first obtain the performance of the learning system on each test example and finally returns the average of the entire test set. Unlike the above example-based metric, the label-based metric first returns the performance of the system on each label and finally gets the macro/microaveraged $F1$ value on all labels.

In this paper, *one-error*, *coverage*, *ranking loss*, and *average precision* are employed for example-based performance evaluation. And *macroaveraging* and *microaveraging* $F1$ are label-based metrics. For example-based metrics except *average precision*, as their values increase, it means that the performance of the algorithm is worse. For the remaining metrics, their values are proportional to the performance of the algorithm.

Let $T = \{(x_i, y_i)\}_{i=1}^{m} \subset \mathbb{R}^d \times \{+1, -1\}^q$ be the multilabel test set and $f(x, l)$ can be seen as the confidence of l being the corresponding label associating with x. In addition, $f(x, l)$ can be converted into a ranking function $\text{rank}_f(\mathbf{x}, l)$. If $f(x, l_1) > f(x, l_2)$ holds, then the corresponding ranking function has $\text{rank}_f(x, l_1) < \text{rank}_f(x, l_2)$.

The six evaluation criteria for the algorithm used in the paper are defined as follows:

(1) *One-error*:

$$\text{one-error}(f) = \frac{1}{m} \sum_{i=1}^{m} \left[\left[\arg\max_{l \in \mathcal{Y}} f(x_i, l) \right] \notin y_i \right]. \tag{4}$$

(2) *Coverage*:

$$\text{coverage}(f) = \frac{1}{m} \sum_{i=1}^{m} \max_{l \in y_i} \text{rank}_f(x_i, l) - 1. \tag{5}$$

(3) *Ranking loss*:

$$\text{rloss}(f) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{|y_i||\overline{y}_i|} \left| \{(l', l'') | f(x_i, l') \le f(x_i, l''), \cdot (l', l'') \in y_i \times \overline{y}_i\} \right|. \tag{6}$$

(4) *Average precision*:

$$\text{avgprec}(f) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{|y_i|} \sum_{l \in y_i} \frac{\left| \{l' | \text{rank}_f(x, l') \le \text{rank}_f(x_i, l), l' \in y_i\} \right|}{\text{rank}_f(x_i, l)}. \tag{7}$$

(5) *Macroaveraging* $F1$:

$$F1_{\text{macro}}(h) = \frac{1}{q} \sum_{j=1}^{q} \frac{2\text{TP}_j}{2\text{TP}_j + \text{FN}_j + \text{FP}_j}. \tag{8}$$

(6) *Microaveraging* $F1$:

$$F1_{\text{micro}}(h) = \frac{2 \sum_{i=1}^{q} \text{TP}_j}{2 \sum_{i=1}^{q} \text{TP}_j + \sum_{i=1}^{q} \text{FN}_j + \sum_{i=1}^{q} \text{FP}_j}, \tag{9}$$

where FN_j, TN_j, FP_j, and TP_j indicate the number of false-negative, true-negative, false-positive, and true-positive instances with regard to l_j.

In order to test the effectiveness of LLRD, we chose six multilabel learning algorithms MLFE [33], RAKEL [34], ML^2 [35], CLR [36], LIFT [18], and RELIAB [37] for performance comparison. MLFE makes full use of the intrinsic information in feature space, making the semantics of the label space more abundant. The specific parameters of MLFE are set as follows: $\rho = 1$, $c_1 = 1$, $c_2 = 2$, and β_1, β_2, and β_3 searched from {1, 2, . . .,10}, {1, 10, 15}, and {1, 10}. RAKEL is a high-order approach. The basic idea of the algorithm is to transform the multilabel learning problem into integration of multiclass classification

Table 2: Performance of each multilabel algorithm (mean ± std. deviation) on the regular-scale datasets.

