

# 一种基于全局和局部标记相关性的 多标记分类算法

朱赛赛, 贾修一, 李泽超

(南京理工大学计算机科学与工程学院, 江苏南京 210094)

**摘要:** 多标记学习用于处理一个示例同时与多个类别标记相关的问题. 在多标记学习中, 标记相关性能够显著提升学习算法的性能. 大多数现有的多标记学习算法在利用标记的相关性时, 要么只使用被所有示例所共享的全局标记相关性, 要么就使用局部标记相关性, 它们认为不同簇中的示例应该存在不同的标记相关性. 本文中, 我们提出了一种同时利用全局和局部标记相关性的多标记学习算法, 从而为学习进程提供更全面的标记信息. 在计算全局和局部标记相关性时, 我们使用了余弦相似性来获取不同标记之间的正相关性和负相关性, 这样有助于我们进一步实现更可靠的多标记学习. 我们在多种类型的数据集上进行了广泛的对比实验来验证所提算法的有效性. 实验结果表明, 该算法显著优于大多数对比算法, 展现出其在多标记学习中的突出性能.

**关键词:** 多标记学习; 标记相关性; 余弦相似性

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 0372-2112 (2020)12-2345-07

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.3969/j.issn.0372-2112.2020.12.009

## Exploiting Global and Local Label Correlations for Multi-label Classification

ZHU Sai-sai, JIA Xiu-yi, LI Ze-chao

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China)

**Abstract:** Multi-label learning deals with the problem where each instance has a set of class labels simultaneously. In multi-label learning, label correlations have shown promising strength in improving multi-label learning. Most of the existing multi-label learning algorithms exploited either global label correlations shared among all instances, or local label correlations varied across different clusters of instances. In this study, we propose a novel multi-label learning method by simultaneously taking into account both the global and local label correlations to capture more comprehensive label information during the learning process. To calculating global and local label correlations, we utilize cosine similarity to obtain positive and negative correlations between different labels, which helps us to further achieve more reliable multi-label learning. We implemented extensive experimental comparisons based on various data sets to validate the effectiveness of our algorithm. The experimental results show that the proposed algorithm significantly outperforms most of the state-of-the-art approaches, demonstrating its prominent performance for multi-label learning.

**Key words:** multi-label learning; label correlations; cosine similarity

### 1 引言

在机器学习领域中, 多标记学习框架被提出用于解决一个示例同时与多个标记相关的问题<sup>[1,2]</sup>. 多标记学习的目标是通过训练数据学习出一个能够给未见示例分配一组合适的标记集合的模型. 近年来, 多标记学习被广泛应用于文本分类<sup>[3]</sup>、图像自动标注<sup>[4]</sup>、社

网络分析<sup>[5]</sup>、音乐情感分类<sup>[6]</sup>等多个领域.

解决多标记学习问题的最直接和最简单的方法是其分解为多个单标记分类问题, 然后直接使用传统的分类算法为每个标记训练一个二分类模型. 然而, 这种方式完全忽略了不同标记之间存在的相关性, 近年来的许多研究表明, 利用标记相关性能够显著提升多标记学习算法的分类性能<sup>[7]</sup>. 因此, 在多标记学习中,

收稿日期: 2019-09-23; 修回日期: 2020-05-02; 责任编辑: 梅志强

基金项目: 国家重点研发计划资助 (No. 2019YFB1707300); 国家自然科学基金项目 (No. 61773208, No. 61772275); 江苏省自然科学基金项目 (No. BK20191287); 中央高校基本科研业务费专项资金 (No. 30920021131)

研究如何有效地探索并利用标记相关性已经成为最具吸引力的研究方向之一。

到目前为止,学者们已经提出了很多基于利用二阶或者高阶标记相关性的多标记学习算法.但是这些算法大多数使用的都是全局标记相关性,它们认为标记间的相关性是被所有示例共享的<sup>[8,9]</sup>.然而,只考虑这种全局的标记相关性很难正确捕获到一些仅被部分示例所共享的局部相关性.为了解决这个问题,许多研究开始致力于如何利用多标记学习中的局部标记相关性<sup>[10-12]</sup>.在这些研究中,它们大多都是使用某种聚类算法估计不同聚类簇中示例所共享的不同的局部标记相关性.然而,局部标记相关性的估计高度依赖于聚类的准确度.较差的聚类结果可能导致对局部标记相关性的估计不准确,甚至有可能破坏一些明显的全局标记相关性.因此,将两者结合起来而不是单独使用全局或者局部标记相关性或许能够提供一种更合理有效的多标记学习策略.

