

回归核极限学习机的多标记学习算法

王一宾^{1 2} 程玉胜^{1 2} 何 月¹ 裴根生¹

摘 要 基于极限学习机(ELM)的多标记学习算法多使用 ELM 分类模式,忽略标记之间存在的相关性.为此,文中提出结合关联规则与回归核极限学习机的多标记学习算法(ML-ASRKELM).首先通过关联规则分析标记空间,提取标记之间的规则向量.然后通过提出的多标记回归核极限学习机(ML-RKELM)得出预测结果.若规则向量不为空,将规则向量与预测结果运算得出最终预测结果,否则最终结果即为 ML-RKELM 的预测结果.对比实验表明 ML-ASRKELM 与 ML-RKELM 性能较优,统计假设检验进一步说明文中算法的有效性.

关键词 多标记学习,极限学习机(ELM),标记相关性,关联规则,回归拟合

引用格式 王一宾,程玉胜,何月,裴根生.回归核极限学习机的多标记学习算法.模式识别与人工智能,2018,31(5):419-430.

DOI 10.16451/j.cnki.issn1003-6059.201805004

中图法分类号 TP 391

Multi-label Learning Algorithm of Regression Kernel Extreme Learning Machine

WANG Yibin^{1 2}, CHENG Yusheng^{1 2}, HE Yue¹, PEI Gensheng¹

ABSTRACT In the multi-label learning algorithms based on extreme learning machine(ELM), the ELM classification model is often used, and the correlation between labels is ignored. Accordingly, a multi-label learning algorithm of regression kernel extreme learning machine with association rules(ML-ASRKELM) is proposed in this paper. Firstly, the rule vectors between labels are extracted by analyzing the association rules of label space. Then, the prediction results are obtained by the proposed multi-label regression kernel extreme learning machine(ML-RKELM). Eventually, if the rule vectors are not empty, the final results are calculated by the rule vectors and the prediction results of ML-RKELM. Otherwise, the final results are predicted by ML-RKELM. The experimental results show that ML-ASRKELM and ML-RKELM are superior to other algorithms, and the effectiveness of the proposed algorithms are illustrated by the statistical hypothesis test.

Key Words Multi-label Learning, Extreme Learning Machine(ELM), Label Correlations, Association Rules, Regression Fitting

Citation WANG Y B, CHENG Y S, HE Y, PEI G S. Multi-label Learning Algorithm of Regression Kernel Extreme Learning Machine. Pattern Recognition and Artificial Intelligence, 2018, 31(5): 419-430.

收稿日期: 2017-12-18; 录用日期: 2018-04-08

Manuscript received December 18, 2017;

accepted April 8, 2018

安徽省高校重点科研项目(No.KJ2017A352)、安徽省高校重点实验室基金项目(No.ACAIM160102)资助

Supported by Natural Science Research Funds of Education Department of Anhui Province(No.KJ2017A352), University Key Laboratory of Intelligent Perception and Computing of Anhui Province(No.ACAIM160102)

本文责任编辑 陈松灿

Recommended by Associate Editor CHEN Songcan

1. 安庆师范大学 计算机与信息学院 安庆 246133

2. 安庆师范大学 安徽省高校智能感知与计算重点实验室 安庆 246133

1. School of Computer and Information, Anqing Normal University, Anqing 246133

2. University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246133

在多标记学习 (Multi-label Learning, ML)^[1] 中, 每个样本与多个标记相关联, 该方法已应用于图像、文本、生物、情感分析等多种实际问题^[2-6].

标记间可能存在某种关系, 利用这种标记关系信息可有效提高多标记分类精度. 目前, 针对多标记学习中标记相关性问题, 根据标记相关性程度可分为一阶、二阶和高阶等策略^[1]. Zhang 等^[7] 扩展 k -最近邻 (KNN) 算法到多标记学习中, 形成 ML-KNN, 采用最大后验 (Maximum a Posteriori, MAP) 原理, 通过相邻样本的标记集合获取信息以预测未标记样本的标记集合. 但 ML-KNN 未考虑标记间的相关性, 属于一阶策略. Elisseeff 等^[5] 提出核技术多标记算法 (Rank-SVM), 利用 SVM 中最大间隔策略以适应多标记学习, 根据排序损失函数为每个标记构造 SVM 分类器, 解决多标记分类问题. Rank-SVM 考虑标记对间的相关性, 属于二阶策略.

在多标记中, 不同标记可能具有自身的一些特征, 因此 Zhang 等^[8] 提出标记特定特征的多标记学习算法 (Multi-label Learning with Label Specific Features, LIFT), LIFT 属于高阶策略, 分别在正例和负例样本中通过聚类分析各标记构建特定标记特征, 但这种需特征空间参与的算法会因冗余特征及特征维度增加而影响分类精度及效率.

为了解决上述算法耗时较大的问题, 部分学者将极限学习机 (Extreme Learning Machine, ELM) 理论成功应用到多标记学习中. Huang 等^[9] 提出的 ELM 是一种快速单隐藏层前馈神经网络算法, 在网络参数确定过程中无需任何迭代. 相比神经网络算法, ELM 具有训练速度更快、泛化能力更强等优点. Sun 等^[10]、Meng 等^[11] 和 Venkatesan 等^[12] 将 ELM 理论应用到多标记分类中, 耗时较低, 分类效果理想. Zhang 等^[13] 提出多层径向基函数极限学习机 (Multi-layer Radial Basis Function Extreme Learning Machine, ML-ELM-RBF), 改用多隐藏层 ELM 实现多标记分类, 提高分类效果. 上述 ELM 需要设置隐藏层数目, 并且需要随机初始化权值与偏置, 导致算法结果不稳定. 对此 Luo 等^[14] 使用核 ELM 处理多标记问题, 通过核函数代替隐藏层节点映射, 保证网络的稳定性^[15].

基于上述研究成果, 为了降低学习耗时, 本文提出采用回归核极限学习机处理多标记学习问题 (Multi-label Learning of Regression Kernel Extreme Learning Machine, ML-RKELM). 相比其它分类模式的 ELM 处理多标记问题, ML-RKELM 使用回归模式的核 ELM, 具有更稳定高效的预测结果. 同时为了考

虑标记间相关性, 采用成熟的关联规则分析算法对标记空间进行规则提取. 在设定阈值范围内, 频繁模式生长 (Frequent Patterns Growth, FP-Growth) 算法未得到有效标记规则, 认为标记间为弱相关或不相关, 此时直接使用 ML-RKELM 进行多标记分析. 若得到若干有效标记规则, 认为这些标记间存在强相关性, 实现基于结合关联规则分析的 ML-RKELM (ML-RKELM with Association Rules, ML-ASRKELM). ML-ASRKELM 为高阶策略的多标记算法, 保留标记间关联信息, 同时具备 ELM 耗时较短、求解性能较高和泛化能力较强等特点. 本文将 ML-RKELM 结合关联规则, 提供相关性分析的途径. 为了验证算法, 在多个公开的多标记数据集上进行对比实验, 实验表明, 本文提出的两种算法在多标记分类中都具备高效性和稳定性.

1 基本理论

1.1 基于 FP-Growth 关联规则

关联分析是一种寻找数据集中变量之间隐含有趣关系的方法, 其中频繁项集 (Frequent Itemsets, FI) 和关联规则 (Association Rules, AR) 是 2 种常用分析方式^[16]. FP-Growth 关联规则首先构建 FP-Tree, 然后通过 FP-Tree 挖掘频繁项集, 最后构建强关联规则. 其中 FP 表示为频繁模式 (Frequent Patterns). 为了更深入了解 FP-Growth, 对关联规则进行如下相关定义.

设

$$I = \{i_1, i_2, \dots, i_n\}$$

为 n 个项目总集合,

$$T = \{t_1, t_2, \dots, t_m\}$$

为 m 个事务总集合. 在集合 T 中每个事务 t_i 都有独一无二事务 ID (简称 TID), 且包含集合 I 的一个非空子集. P 表示概率.

定义 1 关联规则 令集合 $A \subseteq I, B \subseteq I, A \cap B = \emptyset$, 规则 R 定义为 $R: A \rightarrow B$ 的蕴含式, 其中 A 称为规则的先导 (Antecedent), B 称为规则的后继 (Consequent).

