Adaptive Weighted Ranking-Oriented Label Distribution Learning

Xiuyi Jia, *Member, IEEE*, Tian Qin, Yunan Lu, and Weiwei Li

***Abstract*— Label distribution learning (LDL) is a novel machine-learning paradigm generalized from multilabel learning (MLL). LDL attaches a label distribution to each instance, giving the description degree of different labels. In many real-world applications, key labels, that is, labels with relatively higher description degrees, are preferable to be better predicted. Unfortunately, existing LDL metrics measure the distance or similarity between label distributions from a global perspective, failing to give sufficient attention to key labels. Therefore, we design a novel LDL metric, the description-degree percentile average (DPA), which simultaneously integrates both the exact ranking value and the description degree of each label. The DPA can enhance accuracy in predicting key labels. Furthermore, noting the shape characteristics of the label distributions, we minimize the variance distance between the predicted and the ground-truth label distributions, to better maintain the distinguishability of labels. Finally, we propose an adaptive weighted ranking-oriented LDL algorithm, which is more suitable for realistic LDL problems that require higher accuracy in predicting key labels. We conduct extensive comparison experiments on various types of LDL datasets. Experimental results on both traditional and newly introduced metrics demonstrate the effectiveness of our proposal.**

***Index Terms*— Evaluation measures, key labels, label distribution learning (LDL), label ranking.**

I. Introduction

LABEL distribution learning (LDL) [1] is a novel machine-learning paradigm to address label polysemy. Unlike multilabel learning (MLL) [2], LDL not only can associate a certain number of labels to each instance, but also can give the relative importance of each label to a particular instance. Such relative importance is called the label description degree, which allows the LDL paradigm to provide richer semantic information. As shown in Fig. 1, when describing a facial expression image, LDL assigns a label distribution to each instance. In addition to the description degree of each expression, the label distribution implies the relevance intensities of different expression labels in qualitative terms, that is, label ranking relation [3].

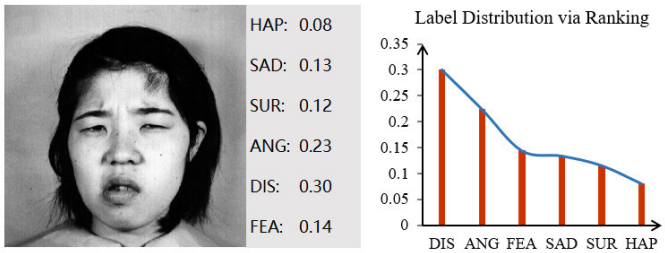

Fig. 1. Example of facial expression recognition, illustrating the fact that the label distribution implies label ranking relation by attaching the relative importance to each label.

Since Jia et al. [4] first explicitly considered the correlation between labels in LDL models, successive works inspired by various label correlations have been proposed, all with satisfactory empirical results [5], [6], [7], [8]. Similar to various label correlations, the label ranking relation is also a reflection of the relationship between labels. The label ranking relation is a qualitative representation of label distribution, which is crucial for many real-world applications. For example, a book or a movie is usually attached to several labels, but only the first few are shown on recommendation sites, such as Douban and YouTube. Similarly, when product labels that are top ranked in terms of user interest are extracted, shopping software, such as eBay and Amazon, will rely on these labels to recommend corresponding products to users. That is, a good prediction of the label ranking relation can help extract the key labels needed for realistic applications.

Intuitively, there are two ways to extract such key labels by using LDL. One is to transform LDL into a classification problem and obtain the corresponding class labels. The other is to maintain the ranking relation during the LDL and then extract the top labels. For the first solution, LDL4C [9] noticed the mismatch between the objective of LDL and that of classification and tried to mitigate such mismatch by reweighing with respect to information entropy. Subsequently, Wang and Geng [10] proposed a general LDL machine (LDLM) method with a stronger theoretical support for this issue. LDLM considers the margin between the top-2 labels and other labels during prediction; consequently, LDLM

Manuscript received 18 May 2022; revised 4 November 2022; accepted 13 March 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1706900; and in part by the National Natural Science Foundation of China under Grant 62176123, Grant 61906090, and Grant U20B2064. The work of Xiuyi Jia was supported by Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications. *(Corresponding author: Weiwei Li.)*

Xiuyi Jia, Tian Qin, and Yunan Lu are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: jiaxy@njust.edu.cn; qintian@njust.edu.cn; luyn@njust.edu.cn).

Weiwei Li is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: liweiwei@nuaa.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2023.3258976.

Digital Object Identifier 10.1109/TNNLS.2023.3258976

2162-237X © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

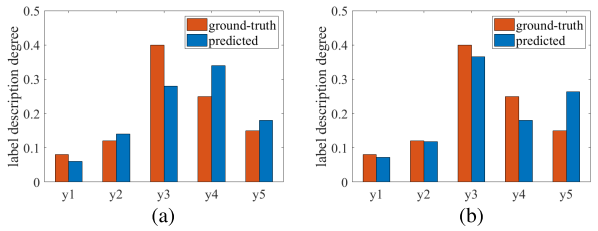

Fig. 2. Example of the fact that DPA can reflect different impacts of predicting the wrong ranking position of different labels. The red and blue bars are the ground-truth and the predicted label distributions, respectively. Fig. 2(a) and (b) shares the same Kullback–Leibler (K–L) divergence and Spearman's rank correlation while Fig. 2(b) is better measured by DPA. (a) DPA = 0.724, K–L = 0.043, Spearman = 0.9. (b) DPA = 0.734, K–L = 0.043, Spearman = 0.9.

may achieve relatively poor results when the number of key labels exceeds 2. However, the above works only validated the classification effect of their models without fitting the original label distributions well. For the second solution, to maintain the label ranking relation, LDL-LRR [3] added a ranking loss term to the objective function. Inspired by the pairwise ranking method RankNet [11], LDL-LRR considered the number of ranking errors in all label pairs and introduced Spearman's rank correlation [12] and Kendall tau as ranking metrics. However, in practice, the impacts of predicting the wrong ranking value vary from label to label. LDL-LRR aims to optimize the overall label ranking, failing to give any preference to key labels which may be crucial in practice. Additionally, LDL-LRR considers all labels to be key labels and assigns the same weight.

Therefore, we aim to design a method that is more suitable for LDL problems preferable to key labels, considering all possible key labels and giving them higher weights in the learning process. Inspired by the fault percentile average (FPA) [13], an evaluation metric for cases where fewer modules have more defects in software defect prediction tasks, we design a novel metric, namely, the description-degree percentile average (DPA) for ranking-oriented LDL, which can measure the prediction effect on key labels with higher weights from the ranking perspective. Unlike existing LDL metrics and ranking-oriented metrics, DPA integrates the ground-truth description degree and the predicted ranking of each label, giving preference to key labels in measurement. As shown in Fig. 2, the two predicted label distributions in Fig. 2(a) and (b) share the same K–L divergence and Spearman's rank correlation. However, the prediction in Fig. 2(b) is superior to that in Fig. 2(a) when the highest label plays a decisive role. DPA, whose value is greater in Fig. 2(b), can reflect such strength. Relying on the proposed DPA, we design a novel regularization term to characterize the property that the labels should have different weights due to different rankings and combine the term with the K–L divergence, hoping to improve the prediction of key labels while fitting original label distributions.

In addition, we found that existing LDL algorithms, especially those based on minimizing/maximizing the distance/similarity of two distributions, inevitably predict the label distribution as an average distribution during learning, possibly yielding barely satisfactory results in terms of

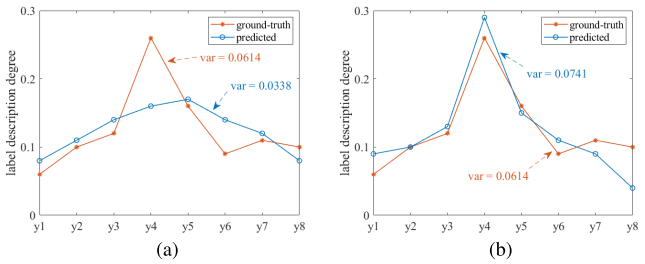

Fig. 3. Example to visualize the fact that the label distribution with the variance closer to the ground-truth predicts key labels better. The red and blue lines are the ground-truth and the predicted label distributions, respectively. $\mathrm{Dis}_{\mathrm{var}}$ is the distance of variances of two label distributions. (a) K–L = 0.044, $\mathrm{Dis}_{\mathrm{var}}$ = 0.0276. (b) K–L = 0.044, $\mathrm{Dis}_{\mathrm{var}}$ = 0.0127.

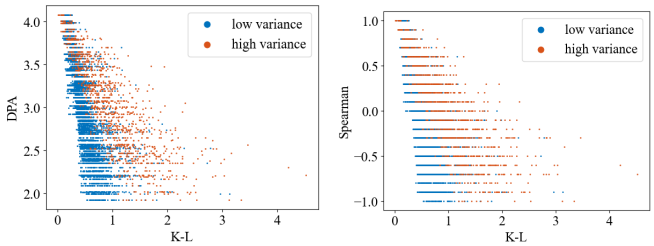

Fig. 4. Relationship between K–L loss and ranking quality (i.e., DPA and Spearman's rank correlation) in the cases of low and high variances. If the variance of a point ranks in the top 80 percent of the 10 000 points in ascending order, we classify it as low variance, otherwise high variance.

common metrics. However, the distinguishability among labels is consequently weakened, and extracting key labels becomes difficult. On further analysis, this distinguishability among labels can be reflected, to some extent, by the shape of the label distribution, and the variance of the label distribution can quantify the shape of the label distribution very well. As shown in Fig. 3, the large variance reflects the large variability among the labels. Although the predicted distributions in Fig. 3(a) and (b) have no difference in K–L divergence, the variance of Fig. 3(b) is closer to the ground truth, and the key labels are more obvious. Such prediction results are more favorable for us to extract the key labels. Moreover, the label distinguishability is somehow beneficial for more accurate top-k prediction, since a ranking error in a high variance label distribution will lead to a greater K–L loss. As shown in Fig. 4, there are 10 000 points in each figure, and each point represents a randomly generated label distribution. Besides, we randomly generated 30 ground-truth label distributions, and the x- and y-axes corresponding to each point are the mean of the K–L divergence and ranking metric values between this point and the 30 ground-truth label distributions, respectively. It can be seen that for label distributions with low variance, large changes in DPA or Spearman's rank correlation cause only small K–L fluctuations, that is, K–L is not sensitive to ranking errors with respect to low-variance label distributions. In contrast, K–L is sensitive to ranking errors for label distributions with high variance. Therefore, we attempt to maintain the distinctness by minimizing the distance between the variances of the predicted and the ground-truth label distributions.

In summary, to better extract key labels while predicting the description degree of each label as accurately as possible, we propose a novel adaptive weighted ranking-oriented LDL

method, LDL_DPA. Using the K–L divergence as the basis loss, we transform the DPA into a more optimization-friendly loss term to improve the ranking performance of the algorithm. In addition, we add a variance distance term to maintain the distinguishability of labels by reducing the variance distance between the predicted and the ground-truth distributions.

Our main contribution is threefold.

1) We design a novel metric, the DPA, for measuring the association of label ranking and weights. In addition to measuring the performance of LDL algorithms in maintaining label ranking, the DPA can also reflect the key label prediction accuracy.
2) We discuss a kind of LDL problem oriented to different ranked labels having different weights and propose an adaptive weighted ranking-oriented LDL algorithm (LDL_DPA), by which we can predict the description degree of each label while preferring the key labels.
3) We investigate the relationship between label distinguishability and the shape of the label distribution and provide a quantitative method of variance to portray the label distribution shape. Our proposed LDL_DPA, which can help maintain label distinguishability, has significant advantages over both traditional LDL evaluation metrics, learning-to-rank metrics, and the DPA.

II. Related Works

To estimate the age distribution, Geng et al. [14] first proposed LDL and then formally extended it to a machine-learning paradigm [1]. In recent years, much effort has been devoted to enhancing theoretical LDL models.

Among these works, several methods have achieved satisfactory results by applying label correlation. LDL-LCLR [7] captured the global label correlations based on low-rank approximation. Wang and Geng [15] encoded label correlations with the basic manifold structure of label distributions, which exploited label correlations by learning the manifold for different sample clusters. Qian et al. [16] proposed a feature selection method for LDL that uses neighborhood granularity to mine feature similarity while generating label correlations with correlation coefficients. Noting the label ranking relation implied in LDL, Jia et al. [3] introduced a ranking loss term to the objective function, thereby improving prediction results on both traditional and ranking metrics. In fact, the ranking relation can also be viewed as a form of label correlation. According to extensive experimental results, LDL models optimized with the distance metric alone often cannot compare with those taking advantage of label correlations.

The optimization objectives of the original LDL algorithms are based on evaluation metrics, with SA-IIS and SA-BFGS [1] aiming to minimize the K–L divergence between the predicted and ground-truth label distributions. LALOT [17] exploited optimal transport theory and designed the OT distance to capture the geometric information of the label space. Geng and Xia [18] selected the Jeffrey divergence as the optimization objective in head pose estimation. EDL [19] used the weighted J-divergence as the loss function for emotion distribution learning. Zhou et al. [20] defined a balanced divergence distance that maintains symmetry. Noting that the commonly used K–L divergence does not capture the similarity between age labels, Akbari et al. [5] designed the γ-Wasserstein loss based on the specific geometric structure of the age label space to increase the impact of highly correlated ages. The metrics used in the above work calculate the prediction error for each label and then sum it up. However, the label distribution of a sample can be considered as a whole. To address the situation where the prediction is overconfident and the entropy of the label distribution is much smaller than the ground truth, Zhang et al. [21] used two smoothing measures to reduce the difference in entropy. This work exploits entropy to describe the entire label distribution and optimize the prediction effect. Wen et al. [22] proposed an adaptive variance method in age distribution estimation, which adaptively adjusts the variance of the distribution for different samples. This work reflects the idea of using variance to characterize label distributions.

III. DPA Metric

In software defect prediction, a few software modules tend to contain most of the defects. To help allocate testing resources efficiently, modules with more defects should receive more attention. Therefore, the FPA is proposed as a metric to measure the ability of the algorithm to predict key modules. Inspired by the FPA, we design a novel metric, the DPA, for the ranking-oriented LDL problem to evaluate the prediction effect of LDL models for key labels.

