

近邻标签空间非平衡化标签补全的多标签学习

程玉胜^{1,2} 赵大卫¹ 钱 坤¹

摘 要 研究者目前通常通过标注标签之间的相关信息研究标签之间的相关性,未考虑未标注标签与标注标签之间的关系对标签集质量的影响.受 K 近邻的启发,文中提出近邻标签空间的非平衡化标签补全算法(NeLC-NLS),旨在充分利用近邻空间中元素的相关性,提升近邻标签空间的质量,从而提升多标签分类性能.首先利用标签之间的信息熵衡量标签之间关系的强弱,获得基础标签置信度矩阵.然后利用提出的非平衡标签置信度矩阵计算方法,获得包含更多信息的非平衡标签置信度矩阵.继而度量样本在特征空间中的相似度,得到 k 个近邻标签空间样本,并利用非平衡标签置信度矩阵计算得到近邻标签空间的标签补全矩阵.最后利用极限学习机作为线性分类器进行分类.在公开的 8 个基准多标签数据集上的实验表明,NeLC-NLS 具有一定优势,使用假设检验和稳定性分析进一步说明算法的有效性.

关键词 多标签学习, 标签相关性, 信息熵, 标签补全, 极限学习机

引用格式 程玉胜,赵大卫,钱 坤.近邻标签空间非平衡化标签补全的多标签学习.模式识别与人工智能,2018,31(8):740-749.

DOI 10.16451/j.cnki.issn1003-6059.201808006

中图法分类号 TP 18

Multi-label Learning for Non-equilibrium Labels Completion in Neighborhood Labels Space

CHENG Yusheng^{1,2}, ZHAO Dawei¹, QIAN Kun¹

ABSTRACT The correlation between labels are studied through the related information about the marked labels. However, the influence of the relationship between unmarked and marked labels on the quality of the multi-label set is not taken into account. Inspired by k -nearest neighbors(KNN), a non-equilibrium labels completion of neighboring labels space(NeLC-NLS) is proposed to improve the quality of the neighboring label space and the performance of the multi-label classification. Firstly, the information entropy between labels is utilized to measure the strength of the relationship between labels, and the confidence matrix of the basic label is obtained. Then, the confidence matrix of non-equilibrium labels containing more information is obtained via the proposed non-equilibrium label confidence matrix. Secondly, the similarity of samples is measured in the feature space and the k -nearest neighbors are obtained. Then, the non-equilibrium labels completion matrix is employed to calculate the label completion matrix of the neighboring labels space. Finally, the extreme learning machine is adopted as a linear classifier. The experimental results of the proposed algorithm on 8 public multi-label datasets show that NeLC-NLS is superior to other multi-label learning algorithms. The effectiveness of NeLC-NLS is further illustrated by using hypothesis testing and stability analysis.

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

收稿日期:2018-04-18;录用日期:2018-07-21

Manuscript received April 18, 2018;

accepted July 21, 2018

安徽省高校自然科学基金重点项目(No. KJ2017A352)、安徽省高校重点实验室基金项目(No. ACAIM160102)资助

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

Key Words Multi-label Learning, Label Correlations, Information Entropy, Label Completion, Extreme Learning Machine

Citation CHENG Y S, ZHAO D W, QIAN K. Multi-label Learning for Non-equilibrium Labels Completion in Neighborhood Labels Space. Pattern Recognition and Artificial Intelligence, 2018, 31(8): 740-749.

多标签学习^[1]一直是机器学习、模式识别领域的研究热点之一。在多标签学习任务中,单个实例与多个标签相关联,通过训练集训练有效模型,有效预测未知实例所属的标签集合。对此学者们提出大量的多标签学习算法。二级分类 BR 算法(Binary Relevance)^[2]、标签幂级 LP 算法(Label Power Set)等通过增加分类器个数或标签的种类以解决多标签问题。反向传播多标签学习(Back-Propagation for Multi-label Learning, BP-MLL)^[3]引入排序损失(Ranking Loss)因素,减少运算时间,大幅增加计算复杂度。多标签 K 近邻算法(K -Nearest Neighbor for Multi-label Learning, ML-KNN)^[4]利用最大化后验概率(Maximum a Posterior, MAP)预测待测样本的标签集,时间复杂度较低,但是并不能保证近邻样本能较好地预测最终的标签集合。

包括上述算法在内的大多数多标签分类学习方法通常忽略标签之间的相关性。然而,真实世界中的标签和标签之间并非相互独立,它们之间存在一定的关联性。可见在没有足够多的训练集时,充分利用标签间相关性以丰富训练集的标签集是必要的。

目前,许多学者利用标签间的相关性提高多标签学习系统的性能。链性分类(Classifier Chains, CC)^[5]将多标签学习问题转化为基于 BR 的分类器链,但是链是随机排列的,标签间相关性的考虑也是随机的。校准标记排序算法(Calibrated Label Ranking, CLR)^[6]利用人工校准标签进行排序,但是遇到数据集较大、类别标签较多等问题时,不仅耗时耗力,而且得到的子分类器过多,增加算法的复杂度。随机 k 标签集(Random k -Labelsets, RAKEL)^[7]为基于 LP 分类器的集成算法,在保留 LP 优点的同时考虑标签之间的相关性,但会导致算法更复杂。多标记分类的核方法(A Kernel Method for Multi-labelled Classification, RankSVM)^[8]利用最大间隔准则策略适应多标签学习,在建模过程中对样本对应的相关标签和不相关标签间的排序损失构建 SVM 分类器。因为需要计算大量的变量,所以耗时较大。Zhang 等^[9]提出利用标签之间的协方差矩阵描述成对标签之间的相关性(可以正相关、不相关或负相关),但是只能求出成对标签之间的相关性。

同时,作为不确定性的一种有效度量方法,信息熵^[10]等相关信息论知识广泛应用于标签相关性研究中。基于该理论,张振海等^[11]提出基于相关信息熵的多标签分类算法,在 RAKEL 基础上考虑到其没有衡量标签之间的相关关系,使用相关信息熵衡量标签之间相关关系的强弱,提高多标签分类性能。Lee 等^[12]提出基于 CC 的多标签学习方法,使用有向无环图对标签的相关性进行建模,利用条件熵设计使标签之间相关性最大化的多标签学习方法,取得不错效果。

不难发现,利用标签之间的信息熵表示标签间相关性是可行的。但是 Huang 等^[13]通过重用权值计算发现标签关系不对称,传统的互信息计算的置信度矩阵不可行,同时这些方法通常度量标注标签间的信息熵,依此度量标注标签间的互相影响。由此可见,利用这种关系衡量标签之间的相关信息,仅考虑标注标签间的互相影响,却忽视样本标签集中未标注标签与标注标签间的影响。然而在样本标签集中未标注标签中可能包含大量有价值信息。

