第48卷第1期 中国科学技术大学学报 Vol.48,No.1
2018年1月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Jan. 2018

文章编号:0253-2778(2018)01-0065-10

一种基于朴素贝叶斯的校准标签排序方法

张其龙,邓维斌,胡 峰,瞿 原,胡宗容

(重庆邮电大学计算智能重庆市重点实验室,重庆 400065)

摘要:传统的校准标签排序算法(calibrated label ranking, CLR)利用成对标签关联进行转化来预测结果.该算法的校准是在二元关系算法(binary relevance, BR)基础上进行比较产生结果,其预测对BR产生结果具有一定的依赖性,因此该算法在预测某些数据集时具有一定的局限性.为了更好地区分标签的相关性和不相关性,提出了一种用于标签边界域的校准方法,对处于相关性标签和不相关性标签的边界部分采用贝叶斯概率进一步校正,从而提高边界域部分分类的准确性.基于朴素贝叶斯校准的标签排序方法(calibrated lable ranking method based on naive bayes, NBCLRM)与校准标签排序等7种传统的方法进行对比,实验结果表明,本文提出的算法不仅可以根据需求修改阈值ε和μ来调节预测结果,而且能够有效地提升传统多标签学习方法的性能.

关键词:数据挖掘;朴素贝叶斯;校准标签排序算法;多标签学习算法

中图分类号:TP391 **文献标识码**:A doi:10.3969/j.issn.0253-2778.2018.01.009

引用格式:张其龙,邓维斌,胡峰,等.一种基于朴素贝叶斯的校准标签排序方法[J].中国科学技术大学学报,2018,48(1):65-74.

ZHANG Qilong, DENG Weibin, HU Feng, et al. A calibrated lable ranking method based on naive Bayes[J]. Journal of University of Science and Technology of China, 2018,48(1):65-74.

A calibrated lable ranking method based on naive Bayes

ZHANG Qilong, DENG Weibin, HU Feng, QU Yuan, HU Zongrong

(*Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: The traditional calibrated label ranking algorithm (calibrated label ranking, CLR) uses pairs of label associations to transform and predict results. Its algorithmic calibration is achievely comparing it with the basis of binary relevance (BR). Its prediction has a certain dependence on the results of BR, thus incurring some limitations on the prediction of some datasets. To better distinguish between the relevance and irrelevance of the label, a method is presented for calibrating label boundary regions, which further corrects the boundary portion of the relevant label and the irrelevant label using Bayesian probability, thereby improving the accuracy of the classification of the boundary domain. CLR method based on naive Bayes(NBCLRM) presented is compared with seven traditional methods such as calibrated label ranking. Experimental results show that the proposed algorithm can not only adjust prediction results by modifying the thresholds ε and μ, but also effectively improve the performance of traditional multi-label learning

收稿日期:2017-05-22;**修回日期**:2017-06-23

基金项目:国家自然科学基金(61473001,71071045,71131002)资助.

作者简介:张其龙,男,1989年生,硕士生.研究方向:计算智能、数据挖掘.E-mail:814150638@qq.com

通讯作者:胡峰,博士/教授.E-mail:hufeng@cqupt.edu.cn

methods.

Key words: data mining; Naive Bayes; calibrated label ranking; multi-label learning algorithm

0 引言

随着信息化的普及,各领域都积累和收集了大量的数据信息,激增的数据包含许多的信息,传统的简单查询和统计已不能满足日益增长的需求,需要一些新的技术来挖掘数据隐含的价值,提取大数据中有价值的信息是当前发展的一个热点.数据挖掘从简单统计发展成包括分类和回归、聚类分析、关联规则、时序模式等多方位技术[1].

随着事物间的复杂化,事物也开始具有复杂的语义信息.事物的划分由原来的单一标签逐渐转化到多个标签,例如一张图片既可以同时出现大海、蓝天、沙滩,还可以出现房屋等信息.在多标签学习中,每个样本可能包含一个或者多个标签,含有更多标签信息的样本能够更好地表现事物语义信息的多样性[2]。多标签学习的任务主要包括多标签的分类和多标签的排序,这两个任务也是多标签进行评估解决问题方法性能的两个方面.前一种任务是确定和测试样本标签的关联的标签集合,后一种任务是根据多标签的所有确定标签和测试样本的关联度来确定标签的序列[3].随着机器学习发展,多标签学习已经逐渐的取代了单标签,成为机器学习及其相关领域研究的一个热点方向,并大量应用于文本分类[4]、图像视屏标注[5]、音乐情感分类[6]、生物信息[7-8]等领域.

多标签数据挖掘以机器学习为基础,通过机器学习技术将数据转化为信息,再由数据挖掘方法将信息转化为知识。机器学习的方法需要通过学习已有的数据集,构建一个合适的模型来处理未知数据[3].多标签学习算法主要分为:问题转化和算法适应[2],其中问题转化是将多标签转化为多个单标签,利用成熟的单标签学习方法进行标签的预测,如二元关系法、校准的标签排序、分类器链、随机 k 标签集、层次多标签分类等方法.算法适应是直接改变现已存在的单标签算法来适应多标签的处理,如多标签 k 近邻、反向传播多标签学习法等方法.