A. Preliminaries

Let $\mathcal{X} = \mathbb{R}^m$ denote the m-dimensional feature space, and $\mathcal{D} = \{d^{y_1}, d^{y_2}, \ldots, d^{y_c}\}$ denote the label space, where c is the number of labels and $\mathcal{Y} = \{y_1, y_2, \ldots, y_c\}$ is the finite label set. Given a training set $T = \{(\mathbf{x}_1, \mathbf{d}_1), (\mathbf{x}_2, \mathbf{d}_2), \ldots, (\mathbf{x}_n, \mathbf{d}_n)\}$, where $\mathbf{x}_i = \{x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(m)}\} \in \mathcal{X}$ is the feature vector of the ith instance and $\mathbf{d}_i = \{d_i^{y_1}, d_i^{y_2}, \ldots, d_i^{y_c}\}$ is the label distribution associated with $\mathbf{x}_i$. $d_i^{y_j}$ is called the description degree of label y_j with respect to $\mathbf{x}_i$ and satisfies $\sum_j d_{x_i}^{y_j} = 1$, $d_{x_i}^{y_j} \geq 0$. The main goal of LDL is to learn a mapping $F : \mathcal{X} \to \mathcal{D}$ from the training set T for predicting the corresponding label distribution of input instances.

B. Formula

Before defining the DPA, we first briefly introduce the FPA. The FPA is used to measure the ranking performance of software defect prediction models. For the given c software modules, we first need to rank these modules in ascending order of the defect number they are predicted to contain. The percentage of real defects in the predicted top k modules is $(1/n)\sum_{i=c-k+1}^{c} n_i$, where n_i is the number of real defects in the module ranked i in the above-predicted order, and n is the total number of real defects. Then FPA is defined as follows:

$$\mathrm{FPA} = \frac{1}{c}\sum_{k=1}^{c}\frac{1}{n}\sum_{i=c-k+1}^{c} n_i. \tag{1}$$

Briefly, FPA is the average value of the percentage of real defects obtained when k takes different values from 1 to c. A larger FPA means a better-predicted ranking.

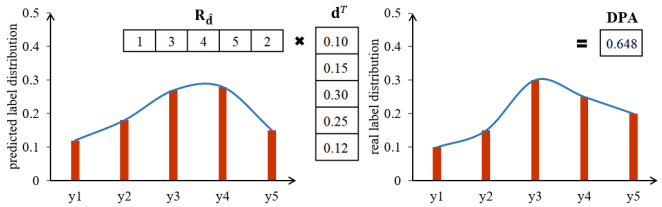

Fig. 5. Example to the illustration of the calculation process of DPA.

Inspired by the FPA, we can design a novel metric DPA for LDL to measure the ranking of the importance of each label in the label distribution. First, since the sum of the label distribution for each instance equals 1, we set n in (1) to the constant 1. The sum of the ground-truth description degrees of the labels predicted to rank in the top k is $\sum_{j=c-k+1}^{c} d_i^{y_j}$. Then, the DPA can be defined as follows:

$$\mathrm{DPA}\left(\mathbf{d}_i, \widehat{\mathbf{d}}_i\right) = \frac{1}{c} \sum_{k=1}^{c} \sum_{j=c-k+1}^{c} d_i^{y_j} \tag{2}$$

where c is the number of labels and $\mathbf{d}_i$ and $\widehat{\mathbf{d}}_i$ are the ground-truth and predicted label distributions, respectively. After exchanging the summation order, the DPA can be written as follows:

$$\begin{aligned}\mathrm{DPA}\left(\mathbf{d}_i, \widehat{\mathbf{d}}_i\right) &= \frac{1}{c} \sum_{j=1}^{c} \sum_{k=c-j+1}^{c} d_i^{y_j} \\ &= \frac{1}{c} \sum_{j=1}^{c} d_i^{y_j} \sum_{k=c-j+1}^{c} 1 \\ &= \frac{1}{c} \sum_{j=1}^{c} j * d_i^{y_j}.\end{aligned} \tag{3}$$

The weighted summation is carried out after ranking by predicted label distributions. Applying matrix operations, DPA can be obtained by the following equation:

$$\mathrm{DPA}\left(\mathbf{d}_i, \widehat{\mathbf{d}}_i\right) = \frac{1}{c} \mathbf{R}_{\widehat{\mathbf{d}}_i} \cdot \mathbf{d}_i{}^T \tag{4}$$

where $\mathbf{R}_{\widehat{\mathbf{d}}_i}$ is the ranking matrix corresponding to the predicted label distribution $\widehat{\mathbf{d}}_i$. The formula can be interpreted as the mean of the ground-truth description degrees weighted by the predicted label ranking. If the label distribution is predicted well on ranking, key labels with larger description degrees will receive preference and the DPA obtained is larger. Two or more labels with the same predicted description degrees should be weighted equally. For example, if y_{j_1} and y_{j_2} are predicted to be the same, ranked third and fourth, they will both be weighted with $\mathbf{R}_{\widehat{\mathbf{d}}_i}^{j_1} = \mathbf{R}_{\widehat{\mathbf{d}}_i}^{j_2} = (3+4)/2 = 3.5$.

From (2), DPA is the average of the description degrees obtained when k takes from 1 to c. In real-world applications, the number of key labels of the LDL problem can be interpreted to be any number from 1 to c. Fig. 5 is an intuitive example to illustrate the calculation process of the DPA.

C. Mathematical Properties

The CLC is the area under the curve of the Alberg plot [23], and like FPA, is an evaluation metric used to measure the prediction effect of software module defects. The FPA is chosen over the CLC because the former has a simpler and easier-to-understand calculation method. There is a linear relationship between the FPA and the CLC as follows [13]:

label	GT	pred-1	pred-2
1	0.1	0.08	0.12
2	0.15	0.13	0.09
3	0.3	0.35	0.25
4	0.25	0.24	0.28
5	0.2	0.22	0.26

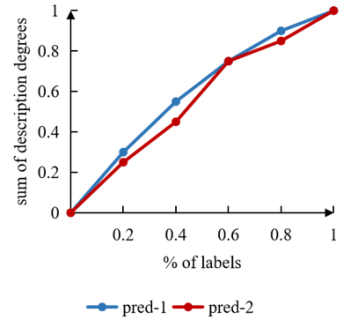

Fig. 6. Example of the Alberg plot for LDL. The red and blue curves represent two different predictions.

$$\mathrm{FPA} = \mathrm{CLC} + \frac{1}{2c}. \tag{5}$$

We can also refer to the Alberg plot in the LDL problem. First, labels are ranked according to the predicted distributions. Second, we calculate the sum of the description degrees of the top k labels. Use the percentage of the top k labels as the horizontal axis and the sum of the description degrees as the vertical axis, where k ranges from 1 to c. Finally, the corresponding points are drawn and connected. The larger the area under the curve is, the better the prediction. Fig. 6 is an example to illustrate the drawing process. Then, the DPA is constantly equal to the area under the curve plus $1/2c$ (c is the number of labels). Apparently, the range of the area under the curve is $(0, 1)$, and the range of the DPA is $((1/2c), 1 + (1/2c))$. Since $c \geq 2$ in LDL problems, the maximum range of the DPA is $(0, (5/4))$.

IV. LDL_DPA Approach

To better solve LDL problems preferable to key labels, we propose an adaptive weighted ranking-oriented LDL approach, LDL_DPA. We modify the DPA as a regularization term of the objective function, hoping to improve key label prediction while fitting original label distributions. In addition, to maintain the distinguishability of labels during learning, we add a variance distance minimization term.

A. Framework

From among the various metrics for measuring the distance between two distributions [24], we select the K–L divergence as the basic loss term for its stability in LDL experiments [25]

$$\mathrm{Dis}_{\mathrm{K\text{–}L}}\left(\mathbf{d}_i, \widehat{\mathbf{d}}_i\right) = \sum_j d_i^{y_j} \ln \frac{d_i^{y_j}}{\widehat{d_i^{y_j}}}. \tag{6}$$

Similar to previous works [1], we assume that the prediction model follows the maximum entropy model [26]. Then the predicted label distribution $\widehat{\mathbf{d}}_i = \{\widehat{d_i^{y_1}}, \widehat{d_i^{y_2}}, \ldots, \widehat{d_i^{y_c}}\}$ can be

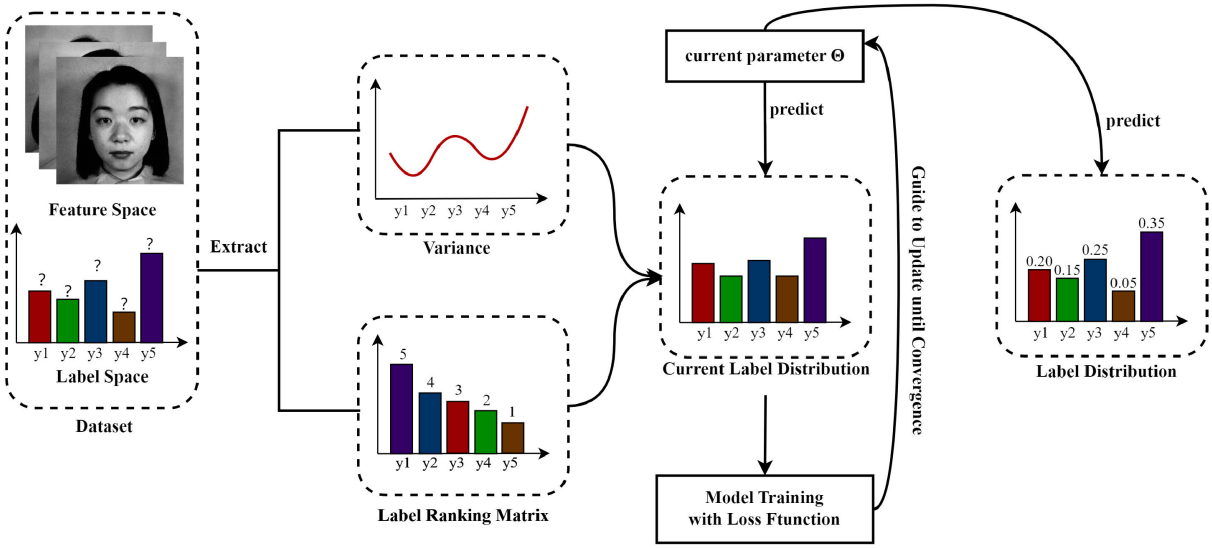

Fig. 7. Overall framework of LDL_DPA.

obtained by the following equation:

$$p\left(\widehat{d}_i^{y_j} \mid \mathbf{x}_i; \Theta\right) = \frac{\exp\left(\sum_r \Theta_{jr} x_i^{(r)}\right)}{\sum_k \exp\left(\sum_r \Theta_{kr} x_i^{(r)}\right)} \tag{7}$$

where Θ is the parameter matrix to be learned and $x_i^{(r)}$ is the rth element of the feature vector $\mathbf{x}_i$.

Relying on the aforementioned analysis, we would like to introduce the DPA as a regularization term into the LDL model. However, the predicted ranking matrix used in the DPA fails to contain parameters explicitly, and the matrix cannot be directly adopted as a regularization term to constrain the parameters. To solve this problem and for easy optimization, we swap the order of ranking and summation, that is, we rank the labels by ground-truth label distributions and then find the weighted mean of predicted description degrees. Then, the ranking optimization term based on the DPA is calculated by the following equation:

$$\mathrm{Rank}_{\mathrm{DPA}}\left(\mathbf{R}_{\mathbf{d}_i}, \widehat{\mathbf{d}}_i\right) = \frac{1}{c} \sum_i \mathbf{R}_{\mathbf{d}_i} \cdot \widehat{\mathbf{d}}_i^{\,T} \tag{8}$$

where c is the number of labels and $\mathbf{R}_{\mathbf{d}_i}$ is the ranking matrix obtained from the ground-truth label distribution $\mathbf{d}_i$.

In the learning process, to maintain the distinguishability of labels and the shape of label distributions, we introduce the distance between the variance of the predicted and the ground-truth distributions to the minimization objective

$$\mathrm{Dis}_{\mathrm{var}}\left(\mathbf{d}_i, \widehat{\mathbf{d}}_i\right) = \left(\mathrm{var}(\mathbf{d}_i) - \mathrm{var}\left(\widehat{\mathbf{d}}_i\right)\right)^2 \tag{9}$$

where $\mathrm{var}(\mathbf{d}_i) = 1/c * \sum_{j=1}^{c} (d_i^{y_j} - \mu)^2$ is the variance of $\mathbf{d}_i$, which can reflect its distinctness quantitatively to some extent and μ is the mean of label description degrees and is constantly equal to $1/c$ in LDL.

Algorithm 1 LDL_DPA Framework

Input: training set $\{\mathbf{x}, \mathbf{d}\}$, parameters λ_1, λ_2, λ_3.
Output: $p\left(\widehat{d}_i^{y_j} \mid \mathbf{x}_i; \Theta\right)$.
Initialization the model parameter vector Θ;
Compute the ranking matrix $\mathbf{R}_{\mathbf{d}_i}$;
repeat
  update Θ by solving (11) using L-BFGS;
until *stopping criterion is satisfied*;
return the predicted $\widehat{d}_i^{y_j}$ by solving Eq. (7).

In summary, the framework of LDL_DPA can be formulated as follows:

$$\begin{aligned}\Theta^* = \arg\min_{\Theta} & \sum_i \mathrm{Dis}_{\mathrm{K-L}}\left(\mathbf{d}_i, \widehat{\mathbf{d}}_i\right) + \lambda_1 \|\Theta\|_F^2 \\ & - \lambda_2 \sum_i \mathrm{Rank}_{\mathrm{DPA}}\left(\mathbf{R}_{\mathbf{d}_i}, \widehat{\mathbf{d}}_i\right) \\ & + \lambda_3 \sum_i \mathrm{Dis}_{\mathrm{var}}\left(\mathbf{d}_i, \widehat{\mathbf{d}}_i\right)\end{aligned} \tag{10}$$

where λ_1, λ_2 and λ_3 are the balance factors. As a larger DPA means better label ranking, the third term of the objective function is a maximization problem. The overall procedure of LDL_DPA is summarized in Algorithm 1. To facilitate understanding, we have drawn the overall framework of LDL_DPA, as shown in Fig. 7.