此外,研究发现,近邻空间多标签学习可以提高近邻标签空间质量,提升分类器性能。张敏灵^[14]在 ML-KNN 基础上提出新型多标签懒惰学习算法(Improved Multi-label Lazy Learning Approach, IMLLA),首先统计测试样本的 k 个近邻的标签信息,利用这些信息对每个标签构建一个标签计数向量,最后使用已训练的线性分类器计算该向量,用于分类。算法考虑到标签相关性对分类器性能的提高,但构建标签计数向量并不能准确反映标签之间存在的复杂关系。Younes 等^[15]在 ML-KNN 的 MAP 基础上加入邻域标签数量关系,依此考虑标签相关性。Gweon 等^[16]提出多标签学习方法,使用双距离的最近邻标签集,预测在训练数据中观察到的标签集,用于最小化特征空间和新实例标签空间中距离的加权和,但仅隐含考虑标签间的相关性。

基于上述方法的考虑,本文针对标签之间的相关性,考虑到近邻标签空间对泛化性能的作用,在近邻标签空间中引入非平衡参数,提出近邻标签空间的非平衡化标签补全算法(Non-equilibrium Labels Completion of Neighboring Labels Space, NeLC-

NLS), 使用极限学习机 (Extreme Learning Machine, ELM) [17-19] 求解权重, 提高算法效率和分类精度. 在多个公开的多标签数据集上的实验和统计假设检验及其稳定性分析不仅验证 NeLC-NLS 的有效性, 同时也证实近邻标签空间考虑标签间相关性以提升算法性能的合理性.

1 非平衡化标签置信度矩阵建模

1.1 多标签理论及其信息熵度量

定义 1 给定样本特征矩阵

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T \in \mathbf{R}^{N \times d},$$

其中, N 表示训练数据中样本的数目, d 表示特征的维度, $\mathbf{x}_i \in \mathbf{R}^d$ 表示对应的第 i 个样本的特征向量,

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]^T \in \mathbf{R}^{N \times L}$$

表示样本对应的标签矩阵, 其中, L 表示数据中包含标签的数目, $\mathbf{y}_i = \{1, -1\}^L$ 表示第 i 个样本对应的二元标签指示向量, 含有 N 个样本的多标签训练数据集可表示为

$$\mathbf{D} = \{(\mathbf{x}_i, \mathbf{y}_i) \mid 1 \leq i \leq N\} \subset \mathbf{R}^d \times \{+1, -1\}^L.$$

定义 2 设集合 $A = \{a_1, a_2, \dots, a_m\}$, $p(a_i)$ 为元素 a_i 的先验概率, 称

$$H(A) = - \sum_{i=1}^m p(a_i) \log_2 p(a_i)$$

为集合 A 的信息熵, 信息熵的值越大说明集合的不确定程度越大.

定义 3 设集合 $A = \{a_1, a_2, \dots, a_m\}$, 集合 $B = \{b_1, b_2, \dots, b_n\}$, 在给定集合 A 条件下集合 B 的条件熵为

$$H(B|A) = - \sum_{i=1}^m \sum_{j=1}^n p(b_j|a_i) \log_2 p(b_j|a_i),$$

其中, $H(b_j|a_i)$ 为条件信息量, 用于描述在元素 a_i 出现时, 元素 b_j 的不确定程度大小. 值越大表示两个元素之间的不确定程度越大, 否则表示两个元素之间存在某种程度的联系. 记为

$$H(b_j|a_i) = - p(a_i b_j) \log_2 p(b_j|a_i). \quad (1)$$

可见, 条件熵用于描述在集合 A 出现时, 集合 B 的不确定程度大小.

1.2 非平衡化标签补全矩阵建模

在真实世界中许多额外的信息蕴含在未标注标签集中, 因此利用未标注标签与标注标签之间关系分析标签之间关系应能得到更多有价值的信息. 基于这种考虑, 利用式(1)分别计算每个实例的标签集合中未标注标签和标注标签元素之间的信息量, 得到未标注标签相关的基础标签置信度矩阵.

定义 4 假设训练样本矩阵

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]^T \in \mathbf{R}^{N \times L}, \mathbf{y}_i = \{1, -1\}^L.$$

考虑到标签间关系并非对称存在, 本文定义如下基础标签置信度矩阵, 使用条件信息量的倒数描述元素间的置信度大小:

$$a_{ij} = \frac{1}{H(\bar{l}_j | l_i)}, b_{ij} = \frac{1}{H(l_j | \bar{l}_i)},$$

其中, l_i 表示 y_i 值为 1, \bar{l}_i 表示 y_i 值为 -1, $i = 1, 2, \dots, L, j = 1, 2, \dots, L, i \neq j$. a_{ij} 侧重表示标注标签对未标注标签的置信度大小, b_{ij} 侧重表示未标注标签对标注标签的置信度大小.

在众多的多标签相关性研究中, 基本上都是计算两两标签相关性的大小. 然而这些方法都未考虑未标注标签和标注标签之间的关系, 更未考虑标签关系不成对的现象. 而对“Missing”标签补全的方法在应用于多标签学习算法时, 也仅考虑标注标签数据对未标注标签数据的影响 [20]. 然而在未标注标签中可能包含大量有价值信息, 受其启发, 需要加大对未标注信息的关注. 因此, 增加未标注相关标签的权重有利于缺省标签的补全. 基于这种考虑, 本文引入非平衡参数 α , 通过减小 a_{ij} 和增加 b_{ij} 的基础标签置信度矩阵加权计算方法, 提出非平衡化的标签置信度矩阵计算方法:

$$\text{Conf}_{ij} = -\alpha a_{ij} + (1 - \alpha) b_{ij}. \quad (2)$$

受标签传播依赖思想 [21] 启发, 定义如下的 $L \times N$ 维标签补全矩阵:

$$\hat{\mathbf{Y}} = \text{Conf} \times \mathbf{Y}. \quad (3)$$

引入非平衡参数后非平衡化标签补全矩阵算法 (Non-equilibrium Labels Completion, NeLC) 步骤如下.

算法 1 NeLC

输入 训练样本矩阵 \mathbf{Y} , 非平衡参数 α

输出 非平衡化标签补全矩阵 $\hat{\mathbf{Y}}$
 $\mathbf{Y} = \{\mathbf{Y}_i \mid i = 1, 2, \dots, L\}$ /* 训练集标签集合 */
 for each l_i, l_j
 if $i \neq j$

$$a_{ij} = \frac{1}{H(\bar{l}_j | l_i)}, b_{ij} = \frac{1}{H(l_j | \bar{l}_i)}$$

/* 利用式(1)分别计算得出 a_{ij} 和 b_{ij} */

else if $i = j$

$$a_{ij} = 1, b_{ij} = 1$$

end

/* 将对角线元素置为 1 */

$$\text{Conf}_{ij} = -\alpha a_{ij} + (1 - \alpha) b_{ij};$$

/* 利用式(2) 求出置信度矩阵 */

$\hat{Y} = \text{Conf} \times Y$

/* 利用式(3) 求出非平衡化标签补全矩阵 */

end

return \hat{Y} /* 返回非平衡化标签补全矩阵 */

2 近邻标签空间的非平衡化

2.1 近邻标签补全矩阵构建

本文在 KNN 思想基础上首先利用曼哈顿距离度量训练集中每个特征空间样本之间的相似度. 基于训练集中 k 个近邻特征空间得到每个样本对应的近邻标签矩阵 $N_{x_{train}}^l, x_{train} \in \mathbf{x}_i$, 根据式(3) 定义 $L \times k$ 维标签补全矩阵:

$$LC_{train_k} = \text{Conf} \times N_{x_{train}}^l. \quad (4)$$

计算训练集合下的近邻标签空间下 LC_{train_k} 的和, 得到 $N \times L$ 的近邻标签空间补全矩阵 M_{train} :

$$M_{train} = \frac{1}{k} \sum_{i=1}^k LC_{train_k}(i, N). \quad (5)$$

对于测试集中特征空间的每个样本, 本文算法首先确定该样本在训练样本中对应的近邻集合 $N_{x_{test}}^l$. 类似地使用式(4) 计算得到对应的补全近邻矩阵 LC_{test_k} , 再由式(5) 获得矩阵 M_{test} 提供的信息确定第 l 个标签的输出. 本文采用如下线性分类器确定样本 \mathbf{x} 在第 l 类标签上的输出:

$$f_l(\mathbf{x}) = M\beta. \quad (6)$$

2.2 极限学习机

ELM 为一种快速的前馈单隐藏层神经网络学习算法, 网络结构中的隐藏层的参数随机选择, 只需设置隐藏层神经网络神经元的个数, 最后使用最小二乘法求出隐藏层的输出权重. 因此, 相比神经网络算法, ELM 可以保证网络具有良好的泛化能力, 同时避免传统神经网络算法迭代学习带来的问题, 如网络权重和偏置参数的调整、陷入局部最优解等. 而 ELM 输出权值是通过奇异值表现输入数据, 在解决大型的非结构化数据时取得较好的泛化性能.

定义 5 设有 N 个随机样本 $\{(\mathbf{X}_i, \mathbf{Y}_i)\}$, 其中

$$\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{in}]^T,$$

$$\mathbf{Y}_i = [\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{im}]^T,$$

$$i = 1, 2, \dots, N.$$

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

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

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

定义 6 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 7 通过式(7)和式(8)可知, 在确定 β_i, ω_i 和 b_i 时可得

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

将式(9) 使用矩阵表示为

$$H\beta = Y. \quad (10)$$

2.3 近邻标签空间非平衡化权重求解

根据式(6) 和式(10), 利用 ELM 求解权重, 在 ELM 中输出函数 $f_l(\mathbf{x}_i)$ 可表示为

$$f_l(\mathbf{x}) = H\beta = \begin{bmatrix} h(\mathbf{x}_1) \\ h(\mathbf{x}_2) \\ \vdots \\ h(\mathbf{x}_N) \end{bmatrix}_{N \times d} \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{d \times L}. \quad (11)$$

根据式(8) 可得

$$\begin{aligned} \beta &= H^+ Y, \\ \text{s. t. } H^+ &= \begin{cases} (H^T H)^{-1} H^T \\ H^T (H H^T)^{-1} \end{cases} \end{aligned} \quad (12)$$

如果 $H^T H$ 为非奇异矩阵, $H^+ = (H^T H)^{-1} H^T$; 如果 $H H^T$ 为非奇异矩阵, $H^+ = H^T (H H^T)^{-1}$, H^+ 为 H 的 Moore-Penrose 广义逆矩阵. 根据岭回归理论 (Ridge Regression Theory), 建议在 $H H^T$ 或 $H^T H$ 对角线上加上正则项 C , 提高算法的稳定性和泛化能力, 式(11) 最小化目标为

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

$$\text{s. t. } \xi_i = Y_i - f_i(\mathbf{x}_i), \quad i = 1, 2, \dots, N.$$

根据 KKT 最佳条件, 隐藏输出权重 β 表示为

$$\beta = H^T \left(\frac{I}{C} + H H^T \right)^{-1} Y, \quad (13)$$

此时多标签输入函数可表示为

$$f_l(\mathbf{x}) = H\beta = H H^T \left(\frac{I}{C} + H H^T \right)^{-1} Y. \quad (14)$$

在 ELM 中需要设置隐藏层节点数, 随机设定权重和偏置, 会导致计算结果不稳定, 容易受到随机设定值的影响. 使用核函数可以避免神经元参数设置的随机性, 解决这一问题. 当映射函数 $h(\mathbf{x})$ 未知时, 引入核矩阵 (本文采用径向基函数 (Radial Basis

Function, RBF) 核):

$$\boldsymbol{\Omega}_{\text{ELM}} = \boldsymbol{H}\boldsymbol{H}^T: \boldsymbol{\Omega}_{\text{ELM}(i,j)} = h(\boldsymbol{x}_i, \boldsymbol{x}_j), \quad (15)$$

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|).$$

根据式(13), 可得

$$\boldsymbol{H}\boldsymbol{H}^T = \begin{bmatrix} K(\boldsymbol{x}, \boldsymbol{x}_1) \\ K(\boldsymbol{x}, \boldsymbol{x}_2) \\ \vdots \\ K(\boldsymbol{x}, \boldsymbol{x}_N) \end{bmatrix}^T,$$

这样式(14) 就可表示为

$$f_i(\boldsymbol{x}) = h(\boldsymbol{x})\boldsymbol{H}^T \left(\frac{\boldsymbol{I}}{C} + \boldsymbol{H}\boldsymbol{H}^T \right)^{-1} \boldsymbol{Y} = \begin{bmatrix} K(\boldsymbol{x}, \boldsymbol{x}_1) \\ K(\boldsymbol{x}, \boldsymbol{x}_2) \\ \vdots \\ K(\boldsymbol{x}, \boldsymbol{x}_N) \end{bmatrix} \left(\frac{\boldsymbol{I}}{C} + \boldsymbol{\Omega}_{\text{ELM}} \right)^{-1} \boldsymbol{Y}. \quad (16)$$

本文通过使用拟合回归训练得出权重 $\boldsymbol{\beta}$. 结合式(8)、式(12) 和式(16), 已知训练数据标签集合 \boldsymbol{Y} 和拟合训练得到的输出权重 $\boldsymbol{\beta}$, 即可预测得到标签集合. 最小化目标函数为

$$f_E = \sum_i^N \|f_i(\boldsymbol{x}_i) - \boldsymbol{Y}_i\|^2.$$

近邻标签空间的非平衡化标签补全算法 (NeLC-NLS) 步骤如下.

算法 2 NeLC-NLS

输入 多标签训练数据集 $\boldsymbol{D} = \{(\boldsymbol{x}_i, \boldsymbol{Y}_i) \mid 1 \leq i \leq N\}$,
测试数据集 $\boldsymbol{D}^* = \{(\boldsymbol{x}_j, \boldsymbol{Y}_j) \mid 1 \leq j \leq T\}$,
正则参数 C , RBF 核参数 α ,
非平衡参数 α , 近邻个数 k

输出 预测标签 \boldsymbol{Y}^*

$\boldsymbol{Y} = \{\boldsymbol{Y}_i \mid i = 1, 2, \dots, k\}$ /* \boldsymbol{D} 中的训练标签集合 */

for each $\boldsymbol{Y}_i \in \mathbf{R}^k$

\boldsymbol{Conf} /* 根据算法 1 得出标签置信度矩阵 \boldsymbol{Conf} */
end

$\boldsymbol{LC_train}_k = \boldsymbol{Conf} \times \boldsymbol{N}_{x_train}^l$.