定义 2 支持度 项目集合 A 和 B 在事务总集合 T 中出现的概率表示为支持度 (Support), 形式化定义为

$$Sup(A \rightarrow B) = P(A \cup B).$$

定义 3 置信度 在事务总集合 T 中包含项目集合 A 的条件下, 包含项目集合 B 的概率表示为置信度 (Confidence). 置信度可使用条件概率形式化定

义为

$$\text{Conf}(A \rightarrow B) = \frac{P(A \cup B)}{P(A)}.$$

在关联分析时需事先设定最小支持度(minsup 阈值)和最小置信度(minconf 阈值),关联规则同时满足这两个条件称为强关联规则.项目集合满足最小支持度称为频繁项集.在实际过程中,同时满足最小支持度和最小置信度的规则可能过多,置信度的度量未考虑规则后继中出现的项目集合的支持度,这可能会导致高置信度的规则并不是所需的规则.这时就需要加入额外的度量方式进一步过滤多余的规则,其中最通用的是提升度(lift)^[17]度量,通过提升度使挖掘的规则更符合所需条件.

定义4 提升度 提升度表示为整体的支持度与预期的支持度的偏差.当规则的 $\text{lift} < 1$ 时,表示此规则相互排斥;当 $\text{lift} = 1$ 时,表示规则无任何关联;当 $\text{lift} > 1$ 时,这种规则才有意义.其形式化定义为

$$\text{lift}(A \rightarrow B) = \frac{P(A \cup B)}{P(A)P(B)}.$$

1.2 极限学习机

神经网络算法需要较多的网络参数设置,在求解最优解时可能出现局部最优解,无法得到全局最优解.极限学习机理论是一种高效求解单隐藏层前馈神经网络的算法,求解时只需设置隐藏层节点数,并随机初始化权值和偏置就能高效求解全局最优解,具有运行速度较快、泛化性能较高等特点^[9].

在对 ELM 进行形式化定义前需进行如下规定.设有 N 个随机样本

$$\{(X_i, Y_i) \mid i = 1, 2, \dots, N\},$$

其中

$$X_i = [x_{i1} \ x_{i2} \ \dots \ x_{in}]^T,$$

$$Y_i = [y_{i1} \ y_{i2} \ \dots \ y_{im}]^T,$$

对于具有 L 个隐藏节点的单隐藏层神经网络形式化定义为

$$\sum_{i=1}^L \beta_i g_i(X_j) = \sum_{i=1}^L \beta_i g(w_i \cdot X_j + b_i) = o_j, \quad (1)$$

其中

$$\beta_i = [\beta_{i1} \ \beta_{i2} \ \dots \ \beta_{im}]^T$$

表示输出权值 $g(x)$ 表示激活函数,

$$w_i = [w_{i1} \ w_{i2} \ \dots \ w_{in}]^T$$

表示输入权值 b_i 表示为第 i 个隐藏神经元的偏置, \cdot 表示为点积.

单隐藏层神经网络的目标是为了输出的误差最小,即

$$\sum_{j=1}^N \|o_j - y_j\| = 0, \quad (2)$$

通过式(1)和式(2)可知,在确定 β_i 、 w_i 和 b_i 后可得

$$\sum_{i=1}^L \beta_i g(w_i \cdot X_j + b_i) = y_j, \quad j = 1, 2, \dots, N. \quad (3)$$

将式(3)使用矩阵表示为

$$H\beta = Y. \quad (4)$$

2 标记学习

单标记学习无法处理真实世界对象多语义性、概念复杂性,无法满足目前机器学习的高要求,由此建立多标记学习框架.该框架通过对任意一个对象进行特征向量的描述,根据特征向量尽可能将对象进行合适的类别标记和精准分类.

定义5 多标记学习 假定 X 为 n 维的示例空间 \mathbf{R}^n ,在给定数据集

$$\{(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)\},$$

其中 $x_i \in X$ 为一个示例, $Y_i \in Y$ 为一组标记集合

$$\{y_1^{(i)} \ y_2^{(i)} \ \dots \ y_l^{(i)}\},$$

且

$$y_l^{(i)} \in Y,$$

多标记学习就是为得到映射关系

$$f: X \rightarrow 2^Y.$$

在多标记学习评价中,单标记评价指标无法直接进行多标记指标评价,如 Accuracy、Precision 和 Recall.本文使用 5 种多标记通用评价指标进行算法性能评价,评价指标包括:海明损失(Hamming Loss, HL)、1-错误率(One-Error, OE)、覆盖率(Coverage, CV)、排序损失(Ranking Loss, RL)和平均精度(Average Precision, AP)^[18].

设多标记分类器为 $h(\cdot)$,预测函数为 $f(\cdot, \cdot)$,排序函数为 rank_f .多标记数据集

$$D = \{(x_i, Y_i) \mid 1 \leq i \leq n\}.$$

上述 5 种评价指标 HL、OE、CV、RL 和 AP 形式化定义如下.

海明损失评估对象标记被错误分类标记的次数情况、正确的标记被错误预测情况:

$$HL_D(h) = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{|Y|} |h(x_i) \Delta Y_i| \right),$$

其中 Δ 表示两个集合之间的对称差.当 $HL_D(h) = 0$ 时为最好的情况,即 $HL_D(h)$ 越小 $h(\cdot)$ 的性能越高.

1-错误率评估对象最高排位标记并未正确标记

的次数情况:

$$OE_D(f) = \frac{1}{n} \sum_{i=1}^n \left[\left[\arg \max_{y \in Y} f(x_i, y) \right] \notin Y_i \right].$$

当 $OE_D(f) = 0$ 时为最好的情况, 即 $OE_D(f)$ 越小, $f(\cdot, \cdot)$ 性能越高.

覆盖率评估对象标记序列中所需标记数达到覆盖全部标记:

$$CV_D(f) = \frac{1}{n} \sum_{i=1}^n \max_{y \in Y_i} \text{rank}_f(x_i, y) - 1.$$

$CV_D(f)$ 越小, $f(\cdot, \cdot)$ 的性能越高.

排序损失评估对象非属标记的排位高于所属标记的次数情况:

$$RL_D(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|Y_i| \left| \bar{Y}_i \right|} \cdot$$

$\left| \left\{ (y_1, y_2) \mid f(x_i, y_1) \leq f(x_i, y_2), (y_1, y_2) \in Y_i \times \bar{Y}_i \right\} \right|$.
当 $RL_D(f) = 0$ 时为最好情况, 即 $RL_D(f)$ 越小, $f(\cdot, \cdot)$ 的性能越高.

平均精度评估在特定标记 $y \in Y_i$ 排列的正确标记的平均分数:

$$AP_D(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|Y_i|} \cdot \sum_{y \in Y_i} \frac{|\{y' \mid \text{rank}_f(x_i, y') \leq \text{rank}_f(x_i, y), y' \in Y_i\}|}{\text{rank}_f(x_i, y)}.$$

当 $AP_D(f) = 1$ 时为最好, 即 $AP_D(f)$ 越大, $f(\cdot, \cdot)$ 的性能越高.

3 基于关联规则极限学习机的多标记学习

3.1 标记间关联规则分析

在多标记学习问题中, 每个实例样本被多个标签标记, 而部分标记之间存在一定的关联性, 利用标记之间的关联性获得额外的分类信息, 更有利于多标记的分类. 因此, 研究标记之间的相关性也是多标记学习的重要内容.

关联规则对于寻找标记空间中各标记内在关联性是一种可行方法^[19], 它属于高阶策略. 为了更好地理解多标记数据集中标记空间关联分析过程, 假设标记空间为 m 维、 n 个示例, 表示为

$$Y = \{Y_1, Y_2, \dots, Y_n\} \in \mathbf{R}^m.$$

表 1 为标记空间模拟数据, 其中 9 个示例, 5 种类标记分别为 A, B, C, D 和 E . 将标记空间数据按关联事务表改写得到标记转换集 (Label Conversion Set, LCS)

表 1 标记空间虚拟数据集

标记	A	B	C	D	E	LCS
Y_1	1	1	0	0	1	{1 2 5}
Y_2	0	1	0	1	0	{2 4}
Y_3	0	1	1	0	0	{2 3}
Y_4	1	1	0	1	0	{1 2 4}
Y_5	1	0	1	0	0	{1 3}
Y_6	0	1	1	0	0	{2 3}
Y_7	1	0	1	0	0	{1 3}
Y_8	1	1	1	0	1	{1 2 3 5}
Y_9	1	1	1	0	0	{1 2 3}

设第 i 个示例标记集合

$$Y_i = \{y_1^{(i)}, y_2^{(i)}, \dots, y_l^{(i)}\} \in Y,$$

则标记转换集

$$LCS = \{Y_1^{\#}, Y_2^{\#}, \dots, Y_l^{\#}\},$$

其中 $Y_i^{\#}$ 表示标记集合 Y_i 中标记值为 1 的标记编号, 即

$$Y_i^{\#} = \{j \mid y_j^{(i)} = 1, j \in l\}.$$

标记关联规则分析算法采用 FP-Growth 算法, 首先构建 FP-Tree, 通过 FP-Growth 算法得出所需规则. 基于 FP-Growth 标记关联规则算法 (Label Association Rules, LAS) 参考文献 [20], 在产生规则后加入最小置信度 minconf 和提升度 lift 约束规则, 最终产生标记规则 LCS 列项模拟数据在 $\text{minsup} = 0.3$ 和 $\text{minconf} = 0.6$ 时, 标记空间中各标记对的最终规则如表 2 所示.

