基于标记权重的多标记特征选择算法

林梦雷[1] 刘景华[2] 王晨曦[3] 林耀进[3]
(闽南师范大学数学与统计学院 漳州 363000)[1] (厦门大学自动化系 厦门 361000)[2]
(闽南师范大学计算机学院 漳州 363000)[3]

摘 要 在多标记学习中,特征选择是解决多标记数据高维性的有效手段。每个标记对样本的可分性程度不同,这可能会为多标记学习提供一定的信息。基于这一假设,提出了一种基于标记权重的多标记特征选择算法。该算法首先利用样本在整个特征空间的分类间隔对标记进行加权,然后将特征在整个标记集合下对样本的可区分性作为特征权重,以此衡量特征对标记集合的重要性。最后,根据特征权重对特征进行降序排列,从而得到一组新的特征排序。在6个多标记数据集和4个评价指标上的实验结果表明,所提算法优于一些当前流行的多标记特征选择算法。

关键词 特征选择,标记权重,分类间隔,多标记分类
中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.10.052

Multi-label Feature Selection Algorithm Based on Label Weighting

LIN Meng-lei[1] LIU Jing-hua[2] WANG Chen-xi[3] LIN Yao-jin[3]
(School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China)[1]
(Department of Automation, Xiamen University, Xiamen 361000, China)[2]
(School of Computer Science, Minnan Normal University, Zhangzhou 363000, China)[3]

Abstract In multi-label learning, each sample is described as a feature vector and simultaneously associated with multiple class labels. Feature selection is able to remove irrelevant and redundant features, which is an efficient measure of overcoming the curse of dimensionality for multi-label data. Label has different separability with sample, which may provide some usefull informations for multi-label learning. Based on this assumption, a multi-label feature selection algorithm based on label weighting was proposed in this paper. First, the margin of sample in all feature space is calculated and it is used as label weighting. Then, the distinguishability of feature is adopted based on label set for calculating feature weighting, which will measure the importance degree of feature. Finally, all features are sorted by the value of feature weighting. Experiment was conducted on four multi-label datasets, and four evaluation criteria were used to measure the effectiveness of our method. Experimental results show that the proposed algorithm is superior to several state-of-the-art multi-label feature selection algorithms.

Keywords Feature selection, Label weighting, Classification margin, Multi-label classification

1 引言

多标记学习是机器学习、数据挖掘和模式识别等领域的研究热点之一[1-5]。在多标记学习框架中,每个样本不仅由一组特征向量描述,而且可能同时隶属于多个类别标记。例如:在文本分类问题[1-2]中,一篇文档可能同时与多个主题相关,如“羽毛球赛”、“运动”和“体育”;在图像标注[3]中,一幅图像可能同时具有多个语义标注,如“湖泊”、“树木”和“风景”等;在生物信息学[2]中,每段基因可能同时具有多种功能,如“蛋白质合成”、“新陈代谢”和“转录”等。

在多标记数据中,通常不可避免地涉及到数据的高维性。数据的高维性可能会造成维数灾难,严重干扰多标记分类器的分类性能。降维技术是解决该问题的有效手段。常见的多标记特征降维方法主要包括特征提取和特征选择。特征提取是将原始高维的特征空间进行转换或者映射到一个低维的特征空间的过程。常见的特征提取方法包括线性判别分析(LDA)[6]、特征抽取方法(MDDM)[7]和多标记潜在语义搜索(MLSI)[8]等。特征提取方法虽然能够提高多标记分类器的

到稿日期:2016-09-05 返修日期:2017-02-13 本文受国家自然科学基金(61303131,61379021,61603173),福建省自然科学基金项目(2013J01028),福建省高校新世纪优秀人才支持计划资助。

林梦雷(1963-),男,教授,硕士生导师,主要研究方向为机器学习、粒计算,E-mail:menglei36@126.com;刘景华(1989-),女,博士生,主要研究方向为数据挖掘、机器学习;王晨曦(1981-),女,硕士,讲师,主要研究方向为数据挖掘;林耀进(1980-),男,博士,副教授,主要研究方向为数据挖掘、粒计算。

分类性能，但该方法破坏了原始特征空间，导致新的特征空间失去了物理意义。与特征提取方法不同，特征选择方法是在原始特征空间中利用某种特定的评价准则选择一组最优的特征子集的过程[9-18]。常见的评价准则包括信息度量[9-12]、依赖性度量[13-14]和大间隔[15-19]等。其中，大间隔是从特征对样本的可区分性出发，不仅可以处理混合型数据，而且具有很好的泛化能力。例如，Spolaor 等人[16]基于二元关系方法(BR)和标签幂集法(LP)，利用 Relief (RF)和信息增益(IG)作为评价准则，提出了 4 种基于多标记的特征选择方法(RF-BR，RF-LP，IG-BR 和 IG-LP)；Spolaor 等人[17]将 ReliefF 方法运用于多标记学习中，提出了一种新的多标记特征选择算法(RF-ML)；Reyes 等人[18]将 ReliefF 算法进行扩展，提出了 3 种多标记特征选择算法(ReliefF-ML，PPT-ReliefF 和 RReliefF-ML)；此外，Lin 等人[11]利用大间隔将邻域信息熵扩展至多标记，提出了 3 种多标记特征选择算法(NFNMIopt，NFNMIneu，NFNMIpes)。虽然上述方法取得了较好的实验结果，但是并未考虑标记空间中标记对样本具有不同程度的可分性。

在多标记学习中，每个样本可能隶属于多个不同的标记，每个标记对样本的划分能力也具有明显差异，这可能会为多标记学习提供一定的信息。目前，针对标记对样本的可分性进行特征选择的研究较少，因此本文提出了基于标记权重的多标记特征选择算法(LWMF)。首先，根据样本在特征空间的分类间隔对标记赋予权重，由于不同标记对样本会产生不同的分类，那么在同一特征空间中，样本在各类别标记中的分类间隔也各不相同。在同一特征空间中，在某个类别标记下样本的分类间隔越大，说明该类别标记对样本的可分性越强；反之，分类间隔越小，说明该类别标记对样本的可分性越弱。其次，利用特征加权的方法，构造每个特征在标记集合下对样本的可区分性，以此评估特征质量的优劣。最后，根据权重对特征进行降序排列，从而产生一组新的特征排序。LWMF 算法的主要贡献有：1)通过赋予标记权重的方法，考虑了标记对样本的不同区分性；2)在标记集合中度量特征对样本的可区分性，体现了特征与标记集合之间的相关性；3)该算法的计算复杂度不依赖于任何分类器，且时间复杂度较低。实验结果表明，LWMF 算法能够有效提升多标记分类器预测的分类性能。

