

非平衡化标签补全核极限学习机多标签学习

程玉胜^{1,2}, 赵大卫¹, 王一宾^{1,2}, 裴根生¹

(1. 安庆师范大学计算机与信息学院, 安徽安庆 246011; 2. 安徽省高校智能感知与计算重点实验室, 安徽安庆 246011)

摘 要: 目前众多的研究者通常直接将标签置信度矩阵作为先验知识直接加入到分类模型中, 并没有考虑未标注先验知识对标签集质量的影响. 基于此, 引入非平衡参数的方法, 将先验知识获得的基础置信度矩阵进行非平衡化, 从而提出一种非平衡化的标签补全的核极限学习机多标签学习算法 (KELM-NeLC): 首先使用信息熵计算标签之间的相关关系得到标签置信度矩阵, 然后利用非平衡参数方法对基础的标签置信度矩阵进行改进, 构建出一个非平衡的标签补全矩阵, 最后为了学习获得更加准确的标签置信度矩阵, 将非平衡化的标签补全矩阵与核极限学习机进行联合学习, 依此解决多标签分类问题. 提出的算法在公开的多个基准多标签数据集上的实验结果表明, KELM-NeLC 算法较其他对比的多标签学习算法有一定优势, 使用统计假设检验进一步说明所提出算法的有效性.

关键词: 机器学习; 多标签学习; 标签相关性; 信息熵; 标签补全; 极限学习机

中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112 (2019)03-0719-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2019.03.029

Multi-label Learning of Kernel Extreme Learning Machine with Non-Equilibrium Label Completion

CHENG Yu-sheng^{1,2}, ZHAO Da-wei¹, WANG Yi-bin^{1,2}, PEI Gen-sheng¹

(1. School of Computer and Information, Anqing Normal University, Anqing, Anhui 246011, China;

2. The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing, Anhui 246011, China)

Abstract: At present, many researchers usually directly add the label confidence matrix as a priori knowledge to the classification model, and do not consider the influence of non-equilibrium prior knowledge on the quality of the label set. Based on this, the method of non-equilibrium parameters is introduced, and the basis confidence matrix obtained from the prior knowledge is non-equilibrium. Therefore, a multi-label learning algorithm is proposed, which uses kernel extreme learning machine with non-equilibrium label completion (KELM-NeLC). Firstly, information entropy is used to measure the correlation between labels which gets the basic label confidence matrix. Secondly, the basic label confidence matrix is improved to construct non-equilibrium label completion matrix by the non-equilibrium parameter. Finally, in order to learn to obtain a more accurate label confidence matrix, the non-equilibrium label completion matrix is introduced with the kernel extreme learning machine to solve the multi-label classification problem. The experimental results of the proposed algorithm in the opening benchmark multi-label datasets show that the KELM-NeLC algorithm has some advantages over other comparative multi-label learning algorithms and the statistical hypothesis test further illustrates the effectiveness of the proposed algorithm.

Key words: machine learning; multi-label learning; label correlations; information entropy; label completion; extreme learning machine

1 引言

多标签学习^[1]是处理具有丰富语义的真实世界对象的主要学习框架之一. 而大多数多标签学习方法认

为标签间是互相独立的, 忽略了标签间的相关性. 例如 Zhou^[2]等提出的多标签懒惰学习算法 (ML-KNN). 然而, 真实世界中的标签和标签间存在一定的关联性. 如何充分利用标签间的相关性, 构造强泛化性能的系统

是多标签分类学习算法的关键之一。

基于此,许多学者提出了相关算法. Elisseeff^[3]等提出一种核技术多标签算法 RankSVM,利用最大间隔准则策略来适应多标签学习. 张敏灵^[4]等提出基于标签类属性多标签学习算法 LIFT,为每个标签生成专属特征. 信息熵^[5]等相关信息论知识也被广泛应用到标签相关性研究当中. 基于该理论, Lee^[6]等提出了一种基于 CC(Classifier Chains)算法的多标签学习方法,利用条件熵设计出使标签间的相关性最大化的多标签学习方法. 但包括上述算法在内的大多数算法均忽视了未知标签集与已知标签集的相互影响。

实际上,在多标记学习数据集中,标签集中的已知标签数不应大于未知标签,否则对象的多标签就失去意义. 但是不可否认在未知标签中可能包含了大量有价值信息,正如关于对“异常”研究那样. 基于这种考虑,本文引入非平衡参数,提出非平衡化标签置信度矩阵计算方法. 首先利用标签之间的信息熵来衡量标签之间关系的强弱获得基础标签置信度矩阵,然后提出非平衡化标签置信度矩阵计算方法,获得数据集中更准确的标签置信度矩阵,最后利用标签置信度矩阵对初始的不完全标签矩阵进行增强补全。