Comparing algorithms	cal500	Emotions	Medical	Llog	Image	Scene	Yeast
One-error ↓							
LLRD	0.136 ± 0.041	**0.248 ± 0.048**	**0.125 ± 0.031**	**0.657 ± 0.038**	**0.244 ± 0.018**	**0.116 ± 0.019**	**0.217 ± 0.013**
MLFE	0.168 ± 0.049	0.259 ± 0.050	0.131 ± 0.030	0.672 ± 0.041	0.257 ± 0.031	0.127 ± 0.022	0.233 ± 0.026
LIFT	0.125 ± 0.049	0.251 ± 0.027	0.156 ± 0.041	0.664 ± 0.034	0.276 ± 0.026	0.132 ± 0.012	0.226 ± 0.021
RELIAB	**0.116 ± 0.030**	0.255 ± 0.041	0.163 ± 0.028	0.754 ± 0.035	0.342 ± 0.032	0.258 ± 0.011	0.255 ± 0.016
ML^2	0.201 ± 0.090	0.261 ± 0.045	0.135 ± 0.032	0.674 ± 0.051	0.260 ± 0.027	0.144 ± 0.019	0.246 ± 0.034
CLR	0.243 ± 0.058	0.310 ± 0.019	0.362 ± 0.009	0.841 ± 0.036	0.449 ± 0.013	0.331 ± 0.031	0.234 ± 0.022
RAKEL	0.622 ± 0.065	0.289 ± 0.032	0.237 ± 0.032	0.871 ± 0.028	0.397 ± 0.019	0.314 ± 0.030	0.291 ± 0.031
Coverage ↓							
LLRD	0.774 ± 0.021	0.282 ± 0.034	**0.029 ± 0.009**	0.194 ± 0.025	**0.157 ± 0.010**	**0.008 ± 0.009**	**0.447 ± 0.010**
MLFE	0.769 ± 0.024	0.283 ± 0.030	0.033 ± 0.010	0.200 ± 0.027	0.162 ± 0.018	0.012 ± 0.008	0.449 ± 0.011
LIFT	0.753 ± 0.015	**0.271 ± 0.023**	0.040 ± 0.014	0.164 ± 0.007	0.172 ± 0.013	0.026 ± 0.007	0.454 ± 0.017
RELIAB	**0.746 ± 0.019**	0.306 ± 0.020	0.044 ± 0.013	**0.155 ± 0.013**	0.185 ± 0.007	0.114 ± 0.004	0.457 ± 0.015
ML^2	0.814 ± 0.033	0.292 ± 0.044	0.035 ± 0.013	0.201 ± 0.026	0.164 ± 0.009	0.010 ± 0.007	0.461 ± 0.016
CLR	0.789 ± 0.010	0.330 ± 0.011	0.073 ± 0.041	0.182 ± 0.050	0.233 ± 0.017	0.122 ± 0.011	0.484 ± 0.020
RAKEL	0.958 ± 0.011	0.335 ± 0.031	0.077 ± 0.014	0.332 ± 0.021	0.249 ± 0.006	0.161 ± 0.007	0.553 ± 0.016
Ranking loss ↓							
LLRD	0.185 ± 0.011	**0.144 ± 0.028**	0.018 ± 0.007	0.185 ± 0.022	**0.129 ± 0.010**	**0.042 ± 0.008**	**0.163 ± 0.008**
MLFE	0.188 ± 0.010	0.146 ± 0.030	0.014 ± 0.007	0.191 ± 0.025	0.134 ± 0.017	0.046 ± 0.010	0.167 ± 0.011
LIFT	**0.178 ± 0.008**	**0.144 ± 0.026**	0.029 ± 0.009	0.148 ± 0.014	0.148 ± 0.012	0.054 ± 0.015	0.164 ± 0.013
RELIAB	0.182 ± 0.007	0.165 ± 0.021	0.026 ± 0.008	**0.134 ± 0.011**	0.176 ± 0.008	0.076 ± 0.007	0.185 ± 0.021
ML^2	0.205 ± 0.021	0.153 ± 0.033	**0.011 ± 0.009**	0.194 ± 0.027	0.136 ± 0.012	0.050 ± 0.007	0.175 ± 0.015
CLR	0.231 ± 0.020	0.181 ± 0.020	0.072 ± 0.051	0.137 ± 0.028	0.241 ± 0.015	0.098 ± 0.021	0.196 ± 0.009
RAKEL	0.359 ± 0.012	0.213 ± 0.019	0.066 ± 0.019	0.281 ± 0.034	0.244 ± 0.016	0.155 ± 0.023	0.243 ± 0.010