本文第 2 节介绍了大间隔的相关知识；第 3 节设计了基于标记权重的多标记特征选择模型；第 4 节针对所提算法进行实验分析；最后总结全文。

2 大间隔

给定一个决策系统 $NDT=\langle U,F,D\rangle$，$U=\{x_1,x_2,\cdots,x_n\}$表示样本的集合，$F=\{f_1,f_2,\cdots,f_m\}$是用于描述样本的一组特征，D 是类别标记。

定义 1[12] 设 U 是非空样本空间，若 $\forall x_i,x_j,x_k\in U$，都存在唯一确定的实函数 Δ 与之对应，而且 Δ 满足：

(1)$\Delta(x_i,x_j)\geqslant 0$ 当且仅当 $x_i=x_j$，$\Delta(x_i,x_j)=0$；

(2)$\Delta(x_i,x_j)=\Delta(x_j,x_i)$；

(3)$\Delta(x_i,x_k)\leqslant\Delta(x_i,x_j)+\Delta(x_j,x_k)$。

则称$\langle U,\Delta\rangle$是度量空间。其中，Δ 是 U 上的距离函数。

在 m 维特征空间中，给定任意两点 $x_i=(x_{1i},x_{2i},\cdots,x_{mi})$和 $y_i=(y_{1i},y_{2i},\cdots,y_{mi})$，一般将两点的距离函数定义为闵科夫斯基距离：

$$\Delta_P(x_i,x_j)=\left[\sum_{l=1}^{N}(x_{li},x_{lj})^P\right]^{\frac{1}{P}}$$

当 $P=1$ 时，Δ 函数表示曼哈顿距离；当 $P=2$ 时，Δ 为欧氏距离；当 $P\to\infty$时，$\Delta_P(x_i,x_j)=\max_l|x_{li},x_{lj}|$。

定义 2[21] 设 U 表示样本空间，x 是给定的样本，则样本 x 的分类间隔定义为：

$$margin(x)=\Delta(x,NM(x))-\Delta(x,NH(x)) \tag{1}$$

其中，$NH(x)$表示在样本空间 U 中与 x 距离最近的同类样本，称为 x 的 Nearest Hit(NH)。而 $NM(x)$表示在样本空间中与样本 x 最近的异类样本，称为 x 的 Nearest Miss (NM)。$\Delta(x-NM(x))$和 $\Delta(x-NH(x))$分别表示样本点 x 到 $NM(x)$和 $NH(x)$的距离(见图 1)。

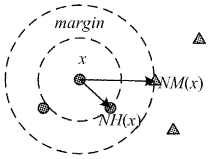


图 1 样本 x 的分类间隔 $margin(x)$

Relief 算法[20]作为经典的基于大间隔的特征选择方法，其主要是将间隔对样本的可区分性用于评估特征质量的优劣。Relief 算法通过样本的间隔迭代得到特征的权：

$$w_i=w_i+\|x_i-NM(x_i)\|-\|x_i-NH(x_i)\| \tag{2}$$

其中，$\|x_i-NM(x_i)\|-\|x_i-NH(x_i)\|$表示样本在第 i 个特征分量上的间隔的 2 倍。

为了动态衡量特征权对间隔的影响，Simba 算法[20]利用每个特征权分量的梯度来迭代更新特征的权重，迭代过程如下：

$$\begin{aligned}w_i&=w_i+\frac{\partial\theta^w(x)}{\partial w_i}\\&=w_i+\frac{1}{2}\left(\frac{(x_i-NM(x_i))^2}{\|x_i-NM(x_i)\|_w}-\frac{(x_i-NH(x_i))^2}{\|x_i-NH(x_i)\|_w}\right)\end{aligned} \tag{3}$$

其中，$\theta^w(x)=\frac{1}{2}(\|x-NM(x)\|_w-\|x-NH(x)\|_w)$是带有权变量的假设间隔，其中 $\|z\|_w=\sqrt{\sum_i w_i^2\cdot z_i^2}$。

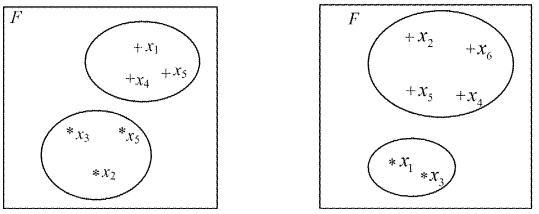
3 基于标记权重的多标记特征选择模型

3.1 标记权重模型

在多标记学习框架中，每个样本可能同时隶属于多个类别标记，而每个类别标记在同一特征空间中会对样本产生不同的分类。

为了更加形象地刻画不同标记在同一特征空间中对样本产生的分类情况，本文通过以下实例加以说明。设给定训练样本 $U=\{x_1,x_2,\cdots,x_6\}$，描述样本的特征空间 $F=\{f_1,f_2,\cdots,f_m\}$，样本可能隶属于标记集合 $L=\{l_1,l_2\}$。图 2 给出了样本在同一特征空间、不同类别标记下的划分情况。假设

标记为“+”的样本属于第一类，标记为“*”的样本属于第二类。图 2(a)中，在特征空间 F 下，训练样本 U 在类别标记 l_1 中可被划分为两类，$\{x_1,x_4,x_5\}$属于第一类，$\{x_2,x_3,x_6\}$属于第二类；从图 2(b)可以看出，对于类别标记 l_2，在同一特征空间 F 下，划分得到的第一类样本为$\{x_2,x_4,x_5,x_6\}$，相应的第二类样本为$\{x_1,x_3\}$。



(a)标记 l_1 在 F 下对样本的分类　　(b)标记 l_2 在 F 下对样本的分类

图 2　样本在同一特征空间、不同类别标记下的划分情况

显然，不同类别标记在同一特征空间 F 下对样本的可分性是各不相同的。根据每个类别标记对样本可分性的不同，本文考虑利用样本在整个特征空间的分类间隔来对标记赋予一定的权重，以此衡量每个标记对样本的可区分性程度。具体见定义 3。