B. Optimization

Substituting (7) into (10) yields the objective function of Θ

$$\begin{aligned}T(\Theta) = & \sum_i \mathbf{d}_i \ln \mathbf{d}_i - \sum_i \mathbf{d}_i \ln \widehat{d}_i^{y_j} \\ & + \lambda_1 \|\Theta\|_F^2 - \frac{\lambda_2}{c} \sum_i \mathbf{R}_{\mathbf{d}_i} \widehat{d}_i^{y_j} \\ & + \lambda_3 \sum_i \left(\mathrm{var}(\mathbf{d}_i) - \frac{1}{c} \sum_j \left(\widehat{d}_i^{y_j} - \frac{1}{c}\right)^2\right)^2.\end{aligned} \tag{11}$$

The minimization of the function $T(\Theta)$ can be effectively solved by L-BFGS [27], and the gradient is calculated by the following equation:

$$\begin{aligned}\frac{\partial T}{\partial \Theta_{jr}} = &\sum_i \frac{\exp\left(\sum_r \Theta_{jr} x_i^{(r)}\right)}{\sum_k \exp\left(\sum_r \Theta_{kr} x_i^{(r)}\right)} - \sum_i d_{x_i}^{y_j} x_i^{(r)} \\ &+ 2\lambda_1 \Theta_{jr} - \frac{\lambda_2}{c} \sum_i \mathbf{R}_{\mathrm{d}_i} \frac{\partial \widehat{d_i^{y_j}}}{\partial \Theta_{jr}} \\ &+ \lambda_3 \sum_i 2\left(\mathrm{var}(\mathbf{d}_i) - \frac{1}{c}\sum_j \left(\widehat{d_i^{y_j}} - \frac{1}{c}\right)^2\right) \\ &\cdot \frac{\partial\, \mathrm{var}\left(\widehat{\mathbf{d}}_i\right)}{\partial \widehat{d_i^{y_j}}} \frac{\partial \widehat{d_i^{y_j}}}{\partial \Theta_{jr}} \end{aligned} \tag{12}$$

where

$$\frac{\partial\, \mathrm{var}\left(\widehat{\mathbf{d}}_i\right)}{\partial \widehat{d_i^{y_j}}} = \frac{2}{c}\sum_j \left(\frac{\exp\left(\sum_r \Theta_{jr} x_i^{(r)}\right)}{\sum_k \exp\left(\sum_r \Theta_{kr} x_i^{(r)}\right)} - \frac{1}{c}\right) \tag{13}$$

$$\begin{aligned}\frac{\partial \widehat{d_i^{y_j}}}{\partial \Theta_{jr}} = &\frac{x_i^{(r)} \exp\left(\sum_r \Theta_{jr} x_i^{(r)}\right)}{\sum_k \exp\left(\sum_r \Theta_{kr} x_i^{(r)}\right)} \\ &\cdot \left(1 - \frac{\exp\left(\sum_r \Theta_{jr} x_i^{(r)}\right)}{\sum_k \exp\left(\sum_r \Theta_{kr} x_i^{(r)}\right)}\right). \end{aligned} \tag{14}$$

V. Experiments

A. Datasets

The datasets used in our experiments were collected from different sources, including two facial expression datasets (SJAFFE [28] and SBU [29]); two image emotion datasets (Twitter and Flickr [30]); two datasets associating face images and age (MORPH [31] and FG-Net [32]); a movie ratings dataset (Movie [33]); a natural scene dataset transformed from multilabel image dataset (Scene [34]); a dataset reflecting the relationship between human genes and diseases (Gene); and nine biological experiments datasets (Yeast [35]). We attempt to present the practicality of the algorithm by validating it on different types of datasets. The characteristics of the 18 datasets are summarized in Table I.

Since the feature dimensions of Twitter and Flickr are too high, we reduce them to 168 dimensions by using principal component analysis (PCA). The ages of face images in MORPH range from 16 to 77, while in FG-Net, the ages range from 0 to 69. We generate the ground truth of age distributions in these two datasets by using the Gaussian distribution, centered on the true age, with a standard deviation of 3. The feature vectors of MORPH are biologically inspired features (BIFs) [36]. We reduced their dimension to 200 by using marginal fisher analysis (MFA) [37] for easier experiments. There are ten datasets in total in the Yeast series. We select nine of them (10–18 in Table I) because the remaining dataset has only two labels, which may not reflect the label ranking relation well.

TABLE I
Statistics of 18 Real-World Datasets

Index	Datasets	#Examples	#Features	#Labels
1	SJAFFE	213	243	6
2	SBU	2500	243	6
3	Twitter	10045	168	8
4	Flickr	11150	168	8
5	MORPH	55132	200	62
6	FG-Net	1002	251	70
7	Movie	7755	1869	5
8	Gene	17892	36	68
9	Scene	2000	294	9
10	Alpha	2465	24	18
11	Cdc	2465	24	15
12	Cold	2465	24	4
13	Diau	2465	24	7
14	Dtt	2465	24	4
15	Elu	2465	24	14
16	Heat	2465	24	6
17	Spo	2465	24	6
18	Spo5	2465	24	3

TABLE II
Evaluation Measures of Different Types for LDL Algorithms, Where "↓" Indicates "the Smaller the Better," "↑" Indicates "the Larger the Better"

	Name	Formula
Distance	Euclidean ↓	$\mathrm{Dis}_1(\mathbf{d}, \hat{\mathbf{d}}) = \sqrt{\sum_{j=1}^{c} \left(d_j - \hat{d}_j\right)^2}$
	Canberra ↓	$\mathrm{Dis}_2(\mathbf{d}, \hat{\mathbf{d}}) = \sum_{j=1}^{c} \frac{\lvert d_j - \hat{d}_j \rvert}{\lvert d_j + \hat{d}_j \rvert}$
	K-L ↓	$\mathrm{Dis}_3(\mathbf{d}, \hat{\mathbf{d}}) = \sum_{j=1}^{c} d_j \ln \frac{d_j}{\hat{d}_j}$
	Squared-chord ↓	$\mathrm{Dis}_4(\mathbf{d}, \hat{\mathbf{d}}) = \sum_{j=1}^{c} \left(\sqrt{d_j} - \sqrt{\hat{d}_j}\right)^2$
Similarity	Intersection ↑	$\mathrm{Sim}_1(\mathbf{d}, \hat{\mathbf{d}}) = \sum_{j=1}^{c} \min\left(d_j, \hat{d}_j\right)$
	Cosine ↑	$\mathrm{Sim}_2(\mathbf{d}, \hat{\mathbf{d}}) = \frac{\sum_{j=1}^{c} d_j \hat{d}_j}{\sqrt{\sum_{j=1}^{c} d_j^2}\sqrt{\sum_{j=1}^{c} \hat{d}_j^2}}$
Ranking	Spearman's rank ↑	$\mathrm{Spearman}(\mathbf{d}, \hat{\mathbf{d}}) = 1 - \frac{6\sum_{j=1}^{c}(\mathbf{R}_{\mathbf{d}}^j - \mathbf{R}_{\hat{\mathbf{d}}}^j)^2}{c(c^2-1)}$
	DPA ↑	$\mathrm{DPA}(\mathbf{d}, \hat{\mathbf{d}}) = \frac{1}{c}\sum_{j=1}^{c} \mathbf{R}_{\hat{\mathbf{d}}}^j \cdot \mathbf{d}_j$

B. Evaluation Metrics

Existing evaluation metrics for LDL are borrowed from the distance or similarity metrics between probability distribution functions. According to the summary of Cha [24], these metrics belong to different syntactic families and have various semantics. To evaluate experimental results from multiple perspectives, we first select a metric from each syntactic family separately, where the widely used metrics are preferred. Because there exists a separate L_1 family, L_2 (Euclidean) is chosen from among the L_p Minkowski family that contains L_1 (City block). Such selection criteria can ensure the syntactic diversity of the evaluation system. To compare the ranking performance of different LDL methods, Spearman's rank correlation (Spearman) is included in our evaluation system, which compares two lists of definite ranking values. While Spearman measures the overall ranking effect of the algorithms, the DPA we designed prefers to measure the prediction of key labels.

Finally, eight evaluation measures are selected, of which four are distance-based metrics Euclidean, Canberra, K-L, and

TABLE III

COMPARISON RESULTS OF DIFFERENT LABEL DISTRIBUTION ALGORITHMS ON NINE DATASETS, WHERE "↓" INDICATES "THE SMALLER THE BETTER," "↑" INDICATES "THE LARGER THE BETTER." THE BEST PERFORMANCE AMONG THE COMPARISON ALGORITHMS IS HIGHLIGHTED IN BOLDFACE