1 多标签学习

1.1 多标签学习问题描述

多标签分类是一个多维特征与多个标签相关联的有监督学习问题,是对简单的单标签问题进行转化为.设 $X=F^m$ 是 m 维特征向量空间,$\Lambda=\{\lambda_1,\lambda_2,\cdots,\lambda_q\}$ 为整个标签集合.数据样本 $DS^{(k)}=\{x^{(k)},Y^{(k)}\}$ 由 m 维特征向量 $x^{(k)}$ 和 q 维标签子集 $Y^{(k)}\in 2^A$ 组成,其中 $Y^{(k)}=\{y_i^{(k)}\in\{0,1\}\mid 1\leqslant i\leqslant q\}$.当 $q=1$ 时,多标签学习问题转化为单标签问题.对于测试样本 $TD^{(s)}$,通过多标签学习算法的可输出对应的标签集合($y^{(t)}=\{y_i^{(t)}\in\{0,1\}\mid 1\leqslant i\leqslant q\}$)和每个样本与标签的关联度($cy^{(t)}=\{cy_i^{(t)}\mid 1\leqslant i\leqslant q\}$).根据样本的预测标签集合与对应的样本和标签的关联度可进行一些性能的评估.

1.2 多标签学习方法

基于多标签学习方法中的标签之间的关联性,已有的求解策略大致可以分为三类:一阶策略,二阶策略和高阶策略[2].

一阶策略方法忽略标签之间的关联性,认为标签是独立存在的.该方法实现简单,但是泛化能力较低.二元关系(BR)直接把多标签学习问题中的多个标签分解为独立的单标签问题进行求解.Taha等[9]将二元关系方法运用到了阿拉伯文本的分类.

二阶策略方法主要考虑了成对标签之间的关系.不能全面地解决二阶以上的标签之间关联.Fürnkranz等提出了将多标签成对关联并引入人工校准的校准标签排序(CLR)[10]以及改进传统的神经网络模型的反向传播多标签学习算法(BP-MLL)[11].

高阶策略方法考虑了多个标签之间的关联信息.针对标签之间的关联考虑得比较全面,但其计算的复杂度难以估计,并且难以处理规模巨大的多标签学习问题.Tsoumakas等提出了将标签集化分成多个标签子集并采用标签幂集法进行建模的随机 k 标签集(RAkEL)[12]以及构建一组有序二分类器并输入依赖输出的分类器链(CC)[13].

2 校准标签排序算法

校准标签排序法是基于成对比较排序法(ranking by pairwise comparison, RPC)[3]的一个拓展.在RPC的基础上加入人工校准标签,该标签区分相关标签和不相关标签集合,校准标签和相关标签组合产生正例,校准标签则为负例;校准标签和

不相关标签组合产生负例，校准标签则为正例. 经过校准标签排序方法处理多标签数据将会同时产生二元关系法和成对比较排序法标签转化的结果.

$$y_{ij}=\{\varphi(y_i,y_j)\mid\varphi(y_i)\neq\varphi(y_j),1\leqslant i\leqslant j\leqslant n\},$$

$$\varphi(y_i,y_j)=\begin{cases}1, & \text{if}(\varphi(y_i)>\varphi(y_j))\\ 0, & \text{if}(\varphi(y_i)<\varphi(y_j))\end{cases}\tag{1}$$

$$\zeta(x,y_i)=\sum_{k=1}^{j-1}[\![t_{jk}=0]\!]+\sum_{i=j}^{n}[\![t_{ij}=1]\!],\quad 1\leqslant i\leqslant n\tag{2}$$

$$\xi(x,t)=\sum_{l=1}^{n}[\![t_{ll}=0]\!]\tag{3}$$

$$y_i=\{t_{ij}\mid\zeta(x,t_{ij})>\xi(x,t),0\leqslant i,j\leqslant n\}\tag{4}$$

$$fy_j=\frac{\zeta(x,t_j)}{n},\ (1\leqslant j\leqslant n)\tag{5}$$

算法 2.1　校准标签排序法(CLR)

输入：

n：多标签的维数

y_i：{y_1，y_2，…，y_n}训练集的 n 维标签

$X^{(s)}$：{x_1，x_2，…，x_m}训练数据集的特征

$T^{(t)}$：{t_1，t_2，…，t_m}预测数据集的特征

输出：

ty：预测样本的标签集

fy：预测样本与标签的关联度

/ * 每一个样本都执行以下的循环将成对标签转化新标签 λ_{ij} * /

for i in range(1，n)

　for j in range(i，n)

　　if(i = j) λ_{ii} = y_i

/ * 根据公式(1)，可以得到 * /

　　else λ_{ij} = $\varphi(y_i,y_j)$

End if

End for

End for

/ * 1、新标签和对应的数据集特征对应，合并相同维数产生的标签，进行分类器的学习，得到学习模型. 2、将待预测的数据集 $T_{(s)}$ 对训练好的模型进行预测，得到预测结果标签 t_{ij} * /

for i in range(1，n)

　for j in range(i，n)

M_{ij} = Train(X，Y_{ij})

t_{ij} = M_{ij}.pred($T^{(t)}$)

End for

End for

/ * 计算最终的预测结果 * /

/ * 根据公式(3)进行人工校准标签的结果统计 * /

$\zeta=\xi(x,t)$

for i in range(1，n)

/ * 根据公式(2)进行成对标签预测结果进行每种可能的类别的投票统计 * /

$\xi_i=\zeta(x,t_i)$

/ * 根据公式(4)进行对比，得出最终预测的标签集和该标签的关联度 * /

$ty_i=y_i(\zeta,\xi_i)$

$fy_i=fy_j(i=j)$

End for

/ * 返回标签集合和标签的关联度 * /

Return ty，fy

3　NBCLRM 算法

NBCLRM 的主要思想是朴素贝叶斯利用样本信息和参数的先验信息得出预测的后验概率，通过后验概率对校准的标签排序方法产生的结果进行校准. NBCLRM 算法主要包括：训练预测数据集的特征处理、NB 的概率计算、校准 CLR 多标签学习算法、阈值 ε 和 μ 的确定. 朴素贝叶斯法认为各特征相互独立，不存在任何的联系，是基于贝叶斯定理和特征条件独立假设的分类方法. 通过已知的训练数据集先计算出特征条件独立假设学习输入、输出的联合概率分布；然后基于算出的联合概率分布. 对预测的数据集，利用贝叶斯定义求出相关标签的后验概率. 再用相关标签的后验概率进行 CLR 预测的关联度进行校准. 对本文选取的多标签数据，采用纯 NB 分类器进行预测分类，其性能效果很不理想，其主要原因是忽略了多标签之间的相关性.