表 2 最终规则集

标记集	支持度	置信度
{1 3}	0.4444	0.6
{3 1}	0.4444	0.6

从表 2 可看出, 标记 1 与标记 3 存在强相关性, 标记关联规则算法步骤如下.

算法 1 基于 FP-Growth 的标记关联规则算法 (LAS)

输入 标记空间数据转换集 LCS ,

最小支持度 minsup , 最小置信度 minconf

输出 最终标记规则 R_l

$FLR = \emptyset$;

$FT = \text{FP-Tree}(LCS, \text{minsup})$;

$LR = \text{FP-Growth}(FT, \emptyset)$;

for each LR

if $\text{conf}(LR) > \text{minconf}$ and $\text{lift}(LR) \geq 1$ then

$$FLR = LR;$$

end for

for each $y_j \in Y$

if $y_j \in FLR$ then $R_l^{(j)} = FLR_{conf}$;

else $R_l^{(j)} = 1$;

end for

return R_l .

通过标记关联规则算法后产生的 R_l 即为标记规则向量.

3.2 多标记学习的极限学习机

在多标记学习中, 设有 N 个样本的训练数据集

$$D = \{x_i, Y_i\}_{i=1}^N,$$

其中 $x_i \in \mathbf{R}^n$ 为 n 维特征向量, 函数 $h(x_i)$ 将 x_i 从输入空间映射到 L 维的特征空间, $Y_i \in \mathbf{R}^m$ 为输出标记集合. 根据式 (1) 和式 (4), 多标记 ELM 的输出函数 $f_l(x)$ 为

$$f_l(x) = H\beta = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_N) \end{bmatrix}_{N \times L} \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_L^T \end{bmatrix}. \quad (5)$$

根据式 (4) 可得

$$\beta = H^+ Y,$$

$$\text{s.t. } H^+ = \begin{cases} (H^T H)^{-1} H^T, & H^T H \text{ 非奇异} \\ H^T (H H^T)^{-1}, & H H^T \text{ 非奇异} \end{cases} \quad (6)$$

H^+ 为 H 的广义逆矩阵, 根据岭回归理论 (Ridge Regression Theory), 为了提高算法的稳定性和泛化性, 添加正则项 C , 则式 (5) 最小化目标为

$$\min L_f = \|\beta\|^2 + C \sum_{i=1}^N \|\xi_i\|^2,$$

$$\text{s.t. } \xi_i = Y_i - f_l(x_i), \quad i = 1, 2, \dots, N.$$

根据 KKT 条件, 求出输出权值 β ,

$$\beta = H^T \left(\frac{I}{C} + H H^T \right)^{-1} Y. \quad (7)$$

此时多标记输出函数就可表示为

$$f_l(x) = H\beta = H H^T \left(\frac{I}{C} + H H^T \right)^{-1} Y. \quad (8)$$

ELM 需设置隐藏层节点数, 每次需随机设定权值和偏置, 导致计算结果并不稳定, 容易受到随机设定值的影响. 采用核 ELM 可以解决这一问题, 当映射函数 $h(x)$ 未知时, 即引入核矩阵 (本文采用 RBF 核):

$$\begin{aligned} \Omega_{ELM} &= H H^T: \Omega_{ELM(i,j)} = h(x_i, x_j), \\ K(x_i, x_j) &= \exp(-\gamma \|x_i - x_j\|^2). \end{aligned} \quad (9)$$

根据式 (9), 可得

$$H H^T = \begin{bmatrix} K(x, x_1) \\ K(x, x_2) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T.$$

这样式 (8) 就可表示为

$$\begin{aligned} f_l(x) &= h(x) H^T \left(\frac{I}{C} + H H^T \right)^{-1} Y = \\ &= \begin{bmatrix} K(x, x_1) \\ K(x, x_2) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{I}{C} + \Omega_{ELM} \right)^{-1} Y. \end{aligned} \quad (10)$$

本文在核 ELM 中并未使用传统分类模式进行多标记分类, 通过 ELM 回归模式训练得出输出权值 β . 由于 ELM 原本是处理多类标记问题 (属于单标记), 所以该算法分类模式需统计每个示例的标记, 得出总标记类别数. 但对于多标记问题, 标记的类别数已知. 若在多标记学习中采用分类模式的 ELM (或核 ELM) 处理低维标记空间耗时可能与回归模式相差不多, 但在处理高维标记空间耗时较高. 而回归模式 ELM 避免这一问题, 同时回归核 ELM 直接将特征空间映射后, 直接求解目标函数矩阵解析解, 算法耗时短于分类模式.

通过结合式 (2)、式 (4) 和式 (10), 已知训练数据标记集合 Y 和拟合输出权值 β , 最后将未标记特征值映射到 Ω_{ELM} 中, 根据训练得到的 β 可预测标记集合. 最小化目标函数为

$$f_E = \sum_{i=1}^N \|f_l(x_i) - Y_i\|^2.$$

基于上述推论, 结合关联规则得出标记规则, 再采用回归模式的核 ELM 拟合, 最终得出标记预测结果.

ML-RKELM 和 ML-ASRKELM 首先都需要算法 1 对标记空间进行关联规则分析, 得到规则向量集 (Ruleset), 然后需判断该集合是否为空: 若规则向量集不为空则为 ML-ASRKELM; 若规则向量集为空则为 ML-RKELM. 算法步骤如下.

算法 2 ML-ASRKELM

输入 多标记训练数据 $D = \{x_i, Y_i\}_{i=1}^N$,

训练集标记关联规则 R_l ,

测试集 $D^* = \{x_j^*\}_{j=1}^M$,

正则参数 C , RBF 核参数 σ

输出 预测标记 Y^*

for training dataset D

```

    Calculate kernel matrix  $\Omega_{ELM}$  // 式(9)
    Calculate output weights  $\beta$  // 式(7)、式(10)
end for
for testing dataset  $D^*$ 
    Calculate outputs  $f_i(\mathbf{x}) = \Omega_{ELM} \beta$  // 式(10)
end for
if input rule set  $R_l \neq \emptyset$ 
     $f_l(\mathbf{x}^*) = f_l(\mathbf{x}) \otimes R_l$  // ML-ASRKELM
else
     $f_l(\mathbf{x}^*) = f_l(\mathbf{x})$  // ML-RKELM
return  $\mathbf{Y}^* = \{y_j^* | f_l(\mathbf{x}_j^*) > 0, j = 1, 2, \dots, M\}$ 

```

算法中 \otimes 运算表示 R_l 向量第 k 个值与 $f_l(\mathbf{x}_j^*)$ 矩阵第 k 行相乘. 当 $R_l = \emptyset$ 时, 标记相关性规则不参与算法, 最终结果即为 ML-RKELM 结果. 当 $R_l \neq \emptyset$ 时, 关联规则得到的多标记规则加入 ML-RKELM 中, 即最终结果为 ML-ASRKELM 结果.

4 实验及结果分析

为了说明本文算法的有效性, 选取 Yeast Gene、Natural Scene、Delicious、Yahoo Web Pages(包含 11 个子数据集) 14 个数据集.

Yeast Gene^[5,13] 包含 2 407 个样本, 训练数据集有 1 500 个样本, 测试数据集有 917 个样本, 每个样本包含 103 个属性值, 所有的样本大致有 14 种类别属性, 每个样本对应的平均标记数为 4.24 ± 1.57 .

Natural Scene 数据集^[7,13] 由 2 000 幅图像组成, 人工手动标记图像, 5 类标记分别为: Desert、Mountains、Sea、Sunset 和 Trees. 标记从 1 个到 3 个不等, 平均每幅图像有 1.24 ± 0.44 个类标记, 特征向量维度为 294, 包括 1 000 个训练集和 1 000 个测试集.

Delicious 数据集^[13] 收集书签站点文本数据, 形成文本多标记数据, 包括 12 920 个训练集和 3 185 个测试集, 数据包含离散特征 500 维, 共有 983 个标记, 平均每个样本标记数为 19.02 个.

雅虎网页数据集^[7,13] 收集自雅虎网站, 包括 11 个板块(艺术、商业、计算机等), 各数据子集特征数从 438~1047 不等, 每个数据子集包含 2 000 个训练集和 3 000 个测试集, 具体信息见表 3.