定义 3　给定样本空间$U=\{x_1,x_2,\cdots,x_n\}$，描述样本的一组特征 $F=\{f_1,f_2,\cdots,f_d\}$和标记集合 $L=\{l_1,l_2,\cdots,l_t\}$。对于$\forall l\in L$，在特征空间 F 下对类别标记 l 赋予的权重可定义为：

$$w_l=\sum_{i=1}^{n}(\Delta_F(x_i,NM(x_i))-\Delta_F(x_i,NH(x_i))) \tag{4}$$

其中，距离函数定义为：

$$\Delta_F(x,y)=\sqrt{\sum_{f=1}^{d}(x(f)-y(f))^2} \tag{5}$$

式(4)中，$\Delta_F(x,y)$表示样本 x 和样本 y 在特征空间 F 上的距离，x_i 表示第 i 个样本。式(5)中，$x(f)$和 $y(f)$分别表示样本 x 和样本 y 在特征 f 上的特征值。类别标记在特征空间中对样本的分类间隔越大，对应的标记权重越大，说明该标记对样本的可分性越强；反之，若在同一特征空间中样本的分类间隔越小，则对应的标记权重也相应较小，说明该类别标记对样本的可分性越弱。

由于不同类别标记会对样本产生不同的分类，因此每个类别标记也会诱导样本产生不同的分类间隔，见定义 4。

定义 4　设 U 是样本空间，对应的样本可能隶属于标记集合 $L=\{l_1,l_2,\cdots,l_t\}$中，对于$\forall l\in L$，给定样本 x，则样本 x 在标记 l 下的分类间隔定义为：

$$m^l(x)=\Delta(x,NM^l(x))-\Delta(x,NH^l(x)) \tag{6}$$

其中，$NH^l(x)$表示在类别标记 l 下样本空间 U 中与 x 最近的同类样本；$NM^l(x)$表示在类别标记 l 下与样本 x 最近的异类样本。$\Delta(x-NM^l(x))$和 $\Delta(x-NH^l(x))$则分别表示样本点 x 到 $NM^l(x)$和 $NH^l(x)$的距离。

定义 5　给定训练样本 U，$\forall x\in U$，w 为特征的权重向量，则特征子集的评价函数为：

$$e(w)=\sum_{l\in L}w_l\cdot\sum_{x\in U}m^l_{U\setminus x}(x) \tag{7}$$

其中，最大化的 $w^2=1$。在类别标记集合 L 下，通过最大化假设间隔来对特征赋予权值，那么特征 f 的权值 w_f 的计算方式可定义为：

$$w_f=\sum_{l=1}^{t}w_l\cdot\sum_{i=1}^{n}d_f(x_i,NM^l(x_i))-\sum_{l=1}^{t}w_l\cdot\sum_{i=1}^{n}(d_f(x_i,NH^l(x_i)) \tag{8}$$

又可以简化式(8)为：

$$w_f=\sum_{l=1}^{t}w_l\cdot\sum_{i=1}^{n}(d_f(x_i,NM^l(x_i))-d_f(x_i,NH^l(x_i))) \tag{9}$$

其中，$d_f(x_i,NM^l(x_i))$和 $d_f(x_i,NH^l(x_i))$分别表示在特征 f 下样本 x_i 在类别标记 l 中与其最近的异类样本的距离和最近的同类样本的距离。若在类别标记 l 下，样本 x_i 不存在与其最近的异类样本，则令 $d_f(x_i,NM^l(x_i))=0$；若样本 x_i 在类别标记 l 下不存在与其最近的同类样本，则令 $d_f(x_i,NH^l(x_i))=0$。

本文将距离 $d_f(x,y)$定义为：

$$d_f(x,y)=\frac{|x(f)-y(f)|}{\max(f)-\min(f)} \tag{10}$$

其中，$x(f)$和 $y(f)$分别表示样本 x 和样本 y 在特征 f 上的特征值。$\max(f)$和 $\min(f)$分别表示特征 f 在样本空间中取得的最大值与最小值。

3.2　基于标记权重的多标记特征选择算法

在多标记数据中，每个样本可能同时隶属于多个类别标记，而每个类别标记对样本有着不同程度的可辨别性。正是基于此，本文通过对标记赋予一定的权重来探索各类别标记对样本的可区分性程度，从而提出了基于标记权重的多标记特征选择算法。由于不同类别标记在同一特征空间对样本会产生不同的分类，因此样本在每个类别标记下均能计算得到不同的分类间隔。同一特征空间中样本的分类间隔越大，说明该类别标记对样本的可分性越强；反之，分类间隔越小，表明该类别标记越难以区分样本。该算法的主要思想为：首先，针对给定的类别标记 l，计算各样本在整个特征空间下的分类间隔，并将分类间隔作为类别标记 l 的权重；其次，结合标记权重，利用特征对样本的可分性来计算特征的权重，通过特征权重可以有效地衡量特征对整个标记集合的重要性程度；最后，对特征权重进行排序，从而得到一组新的特征排序。

根据以上分析，基于标记权重的多标记特征选择算法(LWMF 算法)的具体描述如算法 1 所示。

算法 1　LWMF 算法

输入：多标记数据集 D

输出：特征排序 rank

1. for each l∈L
2.　根据定义 3 计算每个类别标记 l 的权重 w_l；
3. end
4. 对标记集合的权重 $w_L=\{w_1,w_2,\cdots,w_t\}$归一化；
5. for each f∈F
6.　for each l∈L
7.　　结合定义 4 和式(9)、式(10)计算每个特征 f 的权重；
8.　end
9. end
10. 对得到的特征权重进行排序，从而得到一组特征排序 rank。

在算法 1 中，假设特征集合包含 d 个特征和 t 个类别标记。步骤 1－步骤 3 计算标记的权重，计算代价为 $O(t)$；步骤 5－步骤 9 计算每个特征的权重，计算的时间复杂度为 $O(d\times t)$；步骤 10 对特征排序的计算代价为 $O(d\log d)$；LWMF 算法的计算代价主要在于计算特征权重 $O(t)$ 和标记权重 $O(d\times t)$，并不依赖于任何分类器。