$\boldsymbol{LC_test}_k = \boldsymbol{Conf} \times \boldsymbol{N}_{x_train}^l$
/* 根据式(4) 得出标签补全矩阵 */

$$\boldsymbol{M}_{\text{train}} = \frac{1}{k} \sum_{i=1}^k \boldsymbol{LC_train}_k(i, N)$$

$$\boldsymbol{M}_{\text{test}} = \frac{1}{k} \sum_{i=1}^k \boldsymbol{LC_test}_k(i, N)$$

/* 根据式(5) 得出近邻标签补全矩阵 */

for each $\boldsymbol{x}_i \in \boldsymbol{M}_{\text{train}}$

$$\boldsymbol{\Omega}_{\text{ELM}} = \exp \left(-\frac{\|\boldsymbol{x}_i\|^2}{\sigma} \right)$$

/* 输入特征进行核映射 */

end

$$\boldsymbol{\beta} = \left(\frac{\boldsymbol{I}}{C} + \boldsymbol{\Omega}_{\text{ELM}(i)} \right)^{-1} \boldsymbol{Y}$$

/* 通过式(13) 和式(16), 求解输出权重 */

for each $\boldsymbol{x}_i \in \boldsymbol{M}_{\text{train}}$

for each $\boldsymbol{x}_j^* \in \boldsymbol{M}_{\text{test}}$

$$\boldsymbol{\Omega}_{\text{ELM}(i,j)} = \exp \left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j^*\|^2}{\sigma} \right)$$

end

end

$f_i(\boldsymbol{x}_j^*) = \boldsymbol{\Omega}_{\text{ELM}(i,j)} \boldsymbol{\beta}$; /* 测试集预测标签结果 */

$\boldsymbol{Y}^* = \{\boldsymbol{y}_j^* \mid f_i(\boldsymbol{x}_j^*) > 0, j = 1, 2, \dots, T\}$

/* $f_i(\boldsymbol{x}_j^*) > 0$ 表示该标签存在 */

return \boldsymbol{Y}^*

3 实验及结果分析

3.1 实验环境

为了说明本文算法的有效性, 选取 8 个包含多个领域的多标签数据集. 数据集均来自 <http://mulan.sourceforge.net/datasets-mlc.html>. 具体描述见表 1.

表 1 多标签数据集的详细描述

Table 1 Detailed descriptions of multi-label datasets

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
Birds	322	323	20	260	1.470	0.074	Audio
Cal500	351	151	174	68	26.044	0.150	Music
Emotions	391	202	6	72	1.868	0.311	Music
Enron	1123	579	53	1001	3.378	0.064	Text
Flags	129	65	7	19	3.392	0.485	Images
Natural scene	1000	1000	5	294	1.236	0.247	Images
Image	1000	1000	5	294	1.236	0.245	Images
Yeast	1500	917	14	103	4.237	0.303	Biology

实验代码均在 Matlab 2016a 中运行,硬件环境为 Intel® Core_(TM) i5-2380p 3.10 GHz CPU,8 GB 内存,操作系统为 Windows 7. 本文选取常用的 5 种评价准则:平均精度(Average Precision, AP),覆盖率(Coverage, CV),海明损失(Hamming Loss, HL),1- 错误率(One-Error, OE),排序损失(Ranking Loss, RL)^[22],综合评价多标签学习算法的性能. 为了方便,分别简写为 AP↑、CV↓、HL↓、OE↓ 和 RL↓,其中,↑表示指标数值越高越好,↓表示指标数值越低越好. 设多标签分类器为 $h(\cdot)$,预测函数为 $f(\cdot, \cdot)$,排序函数为 $rank_f$,多标签数据集 $D = \{(\mathbf{x}_i, \mathbf{Y}_i \mid 1 \leq i \leq n)\}$. 上述 5 种评价指标形式化定义如下.

1) AP. 评估特定标签 $y \in \mathbf{Y}_i$ 排列的正确标签的平均分数:

$$AP_D(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|\mathbf{Y}_i|} \cdot \sum_{y \in \mathbf{Y}_i} \frac{|\{rank_f(\mathbf{x}_i, y') \leq rank_f(\mathbf{x}_i, y), y' \in \mathbf{Y}_i\}|}{rank_f(\mathbf{x}_i, y)}.$$

2) CV. 度量平均需要多少步才能遍历样本所有的相关标签:

$$CV_D(f) = \frac{1}{n} \sum_{i=1}^n \max_{y \in \mathbf{Y}_i} rank_f(\mathbf{x}_i, y) - 1.$$

3) HL. 度量样本在单个标签情况下的真实标签和预测标签的错误匹配情况:

$$HL_D(h) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|\mathbf{Y}|} |\{h(\mathbf{x}_i) \neq \mathbf{Y}_i\}|.$$

4) OE. 评估对象最高排位标签并非正确标签的次数:

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

5) RL. 考察样本的不相关标签的排序低于相关标签的排序情况:

$$RL_D(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|\mathbf{Y}_i| \cdot |\overline{\mathbf{Y}}_i|} \cdot |\{(y_1, y_2) \mid$$

$$f(\mathbf{x}_i, y_1) \leq f(\mathbf{x}_i, y_2), (y_1, y_2) \in \mathbf{Y}_i \times \overline{\mathbf{Y}}_i\}|.$$

为了验证本文算法性能,将 NeLC-NLS 与 1 个基分类器为 ML-KNN 的 NLS-MLKNN 及 5 个多标签分类算法进行对比实验,分别为 RELM(Regularized ELM)^[23]、ML-KNN、IMLLA、RankSVM 和 ML-KELM (Multi-label Learning of Kernel Extreme Learning Machine)^[24].

在 NeLC-NLS 中正则化系数设为 1,核函数选择 RBF,核参数设在 $[1, 100]$ 内,训练方式选择回归拟合. 在 NLS-MLKNN 中近邻个数 $k = 15$,平滑参数 $s = 1$. 在 RELM 中正则化系数设为 1,隐藏层神经元个数设为 100,因为 RELM 不够稳定,所以运行 10 次,并给出平均值和标准差,以便得到较准确的数值. 在 ML-KNN 中近邻个数 $k = 15$,平滑参数 $s = 1$. 在 IMLLA 中近邻个数 $k = 15$. 在 RankSvm 中,代价参数设为 1,同时选择 RBF 为核函数. 在 ML-KELM 中正则化系数设为 1,核函数选择 RBF,核参数设在 $[1, 100]$ 内.

3.2 实验结果

表 2 ~ 表 6 给出本文算法和其它 6 种算法在 8 个基准多标签数据集上对应的实验结果,在每个评价指标数据下标注排位顺序,最优结果使用黑体数字表示,同时,每种方法在所有数据集上的平均排序结果列在最后一行,平均排序越小,算法性能越优.