同时,为了将多标签分类和标签相关性进行联合学习,选择合适的分类器模型,直接关系到算法的性能和效率. 由黄广斌^[7]等提出的极限学习机(Extreme Learning Machine, ELM) 分类器模型具有训练速度快,泛化能力强等优点,许多学者将其引入到分类,回归和聚类学习方法^[8-10]中,取得了不错的效果. 例如, Luo^[11]等使用核 ELM 算法处理多标签问题,通过核函数代替隐藏层映射保证网络的稳定性,但是该算法并没有考虑标签间的相关性。

因此本文选择核 ELM 作为分类器模型,将非平衡化标签补全矩阵和核 ELM 算法联合起来构建新的多标签分类模型,提出基于非平衡化标签补全的多标签核极限学习机算法(multi-label learning of Kernel Extreme Learning Machine with Non-equilibrium Label Completion, KELM-NeLC). KELM-NeLC 算法在多个公开的多标签数据集上的实验和统计假设检验,证实了考虑标签间相关性来提升算法的合理性。

2 非平衡化标签补全矩阵建模过程

2.1 多标签理论及其信息熵度量

定义 1 样本特征矩阵 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$, 其中 N 表示训练集中样本的数目, $\mathbf{x}_i \in \mathbb{R}^d$ 表示对应的第 i 个样本的特征向量, d 表示训练集中特征的维度. $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T$ 表示样本所对应的标签矩阵, $\mathbf{y}_i = \{1, -1\}^k$ 表示第 i 个样本的标签值,其中 k 表示标签的

数目. 则含有 N 个样本的多标签训练集可以表示为:

$$D = \{(\mathbf{x}_i, \mathbf{Y}_i) \mid 1 \leq i \leq N, \mathbf{x}_i \in \mathbb{R}^d, \mathbf{Y}_i = \{1, -1\}^k\} \quad (1)$$

定义 2 设集合 $A = \{a_1, \dots, a_m\}$, $p(a_i)$ 为元素 a_i 的先验概率,则集合 A 的信息熵为:

$$I(A) = - \sum_{i=1}^n p(a_i) \log_2 p(a_i) \quad (2)$$

定义 3 设集合 $A = \{a_1, \dots, a_m\}$, 集合 $B = \{b_1, \dots, b_n\}$, 则在给定集合 A 条件下集合 B 的条件熵为:

$$I(B|A) = - \sum_{i=1}^m \sum_{j=1}^n I(b_j|a_i) \quad (3)$$

其中, $I(b_j|a_i)$ 条件信息量,记为:

$$I(b_j|a_i) = -p(a_i b_j) \log_2 p(b_j|a_i) \quad (4)$$

2.2 非平衡化标签补全矩阵建模

真实世界中许多有价值的信息蕴含在未知标签集中,利用未知标签与已知标签间关系来分析标签间的关系能得到更多有价值的信息. 把“Missing”补全方法应用到多标签的学习,也仅仅考虑已知标签集合同的影响^[12]. 基于这种考虑,依据式(4)计算每个样本标签集中未知标签和已知标签元素间的条件信息量,得到基础标签置信度矩阵,即:

$$\mathbf{a}_{ij} = 1/I(\bar{l}_j|\bar{l}_i), \mathbf{b}_{ij} = 1/I(l_j|\bar{l}_i)$$

其中, l_i 和 \bar{l}_i 分别表示为 \mathbf{y}_i 值为“1”,和 \mathbf{y}_i 值为“-1”, $i = 1, \dots, k, j = 1, \dots, k$ 且 $i \neq j$. 不可否认在未知标签中可能包含了大量有价值信息,正如关于对“异常”研究那样. 受其启发,增加未知相关标签的权重,这有利于缺省标签的补全. 基于此,本文引入非平衡参数 α ,通过对基础标签置信度矩阵 \mathbf{a}_{ij} 权重的减少和 \mathbf{b}_{ij} 权重增大提出非平衡化的标签置信度矩阵计算方法:

$$\mathbf{M}_{ij} = -\alpha \times \mathbf{a}_{ij} + (1 - \alpha) \times \mathbf{b}_{ij} \quad (5)$$

受标签传播依赖思想启发^[13]定义如下的非平衡化

标签补全矩阵: $\hat{\mathbf{Y}} = \mathbf{Y} \times \mathbf{M} \quad (6)$

非平衡化标签补全矩阵算法,如算法 1 所示:

算法 1 非平衡化标签补全矩阵算法

输入: 训练样本标签矩阵 \mathbf{Y} , 非平衡参数 α ;