文献[7]中首次提出了在多标记学习中同时利用全局和局部标记相关性的方法 GLOCAL. 他们的实验结果表明了同时利用全局和局部标记相关性的有效性和合理性.然而,GLOCAL 是基于低秩建模和流形正则化的,这使得模型的学习较为复杂,并且需要优化多个参数,从而导致算法的时间复杂度较高.为此,在本文中,提出了一种新的更为简单的同时利用全局和局部标记相关性的多标记分类算法 (Exploiting Global and Local Label Correlations for multi-label classification, EGLLC). 在利用全局标记相关性时,我们使用余弦相似性来获取任意两个标记之间不同类型的相关性,其中包括正相关、负相关和不相关.大多数已有的工作在利用标记相关性时没有考标记的负相关性,实际上,标记之间的负相关性同样能够为多标记学习提供有用的信息.与已有工作类似<sup>[7,10]</sup>,我们也采取聚类的方式来计算局部标记相关性.具体来说,我们首先利用  $k$ -means 聚类算法将训练集中的所有示例划分到不同的簇中,然后基于每个簇中的示例使用余弦相似性来估计局部标记相关性.最后,在多个领域的 12 个基准数据集上,我们对所提出的算法和其它方法进行了广泛的对比实验.实验结果表明,EGLLC 性能优于其它方法,具有较强的多标记学习能力.

## 2 利用全局和局部标记相关性的多标记分类算法

### 2.1 多标记学习

在多标记学习中,假设  $X \in \mathbf{R}^d$  是  $d$  维的输入空间, $\mathbf{Y} = (l_1, l_2, \dots, l_q)$  表示包含  $q$  个类别标记的标记空间. $D = \{(x_i, y_i)\}_{i=1}^n$  是包含  $n$  个样本的训练集,令  $X$  为特征矩阵

$X$  包含样本的集合, $\mathbf{Y}$  为  $\mathbf{Y}$  的标记集合,其中  $x_i \in X$  是一个  $d$  维的特征向量  $[x_{i1}, x_{i2}, \dots, x_{id}]$ ,  $y_i = [y_{i1}, y_{i2}, y_{iq}] \in \{0, 1\}^q$  是一个与  $x_i$  相关的二值标记向量.对于任一标记  $l_j$ ,当标记  $l_j$  与样本  $x_i$  相关时  $y_{ij} = 1$ ,否则  $y_{ij} = 0$ .

多标记学习的目标是学习一个能够为未见示例预测一组适当的标记的分类器:  $h: X \rightarrow 2^Y$ . 在多标记学习中,我们通常先学习一个实值函数  $f: X \times Y \rightarrow \mathbf{R}$ , 实值函数  $f(x, l)$  的值可以看作标记  $l$  属于示例  $x$  的置信度.然后,多标记分类器  $h(x)$  可以被定义为  $h(x) = \{l | f(x, l) > t, l \in Y\}$ , 其中  $t = 0.5$  是预先设置的阈值.这里我们假设  $f$  由  $q$  个子函数组成,每个标记对应一个函数.例如,  $f = [f_1, f_2, \dots, f_q]$ . 为了简单起见,本文中对于每个子函数  $f_i$ ,我们使用的是逻辑斯蒂回归模型:

$$f_i(x_i) = \frac{1}{1 + \exp(-x_i \mathbf{W}_i)}$$

其中  $\mathbf{W}_i = [W_{i1}, W_{i2}, \dots, W_{id}, b_i]^T$  表示我们要学习的第  $i$  个标记对应的子函数的参数向量,为了表示方便,偏置项  $b_i$  已经被扩充到  $\mathbf{W}_i$  中,同时将常值 1 作为额外的维度增加到每个示例  $x_i$  的最后.对于逻辑斯蒂回归模型的最优解,一般可以通过优化下面的目标函数得到:

$$\min_{\mathbf{W}_i} \sum_{i=1}^n [\log(1 + \exp(x_i \mathbf{W}_i)) - y_{ii}(x_i \mathbf{W}_i)] \quad (1)$$

### 2.2 全局标记相关性

在多标记学习的训练过程中,不同标记之间的相关性能够提供额外的信息,这些信息能够显著提升模型的性能<sup>[9]</sup>.为了计算各个标记之间的相关性,我们在标记空间利用余弦相似性计算标记之间的相似程度作为标记相关性.在标记空间中与样本不相关的标记取值为 0,为方便计算余弦相似性,我们将数据中标记为 0 的值都用 -1 代替,从而可以得到标记空间中的标记之间是正相关的,负相关的还是不相关的.对于两个正相关的标记,我们令它们对应的子分类器尽可能地相似,反之则令它们尽可能不同.这里我们用参数矩阵  $\mathbf{W}$  中任意两列之间的距离表示对应的两个子分类器之间的相似性.如果标记相关性是基于所有的样本  $X$  计算得到,我们称之为全局标记相关性.因此,在模型中加入的全局标记相关性的正则化项可以表示为:

$$\min_{\mathbf{W}} \sum_{j=1}^q \sum_{l=1}^q \theta_{jl} \text{Dis}(\mathbf{W}_j, \mathbf{W}_l) \quad (2)$$