首先计算正态分布的参数，合并训练集和测试集需要对原始样本数据特征进行归一化，将结果归一化到[1，2]区间，但消除特征之间的差异过大会影响概率的计算，见公式(6).其中常数 λ 的加入是为了防止分母为 0. 样本数据拆分为原训练集和原测试集. 训练集分别计算每列特征值的均值和方差，得到每列特征值的正态分布参数，见公式(7)，(8). 其次计算条件概率和先验概率，根据训练集计算的均值和方差以及测试集的特征可以计算预测实例的条件概率，见公式(9). 已知训练集数据标签可以计算出每种类别的先验概率，见公式(10)，其中引入一个正数 λ，主要目的是防止先验概率值可能出现为 0 的情况影响到后验概率的计算，使预测结果产生较大误差.取 $\lambda=1$，这时称为拉普拉斯平滑

(Laplace smoothing). 常数 k 为数据样本中类别的种类数量,本次实验的数据集为二分类数据样本,所以 $k=2$. 通过先验概率和条件概率计算出类别标签类别的概率, 见公式(11). 通过类别标签的概率可计算出相关标签的相对概率, 见公式(12), 其中 α 和 ω 是防止数据的不平衡导致数据预测结果的倾斜而引入的权重. 然后通过 CLR 方法进行模型训练预测出每个预测集样本的相关标签概率并进行决策,见公式(13). 最后将不确定部分即概率为[1−ε, ε], 对处于不确定区间的预测结果通过阈值 μ 进行二次划分确定其最终结果, 见公式(14). 其具体的算法流程如算法 3.1 所示.

$$X^{*}=\frac{x-\min+\lambda}{\max-\min+\lambda}+1,\ (\lambda>0) \tag{6}$$

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{7}$$

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})^2 \tag{8}$$

$$\rho(X=x_i\underset{y\in\{0,1\}}{\mid}Y=y)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-\overline{x})^2}{2\sigma^2}\right)+\lambda,\ (1\leqslant i\leqslant n) \tag{9}$$

$$\varphi(\underset{y\in\{0,1\}}{Y=y})=\frac{\lambda+\sum_{i=1}^{m}|y|}{n+k\lambda} \tag{10}$$

$$\omega(\underset{y\in\{0,1\}}{Y=y})=\varphi(Y=y)\prod_{i=1}^{n}\rho(X=x_i\mid Y=y) \tag{11}$$

$$\theta=\frac{\alpha*\omega(Y=1)}{\alpha*\omega(Y=1)+\beta*\omega(Y=0)} \tag{12}$$

$$p(y^{*})=\begin{cases}1 & fy_i>\varepsilon\\ 0 & fy_i\leqslant 1-\varepsilon\\ \partial_i & \text{otherwise}\end{cases} \tag{13}$$

$$\partial_i=\mu*\theta_i+(1-\mu)*fy_i \tag{14}$$

算法 3.1 基于朴素贝叶斯校准标签排序(NBCLRM)

输入:

$D^{(s)}$:训练集数据集的特征

n:训练集标签维数

m:数据集的特征维数

Y:{y_1, y_2,…,yn_i 训练集标签集

$T^{(s)}$:待预测样本数据集的特征

输出:

ty:最终预测的标签集合

/* 训练集数据集和测试集数据集进行样本合并,计算每维数据集特征的最大值 max 和最小值 min,然后拆分训练集和预测集得到 $D^{(s)*}$ 和 $T^{(s)*}$,最后并将训练集和预测集特征的归一化到[1,2] */

for i in range(m)

maxD$_i$ = max(D$_i$ (s) *)

minD$_j$ = min(D$_i$ (s) *)

/* 利用公式(6)进行归一化 */

X$_i$ = X$_i$ *

End for/* 利用公式(7)和(8)求解训练集的均值 $\overline{X}$ 和方差 σ^2,再利用公式(9)求解条件概率 */

for i in range(m)$\overline{X_i}=\overline{X}$

$\sigma_i^2=\sigma^2$

$\rho_i(X=x_i\underset{y\in\{0,1\}}{\mid}Y=y)=\rho(X=x_i\underset{y\in\{0,1\}}{\mid}Y=y)$

End for

/* 使用公式(10)计算标签的先验概率 */

for j in range(n)

$\varphi_j\underset{y\in\{0,1\}}{(Y=y)}=\varphi(Y=y)$

End for

/* 使用公式(11)计算联合概率分布 */

$\omega_j\underset{y\in\{0,1\}}{(Y=y)}=\omega(Y=y)$

/* 公式(12)计算预测结果为相关标签的概率 */

$\theta^{*}=\theta$

/* 将归一化的数据集传入到算法 1 中,得到标签关联度 fy, 利用公式(13)设定的 ε 将结果化分为三个区间域:正域、负域和中间域,根据公式(14)计算中间域 */

for k in range(n):

if (fy$_k$ > ε):

p(y$_k$ *) = 1

elif(fy$_k$ < 1 − ε):

p(y$_k$ *) = 0

else p(y$_k$ *) = ∂k

if(p(y$_k$ *) > 0.5):

ty$_k$ = 1

else ty$_k$ = 0

else ty$_k$ = 0

End if

End for

/* 结果返回标签集合 */

Return ty

4 实验结果

4.1 实验设置

实验全部运行在 i7-4720HQ @ 2.6 GHz 的处理器、12 G 内存.本实验中,BR、BPMLL、RAkEL、

CLR、MLkNN、CC 及 HOMER 等 7 种传统的多标签学习方法采用 Mulan[14] 多标签学习开源项目提供的算法，需要的基分类器采用随机森林.