输出: 非平衡化标签补全矩阵 $\hat{\mathbf{Y}}$;

```

(1)  $\mathbf{Y} = \{\mathbf{Y}_i \mid i = 1, \dots, k\}$ 
(2) for each  $l_i, \bar{l}_j$ 
(3) While  $i \neq j$ 
(4)  $\mathbf{a}_{ij} = 1/I(\bar{l}_j|\bar{l}_i)$   $\mathbf{b}_{ij} = 1/I(l_j|\bar{l}_i)$  /* 利用式(4)分别计算得出  $\mathbf{a}_{ij}$  和  $\mathbf{b}_{ij}$  */
(5) if  $i = j$ 
(6)  $\mathbf{a}_{ij} = \mathbf{b}_{ij} = 1$ ; /* 将对角线元素置为 1 */
(7) end
(8) end
(9)  $\mathbf{M}_{ij} = -\alpha \times \mathbf{a}_{ij} + (1 - \alpha) \times \mathbf{b}_{ij}$ ; /* 利用式(5)求出置信度矩阵 */

```

(10) end

(11) $\hat{\mathbf{Y}} = \mathbf{Y} \times \mathbf{M} / *$ 利用式(6)求出非平衡化标签补全矩阵 $*$ /(12) return $\hat{\mathbf{Y}} / *$ 返回非平衡化标签补全矩阵 $*$ /

3 非平衡化标签补全的极限学习机算法建模

3.1 极限学习机理论

ELM 算法是一种快速的前馈单隐藏层神经网络学习算法,过程中无需对网络权重和偏置进行任何迭代调整,与传统的神经网络算法相比它的训练速度快,泛化能力强.

定义 4 设有 N 个随机样本 $\{(\mathbf{X}_i, \mathbf{Y}_i)\}$, 其中 $i = \{1, 2, \dots, N\}$, $\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{im}]^T$, $\mathbf{Y}_i = [\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{im}]^T$ 则对于具有 L 个隐藏节点的单隐藏层神经网络形式化定义为:

$$\sum_{i=1}^L \beta_i g_i(\mathbf{X}_j) = \sum_{i=1}^L \beta_i g(\omega_i \cdot \mathbf{X}_j + b_i) = o_j \quad (7)$$

在式(7)中, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ 表示输出权重, $g(x)$ 为激活函数, $\omega_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{im}]^T$ 为输入权重, b_i 表示为第 i 个隐藏神经元的偏置, “ \cdot ” 表示为点积.

定义 5 对于单隐藏层神经网络的目标是输出误差最小,表示为:

$$\sum_{j=1}^N \|o_j - y_j\| = 0 \quad (8)$$

定义 6 通过式(6)和(7)可知在确定 β_i, ω_i 和 b_i 时可得到:

$$\sum_{i=1}^L \beta_i g(\omega_i \cdot \mathbf{X}_j + b_i) = y_j, j = 1, 2, \dots, N \quad (9)$$

式(9)用矩阵表示为:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y} \quad (10)$$

3.2 标签补全的极限学习机多标签算法

设多标签训练数据集 $D = \{(\mathbf{x}_i, \mathbf{Y}_i) \mid 1 \leq i \leq N\}$, 利用函数 $h(\mathbf{x}_i)$ 将 \mathbf{x}_i 从输入空间映射到 L 维特征空间, \mathbf{Y}_i 为与 \mathbf{x}_i 对应的一组输出标签. 对于一个新的待分类对象 \mathbf{x} , 先通过算法 1 得到非平衡化标签补全矩阵 $\hat{\mathbf{Y}}$, 再自动学习获得权重调整非平衡化标签置信度矩阵, 得到标签置信度矩阵 \mathbf{M}^* , 最后结合核极限学习机理论模型, 将非平衡化的标签间置信度矩阵 \mathbf{M} 加入到该模型中来预测标签集合:

$$\begin{aligned} \mathbf{M}^* &= \mathbf{M} \times \boldsymbol{\beta}_1 \\ \text{s. t. } \hat{\mathbf{Y}} &= \mathbf{Y} \times \mathbf{M} \end{aligned} \quad (11)$$

由式(10)和(11)可以得到:

$$\begin{aligned} \mathbf{Y} \times \mathbf{M} \times \boldsymbol{\beta}_1 &= \mathbf{H}\boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_2 \boldsymbol{\beta}_1^{-1} &= \mathbf{H}^\dagger \hat{\mathbf{Y}} \end{aligned} \quad (12)$$