$\theta$  表示全局标记相关性矩阵,通过计算标记空间中各个标记向量的余弦相似性获得. $\theta_{jl}$  表示第  $j$  个标记和第  $l$  个标记之间的相关性,其取值范围在 -1 到 1 之间.当  $\theta_{jl} > 0$  时表示标记  $j$  和标记  $l$  是正相关的,当  $\theta_{jl} < 0$  时表示标记  $j$  和标记  $l$  是负相关的,当  $\theta_{jl} = 0$  时表示标记  $j$  和标记  $l$  不相关.  $|\theta_{jl}|$  值越大说明相关程度越大.  $\text{Dis}(\mathbf{W}_j, \mathbf{W}_l)$  表示两个参数向量之间的距离,为方便起见,我们

采用欧氏距离的平方  $\text{Dis}(\mathbf{W}_j, \mathbf{W}_l) = \|\mathbf{W}_{jk} - \mathbf{W}_{lk}\|^2$  来度量相似性.

### 2.3 局部标记相关性

实际上,不同的示例可能共享不同的标记相关性,和全局标记相关性不同,如果标记相关性是由部分样本  $\mathbf{X}_i$  计算得出,我们称之为局部标记相关性.不同的样本子集可能会得到不同的局部标记相关性.为了有效地利用局部标记相关性,我们通过聚类方法将训练数据划分到  $m$  个簇中,分别在每个簇中计算标记之间的相关性.用  $\mathbf{X}_i$  表示第  $i$  个簇  $g_i$  中的示例,  $\mathbf{Y}_i$  是与之对应的标记矩阵.在我们实验过程中,出于简单考虑,我们选择标准的  $k$ -means 聚类算法来划分训练数据.我们假设相似的样本有相似的标记集合,因此我们在特征空间上进行聚类.在每个簇中,标记的相关性同样是从数据的标记空间中获取的,并采用余弦相似性来计算局部标记相关性.那么对于两个局部相关的标记,我们采取与全局相关的标记相同的处理方式,即正相关的标记对应的子分类器应该尽可能相似,反之则应尽可能不同.局部标记相关性可以表示为:

$$\min_{\mathbf{W}} \sum_{i=1}^m \sum_{j=1}^q \sum_{l=1}^q R_{jil} \|\mathbf{W}_j - \mathbf{W}_l\|^2 \quad (3)$$

$\mathbf{R}_i$  表示第  $i$  个簇中的示例的标记相关性矩阵,  $R_{jil}$  表示在第  $i$  个簇中第  $j$  个标记与第  $l$  个标记之间的余弦相似性.

### 2.4 最终优化目标

联合式(1)~(3),同时考虑全局和局部标记相关性,我们将最终需要优化的目标定义为:

$$\begin{aligned} G(\mathbf{W}) = & \sum_{j=1}^q \sum_{i=1}^n (\log(1 + \exp(\mathbf{x}_i \mathbf{W}_j)) - y_{ij}(\mathbf{x}_i \mathbf{W}_j)) \\ & + \frac{\lambda_1}{2} \|\mathbf{W}\|_F^2 + \frac{\lambda_2}{2} \text{tr}(\mathbf{W} \mathbf{L} \mathbf{W}^T) \\ & + \frac{\lambda_3}{2} \sum_{i=1}^m \text{tr}(\mathbf{W} \mathbf{L}_i \mathbf{W}^T) \end{aligned} \quad (4)$$

其中,  $\mathbf{L} = \mathbf{D} - \boldsymbol{\theta}$ ,  $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_q)$  是一个对角矩阵, 对角线元素  $d_i = \sum_{j=1}^q \theta_{ij}$ . 同理,  $\mathbf{L}_i = \mathbf{D}_i - \mathbf{R}_i$ ,  $\mathbf{D}_i = \text{diag}(d_{i1}, d_{i2}, \dots, d_{iq})$ , 对角矩阵中的对角元素根据每个簇中的标记相关性矩阵  $\mathbf{R}_i$  计算:  $d_{ij} = \sum_{k=1}^q R_{ijk}$ . 式(4)中的第二项使用来防止过拟合的正则化项, 第三项能够在模型训练过程中考虑全局标记相关性, 最后一项是在训练模型时考虑不同示例中的局部标记相关性.  $\lambda_1$ ,  $\lambda_2$  和  $\lambda_3$  是平衡因子.

### 2.5 优化方法

我们使用拟牛顿下降法 L-BFGS 算法<sup>[13]</sup>来求解目标函数  $G(\mathbf{W})$  的最小化问题. BFGS 算法是通过使用一个迭代增长的矩阵避免计算海塞矩阵的逆来减小牛顿法的计

算复杂度,而 L-BFGS 算法是对 BFGS 的进一步改进,利用 L-BFGS 算法能够快速有效地求解目标函数  $G(\mathbf{W})$  的最小值,并且 L-BFGS 算法地迭代过程仅与目标函数的一阶梯度有关,本文中目标函数  $G(\mathbf{W})$  的一阶梯度为:

$$\begin{aligned} \frac{\partial G(\mathbf{W})}{\partial \mathbf{W}} = & \mathbf{X}^T \left( \frac{1}{1 + \exp(-\mathbf{X} \mathbf{W})} - \mathbf{Y} \right) \\ & + \lambda_1 \mathbf{W} + \lambda_2 \mathbf{W} \mathbf{L} + \lambda_3 \sum_{i=1}^m \mathbf{W} \mathbf{L}_i \end{aligned} \quad (5)$$

受篇幅所限,本文不再对 L-BFGS 算法步骤展开赘述,读者可以参阅文献[13].

