

## 基于标记增强和模糊辨识度的标记分布特征选择

熊传镇<sup>1</sup>, 钱文彬<sup>1,2</sup>, 王映龙<sup>1</sup>

(1. 江西农业大学计算机与信息工程学院, 南昌 330045; 2. 江西农业大学软件学院, 南昌 330045)

**摘要:** 特征选择作为多标记学习任务中关键预处理步骤, 能够有效地解决高维多标记数据存在的维度灾难问题。在现有大部分的多标记学习中, 标记是以逻辑分布的形式刻画, 即示例中相关标记的重要性相同; 然而, 在许多现实生活中, 每个示例的标记重要程度呈现差异性。本文提出了一种基于模糊相似性的标记增强算法, 通过衡量示例中标记的模糊相关性, 将传统的多标记数据转换为标记分布数据; 分析了标记分布数据中在标记上的标记差异性和在特征上的模糊相对辨识关系, 给出了在标记空间和特征空间上的模糊辨识度, 并构造了衡量特征辨识能力的特征重要度; 在此基础上, 构建面向标记分布数据的特征选择算法, 能获得按特征重要度降序的特征选择结果。最后通过在多个多标记数据集上实验对比和分析, 进一步验证了算法的有效性和可行性。

**关键词:** 特征选择; 粒计算; 粗糙集; 标记分布; 模糊辨识度

**中图分类号:** TP391      **文献标志码:** A

## Label Enhancement and Fuzzy Discernibility Based Label Distribution Feature Selection

XIONG Chuanzhen<sup>1</sup>, QIAN Wenbin<sup>1,2</sup>, WANG Yinglong<sup>1</sup>

(1. School of Computer and Information Engineering, Jiangxi Agricultural University, Nanchang 330045, China; 2. School of Software, Jiangxi Agricultural University, Nanchang 330045, China)

**Abstract:** Feature selection is the key pre-processing step of multi-label learning tasks. It can efficiently solve the problem of the “curse of dimensionality”, which is existed in the high-dimensional multi-label data. In multi-label learning, the label is described as the form of logical distribution, in which the importance of each label associated with the instance is equivalent. However, the label importance of each label is usually different in many fields. For this issue, a label enhancement algorithm is proposed in this paper. By evaluating the fuzzy similarity relation on labels among instances, it transforms the multi-label data to the label distribution data. The discernibility relation on labels and the fuzzy relative discernibility relation on features are analyzed in details for label distribution data, then the fuzzy discernibility on the label space and the feature space is defined, and the significance of feature is constructed to assess the discernibility ability of the feature. On this basis, a feature selection algorithm is proposed for label distribution data, which can obtain the result of feature selection in descending order of feature significance. Finally, the experimental results show that the proposed algorithm is effective and feasible on several multi-label datasets.

**基金项目:** 国家自然科学基金(61966016)资助项目; 江西省自然科学基金(20192BAB207018)资助项目; 江西教育厅科学技术研究基金(GJJ180200)资助项目。

**收稿日期:** 2020-04-12; **修订日期:** 2020-10-10

**Key words:** feature selection; granular computing; rough set; label distribution; fuzzy discernibility

## 引言

在传统的监督学习框架中,训练数据中的示例是由单个标记所描述。这种监督学习被称作单标记学习。然而,在现实应用领域中,许多数据往往同时含有多个类别标记,即每个示例可能存在多个语义与之相关联。例如,一个人在某个时刻的表情可能包含多种复杂情绪(快乐、悲伤和愤怒等);一幅美丽的风景画可能描绘了多重的景色(天空、海水和沙滩等);一本杂志中可能包含着丰富多样的内容(时尚、运动和科技等)。由于数据的语义多义性,使得传统的单标记学习方法具有局限性。为此,多标记学习作为机器学习的重要范式<sup>[1]</sup>,它能有效处理语义丰富的多标记数据。近年来,多标记学习已经成为了文本分类<sup>[2]</sup>、图像识别<sup>[3]</sup>等多种应用领域中的研究热点。对于多标记数据而言,示例的特征往往呈现高维性特点。在多标记数据的高维特征中,可能存在某些不相关或者冗余的特征,这些特征使得模型的复杂度和计算的时间消耗显著提升,从而影响多标记学习的分类性能和泛化能力。为了有效地解决高维多标记数据存在的维度灾难问题,可在学习任务前采用特征选择方法来分析高维特征中的无关或冗余特征,可缩短学习模型的计算时间,并使模型更具有可解释性和可扩展性。在多标记分类中,特征选择已成为数据预处理的重要步骤。

粗糙集理论作为代表性的粒计算模型<sup>[4]</sup>,一直以来在特征选择任务中受到广泛关注和研究<sup>[5-6]</sup>,其特点是无需提供数据本身以外的任何先验知识,可直接从数据中剔除大量冗余或不相关的特征,对知识的描述和分析也更客观。由于经典的粗糙集只能处理名义型数据,难于直接处理连续型数据。为此,Dubois和Prade提出了模糊粗糙集用以处理混合型数据问题<sup>[7]</sup>。与粗糙集中的逻辑等价关系不同,模糊粗糙集使用模糊相似关系进行信息粒度的划分,无需对数据离散化,可直接描述连续型数据中示例的相似性。由于模糊粗糙集模型的灵活性和普适性,在知识发现领域中取得了许多有意义的研究成果<sup>[8-10]</sup>。

对于多标记学习而言,假定数据中示例的相关标记之间的重要度都是相等,标记是以逻辑值的形式描述;然而,在许多应用领域中,每个示例含有的标记的重要程度往往不同。为解决上述问题,耿新等提出了标记分布学习的机器学习范式<sup>[11-12]</sup>,能有效地解决重要程度不同的标记多义性问题。相较于传统的逻辑值标记,标记分布数据中包含更丰富的分类信息。但由于现实世界中往往难以直接获取标记的分布信息,需利用标记增强方法把多标记数据的逻辑标记转化为标记分布数据。为此,本文提出了一种基于模糊相似性的标记增强算法,将传统的多标记数据转化为标记分布数据;分析了标记分布数据中标记的差异性和特征的模糊相对辨识关系,给出了在标记空间和特征空间上的模糊辨识度,并构造了衡量特征关于标记分布数据辨识能力的特征辨识重要度;在此基础上,设计了面向标记分布数据的特征选择算法。在多个Mulan数据集上的实验分析表明,经过标记增强后的特征选择算法能够有效地提升分类效果。

