

多标签分类中标签检测技术的实验比较

刘佳丽, 许建华

(南京师范大学 计算机科学与技术学院, 江苏 南京 210023)

[摘要] 当前的部分多标签分类算法本质上由两项分类技术级联而成, 前一级建立标签排序系统, 后一级检测相关标签, 兼顾进一步改善分类性能. 本文针对不同标签检测技术开展研究, 收集并实现 4 种通用标签检测技术: 线性回归阈值法、多输出线性回归法、Logistic 回归法以及离散 Bayes 规则, 以 k 近邻算法作为基线算法, 在 10 个基准数据集上进行实验比较. 实验结果表明, 从计算时间与分类性能两个方面来说, 多输出线性回归法是值得推荐的方法.

[关键词] 多标签分类 k 近邻法 线性回归阈值函数 多输出线性回归 Logistic 回归 离散 Bayes 规则

[中图分类号] TP18 [文献标志码] A [文章编号] 1672-4292(2012)04-0055-07

An Empirical Comparison of Label Detection Techniques for Multi-Label Classification

Liu Jiali, Xu Jianhua

(School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China)

Abstract: Now some multi-label classification methods cascade two different classification techniques in essence. The former is to build a label ranking system and the latter to detect relevant labels effectively and improve classification performance further. To compare the different detection techniques, we collect four general label detection approaches: linear regression threshold, multiple output linear regression, logistic regression and discrete Bayesian methods. With k -nearest neighbor algorithm as a baseline method, we conduct an extensive experimental comparison on ten benchmark data sets. Our experimental results demonstrate that multiple output linear regression technique is recommendable, according to both computational time and classification performance.

Key words: multi-label classification, k -nearest neighbor algorithm, linear regression threshold function, multi-output linear regression, logistic regression, discrete Bayesian rule

传统的(两类或者多类)单标签分类仅仅需要预测未知样本的一个标签, 它的一个推广是多标签分类, 则需要确定未知样本的一组标签(称之为相关标签子集). 当前, 多标签分类在图像与视频标注、文本分类、基因与蛋白质功能分类、音乐情调分类中得到广泛的应用. 例如, 一幅海滨拍摄的相片可能会标注为沙滩、大海和蓝天; 一篇关于刘翔的新闻可能归属时政、体育、社会新闻栏目. 近 5 年来, 许多学者提出来了大量的多标签分类算法, 大体上可以分成四大类: 基于数据集分解的方法^[1-2], 利用分解手段(一对多、一对一和标签幂集)将多标签数据集分解成一个或者多个两类或者多类数据子集, 然后为每一个子集设计一个子分类器, 最后组装成一个多标签分类算法, 其优点是可以利用现有的单标签算法及其开源软件快速地构造一个多标签分类算法; 算法扩展方法^[1-2], 扩展现有的多类算法直接处理多标签数据集, 可以充分考虑标签相关性, 但高计算复杂性是这类算法的最大弱点; 混合方法^[3], 在扩展某一个算法的同时采用一个或者两个分解手段, 在保持低计算复杂性的基础上尽可能考虑标签相关性; 集成算法^[4], 利用前 3 类算法作为基分类器构造集成算法, 面临的困难仍然是高计算复杂性.

对于前 3 类算法, 大多是从经典的单标签分类发展而来的, 其中的部分算法只能产生一个标签排序系

收稿日期: 2012-04-28.

基金项目: 国家自然科学基金(60875001).

通讯联系人: 许建华, 博士, 教授, 研究方向: 模式识别、机器学习和生物信息学. E-mail: xujianhua@njnu.edu.cn

统,如多标签支持向量机与多标签神经网络^[5,6];另外的其他算法本身可以直接完成多标签分类,但是其性能相对欠佳,如 k 近邻法^[1,7,8]. 因此,这些算法需要后接一个相对简单的算法来完成相关标签的检测,兼顾进一步改善分类性能. 从算法结构来看,这类算法由两个算法串连而成,可以称之为两阶段算法^[9],其中的第二阶段算法称之为标签检测技术.