上述同时利用全局和局部标记相关性的算法的详细步骤如算法 1 所示.

#### 算法 1 基于全局和局部标记相关性的多标记分类算法 (EGLLC 算法)

输入: 训练数据集  $D = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^n$ , 参数  $\lambda_1, \lambda_2, \lambda_3, m$  以及收敛阈值  $\xi$ .  
输出: 模型参数矩阵  $\mathbf{W}$ .

1. 在  $D = \{\mathbf{x}_i\}_{i=1}^n$  上应用  $k$ -means 聚类得到  $m$  个簇;
2. 在每个簇里计算对应的标记空间  $\{\mathbf{y}_i\}_{i=1}^n$  的余弦相似性矩阵  $\mathbf{R}_i$ ;
3. 利用 L-BFGS 算法迭代优化求解  $\mathbf{W}$ , 其中式(5)计算目标函数梯度  $\nabla G(\mathbf{W})$ ;
4. 返回参数矩阵  $\mathbf{W}$ .

## 3 实验

### 3.1 评价指标

本文中,我们使用五个常用的多标记评价指标来评价各个对比算法的性能,分别为 Hamming loss, Ranking loss, One error, Coverage 和 Average precision<sup>[2]</sup>. 对于前 4 种评价指标来说,指标的取值越小,则表明算法的性能越好;对最后一个评价指标而言,当其取值越大时,表明算法的性能越好.篇幅所限,各指标具体计算公式可见文献[2].

### 3.2 实验配置

本文中,我们在多种多标记基准数据集上与其它算法进行了对比实验来验证我们所提算法的有效性.实验中用到的数据集的详细信息如表 1 所示.

我们将 EGLLC 与 6 种多标记学习算法进行了对比: ML-kNN<sup>[14]</sup>, RankSVM<sup>[15]</sup>, ML-LOC<sup>[10]</sup>, GLOCAL<sup>[7]</sup>, MLFE-MSVR<sup>[16]</sup> 和 MLFE-LR<sup>[16]</sup>. 对比算法涉及到的参数都使用相应文献的建议或者采用交叉验证进行参数调优,受篇幅所限在此不予赘述.

在我们所提出的算法 EGLLC 中涉及到的参数  $\lambda_1$ ,  $\lambda_2$  和  $\lambda_3$  的值都是从  $[10^{-4}, 10^1]$  中选择的,另外,聚类簇的数量  $m$  的取值范围是  $[5, 30]$ . 每个数据集的最佳参数设置都是依据算法在训练数据上的 5 折交叉验证的结果确定的.

表 1 数据集基本信息

名称	示例数目	示例维度	标记数目	标记密度	数据类型
arts	5000	462	26	1.64	text
business	5000	438	30	1.59	text
cal500	502	68	174	26.04	music
enron	1702	1001	53	3.38	text
flags	194	19	7	3.39	image
genbase	662	1185	27	1.25	biology
image	2000	294	5	1.24	image
llog	1460	1004	75	1.18	text
medical	978	1449	45	1.25	text
slashdot	3782	1079	22	1.18	text
society	5000	636	27	1.69	text
yeast	2417	103	14	4.237	biology

### 3.3 实验结果与分析

对于每个算法,我们在所有数据集上进行了 10 次 5 折交叉验证,并记录了实验结果的均值与方差.受篇幅所限,表 2 和表 3 给出了 Ranking loss 和 Average precision 指标下,各对比算法在所有数据集上的实验结果,其中最优实验结果已用粗体标识出来,最后一行为各算法的性能排序平均结果,在其他指标上的结果我们只给出统计信息.

在 60 组(5 评价指标  $\times$  12 数据集)对比实验结果中,EGLLC 共有 38 次排名第一(含并列),14 次排名第二(含并列),排在 3 和 4 共有 8 次,并且 EGLLC 没有排在第 5 之后的情况.