4.2 实验数据集

从 Mulanlibrary[15] 下载了 8 个多标签数据集与多个传统的算法进行了比较。其中 Nuwide 数据集是从 Nuswide-cVLADplus 数据集随机抽样选取的一部分.标签基数表示每个样本的相关标签的平均数量，其求解见公式(15)，标签差异表示多标签数据集中不同样本出现不同标签集合的数量，其求解见公式(16). 具体数据集的信息见表 1.

$$LCard(D)=\frac{1}{m}\sum_{i=1}^{m}\llbracket Y_i=1\rrbracket \tag{15}$$

$$LDis(D)=|\{\exists x:(x,Y)\in D\}| \tag{16}$$

表 1　数据集描述

Tab.1　Description of datasets

数据集	样本数	属性数	标签数	标签基数	标签差异
Flags	194	19	7	3.392	54
Emtions	593	72	6	1.869	27
Enron	1 702	1 001	53	3.378	753
Scene	2 407	294	6	1.074	15
Birds	645	260	19	1.014	133
Yeast	2 417	103	14	4.237	198
CAL500	502	68	174	26.044	502
Nuwide	2 794	128	81	1.85	788

4.3 评价指标

与单标记学习系统相比，多标签学习的评价准则更加复杂. 多标签学习算法的评估指标主要分为两类：基于标签分类的方法和基于标签排的方法. 本文实验采用了 5 种多标签评估指标，分别为：汉明损失(hamming loss，HL)，覆盖范围(coverage)，排序损失(ranking loss，RL)，平均查准率(average precision，AvePre) 和微观 F_1 值(micro F_1-measure，MF1)[16]. 其中 Hamming loss 和 micro F_1-measure 作为多标签分类的评价指标，而 coverage，ranking loss 和 average precision 作为多标签排序性能的评价指标. 其中 HL、coverage、RL 三个评价指标越小其性能越好，而 AvePre 和 MF1 越大其性能越好.

4.3.1 基于标签分类的方法

基于标签分类的方法主要是对单个标签进行评估，然而在多标签中，将多个标签组成的标签集的每一个标签进行单独评估，可以评估多标签中的每个单个标签，最终将每一个单标签的评估结果的平均值作为多标签的预测结果.有两种方法实现该平均：宏观平均和微观平均.

对于二分类问题，可以将样例的真实标签分类和预测标签分类进行组合化分为真正例(TP)、假正例(FP)、真反例(TN)、假反例(FN)四种情形[17]，TP、FP、TN、FN 分别代表其对应的样例数，分类结果的混淆矩阵如表 2 所示.

表 2　分类结果混淆矩阵

Tab.2　the confusion matrix of classification results

类别标签 λ		真实标签分类	
		正例	负例
预测标签分类	正例	TP(真正例)	FN(假反例)
	负例	FP(假正例)	TN(真反例)

设 m 个样本 n 个标签的测试集

$DS=(X_i,Y_1^j,l_1^j)$，$1\leqslant i\leqslant m$，$1\leqslant j\leqslant n$.

其中，X_i 表示第 i 个样本，Y_j^i 表示样本实际的第 i 个样本的第 j 个标签，L_i^j 表示样本预测结果中的第 i 个样本的第 j 个标签.

Hamming loss：用于度量样本在单个标签上的真实标签和其预测对应标签的错误匹配情况.

$$HL=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{j}^{n}[Y_i^j\neq L_i^j]=\frac{1}{n}\sum_{j}^{n}(FP^j+FN^j) \tag{17}$$

Micro F_1-measure：用于综合衡量标签的微观精度和微观召回率.

$$microP=\frac{\sum_{j}^{n}TP^j}{\sum_{j}^{n}TP^j+\sum_{j}^{n}FP^j} \tag{18}$$

$$microR=\frac{\sum_{j}^{n}TP^j}{\sum_{j}^{n}TP^j+\sum_{j}^{n}FN^j} \tag{19}$$

$$microF_1=\frac{2\times microP\times microR}{microP+microR} \tag{20}$$

4.3.2 基于标签排序的方法

Coverage 用于预测标签的排序中，找出所有相

关标签需要的平均次数.

$$\text{coverage}=\frac{1}{m}\sum_{i}^{m}\max_{\lambda\in L_i}(R_i(\lambda))-1 \tag{21}$$

式中,$R_i(\lambda)$表示第 i 个样本中标签 λ 所有标签中的排位,m 为样本集的数量.

Ranking loss:表示不相关标签排在相关标签前面的概率的平均.

$$\text{RL}=\frac{1}{m}\sum_{i}^{m}\frac{1}{|Y_i||L_i|}|\{l_a,l_b:R_i(l_a)>R_i(l_b),(l_a,l_b)\in Y_i\times L_i\}| \tag{22}$$

Average precision:该方法反映了所有样本的预测标签排序中,排在相关标签前面的也是相关标签的概率的平均.

$$\text{AvePre}=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{|R_i|}\sum_{l\in R_i}\frac{|L|\text{rank}(X_i,L)\leqslant\text{rank}(X_i,L),L\in R_i|}{\text{rank}(X_i,L)} \tag{23}$$