表 2 各算法在 8 个数据集上的平均精度
Table 2 AP results of different algorithms on 8 datasets

数据集	NeLC-NLS	NLS-MLKNN	RELM	ML-KNN	IMLLA	RankSVM	ML-KELM
Birds	0.6823 ₍₃₎	0.6867 ₍₂₎	0.7124 ± 0.0106 ₍₁₎	0.6750 ₍₅₎	0.6649 ₍₆₎	0.6112 ₍₇₎	0.6807 ₍₄₎
Cal500	0.5171 ₍₁₎	0.4922 _(5,5)	0.5015 ± 0.0035 ₍₄₎	0.4922 _(5,5)	0.3409 ₍₇₎	0.5029 ₍₃₎	0.5108 ₍₂₎
Emotions	0.8222 ₍₁₎	0.8054 ₍₂₎	0.7640 ± 0.0162 ₍₆₎	0.7878 _(3,5)	0.7878 _(3,5)	0.7503 ₍₇₎	0.7867 ₍₅₎
Enron	0.6615 ₍₁₎	0.5833 ₍₆₎	0.6367 ± 0.0074 ₍₄₎	0.6332 ₍₅₎	0.6476 ₍₃₎	0.5225 ₍₇₎	0.6591 ₍₂₎
Flags	0.8050 ₍₁₎	0.7921 ₍₃₎	0.7749 ± 0.0130 ₍₆₎	0.7856 ₍₅₎	0.7875 ₍₄₎	0.8007 ₍₂₎	0.7400 ₍₇₎
Natural scene	0.8043 ₍₁₎	0.7765 ₍₄₎	0.7531 ± 0.0065 ₍₇₎	0.7649 ₍₆₎	0.7987 ₍₂₎	0.7689 ₍₅₎	0.7858 ₍₃₎
Image	0.8139 ₍₁₎	0.7929 ₍₄₎	0.7712 ± 0.0056 ₍₇₎	0.7908 ₍₅₎	0.8082 ₍₂₎	0.7894 ₍₆₎	0.8052 ₍₃₎
Yeast	0.7676 ₍₁₎	0.7576 ₍₃₎	6.5815 ± 0.0262 ₍₇₎	0.7567 _(4,5)	0.7567 _(4,5)	0.7566 ₍₆₎	0.7623 ₍₂₎
平均排序	1.25	3.6875	5.25	4.9375	4.0	5.375	3.50

表 3 各算法在 8 个数据集上的覆盖率

Table 3 CV results of different algorithms on 8 datasets

数据集	NeLC-NLS	NLS-MLKNN	RELM	ML-KNN	IMLLA	RankSVM	ML-KELM
Birds	3.2693 ₍₁₎	3.6037 ₍₃₎	3.6325 ± 0.1759 ₍₄₎	3.6563 ₍₅₎	3.9102 ₍₆₎	4.2446 ₍₇₎	3.3375 ₍₂₎
Cal500	127.7815 ₍₁₎	129.2848 ₍₂₎	134.4053 ± 2.9534 ₍₆₎	129.5629 ₍₄₎	165.5894 ₍₇₎	129.3377 ₍₃₎	135.2649 ₍₅₎
Emotions	1.7525 ₍₁₎	1.8218 ₍₂₎	2.0965 ± 0.0733 ₍₆₎	1.8762 ₍₄₎	1.8663 ₍₃₎	2.2426 ₍₇₎	1.9257 ₍₅₎
Enron	13.6321 ₍₃₎	13.9102 ₍₄₎	16.7067 ± 0.6222 ₍₇₎	13.3713 ₍₂₎	15.1537 ₍₆₎	14.8411 ₍₅₎	13.3264 ₍₁₎
Flags	3.8615 ₍₂₎	3.8923 ₍₃₎	4.1215 ± 0.0344 ₍₆₎	4.0462 ₍₅₎	3.9385 ₍₄₎	3.6923 ₍₁₎	4.3077 ₍₇₎
Natural scene	0.9080 ₍₁₎	1.0470 ₍₄₎	1.0892 ± 0.0271 ₍₇₎	1.0520 ₍₅₎	0.9430 ₍₂₎	1.0550 ₍₆₎	0.9630 ₍₃₎
Image	0.8800 ₍₁₎	0.9520 ₍₄₎	1.0194 ± 0.0195 ₍₇₎	0.9530 ₍₅₎	0.9000 ₍₃₎	0.9760 ₍₆₎	0.8910 ₍₂₎
Yeast	6.2563 ₍₂₎	6.5627 ₍₆₎	6.5815 ± 0.0262 ₍₇₎	6.4318 ₍₅₎	6.2672 ₍₃₎	6.2475 ₍₁₎	6.3599 ₍₄₎
平均排序	1.5	3.5	6.25	4.375	4.25	4.5	3.625

表 4 各算法在 8 个数据集上的海明损失

Table 4 HL results of different algorithms on 8 datasets

数据集	NeLC-NLS	NLS-MLKNN	RELM	ML-KNN	IMLLA	RankSVM	ML-KELM
Birds	0.0550 ₍₃₎	0.0559 ₍₄₎	0.0543 ± 0.0018 ₍₁₎	0.0579 ₍₆₎	0.0577 ₍₅₎	0.0893 ₍₇₎	0.0545 ₍₂₎
Cal500	0.1393 ₍₁₎	0.1430 ₍₄₎	0.1400 ± 0.0008 ₍₃₎	0.1438 ₍₅₎	0.1941 ₍₇₎	0.1477 ₍₆₎	0.1399 ₍₂₎
Emotions	0.1889 ₍₁₎	0.2203 ₍₅₎	0.2381 ± 0.0069 ₍₆₎	0.2195 ₍₄₎	0.2030 ₍₂₎	0.2946 ₍₇₎	0.2079 ₍₃₎
Enron	0.0489 ₍₁₎	0.0599 ₍₅₎	0.0510 ± 0.0006 ₍₂₎	0.0517 ₍₄₎	0.0511 ₍₃₎	0.0641 ₍₇₎	0.0533 ₍₆₎
Flags	0.3077 ₍₂₎	0.3275 ₍₄₎	0.3255 ± 0.0157 ₍₃₎	0.3429 ₍₅₎	0.2989 ₍₁₎	0.3560 ₍₇₎	0.3516 ₍₆₎
Natural scene	0.1670 ₍₂₎	0.1856 ₍₄₎	0.1991 ± 0.0046 ₍₇₎	0.1866 ₍₅₎	0.1668 ₍₁₎	0.1854 ₍₃₎	0.1928 ₍₆₎
Image	0.1576 ₍₁₎	0.1926 ₍₆₎	0.1940 ± 0.0040 ₍₇₎	0.1740 ₍₃₎	0.1640 ₍₂₎	0.1756 ₍₄₎	0.1906 ₍₅₎
Yeast	0.1936 ₍₂₎	0.2013 ₍₅₎	0.2016 ± 0.0017 ₍₇₎	0.1987 ₍₄₎	0.1948 ₍₃₎	0.2024 ₍₆₎	0.1919 ₍₁₎
平均排序	1.625	4.625	4.5	4.5	3.0	5.875	4.25