设 $\boldsymbol{\beta} = \boldsymbol{\beta}_2 \boldsymbol{\beta}_1^{-1}$, 则多标签 ELM 的输出函数 $f_i(\mathbf{x})$ 为:

$$f_i(\mathbf{x}) = \mathbf{H}\boldsymbol{\beta} = \begin{bmatrix} h(\mathbf{x}_1) \\ \vdots \\ h(\mathbf{x}_N) \end{bmatrix}_{N \times L} \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix} \quad (13)$$

由式(12)可以得到:

$$\begin{aligned} \boldsymbol{\beta} &= \mathbf{H}^\dagger \hat{\mathbf{Y}}; \\ \text{s. t. } \mathbf{H}^\dagger &= \mathbf{H}^T (\mathbf{H}\mathbf{H}^T)^{-1} \end{aligned} \quad (14)$$

其中 $\mathbf{H}\mathbf{H}^T$ 是非奇异矩阵, \mathbf{H}^\dagger 是 \mathbf{H} 的 Moore-Penrose 广义逆矩阵. 根据岭回归理论, 在 $\mathbf{H}\mathbf{H}^T$ 对角线上加上正则项 C , 可提高算法的稳定性和泛化能力. 则式(13)最小化目标为:

$$\begin{aligned} \min L_f &= \|\boldsymbol{\beta}\|^2 + C \sum_{i=1}^N \|\boldsymbol{\xi}_i\|^2; \\ \text{s. t. } \boldsymbol{\xi}_i &= \mathbf{Y}_i - f_i(\mathbf{x}_i), i = 1, 2, \dots, N \end{aligned} \quad (15)$$

根据 KKT 条件 (Karush-Kuhn-Tucker, KKT), 隐藏输出权重 $\boldsymbol{\beta}$ 表示为:

$$\boldsymbol{\beta} = \mathbf{H}^T \left(\frac{\mathbf{E}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \hat{\mathbf{Y}} \quad (16)$$

此时多标签输入函数表示为:

$$f_i(\mathbf{x}) = \mathbf{H}\boldsymbol{\beta} = \mathbf{H}\mathbf{H}^T \left(\frac{\mathbf{E}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \hat{\mathbf{Y}} \quad (17)$$

在传统的 ELM 算法中计算结果容易受到随机设定值的影响. 为此引入核矩阵解决这一问题:

$$\begin{aligned} \boldsymbol{\Omega}_{\text{ELM}} &= \mathbf{H}\mathbf{H}^T; \boldsymbol{\Omega}_{\text{ELM}(i,j)} = h(\mathbf{x}_i, \mathbf{x}_j); \\ \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) &= \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|) \end{aligned} \quad (18)$$

$$\text{根据式(18), 可以得出 } \mathbf{H}\mathbf{H}^T = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T, \text{ 则式}$$

(17)表示为:

$$\begin{aligned} f_i(\mathbf{x}) &= h(\mathbf{x}) \mathbf{H}^T \left(\frac{\mathbf{E}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \hat{\mathbf{Y}} \\ &= \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix} \left(\frac{\mathbf{E}}{C} + \boldsymbol{\Omega}_{\text{ELM}} \right)^{-1} \hat{\mathbf{Y}} \end{aligned} \quad (19)$$

最小化目标函数表示为:

$$f_{\text{SSE}} = \sum_{i=1}^N \|f_i(\mathbf{x}_i) - \mathbf{Y}_i\|^2 \quad (20)$$

非平衡化标签补全的多标签核极限学习机 (KELM-NeLC) 算法, 如算法 2 所示:

算法 2 基于非平衡化标签补全的多标签核极限学习机 (KELM-NeLC) 算法

输入: 多标签训练数据集 $D = \{(\mathbf{x}_i, \mathbf{Y}_i) \mid 1 \leq i \leq N\}$, 测试数据集 $D^* = \{(\mathbf{x}_j, \mathbf{Y}_j) \mid 1 \leq j \leq T\}$, 正则项 C , 核参数 σ , 非平衡参数 α ;

输出: 预测标签 \mathbf{Y}^* ;

(1) $\mathbf{Y} = \{\mathbf{Y}_i \mid i = 1, \dots, k\} / * D$ 中的训练标签矩阵 $*$ /

(2) for each $Y_i \in \mathbb{R}^k$
 (3) M / * 根据算法 1 得出标签置信度矩阵 M */
 (4) end
 (5) $\hat{Y} = Y \times M$; / * 根据式(6)得出标签补全矩阵 */
 (6) for each $x_i \in \mathbb{R}^d$
 (7) $\Omega_{ELM} = \exp(-\|x_i\|^2/\sigma)$; / * 通过式(18), 特征核映射 */
 (8) end
 (9) $\beta = (\frac{E}{C} + \Omega_{ELM(i)})^{-1} \hat{Y}$; / * 通过式(16)和(20), 求解出输出权重 */
 (10) for each $x_i \in \mathbb{R}^d$
 (11) for each $x_j^* \in \mathbb{R}^d$
 (12) $\Omega_{ELM(i,j)} = \exp(-\|x_i - x_j^*\|^2/\sigma)$;
 (13) end
 (14) end