表 2 在 Ranking loss 评价指标下,对比算法在 12 个数据集上的实验结果“均值  $\pm$  方差(排名)”(取值越小表示算法性能越好)

数据集	GLOCAL	ML-kNN	ML-LOC	MLFE-MSVR	MLFE-LR	RankSVM	EGLLC
arts	.137 $\pm$ .001(3)	.148 $\pm$ .001(4)	.152 $\pm$ .002(6)	.122 $\pm$ .001(2)	.149 $\pm$ .000(5)	.179 $\pm$ .000(7)	<b>.107 <math>\pm</math> .001(1)</b>
business	.040 $\pm$ .001(5)	.038 $\pm$ .001(3.5)	.052 $\pm$ .002(7)	.036 $\pm$ .001(2)	.038 $\pm$ .001(3.5)	.049 $\pm$ .000(6)	<b>.029 <math>\pm</math> .000(1)</b>
cal500	<b>.178 <math>\pm</math> .001(1)</b>	.183 $\pm$ .001(4)	.225 $\pm$ .002(7)	.184 $\pm$ .001(6)	.183 $\pm$ .001(4)	.183 $\pm$ .000(4)	.179 $\pm$ .001(2)
enron	.114 $\pm$ .002(5)	.094 $\pm$ .001(4)	.157 $\pm$ .002(7)	.080 $\pm$ .001(2)	.082 $\pm$ .001(3)	.118 $\pm$ .000(6)	<b>.068 <math>\pm</math> .001(1)</b>
flags	.240 $\pm$ .001(4)	.231 $\pm$ .006(3)	.243 $\pm$ .010(5)	.226 $\pm$ .007(2)	.416 $\pm$ .014(7)	.281 $\pm$ .008(6)	<b>.224 <math>\pm</math> .005(1)</b>
genbase	.004 $\pm$ .001(5.5)	.006 $\pm$ .001(7)	<b>.002 <math>\pm</math> .000(2.5)</b>	<b>.002 <math>\pm</math> .000(2.5)</b>	.004 $\pm$ .001(5.5)	<b>.002 <math>\pm</math> .001(2.5)</b>	<b>.002 <math>\pm</math> .000(2.5)</b>
image	.199 $\pm$ .003(5)	.172 $\pm$ .002(4)	.311 $\pm$ .005(7)	<b>.135 <math>\pm</math> .002(1)</b>	.171 $\pm$ .002(3)	.233 $\pm$ .029(6)	.163 $\pm$ .002(2)
llog	.182 $\pm$ .004(5)	.165 $\pm$ .001(4)	.273 $\pm$ .010(7)	.150 $\pm$ .004(2)	.161 $\pm$ .004(3)	.186 $\pm$ .001(6)	<b>.120 <math>\pm</math> .002(1)</b>
medical	.058 $\pm$ .003(7)	.044 $\pm$ .002(4)	.023 $\pm$ .001(3)	.019 $\pm$ .001(2)	.046 $\pm$ .001(5)	.052 $\pm$ .001(6)	<b>.016 <math>\pm</math> .001(1)</b>
slashdot	.057 $\pm$ .001(5)	.047 $\pm$ .001(2.5)	.060 $\pm$ .002(7)	.056 $\pm$ .001(4)	.058 $\pm$ .001(6)	.047 $\pm$ .001(2.5)	<b>.039 <math>\pm</math> .001(1)</b>
society	.138 $\pm$ .001(4)	.132 $\pm$ .001(2)	.157 $\pm$ .004(6)	.134 $\pm$ .001(4)	.144 $\pm$ .000(5)	.161 $\pm$ .000(7)	<b>.115 <math>\pm</math> .001(1)</b>
yeast	.349 $\pm$ .002(7)	<b>.169 <math>\pm</math> .001(1)</b>	.186 $\pm$ .003(4)	.178 $\pm$ .001(3)	.347 $\pm$ .002(6)	.295 $\pm$ .005(5)	.174 $\pm$ .001(2)
AvgRank	4.708	3.583	5.667	2.708	4.667	5.292	1.375

表 3 在 Average precision 评价指标下,对比算法在 12 个数据集上的实验结果“均值  $\pm$  方差(排名)”(取值越大表示算法性能越好)