当经典的多类支持向量机、多层感知器扩展成相应的多标签算法(Rank-SVM 和 BP-MLL)时,这些算法只能首先建立一个好的标签排序系统^[5,6]. 为了确定相关标签子集,利用排序算法的 c (类别数) 个判别函数将原始 d 维训练样本转化成 c 维向量,其对应的阈值使 Hamming 损失最小,再级联一个线性回归模型来训练一个阈值函数. 这一技术可称为线性回归阈值法(Linear Regression Threshold Method, LRTM). 在文献[10]中,将相关与不相关标签设置为 $+1/-1$,为每一类建立一个线性回归模型来检测样本相关标签,本质上实现一个多输出线性回归^[11],本文称之为多输出线性回归法(Multiple Output Linear Regression Method, MOLRM). 基于多标签感知器、两类感知器集成算法和 k NN 算法在 3 个数据集上进行实验比较,说明 Hamming 损失可以得到稍微的降低.

经典的 k 近邻法(k NN)是最常用的单标签分类算法之一^[12,13],它的最简单的多标签形式是利用一对多策略分解数据集、 $k/2$ 作为阈值来确定相关标签,但是性能相对较差^[1,14]. 为了进一步改进基于 k NN 多标签分类算法的性能,ML- k NN^[8]和 IBLR-ML^[14]增加一个基于留一法的训练过程,并将判别函数设计成概率形式. ML- k NN 综合 k NN 和留一法将原始 d 维训练样本向量转化成 c 维整数向量,表示每一类对应的近邻标签数;针对每一个类,利用 Laplace 估计子计算离散 Bayes 公式的类先验概率与类条件概率. 文献[14]中 IBLR-ML 采用距离倒数加权的 k NN 将训练样本转化成 c 维的实数向量;利用 Logistic 模型^[14,15]估计每一类的后验概率,并且依赖于所有类的值. 这两种方法中的技术被称为离散 Bayes 规则和 Logistic 回归法. 由于 IBLR-ML 中采用加权策略,这两种标签检测技术的优劣仍然值得关注.

本文以多标签 k NN 为基算法,在 10 个基准数据集上比较上述 4 种标签检测技术.

1 原始的多标签 k 近邻法

假设独立同分布的多标签训练样本集为:

$$\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l)\}, \quad (1)$$

其中 $x_i \in \mathbf{R}^d$ 是 d 维的特征向量; $y_i = [y_{i1}, \dots, y_{ic}]^T = \{+1, -1\}^c$ 为 c 维的二进制标签向量, $y_{ij} = \pm 1$ 表示第 j 个标签是样本 x_i 的相关或者不相关标签; c 为类别或者标签总数. 原始的多标签 k 近邻法^[7]是单标签 k NN 算法的一个简单扩展,将多标签样本看作同时属于多个类来计算每一个类的近邻样本数. 对于某一待分类样本 x ,计算其与训练样本集(1)中所有样本的某一距离度量,并按照距离递增的顺序排列训练样本,再取出最近的 k 个近邻样本,记为,

$$\{(\bar{x}_1, \bar{y}_1), \dots, (\bar{x}_i, \bar{y}_i), \dots, (\bar{x}_k, \bar{y}_k)\}, \quad (2)$$

其中 (\bar{x}_i, \bar{y}_i) 表示 x 的第 i 个近邻样本及其标签, $i=1, 2, \dots, k$. 根据式(2),计算属于每一个类的近邻样本数作为判别函数,即,

$$g_j(x) = |\{i | \bar{y}_{ij} = +1, i=1, \dots, k\}|, j=1, \dots, c. \quad (3)$$

即计算在 k 个近邻中第 j 个类标签出现的次数. 由于这一过程等价于基于一对多分解的 k NN 算法,所有可以用 $k/2$ 作为阈值来确定样本的二进制标签向量 $y = [y_1, \dots, y_c]^T$,即

$$\text{当 } g_j(x) \geq k/2 \text{ 有 } y_j = +1; j=1, \dots, c. \quad (4)$$

文献[1,14]的实验说明,原始多标签 k NN 算法的性能相对欠佳,需要后接一个标签检测算法来进一步改善分类性能,并更有效地检测相关标签. 为此,利用留一法,根据式(2)和式(3)将训练集(1)转化成一个新的训练集:

$$\{(z_1, y_1), \dots, (z_i, y_i), \dots, (z_l, y_l)\}, \quad (5)$$

其中 $z_i = [z_{i1}, \dots, z_{ic}]^T$ 是 c 维的整数向量,表示每一个类的近邻样本个数,即 $z_{ij} = g_j(x_i)$. 为了便于推导,本文将式(5)改写成矩阵形式:

$$Z = \begin{bmatrix} z_1^T & 1 \\ \cdots & \\ z_l^T & 1 \end{bmatrix}, Y = \begin{bmatrix} y_1^T \\ \cdots \\ y_l^T \end{bmatrix}. \quad (6)$$