通过对多标记高维数据进行特征选择,可以减少算法的训练时间和增强学习算法的分类性能。近年来,多标记特征选择方面已取得许多有意义的研究成果。Spolaor等将ReliefF算法应用于多标记学习中,提出了Relief-ML特征选择算法<sup>[13]</sup>。Zhang等通过利用最大化特征和相关类之间的依赖关系,构造了MDDM算法用以实现特征选择<sup>[14]</sup>。此外,Lee等通过所选特征与标记间的互信息设计了PMU特征选择算法<sup>[15]</sup>。Xu等基于排序损失提出了标记对比转换方法,并与卡方统计结合设计了特征选择方法<sup>[16]</sup>。Teisseyre基于马尔科夫网络构建了标记与特征之间的依赖关系来获得特征子集<sup>[17]</sup>。近年来,

模糊粗糙集作为粒计算理论中的重要工具,在多标记学习中也受到了越来越多研究人员的关注。例如,为解决 LIFT 算法中特征冗余的问题,Xu 等将 LIFT 算法与模糊粗糙集结合,提出了 FRS-LIFT 特征选择算法<sup>[18]</sup>。同时,Lin 等引入模糊互信息用以评估多标记数据中特征的重要性,并设计了一种考虑标记相关性的多标记特征选择算法<sup>[19]</sup>。可知,模糊粗糙集是一种有效的多标记特征选择模型。

上述的多标记特征选择的研究中,对于多标记数据采取的都是均匀分布假设,即每个示例含有的标记重要程度相同。事实上,对于现实生活中的每个示例来说,不同的标记对于示例的描述程度是不一样的。为此,耿新等<sup>[11]</sup>提出了标记分布学习的概念,主要研究现实世界的多标记学习中更深一层的学习问题“每个标记如何描述该示例?”,即每个示例中相关标记的相对重要度该如何描述。在文献<sup>[12]</sup>中,Xu 等归纳总结了多种标记增强算法,将多标记数据中的逻辑标记转换为由实值表达的标记分布数据。相较于传统的逻辑标记,标记分布能够在多标记数据中挖掘出更多隐藏的监督信息,能够有效地提升多标记特征选择中的分类效果。作为一种新的机器学习学习范式,标记分布学习已在各个领域受到越来越多的关注。在文献<sup>[20]</sup>中,He 等提出了一种完全数据驱动的标记分布学习方法来自适应地学习潜在的标记分布,该方法能够根据人脸示例的局部上下文结构,发现跨年龄相关分析中的内在年龄分布模式。与此同时,Liu 等采用结构化的低秩逆协方差方法来预测视觉情感的标记分布<sup>[21]</sup>。此外,Zhao 等提出了一种基于最优传输理论的标记分布学习和标记相关性挖掘的方法<sup>[22]</sup>,并基于 Sinkhorn 距离提出了标记分布学习的数据依赖风险的边界分析。这些研究表明,标记分布能够挖掘出更加深层的潜在信息,能更好描述现实数据中的相对关系。

## 1 基础知识

### 1.1 多标记学习与标记增强

为研究标记分布下的特征选择算法,首先对多标记学习和标记分布学习框架进行简要介绍。令  $X = \mathbf{R}^d = \{x_1, x_2, \dots, x_n\}$  为数据规模为  $n$  的  $d$  维示例空间,每个示例表示为  $x_i = \{x_i^1, x_i^2, \dots, x_i^d\}$ ,  $L = \{l_1, l_2, \dots, l_q\}$  表示  $q$  个候选标记的集合。定义多标记训练集为:  $T = \{(x_i, Y_i) | x_i \in X, Y_i \subseteq L\}$ , 其中  $Y_i$  表示示例  $x_i$  的相关标记的集合。若  $l_k \in Y_i$ , 则  $l_k^i = 1$ ; 否则  $l_k^i = 0$ 。多标记学习的目的是在训练集上寻找一个评价函数  $f: X \times L \rightarrow R$  用来表示示例和标记之间的关系,  $f(x_i, l_j)$  的值越大, 则示例  $x_i$  与标记  $l_j$  的关联性越强。结合阈值函数  $t(x_i)$ , 评价函数可转化为多标记分类器  $h(\cdot): h(x_i) = \{l_k | f(x_i, l_k) > t(x_i)\}$ 。

作为多标记学习的拓展,标记分布学习可挖掘出更多示例和标记之间的关联信息。不同于多标记数据中的逻辑值标记,在标记分布学习中,通常使用实值表示标记对于示例的相对重要性。在标记分布数据中,示例  $x_i$  在标记集  $L$  上的标记分布表示为:  $D_i = \{d_{x_i}^1, d_{x_i}^2, \dots, d_{x_i}^q\}$ , 其中  $d_{x_i}^j$  表示标记  $l_j$  对示例  $x_i$  的描述程度。对于任意示例  $x_i$ , 其标记分布满足一般性:  $0 \leq d_{x_i}^j \leq 1$ , 并且是一个完备集:  $\sum_{1 \leq j \leq q} d_{x_i}^j = 1$ 。

在标记分布学习中,通过对标记的更加丰富的描述,不仅能够最大限度地描述数据和标记之间的关联性,而且能够挖掘出更多的潜在信息。由于难以直接获取标记的分布信息,这些标记分布主要是在多标记数据集上通过标记增强获得。对于多标记数据集  $T = \{(x_i, Y_i) | 1 \leq i \leq n\}$ , 标记增强的目标是通过标记之间的相关性将原先的逻辑标记  $Y_i$  转化为标记分布  $D_i$ , 获得标记分布数据  $S = \{(x_i, D_i) | 1 \leq i \leq n\}$ 。

### 1.2 模糊粗糙集与模糊辨识矩阵

在传统的粗糙集理论中,主要是通过等价关系  $\text{IND}(B) = \{(x, y) | \forall a \in B, a(x) = a(y)\}$  将论域划分为多个等价类  $[x]_B = \{y | (x, y) \in \text{IND}(B)\}$ 。然而,在许多现实应用中,数据往往是数值型或混合型

的。使得等价关系  $\text{IND}(B)$  的处理能力存在一定局限性。为此,模糊粗糙集可有效处理此类数据。

**定义 1** 定义决策表为  $\langle U, C \cup D \rangle$ ,  $U$  表示数据集的论域,  $C$  和  $D$  分别为条件特征集合和决策标记。设  $A$  为论域上的一个模糊集,  $A(x)$  表示示例  $x$  对模糊集  $A$  的隶属度, 值域为  $[0, 1]$ 。  $A(x)$  的值越大, 示例  $x$  属于  $A$  的可能性越大。当  $A(x) = 1$  时, 示例  $x$  完全属于模糊集  $A$ 。

结合模糊隶属度, 对于任意特征  $a \in C$ , 可定义其对应模糊二元关系  $R_a$ 。如果模糊二元关系  $R_a$  满足下列条件, 可称之为模糊相似关系<sup>[8]</sup>: (1) 自反性:  $R_a(x, x) = 1$ ; (2) 对称性:  $R_a(x, y) = R_a(y, x)$ 。