表 5 各算法在 8 个数据集上的 1- 错误率

Table 5 OE results of different algorithms on 8 datasets

数据集	NeLC-NLS	NLS-MLKNN	RELM	ML-KNN	IMLLA	RankSVM	ML-KELM
Birds	0.4025 ₍₄₎	0.3870 ₍₂₎	0.3458 ± 0.0179 ₍₁₎	0.3994 ₍₃₎	0.4272 ₍₆₎	0.5325 ₍₇₎	0.4056 ₍₅₎
Cal500	0.1325 ₍₂₎	0.1391 ₍₄₎	0.1457 ± 0.0117 ₍₅₎	0.1325 ₍₂₎	0.4636 ₍₇₎	0.1325 ₍₂₎	0.1656 ₍₆₎
Emotions	0.2574 ₍₁₎	0.2723 ₍₂₎	0.3287 ± 0.0308 ₍₇₎	0.3168 _(4,5)	0.3218 ₍₆₎	0.3168 _(4,5)	0.3020 ₍₃₎
Enron	0.2625 ₍₂₎	0.3541 ₍₆₎	0.2874 ± 0.0148 ₍₄₎	0.2936 ₍₅₎	0.2867 ₍₃₎	0.4870 ₍₇₎	0.2556 ₍₁₎
Flags	0.2154 ₍₁₎	0.2308 _(2,5)	0.2585 ± 0.0496 ₍₅₎	0.2308 _(2,5)	0.2769 _(6,5)	0.2462 ₍₄₎	0.2769 _(6,5)
Natural scene	0.3000 ₍₁₎	0.3320 ₍₃₎	0.3852 ± 0.0107 ₍₇₎	0.3640 ₍₆₎	0.3070 ₍₂₎	0.3500 ₍₅₎	0.3340 ₍₄₎
Image	0.2880 ₍₁₎	0.3180 ₍₅₎	0.3549 ± 0.0162 ₍₇₎	0.3230 ₍₆₎	0.2960 ₍₂₎	0.3160 ₍₄₎	0.3040 ₍₃₎
Yeast	0.2301 ₍₁₎	0.2366 ₍₄₎	0.2395 ± 0.0051 ₍₆₎	0.2410 ₍₇₎	0.2312 ₍₂₎	0.2366 ₍₄₎	0.2366 ₍₄₎
平均排序	1.625	3.5625	5.25	4.5	4.3125	4.6875	4.0625

表 6 各算法在 8 个数据集上的排序损失

Table 6 RL results of different algorithms on 8 datasets

数据集	NeLC-NLS	NLS-MLKNN	RELM	ML-KNN	IMLLA	RankSVM	ML-KELM
Birds	0.1212 ₍₂₎	0.1349 ₍₄₎	0.1308 ± 0.0086 ₍₃₎	0.1358 ₍₅₎	0.1455 ₍₆₎	0.1629 ₍₇₎	0.1203 ₍₁₎
Cal500	0.1751 ₍₁₎	0.1837 ₍₃₎	0.1880 ± 0.0030 ₍₆₎	0.1854 ₍₄₎	0.3498 ₍₇₎	0.1811 ₍₂₎	0.1879 ₍₅₎
Emotions	0.1475 ₍₁₎	0.1591 ₍₂₎	0.2020 ± 0.0143 ₍₆₎	0.1692 ₍₃₎	0.1693 ₍₄₎	0.2244 ₍₇₎	0.1710 ₍₅₎
Enron	0.0920 ₍₁₎	0.1036 ₍₄₎	0.1183 ± 0.0061 ₍₇₎	0.0944 ₍₃₎	0.1039 ₍₅₎	0.1132 ₍₆₎	0.0934 ₍₂₎
Flags	0.2272 ₍₂₎	0.2449 ₍₃₎	0.2680 ± 0.0117 ₍₅₎	0.3429 ₍₆₎	0.2479 ₍₄₎	0.2262 ₍₁₎	0.3259 ₍₇₎
Natural scene	0.1593 ₍₁₎	0.1928 ₍₄₎	0.2057 ± 0.0062 ₍₇₎	0.1952 ₍₅₎	0.1667 ₍₂₎	0.1965 ₍₆₎	0.1726 ₍₃₎
Image	0.1550 ₍₁₎	0.1723 ₍₅₎	0.1894 ± 0.0047 ₍₇₎	0.1715 ₍₄₎	0.1593 ₍₃₎	0.1766 ₍₆₎	0.1563 ₍₂₎
Yeast	0.1643 ₍₁₎	0.1759 ₍₆₎	0.1786 ± 0.0012 ₍₇₎	0.1733 ₍₅₎	0.1662 ₍₂₎	0.1696 ₍₄₎	0.1669 ₍₃₎
平均排序	1.25	3.875	6.0	4.375	4.125	4.875	3.375

在 AP 指标上,NeLC-NLS 除了在 Birds 数据集上性能排第三,在其它数据集上均为最优. 在 CV 指标上,NeLC-NLS 在 Enron、Flags 和 Yeast 数据集上性能不是最优,在其它数据集上均为第一. 在 HL 指标上,NeLC-NLS 在 Birds、Flags 和 Natural scene 数据集上性能排名第二或第三,在其它数据集上均为最优. 在 OE 指标上,NeLC-NLS 在 Birds 数据集上性能排第四,在 Enron 数据集上排第二,在其它数据集均为最佳. 在 RL 指标上,NeLC-NLS 在 Birds、Flags 数据集上性能排名第二,在其它数据集均为最佳.

相比 RELM 和 MLKELM 以 ELM 为基分类器的算法,NeLC-NLS 整体性能表现更优,证明非平衡化标签补全算法的有效性. 此外本文设计的 NLS-MLKNN 整体性能上优于 ML-KNN,也证明非平衡化标签补全算法的有效性. 同时 NeLC-NLS 表现优于 NLS-MLKNN,说明统一学习框架的学习优于分段式学习.

3.3 算法性能及其稳定性统计分析

为了更好地对比 NeLC-NLS 与其它对比实验算法在 8 个数据集上的实验结果,采用显著性水平为 5% 的 Nemenyi 检验^[25-26]. 当 2 个对比算法在所有数据集上的平均排序的差值大于临界差值(Critical Difference, CD),认为这 2 种算法存在显著性差异,否则认为无显著差异. 图 1 给出在不同评价指标下的每种算法之间的对比,对于没有显著性差异的算法使用彩色实线相连,各评价指标中从左到右,算法性能排序依次降低.

对于每种算法,都有 30 种实验对比结果(6 种对比算法,5 种评价准则). 从图 1 得如下结论.

1) 在 27% 情况下,NeLC-NLS 在统计上与其它算法有显著差异,而且不差于其它算法性能. 在 AP 指标上,NeLC-NLS 与 ML-KNN、RELM 和 RankSVM 有显著差异. 在 CV 指标上,NeLC-NLS 与 RELM 有显著差异. 在 HL 评价指标上,NeLC-NLS 与 RankSVM 有显著差异. 在 OE 指标上,NeLC-NLS 与 RELM 有显著差异. 在 RL 指标上,NeLC-NLS 与 RankSVM 和 RELM 有显著差异.

2) 对比 NLS-MLKNN 和 ML-KNN,在统计上它们没有显著差异,但整体上 NLS-MLKNN 优于 ML-KNN.

3) 对比 NeLC-NLS 和 NLS-MLKNN,在统计上没有显著差异,但是整体上 NeLC-NLS 占优. 这说明统一框架下学习的方式优于分段式学习.