Data	Algorithm	Euclidean ↓	Canberra ↓	K-L ↓	Squared ↓	Intersection ↑	Cosine ↑	Spearman ↑	DPA ↑
SJAFFE	LDL_DPA	**.1118±.0065**	**.6453±.0286**	**.0396±.0043**	**.0197±.0020**	**.8909±.0061**	**.9625±.0042**	**.5234±.0469**	**.6455±.0083**
	LDL-LRR	.1172±.0051	.6891±.0223	.0437±.0037	.0218±.0018	.8841±.0045	.9589±.0034	.4358±.0352	.6386±.0049
	LDL-LCLR	.1315±.0088	.7593±.0322	.0526±.0059	.0260±.0027	.8711±.0068	.9501±.0059	.4454±.0324	.6367±.0063
	LDLSF	.1143±.0050	.6993±.0369	.0485±.0073	.0223±.0015	.8870±.0061	.9609±.0034	.4996±.0645	.6442±.0088
	LALOT	.1468±.0002	.8567±.0014	.0677±.0001	.0332±.0001	.8552±.0002	.9361±.0001	-.0302±.0829	.5834±.0074
	BFGS-LLD	.1547±.0107	.9004±.0434	.0738±.0089	.0359±.0040	.8467±.0088	.9304±.0084	-.0736±.0342	.5816±.0058
	AA-kNN	.1286±.0090	.7280±.0382	.0535±.0063	.0266±.0029	.8749±.0076	.9489±.0059	.4017±.0374	.6299±.0072
	PT-Bayes	.1548±.0057	.9007±.0313	.0738±.0055	.0359±.0025	.8466±.0052	.9304±.0048	.5040±.0015	.5833±.0000
SBU	LDL_DPA	**.1344±.0020**	**.7272±.0128**	**.0541±.0016**	**.0266±.0008**	**.8709±.0023**	**.9469±.0014**	**.4042±.0176**	**.6375±.0018**
	LDL-LRR	.1595±.0025	.8484±.0078	.0728±.0021	.0353±.0012	.8472±.0017	.9280±.0023	.1248±.0530	.6149±.0052
	LDL-LCLR	.1551±.0034	.8345±.0174	.0698±.0032	.0336±.0015	.8504±.0033	.9314±.0030	.2473±.0221	.6224±.0025
	LDLSF	.1400±.0022	.7707±.0120	.0602±.0018	.0286±.0007	.8646±.0023	.9442±.0014	.3714±.0159	.6358±.0019
	LALOT	.1918±.0070	1.0620±.0713	.1147±.0114	.0538±.0050	.8118±.0098	.8949±.0076	.0527±.0853	.5833±.0055
	BFGS-LLD	.1668±.0022	.9028±.0111	.0851±.0021	.0398±.0010	.8387±.0021	.9177±.0018	-.0468±.0157	.5858±.0020
	AA-kNN	.1591±.0045	.8268±.0230	.0784±.0043	.0377±.0019	.8495±.0043	.9233±.0038	.2423±.0149	.6122±.0026
	PT-Bayes	.1668±.0041	.9028±.0206	.0851±.0036	.0398±.0016	.8387±.0039	.9177±.0034	.3610±.0035	.5833±.0000
Twitter	LDL_DPA	**.3778±.0042**	6.1963±.0091	**.6337±.0098**	.4131±.0048	.6235±.0037	.8236±.0044	**.6047±.0029**	**.8529±.0015**
	LDL-LRR	.3989±.0082	6.2440±.0215	.6826±.0152	.4594±.0116	.5932±.0081	.8136±.0060	.5114±.0037	.8400±.0021
	LDL-LCLR	.3820±.0041	6.3819±.0124	.7079±.0189	**.3971±.0038**	**.6345±.0023**	.8131±.0049	.5924±.0016	.8468±.0017
	LDLSF	.3784±.0043	5.8050±.0327	.7462±.0184	.4182±.0053	.6234±.0040	**.8249±.0041**	.6040±.0040	.8524±.0014
	LALOT	.6554±.0460	6.3049±.1150	1.3291±.2356	.8897±.1028	.3168±.0476	.4889±.0988	.2298±.1936	.6082±.1385
	BFGS-LLD	.6381±.0024	6.1963±.0091	1.3088±.0055	.8991±.0027	.3439±.0014	.5034±.0013	.6046±.0030	.8528±.0015
	AA-kNN	.4036±.0032	**5.4005±.0264**	1.0959±.0451	.4567±.0037	.6047±.0019	.8043±.0029	.5780±.0042	.8356±.0014
	PT-Bayes	.6707±.0124	7.2414±.0127	4.0197±.0422	.7594±.0193	.4555±.0104	.5750±.0131	.5846±.0075	.8033±.0041
Flickr	LDL_DPA	**.3909±.0037**	5.2539±.0171	**.6396±.0075**	**.4372±.0048**	**.5950±.0033**	.7767±.0039	.5520±.0036	.7980±.0017
	LDL-LRR	.4210±.0033	5.4967±.0328	.7226±.0119	.4875±.0052	.5639±.0028	.7441±.0056	.4427±.0105	.7707±.0020
	LDL-LCLR	.3955±.0039	5.5297±.0170	.6609±.0082	.4410±.0041	.5922±.0032	.7688±.0043	.5390±.0031	.7923±.0011
	LDLSF	.3933±.0039	**5.2231±.0261**	.7712±.0192	.4445±.0055	.5937±.0034	**.7771±.0044**	**.5521±.0025**	**.7988±.0018**
	LALOT	.5857±.0604	5.6365±.1107	1.3140±.3283	.8171±.1384	.3751±.0670	.4709±.1205	.0607±.2240	.5471±.1247
	BFGS-LLD	.5478±.0016	5.5373±.0204	1.0541±.0046	.7379±.0034	.4106±.0017	.5575±.0013	.5518±.0036	.7979±.0018
	AA-kNN	.4251±.0028	5.2889±.0177	.9078±.0119	.4901±.0043	.5654±.0022	.7359±.0038	.4902±.0052	.7703±.0020
	PT-Bayes	.7085±.0093	7.1203±.0188	4.1765±.1008	.8186±.0165	.3964±.0071	.5329±.0092	.4958±.0060	.7535±.0031
MORPH	LDL_DPA	**.1919±.0003**	57.9896±.0071	**.6432±.0018**	**.3961±.0003**	**.5742±.0004**	**.7402±.0005**	**.8847±.0011**	**.8769±.0002**
	LDL-LRR	.2628±.0007	60.2786±.0290	1.2805±.0126	.8179±.0039	.3328±.0025	.5095±.0051	.6364±.0097	.7461±.0025
	LDL-LCLR	.1965±.0002	58.3353±.0046	.6669±.0020	.4257±.0005	.5535±.0003	.7353±.0003	.8517±.0015	.8756±.0002
	LDLSF	.2187±.0002	59.3533±.0040	.8594±.0056	.5410±.0012	.4804±.0006	.6841±.0006	.7577±.0018	.8557±.0006
	LALOT	.3667±.4783	61.1598±4.6862	1.6719±.0002	1.0504±.0052	.3948±.0009	.3912±.0015	.0200±.0001	.5026±.0002
	BFGS-LLD	.2825±.0000	61.5961±.0014	1.7318±.0001	1.0734±.0001	.2151±.0000	.3897±.0000	.0066±.0017	.5125±.0009
	AA-kNN	.2121±.0006	**53.9969±.0876**	.8839±.0113	.5210±.0034	.5067±.0015	.6786±.0026	.8215±.0019	.8432±.0015
	PT-Bayes	.2162±.0009	61.5828±.0578	1.1380±.0103	1.0225±.0016	.3746±.0009	.6876±.0008	.5287±.0002	.5081±.0000
FG-Net	LDL_DPA	**.1989±.0066**	**59.0951±.2174**	**.8011±.0777**	**.4091±.0243**	**.5902±.0137**	**.7122±.0159**	**.8220±.0114**	**.8662±.0077**
	LDL-LRR	.2644±.0022	61.4232±.0862	1.3011±.0363	.7930±.0190	.3553±.0108	.5411±.0067	.7105±.0316	.7779±.0099
	LDL-LCLR	.2388±.0051	59.5371±.2315	1.1752±.0784	.4756±.0156	.5500±.0099	.6482±.0112	.7938±.0205	.8607±.0064
	LDLSF	.2359±.0060	63.1485±.1237	3.4502±.1889	.7275±.0193	.4732±.0094	.6083±.0117	.3821±.0074	.7571±.0061
	LALOT	.3430±.0081	60.1758±.2217	1.1156±.0330	1.0290±.0992	.2578±.0051	.6874±.0050	.0642±.0083	.5072±.0001
	BFGS-LLD	.2953±.0010	62.5384±.0785	1.8134±.0058	1.1153±.0030	.2002±.0011	.3773±.0009	-.0146±.0053	.4979±.0092
	AA-kNN	.2553±.0029	60.4329±.3074	1.2364±.0560	.7471±.0123	.3791±.0069	.5688±.0128	.7290±.0134	.7871±.0089
	PT-Bayes	.8891±.0130	68.4146±.3582	1.6970±.0924	1.6646±.0118	.0419±.0052	.1433±.0052	.3664±.0165	.5073±.0000
Movie	LDL_DPA	**.1640±.0017**	**.9818±.0074**	**.0959±.0024**	**.0498±.0012**	**.8381±.0017**	**.9368±.0018**	**.7120±.0062**	**.7179±.0016**
	LDL-LRR	.1702±.0015	1.0201±.0072	.1017±.0021	.0531±.0010	.8316±.0013	.9330±.0015	.7013±.0052	.7156±.0013
	LDL-LCLR	.1868±.0026	1.0887±.0129	.1269±.0036	.0636±.0018	.8176±.0025	.9195±.0023	.6623±.0091	.7098±.0018
	LDLSF	.1904±.0026	1.2064±.0134	.2753±.0196	.0763±.0023	.8127±.0026	.9174±.0024	.6527±.0067	.7081±.0017
	LALOT	.4041±.0902	1.9512±.3976	.5109±.1994	.2432±.0868	.6194±.0885	.6858±.1104	-.1737±.3555	.5715±.0635
	BFGS-LLD	.2650±.0008	1.4372±.0032	.1967±.0012	.1079±.0007	.7389±.0004	.8585±.0006	.0697±.0127	.6080±.0022
	AA-kNN	.1785±.0012	1.0509±.0076	.1130±.0022	.0588±.0011	.8240±.0013	.9253±.0014	.6569±.0117	.7094±.0021
	PT-Bayes	.2849±.0039	1.5567±.0206	.5732±.0562	.1378±.0050	.7235±.0027	.8501±.0020	.5087±.0009	.6028±.0006
Gene	LDL_DPA	**.0864±.0004**	14.4347±.1423	.2360±.0026	.1069±.0015	**.7846±.0018**	**.8347±.0008**	**.2070±.0035**	**.5292±.0005**
	LDL-LRR	.0874±.0003	14.6225±.1358	.2392±.0021	.1085±.0016	.7812±.0016	.8319±.0006	.1061±.0157	.5215±.0010
	LDL-LCLR	**.0864±.0004**	14.4493±.1445	.2360±.0027	.1070±.0016	.7843±.0018	.8346±.0009	.1641±.0024	.5288±.0005
	LDLSF	**.0864±.0004**	14.4492±.1416	.2369±.0023	.1071±.0015	.7844±.0018	**.8347±.0009**	.1680±.0025	**.5292±.0005**
	LALOT	.0872±.0022	**14.2661±.1110**	**.2313±.0231**	**.1068±.0099**	.7836±.0049	.8310±.0071	.0287±.0015	.5053±.0004
	BFGS-LLD	.0867±.0004	14.5309±.1445	.2386±.0027	.1082±.0016	.7833±.0018	.8332±.0009	.0480±.0012	.5118±.0006
	AA-kNN	.0876±.0004	14.5922±.1417	.2390±.0027	.1086±.0016	.7810±.0017	.8313±.0009	.1466±.0024	.5267±.0010
	PT-Bayes	.2671±.0332	34.5712±1.3736	1.6615±.1081	.5190±.0451	.4609±.0259	.4404±.0518	.0400±.0057	.5109±.0005
Scene	LDL_DPA	**.4179±.0094**	6.6686±.0331	**.7741±.0219**	**.4952±.0114**	.5554±.0059	**.7351±.0086**	**.5761±.0099**	**.7811±.0060**
	LDL-LRR	.4345±.0081	6.7292±.0273	.7997±.0217	.5476±.0126	.5210±.0059	.7276±.0129	.4926±.0207	.7743±.0099
	LDL-LCLR	.4244±.0102	6.6540±.0285	.8036±.0250	.4988±.0124	.5530±.0065	.7257±.0092	.5701±.0099	.7784±.0056
	LDLSF	.4326±.0063	6.7962±.0236	1.3030±.0685	.5398±.0103	.5426±.0050	.7238±.0112	.5467±.0131	.7690±.0089
	LALOT	.5366±.0087	7.0573±.0435	1.1931±.0501	.7933±.0211	.3617±.0123	.5727±.0207	.2597±.1283	.6206±.0753
	BFGS-LLD	.5252±.0086	6.9997±.0525	1.1725±.0217	.7865±.0108	.3630±.0064	.5748±.0055	.1167±.0154	.5349±.0063
	AA-kNN	.5162±.0054	**5.4043±.0356**	.8581±.0226	.5099±.0099	**.5594±.0057**	.7265±.0109	.5655±.0130	.7745±.0089
	PT-Bayes	.5647±.0071	7.1495±.0203	2.0575±.0948	.8203±.0064	.3497±.0036	.5585±.0036	.5552±.0038	.5606±.0011

Squared-chord (Squared), two similarity-based metrics (i.e., intersection and cosine), a label ranking metric, Spearman, and the DPA were introduced. The specific names and formulas of these metrics are displayed in Table II.

C. Experimental Setting

We have carried out comparative experiments with seven types of LDL algorithms: PT-Bayes, AA-kNN, BFGS-LLD [1], LDL-LRR [3], LDL-LCLR [7], LDLSF [38], and

TABLE IV

COMPARISON RESULTS OF DIFFERENT LABEL DISTRIBUTION ALGORITHMS ON YEAST DATASETS, WHERE "↓" INDICATES "THE SMALLER THE BETTER," "↑" INDICATES "THE LARGER THE BETTER." THE BEST PERFORMANCE AMONG THE COMPARISON ALGORITHMS IS HIGHLIGHTED IN BOLDFACE