数据集	GLOCAL	ML-kNN	ML-LOC	MLFE-MSVR	MLFE-LR	RankSVM	EGLLC
arts	.607 $\pm$ .003(4)	.524 $\pm$ .002(5)	.623 $\pm$ .004(3)	.635 $\pm$ .002(2)	.508 $\pm$ .001(6)	.433 $\pm$ .000(7)	<b>.637 <math>\pm</math> .002(1)</b>
business	.883 $\pm$ .001(4)	.882 $\pm$ .001(5)	.878 $\pm$ .002(6)	.893 $\pm$ .001(2)	.890 $\pm$ .001(3)	.861 $\pm$ .000(7)	<b>.898 <math>\pm</math> .001(1)</b>
cal500	.507 $\pm$ .001(2)	.493 $\pm$ .002(6)	.473 $\pm$ .003(7)	<b>.510 <math>\pm</math> .002(1)</b>	.500 $\pm$ .005(4)	.498 $\pm$ .001(5)	.505 $\pm$ .002(3)
enron	.669 $\pm$ .002(4)	.629 $\pm$ .002(5)	.599 $\pm$ .003(6)	<b>.717 <math>\pm</math> .002(1.5)</b>	.672 $\pm$ .002(3)	.513 $\pm$ .000(7)	.717 $\pm$ .001(1.5)
flags	.790 $\pm$ .001(5)	.797 $\pm$ .005(3)	.791 $\pm$ .009(4)	.802 $\pm$ .007(2)	.677 $\pm$ .010(7)	.760 $\pm$ .006(6)	<b>.809 <math>\pm</math> .004(1)</b>
genbase	.991 $\pm$ .001(5)	.987 $\pm$ .002(7)	.995 $\pm$ .001(3.5)	.996 $\pm$ .001(2)	.989 $\pm$ .002(6)	.995 $\pm$ .001(3.5)	<b>.997 <math>\pm</math> .001(1)</b>
image	.766 $\pm$ .004(5)	.795 $\pm$ .003(3)	.650 $\pm$ .005(7)	<b>.835 <math>\pm</math> .002(1)</b>	.792 $\pm$ .003(4)	.725 $\pm$ .029(6)	.800 $\pm$ .002(2)
llog	.316 $\pm$ .003(5)	.306 $\pm$ .004(6)	.339 $\pm$ .005(4)	.419 $\pm$ .006(2)	.392 $\pm$ .003(3)	.246 $\pm$ .001(7)	<b>.428 <math>\pm</math> .004(1)</b>
medical	.815 $\pm$ .007(5)	.802 $\pm$ .005(6)	.894 $\pm$ .003(3)	.902 $\pm$ .004(2)	.692 $\pm$ .003(7)	.826 $\pm$ .003(4)	<b>.904 <math>\pm</math> .003(1)</b>
slashdot	.885 $\pm$ .001(3)	.884 $\pm$ .001(4)	.880 $\pm$ .002(6)	.889 $\pm$ .002(2)	.882 $\pm$ .001(5)	.888 $\pm$ .002(2)	<b>.896 <math>\pm</math> .001(1)</b>
society	.596 $\pm$ .003(5)	.612 $\pm$ .001(4)	.637 $\pm$ .003(3)	.647 $\pm$ .002(2)	.566 $\pm$ .000(6)	.546 $\pm$ .000(7)	<b>.650 <math>\pm</math> .001(1)</b>
yeast	.609 $\pm$ .001(6)	.762 $\pm$ .002(1)	.757 $\pm$ .003(3)	.756 $\pm$ .002(4)	.606 $\pm$ .002(7)	.626 $\pm$ .006(5)	.758 $\pm$ .001(2)
AvgRank	4.417	4.333	4.625	1.958	5.083	5.542	1.375

另外,如图 1 所示,我们给出了 EGLLC 和 GLOCAL 算法在所有数据集上一次实验所需的训练时间. 我们可以看到,EGLLC 的训练速度明显快于 GLOCAL,特别是当数据集的特征和标记数量较多的时候.

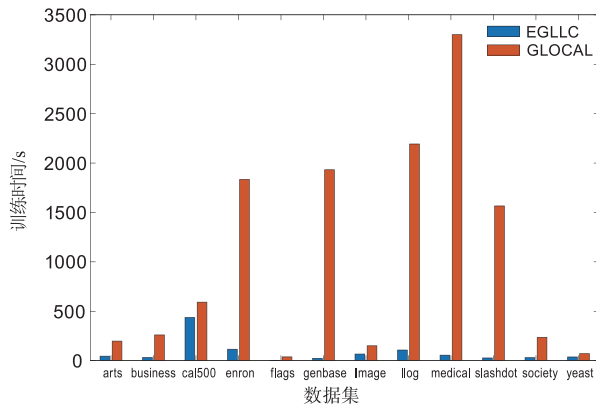


图1 EGLLC算法和GLOCAL算法在所有数据集上所需的训练时间

本文中,我们利用 Friedman 检验来进一步分析了各个对比算法之间的相对性能. 表 4 中给出了各个评价指标的 Friedman 统计值  $F_F$ ,以及相应的临界值. 对于每个评价指标,都在 0.05 显著水平下拒绝各比较算法的性能之间无区别的假设. 因此,我们继续使用 Bonferroni-Dunn 检验来验证我们所提的算法是否显著优于其它对比算法. 实验中将 EGLLC 看作控制算法,并利用临界值域(CD)来度量各个算法的平均排序与 EGLLC 之间的差异,当对比算法的平均排序值与 EGLLC 相差超过 CD 值时(本文中  $CD = 2.327$ ;对比算法数量  $k = 7$ ,数据集数量  $N = 12$ ),说明 EGLLC 算法的性能是显著区别于该算法.

表 4 各评价指标的 Friedman 统计值  $F_F$  以及在 0.05 显著性水平下相应的临界值

评价指标	$F_F$	临界值
Hamming loss	12.405	2.240
Ranking loss	10.351	
One error	9.543	
Coverage	60.231	
Average precision	11.344	

图 2 中给出了各评价指标的 CD 图,其中各个算法的平均排序值在水平轴中标出,越靠近右边说明排名越靠前. 在每一个图中,与 EGLLC 相差不到一个 CD 值得算法都已用一条黑色得加粗线连接,不相连的算法就说明它们之间的性能有显著性差异.