Average precision ↑							
LLRD	**0.506 ± 0.018**	0.819 ± 0.031	**0.905 ± 0.020**	**0.421 ± 0.033**	**0.841 ± 0.009**	**0.934 ± 0.010**	**0.775 ± 0.008**
MLFE	0.490 ± 0.017	0.812 ± 0.032	0.901 ± 0.021	0.410 ± 0.029	0.835 ± 0.019	0.928 ± 0.013	0.766 ± 0.016
LIFT	0.502 ± 0.021	**0.824 ± 0.024**	0.880 ± 0.030	0.416 ± 0.031	0.820 ± 0.018	0.922 ± 0.008	0.768 ± 0.018
RELIAB	0.497 ± 0.016	0.801 ± 0.021	0.869 ± 0.020	0.405 ± 0.041	0.781 ± 0.009	0.851 ± 0.008	0.751 ± 0.010
ML^2	0.481 ± 0.030	0.816 ± 0.031	0.898 ± 0.022	0.404 ± 0.031	0.832 ± 0.014	0.930 ± 0.009	0.759 ± 0.020
CLR	0.425 ± 0.034	0.770 ± 0.019	0.695 ± 0.032	0.312 ± 0.059	0.722 ± 0.015	0.801 ± 0.012	0.755 ± 0.006
RAKEL	0.343 ± 0.009	0.772 ± 0.037	0.798 ± 0.018	0.228 ± 0.020	0.731 ± 0.017	0.777 ± 0.023	0.717 ± 0.007
Macroaveraging $F1$ ↑							
LLRD	0.231 ± 0.026	**0.676 ± 0.051**	**0.736 ± 0.050**	0.408 ± 0.028	**0.666 ± 0.024**	**0.800 ± 0.016**	0.420 ± 0.030
MLFE	0.239 ± 0.025	0.668 ± 0.050	0.702 ± 0.056	**0.415 ± 0.041**	0.655 ± 0.021	0.787 ± 0.015	0.430 ± 0.024
LIFT	0.179 ± 0.014	0.651 ± 0.035	0.694 ± 0.052	0.392 ± 0.045	0.624 ± 0.033	0.788 ± 0.018	0.377 ± 0.019
RELIAB	**0.288 ± 0.015**	0.639 ± 0.038	0.686 ± 0.058	0.394 ± 0.031	0.568 ± 0.030	0.671 ± 0.021	0.409 ± 0.023
ML^2	0.226 ± 0.024	0.656 ± 0.045	0.686 ± 0.058	0.382 ± 0.035	0.652 ± 0.018	0.783 ± 0.015	0.438 ± 0.017
CLR	0.220 ± 0.017	0.604 ± 0.032	0.616 ± 0.118	0.402 ± 0.056	0.523 ± 0.027	0.635 ± 0.013	0.386 ± 0.016
RAKEL	0.195 ± 0.010	0.615 ± 0.030	0.679 ± 0.037	0.377 ± 0.054	0.545 ± 0.018	0.654 ± 0.012	**0.441 ± 0.011**
Microaveraging $F1$ ↑							
LLRD	0.325 ± 0.011	**0.692 ± 0.048**	**0.814 ± 0.030**	0.126 ± 0.027	**0.665 ± 0.024**	**0.792 ± 0.017**	**0.656 ± 0.011**
MLFE	0.384 ± 0.017	0.683 ± 0.047	0.785 ± 0.031	0.137 ± 0.032	0.653 ± 0.024	0.781 ± 0.015	0.643 ± 0.013
LIFT	0.313 ± 0.013	0.664 ± 0.015	0.763 ± 0.031	0.168 ± 0.034	0.625 ± 0.031	0.779 ± 0.022	0.650 ± 0.016
RELIAB	**0.454 ± 0.011**	0.647 ± 0.038	0.748 ± 0.024	**0.188 ± 0.028**	0.562 ± 0.021	0.639 ± 0.013	0.631 ± 0.015
ML^2	0.366 ± 0.013	0.674 ± 0.042	0.780 ± 0.021	0.074 ± 0.031	0.650 ± 0.019	0.776 ± 0.018	0.635 ± 0.018
CLR	0.330 ± 0.012	0.626 ± 0.029	0.606 ± 0.143	0.165 ± 0.050	0.531 ± 0.008	0.634 ± 0.017	0.623 ± 0.010
RAKEL	0.356 ± 0.025	0.648 ± 0.024	0.669 ± 0.016	0.155 ± 0.019	0.533 ± 0.005	0.645 ± 0.009	0.637 ± 0.011