4.4 实验结果与分析

BR、BPMLL、RAkEL、CLR、MLkNN、CC 和 HOMER 7 种传统的多标签学习方法以及 NBCLRM 在 5 种评价指标上性能如表 3~7 所示.以下是对 8 个数据集进行了 5 项性能指标的比较,NBCLRM 通过对原有的 CLR 进行朴素贝叶斯校准,具有一定程度的提高.图 1 是传统的 7 种算法和改进的 NBCLRM 通过比较显示 NBCLRM 在预测结果的评估上占有一定的优势,在不同的评价指标上的较好数据集所占的比列分别为,HL:37.5%,coverage 和 ranking loss:50%,average precision 和 micro F_1:75%,尤其是 average precision 和 micro F_1-measure 两个指标的提升效果表现得比较明显.图 2 为传统的 CLR 方法和 NBCLRM 方法进行性能对比的结果.传统的 CLR 方法在 HL 上优于 NBCLRM 方法,即在某些数据集上预测的准确性略高于 NBCLRM 方法,其主要原因是 NBCLRM 方法为了大幅提升主要的评估指标而降低了其他的评价指标.对于数据集 scene,5 个评价指标只有一个 average precision 高于其他 7 种传统的算法,而且,这主要是因为校准标签排序算法 CLR 对于在 scene 数据集上表现的没有其他 6 种传统算法对预测分类的精准性,算法本身局限了自己的预测结果,再加上 NBCLRM 没有准确地校准预测结果,导致了 NBCLRM 的评估性能没有其他算法的好,甚至有些指标比 CLR 表现得差.对于数据集 CAL500,CLR 在预测结果上明显好于 NBCLRM,这主要是 NBCLRM 将更多的 CLR 从错误的分类校准过来,从而导致了误分类的标签数量过多,进而影响了利用 NB 的校准结果,反而表现得比原 CLR 更差.传统的 7 种算法在 Nuwide 数据集上表现的都比较差,这主要原因是标签基数较低,数据不平衡,数据样本的相关标签的数量很少,导致数据预测的区分度很低,大多数被判别到了不相关标签,然而对 NBCLR 在 Nuwide 数据集上表现得好于其他算法,其主要原因是朴素贝叶斯很大程度上由特征属性决定,利用朴素贝叶斯进行校准,可以将不确定部分有效地进行二次划分,提升了不确定部分的校准.

表 3 不同方法在 Hamming loss 指标下的实验结果

Tab.3 Experimental results of different methods in Hamming loss

BR	RAkEL	CLR	MLkNN	CC	HOMER	BPMLL	NBCLRM	
flags	0.295	0.259	0.251	0.268	0.27	0.259 34	0.299	**0.242**
emotions	0.218	0.201	**0.193**	0.209	0.197	0.204 62	0.218	0.204
scene	0.099	0.088 4	0.088 6	0.095	0.09	**0.087 38**	0.227	0.103
birds	0.044	0.042	0.041	0.05	0.041	0.039 6	0.125	**0.035 7**
yeast	0.217	**0.193**	0.194	0.198	0.194	0.217 95	0.23	0.223
CAL500	0.158	0.144	0.142	**0.139**	0.144	0.210 8	0.245	0.202
Nuwide	0.023 5	0.022 9	0.022 9	0.022 58	0.023 2	0.036 16	0.037 5	0.019 2
Enron	0.05	0.047	0.047	0.051	**0.046 8**	0.049 37	0.236	0.052
The number of better dataset	0	1	1	1	1	1	0	3

表 4　不同方法在 coverage 指标下的实验结果

Tab.4　Experimental results of different methods in coverage

BR	RAkEL	CLR	MLkNN	CC	HOMER	BPMLL	NBCLRM	
flags	3.83	4.138	3.569	3.615	3.662	3.969	3.8	**3.47**
emotions	1.936	2.485	1.812	1.876	1.807	2.203	1.92	**1.801**
scene	0.099	0.089	0.089	0.095	0.09	**0.087 4**	0.227	0.1051
birds	2.978	5.672	2.319	2.7	**1.92**	3.993 8	5.51	2.03
yeast	7.22	9.093	6.233	6.364	6.318	7.744 8	6.69	**6.206**
CAL500	155.966	168.867	**129.34**	130.32	136.28	167.113	131.45	130.01
Nuwide	23.19	43.813	13.384	13.782 9	15.92	31.384	15.78	**13.04**
enron	18.63	31.58	**11.734**	13.181	12.149	23.226 25	16.51	11.96
The number of better dataset	0	0	2	0	1	1	0	4

表 5　不同方法在 ranking loss 指标下的实验结果

Tab.5　Experimental results of different methods in ranking loss

BR	RAkEL	CLR	MLkNN	CC	HOMER	BPMLL	NBCLRM	
flags	0.232	0.294	0.207	0.187	0.213	0.248 462	0.235	**0.168**
emotions	0.171	0.26	0.155	0.159	0.155	0.203 5	0.178	**0.15**
scene	0.091	0.176 401	0.072	0.092	**0.064**	0.134 6	0.176	0.075
birds	0.105	0.25	0.079	0.098	**0.063**	0.143 9	0.146	0.087
yeast	0.203	0.328	0.165	0.171	0.169	0.231 8	0.193	**0.162**
CAL500	0.25	0.489	**0.186**	0.189	0.204	0.316 1	0.193	0.192
Nuwide	0.1819	0.4128	0.086 9	0.089 7	0.100 8	0.237 8	0.102 2	**0.085**
enron	0.268	0.333	0.214	0.28	**0.209**	0.3351	0.865	0.215
The number of better dataset	0	0	1	0	3	0	0	4