对于任意特征子集  $B \subseteq C$ , 其模糊二元关系为  $R_B = \bigcap_{a \in B} R_a$ 。若满足  $\forall a \in B, R_a$  为模糊相似关系, 则  $R_B$  也是模糊相似关系。

基于模糊相似关系  $R_B$ , 可以将论域划分为多个模糊集:  $U/R_B = \{[x_i]_B\}_{i=1}^n$ , 每个模糊集可表示为:

$$[x_i]_B = \frac{r_{i_1}}{x_1} + \frac{r_{i_2}}{x_2} + \dots + \frac{r_{i_n}}{x_n}, r_{i_j} = R_B(x_i, x_j) \text{ 为示例 } x_i \text{ 和 } x_j \text{ 在模糊相似关系 } R_B \text{ 下的模糊相似度。}$$

**定义 2** 给定非空论域  $U$ , 条件特征集合  $C$ , 决策标记  $D$ 。设  $U/D = \{D_1, D_2, \dots, D_q\}$  为论域的决策划分。对于每个决策类  $D_i$ , 其模糊上近似与模糊下近似定义为<sup>[8]</sup>

$$\begin{aligned} \underline{R}_B D_i(x) &= \inf_{u \in U} \max \{1 - R_B(x, u), D_i(u)\} \\ \overline{R}_B D_i(x) &= \sup_{u \in U} \min \{R_B(x, u), D_i(u)\} \end{aligned} \quad (1)$$

式中:  $D_i(u)$  表示示例  $u$  对决策类  $D_i$  的隶属度, 一般而言,  $\forall u \in U$ , 存在

$$D_i(u) = \begin{cases} 1 & u \in D_i \\ 0 & u \notin D_i \end{cases} \quad (2)$$

因此, 可将模糊上下近似简化为

$$\begin{aligned} \underline{R}_B D_i(x) &= \inf_{u \notin D_i} \{1 - R_B(x, u)\} \\ \overline{R}_B D_i(x) &= \sup_{u \in D_i} \{R_B(x, u)\} \end{aligned} \quad (3)$$

对于任意的决策类  $D_i$  而言, 其模糊粗糙集  $\langle \underline{R}_B D_i(x), \overline{R}_B D_i(x) \rangle$  从肯定和可能两个不同的视角描述了示例  $x$  对于决策类  $D_i$  的隶属度。相对于经典粗糙集, 模糊粗糙集的描述更加具体。

作为经典的模糊粗糙集模型, 模糊辨识矩阵也有着广泛的应用。相对于传统的辨识矩阵, 模糊辨识矩阵能够更加具体地描述数值型示例之间的差异性。

**定义 3** 设决策表为  $\langle U, C \cup D \rangle$ ,  $C$  为条件特征的集合。给定模糊辨识矩阵  $M = \{M(x, y)\}$ , 其大小为  $|U| \times |U|$ 。在此辨识矩阵中, 每个项集  $M(x, y)$  都表示各个特征对于示例对  $(x, y)$  的辨识度的集合, 具体定义为<sup>[23]</sup>

$$M(x, y) = \{a_{N(R_a(x, y))} | a \in C\} \quad \forall x, y \in U \quad (4)$$

式中:  $R_a(x, y)$  为示例对  $(x, y)$  关于特征  $a$  的模糊相似度, 也记作  $\text{sim}_a(x, y)$ ;  $N(R_a(x, y)) = 1 - R_a(x, y)$  表示示例对  $(x, y)$  关于特征  $a$  的模糊辨识度  $\text{dis}_a(x, y)$ 。

## 2 模糊标记增强和相对模糊辨识关系

### 2.1 基于模糊相似性的标记增强

为了将原始的逻辑标记空间转化成标记分布空间, 需利用标记增强方法对原始的标记集合进行转换。为此, 本文引入了不同示例间关于标记的模糊相似性来刻画原逻辑标记的模糊性。

**定义 4** 设多标记决策表为  $\langle U, C \cup L \rangle$ ,  $U$  为包含  $n$  个示例的非空有限论域,  $C$  为条件特征集

合,标记的集合记为  $L = \{l_1, l_2, \dots, l_q\}$ 。  $\forall x_i, x_j \in U$ , 其模糊标记相似性定义为

$$\text{FLS}(x_i, x_j) = \frac{\sum_{k=1}^q (l_k^i \cdot l_k^j)}{\sqrt{|x_i^L| \cdot |x_j^L|}} \quad (5)$$

式中  $x_i^L = \{l_1^i, l_2^i, \dots, l_q^i\}$  表示示例  $x_i$  在标记  $L$  上的标记值,  $|x_i^L| = \sum_{k=1}^q l_k^i$ 。

这样示例  $x_i$  与论域中所有示例的模糊标记相似性可以用模糊标记相似性集合  $\text{FLS}(x_i)$  表示为

$$\text{FLS}(x_i) = [\text{FLS}(x_i, x_1), \text{FLS}(x_i, x_2), \dots, \text{FLS}(x_i, x_n)] \quad (6)$$

然后,构建一个模糊标记关联矩阵  $\text{FLA}$ , 其大小为  $|U| \times |L|$ 。矩阵中的项集  $\text{FLA}(i, k)$  表示示例  $x_i$  与标记  $l_k$  的模糊关联强度, 其定义为

$$\text{FLA}(i, k) = \text{FLS}(x_i) \times L(k) = \sum_{j=1}^n \text{FLS}(x_i, x_j) \cdot l_k^j \quad (7)$$

式中:  $L(k)$  表示标记  $l_k$  的原始逻辑分布。将矩阵中每个示例的关联度归一化, 即可得到标准化的模糊标记关联度矩阵  $\text{SFLA}$ , 其描述为

$$\text{SFLA}(i, k) = \frac{\text{FLA}(i, k)}{|\text{FLA}(i)|} = \frac{\text{FLA}(i, k)}{\sum_{j=1}^q \text{FLA}(i, j)} \quad (8)$$

在关联度矩阵中, 重复利用了示例关于标记的模糊相似性, 将原始的逻辑标记上产生更加丰富的信息。在此基础上, 将这些模糊性与原始的逻辑标记相结合, 进而对示例和标记之间的关联度进行了具体的描述。然而, 由于多标记数据之间的不平衡性, 可能存在着原始逻辑标记中的无关标记的关联度失衡的现象, 需要对关联度矩阵  $\text{SFLA}$  进行合理规约。