Data	Algorithm	Euclidean ↓	Canberra ↓	K-L ↓	Squared ↓	Intersection ↑	Cosine ↑	Spearman ↑	DPA ↑
Alpha	LDL_DPA	**.0230±.0003**	**.6788±.0087**	**.0055±.0002**	**.0028±.0001**	**.9625±.0005**	**.9946±.0001**	**.2185±.0101**	**.5331±.0003**
	LDL-LRR	.0235±.0004	.6969±.0116	.0057±.0002	.0029±.0001	.9615±.0006	.9944±.0002	.1465±.0384	.5314±.0010
	LDL-LCLR	.0231±.0003	.6821±.0088	**.0055±.0002**	**.0028±.0001**	.9623±.0005	**.9946±.0001**	.2044±.0091	.5328±.0003
	LDLSF	.0231±.0003	.6825±.0090	**.0055±.0002**	**.0028±.0001**	.9623±.0005	**.9946±.0001**	.2027±.0077	.5328±.0002
	LALOT	.0245±.0002	.7302±.0039	.0060±.0001	.0030±.0000	.9596±.0002	.9941±.0001	.0197±.0133	.5283±.0003
	BFGS-LLD	.0236±.0003	.6995±.0084	.0057±.0002	.0029±.0001	.9614±.0005	.9944±.0001	-.0718±.0068	.5259±.0002
	AA-kNN	.0233±.0003	.6909±.0081	.0056±.0002	**.0028±.0001**	.9618±.0004	.9945±.0001	.1930±.0114	.5327±.0003
	PT-Bayes	.1603±.0132	4.5161±.3053	.3211±.0449	.1310±.0171	.7557±.0167	.8335±.0158	.0510±.0313	.5291±.0008
Cdc	LDL_DPA	**.0278±.0006**	**.6449±.0118**	**.0069±.0003**	**.0036±.0002**	**.9576±.0007**	**.9933±.0003**	**.2019±.0120**	**.5405±.0004**
	LDL-LRR	.0285±.0004	.6627±.0086	.0072±.0003	.0037±.0002	.9564±.0006	.9931±.0002	.1462±.0272	.5391±.0007
	LDL-LCLR	.0279±.0006	.6475±.0114	.0070±.0003	**.0036±.0002**	.9574±.0007	**.9933±.0003**	.1952±.0151	.5402±.0005
	LDLSF	.0279±.0006	.6478±.0114	.0070±.0003	**.0036±.0002**	.9574±.0007	**.9933±.0003**	.1944±.0144	.5402±.0005
	LALOT	.0287±.0002	.6651±.0045	.0074±.0001	.0037±.0000	.9562±.0003	.9929±.0001	-.0191±.0232	.5325±.0008
	BFGS-LLD	.0284±.0006	.6551±.0133	.0073±.0004	.0037±.0002	.9569±.0008	.9930±.0003	-.0371±.0166	.5328±.0006
	AA-kNN	.0284±.0005	.6581±.0110	.0072±.0003	.0037±.0002	.9567±.0007	.9931±.0003	.1710±.0160	.5394±.0007
	PT-Bayes	.1678±.0086	3.6812±.1979	.3051±.0343	.1238±.0105	.7610±.0114	.8410±.0093	.0077±.0246	.5338±.0007
Cold	LDL_DPA	**.0680±.0014**	**.2396±.0049**	**.0121±.0007**	**.0062±.0004**	**.9410±.0012**	**.9886±.0006**	**.2575±.0163**	**.6377±.0009**
	LDL-LRR	.0687±.0014	.2421±.0047	.0124±.0006	.0063±.0004	.9403±.0011	.9883±.0005	.2390±.0290	.6366±.0012
	LDL-LCLR	.0682±.0013	.2402±.0047	.0122±.0006	**.0062±.0004**	.9408±.0012	**.9886±.0005**	.2554±.0156	.6376±.0010
	LDLSF	.0682±.0013	.2402±.0047	.0122±.0006	**.0062±.0004**	.9408±.0012	**.9886±.0005**	.2568±.0167	.6376±.0010
	LALOT	.0734±.0002	.2583±.0008	.0142±.0001	.0072±.0000	.9361±.0002	.9866±.0001	-.0510±.0361	.6218±.0015
	BFGS-LLD	.0719±.0012	.2528±.0041	.0135±.0007	.0069±.0004	.9376±.0009	.9873±.0006	-.0031±.0268	.6259±.0017
	AA-kNN	.0721±.0017	.2539±.0064	.0135±.0006	.0068±.0004	.9373±.0016	.9873±.0006	.2110±.0229	.6356±.0011
	PT-Bayes	.2480±.0105	.9133±.0411	.2483±.0335	.0965±.0084	.7858±.0085	.8841±.0068	.0167±.0265	.6259±.0015
Diau	LDL_DPA	**.0543±.0009**	.4311±.0075	**.0131±.0005**	**.0067±.0003**	**.9403±.0011**	**.9880±.0004**	**.4173±.0183**	**.5908±.0009**
	LDL-LRR	.0547±.0010	.4342±.0078	.0133±.0005	.0068±.0003	.9398±.0011	.9878±.0004	.4102±.0156	.5904±.0008
	LDL-LCLR	**.0543±.0009**	.4311±.0077	**.0131±.0005**	**.0067±.0003**	.9402±.0011	.9879±.0004	.4152±.0193	**.5908±.0009**
	LDLSF	**.0543±.0009**	**.4308±.0078**	**.0131±.0005**	**.0067±.0003**	**.9403±.0011**	.9879±.0004	.4167±.0177	**.5908±.0007**
	LALOT	.0630±.0002	.4968±.0018	.0171±.0001	.0088±.0001	.9310±.0003	.9843±.0001	.0555±.0288	.5759±.0018
	BFGS-LLD	.0605±.0009	.4777±.0087	.0159±.0005	.0082±.0003	.9335±.0012	.9853±.0004	-.0104±.0201	.5710±.0011
	AA-kNN	.0554±.0005	.4391±.0038	.0137±.0004	.0070±.0003	.9391±.0005	.9873±.0003	.3927±.0079	.5901±.0005
	PT-Bayes	.2264±.0132	1.7781±.1052	.2993±.0246	.1198±.0085	.7600±.0137	.8545±.0097	.1208±.0396	.5788±.0021
Dtt	LDL_DPA	**.0479±.0005**	**.1686±.0020**	**.0062±.0003**	**.0032±.0002**	**.9584±.0005**	**.9941±.0002**	**.2117±.0214**	**.6317±.0004**
	LDL-LRR	.0484±.0009	.1708±.0037	.0064±.0004	.0033±.0003	.9579±.0009	.9940±.0003	.1771±.0374	.6300±.0009
	LDL-LCLR	.0480±.0005	.1691±.0019	.0063±.0003	**.0032±.0002**	.9583±.0005	**.9941±.0002**	.2082±.0223	**.6317±.0004**
	LDLSF	.0480±.0005	.1691±.0019	.0063±.0003	**.0032±.0002**	.9583±.0005	**.9941±.0002**	.2089±.0231	**.6317±.0004**
	LALOT	.0505±.0002	.1775±.0010	.0068±.0000	.0035±.0000	.9561±.0002	.9936±.0000	-.0116±.0227	.6246±.0006
	BFGS-LLD	.0495±.0004	.1738±.0017	.0066±.0003	.0034±.0002	.9571±.0004	.9938±.0002	.0520±.0155	.6267±.0004
	AA-kNN	.0509±.0012	.1789±.0039	.0070±.0003	.0036±.0002	.9559±.0011	.9934±.0003	.1706±.0369	.6305±.0010
	PT-Bayes	.2301±.0085	.8604±.0292	.2217±.0126	.0868±.0047	.8002±.0064	.8962±.0045	.0843±.0436	.6278±.0017
Elu	LDL_DPA	**.0276±.0006**	**.5812±.0122**	**.0061±.0002**	**.0031±.0001**	**.9590±.0008**	**.9941±.0002**	**.2302±.0118**	**.5424±.0003**
	LDL-LRR	.0279±.0006	.5892±.0116	.0063±.0002	.0032±.0001	.9584±.0008	.9939±.0002	.1950±.0096	.5415±.0004
	LDL-LCLR	.0277±.0006	.5835±.0125	.0062±.0003	.0032±.0001	.9588±.0009	.9940±.0002	.2223±.0096	.5422±.0003
	LDLSF	.0277±.0006	.5836±.0124	.0062±.0003	.0032±.0001	.9588±.0009	.9940±.0002	.2214±.0095	.5422±.0004
	LALOT	.0294±.0003	.6204±.0060	.0069±.0001	.0035±.0001	.9562±.0004	.9934±.0001	.0233±.0271	.5363±.0008
	BFGS-LLD	.0284±.0006	.6019±.0133	.0065±.0003	.0033±.0001	.9575±.0009	.9938±.0003	-.0260±.0059	.5351±.0001
	AA-kNN	.0282±.0005	.5938±.0103	.0064±.0002	.0032±.0001	.9581±.0007	.9938±.0002	.2007±.0066	.5418±.0003
	PT-Bayes	.1714±.0121	3.4322±.2381	.3067±.0371	.1241±.0129	.7625±.0163	.8428±.0143	.0424±.0140	.5369±.0005
Heat	LDL_DPA	**.0592±.0012**	**.3637±.0072**	**.0126±.0005**	**.0064±.0003**	**.9403±.0012**	**.9880±.0005**	**.1735±.0182**	**.5924±.0008**
	LDL-LRR	.0600±.0011	.3687±.0060	.0129±.0005	.0066±.0002	.9395±.0010	.9877±.0005	.1544±.0330	.5914±.0010
	LDL-LCLR	.0594±.0012	.3651±.0076	.0127±.0005	.0065±.0003	.9401±.0012	.9879±.0005	.1644±.0147	.5920±.0007
	LDLSF	.0594±.0012	.3653±.0076	.0127±.0005	.0065±.0003	.9400±.0012	.9879±.0005	.1620±.0139	.5919±.0007
	LALOT	.0614±.0002	.3792±.0010	.0138±.0001	.0072±.0000	.9378±.0002	.9870±.0001	.0262±.0260	.5835±.0013
	BFGS-LLD	.0609±.0013	.3735±.0081	.0134±.0006	.0068±.0003	.9386±.0013	.9872±.0005	-.0341±.0165	.5805±.0008
	AA-kNN	.0611±.0016	.3772±.0100	.0135±.0007	.0069±.0004	.9381±.0016	.9872±.0006	.1680±.0265	.5919±.0010
	PT-Bayes	.2426±.0046	1.5151±.0293	.2881±.0133	.1157±.0041	.7589±.0039	.8555±.0034	.0558±.0231	.5867±.0010
Spo	LDL_DPA	**.0819±.0011**	**.5123±.0109**	**.0245±.0008**	**.0126±.0005**	**.9157±.0016**	**.9770±.0005**	**.2101±.0277**	**.6015±.0020**
	LDL-LRR	.0824±.0012	.5156±.0108	.0248±.0008	.0128±.0005	.9151±.0016	.9768±.0005	.1747±.0423	.5999±.0023
	LDL-LCLR	.0821±.0011	.5137±.0108	.0247±.0008	.0127±.0005	.9154±.0016	.9769±.0005	.2000±.0220	.6008±.0018
	LDLSF	.0822±.0011	.5139±.0108	.0247±.0008	.0127±.0005	.9154±.0016	.9769±.0005	.1996±.0224	.6008±.0018
	LALOT	.0870±.0001	.5436±.0007	.0284±.0001	.0143±.0000	.9107±.0001	.9736±.0001	-.0363±.0138	.5790±.0003
	BFGS-LLD	.0845±.0015	.5283±.0121	.0271±.0009	.0137±.0005	.9129±.0019	.9747±.0008	.0320±.0181	.5846±.0008
	AA-kNN	.0860±.0012	.5364±.0103	.0273±.0009	.0140±.0005	.9115±.0015	.9743±.0006	.1562±.0279	.5961±.0016
	PT-Bayes	.2370±.0090	1.4803±.0560	.2927±.0256	.1162±.0068	.7639±.0095	.8567±.0068	.0723±.0452	.5891±.0035
Spo5	LDL_DPA	**.1169±.0034**	**.2826±.0083**	**.0293±.0017**	**.0152±.0009**	**.9087±.0026**	**.9741±.0014**	**.1240±.0264**	**.6752±.0018**
	LDL-LRR	.1172±.0030	.2835±.0073	.0296±.0015	.0153±.0008	.9084±.0023	.9739±.0012	.0932±.0331	.6732±.0016
	LDL-LCLR	**.1169±.0033**	.2828±.0083	**.0293±.0017**	**.0152±.0009**	**.9087±.0026**	**.9741±.0014**	.1214±.0258	.6751±.0019
	LDLSF	**.1169±.0033**	.2828±.0082	**.0293±.0017**	**.0152±.0009**	.9086±.0026	**.9741±.0014**	.1178±.0297	.6750±.0022
	LALOT	.1178±.0002	.2843±.0006	.0303±.0001	.0157±.0000	.9083±.0002	.9734±.0001	.0160±.0147	.6666±.0010
	BFGS-LLD	.1179±.0031	.2848±.0075	.0299±.0015	.0155±.0008	.9079±.0024	.9736±.0013	.0055±.0267	.6650±.0019
	AA-kNN	.1240±.0027	.2988±.0065	.0336±.0006	.0171±.0003	.9033±.0021	.9702±.0006	.1173±.0171	.6747±.0017
	PT-Bayes	.2823±.0200	.7186±.0508	.2423±.0310	.0959±.0118	.7786±.0150	.8833±.0132	.0302±.0523	.6695±.0029

LALOT [17]. The first three are basic LDL algorithms proposed earliest. The maximum likelihood estimation is exploited in PT-Bayes. The number of neighbors k in AA-kNN is set to $c+1$, where c is the number of labels. BFGS-LLD exploits K–L divergence as the loss function and is optimized by the quasi-Newton method BFGS.

TABLE V

ABLATION RESULTS ON DIFFERENT METRICS, WHERE "↓" INDICATES "THE SMALLER THE BETTER," "↑" INDICATES "THE LARGER THE BETTER." THE BEST RESULTS AMONG THREE VERSIONS ARE HIGHLIGHTED IN BOLDFACE

Data	Algorithm	Euclidean ↓	Canberra ↓	K-L ↓	Squared ↓	Intersection ↑	Cosine ↑	Spearman ↑	DPA ↑
SJAFFE	LDL_DPA	**0.1118**	**0.6453**	**0.0396**	**0.0197**	**0.8909**	**0.9625**	**0.5234**	**0.6455**
	LDL_DPA-Var	0.1311	0.7562	0.0523	0.0258	0.8716	0.9503	0.4503	0.6373
	LDL_DPA-Rank	0.1486	0.8617	0.0681	0.0332	0.8534	0.9356	0.2736	0.6157
	LDL_DPA-Var-Rank	0.1547	0.9004	0.0738	0.0359	0.8467	0.9304	-0.0736	0.5816
SBU	LDL_DPA	**0.1344**	**0.7272**	**0.0541**	**0.0266**	**0.8709**	**0.9469**	**0.4042**	**0.6375**
	LDL_DPA-Var	0.1435	0.7722	0.0597	0.0290	0.8619	0.9412	0.3450	0.6318
	LDL_DPA-Rank	0.1544	0.8312	0.0692	0.0333	0.8511	0.9320	0.2545	0.6233
	LDL_DPA-Var-Rank	0.1668	0.9028	0.0851	0.0398	0.8387	0.9177	-0.0468	0.5858
Twitter	LDL_DPA	**0.3778**	**6.1963**	**0.6337**	**0.4131**	**0.6235**	**0.8236**	**0.6047**	**0.8529**
	LDL_DPA-Var	0.3806	6.2339	0.6340	0.4239	0.6168	0.8234	0.6046	0.8528
	LDL_DPA-Rank	0.3898	6.2294	0.6560	0.4447	0.6030	0.8180	0.5965	0.8488
	LDL_DPA-Var-Rank	0.6381	6.2593	1.3088	0.8991	0.3439	0.5034	0.6037	0.8524
Flickr	LDL_DPA	**0.3909**	**5.3539**	**0.6396**	**0.4372**	**0.5950**	**0.7767**	**0.5520**	**0.7980**
	LDL_DPA-Var	0.4167	5.4952	0.6826	0.4948	0.5625	0.7622	0.5518	0.7979
	LDL_DPA-Rank	0.5187	5.4388	0.9412	0.6809	0.4441	0.6217	0.4898	0.7719
	LDL_DPA-Var-Rank	0.5478	5.5373	1.0541	0.7379	0.4106	0.5575	0.5328	0.7887
MORPH	LDL_DPA	**0.1919**	**57.9896**	**0.6432**	**0.3961**	**0.5742**	**0.7402**	**0.8847**	**0.8769**
	LDL_DPA-Var	0.1969	58.3868	0.6699	0.4285	0.5516	0.7347	0.8559	0.8754
	LDL_DPA-Rank	0.2628	60.6999	1.3459	0.8775	0.3012	0.5175	0.6556	0.7646
	LDL_DPA-Var-Rank	0.2825	61.5961	1.7318	1.0734	0.2151	0.3897	0.0066	0.5125
FG-Net	LDL_DPA	**0.1989**	**59.0951**	**0.8011**	**0.4091**	**0.5902**	**0.7122**	**0.8224**	**0.8662**
	LDL_DPA-Var	0.2064	60.1440	0.8529	0.5415	0.5052	0.7019	0.8220	0.8429
	LDL_DPA-Rank	0.2628	61.4435	1.2771	0.8005	0.3488	0.5487	0.7437	0.7822
	LDL_DPA-Var-Rank	0.2953	62.5384	1.8134	1.1153	0.2002	0.3773	-0.0146	0.4979
Movie	LDL_DPA	**0.1640**	**0.9818**	**0.0959**	**0.0498**	**0.8381**	**0.9368**	**0.7120**	**0.7179**
	LDL_DPA-Var	0.1817	1.0580	0.1180	0.0598	0.8225	0.9237	0.6727	0.7115
	LDL_DPA-Rank	0.1859	1.0808	0.1243	0.0625	0.8187	0.9205	0.6645	0.7102
	LDL_DPA-Var-Rank	0.2650	1.4372	0.1967	0.1079	0.7389	0.8585	0.0697	0.6080
Gene	LDL_DPA	**0.0864**	**14.4347**	**0.2360**	**0.1069**	**0.7846**	**0.8347**	**0.2070**	**0.5292**
	LDL_DPA-Var	0.0866	14.4710	0.2379	**0.1069**	0.7842	0.8336	0.1736	0.5243
	LDL_DPA-Rank	0.0867	14.5318	0.2386	0.1082	0.7833	0.8332	0.1351	0.5236
	LDL_DPA-Var-Rank	0.0867	14.5309	0.2386	0.1082	0.7833	0.8332	0.0480	0.5118
Scene	LDL_DPA	**0.4179**	**6.6560**	**0.7741**	**0.4952**	**0.5554**	**0.7351**	**0.5761**	**0.7811**
	LDL_DPA-Var	0.4331	6.6686	0.8354	0.5074	0.5469	0.7151	0.5614	0.7744
	LDL_DPA-Rank	0.4355	6.7326	0.8116	0.5203	0.5378	0.7187	0.5575	0.7731
	LDL_DPA-Var-Rank	0.5252	6.9997	1.1725	0.7865	0.3630	0.5748	0.1167	0.5349
Alpha	LDL_DPA	**0.0230**	**0.6788**	**0.0055**	**0.0028**	**0.9625**	**0.9946**	**0.2185**	**0.5331**
	LDL_DPA-Var	0.0231	0.6825	**0.0055**	**0.0028**	0.9623	**0.9946**	0.2056	0.5328
	LDL_DPA-Rank	0.0234	0.6945	0.0056	0.0029	0.9617	0.9945	0.1990	0.5325
	LDL_DPA-Var-Rank	0.0236	0.6995	0.0057	0.0029	0.9614	0.9944	-0.0718	0.5259
Cdc	LDL_DPA	**0.0278**	**0.6449**	**0.0069**	**0.0036**	**0.9576**	**0.9933**	**0.2019**	**0.5405**
	LDL_DPA-Var	0.0281	0.6521	0.0070	**0.0036**	0.9571	**0.9933**	0.1998	0.5402
	LDL_DPA-Rank	0.0283	0.6541	0.0073	0.0037	0.9569	0.9930	0.1854	0.5401
	LDL_DPA-Var-Rank	0.0284	0.6551	0.0073	0.0037	0.9569	0.9930	-0.0371	0.5328
Cold	LDL_DPA	**0.0680**	**0.2396**	**0.0121**	**0.0062**	**0.9410**	**0.9886**	**0.2575**	**0.6377**
	LDL_DPA-Var	0.0682	0.2402	0.0122	**0.0062**	0.9408	**0.9886**	0.2545	0.6375
	LDL_DPA-Rank	0.0697	0.2455	0.0127	0.0065	0.9395	0.9880	0.2238	0.6363
	LDL_DPA-Var-Rank	0.0719	0.2528	0.0135	0.0069	0.9376	0.9873	-0.0031	0.6259
Diau	LDL_DPA	**0.0543**	**0.4311**	**0.0131**	**0.0067**	**0.9403**	**0.9880**	**0.4173**	**0.5908**
	LDL_DPA-Var	0.0577	0.4569	0.0144	0.0073	0.9365	0.9867	0.4007	0.5898
	LDL_DPA-Rank	0.0601	0.4750	0.0157	0.0081	0.9339	0.9855	0.3942	0.5896
	LDL_DPA-Var-Rank	0.0605	0.4777	0.0159	0.0082	0.9335	0.9853	-0.0104	0.5710
Dtt	LDL_DPA	**0.0479**	**0.1686**	**0.0062**	**0.0032**	**0.9584**	**0.9941**	**0.2117**	**0.6317**
	LDL_DPA-Var	0.0489	0.1721	0.0065	0.0033	0.9575	0.9939	0.2044	0.6316
	LDL_DPA-Rank	0.0491	0.1727	0.0065	0.0034	0.9574	0.9938	0.1557	0.6296
	LDL_DPA-Var-Rank	0.0495	0.1738	0.0066	0.0034	0.9571	0.9938	0.0520	0.6267
Elu	LDL_DPA	**0.0276**	**0.5812**	**0.0061**	**0.0031**	**0.9590**	**0.9941**	**0.2302**	**0.5424**
	LDL_DPA-Var	0.0279	0.5886	0.0063	0.0032	0.9585	0.9939	0.2184	0.5420
	LDL_DPA-Rank	0.0282	0.5970	0.0064	0.0033	0.9579	0.9938	0.2011	0.5418
	LDL_DPA-Var-Rank	0.0284	0.6019	0.0065	0.0033	0.9575	0.9938	-0.0260	0.5351
Heat	LDL_DPA	**0.0592**	**0.3637**	**0.0126**	**0.0064**	**0.9403**	**0.9880**	**0.1735**	**0.5924**
	LDL_DPA-Var	0.0594	0.3652	0.0127	0.0065	0.9401	0.9879	0.1661	0.5920
	LDL_DPA-Rank	0.0603	0.3699	0.0131	0.0067	0.9392	0.9875	0.1052	0.5899
	LDL_DPA-Var-Rank	0.0609	0.3735	0.0134	0.0068	0.9386	0.9872	-0.0341	0.5805
Spo	LDL_DPA	**0.0819**	**0.5123**	**0.0245**	**0.0126**	**0.9157**	**0.9770**	**0.2101**	**0.6015**
	LDL_DPA-Var	0.0827	0.5172	0.0249	0.0128	0.9148	0.9766	0.1975	0.6006
	LDL_DPA-Rank	0.0830	0.5197	0.0258	0.0131	0.9144	0.9759	0.2031	0.6011
	LDL_DPA-Var-Rank	0.0845	0.5283	0.0271	0.0137	0.9129	0.9747	0.0320	0.5846
Spo5	LDL_DPA	**0.1169**	**0.2826**	**0.0293**	**0.0152**	**0.9087**	**0.9741**	**0.1240**	**0.6752**
	LDL_DPA-Var	0.1179	0.2848	0.0299	0.0155	0.9079	0.9736	0.1072	0.6739
	LDL_DPA-Rank	0.1179	0.2849	0.0299	0.0155	0.9079	0.9736	0.0141	0.6698
	LDL_DPA-Var-Rank	0.1185	0.2864	0.0300	0.0155	0.9075	0.9735	0.0055	0.6650