实验代码均在 Matlab2016a 中运行, 硬件环境为 Intel® Core™ i7-4790K 4.0 GHz CPU 8 GB 内存; 操作系统为 Windows 7. 为了验证算法的可靠性、有效性和实用性, 选择多标记的 5 种常用评价指标: HL、OE、CV、RL 和 AP. 通过这些评价指标综合评价

各算法的性能, 得出各算法全面的性能评估结果. 实验中将评价指标分别简写为 HL↓、OE↓、CV↓、RL↓ 和 AP↑, 其中, ↑ 表示指标数值越高越好, ↓ 表示指标数值越低越好. 对比算法采用二阶策略算法(Rank-SVM)^[5]、高阶策略算法(LIFT)^[8]、基于核极限学习机(Kernel Extreme Learning Machine, KELM)的多标记算法(ML-KELM^[14]和 ML-RBF^[21])、ML-KNN^[7]. 通过算法对比可以充分说明本文提出的两种算法在标记相关性和 ELM 分类器中的效果.

表 3 雅虎网页数据集

Table 3 Yahoo Web Pages dataset

数据集	标记数	特征数	训练集 平均标记	测试集 平均标记
Arts	26	462	1.627 ± 0.883	1.642 ± 0.959
Business	30	438	1.590 ± 0.840	1.586 ± 0.849
Computers	33	681	1.487 ± 1.087	1.522 ± 1.047
Education	33	550	1.465 ± 0.764	1.458 ± 0.752
Entertainment	21	640	1.426 ± 0.931	1.417 ± 0.991
Health	32	612	1.667 ± 0.852	1.659 ± 0.901
Recreation	22	606	1.414 ± 0.814	1.429 ± 0.868
Reference	33	793	1.159 ± 0.430	1.177 ± 0.537
Science	40	743	1.489 ± 0.788	1.425 ± 0.758
Social	39	1047	1.274 ± 0.639	1.290 ± 0.617
Society	27	636	1.705 ± 1.206	1.684 ± 1.246

因实验数据样本较多, 考虑到算法验证的可行性和准确性, 为减少随机误差的产生, 在一个数据集中, 对每种测试算法进行 10 次实验, 根据 10 次实验得到的评价指标求出平均值(Mean)和标准差(Standard Deviation). 在每个评价指标数据下标注排位情况, 如 ML-ASRKELM₍₁₎ 表示在某个数据集上该算法最优秀, 同时使用黑体数字表示. 在雅虎网页数据集上给出 11 个子集的平均评价指标数据.

表 4 和表 5 分别为 Yeast Gene 和 Natural Scene 数据集的实验结果, 其中 ML-RKELM、ML-ASRKELM、ML-KNN、ML-KELM 分类器的稳定性 10 次实验结果相同, 标准差均为 0, 后同.

从表 4 可看出, 在 Yeast Gene 数据集上, ML-ASRKELM 在 HL、CV、RL、AP 指标上最优, 在 OE 指标上与 Rank-SVM 并列第二. ML-RKELM 在 OE 指标上性能最优, 其余指标排第二. 在 Natural Scene 数据集上, ML-ASRKELM 在 OE、CV、RL 和 AP 指标上最优, 在 HL 指标上排第二. ML-RKELM 在 HL 指标上性能最优, 其余指标排第二.

表 4 7 种算法在 Yeast Gene 数据集上的实验结果

Table 4 Experimental results of 7 multi-label algorithms on Yeast Gene dataset

算法	HL ↓	OE ↓	CV ↓	RL ↓	AP ↑
ML-RKELM	0.1875±0 ₍₂₎	0.2290±0₍₁₎	6.1112±0 ₍₂₎	0.1582±0 ₍₂₎	0.7712±0 ₍₂₎
ML-ASRKELM	0.1870±0₍₁₎	0.2323±0 _(2.5)	6.1036±0₍₁₎	0.1577±0₍₁₎	0.7726±0₍₁₎
Rank-SVM	0.1982±0 ₍₆₎	0.2323±0 _(2.5)	6.4613±0 ₍₆₎	0.1669±0 ₍₃₎	0.7638±0 ₍₃₎
LIFT	0.1947±0.0009 ₍₃₎	0.2400±0.0040 ₍₆₎	6.4092±0.0415 ₍₄₎	0.1676±0.0012 ₍₄₎	0.7618±0.0010 ₍₄₎
ML-RBF	0.1978±0.0013 ₍₄₎	0.2372±0.0069 ₍₅₎	6.4963±0.0284 ₍₇₎	0.1736±0.0023 ₍₆₎	0.7586±0.0023 ₍₅₎
ML-KNN	0.1980±0 ₍₅₎	0.2345±0 ₍₄₎	6.4144±0 ₍₅₎	0.1715±0 ₍₅₎	0.7585±0 ₍₆₎
ML-KELM	0.2084±0 ₍₇₎	0.2497±0 ₍₇₎	6.3631±0 ₍₃₎	0.1769±0 ₍₇₎	0.7463±0 ₍₇₎

表 5 7 种算法在 Natural Scene 数据集上的实验结果

Table 5 Experimental results of 7 multi-label algorithms on Natural Scene dataset

算法	HL ↓	OE ↓	CV ↓	RL ↓	AP ↑
ML-RKELM	0.1578±0₍₁₎	0.2855±0 ₍₂₎	0.8745±0 ₍₂₎	0.1506±0 ₍₂₎	0.8156±0 ₍₂₎
ML-ASRKELM	0.1586±0 ₍₂₎	0.2850±0₍₁₎	0.8729±0₍₁₎	0.1495±0₍₁₎	0.8188±0₍₁₎
Rank-SVM	0.1884±0 ₍₇₎	0.3540±0 ₍₇₎	1.0820±0 ₍₇₎	0.2017±0 ₍₇₎	0.7646±0 ₍₇₎
LIFT	0.1623±0.0019 ₍₄₎	0.2916±0.0042 ₍₃₎	0.8820±0.0084 ₍₃₎	0.1508±0.0019 ₍₃₎	0.8125±0.0022 ₍₃₎
ML-RBF	0.1614±0.0019 ₍₃₎	0.2916±0.0055 ₍₄₎	0.8865±0.0205 ₍₄₎	0.1553±0.0047 ₍₅₎	0.8112±0.0043 ₍₅₎
ML-KNN	0.1850±0 ₍₆₎	0.3450±0 ₍₆₎	1.0800±0 ₍₆₎	0.1958±0 ₍₆₎	0.7700±0 ₍₆₎
ML-KELM	0.1666±0 ₍₅₎	0.2940±0 ₍₅₎	0.8880±0 ₍₅₎	0.1539±0 ₍₄₎	0.8113±0 ₍₄₎

表 6 为 Delicious 数据集上的实验结果, 因该数据集样本数量较大, 标记数较多, Rank-SVM 和 LIFT

运行时间超过 24 h 且未得出结果, 故使用 N/A 表示算法在该数据集上无结果。

表 6 7 种算法在 Delicious 数据集上的实验结果

Table 6 Experimental results of 7 multi-label algorithms on Delicious dataset

算法	HL ↓	OE ↓	CV ↓	RL ↓	AP ↑
ML-RKELM	0.0214±0_(1.5)	0.1203±0 ₍₂₎	533.1460±0 ₍₃₎	0.0871±0_(1.5)	0.5198±0 ₍₂₎
ML-ASRKELM	0.0214±0_(1.5)	0.1196±0₍₁₎	532.8527±0 ₍₂₎	0.0871±0_(1.5)	0.5203±0₍₁₎
Rank-SVM	N/A	N/A	N/A	N/A	N/A
LIFT	N/A	N/A	N/A	N/A	N/A
ML-RBF	0.0285±0 ₍₄₎	0.1682±0.0022 ₍₅₎	763.2003±3.2410 ₍₇₎	0.1694±0.0011 ₍₇₎	0.4525±0.0008 ₍₅₎
ML-KNN	0.0227±0 ₍₃₎	0.1965±0 ₍₆₎	475.5636±0₍₁₎	0.1007±0 ₍₃₎	0.4423±0 ₍₆₎
ML-KELM	0.0316±0 ₍₅₎	0.1356±0 ₍₃₎	727.5252±0 ₍₆₎	0.1432±0 ₍₆₎	0.4676±0 ₍₃₎

由表 6 可得, 在 Delicious 数据集上, ML-ASRKELM 在 HL、OE、RL、AP 指标上最优, 在 HL 与 RL 指标上与 ML-RKELM 并列第一, 在 CV 指标上排第二。而基于二阶策略的 Rank-SVM 和基于高阶策略的 LIFT 耗时巨大, 在规定时间内未得出结果。