(15) $f_i(x_j^*) = \Omega_{ELM(i,j)} \beta_2 \beta_1^{-1}$; / * 测试集预测标签结果 */
 (16) $Y^* = \{y_j^* | f_i(x_j^*) > 0, j=1, \dots, T\}$; / * 当预测值 $f_i(x_j^*) > 0$ 表示该标签存在 */
 (17) return Y^*

4 实验及其结果分析

4.1 实验数据集描述

本文选取 Natural Scene, Emotions, Yahoo Web Pages (包含 11 个子数据集) 共 13 个数据集. 其中 Natural Scene, Emotions 来自 <http://mulan.sourceforge.net/datasets-mlc.html>. Yahoo Web Pages 来自 <http://www.keel.ntt.co.jp/as/members/ueda/yahoo.tar>. 具体描述见表 1.

表 1 多标签数据集的详细描述

数据集	训练集	测试集	标签数	属性数	平均标签个数	标签密度	所属领域
Emotions	391	202	6	72	1.868	0.311	Music
Natural Scene	1000	1000	5	294	1.236	0.247	Images
Arts	2000	3000	26	462	1.636	0.063	Text
Business	2000	3000	30	438	1.588	0.053	Text
Computers	2000	3000	33	681	1.508	0.046	Text
Education	2000	3000	33	550	1.461	0.044	Text
Entertainment	2000	3000	21	640	1.420	0.068	Text
Health	2000	3000	32	612	1.663	0.052	Text
Recreation	2000	3000	22	606	1.423	0.065	Text
Reference	2000	3000	33	793	1.169	0.035	Text
Science	2000	3000	40	743	1.451	0.036	Text
Social	2000	3000	39	1047	1.283	0.033	Text
Society	2000	3000	26	462	1.692	0.063	Text

4.2 评价指标及算法相关参数设置

本文选取常用的 5 种评价准则, 即 Average Precision, CoVerage, Hamming Loss, One-Error 和 Ranking Loss^[14] 来综合评价多标签学习算法的性能. 分别简写为: AP \uparrow 、CV \downarrow 、HL \downarrow 、OE \downarrow 和 RL \downarrow . 其中 \uparrow 表示指标数值越高越好, \downarrow 表示指标数值越低越好. 将 KELM-NeLC 算法与 RELM (Regularized ELM)^[15], ML-KNN, LIFT, RankSVM 和 MLFE (Feature-induced labeling information Enrichment for Multi-label Learning, MLFE)^[16] 5 种常用算法做对比实验. 在 KELM-NeLC 算法中正则化系数 $C=1$, 核函数选择 RBF, 核参数设在 $[1, 5]$ 间, 非平衡化参数设在 $[0, 0.5]$ 间. 在 RELM 算法中正则化系数 $C=1$, 隐藏层神经元个数设为 100. 在 ML-KNN 算法中近邻个数设为 10 和平滑参数设为 1. 在 LIFT 算法中参数 $r=0.1$, 核函数选择 Linear. 在 RankSvm 中, 代价参

数设为 1, 核函数选择 RBF. 在 MLFE 算法中, 核函数选择 RBF, 核参数 β_1, β_2 和 β_3 选自 $\{1, 2, \dots, 10\}$, $\{1, 10, 15\}$ 和 $\{1, 10\}$ 分别在训练集上进行交叉验证. 为了提高算法的准确率, 检验算法稳定性把每个算法都运行 10 次, 并给出平均值 (mean) 和标准差 (standard deviation).

4.3 实验结果与分析

表 2 到表 4 给出了本文算法和其他 5 种算法在 13 个数据集上的实验结果. 其中表 2 给出每个算法在 Yahoo Web Pages 的 11 个网页数据集上各评价指标的平均实验结果和标准差, 最好的结果加粗表示. 在 Emotions 和 Natural Scene 数据集上最好的实验结果加粗显示, 如表 3 和表 4 所示. (注: 在表 3 和表 4 中 ± 0 表明该算法稳定, 运行 10 次结果没有变化)