LDL-LRR and LDL-LCLR both consider label correlation. The former utilizes the ranking relation, and the latter captures the global correlation through low-rank approximation. For LDL-LRR, the parameters λ and β are selected from $10^{\{-6,-5,...,-2,-1\}}$ and $10^{\{-3,-2,...,1,2\}}$, respectively. For LDL-LCLR, the parameters λ_1, λ_2, λ_3, λ_4, and k are set to

TABLE VI
UPPER AND LOWER BOUNDS FOR DPA VALUES ON DIFFERENT TEST DATASETS

Dataset	upper bound	lower bound
SJAFFE	0.6773	0.4894
SBU	0.6763	0.4903
Twitter	0.9399	0.1851
Flickr	0.9104	0.2146
MORPH	0.9385	0.0758
FG-Net	0.9426	0.0717
Movie	0.7602	0.4398
Gene	0.6524	0.3623
Scene	0.8856	0.2255
Alpha	0.5546	0.5009
Cdc	0.5630	0.5037
Cold	0.6629	0.5871
Diau	0.6142	0.5286
Dtt	0.6512	0.5988
Elu	0.5646	0.5069
Heat	0.6234	0.5433
Spo	0.6387	0.5279
Spo5	0.7197	0.6136

0.0001, 0.001, 0.001, 0.001, and 4, respectively. LDLSF used label-specific features and common features simultaneously; the parameters λ_1, λ_2, and λ_3 were selected from $10^{\{-6,-5,\ldots,-2,-1\}}$, respectively, and ρ was set to 10^{-3}. LALOT adopts the optimal transport distance as the loss function, and the tradeoff parameter C and the entropic regularization coefficient λ are set to 200 and 0.2, respectively.

In LDL_DPA, we select parameters λ_1, λ_2, and λ_3 from $10^{\{-9,-8,\ldots,-1,0\}}$, $10^{\{-7,-6,\ldots,-3,-2\}}$, and $10^{\{-4,-3,\ldots,0,1\}}$ respectively. The maximum iteration is set to 400 and the initialization of other variables is all zero.

D. Results and Discussion

We conducted ten times fivefold cross-validation on each dataset. The experimental results are presented in the form of "mean±std" in Tables III and IV.

In general, our proposed LDL_DPA achieves optimal results on most metrics and suboptimal results for the rest. Except on the Flickr dataset, LDL_DPA significantly outperforms the other algorithms used in our experiments in terms of two ranking metrics. We find that LDLSF performs relatively better on the Twitter and Flickr datasets, because its label feature extraction strategy is suitable for such datasets where many labels are associated with zero description degree.

Analysis of the observations shows that the first three methods, that is, LDL_DPA, LDL-LRR, and LDL-LCLR, which consider ranking relation or other types of label correlations, outperform other methods on most metrics. Moreover, the performance of these three methods is more stable and, unlike AA-kNN, does not fluctuate. Ranking dominant methods LDL_DPA and LDL-LRR clearly have advantages on two ranking metrics and can thus, in turn, verify the validity of our introduced ranking metric DPA. In addition, LALOT shows clear advantages on the Gene and MORPH datasets. Therefore, we speculate that utilizing the optimal transmission distance can help the algorithm improve its performance on datasets with larger sample sizes.

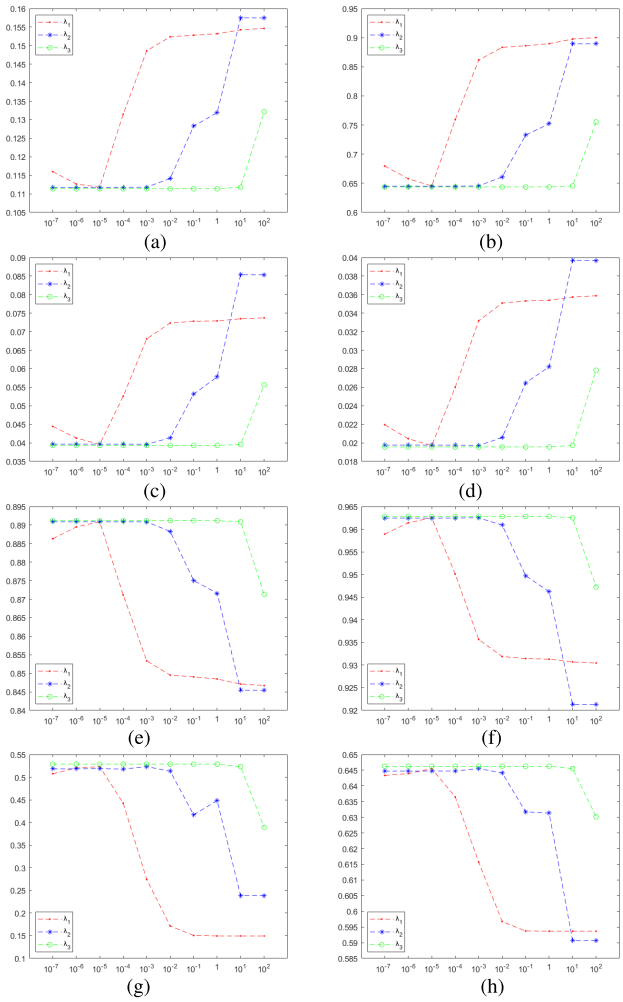

Fig. 8. Influence of $\lambda_1, \lambda_2, \lambda_3$ on SJAFFE. (a) Euclidean ↓. (b) Canberra ↓. (c) K–L ↓. (d) Squared ↓. (e) Intersection ↑. (f) Cosine ↑. (g) Spearman ↑. (h) DPA ↑.

The results in Table IV show that multiple algorithms can reach the optimal value on the Yeast datasets mainly because of the small feature dimension of the datasets; additionally, there is limited space for optimization through learning. Moreover, some existing algorithms have learned well from a limited number of features. For example, these algorithms have achieved a similarity of over 0.97 in terms of the Cosine metric, even more than 0.99 on some Yeast datasets.

1) Ablation Study: We conduct ablation studies to further analyze how our proposal works. The "LDL_DPA-Var" results are obtained by setting the balance factors λ_3 to 0, while keeping the other two parameters the same as those of "LDL_DPA." Similarly, the "LDL_DPA-Rank" results are obtained by setting the balance factor λ_2 to 0, while keeping the other two parameters the same as those of "LDL_DPA." Then, we derive an "LDL_DPA-Var-Rank" version (with $\lambda_2, \lambda_3 = 0$), which involves only basic distance minimization, while keeping λ_1 unchanged.

As shown in Table V, the addition of the DPA-based ranking term can bring significant improvements to the model. The improvement achieved on the Yeast datasets seems to be not significant enough numerically because of the limited range of the DPA itself for this type of dataset. We performed an

TABLE VII

EXPERIMENTAL RESULTS ON DIFFERENT MEASURES UNDER THE PAIRWISE TWO-TAILED t-TEST WITH A 0.05 SIGNIFICANCE LEVEL, WHICH ARE RECORDED IN THE FORM OF WIN/TIE/LOSS NUMBER, SHOWING THE ADVANTAGE OF LDL_DPA COMPARED TO OTHER LDL ALGORITHMS INTUITIVELY

Measure	LDL-LRR	LDL-LCLR	LDLSF	LDLOT	BFGS-LLD	AA-kNN	PT-Bayes
Euclidean	10/8/0	12/6/0	11/7/0	17/1/0	11/7/0	17/1/0	18/0/0
Canberra	12/6/0	11/7/0	8/8/2	11/6/1	13/5/0	9/6/3	16/2/0
K-L	9/9/0	10/8/0	11/7/0	15/2/1	7/11/0	18/0/0	18/0/0
Squared	10/8/0	11/6/1	12/6/0	15/2/1	11/7/0	17/1/0	17/1/0
Intersection	12/6/0	12/5/1	13/5/0	18/0/0	12/6/0	15/2/1	18/0/0
Cosine	13/5/0	14/4/0	12/5/1	17/1/0	13/5/0	17/1/0	18/0/0
Spearman	12/6/0	13/5/0	11/6/1	16/2/0	16/2/0	18/0/0	18/0/0
DPA	15/3/0	11/7/0	9/8/1	16/2/0	16/2/0	18/0/0	18/0/0

TABLE VIII

COMPARISON RESULTS WITH LDL-HR AND LDLM ON DIFFERENT METRICS, WHERE "↓" INDICATES "THE SMALLER THE BETTER," "↑" INDICATES "THE LARGER THE BETTER." THE BEST RESULTS AMONG THREE VERSIONS ARE HIGHLIGHTED IN BOLDFACE