4 实验设计与结果比较

4.1 实验数据

为了验证本文所提算法的有效性，选取 6 个数据集进行实验，各数据集相应的描述信息如表 1 所列，数据集来自 http://mulan.sourceforge.net/datasets.html。

表 1 多标记数据集的描述

数据集	样本数	特征数	类别数	训练样本数	测试样本数
Arts	5000	462	26	2000	3000
Education	5000	550	33	2000	3000
Recreation	5000	606	22	2000	3000
Reference	5000	793	33	2000	3000
Social	5000	1047	39	2000	3000
Yeast	2417	103	14	1499	918

4.2 实验设置

本文实验令测试集为 $Z=\{(x_i,Y_i)\}_{i=1}^{m}\subset R^d\times\{+1,-1\}^L$，根据预测函数 $f_l(x)$ 可定义排序函数为 $rank_f(x,l)\in\{1,\cdots,L\}$。将 Average Precision (AP)，Ranking Loss (RL)，Hamming Loss(HL)和 Coverage(CV)作为分类性能的评价指标[23]。

Average Precision (AP)：用于考察所有样本的预测标记排序中位置排在该样本标记前面的标记仍属于该样本标记的概率的平均，定义为：

$$avgPre(f)=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{|R_i|}\sum_{l\in R_i}\frac{\{k\mid rank_f(x_i,k)\leqslant rank_f(x_i,l),k\in R_i\}}{rank_f(x_i,l)}$$

Ranking Loss (RL)：用来考察所有样本的不相关标记的排序排在相关标记前面的概率的平均，定义为：

$$rLoss(f)=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{|R_i||\overline{R}_i|}|\{(l,k)\mid rank_f(x_i,l)\geqslant rank_f(x_i,k),(l,k)\in R_i\times\overline{R}_i\}|$$

Hamming Loss(HL)：用于度量样本在单个类别标记上误分类的情况，定义为：

$$hLoss(h)=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{L}\sum_{l=1}^{L}[h_l(x_i)\neq Y_{il}]$$

Coverage(CV)：用于度量样本遍历所有与其相关的类别标记平均所需要的步数，定义为：

$$coverage(f)=\frac{1}{m}\sum_{i=1}^{m}\max_{l\in R_i}rank_f(x_i,l)-1$$

其中，$R_i=\{l\mid Y_{il}=+1\}$ 表示与样本 x_i 相关的标记构成的集合，$\overline{R}_i=\{l\mid Y_{il}=-1\}$ 表示与样本 x_i 不相关的标记集合。

AP 指标取值越大，说明分类的性能越优，最优值为 1；RL，HL 以及 CV 的指标取值越小，说明分类的性能越优，最优值为 0。

为了充分验证实验的有效性，本文从不同角度选择了 4 种对比算法，包括 MDDMspc[7]，MDDMproj[7]，MLNB[23] 和 RF-ML[17]。其中，MDDM 是基于最大依赖的多标记维数约简方法，其根据使用线性核和非线性核的情况又可分为 MDDMspc 和 MDDMproj；MLNB 基于主成分分析法 PCA 和遗传算法 GA，同时利用贝叶斯分类器的方法实现特征提取；RF-ML 算法是利用分类间隔对特征加权，然后利用特征对标记的可分性强弱进行特征选择。另外，本实验采用分类算法 ML-kNN 来评估特征选择后的数据。根据文献[24]，将 ML-kNN 的平滑参数 s 设置为 1，近邻个数 k 设置为 10。

4.3 实验结果与分析

为了验证所提算法 LWMF 的有效性，实验首先比较各种算法诱导出来的特征子集的分类性能，并分析各算法的分类性能随特征数目的变化情况。其中，MDDMspc，MDDMproj，RF-ML 和 LWMF 算法得到的是一组特征排序，因此在实验中将取特征排序的前 k 个特征作为特征子集。设 k 等于 MLNB 算法得到的特征数目。

表 2－表 5 列出了 5 种算法分别在 4 种评价指标上的实验结果。

表 2 AP 评价指标下各算法的性能比较(↑)

数据集	MDDMspc	MDDMproj	RF-ML	MLNB	LWMF
Arts	0.5072	0.4943	0.4944	0.4991	**0.5118**
Education	0.5389	0.5425	0.5365	0.5478	**0.5539**
Recreation	0.4717	0.4703	0.4365	0.4790	**0.4859**
Reference	0.6126	0.6106	0.6169	0.6234	**0.6247**
Social	0.6941	0.6914	0.6513	0.7047	**0.7058**
Yeast	0.7213	0.7210	**0.7473**	0.7355	**0.7473**
Average	*0.5910*	*0.5884*	*0.5805*	*0.5983*	***0.6049***

表 3 RL 评价指标下各算法的性能比较(↓)

数据集	MDDMspc	MDDMproj	RF-ML	MLNB	LWMF
Arts	0.1521	0.1555	0.1527	0.1542	**0.1482**
Education	0.0914	0.0924	0.0939	0.0922	**0.0897**
Recreation	0.1838	0.1859	0.1955	0.1879	**0.1834**
Reference	0.0888	0.0889	**0.0856**	0.0889	0.0867
Social	0.0686	0.0682	0.0696	0.0682	**0.0660**
Yeast	0.1990	0.2041	0.1815	0.1871	**0.1808**
Average	*0.1306*	*0.1325*	*0.1298*	*0.1298*	***0.1258***

表 4 HL 评价指标下各算法的性能比较(↓)

数据集	MDDMspc	MDDMproj	RF-ML	MLNB	LWMF
Arts	0.0607	0.0612	0.0615	0.0612	**0.0604**
Education	0.0426	0.0422	0.0425	0.0405	**0.0403**
Recreation	0.0620	0.0616	0.0633	0.0611	**0.0606**
Reference	0.0322	0.0311	0.0306	**0.0296**	0.0303
Social	0.0272	0.0268	0.0303	**0.0248**	0.0252
Yeast	0.2209	0.2246	0.2089	0.2080	**0.2042**
Average	*0.0743*	*0.0746*	*0.0729*	*0.0709*	**0.0702**

表 5 CV 评价指标下各算法的性能比较(↓)