令示例为  $x_i$  的原始逻辑标记中无关标记所占的比例为  $\text{NALR}(i) = \frac{|\{l_k | l_k^i = 0\}|}{q}$ , 无关标记的关联度之和为  $\text{NAS}(i) = \sum_{l_k^i=0} \text{SFLA}(i, j)$ 。假如示例  $x_i$  满足  $\text{NAS}(i) > \text{NALR}(i)$ , 则对其关联度执行如下规约计算

$$\text{SFLA}^*(i, k) = \begin{cases} \text{SFLA}(i, k) \times \text{NALR}(i) & l_k^i = 0 \\ \text{SFLA}(i, k) / \text{AS}(i) \times (1 - \text{NAS}(i) \times \text{NALR}(i)) & l_k^i = 1 \end{cases} \quad (9)$$

式中:  $\text{AS}(i) = 1 - \text{NAS}(i)$  表示相关标记的关联度之和,  $\text{SFLA}^*$  表示规约后的标准化的模糊标记关联度矩阵。

在此基础上, 为了消除关联度较小的标记影响, 需要将原有的关联度进行简化, 以提高算法的计算效率, 具体约简方法为

$$\text{RSFLA}(i, k) = \begin{cases} \text{SFLA}^*(i, k) & \text{SFLA}^*(i, k) / \max\{\text{SFLA}^*(i)\} \geq 0.25 \\ 0 & \text{其他} \end{cases} \quad (10)$$

式中  $\text{RSFLA}$  为简化模糊标记关联度矩阵。

最后, 将经过约简的关联度归一化, 即可获得示例  $x_i$  关于标记  $l_k$  的标记分布

$$d_{x_i}^k = \frac{\text{RSFLA}(i, k)}{|\text{RSFLA}(i)|} = \frac{\text{RSFLA}(i, k)}{\sum_{j=1}^q \text{RSFLA}(i, j)} \quad (11)$$

经过标记增强后, 可获得多标记数据中示例  $x_i$  的标记分布  $D_i = \{d_{x_i}^1, d_{x_i}^2, \dots, d_{x_i}^q\}$ 。

## 2.2 模糊辨识相对关系和特征评价指标

不同于传统辨识矩阵,在模糊辨识矩阵中,通常使用辨识度从不同的角度来衡量特征的重要性,如最大可辨识对<sup>[23]</sup>、最小元素<sup>[24]</sup>等方法。这些方法的核心思想都是衡量特征在不同示例对中的辨识能力。为此,本文提出了一种泛化的模糊辨识模型,通过计算特征关于每个示例对的相对辨识能力(模糊相对辨识度FRD),获得特征在论域上的整体辨识能力。

**定义5** 设多标记决策表 $\langle U, C \cup L \rangle$ ,  $U$ 为非空有限论域,  $C$ 为条件特征的集合。设特征 $a \in C$ 关于示例对 $(x, y)$ 的模糊辨识度为 $\text{dis}_a(x, y)$ ,则模糊相对辨识度 $\text{FRD}_a(x, y)$ 表示为

$$\text{FRD}_a(x, y) = \frac{\text{dis}_a(x, y) - \min_{c \in C} \text{dis}_c(x, y)}{\max_{c \in C} \text{dis}_c(x, y) - \min_{c \in C} \text{dis}_c(x, y)} \quad (12)$$

可知,模糊相对辨识度 $\text{FRD}_a(x, y)$ 满足对称性,且随模糊辨识度 $\text{dis}_a(x, y)$ 单调递增,其值域为 $[0, 1]$ 。给定阈值 $\lambda$ ,可以定义多标记数据中示例对上具有强辨识能力的特征集合,记为相对辨识特征 $\text{RDF}^\lambda(x, y)$ ,其描述如下

$$\text{RDF}^\lambda(x, y) = \{a | \text{FRD}_a(x, y) > \lambda\} \quad (13)$$

**性质1** 多标记决策表 $\langle U, C \cup L \rangle$ ,对于任意的示例对 $(x, y) \in U \times U$ ,其相对辨识特征 $\text{RDF}^\lambda(x, y)$ 满足:(1)  $\text{RDF}^\lambda(x, y) = \text{RDF}^\lambda(y, x)$ ;(2)  $0 \leq \lambda_1 \leq \lambda_2 \leq 1$ ,  $\text{RDF}^{\lambda_1}(x, y) \supseteq \text{RDF}^{\lambda_2}(x, y)$ 。

**证明:** (1)由模糊相对辨识度的对称性可知。

(2)由定义5可知,设阈值为 $0 \leq \lambda_1 \leq \lambda_2 \leq 1$ ,对于相对辨识特征 $\text{RDF}^{\lambda_2}(x, y)$ 中的任意特征 $a$ ,其模糊相对辨识度 $\text{FRD}_a(x, y) > \lambda_2$ ,其一定也满足模糊相对辨识度 $\text{FRD}_a(x, y) > \lambda_1$ ,可得 $a \in \text{RDF}^{\lambda_1}(x, y)$ ;反之,则不成立。故可得出 $\text{RDF}^{\lambda_1}(x, y) \supseteq \text{RDF}^{\lambda_2}(x, y)$ 。

对于任意特征 $a \in C$ ,统计其所在的相对辨识特征集合,可定义相对辨识关系

$$\text{RDP}^\lambda(a, U) = \{(x, y) | a \in \text{RDF}^\lambda(x, y)\} \quad (14)$$

为简化计算,式(14)可表示为

$$\text{RDP}^\lambda(a, U^*) = \{(x, y) | x < y, a \in \text{RDF}^\lambda(x, y)\} \quad (15)$$

对于特征 $a \in C$ ,对应的相对辨识矩阵 $\text{RDM}_a^\lambda$ ,其大小为 $|U| \times |U|$ 。矩阵中的每个项集 $\text{RDM}_a^\lambda(x, y)$ 表示在相对辨识关系 $\text{RDP}^\lambda(a, U^*)$ 下的模糊相对辨识度 $\text{FRD}_a(x, y)$ ,其描述如下

$$\text{RDM}_a^\lambda(x, y) = \begin{cases} \text{FRD}_a(x, y) & (x, y) \in \text{RDP}^\lambda(a, U^*) \\ 0 & \text{其他} \end{cases} \quad (16)$$

通过矩阵 $\text{RDM}_a^\lambda$ ,可以衡量特征 $a$ 对于论域 $U$ 上的示例对的相对辨识能力。辨识能力依靠的是每个示例对的模糊相对辨识度,故减少了数据不平衡性带来的影响,从而能够合理刻画特征对于不同示例对的辨识能力。

然而,由于在标记分布数据中不同的示例对之间的标记差异性往往不同,为此需要在标记分布中定义标记的差异性( $\text{dis}_L$ ),用于衡量不同示例对之间在标记空间上的差异性,定义如下。