表 2 雅虎网页数据集测试结果

算法	AP ↑	CV ↓	HL ↓	OE ↓	RL ↓
KELM-NeLC	0.7026 ± 0.0859	3.5388 ± 1.0891	0.0378 ± 0.0130	0.3656 ± 0.1127	0.0809 ± 0.0330
RELM	0.6904 ± 0.0889	4.4938 ± 1.3178	0.0387 ± 0.0130	0.3768 ± 0.1150	0.1023 ± 0.0366
ML-KNN	0.6248 ± 0.1166	4.0975 ± 1.2372	0.0432 ± 0.0144	0.4714 ± 0.1574	0.1024 ± 0.0453
LIFT	0.6917 ± 0.0910	3.6260 ± 1.1644	0.0385 ± 0.0128	0.3809 ± 0.1192	0.0841 ± 0.0340
RankSVM	0.6258 ± 0.0998	3.8637 ± 1.1340	0.0486 ± 0.0145	0.4955 ± 0.1369	0.0917 ± 0.0324
MLFE	0.6780 ± 0.0931	4.7492 ± 1.3343	0.0398 ± 0.0137	0.3871 ± 0.1206	0.1105 ± 0.0375

表 3 情感数据集测试结果

算法	AP ↑	CV ↓	HL ↓	OE ↓	RL ↓
KELM-NeLC	0.8008 ± 0	1.8465 ± 0	0.2079 ± 0	0.2772 ± 0	0.1599 ± 0
RELM	0.7640 ± 0.0162	2.0965 ± 0.0733	0.2381 ± 0.0069	0.3287 ± 0.0308	0.2020 ± 0.0143
ML-KNN	0.7808 ± 0	1.9158 ± 0	0.2137 ± 0	0.3317 ± 0	0.1729 ± 0
LIFT	0.7430 ± 0.0075	2.1881 ± 0.0329	0.2404 ± 0.0079	0.3639 ± 0.0170	0.2292 ± 0.0082
RankSVM	0.7503 ± 0	2.2426 ± 0	0.2946 ± 0	0.3168 ± 0	0.2244 ± 0
MLFE	0.7822 ± 0	1.9703 ± 0	0.2459 ± 0	0.3069 ± 0	0.1803 ± 0

表 4 自然场景数据集测试结果

算法	AP ↑	CV ↓	HL ↓	OE ↓	RL ↓
KELM-NeLC	0.8285 ± 0	0.8160 ± 0	0.1724 ± 0	0.2710 ± 0	0.1366 ± 0
RELM	0.7531 ± 0.0065	1.0892 ± 0.0271	0.1991 ± 0.0046	0.3852 ± 0.0107	0.2057 ± 0.0062
ML-KNN	0.7615 ± 0	1.0680 ± 0	0.1836 ± 0	0.3670 ± 0	0.1982 ± 0
LIFT	0.8071 ± 0.0037	0.8927 ± 0.0143	0.1654 ± 0.0017	0.3002 ± 0.0059	0.1551 ± 0.0031
RankSVM	0.7689 ± 0	1.0550 ± 0	0.1854 ± 0	0.3500 ± 0	0.1965 ± 0
MLFE	0.8166 ± 0	0.8440 ± 0.0143	0.1624 ± 0	0.2910 ± 0	0.1452 ± 0

实验结果说明:在雅虎 11 个网页数据集上平均结果如表 2 所示,在各个评价指标下的总体实验结果均值均于其他 5 种对比算法.从表 3 中可以看出,在 Emotions 数据集上 KELM-NeLC 算法 5 种评价指标明显优于其他算法;在表 4 中 Natura scene 数据集上,KELM-NeLC 算法仅在海明损失指标上性能排第三,其他指标均为最优.

4.4 非平衡化标签补全与其它分类算法结合

为了验证非平衡化标签补全算法的合理性,本文选取了 RELM 算法和改编的 k 近邻(KNN)算法与算法 1 结合,分别得到 RELM-NeLC 算法和 KNN-NeLC 算法.数据集选择 Business 和 Emotions。RELM-NeLC 算法和 RELM 算法中的正则化系数 $C = 1$,隐藏层神经元个数设为 100. KNN-NeLC 算法和 KNN 算法中近邻数设为 10. 实验结果如表 5 和表 6 所示,数据集上最好的实验结果加粗显示:

表 5 Business 数据集测试结果

算法	AP ↑	CV ↓	HL ↓	OE ↓	RL ↓
RELM-NeLC	0.8687 ± 0.0023	2.4053 ± 0.0362	0.0275 ± 0.0002	0.1277 ± 0.0031	0.0409 ± 0.0009
RELM	0.8674 ± 0.0013	2.8599 ± 0.0798	0.0278 ± 0.0002	0.1316 ± 0.0015	0.0505 ± 0.0019
KNN-NeLC	0.8719 ± 0	2.3957 ± 0	0.0300 ± 0	0.1233 ± 0	0.0408 ± 0
KNN	0.8627 ± 0	3.9267 ± 0	0.0297 ± 0	0.1187 ± 0	0.1005 ± 0