表 7~表 11 为各种算法在雅虎网页数据集上的实验结果。Yahoo Web Pages 数据集包含多个数据集, 每种算法在每个指标上都一一进行对比。对于 HL 指标, 在 Business、Computers、Education、Reference、Science、Social、Society 数据集上, ML-ASRKELM 性能最优, 在 Reference、Science 数据集上, 其与 ML-RKELM 并列第一, 在 Entertainment、Health 数据集

上, ML-ASRKELM 性能第二, 在 Recreation 数据集上, ML-RKELM 与 ML-ASRKELM 并列第二。由 HL 的平均值可看出, ML-ASRKELM 平均性能最优。

由表 8 可以看出, ML-ASRKELM 只在 Health 数据集上的指标均排第二, 在其它数据集上指标均为最优。在 Reference 数据集上, 其与 ML-RKELM 并列第一, 在 Social 数据集上, 其与 ML-KELM 并列第一。ML-ASRKELM 平均性能也最优。

在 CV 覆盖率指标上, ML-ASRKELM 在 Recreation、Reference 数据集上指标性能排第二, 平均指标性能排第三, 与平均指标性能最优的数值相差 29.96%。在 RL 指标上, ML-ASRKELM 在 Entertainment、

Health、Recreation 数据集上指标性能最优,在 Entertainment、Recreation 数据集上,其与 ML-RKELM 并列第一,在 Arts、Computers、Reference、Science、Society 数据集上,指标排第二,在 Computers 数据集上,

其与 ML-RKELM 并列第二,ML-ASRKELM 平均指标性能排第二,与平均指标性能最优的数值仅相差 0.7%.在 AP 指标上,ML-ASRKELM 在各个数据集上的性能指标均最优.

表 7 7 种算法在 Yahoo Web Pages 数据集上的 HL 结果

Table 7 Hamming loss results of 7 multi-label algorithms on Yahoo Web Pages dataset

数据集	ML-RKELM	ML-ASRKELM	Rank-SVM	LIFT	ML-RBF	ML-KNN	ML-KELM
Arts	0.0548±0 ₍₅₎	0.0546±0 ₍₃₎	0.0657±0 ₍₇₎	0.0547±0.0001 ₍₄₎	0.0542±0.0001 ₍₁₎	0.0593±0 ₍₆₎	0.0543±0 ₍₂₎
Business	0.0251±0 ₍₂₎	0.0250±0 ₍₁₎	0.0290±0 ₍₇₎	0.0262±0.0001 ₍₅₎	0.0255±0.0001 ₍₄₎	0.0267±0 ₍₆₎	0.0253±0 ₍₃₎
Computers	0.0347±0 _(4.5)	0.0345±0 ₍₁₎	0.0442±0 ₍₇₎	0.0345±0.0003 _(2.5)	0.0345±0.0003 _(2.5)	0.0381±0 ₍₆₎	0.0347±0 _(4.5)
Education	0.0372±0 ₍₂₎	0.0371±0 ₍₁₎	0.0522±0 ₍₇₎	0.0382±0.0002 ₍₅₎	0.0375±0.0002 ₍₃₎	0.0397±0 ₍₆₎	0.0376±0 ₍₄₎
Entertainment	0.0524±0 ₍₃₎	0.0523±0 ₍₂₎	0.0670±0 ₍₇₎	0.0531±0.0005 ₍₅₎	0.0515±0.0004 ₍₁₎	0.0570±0 ₍₆₎	0.0530±0 ₍₄₎
Health	0.0334±0 ₍₃₎	0.0330±0 ₍₂₎	0.0444±0 ₍₇₎	0.0329±0.0001 ₍₁₎	0.0355±0.0002 ₍₅₎	0.0367±0 ₍₆₎	0.0347±0 ₍₄₎
Recreation	0.0559±0 _(2.5)	0.0559±0 _(2.5)	0.0666±0 ₍₇₎	0.0569±0.0003 ₍₅₎	0.0547±0.0002 ₍₁₎	0.0594±0 ₍₆₎	0.0561±0 ₍₄₎
Reference	0.0251±0 _(1.5)	0.0251±0 _(1.5)	0.0355±0 ₍₇₎	0.0255±0.0002 ₍₄₎	0.0252±0.0001 ₍₃₎	0.0273±0 ₍₆₎	0.0257±0 ₍₅₎
Science	0.0305±0 _(1.5)	0.0305±0 _(1.5)	0.0412±0 ₍₇₎	0.0318±0.0002 ₍₅₎	0.0308±0.0001 ₍₃₎	0.0334±0 ₍₆₎	0.0311±0 ₍₄₎
Social	0.0205±0 _(2.5)	0.0202±0 ₍₁₎	0.0290±0 ₍₇₎	0.0207±0.0001 ₍₅₎	0.0206±0.0001 ₍₄₎	0.0219±0 ₍₆₎	0.0205±0 _(2.5)
Society	0.0505±0 ₍₂₎	0.0500±0 ₍₁₎	0.0607±0 ₍₇₎	0.0523±0.0003 ₍₅₎	0.0515±0.0001 ₍₃₎	0.0543±0 ₍₆₎	0.0517±0 ₍₄₎
平均值	0.0382±0.0130 ₍₂₎	0.0380±0.0130 ₍₁₎	0.0487±0.0146 ₍₇₎	0.0388±0.0132 ₍₅₎	0.0383±0.0126 ₍₃₎	0.0413±0.0140 ₍₆₎	0.0386±0.0130 ₍₄₎

表 8 7 种算法在 Yahoo Web Pages 数据集上的 OE 结果

Table 8 One-error results of 7 multi-label algorithms on Yahoo Web Pages dataset

数据集	ML-RKELM	ML-ASRKELM	Rank-SVM	LIFT	ML-RBF	ML-KNN	ML-KELM
Arts	0.4627±0 ₍₂₎	0.4613±0 ₍₁₎	0.5603±0 ₍₇₎	0.4931±0.0047 ₍₅₎	0.4759±0.0042 ₍₄₎	0.5503±0 ₍₆₎	0.4740±0 ₍₃₎
Business	0.1160±0 ₍₃₎	0.1143±0 ₍₁₎	0.1367±0 ₍₇₎	0.1226±0.0024 ₍₆₎	0.1146±0.0012 ₍₂₎	0.1210±0 ₍₅₎	0.1163±0 ₍₄₎
Computers	0.3523±0 ₍₂₎	0.3520±0 ₍₁₎	0.4813±0 ₍₇₎	0.3621±0.0060 ₍₅₎	0.3585±0.0043 ₍₄₎	0.4143±0 ₍₆₎	0.3577±0 ₍₃₎
Education	0.4533±0 ₍₂₎	0.4520±0 ₍₁₎	0.6323±0 ₍₇₎	0.4799±0.0055 ₍₅₎	0.4763±0.0034 ₍₄₎	0.5123±0 ₍₆₎	0.4663±0 ₍₃₎
Entertainment	0.3977±0 ₍₂₎	0.3967±0 ₍₁₎	0.4970±0 ₍₆₎	0.4061±0.0049 ₍₃₎	0.4067±0.0044 ₍₄₎	0.4987±0 ₍₇₎	0.4110±0 ₍₅₎
Health	0.2533±0 ₍₁₎	0.2540±0 ₍₂₎	0.4803±0 ₍₇₎	0.2591±0.0028 ₍₃₎	0.2602±0.0033 ₍₄₎	0.3040±0 ₍₆₎	0.2637±0 ₍₅₎
Recreation	0.4603±0 ₍₂₎	0.4600±0 ₍₁₎	0.5713±0 ₍₇₎	0.4802±0.0054 ₍₅₎	0.4712±0.0046 ₍₄₎	0.5617±0 ₍₆₎	0.4633±0 ₍₃₎
Reference	0.3583±0 _(1.5)	0.3583±0 _(1.5)	0.5147±0 ₍₇₎	0.3838±0.0057 ₍₅₎	0.3648±0.0040 ₍₃₎	0.4117±0 ₍₆₎	0.3750±0 ₍₄₎
Science	0.4823±0 ₍₂₎	0.4820±0 ₍₁₎	0.6523±0 ₍₇₎	0.5093±0.0039 ₍₅₎	0.5061±0.0036 ₍₄₎	0.5577±0 ₍₆₎	0.4943±0 ₍₃₎
Social	0.2817±0 ₍₃₎	0.2807±0 _(1.5)	0.4337±0 ₍₇₎	0.2941±0.0033 ₍₅₎	0.2819±0.0036 ₍₄₎	0.3210±0 ₍₆₎	0.2807±0 _(1.5)
Society	0.3897±0 ₍₂₎	0.3887±0 ₍₁₎	0.4770±0 ₍₇₎	0.4000±0.0045 ₍₄₎	0.4045±0.0015 ₍₅₎	0.4280±0 ₍₆₎	0.3970±0 ₍₃₎
平均值	0.3643±0.1111 ₍₂₎	0.3636±0.1112 ₍₁₎	0.4943±0.1367 ₍₇₎	0.3809±0.1180 ₍₅₎	0.3746±0.1178 ₍₄₎	0.4255±0.1354 ₍₆₎	0.3727±0.1142 ₍₃₎