**定义6** 设 $U = \{x_1, x_2, \dots, x_n\}$ 为示例空间,  $L = \{l_1, l_2, \dots, l_q\}$ 为标记空间的集合。对于任意示例对 $(x, y) \in U \times U$ ,其标记差异性 $\text{dis}_L(x, y)$ 定义为

$$\text{dis}_L(x, y) = 1 - \frac{\sum_{k=1}^q d_x^k \cdot d_y^k}{\sqrt{\sum_{i=1}^q (d_x^i)^2} \sqrt{\sum_{j=1}^q (d_y^j)^2}} \quad (17)$$

式中:  $d_x^k$  为示例  $x$  在标记  $l_k$  上的标记分布。对于任意示例对  $(x, y)$ , 其标记差异性  $\text{dis}_L(x, y)$  的值域为  $[0, 1]$ , 且满足对称性。事实上, 相对于原始的逻辑标记  $l_k \in \{0, 1\}$ , 标记分布  $d_x^k \in [0, 1]$  更加具有普遍性。通过标记分布, 能够对样本和标记之间的隶属关系进行更加详细地描述, 并且能够通过对应的标记差异性  $\text{dis}_L$  挖掘出更加丰富的区分关系和语义信息。

在论域  $U$  上, 给定标记差异矩阵  $\text{LDM}$ , 其大小为  $|U| \times |U|$ , 其描述如下

$$\text{LDM} = \begin{bmatrix} \text{dis}_L(x_1, x_1) & \text{dis}_L(x_1, x_2) & \cdots & \text{dis}_L(x_1, x_n) \\ \text{dis}_L(x_2, x_1) & \text{dis}_L(x_2, x_2) & \cdots & \text{dis}_L(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{dis}_L(x_n, x_1) & \text{dis}_L(x_n, x_2) & \cdots & \text{dis}_L(x_n, x_n) \end{bmatrix} \quad (18)$$

通过标记差异性矩阵  $\text{LDM}$ , 可以得到整个标记空间上的示例对之间的差异性。结合示例空间关于特征  $a$  的相对辨识度和在标记空间上的标记差异性, 可以衡量特征  $a$  对于不同标记差异性的示例对之间的相对辨识能力, 记作  $\text{RDA}^\lambda(a)$ 。

**定义 7** 给定非空论域  $U$ , 条件特征集合  $C$ 。对于任意特征  $a \in C$ , 其特征相对辨识能力  $\text{RDA}^\lambda(a)$  为

$$\text{RDA}^\lambda(a) = \text{RDM}_a^\lambda \times \text{LDM} = \sum_{x, y \in U, x < y} \text{RDM}_a^\lambda(x, y) \times \text{LDM}(x, y) \quad (19)$$

**性质 2** 设特征集合为  $C$ 。对于任意特征  $a \in C$ , 其相对辨识能力  $\text{RDA}^\lambda(a)$  满足下列性质: (1)  $0 \leq \text{RDA}^\lambda(a) \leq \frac{1}{2}|U|^2$ ; (2)  $0 \leq \lambda_1 \leq \lambda_2 \leq 1, \text{RDA}^{\lambda_1}(a) \geq \text{RDA}^{\lambda_2}(a)$ 。

**证明:** (1) 由定义 7 可知, 对于任意示例对  $(x, y)$ , 其满足  $\text{RDM}_a^\lambda(x, y) \leq 1, \text{LDM}(x, y) \leq 1$ , 同时已知相对辨识矩阵非零项目的数量小于  $\frac{1}{2}|U|^2$ , 由此可证得相对辨识能力  $\text{RDA}^\lambda(a) \leq \frac{1}{2}|U|^2$ 。

(2) 由性质 1 可得, 若  $0 \leq \lambda_1 \leq \lambda_2 \leq 1$ , 则相对辨识特征集合  $\text{RDF}^{\lambda_1}(x, y) \supseteq \text{RDF}^{\lambda_2}(x, y)$ , 由此可得相对辨识关系满足  $\text{RDP}^{\lambda_1}(a, U^*) \supseteq \text{RDP}^{\lambda_2}(a, U^*)$ , 最终可得相对辨识能力满足  $\text{RDA}^{\lambda_1}(a) > \text{RDA}^{\lambda_2}(a)$ 。

基于上述公式, 可以通过特征的相对模糊辨识能力去衡量特征的辨识重要度, 其定义如下

$$\text{DS}^\lambda(a) = 2 \frac{\text{RDA}^\lambda(a)}{|U|^2} \quad (20)$$

**性质 3** 设特征集合为  $C$ 。对于任何特征  $a \in C$ , 其辨识重要度  $\text{DS}^\lambda(a)$  满足如下性质: (1)  $0 \leq \text{DS}^\lambda(a) \leq 1$ ; (2)  $0 \leq \lambda_1 \leq \lambda_2 \leq 1, \text{DS}^{\lambda_1}(a) > \text{DS}^{\lambda_2}(a)$ 。

对于标记分布数据而言, 特征的辨识重要度  $\text{DS}^\lambda(a)$  能够适应标记分布学习中的实值标记, 并在特征空间和标记空间上综合考虑特征的辨识能力。对于特征  $a$  而言, 其辨识重要度  $\text{DS}^\lambda(a)$  的值越大, 其在特征集合中的重要程度越高。

### 3 算法描述

基于上述方法, 可通过基于模糊相关性的标记增强方法将传统多标记数据转化为标记分布数据; 在此基础上, 结合模糊相对辨识关系和标记之间的差异性, 可计算特征在标记分布数据上的辨识重要度; 最后, 构造了一种基于标记增强和模糊辨识关系的标记分布特征选择算法 (Label enhancement and fuzzy discernibility based label distribution feature selection algorithm, LDFS), 算法根据辨识重要度排序, 可获得具有强辨识能力的特征选择结果。算法的具体描述如下:

**算法** 基于标记增强和模糊辨识度的标记分布特征选择算法。

输入:多标记决策表  $\langle U, C \cup L \rangle$ , 阈值  $\lambda$ 。

输出:特征排序集合  $S$ 。

(1) 对多标记数据进行标记增强

①  $\forall x_i \in U$ , 计算其模糊标记相似性集合  $FLS(x_i)$ ;

② 构建标准化的模糊标记关联强度矩阵  $SFLA$ ;

③ 对关联度矩阵  $SFLA$  进行规约和约简, 获得每个示例  $x_i$  关于标记  $l_k \in L$  的标记分布  $d_{x_i}^k$ ;

(2) 构建标记差异性矩阵  $LDM$ ;

(3) 对于  $\forall a \in C$ , 计算其关于示例对模糊相对辨识度, 执行如下操作:

①  $\forall (x, y) \in U \times U$ , 计算其在特征  $a \in C$  的模糊辨识度  $dis_a(x, y)$ ;

② 将每个示例对  $(x, y)$  中的模糊辨识度归一化得模糊相似度  $FRD(x, y)$ ;

(4) 根据阈值  $\lambda$ , 计算每个特征  $a \in C$  的相对辨识关系  $RDP^{\lambda}(a, U^*)$ , 并构建其相对辨识矩阵  $RDM_a^{\lambda}$ ;

(5) 计算每个特征  $a \in C$ , 计算其相对模糊辨识能力  $RDA^{\lambda}(a)$ ;

(6) 计算每个特征  $a \in C$  的辨识重要度  $DS^{\lambda}(a)$ ;

(7) 将特征按照辨识重要度降序排列;

(8) 输出特征排序集合  $S$ 。

(9) 结束

步骤1主要是进行对多标记数据进行标记增强, 首先计算示例模糊相似性向量, 其时间复杂度为  $O(|U|^2)$ , 在此基础上构建关联度矩阵  $SFLA$ , 时间复杂度为  $O(|U||L|)$ , 在进行规约和约简后, 可获得了示例的标记分布信息, 时间复杂度为  $O(|U||L|)$ 。在步骤2中, 构建标记分布数据的标记差异性矩阵  $LDM$ , 时间复杂度为  $O(|U|^2)$ 。在步骤3与步骤4中, 计算了每个特征的模糊相对辨识度和相对辨识矩阵, 时间复杂度为  $O(|C||U|^2)$ 。然后, 在步骤5中运用矩阵运算计算每个特征的相对模糊辨识能力  $RDA^{\lambda}(a)$ , 时间复杂度为  $O(|C||U|^2)$ 。最后, 在步骤7和步骤8中获得了每个特征的辨识重要度, 按重要度对特征进行排序, 时间复杂度为  $O(|C|)$ 。综上所述, 算法的时间复杂度为  $O(|C||U|^2)$ 。

## 4 实验与结果分析

### 4.1 实验数据集

为了验证算法的可行性与有效性, 在Mulan数据集中选取了8个数据集, 这些多标记数据来自不同的应用领域, 其数据描述如表1所示。表中的数据都为数值型数据或混合型数据。其中, Types中的Continuous表示数据集中的数据皆为连续型数据; Both表示数据集中的数据为离散或连续的混合型数据。为了实验的规范性, 对连续型数据进行归一化处理, 使其值域在范围 $[0, 1]$ 中, 离散型数据则无需处理。归一化公式定义如下

$$x_i^a = \frac{x_i^a - \min_{m=1}^n \{x_m^a\}}{\max_{m=1}^n \{x_m^a\} - \min_{m=1}^n \{x_m^a\}} \quad (21)$$

式中:  $x_i^a$  表示样本  $x_i$  在连续型特征  $a$  上的值;  $n$  表

表1 实验数据集描述

Table 1 Description of experimental datasets

Dataset	Instance	Feature	Label	Type
Birds	645	260	19	Both
Emotions	593	72	8	Continuous
Flags	194	19	7	Both
Gnegative	1 392	440	8	Continuous
Gpositive	519	440	4	Continuous
Plant	978	440	12	Continuous
Virus	207	440	6	Continuous
Yeast	2417	103	14	Continuous



示论域的大小。

算法运行的硬件配置为 Inter(R) Core(TM) i5-6500 CPU@ 3.20 GHz 处理器,8 GB 内存,操作系统为 Windows 10,编程环境为 Python 3.7。

## 4.2 评价指标

由于多标记学习标注结构的复杂性,本文选取了5个通用的多标记评价指标来分析算法的分类性能,包括汉明损失(Hamming loss)、1-错误率(One-error)、平均准确率(Average precision)、覆盖率(Coverage)及排序损失(Ranking loss)。通过这些评价指标,可以从多个不同的角度来评价算法的分类性能。

在多标记训练集为  $T = \{(x_i, Y_i) | 1 \leq i \leq n\}$ ,  $Y_i$  表示与示例  $x_i$  相关联的真实标记集合。设在多标记学习中示例  $x_i$  的预测标记集合为  $Y'_i$ ,多标记评价指标定义如下。其中,符号“ $\uparrow$ ”(“ $\downarrow$ ”)表示分类性能随指标值增大而提升(降低)。

(1) 汉明损失:衡量示例的预测标记中标记错误的平均个数

$$\text{Hamming loss}(\downarrow) = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \oplus Y'_i|}{|L|} \quad (22)$$

式中:符号  $\oplus$  表示两个集合之间的对称差, $L$  表示所有标记的集合, $|\cdot|$  表示集合的大小。汉明损失值越小,分类性能越好。

(2) 1-错误率:统计满足预测标记排名中的最优标记不在真实标记集合里的示例占总体示例的比例,即

$$\text{One-error}(\downarrow) = \frac{1}{n} \sum_i^n \langle [\arg \min_{l \in L} r(x_i, l)] \notin Y_i \rangle \quad (23)$$

式中: $r(x_i, l)$  表示通过算法对标记预测获得的排序值,值越小,排名越前; $\langle x \rangle$  表示一种逻辑运算:当逻辑语句  $x$  为真时, $\langle x \rangle = 1$ ,否则, $\langle x \rangle = 0$ 。当1-错误率的值减小时,分类性能提升。

(3) 平均准确率:计算预测标记中排在特定真实标记前面的标记同时也是真实标记的比例,即有

$$\text{Average precision}(\uparrow) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|Y_i|} \sum_{l \in Y_i} \frac{|\{l_k \in Y_i: r(x_i, l_k) < r(x_i, l)\}|}{r(x_i, l)} \quad (24)$$

平均准确率值越大,分类性能越高。

(4) 覆盖率:表示在预测标记序列中搜索到所有真实标记所需要的平均搜索深度

$$\text{Coverage}(\downarrow) = \frac{1}{n} \sum_i^n \max_{l \in Y_i} r(x_i, l) - 1 \quad (25)$$

随着覆盖率的增长,分类性能随之降低。

(5) 排序损失:表示不是真实标记的标记在预测标记序列中的排序值比真实标记的小的比例

$$\text{Ranking loss}(\downarrow) = \frac{1}{n} \sum_{i=1}^n \frac{|\{(l_a, l_b) | r(x_i, l_a) > r(x_i, l_b), (l_a, l_b) \in Y_i \times \overline{Y_i}\}|}{|Y_i| |\overline{Y_i}|} \quad (26)$$