problem. We use the default settings recommended by RAKEL algorithm, namely, $k = 3$, ensemble size $n = 2q$. For ML^2, respective parameter values are recorded as follows: $\lambda = 1$, $K = l + 1$ and C_1 and C_2 selected from $\{1, 2, \ldots, 10\}$. ML^2 is the first multilabel learning algorithm to attempt to explore manifolds at the label level. CLR is a second-order problem transformation method. It solves the problem of multilabel classification by using label ranking, in which ranking among labels is implemented by pairwise comparison. The associated parameter ensemble size is set to $\binom{q}{2}$. LIFT uses different feature sets to distinguish different labels by clustering positive and negative examples. The value of ratio parameter r is 0.1, as suggested in [18]. RELIAB utilizes the implicit relative information of label to achieve the task of multilabel learning. The parameters τ and β take values from $\{0.1, 0.15, \ldots, 0.5\}$ and $\{0.001, 0.01, \ldots, 10\}$, respectively. For LLRD, $\eta = 1$, r can be selected from $\{1, 2, \ldots, q-1\}$. In a word, the parameter settings of the comparison algorithm are as recommended in the related papers.

Table 3: Performance of each multilabel algorithm (mean ± std. deviation) on the large-scale datasets.

Comparing algorithms	Slashdot	rcv1subset1	rcv1subset2	rcv1subset3	rcv1subset4	rcv1subset5
One-error ↓						
LLRD	**0.363 ± 0.026**	0.414 ± 0.013	0.411 ± 0.017	0.416 ± 0.029	**0.317 ± 0.015**	0.401 ± 0.018
MLFE	0.374 ± 0.027	0.406 ± 0.018	0.399 ± 0.013	0.402 ± 0.025	0.328 ± 0.013	0.392 ± 0.008
LIFT	0.393 ± 0.033	0.427 ± 0.011	0.434 ± 0.017	0.441 ± 0.020	0.363 ± 0.019	0.430 ± 0.019
RELIAB	0.508 ± 0.022	0.449 ± 0.015	0.458 ± 0.028	0.454 ± 0.012	0.433 ± 0.024	0.423 ± 0.009
ML^2	0.370 ± 0.025	**0.404 ± 0.017**	**0.395 ± 0.018**	**0.398 ± 0.021**	0.323 ± 0.021	**0.388 ± 0.010**
CLR	0.965 ± 0.013	0.513 ± 0.022	0.515 ± 0.009	0.518 ± 0.028	0.472 ± 0.031	0.521 ± 0.021
RAKEL	0.602 ± 0.009	0.605 ± 0.013	0.574 ± 0.012	0.585 ± 0.022	0.561 ± 0.022	0.614 ± 0.009
Coverage ↓						
LLRD	0.107 ± 0.010	**0.125 ± 0.008**	**0.121 ± 0.009**	**0.123 ± 0.006**	0.092 ± 0.004	**0.116 ± 0.009**
MLFE	0.126 ± 0.013	0.136 ± 0.005	0.130 ± 0.010	0.129 ± 0.007	0.094 ± 0.007	0.124 ± 0.007
LIFT	0.112 ± 0.008	0.144 ± 0.020	0.135 ± 0.008	0.156 ± 0.008	0.113 ± 0.012	0.148 ± 0.013
RELIAB	0.131 ± 0.007	0.152 ± 0.012	0.128 ± 0.014	0.144 ± 0.011	0.105 ± 0.020	0.131 ± 0.014
ML^2	**0.103 ± 0.011**	0.138 ± 0.008	0.132 ± 0.010	0.126 ± 0.006	**0.078 ± 0.006**	0.129 ± 0.009
CLR	0.254 ± 0.003	0.146 ± 0.018	0.141 ± 0.007	0.137 ± 0.010	0.109 ± 0.018	0.136 ± 0.011
RAKEL	0.226 ± 0.020	0.426 ± 0.023	0.372 ± 0.016	0.381 ± 0.014	0.365 ± 0.009	0.388 ± 0.020