表 6　不同方法在 average precision 指标下的实验结果

Tab.6　Experimental results of different methods in average precision

BR	RAkEL	CLR	MLkNN	CC	HOMER	BPMLL	NBCLRM	
flags	0.811	0.776	0.815	0.838	0.815	0.806	0.809	**0.852**
emotions	0.797	0.749	0.806	0.797	0.806	0.784 9	0.78	**0.82**
scene	0.85	0.784	0.874	0.851	0.864	0.832	0.689	**0.876**
birds	0.609	0.329	0.645	0.563	0.705	0.514	0.215	**0.796**
yeast	0.726	0.667	0.76	0.757	0.757	0.7269	0.732	**0.762 8**
CAL500	0.431	0.297	**0.495**	0.483	0.473	0.373 65	0.494	0.486
Nuwide	0.368	0.086 9	0.477 5	0.463 9	0.474 1	0.324 66	0.482	**0.585 6**
enron	0.65	0.486	0.696	0.635	**0.702**	0.587 1	0.307	0.684
The number of better dataset	0	0	1	0	1	0	0	6

表 7 不同方法在 Micro F_1-measure 指标下的实验结果

Tab.7 Experimental results of different methods in micro F_1-measure

BR	RAkEL	CLR	MLkNN	CC	HOMER	BPMLL	NBCLRM	
flags	0.7	0.733	0.746	0.708	0.72	0.74786	0.704	**0.758**
emotions	0.665	0.663	0.681	0.65	0.677	0.705 46	0.68	**0.757**
scene	0.693	0.697	0.699	0.718	0.696	**0.723 42**	0.518	0.719
birds	0.387	0.342	0.365	0.255	0.363	0.53714	0.139	**0.589**
yeast	0.626	0.636	0.638	0.636	0.641	0.665 23	0.635	**0.776 9**
CAL500	0.378	0.339	0.349	0.329	0.341	0.432 82	0.441	**0.668**
Nuwide	0.1263	0.0494	0.0512	0.1637	0.0278	0.314 14	0.355 88	**0.639 6**
enron	0.536	0.54	0.547	0.466	0.551	**0.585 95**	0.306	0.574 2
The number of better dataset	0	0	0	0	0	2	0	6

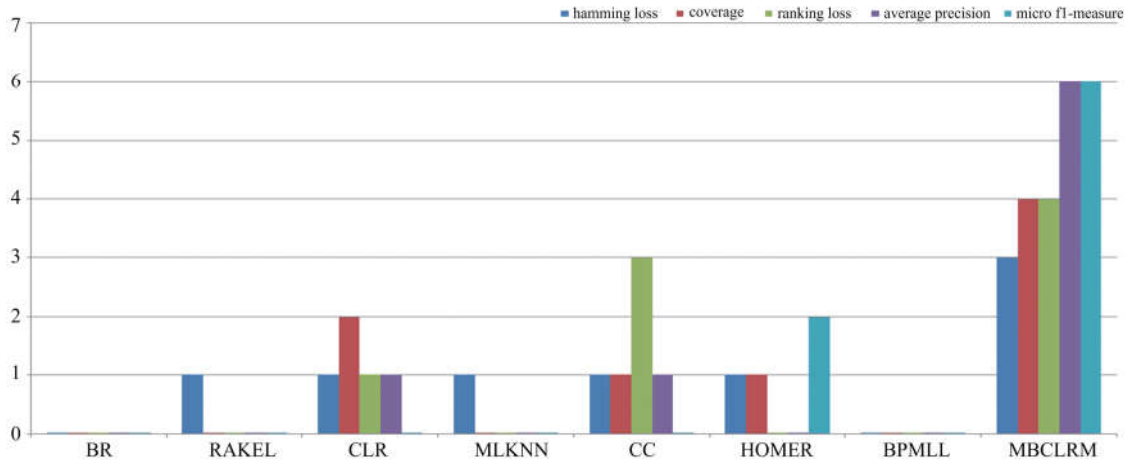

图 1 较佳数据集的数量

Fig.1 the number of better dataset

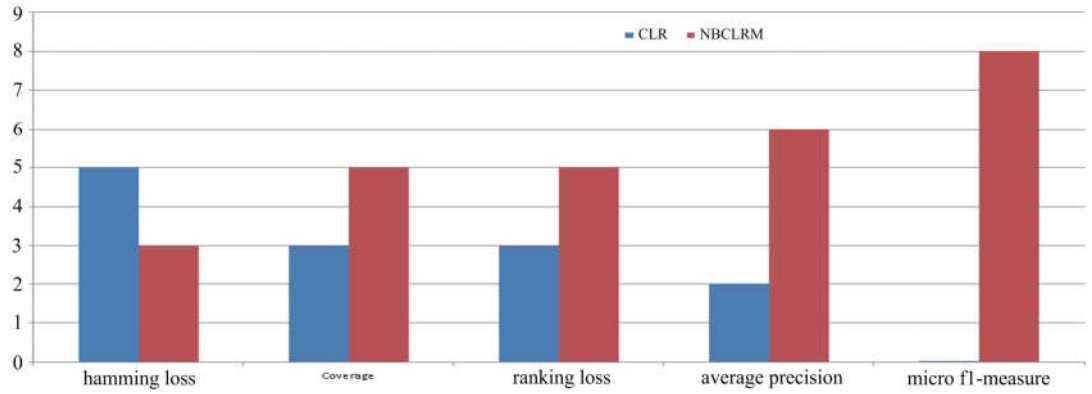

图 2 CLR vs NBCLRM

Fig.2 CLR vs NBCLRM

4.5 阈值的设置

阈值 μ 和阈值 ε 在最终的决策中起着一定的影响作用. 阈值 ε 将数据样本的预测结果划分为了 3 个区间域,即负域$[0,1-\varepsilon)$、中间域$[1-\varepsilon,\varepsilon]$和正域$(\varepsilon,1]$,其中 $\varepsilon \geqslant 0.5$. 本次试验采用 SMO (sequential minimal optimization)的思想来调节多阈值变量,即先固定需要设置的阈值以外的其他阈值,然后设置阈值,求出需求的阈值。首先将阈值 μ 设置为$(0,1)$中的某一个值,本实验设置的中间值,即 $\mu=0.5$,然后调节阈值 ε 的大小,本实验采用暴力方法来求解阈值,阈值 ε 的变化区间在$[0.5,1]$,进行 51 次取值,步长 0.01. 为了防止一些评价性能