式中  $\overline{Y_i}$  表示  $Y_i$  的补集。在排序损失增大时,分类性能降低。

在上述的5个评价指标中,除了覆盖率,其余各项指标的值域都为 $[0,1]$ 。在这些指标中,平均准确率与分类效果呈正相关,其他的4个指标都为指标值越小分类性能越优。根据评价的视角,可以将上述的指标分为两类:一类主要关注的是真实标记与预测标记之间的排序对比,包括1-错误率、平均准确率、覆盖率、排序损失;而汉明损失主要关注的是真实标记与预测标记之间差异性。

### 4.3 实验设置

本节对实验步骤和参数设置做详细描述。首先定义数据集上的模糊相似关系。对于数值型特征,其模糊相似关系定义为

$$R_a(x_i, x_j) = 1 - |x_i^a - x_j^a| \quad (27)$$

对于符号型特征,其模糊相似关系定义为

$$R_a(x_i, x_j) = \begin{cases} 1 & x_i^a = x_j^a \\ 0 & \text{其他} \end{cases} \quad (28)$$

在实验过程中,选取了4个多标记特征选择方法: Relief-ML<sup>[13]</sup>、MDDMproj<sup>[14]</sup>、MDDMspc<sup>[14]</sup>和PMU<sup>[15]</sup>,与本文的算法LDFS进行比较分析。同时选择MLKNN<sup>[25]</sup>作为多标记分类器,设置其邻域粒度 $k=10$ ,平滑参数 $\sigma=1$ 。并采用10折交叉验证的方式来获取对应的实验结果。另外,为了保证所有算法之间的统一性,设原始特征空间大小的 $r\%$ 作为特征子集的大小,其特征选取比例 $r\%$ 定义为<sup>[26]</sup>:(1) 如果 $d < 100$ , $r=40$ ; (2) 如果 $100 \leq d \leq 500$ , $r=30$ ; (3) 如果 $500 \leq d \leq 1000$ , $r=20$ ; (4) 如果 $d \geq 1000$ , $r=10$ 。其中 $d$ 表示原始特征空间大小。

考虑到参数对特征选择算法的影响,对经典特征选择算法的参数设置如下:对于MDDMproj和MDDMspc算法,参考文献[14]所示的默认参数设置,设置正则化参数为0.5;对于算法PMU,按照等宽策略将连续值数值进行两折离散<sup>[15]</sup>。

与此同时,对本文算法LDFS中的阈值 $\lambda$ 进行分析,为了获取合适的阈值,将阈值按步长0.05从0到1进行实验分析以选出合适的阈值。在分析的过程中,当算法在多个评价指标上表现最优时,认定对应的阈值 $\lambda$ 为最佳阈值。图1描绘了数据集Emotions随着阈值变化时分类性能的变化情况。综合考量多标记学习中的5个评价指标,Emotions数据集在阈值 $\lambda=0.6$ 时能获得最佳的分类性能。依据类似的实验分析,可以得到关于其他数据集的最佳阈值参数。

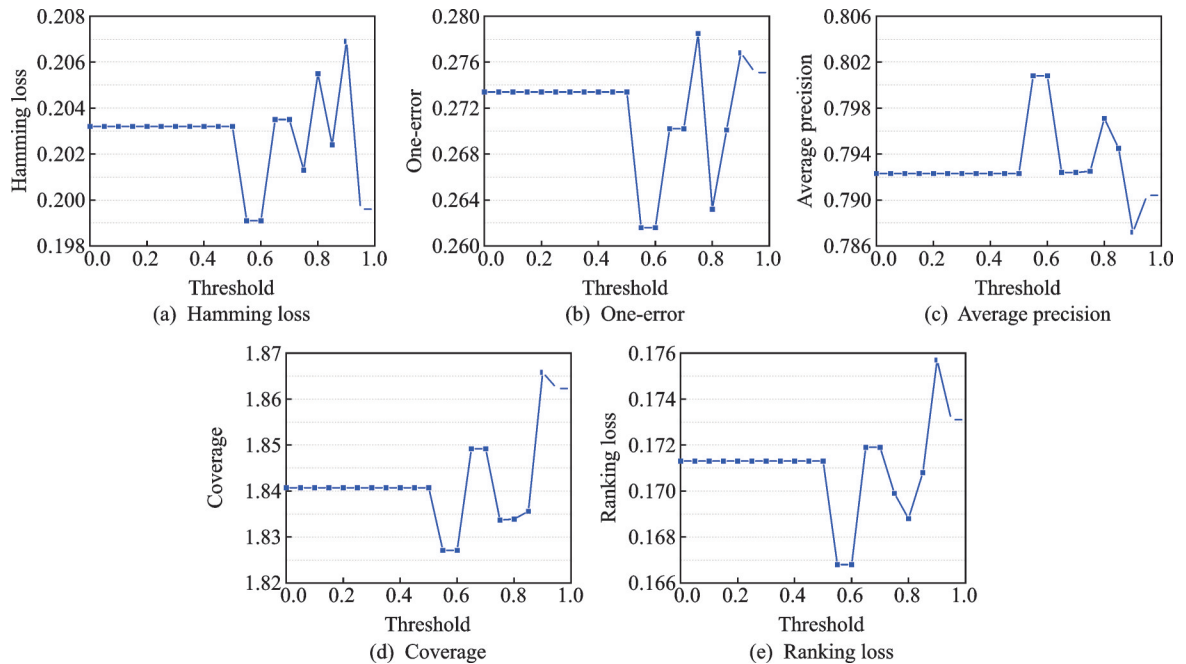


图1 Emotions数据集分类效果随阈值变化情况

Fig.1 Variation of the classification performance with the threshold for Emotions dataset

#### 4.4 实验分析

通过分类器MLKNN,可得出5个特征选择算法在8个数据集上的分类性能。按照汉明损失、1-错误率、平均准确率、覆盖率和排序损失的评价指标,表2~6给出实验的对比结果。