Ranking loss ↓						
LLRD	**0.090 ± 0.010**	**0.049 ± 0.004**	**0.050 ± 0.004**	**0.052 ± 0.002**	0.038 ± 0.002	**0.047 ± 0.003**
MLFE	0.107 ± 0.013	0.052 ± 0.002	0.055 ± 0.007	0.055 ± 0.002	0.040 ± 0.004	0.050 ± 0.003
LIFT	0.098 ± 0.016	0.058 ± 0.007	0.057 ± 0.009	0.068 ± 0.004	0.059 ± 0.010	0.055 ± 0.007
RELIAB	0.124 ± 0.003	0.066 ± 0.010	0.063 ± 0.008	0.062 ± 0.004	0.052 ± 0.006	0.063 ± 0.005
ML^2	0.103 ± 0.012	0.056 ± 0.004	0.057 ± 0.004	0.056 ± 0.003	**0.031 ± 0.003**	0.050 ± 0.004
CLR	0.237 ± 0.008	0.062 ± 0.011	0.066 ± 0.008	0.065 ± 0.012	0.047 ± 0.006	0.071 ± 0.005
RAKEL	0.211 ± 0.019	0.226 ± 0.019	0.215 ± 0.017	0.230 ± 0.015	0.235 ± 0.014	0.214 ± 0.016
Average precision ↑						
LLRD	**0.725 ± 0.019**	0.611 ± 0.010	0.638 ± 0.011	0.634 ± 0.017	**0.717 ± 0.008**	0.643 ± 0.011
MLFE	0.712 ± 0.021	0.618 ± 0.016	0.645 ± 0.009	0.639 ± 0.014	0.708 ± 0.012	0.647 ± 0.012
LIFT	0.703 ± 0.010	0.586 ± 0.009	0.598 ± 0.012	0.595 ± 0.011	0.674 ± 0.013	0.598 ± 0.011
RELIAB	0.624 ± 0.014	0.578 ± 0.021	0.611 ± 0.011	0.614 ± 0.018	0.655 ± 0.018	0.604 ± 0.009
ML^2	0.715 ± 0.022	**0.621 ± 0.012**	**0.647 ± 0.013**	**0.643 ± 0.016**	**0.717 ± 0.013**	**0.650 ± 0.010**
CLR	0.269 ± 0.002	0.575 ± 0.013	0.584 ± 0.021	0.571 ± 0.032	0.614 ± 0.020	0.588 ± 0.013
RAKEL	0.522 ± 0.020	0.395 ± 0.012	0.445 ± 0.018	0.431 ± 0.014	0.450 ± 0.012	0.437 ± 0.016
Macroaveraging $F1$ ↑						
LLRD	0.427 ± 0.035	0.235 ± 0.020	0.259 ± 0.019	0.213 ± 0.031	0.300 ± 0.019	0.211 ± 0.020
MLFE	0.466 ± 0.035	0.198 ± 0.017	0.195 ± 0.056	0.202 ± 0.030	0.249 ± 0.021	0.204 ± 0.021
LIFT	0.429 ± 0.037	0.223 ± 0.025	0.186 ± 0.024	0.200 ± 0.031	0.238 ± 0.013	0.196 ± 0.031
RELIAB	0.425 ± 0.029	**0.342 ± 0.022**	**0.338 ± 0.016**	**0.348 ± 0.014**	**0.342 ± 0.028**	**0.352 ± 0.014**
ML^2	**0.472 ± 0.029**	0.216 ± 0.020	0.206 ± 0.024	0.195 ± 0.030	0.244 ± 0.023	0.208 ± 0.011
CLR	0.174 ± 0.032	0.285 ± 0.032	0.264 ± 0.021	0.272 ± 0.022	0.311 ± 0.031	0.305 ± 0.017
RAKEL	0.354 ± 0.037	0.269 ± 0.030	0.251 ± 0.014	0.255 ± 0.014	0.263 ± 0.014	0.274 ± 0.018
Microaveraging $F1$ ↑						
LLRD	0.496 ± 0.021	0.393 ± 0.013	0.381 ± 0.017	0.406 ± 0.027	0.470 ± 0.013	0.402 ± 0.018
MLFE	0.545 ± 0.019	0.373 ± 0.014	0.375 ± 0.031	0.392 ± 0.024	0.403 ± 0.020	0.381 ± 0.017
LIFT	0.510 ± 0.030	0.320 ± 0.017	0.353 ± 0.014	0.347 ± 0.018	0.342 ± 0.024	0.363 ± 0.008
RELIAB	0.453 ± 0.011	**0.408 ± 0.010**	**0.449 ± 0.008**	**0.451 ± 0.021**	**0.478 ± 0.016**	**0.454 ± 0.012**
ML^2	**0.556 ± 0.022**	0.371 ± 0.014	0.391 ± 0.010	0.383 ± 0.026	0.393 ± 0.022	0.410 ± 0.015
CLR	0.104 ± 0.005	0.367 ± 0.011	0.368 ± 0.024	0.320 ± 0.024	0.381 ± 0.015	0.372 ± 0.008
RAKEL	0.365 ± 0.020	0.359 ± 0.023	0.348 ± 0.016	0.341 ± 0.016	0.371 ± 0.015	0.342 ± 0.006