由于 kNN 是不需要训练模型而直接依据训练样本的算法,所以需要依据留一法将原始训练集转化成新的数据集;对于带模型的排序算法,如多标签支持向量机^[5],则利用 c 个判别函数将原始数据转化成新的训练数据,此时 z_i 是 c 维的实数向量。

2 4 种标签检测技术

2.1 线性回归阈值法

线性回归阈值函数法是 $k/2$ 阈值的推广,采用可变的阈值来确定相关标签,其阈值函数的形式为:

$$t(z) = w^T z + w_0. \quad (7)$$

为了估计式(7)中的参数,对式(5)中每一个训练样本,寻找一个阈值使得 Hamming 损失达到最小^[5],即

$$t_i = \operatorname{argmin}_t (|\{j | z_{ij} \leq t, y_{ij} = 1, j=1, \dots, c\}| + |\{j | z_{ij} > t, y_{ij} = -1, j=1, \dots, c\}|), \quad (8)$$

其中 $|\cdot|$ 表示集合的大小,并构成一个阈值向量:

$$t = [t_1, \dots, t_l]^T. \quad (9)$$

根据 Z 和 t 利用最小二乘线性回归法,得到:

$$[w^T, w_0]^T = (Z^T Z)^{-1} Z^T t. \quad (10)$$

在测试阶段,对于某一个待分类样本 x ,根据训练集(1)和式(2)、(3)将其转化样本为新样本向量 $z = [z_1, \dots, z_c]^T$,代入式(7)后求得该样本对应的阈值,则 x 的标签向量 $y = [y_1, \dots, y_c]^T$ 为:

$$\text{如果 } z_j \geq t(z), \text{ 则 } y_j = +1; \text{ 否则 } y_j = -1; j=1, \dots, c. \quad (11)$$

2.2 多输出线性回归法

线性回归阈值法是设计一个阈值函数来分割相关与不相关标签,而多输出线性回归法为每一个类设计一个线性函数,即:

$$f_j(z) = w_j^T z + w_{j0}, j=1, \dots, c. \quad (12)$$

将式(6)中每 j 列的 $+1/-1$ 值作为第 j 类样本的标识,利用最小二乘回归方法确定式(12)中的参数.将式(12)中的参数记为一个矩阵:

$$W = \begin{bmatrix} w_1 & \cdots & w_c \\ w_{10} & \cdots & w_{c0} \end{bmatrix}, \quad (13)$$

则根据文献[11]有,

$$W = (Z^T Z)^{-1} Z^T Y. \quad (14)$$

从式(14)可以看成,每一个类对应的矩阵相同,因此只需要进行一次计算,从而可以有效地降低计算代价。

在测试阶段,对于某一个待分类的样本 x ,首先转化样本为新样本向量 $z = [z_1, \dots, z_c]^T$,然后代入式(12)计算 $f_j(z)$ ($j=1, \dots, c$),则 x 的标签向量 $y = [y_1, \dots, y_c]^T$ 为:如果 $f_j(z) \geq 0$,则 $y_j = +1$;否则 $y_j = -1$ 。

2.3 Logistic 回归法

多输出线性回归模型计算得到的函数值可以是任意的实数,但研究人员偏爱样本属于每一类的概率,为此可以采用 Logistic 回归模型^[15,16].对于第 j 个类,Logistic 模型的表达式为:

$$P(z_i) = \frac{1}{1 + e^{-(w_j^T z + w_{j0})}} \in [0, 1], j=1, \dots, c, \quad (15)$$

表示样本具有标签 j 的概率大小. Logistic 模型的参数可以通过最大似然法或者最小均方误差两种方法来确定^[15,16].因为最大似然法求解过程中 Hessian 矩阵可能是奇异的,因此本文采用最小均方误差的学习方法,此时每一类的准则函数为:

$$E_j = \frac{1}{2} \sum_{i=1}^l \left(\frac{y_{ij} + 1}{2} - P(z_i) \right)^2, \quad (16)$$