表 12 给出各算法在基准数据集上的耗时对比,由表可知,本文算法优势明显,在 Yeast Gene、Natural Scene 数据集上训练时间与测试时间总和都在 0.2 s 以内,在样本数量及标记数都较大的 Delicious 数据集上,本文两种算法训练时间与测试时间在 2 s 左右,而 Rank-SVM 和 LIFT 实验时间在 24 h 内未得出,故用 N/A 表示,ML-RKELM 和 ML-ASRKELM 在 Yahoo Web Pages 的每个子数据集上训练时间和测试时间总

和都在 0.6 s 以内.由此可以看出,ML-RKELM 和 ML-ASRKELM 平均耗时比 ML-KELM 少 1/3 左右,是 ML-RBF 和 ML-KNN 的 1/12.

实验也进一步证明回归模式核极限学习机在耗时上的优势.由于本文算法使用空间换取时间的策略,对于样本数量一定的数据集,耗时优于传统算法,对于处理大数据集时系统占用内存可能大于传统算法,但在合理范围之内.

表 9 7 种算法在 Yahoo Web Pages 数据集上的 CV 结果

Table 9 Coverage results of 7 multi-label algorithms on Yahoo Web Pages dataset

数据集	ML-RKELM	ML-ASRKELM	Rank-SVM	LIFT	ML-RBF	ML-KNN	ML-KELM
Arts	5.0163±0 ₍₅₎	4.9757±0 ₍₄₎	4.7023±0 ₍₂₎	4.6847±0.0455 ₍₁₎	5.7788±0.0383 ₍₇₎	4.8050±0 ₍₃₎	5.3377±0 ₍₆₎
Business	2.3963±0 ₍₄₎	2.3067±0 ₍₃₎	2.4157±0 ₍₅₎	2.1099±0.0253 ₍₁₎	2.7733±0.0374 ₍₇₎	2.1403±0 ₍₂₎	2.5117±0 ₍₆₎
Computers	4.0440±0 ₍₄₎	4.0353±0 ₍₃₎	4.3030±0 ₍₅₎	3.8271±0.0555 ₍₁₎	5.0634±0.0838 ₍₇₎	4.0233±0 ₍₂₎	4.7067±0 ₍₆₎
Education	4.0860±0 ₍₅₎	4.0067±0 ₍₄₎	3.6107±0 ₍₃₎	3.4911±0.0254 ₍₂₎	5.3682±0.0787 ₍₇₎	3.4037±0 ₍₁₎	4.3647±0 ₍₆₎
Entertainment	2.9010±0 ₍₂₎	2.9013±0 ₍₃₎	2.7920±0 ₍₁₎	2.9415±0.0266 ₍₄₎	3.8990±0.0438 ₍₇₎	3.0103±0 ₍₆₎	2.9770±0 ₍₅₎
Health	3.2877±0 ₍₄₎	3.0070±0 ₍₃₎	3.3110±0 ₍₅₎	2.9338±0.0110 ₍₂₎	3.7748±0.0410 ₍₇₎	2.7877±0 ₍₁₎	3.3787±0 ₍₆₎
Recreation	4.0963±0 ₍₃₎	4.0933±0 ₍₂₎	3.9270±0 ₍₁₎	4.1996±0.0345 ₍₄₎	4.5297±0.0437 ₍₇₎	4.2997±0 ₍₆₎	4.2947±0 ₍₅₎
Reference	3.4190±0 ₍₃₎	3.3270±0 ₍₂₎	3.5710±0 ₍₆₎	3.5089±0.0269 ₍₄₎	3.9425±0.0369 ₍₇₎	2.7697±0 ₍₁₎	3.5163±0 ₍₅₎
Science	6.3967±0 ₍₄₎	6.3820±0 ₍₃₎	6.4407±0 ₍₅₎	5.5721±0.0563 ₍₁₎	7.2828±0.0585 ₍₇₎	5.7880±0 ₍₂₎	6.5573±0 ₍₆₎
Social	3.7170±0 ₍₄₎	3.6833±0 ₍₃₎	3.7360±0 ₍₅₎	2.9659±0.0608 ₍₂₎	4.3701±0.0746 ₍₇₎	2.9280±0 ₍₁₎	3.7523±0 ₍₆₎
Society	5.8927±0 ₍₄₎	5.7827±0 ₍₃₎	5.9367±0 ₍₅₎	5.3436±0.0728 ₍₂₎	6.3341±0.0424 ₍₇₎	5.2487±0 ₍₁₎	6.0293±0 ₍₆₎
平均值	4.1139± 1.2223 ₍₅₎	4.0455± 1.2412 ₍₃₎	4.0678± 1.2287 ₍₄₎	3.7798± 1.0808 ₍₂₎	4.7825± 1.3418 ₍₇₎	3.7459± 1.1709 ₍₁₎	4.3115± 1.2666 ₍₆₎

表 10 7 种算法在 Yahoo Web Pages 数据集上的 RL 结果

Table 10 Ranking loss results of 7 multi-label algorithms on Yahoo Web Pages dataset

数据集	ML-RKELM	ML-ASRKELM	Rank-SVM	LIFT	ML-RBF	ML-KNN	ML-KELM
Arts	0.1256±0 ₍₃₎	0.1250±0 ₍₂₎	0.1240±0 ₍₁₎	0.1267±0.0013 ₍₄₎	0.1498±0.0012 ₍₇₎	0.1279±0 ₍₅₎	0.1358±0 ₍₆₎
Business	0.0395±0 ₍₅₎	0.0376±0 ₍₃₎	0.0390±0 ₍₄₎	0.0348±0.0005 ₍₁₎	0.0463±0.0006 ₍₇₎	0.0359±0 ₍₂₎	0.0415±0 ₍₆₎
Computers	0.0816±0 _(2.5)	0.0816±0 _(2.5)	0.0863±0 ₍₅₎	0.0818±0.0013 ₍₄₎	0.1058±0.0020 ₍₇₎	0.0812±0 ₍₁₎	0.0967±0 ₍₆₎
Education	0.0849±0 ₍₅₎	0.0832±0 ₍₄₎	0.0828±0 ₍₃₎	0.0770±0.0006 ₍₂₎	0.1127±0.0016 ₍₇₎	0.0770±0 ₍₁₎	0.0910±0 ₍₆₎
Entertainment	0.0977±0 _(1.5)	0.0977±0 _(1.5)	0.0979±0 ₍₃₎	0.1010±0.0010 ₍₅₎	0.1178±0.0019 ₍₇₎	0.1081±0 ₍₆₎	0.1007±0 ₍₄₎
Health	0.0497±0 ₍₄₎	0.0459±0 ₍₁₎	0.0568±0 ₍₆₎	0.0481±0.0004 ₍₃₎	0.0587±0.0007 ₍₇₎	0.0473±0 ₍₂₎	0.0517±0 ₍₅₎
Recreation	0.1383±0 _(1.5)	0.1383±0 _(1.5)	0.1404±0 ₍₄₎	0.1401±0.0012 ₍₃₎	0.1549±0.0017 ₍₆₎	0.1550±0 ₍₇₎	0.1460±0 ₍₅₎
Reference	0.0797±0 ₍₃₎	0.0777±0 ₍₂₎	0.0803±0 ₍₅₎	0.0799±0.0007 ₍₄₎	0.0929±0.0010 ₍₇₎	0.0689±0 ₍₁₎	0.0822±0 ₍₆₎
Science	0.1173±0 ₍₃₎	0.1171±0 ₍₂₎	0.1177±0 ₍₄₎	0.1230±0.0015 ₍₆₎	0.1362±0.0013 ₍₇₎	0.1107±0 ₍₁₎	0.1208±0 ₍₅₎
Social	0.0651±0 ₍₅₎	0.0642±0 ₍₄₎	0.0619±0 ₍₃₎	0.0539±0.0013 ₍₂₎	0.0770±0.0013 ₍₇₎	0.0537±0 ₍₁₎	0.0656±0 ₍₆₎
Society	0.1370±0 ₍₃₎	0.1345±0 ₍₂₎	0.1379±0 ₍₄₎	0.1381±0.0018 ₍₅₎	0.1491±0.0011 ₍₇₎	0.1303±0 ₍₁₎	0.1410±0 ₍₆₎
平均值	0.0924± 0.0340 ₍₄₎	0.0912± 0.0346 ₍₂₎	0.0932± 0.0337 ₍₅₎	0.0913± 0.0372 ₍₃₎	0.1092± 0.0374 ₍₇₎	0.0905± 0.0385 ₍₁₎	0.0975± 0.0358 ₍₆₎