3.2. Experimental Results. For each dataset in our experiment, we adopt the tenfold cross-validation strategy. Our experimental results are mainly distributed in Tables 2 and 3, where we record the performance of different algorithms in different multilabel datasets. Specifically, the average and standard deviation of the corresponding evaluation criteria are recorded in the tables. For each evaluation metric, "↓" indicates "the smaller the better" and "↑" indicates "the larger the better". The best results are shown in bold form.

Table 4: The Friedman statistics F_F and the critical value.

Evaluation metric	F_F	Critical value
One-error	34.0909	2.2274
Coverage	20.3765	
Ranking loss	21.1642	
Average precision	39.8409	
Macroaveraging F1	2.6520	
Microaveraging F1	7.6088	

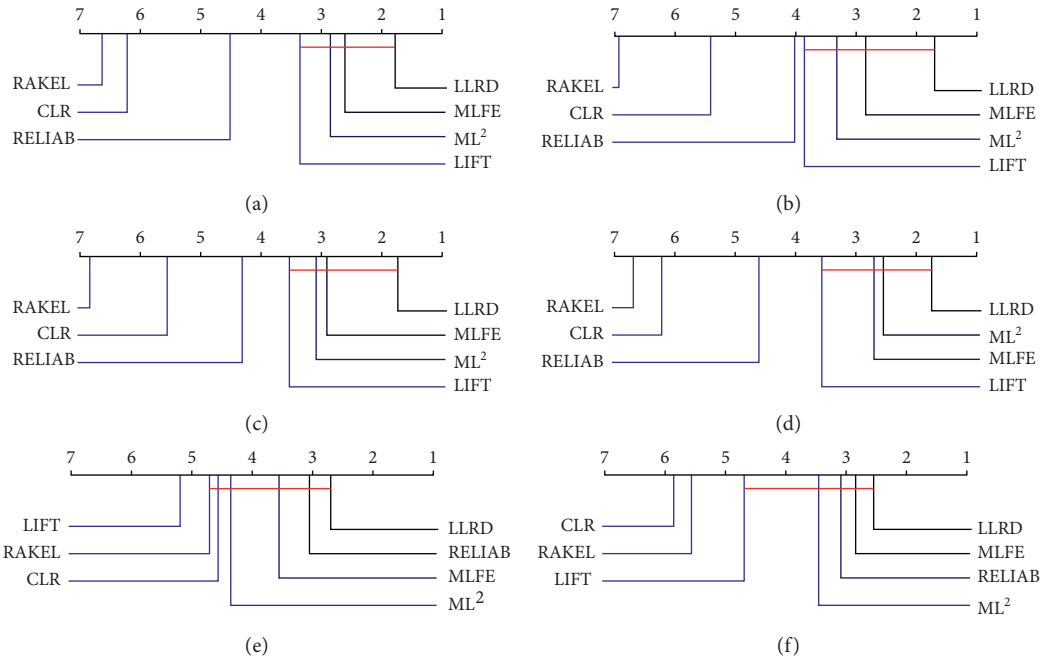

Figure 1: Comparison of LLRD (control algorithm) against other related approaches with the Bonferroni–Dunn test. Approaches that are not connected to LLRD are significantly different in performance from LLRD. (a) One-error. (b) Coverage. (c) Ranking loss. (d) Average precision. (e) Macroaveraging $F1$. (f) Microaveraging $F2$.

We use *Friedman test* [38] based on the average ranks for verifying whether the difference between algorithms is statistically significant. If the assumption that "all algorithms have equal performance" is rejected, it means that the performance of each algorithm is significantly different. As can be seen from the data presented in Table 4, the hypothesis that there is no significant difference among the algorithms is not valid under the condition of 0.05 significance level. Therefore, we need to conduct a post hoc test to further distinguish the various algorithms. Usually, there are two options for post hoc test, one is the Nemenyi test [38] and the other is the Bonferroni–Dunn test [39]. For k algorithms, the former needs to compare $k(k-1)/2$ times, while the latter only needs $k-1$ times in some cases. Thus, we choose the latter. The Bonferroni–Dunn test is used to test whether LLRD is more competitive than the comparative algorithm, in which LLRD plays a role of control algorithm. When the difference of average rank between two algorithms is more than one critical difference CD, the performance of two algorithms is obviously different. The CD value mentioned here can be calculated from $CD = q_\alpha \sqrt{k(k+1)/6N}$, where $k = 7$ and $N = 13$, when the significance level is 0.05, the corresponding $q_\alpha = 2.638$.