对应的梯度向量为:

$$\nabla E_j = \begin{bmatrix} \frac{\partial E_j}{\partial w_j} \\ \frac{\partial E_j}{\partial w_{j0}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^l \left(\frac{y_{ij} + 1}{2} - P_j(z_i) \right) P_j(z_i) (1 - P(z_i)) z_i \\ \sum_{i=1}^l \left(\frac{y_{ij} + 1}{2} - P_j(z_i) \right) P_j(z_i) (1 - P(z_i)) \end{bmatrix}. \quad (17)$$

根据这一梯度,本文构造一个迭代过程来优化每一类参数.值得注意的是这里需要为每一类单独进行训练,因此 Logistic 方法的计算复杂性要比前两种方法高.

在测试阶段,对于某一个待分类的样本 x ,首先转化样本为新样本向量 $z = [z_1, \dots, z_c]^T$,然后代入式(15)计算后验概率 $P_j(z)$ ($j=1, \dots, c$),则 x 的标签向量 $y = [y_1, \dots, y_c]^T$ 为:如果 $P_j(z) \geq 0.5$,则 $y_j = +1$;否则 $y_j = -1$.

2.4 离散 Bayes 规则

由于在文献[5]中的样本为整数向量,因此可以使用离散 Bayes 公式^[12].根据样本集(5),离散 Bayes 规则估计每一个类的先验概率与类条件概率.记 $P_j(+1)$ 和 $P_j(-1)$ 为样本具有或者不具有标签 j 的先验概率, $P_j(m|+1)$ 与 $P_j(m|-1)$ 表示当近邻样本数为 m 时具有或者不具有标签 j 的条件概率.根据 Laplace 估计子有:

$$\begin{aligned} P_j(+1) &= \frac{|i|y_{ij}=1, i=1, \dots, l|+1}{l+2}, \\ P_j(-1) &= 1 - P_j(+1), \\ P_j(m|+1) &= \frac{|i|z_{ij}=m, y_{ij}=1, i=1, \dots, l|+1}{|i|y_{ij}=1, i=1, \dots, l|+k+1}, \\ P_j(m|-1) &= \frac{|i|z_{ij}=m, y_{ij}=-1, i=1, \dots, l|+1}{|i|y_{ij}=-1, i=1, \dots, l|+k+1}, \\ j &= 1, \dots, c; m=0, \dots, k. \end{aligned} \quad (18)$$

在测试阶段,对于某一个待分类的样本 x ,先转化样本为新的样本向量 $z = [z_1, \dots, z_c]^T$,再根据离散 Bayes 公式计算后验概率:

$$P_j(z_j|+1) = \frac{P_j(+1) P_j(z_j|+1)}{P_j(+1) P_j(z_j|+1) + P_j(-1) P_j(z_j|-1)}, \quad j=1, \dots, c, \quad (19)$$

则 x 的标签向量 $y = [y_1, \dots, y_c]^T$ 为:如果 $P_j(z_j|+1) \geq 0.5$,则 $y_j = +1$;否则 $y_j = -1$.

3 实验结果与分析

3.1 10 个基准数据集、3 个性能评价指标与算法参数设置

为比较上述 4 个不同的标签检测算法的性能,本文选取来自文献[17,18]的 10 个基准数据集,表 1 列出其基本信息,包括训练与测试样本数、维数、类别或标签数以及平均标签数.

表 1 实验数据集的基本信息

Table 1 Base information of experimental dataset

数据集	训练样本数	测试样本数	维数	类别数	平均标签数
Bibtex	4 880	2 515	1 836	159	2.38
Cal500	300	202	68	174	26.107
Enron	1 123	579	1 001	53	3.386
Emotions	391	202	72	6	1.813
Image135	1 200	800	135	5	1.245
Medical	645	333	1 449	45	1.24
RCV-s1	3 000	3 000	47 236	101	3.23
Scene	1 211	1 196	294	6	1.062
Tmc2007	21 519	7 077	30 438	22	2.26
Yeast	1 500	917	103	14	4.228