表 6 Emotions 数据集测试结果

算法	AP \uparrow	CV \downarrow	HL \downarrow	OE \downarrow	RL \downarrow
RELM-NeLC	0.7672 ± 0.0136	2.0129 ± 0.0917	0.2407 ± 0.0137	0.3351 ± 0.0175	0.1926 ± 0.0156
RELM	0.7614 ± 0.0138	2.1267 ± 0.1127	0.2422 ± 0.0134	0.3381 ± 0.0270	0.2071 ± 0.0193
KNN-NeLC	0.7984 ± 0	1.8812 ± 0	0.2013 ± 0	0.2871 ± 0	0.1695 ± 0
KNN	0.7913 ± 0	1.9010 ± 0	0.2005 ± 0	0.2921 ± 0	0.2019 ± 0

从表 5 和表 6 中可以看出,RELM-NeLC 和 KNN-NeLC 算法在 5 种多标签评价指标上与 RELM 和 KNN 算法相比均占优. RELM 和 KNN 算法与非平衡化标签补全结合后整体性能有所提升但不能互相促进学习,导致分类性能基本相当.

5 统计假设检验

本文采用显著性水平 5% 的 Nemenyi 检验^[17,18],当两个对比算法在所有数据集上的平均排序的差值小于等于临界差值(critical difference,简称 CD),认为这两

个算法没有显著性差异. 图 1 给出了在不同评价指标下的各算法间的比较,其中最上行为临界值 $CD = 2.091$,对于没有显著性差异的算法用彩色实线相连,各评价指标子图中从左到右,算法性能依次降低.

从图 1 可以发现:对于 KELM-NeLC 算法,在 68% 的情况下,在统计上优于其它算法,32% 情况下与其他算法没有显著差异. 其次是 LIFT 算法,在 36% 的情况下,在统计上优于其它算法. 第三是 RELM 算法,在 16% 的情况下优于其它算法.

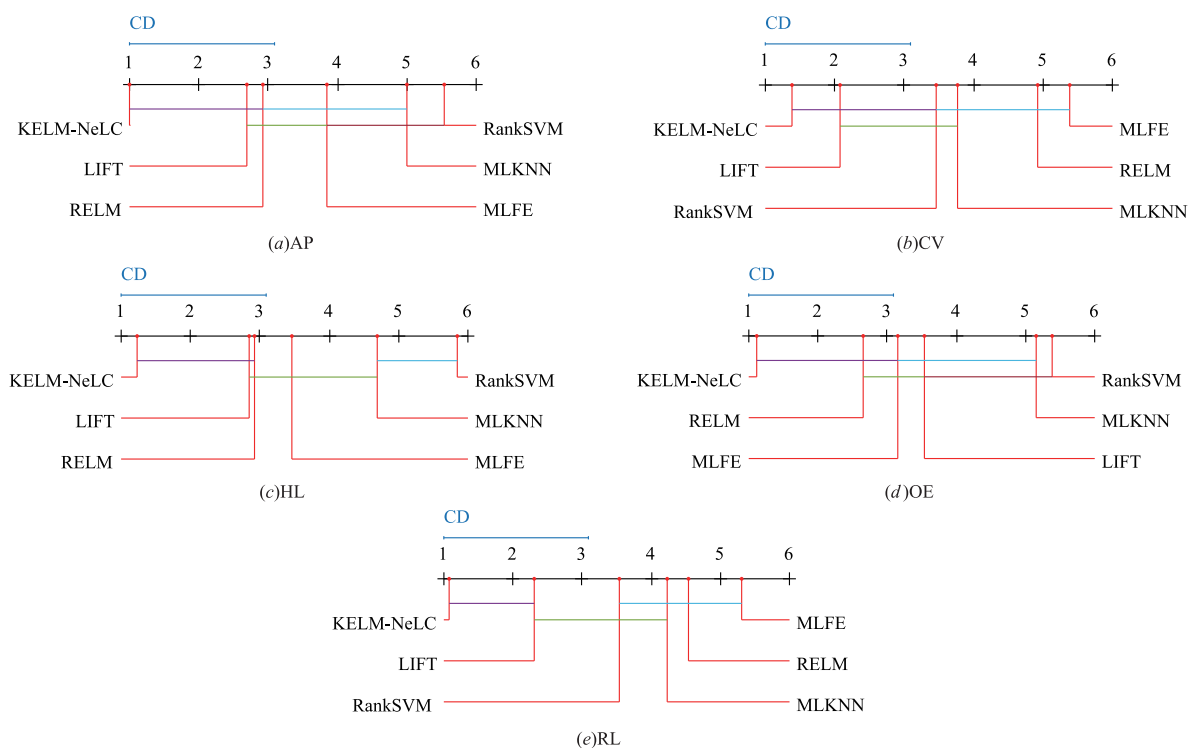


图1 算法综合性能比较

6 结束语

在多标签分类学习中,标签间相关关系的研究是不可忽略的重要因素. 为了充分利用这一关系,本文引入非平衡参数,提出一种非平衡化标签补全核极限学习机多标签学习算法 KELM-NeLC. 将非平衡化标签补全和核 ELM 进行联合学习得到统一的学习框架,从而预测未知样本标签集合. 实验结果充分证明了本文提