表 11 7 种算法在 Yahoo Web Pages 数据集上的 AP 结果

Table 11 Average precision results of 7 multi-label algorithms on Yahoo Web Pages dataset

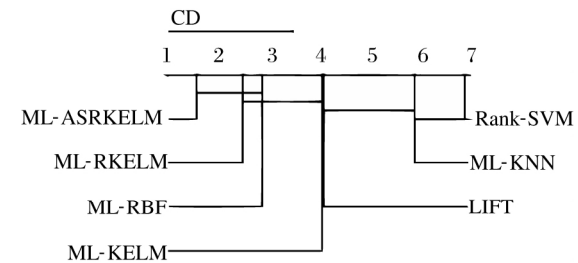
数据集	ML-RKELM	ML-ASRKELM	Rank-SVM	LIFT	ML-RBF	ML-KNN	ML-KELM
Arts	0.6252±0 ₍₂₎	0.6260±0 ₍₁₎	0.5705±0 ₍₆₎	0.6060±0.0028 ₍₅₎	0.6066±0.0018 ₍₄₎	0.5697±0 ₍₇₎	0.6143±0 ₍₃₎
Business	0.8866±0 ₍₂₎	0.8873±0 ₍₁₎	0.8718±0 ₍₇₎	0.8820±0.0014 ₍₅₎	0.8805±0.0012 ₍₆₎	0.8822±0 ₍₄₎	0.8829±0 ₍₃₎
Computers	0.7065±0 ₍₂₎	0.7068±0 ₍₁₎	0.6157±0 ₍₇₎	0.6979±0.0027 ₍₄₎	0.6960±0.0026 ₍₅₎	0.6620±0 ₍₆₎	0.7009±0 ₍₃₎
Education	0.6478±0 ₍₂₎	0.6485±0 ₍₁₎	0.5530±0 ₍₇₎	0.6329±0.0030 ₍₄₎	0.6211±0.0022 ₍₅₎	0.6082±0 ₍₆₎	0.6379±0 ₍₃₎
Entertainment	0.6912±0 ₍₂₎	0.6919±0 ₍₁₎	0.6421±0 ₍₆₎	0.6864±0.0030 ₍₃₎	0.6794±0.0030 ₍₅₎	0.6218±0 ₍₇₎	0.6847±0 ₍₄₎
Health	0.7913±0 ₍₂₎	0.7920±0 ₍₁₎	0.6745±0 ₍₇₎	0.7826±0.0017 ₍₃₎	0.7812±0.0012 ₍₅₎	0.7562±0 ₍₆₎	0.7824±0 ₍₄₎
Recreation	0.6351±0 ₍₂₎	0.6358±0 ₍₁₎	0.5706±0 ₍₆₎	0.6218±0.0027 ₍₄₎	0.6201±0.0026 ₍₅₎	0.5613±0 ₍₇₎	0.6287±0 ₍₃₎
Reference	0.7197±0 ₍₂₎	0.7202±0 ₍₁₎	0.6248±0 ₍₇₎	0.7014±0.0032 ₍₅₎	0.7096±0.0021 ₍₃₎	0.6820±0 ₍₆₎	0.7087±0 ₍₄₎
Science	0.6045±0 ₍₂₎	0.6052±0 ₍₁₎	0.5035±0 ₍₇₎	0.5900±0.0021 ₍₄₎	0.5830±0.0020 ₍₅₎	0.5489±0 ₍₆₎	0.5968±0 ₍₃₎
Social	0.7734±0 ₍₂₎	0.7739±0 ₍₁₎	0.6899±0 ₍₇₎	0.7686±0.0028 ₍₄₎	0.7644±0.0015 ₍₅₎	0.7555±0 ₍₆₎	0.7719±0 ₍₃₎
Society	0.6428±0 ₍₂₎	0.6435±0 ₍₁₎	0.5930±0 ₍₇₎	0.6342±0.0029 ₍₄₎	0.6274±0.0012 ₍₅₎	0.6181±0 ₍₆₎	0.6364±0 ₍₃₎
平均值	0.7022± 0.0860 ₍₂₎	0.7028± 0.0859 ₍₁₎	0.6281± 0.0973 ₍₇₎	0.6913± 0.0372 ₍₄₎	0.6881± 0.0909 ₍₅₎	0.6605± 0.1022 ₍₆₎	0.6951± 0.0874 ₍₃₎

表 12 7 种算法在基准数据集上的耗时结果

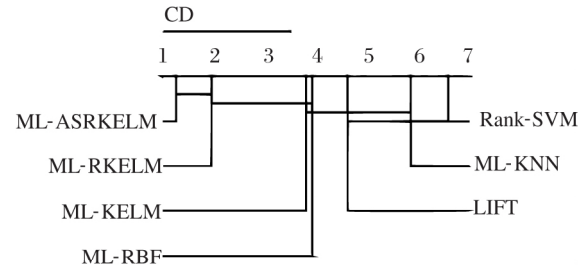
Table 12 Time-consuming results of 7 multi-label algorithms on benchmark dataset

数据集	ML-RKELM	ML-ASRKELM	Rank-SVM	LIFT	ML-RBF	ML-KNN	ML-KELM
Yeast Gene	0.1371	0.1984	46.5140	11.0571	0.1557	0.4963	0.1892
Natural Scene	0.0992	0.1191	2.5620	1.8226	0.5036	1.5642	0.1024
Delicious	1.8289	2.1356	N/A	N/A	95.5469	37.4670	2.3998
Arts	0.3368	0.3502	51.4880	33.2790	1.3086	4.1127	0.6289
Business	0.3229	0.4197	81.0965	33.6562	1.4196	4.1100	0.6821
Computers	0.3922	0.4234	87.1957	66.7561	1.8377	4.4668	0.7543
Education	0.3295	0.3458	82.3137	52.3218	1.6075	4.4722	0.7381
Entertainment	0.3504	0.4175	45.9506	39.9479	1.3568	2.5769	0.5830
Health	0.3592	0.3939	90.7261	57.8998	1.7286	4.0777	0.7452
Recreation	0.3564	0.3757	45.6848	38.6823	1.3211	4.1138	0.5906
Reference	0.4366	0.5400	74.9038	70.3330	2.3278	4.0730	0.7599
Science	0.4355	0.4494	283.5158	74.8062	2.4827	3.7092	0.8612
Social	0.4257	0.5127	104.4253	121.3315	5.3349	4.5668	0.8893
Society	0.3386	0.4004	74.3759	48.5098	1.6436	4.1469	0.6836
平均值	0.4392	0.5058	82.3656	50.0310	8.4697	5.9967	0.7577

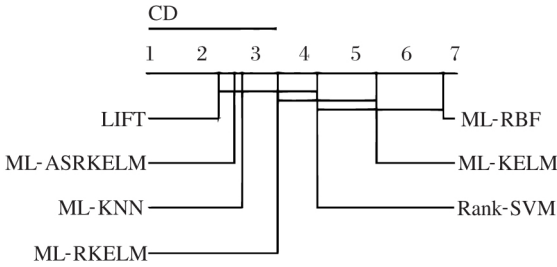
由于 Rank-SVM 和 LIFT 在 Delicious 数据集上未得出有效结果,因此本文测试各算法在除 Delicious 数据集之外的 13 个数据集上的综合性能,使用显著性水平为 5% 以下的 Nemenyi 检验^[8,22].当各对比算法的平均排序差值小于或等于临界差(Critical Difference, CD),认为这 2 种算法无显著性差异,否则具有显著性差异.图 1 给出在不同评价指标下各算法对比,CD 值为 2.497 9,无显著性差异的算法使用实线相连.在各评价指标子图中从左到右,算法性能依次降低.