从图 2 中我们可以得出以下结论:

(1) EGLLC 在所有的评价指标下的平均排名都是最好的;

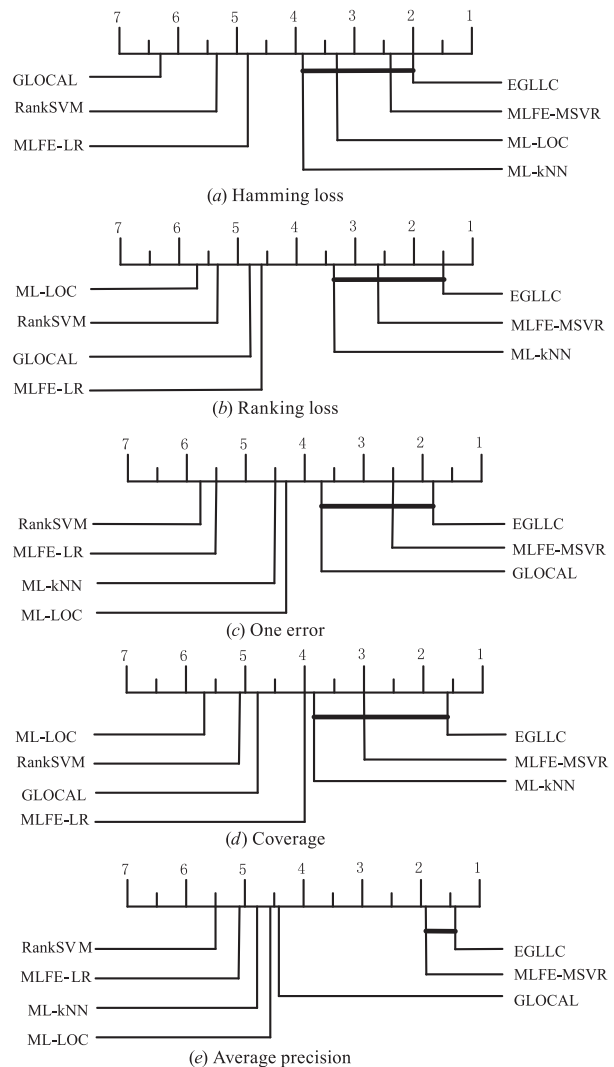


图2 EGLLC(控制算法)与其它六个对比算法在Bonferroni-Dunn 检验下的比较结果. CD图中未与EGLLC相连的算法被认为与控制算法具有显著不同的性能

(2) EGLLC 在所有评价指标下都显著优于 RankSVM 和 MLFE-LR 算法;

(3) EGLLC 在 Hamming loss 指标下与 ML-LOC 算法的性能相当,在 Hamming loss、Ranking loss 和 Coverage 指标下与 ML-kNN 的性能相当,在 One error 指标下与 GLOCAL 性能相当,并在其它所有情况下显著优于 ML-LOC、ML-kNN 和 GLOCAL.

为了进一步分析 EGLLC 算法与 MLFE-MSVR 算法性能的差异,本文通过使用 5% 显著性水平的双尾 t 检验来检验这两个算法之间是否存在显著性差异. 详细的对比结果如表 5 中所示. 在 12 个数据集中,其中 win 表示 EGLLC 显著优于 MLFE-MSVR 的数据集个数, tie 表示 EGLLC 和 MLFE-MSVR 没有显著差别的数据集个数, loss 表示 EGLLC 显著劣于 MLFE-MSVR 的数据集个数. 从对比结果中可以看出我们所提出的 EGLLC 算法

在多数情况下能够显著优于 MLFE-MSVR 算法. 进一步分析, 在 Hamming loss 和 One error 这两个指标上两个算法基本相当, 但在 Ranking loss, Coverage 和 Average precision 这三个指标上, EGLLC 算法明显的优于 MLFE-MSVR 算法. 通过这五个指标的定义我们可以得知, Hamming loss 和 One error 只关注单个标记上的性能, 没有考查标记之间的关联性. 而其他三个指标都是考查的标记排序后标记与标记之间的关系, 而这正是我们所提算法 EGLLC 的优势所在, 充分的考虑了标记之间的相关性, 故能在相应的指标上取得更好的结果.

表 5 95%置信度水平下 EGLLC 算法和 MLFE-MSVR 算法双尾 t 检验统计结果. win/tie/loss 分别表示 EGLLC 显著优于/相当于/差于 MLFE-MSVR 的数据集个数

	Hamming loss	Ranking loss	One error	Coverage	Average precision
win/tie/loss	3/7/2	9/2/1	5/3/4	9/2/1	8/2/2

## 4 总结

本文提出了一种新的多标记学习算法, 在学习过程中同时考虑全局和局部标记相关性. 具体来说, 我们首先整个训练集的标记空间中利用余弦相似性来计算数据中的全局标记相关性, 然后将数据划分为多个簇, 并在每个簇中分别计算标记相关性作为局部标记相关性. 最后将这两种标记相关性同时整合到学习进程中来提升多标记分类模型的性能. 我们在多种类型的数据集上与其它算法进行了对比实验. 实验结果表明, 我们所提的算法在大多数情况下是明显优于其它方法的, 在多标记学习中有着突出的优势. 此外, 我们通过进一步实验验证了本文所提的同时使用全局和局部标记相关性的有效性与合理性.