出的 KELM-NeLC 算法的有效性、稳定性和标签间相关关系语义的合理性.

参考文献

- [1] Zhang M L, Zhou Z H. A review on multi-label learning algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837.
- [2] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach

- to multi-label learning [J]. Pattern Recognition, 2007, 40 (7):2038–2048.
- [3] Elisseeff A, Weston J. A kernel method for multi-labelled classification [A]. Advances in Neural Information Processing Systems 14 [C]. Cambridge, UK: MIT Press, 2002. 681–687.
- [4] Zhang M L, Wu L. Lift: Multi-label learning with label-specific features [J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2015, 37(1):107–120.
- [5] Shannon C E. A mathematical theory of communication [J]. Bell System Technical Journal, 1948, 27(3):379–423.
- [6] Lee J, Kim H, Kim N R, et al. An approach for multi-label classification by directed acyclic graph with label correlation maximization [J]. Information Sciences, 2016, 351(C):101–114.
- [7] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1–3):489–501.
- [8] Huang G B. An Insight into extreme learning machines: random neurons, random features and kernels [J]. Cognitive Computation, 2014, 6(3):376–390.
- [9] Ding S, Xu X, Nie R. Extreme learning machine and its applications [J]. Neural Computing & Applications, 2014, 25(3–4):549–556.
- [10] Huang G, Huang G B, Song S, et al. Trends in extreme learning machines: a review [J]. Neural Networks the Official Journal of the International Neural Network Society, 2015, 61(C):32–48.
- [11] Luo F, Guo W, Yu Y, et al. A multi-label classification algorithm based on kernel extreme learning machine [J]. Neurocomputing, 2017, 260:313–320.
- [12] Liu Y, Wen K, Gao Q, et al. SVM based multi-label learning with missing labels for image annotation [J]. Pattern Recognition, 2018, 78:307–317.
- [13] Pizzuti C. A Multi-objective genetic algorithm for community detection in networks [A]. IEEE International Conference on TOOLS with Artificial Intelligence [C]. Newark, New Jersey, USA: IEEE Computer Society, 2009. 379–386.
- [14] Zhang M L, Zhou Z H. A Review on multi-label learning algorithms [J]. IEEE Transactions on Knowledge & Data Engineering, 2014, 26(8):1819–1837.
- [15] 邓万宇, 郑庆华, 陈琳, 等. 神经网络极速学习方法研究 [J]. 计算机学报, 2010, 33(2):279–287.
Deng W Y, Zheng Q H, Chen L, et al. Research on extreme learning of neural networks [J]. Chinese Journal of Computers, 2010, 33(2):279–287. (in Chinese)
- [16] Zhang Q W, Zhong Y, Zhang M L. Feature-induced labeling information enrichment for multi-label learning [A]. Thirty-Second AAAI Conference on Artificial Intelligence [C]. New Orleans: AAAI in Press, 2018.
- [17] 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(9):1967–1981.
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)
- [18] Ar J. Statistical comparisons of classifiers over multiple data sets [J]. Journal of Machine Learning Research, 2006, 7(1):1–30.

作者简介



程玉胜 男, 1969 年出生于安徽安庆. 现为安庆师范大学计算机与信息学院博士、教授. 主要研究方向包括数据挖掘, 粗糙集和多标签特征选择学习等.
E-mail: chengyusheng@163.com



赵大卫 男, 1993 年出生于安徽芜湖. 安庆师范大学计算机与信息学院研究生. 主要研究方向包括机器学习, 数据挖掘和统计等.
E-mail: like854@qq.com



王一宾 男, 1970 年出生于安徽安庆. 安庆师范大学计算机与信息学院教授. 主要研究方向包括多标签学习, 机器学习和软件安全等.
E-mail: wangyb07@mail.ustc.edu.cn



裴根生 男, 1992 年出生于安徽巢湖. 安庆师范大学计算机与信息学院研究生. 主要研究方向包括机器学习, 数据挖掘和统计等.
E-mail: 509692177@qq.com