在文献[2]中收集 20 个多标签算法的性能评价指标,本文实验采用了 3 个综合性的指标: F1、宏 F1

(Macro F1) 和微 F1 (Micro F1). 令多标签测试样本集为 $\{(x_1, y_n), \dots, (x_i, y_i), \dots, (x_l, y_n)\}$, 其中测试样本 x_i 的真实二进制标签向量为 y_i , 某一个标签检测技术得到的对应预测标签向量为 y_i^p . 对样本 x_i 计算:

$$\begin{aligned} TP_i &= |j| y_{ij}^p = 1, y_{ij} = +1, j=1, \dots, c|, \\ FP_i &= |j| y_{ij}^p = 1, y_{ij} = -1, j=1, \dots, c|, \\ FN_i &= |j| y_{ij}^p = -1, y_{ij} = 1, j=1, \dots, c|. \end{aligned} \quad (20)$$

则 F1 的定义为:

$$F1 = \frac{1}{n} \sum_{i=1}^n \frac{2TP_i}{2TP_i + FP_i + FN_i}. \quad (21)$$

另一方面, 针对标签 j 计算:

$$\begin{aligned} TP^j &= |i| y_{ij}^p = 1, y_{ij} = +1, i=1, \dots, n|, \\ FP^j &= |i| y_{ij}^p = 1, y_{ij} = -1, i=1, \dots, n|, \\ FN^j &= |i| y_{ij}^p = -1, y_{ij} = 1, i=1, \dots, n|. \end{aligned} \quad (22)$$

则宏 F1 和微 F1 的定义分别为:

$$\begin{aligned} \text{Macro F1} &= \frac{1}{c} \sum_{j=1}^c \frac{2TP^j}{(2TP^j + FP^j + FN^j)}, \\ \text{Micro F1} &= \frac{2 \sum_{j=1}^c TP^j}{\sum_{j=1}^c (2TP^j + FP^j + FN^j)}. \end{aligned} \quad (23)$$

从 3 个指标的定义可以看出, F1 先处理每一个样本, 再对所有样本取平均; 宏 F1 先处理每一个类, 再对所有类求平均; 而微 F1 先转化成两类问题再进行计算, 所以三者指标从不同的角度评价多标签分类算法的性能, 另外其取值范围在 0 ~ 1 之间, 且值越大算法性能越好.

在实验中, 本文采用训练-测试模式, 训练集与测试集仅分别用于训练和测试阶段. 对于原始多标签 k NN 算法中 k 值和距离度量, 本文取文献 [8, 14] 推荐的 $k=10$ 和欧氏距离. 对于 Logistic 回归法, 学习步长为 0.1 ~ 0.3, 迭代次数一般为训练集样本总数的 2 倍. 为了避免矩阵 $Z^T Z$ 的奇异性, 在其主对角元素上加上一个微小的常数 0.001. 本实验的微机配置为主频 3.3 GHz 的四核处理器, 内存 2 G.

3.2 计算时间的比较

首先记录 4 种标签检测技术的训练时间和测试时间, 如表 2、表 3 所示. 由于完成整个多标签分类仍需要 k NN 过程, 所以计算时间中包含着 k NN 的计算时间. 为了综合评价 4 项技术的计算代价, 取每一项技术在 10 个数据集的平均值.

从表 2 可以看出, LRTM 的平均计算时间最少, MOLRM 技术次之, Logistic 最差. 在样本数小于 5 000 的 9 个数据集上, MOLRM 与 Bayes 之间的差别不是太明显. 从表 3 可知, 4 项技术的测试时间差别不明显.

表 2 4 种标签检测技术的训练时间比较

Table 2 Train time comparison of four label detection

	techniques				s
数据集	Bayes	Logistic	LRTM	MOLRM	
Bibtex	75.1	7 367.4	94.3	87.3	
Cal500	0.1	329.9	0.3	0.3	
Emotions	0.1	18.9	0.1	0.1	
Enron	3.4	207.3	3.6	3.6	
Image135	3.9	29.2	2.9	4.0	
Medical	2.7	0.9	0.3	0.6	
RCV1-s1	9.5	1 900.9	11.8	10.5	
Scene	4.7	70.5	3.6	4.9	
Tmc2007	681.9	1 494.8	703.3	672.1	
Yeast	8.3	10.0	6.188	8.6	
平均时间	105.3	1 143.0	82.6	99.4	