不难看出,NeLC-NLS 在所有指标上均排在第一. 从上述分析可知,NeLC-NLS 性能最优,统计假

设检验进一步说明 NeLC-NLS 的有效性,近邻标签空间考虑标签非平衡化补全的合理性.

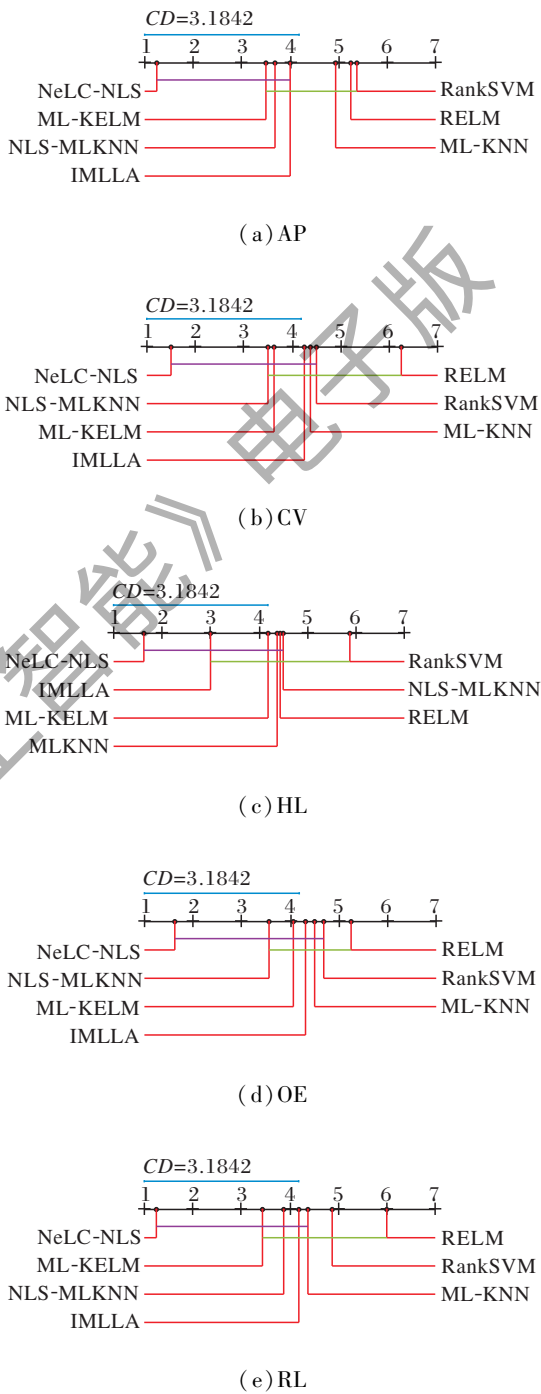


图 1 各算法性能对比

Fig. 1 Performance comparison of different algorithms

为了验证本文算法的稳定性,采用蜘蛛网图显示算法的稳定性^[27]. 因为预测分类在使用不同评价指标的不同数据集上的结果差异很大,所以将实验结果标准化在[0.1,0.5]内,作为通用标准. 最后通过归一化后的数值大小表示稳定指数. 图 2 显示各

评价指标下不同数据集上算法的稳定性.

如图 2 所示,在 AP 指标上,NeLC-NLS 在 7 个数据集的稳定指数值在 $[0.4,0.5]$ 内获得相当稳定的效果. 在 CV 指标上,相比其它算法,NeLC-NLS 在 7 个数据集上的稳定指数值在 $[0.4,0.5]$ 内,可实现相当稳定的解决方案. 在 HL 指标上,NeLC-NLS 可在 7 个数据集上得到更稳定的结果,比 RELM、MLKNN、IMLLA 和 RankSVM 稳定. 在 OE 指标上,NeLC-NLS 可在 7 个数据集上提供更稳定的解决方案. 在 RL 指标上,NeLC-NLS 在所有数据集上均实现稳定的解决方案,比其它对比算法更稳定. 因此,图 2 表明,NeLC-NLS 更稳定并具有更好的预测性能.

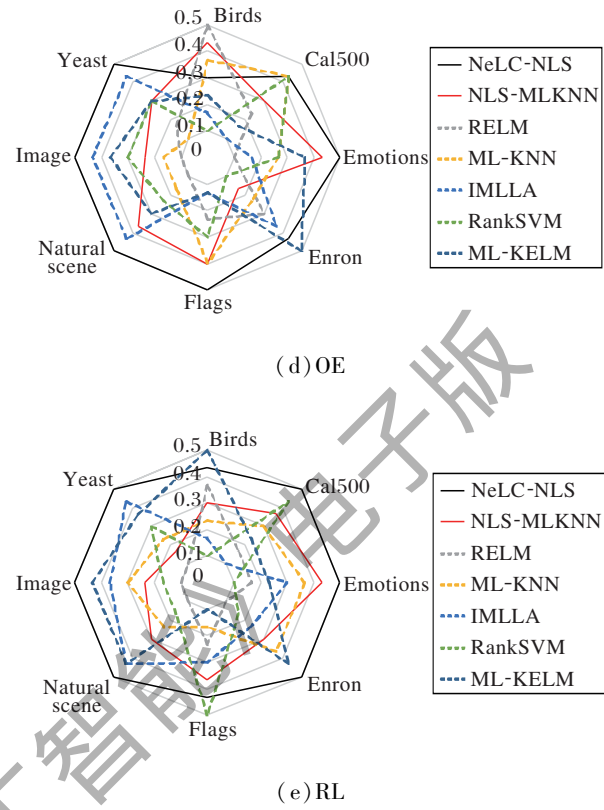
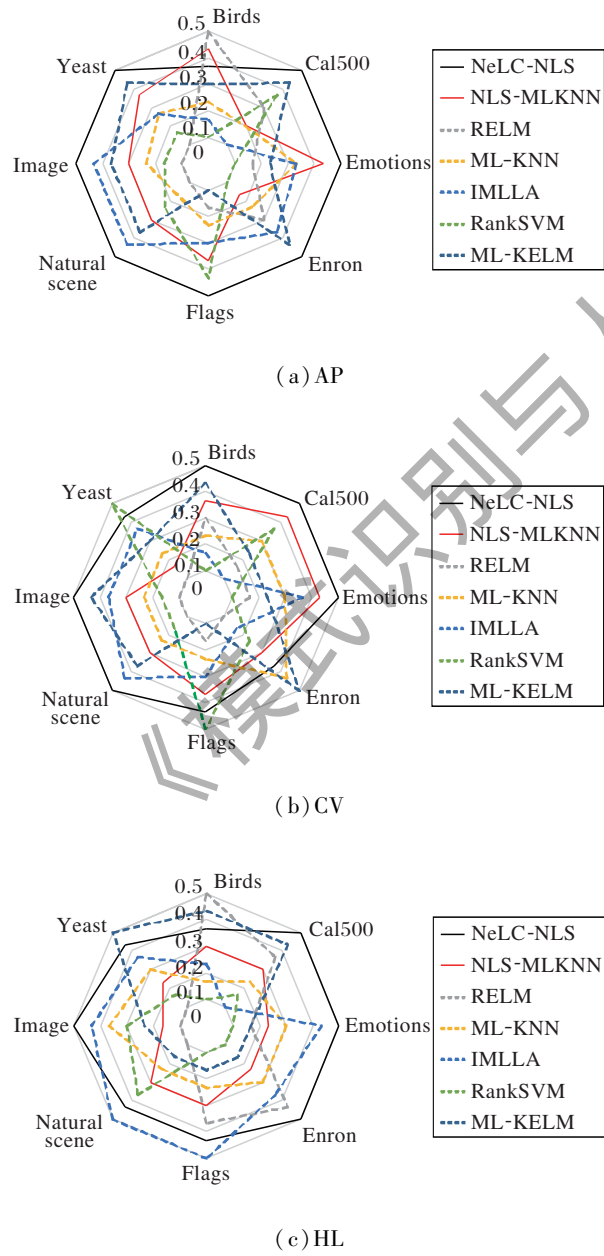