Data	Algorithm	0/1 loss ↓	ep ↓	Euclidean ↓	Canberra ↓	K-L ↓	Squared ↓	Intersection ↑	Cosine ↑	Spearman ↑	DPA ↑
SJAFFE	LDL_DPA	0.4698	**0.7584**	**0.1118**	**0.6453**	**0.0396**	**0.0197**	**0.8909**	**0.9625**	**0.5234**	**0.6455**
	LDL-HR	**0.4505**	0.7657	0.1363	0.8218	0.0599	0.0298	0.8637	0.9460	0.3458	0.6304
	LDLM	0.4835	0.7676	0.1383	0.8201	0.0601	0.0299	0.8627	0.9449	0.3346	0.6297
SBU	LDL_DPA	**0.4980**	**0.7548**	**0.1344**	**0.7272**	**0.0541**	**0.0266**	**0.8709**	**0.9469**	**0.4042**	**0.6375**
	LDL-HR	0.5604	0.7689	0.1557	0.8419	0.0720	0.0343	0.8492	0.9294	0.2857	0.6280
	LDLM	0.5636	0.7697	0.1556	0.8416	0.0719	0.0343	0.8493	0.9295	0.2848	0.6277
Twitter	LDL_DPA	**0.2761**	**0.4256**	**0.3778**	**6.1963**	**0.6337**	**0.4131**	**0.6235**	**0.8236**	**0.6047**	**0.8529**
	LDL-HR	0.2809	0.4296	0.6219	6.1856	1.2290	0.8621	0.3581	0.5421	0.0011	0.8449
	LDLM	0.2805	0.4293	0.6220	6.1856	1.2294	0.8623	0.3581	0.5419	0.5417	0.8449
Flickr	LDL_DPA	0.4273	**0.5637**	**0.3909**	**5.3539**	**0.6396**	**0.4372**	**0.5950**	**0.7767**	**0.5520**	**0.7980**
	LDL-HR	**0.4258**	0.5651	0.5335	5.4867	0.9933	0.7078	0.4273	0.5904	0.4732	0.7915
	LDLM	0.4268	0.5654	0.5336	5.4870	0.9936	0.7080	0.4272	0.5902	0.4731	0.7915
MORPH	LDL_DPA	0.8773	**0.9233**	**0.1919**	**57.9896**	**0.6432**	**0.3961**	**0.5742**	**0.7402**	**0.8847**	**0.8769**
	LDL-HR	0.8065	0.9261	0.2289	59.8373	2.3132	0.6211	0.4455	0.6617	0.7021	0.8418
	LDLM	**0.8062**	0.9260	0.2290	59.9704	2.3756	0.6204	0.4463	0.6613	0.6927	0.8413
FG-Net	LDL_DPA	**0.8723**	**0.9163**	**0.1989**	**59.0951**	**0.8011**	**0.4091**	**0.5902**	**0.7122**	**0.8224**	**0.8662**
	LDL-HR	0.8972	0.9232	0.2320	60.5533	1.0419	0.6349	0.4516	0.6471	0.6860	0.8206
	LDLM	0.9022	0.9231	0.2324	60.5504	1.0433	0.6363	0.4510	0.6467	0.6844	0.8204
Movie	LDL_DPA	**0.4164**	**0.6763**	**0.1640**	**0.9818**	**0.0959**	**0.0498**	**0.8381**	**0.9368**	**0.7120**	**0.7179**
	LDL-HR	0.4610	0.6867	0.2322	1.3008	0.1565	0.0861	0.7718	0.8920	0.6100	0.7040
	LDLM	0.4629	0.6869	0.2318	1.2987	0.1561	0.0858	0.7723	0.8924	0.6092	0.7039
Gene	LDL_DPA	0.9598	**0.9819**	**0.0864**	**14.4347**	**0.2360**	**0.1069**	**0.7846**	**0.8347**	**0.2070**	**0.5292**
	LDL-HR	0.9359	0.9823	0.1144	23.1675	0.5203	0.2030	0.6731	0.7401	-0.0342	0.5152
	LDLM	**0.9358**	0.9824	0.1143	23.1247	0.5277	0.2028	0.6736	0.7406	-0.0342	0.5152
Scene	LDL_DPA	0.6120	**0.6601**	**0.4179**	**6.6560**	**0.7741**	**0.4952**	**0.5554**	**0.7351**	**0.5761**	**0.7811**
	LDL-HR	**0.4640**	0.6751	0.5154	7.0435	1.1431	0.7623	0.3841	0.6006	0.2589	0.6852
	LDLM	0.4675	0.6762	0.5169	7.0556	1.1623	0.7680	0.3822	0.5973	0.2534	0.6830
Alpha	LDL_DPA	0.8949	**0.9426**	**0.0230**	**0.6788**	**0.0055**	**0.0028**	**0.9625**	**0.9946**	**0.2185**	**0.5331**
	LDL-HR	**0.8122**	0.9446	0.0410	1.2701	0.0167	0.0084	0.9309	0.9843	0.0671	0.5293
	LDLM	0.8126	0.9446	0.0410	1.2709	0.0167	0.0084	0.9308	0.9843	0.0672	0.5293
Cdc	LDL_DPA	**0.8243**	**0.9287**	**0.0278**	**0.6449**	**0.0069**	**0.0036**	**0.9576**	**0.9933**	**0.2019**	**0.5405**
	LDL-HR	0.8329	0.9303	0.0422	1.0098	0.0152	0.0076	0.9345	0.9860	0.1359	0.5379
	LDLM	0.8337	0.9303	0.0422	1.0104	0.0152	0.0076	0.9344	0.9860	0.1359	0.5379
Cold	LDL_DPA	**0.5809**	**0.7298**	**0.0680**	**0.2396**	**0.0121**	**0.0062**	**0.9410**	**0.9886**	**0.2575**	**0.6377**
	LDL-HR	0.5830	0.7302	0.0694	0.2445	0.0127	0.0065	0.9397	0.9880	0.2529	0.6367
	LDLM	0.5834	0.7302	0.0694	0.2445	0.0127	0.0065	0.9397	0.9880	0.2537	0.6368
Diau	LDL_DPA	0.7083	**0.8429**	**0.0543**	**0.4311**	**0.0131**	**0.0067**	**0.9403**	**0.9880**	**0.4173**	**0.5908**
	LDL-HR	0.6803	0.8456	0.0576	0.4585	0.0148	0.0077	0.9364	0.9864	0.3642	0.5860
	LDLM	**0.6791**	0.8455	0.0576	0.4586	0.0148	0.0077	0.9364	0.9864	0.3646	0.5861
Dtt	LDL_DPA	**0.6296**	**0.7415**	**0.0479**	**0.1686**	**0.0062**	**0.0032**	**0.9584**	**0.9941**	**0.2117**	**0.6317**
	LDL-HR	0.6381	0.7418	0.0486	0.1707	0.0064	0.0033	0.9579	0.9940	0.1619	0.6306
	LDLM	0.6385	0.7418	0.0486	0.1707	0.0064	0.0033	0.9579	0.9940	0.1612	0.6305
Elu	LDL_DPA	0.9059	**0.9261**	**0.0276**	**0.5812**	**0.0061**	**0.0031**	**0.9590**	**0.9941**	**0.2302**	**0.5424**
	LDL-HR	0.8308	0.9285	0.0366	0.7732	0.0103	0.0052	0.9452	0.9900	0.0793	0.5380
	LDLM	**0.8304**	0.9285	0.0366	0.7734	0.0103	0.0052	0.9453	0.9900	0.0795	0.5380
Heat	LDL_DPA	0.6957	**0.8239**	**0.0592**	**0.3637**	**0.0126**	**0.0064**	**0.9403**	**0.9880**	**0.1735**	**0.5924**
	LDL-HR	**0.6824**	0.8251	0.0610	0.3755	0.0132	0.0068	0.9383	0.9874	0.0966	0.5892
	LDLM	0.6828	0.8252	0.0610	0.3756	0.0132	0.0068	0.9383	0.9874	0.0969	0.5892
Spo	LDL_DPA	0.5513	**0.8102**	**0.0819**	**0.5123**	**0.0245**	**0.0126**	**0.9157**	**0.9770**	**0.2101**	**0.6015**
	LDL-HR	**0.5481**	**0.8102**	0.0831	0.5201	0.0257	0.0131	0.9143	0.9760	0.1449	0.5972
	LDLM	**0.5481**	**0.8102**	0.0831	0.5201	0.0257	0.0131	0.9143	0.9760	0.1449	0.5972
Spo5	LDL_DPA	0.5643	**0.6532**	**0.1169**	**0.2826**	**0.0293**	**0.0152**	**0.9087**	**0.9741**	**0.1240**	**0.6752**
	LDL-HR	**0.5363**	0.6539	0.1176	0.2841	0.0296	0.0154	0.9082	0.9738	0.0481	0.6721
	LDLM	0.5371	0.6539	0.1176	0.2841	0.0296	0.0154	0.9082	0.9738	0.0480	0.6721

experimental analysis of the upper and lower bounds of the DPA values on our test sets. Because the DPA reflects the ranking performance of algorithms, the DPA is taken to be maximum when the predicted label ranking is exactly the same as the ground truth; otherwise, the DPA is taken to be the minimum. The upper and lower bounds of the DPA values on

different datasets are shown in Table VI. The table shows that the range of the DPA is small on the Yeast datasets because the label distributions in these datasets are mostly biased toward uniform distribution, whose ranking relationship is relatively not obvious enough.

Moreover, the generalization effect of the proposed LDL_DPA can be further enhanced after considering variance distance. Comparing the results of LDL_DPA-Var and LDL_DPA-Rank, although the improvement due to adding the variance distance is not as large as that due to considering ranking-oriented weighting in existing metrics, optimizing the algorithm from two angles at the same time seems to be beneficial for further improvement. In addition, optimizing the algorithm from two angles at the same time can bring the shape of the predicted distribution closer to the ground truth. Such improvement cannot be directly measured by existing metrics, but we believe that the shape is an important property of the label distribution and that we are working on it.

2) Significance Test: To verify whether there are significant differences between our proposed LDL_DPA and other existing LDL algorithms, we conduct a pairwise two-tailed t-test with a significance level of 0.05 on eight evaluation metrics. Due to page limitations, only the results of the SJAFFE dataset are presented. As shown in Table VII, the comparison results ruled out the possibility that the advantage of our algorithm is the result of sampling difference.

3) Parameter Sensitivity Analysis: To investigate the sensitivity of the parameters λ_1, λ_2, and λ_3, we tune them from $10^{\{-9,-8,\ldots,1,0\}}$. Due to page limitations, only the results of the SJAFFE dataset are presented. Fig. 8 shows the results on eight metrics by using line graphs. According to our observations, the performance of LDL_DPA is relatively stable if the parameters fall in a certain range (i.e., $\lambda_1 \in [10^{-6}, 10^{-5}]$, $\lambda_2 \in [10^{-7}, 10^{-2}]$, and $\lambda_3 \in [10^{-7}, 10^{1}]$). In addition, the results are less affected by changes in the values of λ_2 and λ_3 than in the value of λ_1 and can thus reflect the robustness of our newly introduced terms.

E. Comparison With Classification Methods

Some state-of-the-art LDL methods worked on the problem of objective mismatch between LDL and classification tasks. These methods took the label with the highest description degree in LDL, that is, the optimal label, as the classification objective and were devoted to improving the prediction accuracy of the optimal label in LDL. LDL-HR [39] learns the optimal label and other labels separately, while LDLM maintains the margin of the top-2 labels and the remaining labels. The 0/1 loss and error probability [9] are used to evaluate classification performance. Such a classification task can be understood as an LDL task with only one key label. To verify the applicability of our method to this case, we conduct comparative experiments on the three methods in terms of the classification metrics and LDL metrics mentioned in Section V-B. The results are shown in Table VIII, where "ep" indicates the error probability.

Table VIII shows that our method achieved the best results on all LDL metrics and error probability. For the 0/1 loss, LDL_DPA ranks first on seven datasets, while LDL-HR and LDLM rank first on seven and five datasets, respectively.

TABLE IX
EXPERIMENTAL RESULTS ON PREDICTING KEY LABELS

Data	Algorithm	$\mathrm{Jac}_{0.5}$	$\mathrm{Jac}_{0.6}$	$\mathrm{Jac}_{0.7}$	$\mathrm{Jac}_{0.8}$
SJAFFE	LDL_DPA	**0.4295**	**0.3679**	0.3043	**0.2099**
	LDL-HR	0.5275	0.4690	0.3710	0.2629
	LDLM	0.5177	0.4778	0.3804	0.2786
	LDL-LRR	0.4646	0.3696	**0.2966**	0.2104
SBU	LDL_DPA	**0.4931**	**0.4186**	0.3593	**0.1953**
	LDL-HR	0.6222	0.4927	**0.3168**	0.2109
	LDLM	0.5635	0.4447	0.3408	0.2026
	LDL-LRR	0.5830	0.4716	0.3994	0.2010
Twitter	LDL_DPA	0.3497	**0.4084**	**0.4723**	**0.5132**
	LDL-HR	0.6993	0.7435	0.7291	0.6889
	LDLM	0.6983	0.7422	0.7284	0.6886
	LDL-LRR	**0.3390**	0.4466	0.5337	0.5584
Flickr	LDL_DPA	**0.5035**	**0.5210**	**0.5405**	**0.5513**
	LDL-HR	0.7008	0.6958	0.6863	0.6505
	LDLM	0.6987	0.6958	0.6858	0.6492
	LDL-LRR	0.6006	0.6222	0.6301	0.6118
MORPH	LDL_DPA	**0.6271**	**0.5485**	**0.5633**	**0.5116**
	LDL-HR	0.7114	0.6518	0.6908	0.6688
	LDLM	0.7013	0.6366	0.6728	0.6451
	LDL-LRR	0.8522	0.8007	0.8033	0.7586
FG-Net	LDL_DPA	**0.6338**	**0.5791**	**0.5439**	**0.4939**
	LDL-HR	0.7056	0.6685	0.6739	0.6737
	LDLM	0.8641	0.8496	0.8437	0.8204
	LDL-LRR	0.8044	0.7766	0.7666	0.7329
Movie	LDL_DPA	**0.3141**	**0.3208**	**0.2841**	**0.2542**
	LDL-HR	0.4229	0.4103	0.3760	0.3085
	LDLM	0.4221	0.4105	0.3760	0.3094
	LDL-LRR	0.3219	0.3301	0.2970	0.2646
Gene	LDL_DPA	**0.7274**	**0.6468**	**0.5508**	**0.4322**
	LDL-HR	0.8004	0.7557	0.6953	0.6049
	LDLM	0.8004	0.7558	0.6951	0.6045
	LDL-LRR	0.7541	0.6731	0.5733	0.4468
Scene	LDL_DPA	**0.5661**	**0.5394**	**0.5153**	**0.4961**
	LDL-HR	0.6954	0.7102	0.7161	0.7020
	LDLM	0.6879	0.7042	0.7120	0.7001
	LDL-LRR	0.5996	0.5799	0.5638	0.5426
Alpha	LDL_DPA	**0.5832**	**0.4854**	**0.3776**	**0.2541**
	LDL-HR	0.6524	0.5712	0.4704	0.3399
	LDLM	0.6523	0.5714	0.4703	0.3397
	LDL-LRR	0.5935	0.4959	0.3881	0.2628
Cdc	LDL_DPA	**0.5644**	**0.4961**	**0.3923**	**0.2977**
	LDL-HR	0.6348	0.5438	0.4569	0.3200
	LDLM	0.6348	0.5437	0.4570	0.3198
	LDL-LRR	0.5861	0.5130	0.4018	0.3032
Cold	LDL_DPA	0.5118	**0.3346**	**0.3170**	**0.0897**
	LDL-HR	0.5165	0.3497	0.3360	**0.0897**
	LDLM	0.5158	0.3497	0.3362	**0.0897**
	LDL-LRR	**0.5106**	0.3459	0.3329	**0.0897**
Diau	LDL_DPA	**0.4186**	0.4060	0.3395	0.1917
	LDL-HR	0.4361	0.4191	0.3933	0.2638
	LDLM	0.4364	0.4191	0.3933	0.2638
	LDL-LRR	0.4228	**0.4044**	**0.3341**	**0.1901**
Dtt	LDL_DPA	**0.5264**	**0.3127**	**0.3100**	**0.0366**
	LDL-HR	0.5280	0.3503	0.3473	**0.0366**
	LDLM	0.5282	0.3513	0.3483	**0.0366**
	LDL-LRR	0.5432	0.3376	0.3351	**0.0366**
Elu	LDL_DPA	**0.5624**	**0.4635**	**0.3666**	**0.2489**
	LDL-HR	0.6233	0.5472	0.4303	0.3378
	LDLM	0.6232	0.5475	0.4304	0.3376
	LDL-LRR	0.5702	0.4741	0.3753	0.2527
Heat	LDL_DPA	**0.5648**	**0.4396**	**0.3409**	**0.2432**
	LDL-HR	0.6151	0.4885	0.3686	0.2830
	LDLM	0.6143	0.4875	0.3683	0.2830
	LDL-LRR	0.5754	0.4498	0.3443	0.2486
Spo	LDL_DPA	**0.5269**	**0.4432**	**0.3590**	**0.2675**
	LDL-HR	0.6079	0.4737	0.3637	0.2761
	LDLM	0.6079	0.4743	0.3637	0.2759
	LDL-LRR	0.5476	0.4644	0.3617	0.2689
Spo5	LDL_DPA	0.4074	0.3778	0.2675	**0.0522**
	LDL-HR	0.5063	0.5047	**0.2644**	**0.0522**
	LDLM	0.5052	0.5033	**0.2644**	**0.0522**
	LDL-LRR	**0.4016**	**0.3720**	0.2650	**0.0522**