表 3 4 种标签检测技术的测试时间比较

Table 3 Test time comparison of four label detection

	techniques				s
数据集	Bayes	Logistic	LRTM	MOLRM	
Bibtex	43.8	17.1	41.1	53.0	
Cal500	0.1	0.2	0.1	0.2	
Emotions	0.1	0.1	0.1	0.1	
Enron	5.2	3.4	4.9	9.0	
Image135	3.2	3.3	2.5	3.3	
Medical	1.0	87.7	0.9	2.1	
RCV1-s1	11.6	14.2	10.3	14.6	
Scene	4.0	4.7	3.2	4.7	
Tmc2007	259.0	299.6	272.7	308.3	
Yeast	9.9	7.9	7.8	10.3	
平均时间	33.8	43.8	34.4	40.6	

3.3 分类性能的比较

表 4~表 6 分别列出了 10 个测试集上的 F1、宏 F1 和微 F1 指标值,其中每一个数据上最好的值用黑体表示.为了排序算法的性能,本文采用置信水平为 5% 的成对 t 检验对每一对技术进行统计检验,如果 A 技术显著地优于 B 技术,则给 A 打 1 分、B 打 0 分;若两个算法不存在显著性差异,则分别得 0.5 分;最后统计每一技术得分,得分值越高,表示算法性能越好.

表 4 4 种标签检测技术的 F1 指标值

数据集	Bayes	Logistic	LRTM	MOLRM
Bibtex	0.258 9	0.160 1	0.186 7	0.290 5
Cal500	0.329 4	0.313 8	0.254 5	0.338 8
Emotions	0.617 7	0.671 1	0.622 6	0.646 7
Enron	0.435 6	0.413 4	0.398 8	0.458 6
Image	0.550 0	0.437 7	0.519 0	0.551 7
Medical	0.684 6	0.645 0	0.575 7	0.678 1
Rcv1-s1	0.694 1	0.629 4	0.630 4	0.674 1
Scene	0.725 2	0.684 4	0.695 9	0.725 5
Tmc2007	0.581 1	0.560 4	0.529 6	0.569 9
Yeast	0.599 3	0.644 3	0.370 1	0.610 3
t 检验总分	2.0	1.0	0.5	2.5

从表 4 和表 6 可以看出, MOLRM 的 t 检验总得分最高, Bayes 方法次之, LRTM 最差. 在表 5 中, 虽然 MOLRM 只有 3 个值是最好的, 但它的均值最高, 检验发现它和 Logistic 和 LRTM 存在显著差异, 与 Bayes 无显著差异; Bayes 方法拥有 4 个最好值, 但它只与 Logistic 存在显著差异, 因此它的得分次之; Logistic 方法最差. 综上, 在比较的 4 种检测算法中, MOLRM 的算法性能是最好的, Bayes 次之, Logistic 和 LRTM 相对较差.

4 结语

当前部分多标签算法本质上由两种分类方法级联而成, 后一项标签检测技术可以有效地检测样本的相关标签, 兼顾进一步改善分类性能. 本文收集与实现 4 种标签检测技术, 并在 10 个基准数据上依据 3 个综合性的评价指标进行实验比较. 本文的研究结果表明, 从计算代价与分类性能两个角度来看, 多输出线性回归法是一项值得推荐的技术, 且同时适用于前一项分类技术的实数与整数输出情况. 当前一项分类器的输出是整数时, 离散 Bayes 规则也是一个值得考虑的技术.

表 5 4 种标签检测技术的宏 F1 指标值

数据集	Bayes	Logistic	LRTM	MOLRM
Bibtex	0.101 5	0.011 6	0.088 5	0.106 0
Cal500	0.055 3	0.044 2	0.108 5	0.163 1
Emotions	0.612 5	0.677 1	0.652 7	0.674 5
Enron	0.088 2	0.055 5	0.110 0	0.108 2
Image135	0.552 4	0.395 7	0.534 1	0.550 9
Medical	0.262 3	0.146 9	0.189 0	0.211 5
Rcv1-s1	0.325 4	0.135 9	0.288 6	0.322 9
Scene	0.722 1	0.692 3	0.696 2	0.724 1
Tmc2007	0.453 1	0.353 9	0.366 1	0.417 6
Yeast	0.336 1	0.384 6	0.212 8	0.355 9
t 检验总分	2.0	0.5	1.0	2.5