过高和过低，进行综合评价后选取阈值 ε. 经过每组数据样本的多次暴力求解可求出阈值 ε. 本轮实验主要参考评价指标 average precision，其他指标为辅的方式进行调节，每组数据样本实验经过 51 次的暴力求解和多个评价指标的参考，这 8 组数据集实验在阈值 ε＝0.75 下综合评估性能较佳，最后将本文实验设置 ε＝0.75.然后调节阈值 μ，阈值 μ 对阈值 ε 划分的不确定中间域[1－ε，ε]进行二次划分. 这里对阈值 μ 进行了 21 次取值，即取值范围为[0，1]，步长为 0.05. 以下列举了 flags、birds 和 emotions 这 3 个数据集的阈值 μ 的变化情况.具体变化情况如图 3～5 所示.

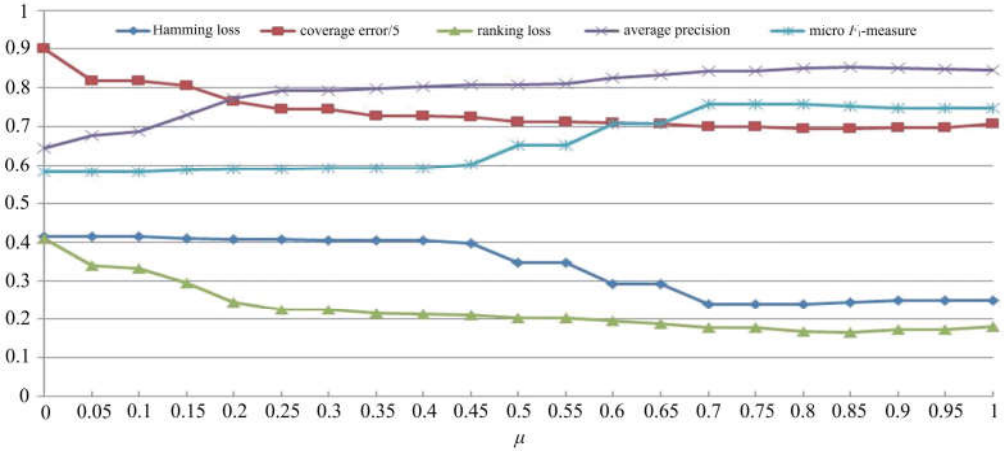

图 3　Flags 数据集的阈值 μ 变化

Fig.3　Flags dataset change of the threshold μ

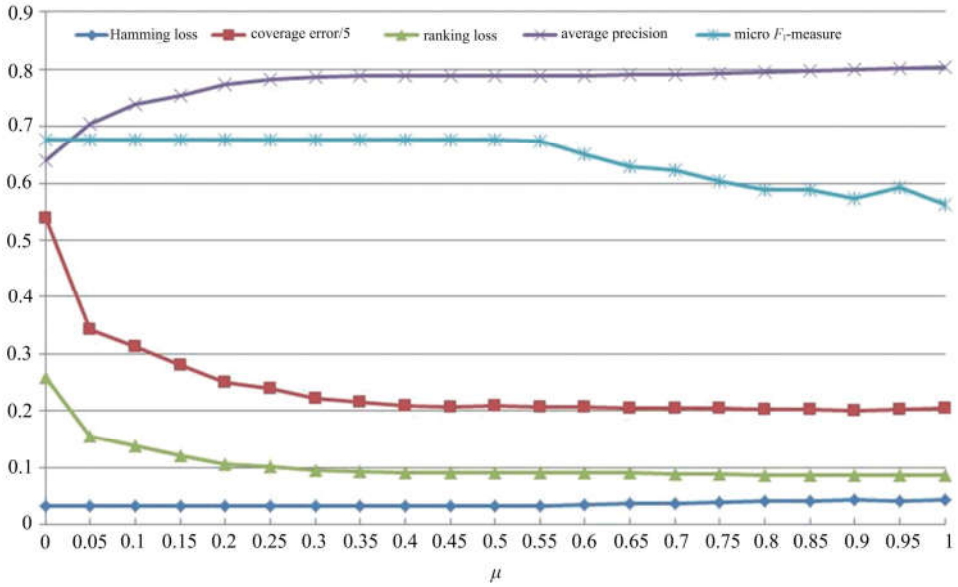

图 4　Birds 数据集的阈值 μ 变化

Fig.4　Birds dataset change of the threshold μ

图 3 随着 μ 的增长，micro-F_1 和 AvePre 逐渐上升，达到最高值后开始往下稍减，HL、coverage 和 RL 先减，然后稍有上升. 图 4 中的 birds 数据集，AvePre 在一直在上升，RL 一直下降，但是 micro-F_1 在下降，HL 在上升，这主要因为将真反例或真正例转化为了假正例或假反例. 对于图 5 的 emotions 数据集，micro-F_1 和 HL 成相反的变化而且走势相反，其主要原因是将假正例或者假反例预测为本应该预测为真反例或真正例的，降低了 HL，并提升了 micro-F_1.图 3～5 表明，一个评价指标达到最佳时，其他的评价指标不一定达到最佳，所以需要根据数据集的用途需求来设置阈值 μ 来达到要求. 以上本文实验结果全部是采用 μ＝0.8 的情况下得出的预测，然后进行的评估指标的评测.