数据集	MDDMspc	MDDMproj	RF-ML	MLNB	LWMF
Arts	5.4740	5.5553	5.4917	5.5040	**5.3647**
Education	3.8987	3.9203	3.9920	3.9183	**3.8513**
Recreation	4.9403	4.9470	5.1367	4.9953	**4.9230**
Reference	3.4390	3.4460	**3.3660**	3.4313	3.3723
Social	3.5803	3.5670	3.6227	3.6007	**3.4850**
Yeast	6.8137	6.8181	6.4913	6.6928	**6.5131**
Average	*4.6910*	*4.7090*	*4.6834*	*4.6904*	***4.5849***

对于给定的评价指标，符号“↑”表示该评价指标的取值越大，分类性能越优；符号“↓”表示该评价指标的取值越小，分类性能越优；此外，用黑体表示各算法中性能最优的结果，斜体表示平均分类性能。

根据表 2－表 5 的结果可以发现：

(1) 对于 AP 指标，LWMF 算法在 6 个数据集上取得的分类精度都是最大值，即分类性能均取得最优；对于 RL 和 CV 指标，除了在 Reference 数据集上 LWMF 略差于 RF-ML 外，在其他 5 个数据集上 LWMF 均取得最优值；对于 HL 指标，LWMF 算法在 4 个数据集上取得最优，在 Reference 和 Social 数据集上其性能略低于 MLNB 算法，但仅相差 0.0007 和 0.0004。

(2)从统计的 6 个数据集、4 个评价指标的 24 种对比结果可知，与 LWMF 对比，MDDMspc 和 MDDMproj 没有胜出的情况，RF-ML 和 MLNB 胜出的结果均占 8.33%。而相比于 4 种对比算法，LWMF 算法胜出的结果占 83.33%。

(3)在平均分类性能方面，LWMF 在 4 种评价指标下均明显优于其他 4 种对比算法。

总之，从特征子集诱导出来的分类性能上看，LWMF 算法的性能排在第一位，其次是 MLNB 和 RF-ML 算法，最后分别是 MDDMproj 和 MDDMspc 算法。但即使得到的特征子集的分类性能最优，也并不能从全局上洞察该算法的分类性能随着特征数目的变化趋势。

为了能够从整体上直观地对比各个算法的分类性能随着特征数目的变化情况，图 3－图 6 分别给出了在 AP，RL，HL 和 CV 性能评价指标下各种算法在数据集上的分类性能随着特征数目的变化趋势。