表 6 4 种标签检测技术的微 F1 指标值

数据集	Bayes	Logistic	LRTM	MOLRM
Bibtex	0.253 7	0.153 6	0.188 0	0.280 7
Cal500	0.325 5	0.309 0	0.290 9	0.340 8
Image	0.546 8	0.443 6	0.528 8	0.547 4
Emotions	0.656 7	0.707 4	0.659 5	0.688 6
Enron	0.462 7	0.434 6	0.387 8	0.469 0
Medical	0.677 8	0.643 5	0.523 2	0.672 6
Rcv1-s1	0.700 5	0.634 4	0.648 7	0.687 5
Scene	0.717 7	0.683 3	0.685 6	0.720 6
Tmc2007	0.612 5	0.600 3	0.559 0	0.604 0
Yeast	0.625 0	0.669 3	0.381 9	0.635 8
t 检验总分	2.0	1.0	0.5	2.5

[参考文献] (References)

- [1] Tsoumakas G, Katakis I. Multi-label classification: an overview [J]. International Journal of Data Warehousing and Mining, 2007, 3(3): 1-13.
- [2] Tsoumakas G, Katakis I, Vlahavas I. Mining multi-label data [C]//Maimon O, Rokach L. Data Mining and Knowledge Discovery Handbook. 3rd ed. New York: Springer, 2010: 667-685.
- [3] Xu J. An extended one-versus-rest support vector machine for multi-label classification [J]. Neurocomputing, 2011, 74(17): 3114-3124.
- [4] Madjarov G, Kocev D, Gjorgjevik D, et al. An extensive experimental comparison of methods for multi-label learning [J]. Pattern Recognition, 2012, 45(9): 3084-3104.
- [5] Elisseeff A, Weston J. A kernel method for multi-labelled classification [C]//Proceedings of the 14th Conference on Neural

- Information Processing Systems. Canada: Vancouver 2001: 681-687.
- [6] Zhang M L ,Zhou Z H. Multilabel neural networks with application to function genomics and text categorization [J]. IEEE Transactions on Knowledge and Data Engineering 2006 ,18(10) : 1 338-1 351.
- [7] Spyromitros E ,Tsoumakas G ,Vlahavas I. An empirical study of lazy multilabel classification algorithms [C]//Proceedings of the 5th Hellenic Conference on Artificial Intelligence. Greece: Syros 2008 ,LNAI 5138: 401-406.
- [8] Zhang M L ,Zhou Z H. ML-kNN: a lazy learning approach to multi-label learning [J]. Pattern Recognition ,2007 ,40 (7) : 2 038-2 048.
- [9] Madjarov G ,Gjorgjevik D ,Dzierski S. Two stage architecture for multi-label learning [J]. Pattern Recognition ,2012 ,45 (3) : 1 019-1 034.
- [10] Petrovskiy M ,Glazkova V. Linear methods for reduction from ranking to multilabel classification [C]//Proceedings of the 19th Australia Joint Conference on Artificial Intelligence. Australia: Hobart 2006 ,LNCS 4304: 1 152-1 156.
- [11] Hastie T ,Tibshirani R ,Friedman J. The Elements of Statistical Learning[M]. New York: Springer 2001.
- [12] Duda R O ,Hart P E ,Stork D G. Pattern Classification [M]. 2nd ed. New York: John Wiley and Sons 2001.
- [13] 张学工. 模式识别 [M]. 3 版. 北京: 清华大学出版社 2010.
Zhang Xuegong. Pattern Recognition [M]. 3rd ed. Beijing: Tsinghua University Press 2010. (in Chinese)
- [14] Cheng W W ,Hullermeier E. Combining instance-based learning and logistic regression for multilabel classification [J]. Machine Learning 2009 ,76(2/3) : 211-225.
- [15] Vach W ,Robner P ,Schumacher M. Neural networks and logistic regression: part I [J]. Computational Statistics and Data Analysis ,1996 21(6) : 661-682.
- [16] Vach W ,Robner P ,Schumacher M. Neural networks and logistic regression: part II [J]. Computational Statistics and Data Analysis ,1996 21(6) : 683-701.
- [17] Tsoumakas. Multi-label data sets [DB/OL]. [2012-04-10]. <http://mulan.sourceforge.net/datasets.html> 2010.
- [18] Zhang M L. Image data set [DB/OL]. [2012-04-10]. <http://cse.seu.edu.cn/people/zhangml> 2009.

[责任编辑: 严海琳]