The CD diagram associated with LLRD and its comparison algorithm is shown in Figure 1. The numbers on the horizontal axis of the coordinate indicate the average rank value of each algorithm under different evaluation criteria. There is no significant difference in performance among the various algorithms connected by solid lines.

Through the analysis of the above experimental results, we can draw the following conclusions:

(1) In terms of the four evaluation criteria of *one-error*, *coverage*, *ranking loss*, and *average precision*, LLRD is obviously superior to RELIAB, RAKEL, and CLR.

(2) The smaller the average rank value, the better the performance of the corresponding. For LLRD, five of the average rank value in the six CD subdiagrams are optimal, which shows LLRD outperforms other algorithms.

(3) For regular-size datasets, LLRD ranks first in 69% of the cases under different evaluation criteria, while for large-scale datasets, it ranks first in 36.1%.

4. Conclusions

In this work, we propose a novel multilabel classification algorithm named LLRD, which adopts the low-rank decomposition to gain the internal information of label and further reduce the information loss of the label transformation via the new feature space. Experimental results show that the performance of the proposed LLRD is better than many state-of-the-art multilabel classification techniques. In the future, we will explore alternative models combining the low-rank decomposition and classification into a joint optimization problem for considering more complex correlation of labels.

Data Availability

The datasets used in our manuscript are all public datasets, which can be downloaded from "http://mulan.sourceforge.net/datasets.html" and "http://meka.sourceforge.net/#datasetsru".

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key R&D Program of China (2019YFC1521400), National Natural Science Foundation of China (61806159 and 61806160, 61972312), and China Postdoctoral Science Foundation (2018M631192).

References

[1] L. Tang, S. Rajan, and V. K. Narayanan, "Large scale multi-label classification via metalabeler," in *Proceedings of the 19th International Conference on World Wide Web*, pp. 211–220, Madrid, Spain, July 2009.

[2] B. Yang, J. T. Sun, T. Wang, and Z. Chen, "Effective multi-label active learning for text classification," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 917–926, Paris, France, July 2009.

[3] H. Kazawa, T. Izumitani, H. Taira, and E. Maeda, "Maximal margin labeling for multi-topic text categorization," in *Advances In Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds., pp. 649–656, MIT Press, Cambridge, MA, USA, 2005.

[4] A. McCallum, "Multi-label text classification with a mixture model trained by EM," in *Proceedings of the Working Notes of the AAAI Workshop on Learning for Text Categorization*, Orlando, FL, USA, July 1999.

[5] R. E. Schapire, Y. Singer, and Boostexter, "A boosting-based system for text categorization," *Machine Learning*, vol. 39, no. 2-3, pp. 135–168, 2000.

[6] N. Ueda and K. Saito, "Parametric mixture models for multi-label text," in *Advances In Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds., pp. 721–728, MIT Press, Cambridge, MA, USA, 2003.

[7] F. Kang, R. Jin, and R. Sukthankar, "Correlated label propagation with application to multi-label learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1719–1726, New York, NY, USA, June 2006.

[8] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, "Multilabel classification of music into emotions," in *Proceedings of the 9th I International Conference on Music Information Retrieval*, pp. 325–330, Philadephia, PA, USA, 2008.

[9] C. Sanden and J. Z. Zhang, "Enhancing multi-label music genre classification through ensemble techniques," in *Proceedings of the 34th SIGIR*, pp. 705–714, Beijing, China, July 2011.

[10] R. C. Hong, M. Wang, Y. Gao et al., "Image annotation by multiple-instance learning with discriminative feature mapping and selection," *IEEE Transactions on Cybernetics*, vol. 44, no. 5, pp. 669–680, 2014.

[11] Y. Xia, L. Nie, L. Zhang, Y. Yang, R. Hong, and X. Li, "Weakly supervised multilabel clustering and its applications in computer vision," *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 3220–3232, 2016.

[12] Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in *Advances In Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds., pp. 681–687, MIT Press, Cambridge, MA, USA, 2002.