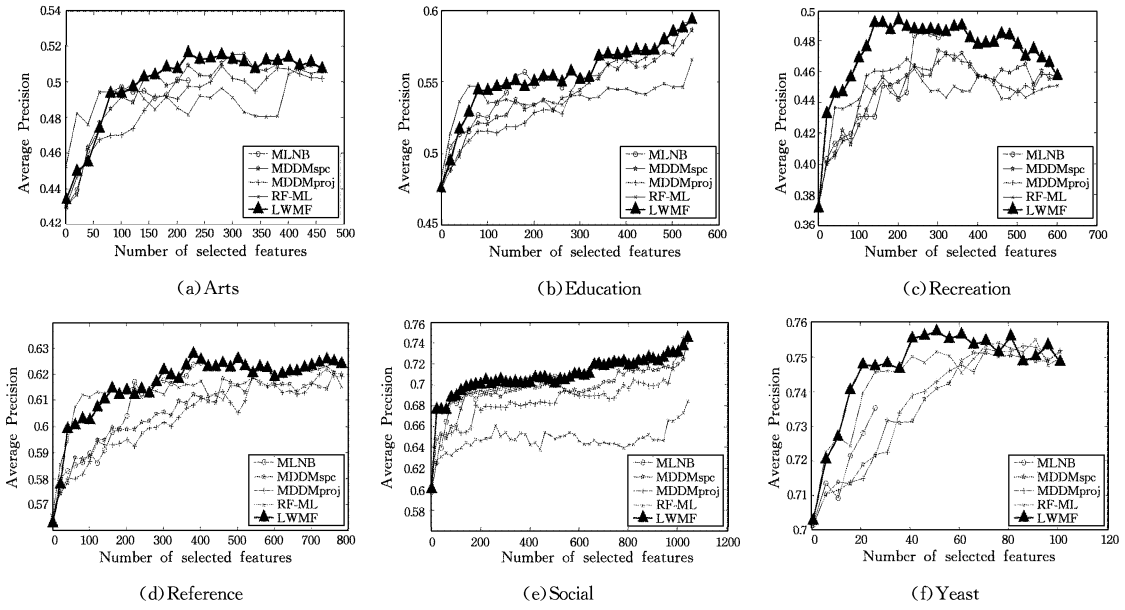


图 3　AP 评价指标下各算法分类性能的变化情况

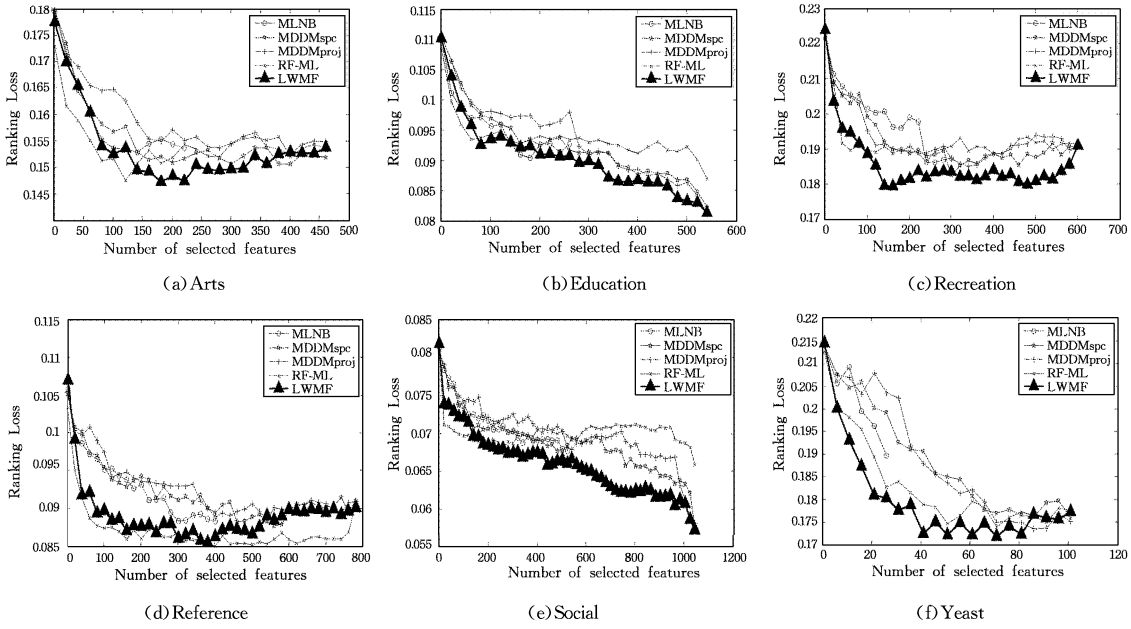


图 4　RL 评价指标下各算法分类性能的变化情况

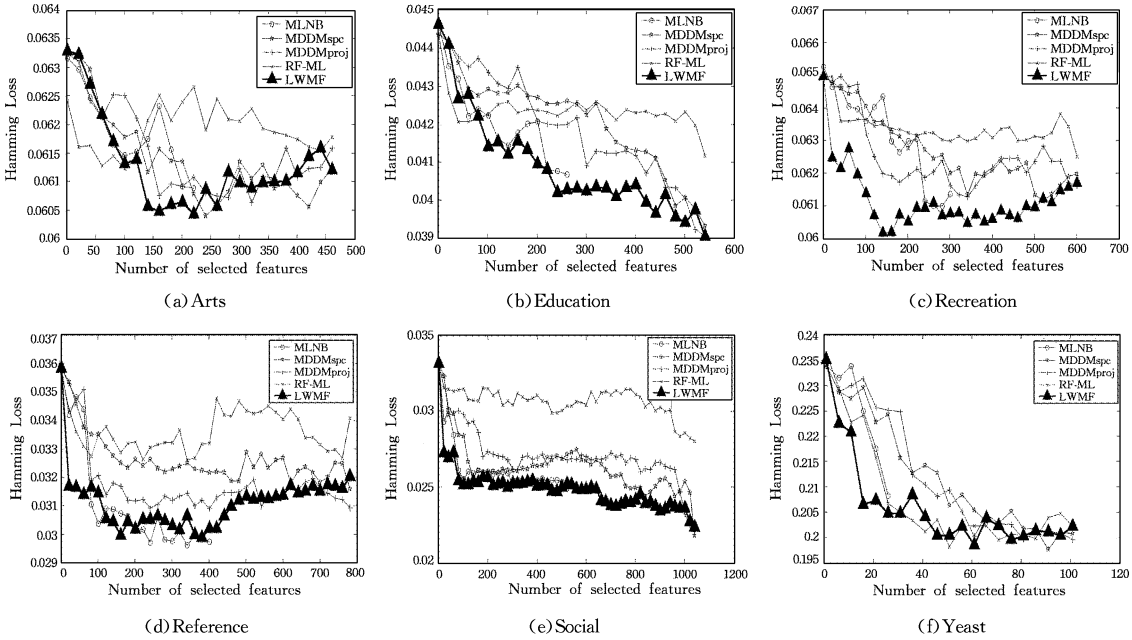


图 5 HL 评价指标下各算法分类性能的变化情况

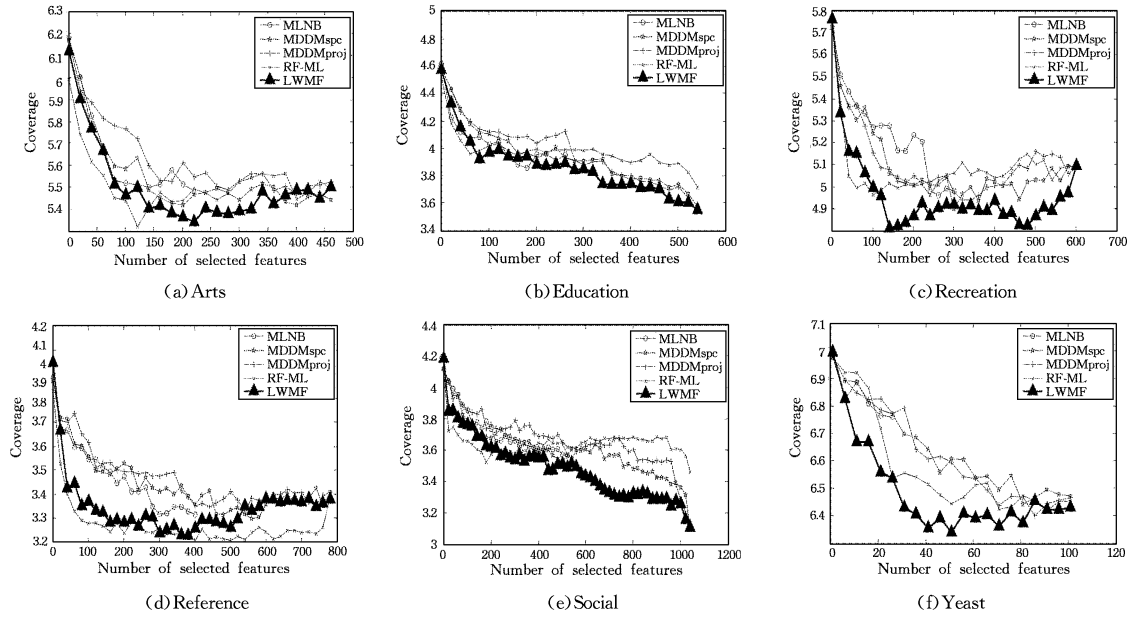


图 6 CV 评价指标下各算法分类性能的变化情况

对于每种评价指标，根据图 3—图 6 可以发现：

(1)对于 AP 指标，LWMF 算法的分类性能曲线都明显位于各算法之上，除在 Education 数据集上特征数据量为150～200时略低于 MLNB；对于 RL 和 CV 指标，除了在 Re-ference 数据集上略低于 RF-ML 算法之外，LWMF 算法的分类性能变化曲线均明显优于其他对比算法。而对于 HL 指标，LWMF 算法在 Education，Recreation 和 Social 数据集上的性能变化曲线明显优于其他对比算法。