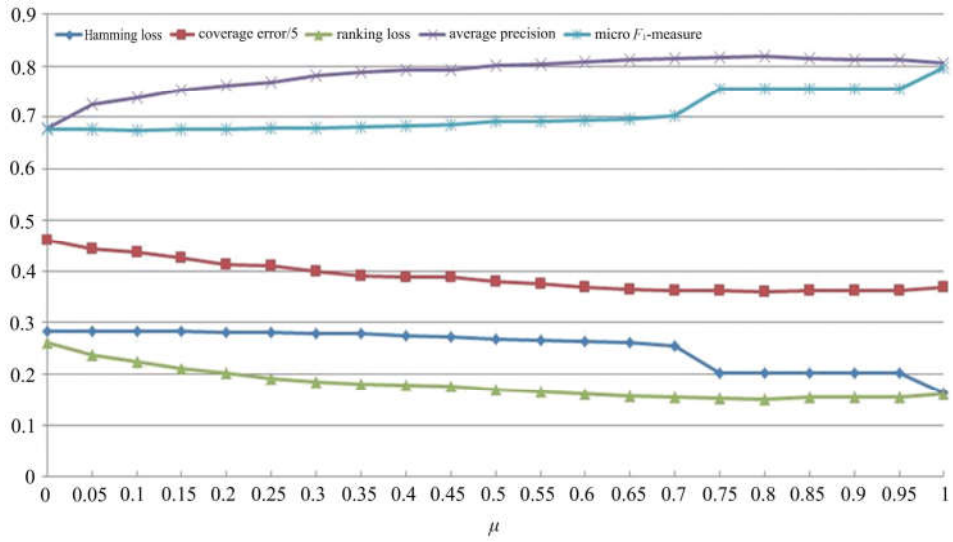

图 5 Emotions 数据集的阈值 μ 变化

Fig.5 Emotions dataset change of the threshold μ

5 结论

本文提出的 NBCLRM 算法是一种基于朴素贝叶斯的二阶策略校准标签排序算法，该算法改变了校准标签排序传统的一次性比较就产生结果的校准方式，使校准标签排序在不确定域中进一步进行校准，降低了错误的预测分类，最终输出尽可能预测准确的分类结果. 通过在以上 8 个多标签数据集中的实验，结果表明，NBCLRM 在大多数情况下比其他的传统的多标签算法表现得好. 而且该校准的思想还可以移植到高阶策略方法的标签校准.

参考文献(References)

[1] 王小妮. 数据挖掘技术[M]. 1 版. 北京：北京航空航天大学出版社，2014.

[2] ZHANG M L, ZHOU Z H. A review on multi-label learning algorithms[J]. Knowledge & Data Engineering IEEE Transactions on, 2014, 26(8): 1819-1837.

[3] 李思男，李宁，李战怀. 多标签数据挖掘技术：研究综述[J]. 计算机科学，2013，40(4)：14-21.
LI Sinan, LI Ning, LI Zhanhuai. Multi-label data mining: A survey[J]. Computer Science, 2013, 40(4): 14-21.

[4] ÑANCULEF R, FLAOUNAS I, CRISTIANINI N. Efficient classification of multi-labeled text streams by clashing[J]. Expert Systems with Applications, 2016, 41(11): 5431-5450.

[5] YU Y, PEDRYCZ W, MIAO D Q. Neighborhood rough sets based multi-label classification for automatic image annotation [J]. International Journal of Approximate Reasoning, 2013, 54(9): 1373-1387.

[6] LO H Y, WANG J C, WANG H M, et al. Cost-sensitive multi-label learning for audio tag annotation and retrieval[J]. IEEE Transactions on Multimedia, 2011, 13(3): 518-529.

[7] YU G X, RANGWALA H, DOMENICONI C, et al. Protein function prediction using multilabel ensemble classification [J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2013, 10(4): 1045-1057.

[8] YU G X, RANGWALA H, DOMENICONI C, et al. Protein function prediction with incomplete annotations [J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2014, 11(3): 579-591.

[9] TAHA A Y, TIUN S. Binary relevance (BR) method classifier of multi-label classification for arabic text[J]. Journal of Theoretical and Applied Information Technology, 2016, 84(3): 414-422.

[10] FÜRNKRANZ J, HÜLLERMEIER E, MENCíA E L, et al. Multilabel classification via calibrated label ranking[J]. Machine Learning, 2008, 73(2): 133-153.

[11] WANG J, HUANG P L, SUN K W, et al. Ensemble of cost-sensitive hypernetworks for class-imbalance learning [C]// Proceedings of the International Conference on Systems, Man, and Cybernetics. Manchester, UK: IEEE Press, 2013: 1883-1888.

[12] TSOUMAKAS G, VLAHAVAS I. Random k-Labelsets: An Ensemble Method for Multilabel Classification[M]// Machine Learning: ECML 2007. Springer, 2007: A122.

[13] READ J, PFAHRINGER B, HOLMES G, et al. Classifier chains for multi-label classification [J]. Machine Learning, 2011, 85(3): 333-359.

[14] TSOUMAKAS G, SPYROMITROS-XIOUFIS E, VILCEK J, et al. MULAN: A Java library for multi-label learning [J]. Journal of Machine Learning Research, 2011, 12(7): 2411-2414.

[15] Mulan: A Java Library for Multi-Label Learning[DB/OL]. [2017-05-06] http://mulan.sourceforge.net/datasets-mlc.html.

[16] HE Z F, YANG M, LIU H D. Joint learning of multi-label classification and label correlations[J]. Journal of Software, 2014, 25(9): 1967-1981.

[17] 周志华. 机器学习[M]. 北京：清华大学出版社，2016.