Comparing the formulas of the two classification metrics shows that 0/1 loss counts the number of prediction errors of the optimal label, while error probability is related to the description degree of the predicted optimal label. That is, 0/1 loss can reflect only the classification performance, while error

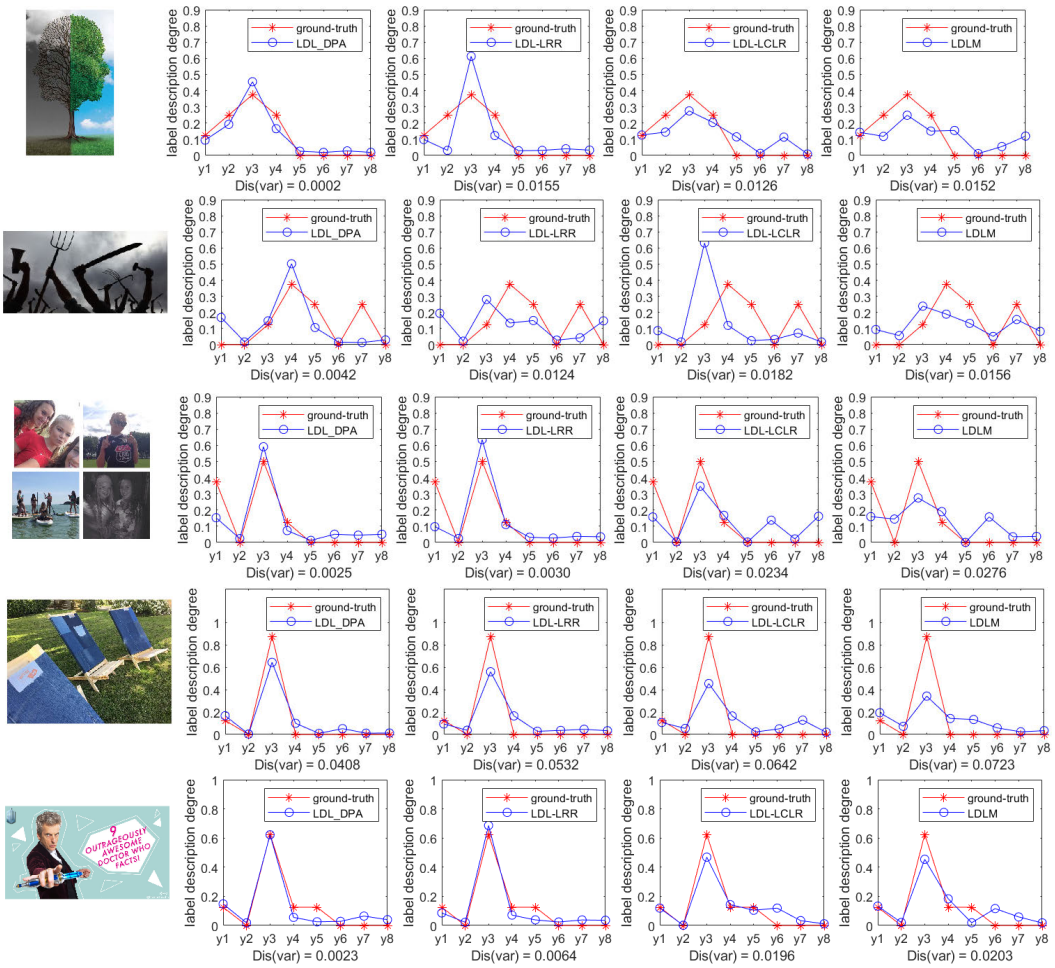

Fig. 9. Several sample visualizations of the prediction results of different algorithms on the Twitter dataset, where Dis(var) denotes the variance distance between the predicted distribution and the ground truth.

probability can also reflect the generalization ability of the algorithm. Our algorithm is essentially an LDL task, which requires not only predicting the optimal label well, but also the corresponding description degrees. Therefore, such results are overall quite considerable. In conclusion, the experimental results demonstrate that our algorithm has better robustness than LDL-HR and LDLM.

F. Performance in Predicting Key Labels

To verify the prediction effectiveness of the proposed algorithm for key labels, we conduct comparison experiments with state-of-the-art algorithms (LDL-HR, LDLM, and LDL-LRR) that consider key labels. We compare the overlap between the predicted and the ground-truth key labels by constructing a key-label set, which consists of the labels with higher description degrees. First, the labels are sorted in descending order of description degrees, and the initial set of key labels contains only the optimal label. Then, the labels are added sequentially until the sum of the description degrees of the key-label set reaches the predefined value, which we take as 0.5, 0.6, 0.7, and 0.8 in the experiments. Finally, we calculate the Jaccard distance of the predicted and the ground-truth key-label sets according to the following equation:

$$\mathrm{Jac}_{\mathrm{val}} = \mathrm{Jaccard}(\mathbf{y}_{\mathrm{val}}, \widehat{\mathbf{y}_{\mathrm{val}}}) = 1 - \frac{|\mathbf{y}_{\mathrm{val}} \cap \widehat{\mathbf{y}_{\mathrm{val}}}|}{|\mathbf{y}_{\mathrm{val}} \cup \widehat{\mathbf{y}_{\mathrm{val}}}|} \tag{15}$$

where $\mathbf{y}_{\mathrm{val}}$ and $\widehat{\mathbf{y}_{\mathrm{val}}}$ denote the ground truth and the predicted key-label sets, respectively, when the predefined value is val. The experimental results are shown in Table IX, where $\mathrm{Jac}_{0.5}$, $\mathrm{Jac}_{0.6}$, $\mathrm{Jac}_{0.7}$, and $\mathrm{Jac}_{0.8}$ indicate the Jaccard distance when the predefined value ranges from 0.5 to 0.8. The smaller the value is, the better the algorithm.

According to the experimental results, our algorithm ranks first in most cases. Especially on datasets with more labels, a clear advantage is achieved, such as in the case of MORPH, FG-Net, and Gene, which contain 62, 70, and 68 labels, respectively. In addition, we find that LDL-LRR, focusing on global label ranking, outperforms LDL-HR and LDLM, in general, which mainly learn the first two labels because in real-world applications, the number of key labels is often greater than two, and optimizing ranking is a more widely applicable method.

G. Performance in Improving Label Distinguishability

To verify the performance of LDL_DPA in maintaining the shape of the label and improving the label distinguishability, we have chosen some examples on Twitter and made a visual comparison of the prediction results of several state-of-the-art algorithms and the prediction results of LDL_DPA. As shown in Fig. 9, the variance of the prediction of LDL_DPA is closer to the ground truth than other algorithms, while the

distribution with smaller variance distance tends to be closer in shape to the ground truth. Compared to LDL_DPA and LDL-LRR, the other two algorithms tend to predict a more even distribution, which is detrimental to improving label distinguishability. The suboptimal performance of LDL-LRR suggests that maintaining label ranking relation also plays a role in enhancing label distinguishability.

VI. Conclusion

LDL is an effective method to address label polysemy. To measure LDL problems preferable to key labels, we introduce a novel ranking metric, DPA. We further propose the LDL_DPA algorithm (which is based on DPA) to better solve such problems. In addition, to improve the distinctness of key labels, we attempt to minimize the variance distance during learning; minimizing the variance can maintain the shape of label distributions to some extent. The experimental results on several datasets validate the effectiveness of our proposal. In the future, we will work on refining the LDL metrics so that we can evaluate the model from more perspectives and help make it more applicable to real-world applications.

References

[1] X. Geng, "Label distribution learning," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 7, pp. 1734–1748, Jul. 2016.

[2] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *Int. J. Data Warehousing Mining*, vol. 3, no. 3, pp. 1–13, 2007.

[3] X. Jia, X. Shen, W. Li, Y. Lu, and J. Zhu, "Label distribution learning by maintaining label ranking relation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1695–1707, Jun. 2023.

[4] X. Jia, W. Li, J. Liu, and Y. Zhang, "Label distribution learning by exploiting label correlations," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 3310–3317.

[5] A. Akbari, M. Awais, S. Fatemifar, S. S. Khalid, and J. Kittler, "A novel ground metric for optimal transport-based chronological age estimation," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 9986–9999, Oct. 2021.

[6] X. Jia, Z. Li, X. Zheng, W. Li, and S.-J. Huang, "Label distribution learning with label correlations on local samples," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1619–1631, Apr. 2021.

[7] T. Ren, X. Jia, W. Li, and S. Zhao, "Label distribution learning with label correlations via low-rank approximation," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 3325–3331.

[8] C. Tan, S. Chen, X. Geng, and G. Ji, "A label distribution manifold learning algorithm," *Pattern Recognit.*, vol. 135, Mar. 2023, Art. no. 109112.

[9] J. Wang and X. Geng, "Classification with label distribution learning," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 3712–3718.

[10] J. Wang and X. Geng, "Label distribution learning machine," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10749–10759.

[11] C. Burges et al., "Learning to rank using gradient descent," in *Proc. 22nd Int. Conf. Mach. Learn. (ICML)*, 2005, pp. 89–96.

[12] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.*, vol. 100, nos. 3–4, pp. 441–471, 1987.

[13] E. J. Weyuker, T. J. Ostrand, and R. M. Bell, "Comparing the effectiveness of several modeling methods for fault prediction," *Empirical Softw. Eng.*, vol. 15, no. 3, pp. 277–295, Jun. 2010.

[14] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2401–2412, Oct. 2013.

[15] J. Wang and X. Geng, "Label distribution learning by exploiting label distribution manifold," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 2, pp. 839–852, Feb. 2023.

[16] W. Qian, Y. Xiong, J. Yang, and W. Shu, "Feature selection for label distribution learning via feature similarity and label correlation," *Inf. Sci.*, vol. 582, pp. 38–59, Jan. 2022.

[17] P. Zhao and Z. Zhou, "Label distribution learning by optimal transport," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 4506–4513.

[18] X. Geng and Y. Xia, "Head pose estimation based on multivariate label distribution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1837–1842.

[19] Y. Zhou, H. Xue, and X. Geng, "Emotion distribution recognition from facial expressions," in *Proc. 23rd ACM Int. Conf. Multimedia*, Oct. 2015, pp. 1247–1250.

[20] D. Zhou, X. Zhang, Y. Zhou, Q. Zhao, and X. Geng, "Emotion distribution learning from texts," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 638–647.

[21] S. Zhang, C. Gong, and E. Choi, "Capturing label distribution: A case study in NLI," 2021, *arXiv:2102.06859*.

[22] X. Wen et al., "Adaptive variance based label distribution learning for facial age estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 379–395.

[23] N. Ohlsson and H. Alberg, "Predicting fault-prone software modules in telephone switches," *IEEE Trans. Softw. Eng.*, vol. 22, no. 12, pp. 886–894, Dec. 1996.

[24] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *Int. J. Math. Model. Meth. Appl. Sci.*, vol. 1, pp. 300–307, Jan. 2007.

[25] Q. Zhao and X. Geng, "Selection of target function in label distribution learning," *Frontiers Comput. Sci. Technol.*, vol. 11, no. 5, pp. 708–719, 2017.

[26] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," *Comput. Linguistics*, vol. 22, no. 1, pp. 39–71, Mar. 1996.

[27] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, no. 1, pp. 503–528, 1989.

[28] M. Lyons, M. Kamachi, and J. Gyoba, "The Japanese female facial expression (JAFFE) database," in *Proc. Int. Conf. Autom. Face Gesture Recognit.*, 1998, pp. 14–16.

[29] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, 2006, pp. 211–216.

[30] J. Yang, M. Sun, and X. Sun, "Learning visual sentiment distributions via augmented conditional probability neural network," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 224–230.

[31] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, 2006, pp. 341–345.

[32] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.

[33] X. Geng and P. Hou, "Pre-release prediction of crowd opinion on movies by label distribution learning," in *Proc. AAAI Conf. Artif. Intell.*, 2015, pp. 3511–3517.

[34] X. Geng and L. Luo, "Multilabel ranking with inconsistent rankers," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3742–3747.

[35] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proc. Nat. Acad. Sci. USA*, vol. 95, no. 25, pp. 14863–14868, 1999.

[36] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 112–119.

[37] S. Yan, D. Xu, B. Zhang, and H.-J. Zhang, "Graph embedding: A general framework for dimensionality reduction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 830–837.

[38] T. Ren, X. Jia, W. Li, L. Chen, and Z. Li, "Label distribution learning with label-specific features," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 3318–3324.

[39] J. Wang and X. Geng, "Learn the highest label and rest label description degrees," in *Proc. 30th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 3097–3103.

Xiuyi Jia (Member, IEEE) received the Ph.D. degree in computer science from Nanjing University, Nanjing, China, in 2011.

He was a Visiting Scholar with the University of Regina, Regina, SK, Canada, from 2008 to 2009. He was a Visiting Scholar with the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, from 2016 to 2017. He is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. His recent research interests include machine learning, granular computing, and computer vision.

Tian Qin is currently pursuing the master's degree with the Nanjing University of Science and Technology, Nanjing, China.

Her research interests include machine learning and data mining.

Yunan Lu is currently pursuing the Ph.D. degree with the Nanjing University of Science and Technology, Nanjing, China.

His research interests include machine learning and data mining.

Weiwei Li received the Ph.D. degree in computer science from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2016.

She is currently an Associate Professor with the Nanjing University of Aeronautics and Astronautics. Her research interests include machine learning, software data mining, and knowledge engineering.