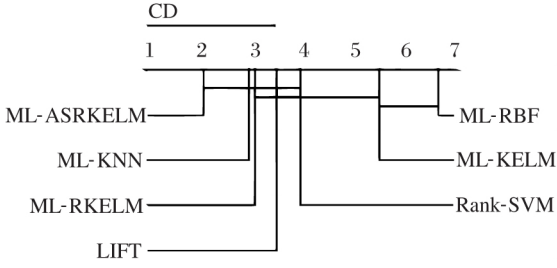
(a) HL



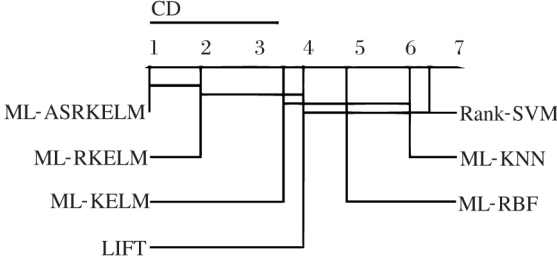
(b) OE



(c) CV



(d) RL



(e) AP

图 1 各算法综合性能对比

Fig.1 Performance comparison of algorithms

由图1可见,对于ML-ASRKELM在CV指标上性能排第二,在其余评价指标上的性能均最优,在40%情况下,与其它算法性能无显著性差异。即在HL指标上,ML-ASRKELM与ML-RKELM、ML-RBF性能相比无显著性差异,在OE指标上,与ML-RKELM无显著性差异,在CV指标上,与LIFT、ML-KNN、ML-RKELM、Rank-SVM无显著性差异,在RL指标上,与ML-KNN、ML-RKELM、LIFT、Rank-SVM无显著性差异,在AP指标上,ML-ASRKELM与ML-RKELM在统计上无显著性差异。而在60%的水平下,在统计上优于其它算法。

对于ML-RKELM在33.3%的情况下,在统计上优于其它算法,在66.7%的情况下与其它算法无显著性差异。

对于LIFT在统计上优于其它算法有13.3%的概率,而在73.3%情况下与其它算法无显著性差异,有13.3%的情况差于其它算法。

对于ML-KNN在统计上优于其它算法有13.3%的概率,在63.3%的情况下与其它算法无显著性差异,在23.3%的情况下差于其它算法。

上述分析可得,ML-ASRKELM性能最优,在60%的水平下,在统计上优于其它算法,在40%水平下,与其它算法性能相当。其次是ML-RKELM,在33.3%的水平下,在统计上优于其它算法,第三为LIFT,在13.3%的水平下优于其它算法。

上述实验分析更进一步表明ML-ASRKELM在各方面性能较优。

5 结束语

目前在多标记学习中考虑标记相关性的算法耗时都十分巨大,本文结合回归核ELM与关联规则,采用分部处理方式,使用关联规则分析标记间相关性,使用回归核ELM解决多标记分类问题,耗时和分类精度都较理想。

关联规则分析在一定条件下得出有效规则标记对,则认为该标记间存在强相关性,再通过本文的ML-ASRKELM得到最终预测标记。而对于未得到有效规则标记,则认为原标记间存在弱相关或不相关,对于这种情况直接采用ML-RKELM进行多标记数据学习。通过多组对比实验,充分表现ML-RKELM和ML-ASRKELM的有效性、高效性和稳定性。

但是,本文只将关联规则分析这一成熟方法结合到ELM中,未对ELM本身结构做出创新。现实世界中对象之间存在大量的弱相关性,本文算法进行

简化,并未对标记空间的弱相关性进行实质分析,导致部分实验中进行相关性分析的结果差于未进行分析的结果。因此,对于标记弱相关性的分析将是本文接下来的研究重点之一。

参 考 文 献

- [1] ZHOU Z H, ZHANG M L. Multi-label Learning // SAMMUT C, WEBB G I, eds. Encyclopedia of Machine Learning and Data Mining. Berlin, Germany: Springer, 2017: 875-881.
- [2] ZHU X F, LI X L, ZHANG S C. Block-Row Sparse Multiview Multilabel Learning for Image Classification. IEEE Transactions on Cybernetics, 2016, 46(2): 450-461.
- [3] ELGHAZEL H, AUSSEM A, GHARROUDI O, et al. Ensemble Multi-label Text Categorization Based on Rotation Forest and Latent Semantic Indexing. Expert Systems with Applications, 2016, 57: 1-11.
- [4] ZHANG M L, ZHOU Z H. Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(10): 1338-1351.
- [5] ELISSEEFF A, WESTON J. A Kernel Method for Multi-labelled Classification // Proc of the 14th International Conference on Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2001: 681-687.
- [6] TOMAR D, AGARWAL S. Multi-label Classifier for Emotion Recognition from Music // Proc of the 3rd International Conference on Advanced Computing, Networking and Informatics. Berlin, Germany: Springer, 2015: 111-123.
- [7] ZHANG M L, ZHOU Z H. ML-KNN: A Lazy Learning Approach to Multi-label Learning. Pattern Recognition, 2007, 40(7): 2038-2048.
- [8] ZHANG M L, WU L. LIFT: Multi-label Learning with Label-Specific Features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 107-120.
- [9] HUANG G B, ZHU Q Y, SIEW C K. Extreme Learning Machine: Theory and Applications. Neurocomputing, 2006, 70(1/2/3): 489-501.
- [10] SUN X, XU J T, JIANG C M, et al. Extreme Learning Machine for Multi-label Classification. Entropy, 2016, 18(6). DOI: 10.3390/e18060225.
- [11] MENG J E, VENKATESAN R, NING W. A High Speed Multi-label Classifier Based on Extreme Learning Machines [C/OL]. [2017-11-25]. <https://arxiv.org/ftp/arxiv/papers/1608/1608.08898.pdf>.
- [12] VENKATESAN R, MENG J E. Multi-label Classification Method Based on Extreme Learning Machines // Proc of the International Conference on Control Automation Robotics & Vision. Washington, USA: IEEE, 2016: 619-624.
- [13] ZHANG N, DING S F, ZHANG J. Multi Layer ELM-RBF for Multi-label Learning. Applied Soft Computing, 2016, 43: 535-

545.

- [14] LUO F F , GUO W Z , YU Y L , *et al.* A Multi-label Classification Algorithm Based on Kernel Extreme Learning Machine. *Neurocomputing* , 2017 , 260: 313-320.
- [15] HUANG G B , ZHOU H M , DING X J , *et al.* Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems , Man , and Cybernetics(Cybernetics)* , 2012 , 42(2) : 513-529.
- [16] 陈 耿 , 朱玉全 , 杨鹤标 , 等. 关联规则挖掘中若干关键技术的研究. *计算机研究与发展* , 2005 , 42(10) : 1785-1789.
(CHEN G , ZHU Y Q , YANG H B , *et al.* Study of Some Key Techniques in Mining Association Rule. *Journal of Computer Research and Development* , 2005 , 42(10) : 1785-1789.)
- [17] WITTEN I H , FRANK E , HALL M A , *et al.* *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco , USA: Morgan Kaufmann , 2016.
- [18] ZHANG M L , ZHOU Z H. A Review on Multi-label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering* 2014 , 26(8) : 1819-1837.
- [19] PARK S H , FÜRNKRANZ J. Multi-label Classification with Label Constraints [C/OL]. [2017 - 11 - 25]. <http://www.ecmlpkdd2008.org/sites/ecmlpkdd2008.org/files/pdf/workshops/pl/11.pdf>.
- [20] BECHINI A , MARCELLONI F , SEGATORI A. A MapReduce Solution for Associative Classification of Big Data. *Information Sciences* , 2016 , 332: 33-55.
- [21] ZHANG M L. ML-RBF: RBF Neural Networks for Multi-label Learning. *Neural Processing Letters* , 2009 , 29(2) : 61-74.
- [22] DEMŠAR J. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* , 2006 , 7: 1-30.

作者简介



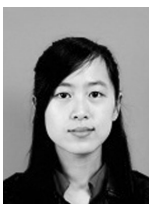
王一宾 硕士 ,教授 ,主要研究方向为多标记学习、机器学习、软件安全. E-mail: wangyb07@mail.ustc.edu.cn.

(**WANG Yibin** , master , professor. His research interests include multi-label learning , machine learning and software security.)



程玉胜(通讯作者) ,博士 ,教授 ,主要研究方向为大数据、粗糙集、特征选择的机器学习. E-mail: Chengyshaq@163.com.

(**CHENG Yusheng**(Corresponding author) , Ph.D. , professor. His research interests include big data , rough sets and machine learning for feature selection.)



何 月 硕士研究生 ,主要研究方向为多标记学习、机器学习、大数据. E-mail: heyuejy@163.com.

(**HE Yue** , master student. Her research interests include multi-label learning , machine learning and big data.)



裴根生 硕士研究生 ,主要研究方向为机器学习、大数据、数据统计. E-mail: peigensheng@163.com.

(**PEI Gensheng** , master student. His research interests include machine learning , big data and data statistics.)