(2)从整体上看，在 24 种(4 个评价准则，6 个数据集)对比结果中，本文所提算法与 MDDMspc 和 MDDMproj 相比，胜出比例占 95.83%；与 MLNB 相比，LWMF 算法胜出的比例占 91.67%，两算法性能相当的比例占 4.16%；与 RF-ML 相比，LWMF 算法胜出的比例占 87.5%，两算法性能相当的比例占 4.16%。

总之，将特征子集诱导出来的分类性能和分类性能随特征数目增加的变化趋势相结合进行分析，更充分地说明了本文所提算法的有效性。

为了从统计上比较 LWMF 算法与其他 4 种对比算法，本文采取了显著性水平为 10% 的 Friedman test[25] 进行检验。对于实验中的每个评价指标，经验证都拒绝了零假设，即所有算法的性能都相等。因此，需要结合特定的 post-hoc test 来进一步分析各算法性能的差异。本文采用显著性水平为 10% 的 Bonferroni-Dunn test[26]，并将本文所提 LWMF 算法作为控制算法。若对比算法与控制算法在所有数据集上的平均排序高于临界差(Critical Difference，CD)，则认为两个算法之间具有显著性差异。表 6 列出了各算法在不同评价指标下的平均排序结果。图 7 给出了在不同评价准则下对比算法与控制算法之间的比较结果。其中，每个子图中左上方的一根

粗线表示临界值 $CD=2.24$（5 种算法、6 个数据集），坐标轴示出了各种算法的平均排序且最左边的平均排序最高，用一根加粗的线连接与控制算法 LWMF 之间的性能没有显著差异的算法。

表 6　各算法在不同评价指标下的平均排序

评价指标	MDDMspc	MDDMproj	RF-ML	MLNB	LWMF
AP	3.33	4.33	3.92	2.33	1.08
RL	2.83	4.00	3.50	3.50	1.17
HL	4.00	3.58	4.17	1.92	1.33
CV	2.83	4.00	3.33	3.50	1.33

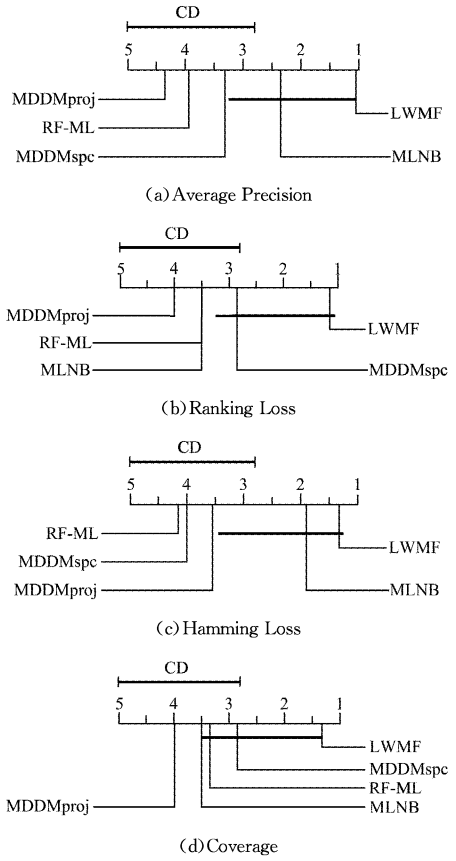


图 7　在 4 种评价准则上的性能比较

对于每种算法，都有 16 种实验比较结果（4 种对比算法、4 种评价准则）。从图 7 可以发现，基于 Average Precision 和 Hamming Loss 指标时，LWMF 算法优于 MDDMspc，MDDMproj 和 RF-ML 算法，与 MLNB 算法的性能相当（见图 7(a)和图 7(c)）；基于 Ranking Loss 指标时，LWMF 算法优于 MDDMproj，RF-ML 和 MLNB 算法，与 MDDMspc 的性能相当（见图 7(b)）；基于 Coverage 指标时，LWMF 算法优于 MDDMproj（见图 7(d)）。

总体来说，LWMF 算法性能最优，在统计上优于其他对比算法性能的比例占 62.5%。这些结果与分析进一步验证了本文算法的有效性。

结束语　本文提出了基于标记权重的多标记特征选择算法（LWMF）。通过样本在整个特征空间中的分类间隔对类别标记加权，从而挖掘类别标记对样本的区分性强弱。在此基础上，构造特征在整个标记集合下对样本的可分性，以此衡量特征的重要性。基于 6 个多标记数据集和 4 种不同的评价准则的实验结果表明，LWML 算法优于其他 4 种当前流行的算法。

参 考 文 献

[1] SCHAPIRE R, SINGER Y. BoosTexter: A boosting-based system for text categorization [J]. Machine Learning, 2000, 39(2/3): 135-168.

[2] ZHANG M, ZHOU Z. Multi label neural networks with applications to functional genomics and text categorization [J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(10): 1338-1351.

[3] BOUTELL M, LUO J, SHEN X, et al. Learning multi-label scene classification [J]. Pattern Recognition, 2004, 37(9): 1757-1771.

[4] ZHENG X Y, ZHANG H X. Multiple Label Approach Based on Local Correlation of Neighbors [J]. Computer Science, 2014, 41(2): 123-126. (in Chinese)
郑希源，张化祥. 基于局部近邻相关性的多标记算法[J]. 计算机科学，2014，41(2)：123-126.

[5] 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. (in Chinese)
何志芬，杨明，刘会东. 多标记分类和标记相关性的联合学习[J]. 软件学报，2014，25(9)：1967-1981.

[6] HOTELLING H. Relations between two sets of variates [J]. Biometrika, 1936, 28(3/4): 321-377.

[7] ZHANG Y, ZHOU Z. Multi-Label dimensionality reduction via dependence maximization [J]. Transactions on Knowledge Discovery from Data, 2010, 4(3): 21-41.

[8] YU K, YU S, TRESP V. Multi-label informed latent semantic indexing [C]// Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY: ACM, 2005: 258-265.

[9] LIU J H, LIN M L, WANG C X, et al. Multi-label Feature Selection Algorithm Based on Local Subspace [J]. Pattern Recognition and Artificial Intelligence, 2016, 29(3): 240-251. (in Chinese)
刘景华，林梦雷，王晨曦，等. 基于局部子空间的多标记特征选择算法[J]. 模式识别与人工智能，2016，29(3)：240-251.