[13] M. L. Zhang and Z. H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 10, pp. 1338–1351, 2006.

[14] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," *Data Mining and Knowledge Discovery Handbook*, Springer pp. 667–685, 2010.

[15] E. Gibaja and S. Ventura, "A tutorial on multilabel learning," *ACM Computing Surveys*, vol. 47, no. 3, pp. 1–38, 2015.

[16] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: a lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[17] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.

[18] M.-L. Zhang and L. Wu, "Lift: multi-label learning with label-specific features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 1, pp. 107–120, 2015.

[19] Y. H. Guo and S. C. Gu, "Multi-label classification using conditional dependency networks," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, Catalonia, Spain, July 2011.

[20] M. L. Zhang and K. Zhang, "Multi label learning by exploiting label dependency," in *Proceedings Of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 999–1008, Washington, DC, USA, 2010.

[21] B. Fu, G. D. Xu, Z. H. Wang, and L. B. Cao, "Leveraging supervised label dependency propagation for multi-label learning," in *Proceedings Of the IEEE 13th International Conference on Data Mining*, pp. 1061–1066, Dallas, TX, USA, December 2013.

[22] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," in *In Lecture Notes in Artificial Intelligence 5782, W*, M. G. Buntine and J. Shawe-Taylor, Eds., pp. 254–269, Springer, Berlin, Germany, 2009.

[23] J. Huang, G. R. Li, S. H. Wang, and Q. M. Huang, "Categorizing social multimedia by neighborhood decision using local pairwise label correlation," in *Proceedings Of the IEEE International Conference on Data Mining Workshop*, pp. 913–920, Shenzhen, China, December 2014.

[24] H. W. Liu, Z. J. Ma, S. C. Zhang, and X. D. Wu, "Penalized partial least square discriminant analysis with ℓ1-norm for multi-label data," *Pattern Recognition*, vol. 48, no. 5, pp. 1724–1733, 2015.

[25] S. J. Huang and Z. H. Zhou, "Multi-label learning by exploiting label correlations locally," in *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pp. 949–955, Toronto, Canada, July 2012.

[26] J. Ling, J. D. Li, and H. Liu, "Exploiting multilabel information for noise-resilient feature selection," *ACM Transactions on Intelligent Systems and Technology*, vol. 9, no. 5, pp. 1–23, 2018.

[27] C. K. Yeh, W. C. Wu, W. J. Ko, and Y. C. F. Wang, "Learning deep latent space for multi-label classification," in *Proceedings Of the 31st AAAI Conference On Artificial Intelligence*, pp. 2838–2844, San Francisco, CA, USA, February 2017.

[28] K.-H. Huang and H.-T. Lin, "Cost-sensitive label embedding for multi-label classification," *Machine Learning*, vol. 106, no. 9-10, pp. 1725–1746, 2017.

[29] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2008.

[30] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 644–658, 2014.

[31] D. Tuia, J. Verrelst, L. Alonso, F. Perez-Cruz, and G. Camps-Valls, "Multioutput support vector regression for remote sensing biophysical parameter estimation," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 804–808, 2011.

[32] M. L. Zhang and Z. H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2013.

[33] Q. W. Zhang, Y. Zhong, and M. L. Zhang, "Feature-induced labeling information enrichment for multi-label learning," in *Proceedings of the IEEE International Conference on Artificial Intelligence*, pp. 4446–4453, Taichung, Taiwan, January 2018.

[34] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-labelsets for multilabel classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 7, pp. 1079–1089, 2010.

[35] P. Hou, X. Geng, and M. L. Zhang, "Multi-label manifold learning," in *Proceedings of the 30th AAAI Conference On Artificial Intelligence*, Phoenix, AZ, USA, February 2016.

[36] J. Fürnkranz, E. Hüllermeier, E. Loza Mencía, and K. Brinker, "Multilabel classification via calibrated label ranking," *Machine Learning*, vol. 73, no. 2, pp. 133–153, 2008.

[37] Y. K. Li, M. L. Zhang, and X. Geng, "Leveraging implicit relative labeling-importance information for effective multi-label learning," in *Proceedings Of 15th IEEE International Conference On Data Mining*, pp. 251–260, Atlantic City, NJ, USA, November 2015.

[38] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[39] A. Benavoli, G. Corani, and F. Mangili, "Should we really use post-hoc tests based on mean-ranks?" *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 152–161, 2016.