图 2 各评估指标下不同数据集上算法的稳定性
Fig.2 Stability of different algorithms on different datasets based on various evaluation indicators

4 结束语

在多标签分类学习中,有必要研究标签间相关关系,而在近邻标签空间中引入对标签相关性的研究可以提升分类器分类性能. 为了充分利用这一关系,本文引入非平衡参数,提出近邻标签空间的非平衡化标签补全算法 (NeLC-NLS). 结合非平衡化标签置信度矩阵与近邻标签空间,提升近邻标签空间的质量,统一在一个学习框架下. 实验表明,NeLC-NLS 优于一些常见的多标签学习算法.

今后将研究对于每类对应的近邻标签空间,采用不同的非平衡化置信度矩阵处理,提高多标签分类器的性能. 另外,如何将近邻标签空间与线性分类器统一成一个多标签学习框架,也是研究的重点.

参 考 文 献

[1] ZHANG M L, ZHOU Z H. Multi-label Learning // SAMMUT C, WEBB G I, eds. Encyclopedia of Machine Learning and Data Mining. Berlin, Germany: Springer, 2017: 875-881.
[2] BOUTELL M R, LUO J B, SHEN X P, et al. Learning Multi-label

- Scene Classification. *Pattern Recognition*, 2004, 37(9): 1757–1771.
- [3] 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.
- [4] ZHANG M L, ZHOU Z H. ML-KNN: A Lazy Learning Approach to Multi-label Learning. *Pattern Recognition*, 2007, 40(7): 2038–2048.
- [5] READ J, PFAHRINGER B, HOLMES G, *et al.* Classifier Chains for Multi-label Classification. *Machine Learning*, 2011, 85(3). DOI: <https://doi.org/10.1007/s10994-011-5256-5>.
- [6] FÜRNKRANZ J, HÜLLERMEIER E, MENCIA E L, *et al.* Multilabel Classification via Calibrated Label Ranking. *Machine Learning*, 2008, 73(2): 133–153.
- [7] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Random k -Labelsets for Multilabel Classification. *IEEE Transactions on Knowledge and Data Engineering*, 2011, 23(7): 1079–1089.
- [8] ELISSEEFF A, WESTON J. A Kernel Method for Multi-labelled Classification // DIETTERICH T G, BECHER S, GHARAMANI Z, eds. *Advances in Neural Information Processing Systems 14*. Cambridge, USA: The MIT Press, 2002: 681–687.
- [9] ZHANG Y, YEUNG D Y. Multilabel Relationship Learning. *ACM Transactions on Knowledge Discovery from Data*, 2013, 7(2). DOI: 10.1145/2499907.2499910.
- [10] SHANNON C E. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 1948, 27(3): 379–423.
- [11] 张振海, 李士宁, 李志刚. 一种基于相关信息熵的多标签分类算法. *西北工业大学学报*, 2012, 30(6): 968–973. (ZHANG Z H, LI S N, LI Z G. A Multi-label Classification Algorithm Using Correlation Information Entropy. *Journal of Northwestern Polytechnical University*, 2012, 30(6): 968–973.)
- [12] LEE J, KIM H, KIM N R, *et al.* An Approach for Multi-label Classification by Directed Acyclic Graph with Label Correlation Maximization. *Information Sciences*, 2016, 351: 101–114.
- [13] HUANG S J, YU Y, ZHOU Z H. Multi-label Hypothesis Reuse // *Proc of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA: ACM, 2012: 525–533.
- [14] 张敏灵. 一种新型多标记懒惰学习算法. *计算机研究与发展*, 2012, 49(11): 2271–2282. (ZHANG M L. An Improved Multi-label Lazy Learning Approach. *Journal of Computer Research and Development*, 2012, 49(11): 2271–2282.)
- [15] YOUNES Z, ABDALLAH F, DENOUEUX T. Multi-label Classification Algorithm Derived from K -Nearest Neighbor Rule with Label Dependencies [C/OL]. [2018-03-15]. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=7080359>.
- [16] GWEON H, SCHONLAU M, STEINER S. Nearest Labelset Using Double Distances for Multi-label Classification [C/OL]. [2018-03-15]. <https://arxiv.org/pdf/1702.04684.pdf>.
- [17] HUANG G, HUANG G B, SONG S J, *et al.* Trends in Extreme Learning Machines: A Review. *Neural Networks*, 2015, 61: 32–48.
- [18] HUANG G B. An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. *Cognitive Computation*, 2014, 6(3): 376–390.
- [19] DING S F, XU X Z, NIE R. Extreme Learning Machine and Its Applications. *Neural Computing and Applications*, 2014, 25(3/4): 549–556.
- [20] LIU Y, WEN K W, GAO Q X, *et al.* SVM Based Multi-label Learning with Missing Labels for Image Annotation. *Pattern Recognition*, 2018, 78: 307–317.
- [21] PIZZUTI C. A Multi-objective Genetic Algorithm for Community Detection in Networks // *Proc of the 21st IEEE International Conference on TOOLS with Artificial Intelligence*. Washington, USA: IEEE, 2009: 379–386.
- [22] 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.
- [23] 邓万宇, 郑庆华, 陈琳, 等. 神经网络极速学习方法研究. *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. *Chinese Journal of Computers*, 2010, 33(2): 279–287.)
- [24] 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.
- [25] 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.
- [26] DEMSAR J. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, 2006, 7: 1–30.
- [27] LIN Y J, LI Y W, WANG C X, *et al.* Attribute Reduction for Multi-label Learning with Fuzzy Rough Set. *Knowledge-Based Systems*, 2018, 152: 51–61.

作者简介



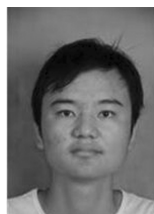
程玉胜(通讯作者), 博士, 教授, 主要研究方向为大数据、粗糙集、特征选择的机器学习. E-mail: chengyushaq@163.com.

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



赵大卫, 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: like854@qq.com.

(ZHAO Dawei, master student. His research interests include machine learning, big data and statistics.)



钱坤, 硕士研究生, 主要研究方向为多标签学习、机器学习、大数据、数据统计. E-mail: 2991359073@qq.com.

(QIAN Kun, master student. His research interests include multi-label learning, machine learning, big data and statistics.)