[10] LIN Y, HU Q, LIU J, et al. Multi-label feature selection based on max-dependency and min-redundancy [J]. Neurocomputing, 2015, 168(c): 92-103.

[11] LIN Y, HU Q, LIU J. et al. Multi-Label Feature Selection Based on Neighborhood Mutual Information [J]. Applied Soft Computing, 2016, 38(c): 244-256.

[12] WANG C X, LIN M L, LIU J H, et al. Multi-label feature selection via fusing feature ranking [J]. Computer Engineering and Applications, 2016, 52(17): 93-100. (in Chinese)
王晨曦，林梦雷，刘景华，等. 融合特征排序的多标记特征选择算法[J]. 计算机工程与应用，2016，52(17)：93-100.

（下转第 317 页）

tern Analysis & Machine Intelligence, 1998, 20(11): 1254-1259.

[16] ESCALERA S, RADEVA P, PUJOL O. Complex salient regions for computer vision problems[C]// Preceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2007: 1-8.

[17] KIM W, JUNG C, KIM C. Spatiotemporal saliency detection and its applications in static and dynamic scenes[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2011, 21(4): 446-456.

[18] MEUR O L, THOREAU D, CALLET P L, et al. A spatio-temporal model of the selective human visual attention[C]// Preceedings of the IEEE International Conference on Image Processing. 2005: 1188-1191.

[19] ZHU Y P, JACOBSON N, PAN H, et al. Motion-decision based spatiotemporal saliency for video sequences[C]// Preceedings of the International Conference on Acoustics, Speech&Signal Processing. 2011: 1333-1336.

[20] ZHOU B L, HOU X D, ZHANG L Q. A phase discrepancy analysis of object motion[C]// Preceedings of the Asian Conference on Computer Vision. 2010: 225-238.

[21] BIAN P, ZHANG L M. Biological plausibility of spectral domain approach for spatiotemporal visual saliency[C]// Preceedings of the International Conference on Neuro-Information Processing, Auckland. 2008: 251-258.

[22] LI F X, KIM T, HUMAYUN A, et al. Video segmentation by tracking many figure-ground segments[C]// Preceedings of the IEEE International Conference on Computer Vision. 2013: 2192-2199.

[23] AKAMINE K, FUKUCHI K, KIMURA A, et al. Fully automatic extraction of salient objects from videos in near real time [J]. Computer Journal, 2010, 55(1): 3-14.

[24] HUANG C R, CHANG Y J, YANG Z X, et al. Video saliency map detection by dominant camera motion removal[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2014, 24(8): 1336-1349.

[25] KREYSZIG E. Introductory mathematical statistics principles and methods[J]. Technometrics, 1971, 13(4): 922-925.

[26] OHTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems Man & Cybernetics, 1979, 9(1): 62-66.

[27] BORJI A, CHENG M M, JIANG H, et al. Salient Object Detection: A Survey[J]. Eprint Arxiv, 2014, 16(7): 3118.

[28] MA Y F, ZHANG H J. A new perceived motion based shot content representation[C]// Preceedings of the IEEE International Conference on Image Processing. 2001: 426-429.

[29] MA Y F, ZHANG H J. A model of motion attention for video skimming [C]// Preceedings of the IEEE International Conference on Image Processing. 2002: 129-132.

[30] LIU Z, YAN H B, SHEN L Q, et al. A motion attention model based rate control algorithm for H. 264/AVC[C]// Preceedings of the IEEE/ACIS International Conference on Computer and Information Science. 2009: 568-573.

（上接第 295 页）

[13] ZHANG L, HU Q, DUAN J, et al. Multi-label Feature Selection with Fuzzy Rough Sets [M]// Rough Sets and Knowledge Technology. Springer International Publishing, 2014: 121-128.

[14] DUAN J, HU Q H, ZHANG L J, et al. Feature Selection for Multi-Label Classification Based on Neighborhood Rough Set [J]. Journal of Computer Research and Development, 2015, 52(1): 56-65. (in Chinese)
段洁，胡清华，张灵均，等. 基于邻域粗糙集的多标记分类特征选择算法[J]. 计算机研究与发展，2015，52(1)：56-65.

[15] SPOLAOR N, CHERMAN E, MONARD M. Using ReliefF for multi-label feature selection[C]// Conferencia Latinoamericana de Informática. 2011: 960-975.

[16] SPOLAOR N, CHERMAN E, MONARD M, et al. A comparison of multi-label feature selection methods using the problem transformation approach [J]. Electronic Notes in Theoretical Computer Science, 2013, 292: 135-151.

[17] SPOLAOR N, CHERMAN E, MONARD M, et al. ReliefF for multi-label feature selection[C]// 2013 Brazilian Conference on Intelligent Systems (BRACIS). IEEE, 2013: 6-11.

[18] REYES O, MORELL C, VENTURA S. Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context [J]. Neurocomputing, 2015, 161: 168-182.

[19] LI J H, FU J F, JIANG W J, et al. Feature Selection Method Based on MRMR for Text Classification[J]. Computer Science, 2016, 43(10): 225-228. (in Chinese)
李军怀，付静飞，蒋文杰，等. 基于 MRMR 的文本分类特征选择方法[J]. 计算机科学，2016，43(10)：225-228.

[20] SUN Y. Iterative RELIEF for feature weighting: algorithms, theories, and applications [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1035-1051.

[21] GILAD-BACHRACH R, NAVOT A, TISHBY N. Margin based feature selection-theory and algorithms [C]// Proceedings of the Twenty-first International Conference on Machine Learning. ACM, 2004: 43.

[22] TSOUMAKAS G, VLAHAVAS I. Random k-label sets: An ensemble method for multi-label classification [C]// European Conference on Machine Learning. 2007: 406-417.

[23] ZHANG M, PEÑA J, ROBLES V. Feature selection for multi-label naive Bayes classification [J]. Information Sciences, 2009, 179(19): 3218-3229.

[24] ZHANG M, ZHOU Z. ML-KNN: A lazy learning approach to multi-label learning [J]. Pattern Recognition, 2007, 40(7): 2038-2048

[25] FRIEDMAN M. A comparison of alternative tests of significance for the problem of m rankings [J]. The Annals of Mathematical Statistics, 1940, 11(1): 86-92.

[26] DUNN O. Multiple comparisons among means [J]. Journal of the American Statistical Association, 1961, 56(293): 52-64.