### 参考文献

- [1] ZHANG Min-ling, ZHOU Zhi-hua. A review on multi-label learning algorithms [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819 – 1837.
- [2] ZHOU Zhi-hua, ZHANG Min-ling. Multi-label learning [A]. Encyclopedia of Machine Learning and Data Mining [C]. New York, NY: Springer, 2017. 875 – 881.
- [3] 程玉胜, 赵大伟, 王一宾, 等. 非平衡化标签补全核极限学习机多标签学习 [J]. 电子学报, 2019, 47(3): 719 – 725.  
CHENG Yu-sheng, ZHAO Da-wei, WANG Yi-bin, et al. Multi-label learning of kernel extreme learning machine with non-equilibrium label completion [J]. Acta Electronica Sinica, 2019, 47(3): 719 – 725. (in Chinese).
- [4] SUN Fu-ming, TANG Jin-hui, LI Hao-jie, et al. Multi-label image categorization with sparse factor representation [J]. IEEE Transactions on Image Processing, 2014, 23(3): 1028 – 1037.
- [5] 吴信东, 赵银凤, 李磊. 基于种子节点选择的网络环境下多标签分类算法研究 [J]. 电子学报, 2016, 44(9): 2074 – 2080.  
WU Xin-dong, ZHAO Yin-feng, LI Lei. Multi-label classification in network environments via seed node selection [J]. Acta Electronica Sinica, 2016, 44(9): 2074 – 2080 (in Chinese).
- [6] WU Bin, ZHONG Er-hong, HORNER Andrew, et al. Music emotion recognition by multi-label multi-layer multi-instance multi-view learning [A]. Proceedings of the ACM International Conference on Multimedia [C]. New York, NY: ACM, 2014. 117 – 126.
- [7] ZHU Yue, KWOK James-T, ZHOU Zhi-hua. Multi-label learning with global and local label correlation [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(6): 1081 – 1094.
- [8] READ Jesse, PFAHRINGER Bernhard, HOLMES Geoff, et al. Classifier chains for multi-label classification [J]. Machine Learning, 2011, 85(3): 333 – 359.
- [9] ZHANG Min-ling, ZHANG Kang. Multi-label learning by exploiting label dependency [A]. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. New York: ACM, 2010. 999 – 1008.
- [10] HUANG Sheng-jun, ZHOU Zhi-hua. Multi-label learning by exploiting label correlations locally [A]. Proceedings of the 26th AAAI Conference on Artificial Intelligence [C]. Palo Alto, CA: AAAI Press, 2012. 949 – 955.
- [11] JIA Xiu-yi, ZHU Sai-sai, LI Wei-wei. Joint label-specific features and correlation information for multi-label learning [J]. Journal of Computer Science and Technology, 2020, 35(2): 247 – 258.
- [12] 刘军煜, 贾修一. 一种利用关联规则挖掘的多标记分类算法 [J]. 软件学报, 2017, 28(11): 2865 – 2878.  
LIU Jun-yu, JIA Xiu-yi. Multi-label classification algorithm based on association rule mining [J]. Journal of Software, 2017, 28(11): 2865 – 2878. (in Chinese).
- [13] YUANYa-xiang. A modified BFGS algorithm for unconstrained optimization [J]. IMA Journal of Numerical Analysis, 1991, 11(3): 325 – 332.
- [14] ZHANG Min-ling, ZHOU Zhi-hua. ML-KNN: A lazy learning approach to multi-label learning [J]. Pattern Recognition, 2007, 40(7): 2038 – 2048.
- [15] ELISSEEFF Andre, WESTON Jason. A kernel method for multi-labelled classification [A]. Proceedings of Advances in Neural Information Processing Systems 14 [C]. New

York:ACM,2001. 681 – 687.

- [16] ZHANG Qian-wen, ZHONG Yun, ZHANG Min-ling. Feature-induced labeling information enrichment for multi-label learning [A]. Proceedings of the 32nd AAAI Conference on Artificial Intelligence [C]. Palo Alto, CA: AAAI Press, 2018. 4446 – 4453.

#### 作者简介



**朱赛赛** 男, 1994 年出生于安徽濉溪, 硕士研究生, 主要研究方向为机器学习, 数据挖掘.  
E-mail: zhusaisworld@163.com



**贾修一 (通信作者)** 男, 1983 年出生于山东日照, 副教授, 博士生导师, 主要研究方向为机器学习, 数据挖掘.  
E-mail: jiaxy@njust.edu.cn



**李泽超** 男, 1985 年出生于河南开封, 教授, 博士生导师, 主要研究方向为图像视频智能分析.  
E-mail: zechao.li@njust.edu.cn