表2 8个数据集上的汉明损失对比

Table 2 Comparison of Hamming loss(↓) on eight datasets

Dataset	MDDMproj	MDDMspc	PMU	Relief-ML	LDFS
Birds	0.049 5±0.002 6(3)	0.050 0±0.004 6(4)	0.047 2±0.003 1(1)	0.051 9±0.003 2(5)	0.049 3±0.006 3(2)
Emotions	0.206 1±0.020 6(3)	0.214 0±0.022 7(5)	0.211 4±0.020 6(4)	0.201 6±0.029 4(2)	0.199 1±0.019 5(1)
Flags	0.272 6±0.056 5(4)	0.265 9±0.050 1(1)	0.266 5±0.046 9(3)	0.301 7±0.049 2(5)	0.266 4±0.061 1(2)
Gnegative	0.082 3±0.007 4(2)	0.085 1±0.008 1(5)	0.083 7±0.007 0(3)	0.084 7±0.008 3(4)	0.082 0±0.008 3(1)
Gpositive	0.161 4±0.024 4(3)	0.180 2±0.020 0(5)	0.172 5±0.021 6(4)	0.151 3±0.025 1(2)	0.150 4±0.023 0(1)
Plant	0.089 6±0.003 4(3)	0.090 4±0.002 8(5)	0.089 8±0.002 7(4)	0.089 1±0.002 1(2)	0.089 0±0.002 3(1)
Virus	0.211 7±0.022 1(4)	0.215 7±0.023 5(5)	0.205 2±0.026 7(3)	0.197 1±0.025 4(2)	0.188 9±0.020 8(1)
Yeast	0.202 6±0.010 1(2)	0.204 0±0.008 7(4)	0.202 8±0.010 3(3)	0.204 1±0.011 0(5)	0.196 7±0.012 5(1)
Avg Rank	3.000	4.250	3.125	3.375	1.250

表3 8个数据集上的1-错误率对比

Table 3 Comparison of One-error(↓) on eight datasets

Dataset	MDDMproj	MDDMspc	PMU	Relief-ML	LDFS
Birds	0.736 4±0.037 2(3)	0.737 9±0.034 1(4)	0.726 9±0.055 7(1)	0.773 7±0.0411(5)	0.731 9±0.066 7(2)
Emotions	0.295 3±0.054 5(3)	0.307 1±0.049 0(5)	0.305 5±0.062 8(4)	0.293 6±0.0817(2)	0.261 6±0.066 5(1)
Flags	0.267 4±0.085 7(5)	0.253 4±0.091 2(4)	0.242 1±0.085 3(2)	0.252 9±0.0980(3)	0.216 6±0.095 3(1)
Gnegative	0.341 9±0.037 0(2)	0.347 0±0.041 6(4)	0.343 4±0.025 2(3)	0.360 7±0.0430(5)	0.340 5±0.045 6(1)
Gpositive	0.333 5±0.061 8(3)	0.368 3±0.058 6(5)	0.360 5±0.050 2(4)	0.323 6±0.0518(2)	0.308 3±0.046 3(1)
Plant	0.670 8±0.051 0(3.5)	0.662 6±0.052 8(2)	0.687 2±0.029 4(5)	0.670 8±0.043 6(3.5)	0.653 3±0.056 2(1)
Virus	0.627 6±0.075 8(5)	0.622 1±0.078 9(4)	0.579 5±0.088 7(3)	0.559 5±0.052 9(2)	0.501 9±0.067 6(1)
Yeast	0.234 6±0.024 8(4)	0.245 3±0.022 2(5)	0.229 2±0.032 6(2.5)	0.227 6±0.024 7(1)	0.229 2±0.030 4(2.5)
Avg Rank	3.562 5	4.125 0	3.062 5	2.937 5	1.312 5

表4 8个数据集上的平均准确率对比

Table 4 Comparison of Average precision(↑) on eight datasets

Dataset	MDDMproj	MDDMspc	PMU	Relief-ML	LDFS
Birds	0.564 8±0.028 9(2)	0.545 9±0.046 7(4)	0.550 6±0.052 1(3)	0.491 5±0.036 2(5)	0.585 1±0.079 3(1)
Emotions	0.784 6±0.036 8(3)	0.773 9±0.031 2(5)	0.774 4±0.029 7(4)	0.788 9±0.043 4(2)	0.800 8±0.039 3(1)
Flags	0.810 9±0.048 7(3)	0.814 0±0.044 6(2)	0.810 3±0.044 9(4)	0.796 3±0.054 0(5)	0.825 0±0.057 2(1)
Gnegative	0.782 2±0.022 2(1)	0.777 9±0.023 8(4)	0.779±0.015 4(3)	0.769 6±0.029 7(5)	0.780 2±0.028 6(2)
Gpositive	0.804 7±0.034 3(3)	0.785 0±0.031 1(5)	0.791 7±0.027 3(4)	0.809 8±0.030 9(2)	0.817 8±0.027 4(1)
Plant	0.530 5±0.034 2(3)	0.536 1±0.032 0(2)	0.524 2±0.017 4(4)	0.522 7±0.026 1(5)	0.540 2±0.038 6(1)
Virus	0.609 6±0.039 8(5)	0.619 9±0.051 7(4)	0.642 8±0.054 9(3)	0.648 8±0.037 2(2)	0.685 3±0.033 6(1)
Yeast	0.750 7±0.022 3(4)	0.746 7±0.018 1(5)	0.754 2±0.018 9(3)	0.754 3±0.018 4(2)	0.758 0±0.022 9(1)
Avg Rank	3.000	3.875	3.500	3.500	1.125

表5 8个数据集上的覆盖率对比

Table 5 Comparison of Coverage(↓) on eight datasets

Dataset	MDDMproj	MDDMspc	PMU	Relief-ML	LDFS
Birds	2.736 0±0.487 9(2)	2.898 9±0.381 2(3)	3.005 2±0.494 7(4)	3.353 9±0.414 8(5)	2.356 9±0.314 5(1)
Emotions	1.854 0±0.185 9(3)	1.929 8±0.124 0(4)	1.932 9±0.159 6(5)	1.818 2±0.174 8(1)	1.827 1±0.195 4(2)
Flags	3.629 7±0.539 6(1)	3.640 5±0.499 4(3)	3.682 4±0.470 4(4)	3.861 1±0.467 8(5)	3.636 8±0.456 8(2)
Gnegative	0.814 7±0.094 9(1)	0.848 5±0.098 2(4)	0.839 2±0.086 4(3)	0.849 3±0.140 5(5)	0.834 2±0.117 8(2)
Gpositive	0.524 3±0.088 5(3)	0.572 5±0.078 2(5)	0.551 2±0.065 1(4)	0.520 1±0.096 4(2)	0.495 1±0.082 2(1)
Plant	2.525 2±0.229 2(4)	2.520 1±0.264 7(3)	2.516 3±0.212 7(2)	2.598 7±0.175 9(5)	2.511 9±0.223 4(1)
Virus	1.501 2±0.135 1(5)	1.461 4±0.171 8(4)	1.394 8±0.153 4(3)	1.368 8±0.216 5(2)	1.254 5±0.160 0(1)
Yeast	6.468 2±0.228 6(4)	6.451 8±0.169 8(5)	6.418 2±0.202 6(3)	6.388 3±0.213 3(2)	6.356 9±0.223 5(1)
AvgRank	3.000	3.750	3.500	3.375	1.375

表6 8个数据集上的排序损失对比

Table 6 Comparison of Ranking loss(↓) on eight datasets

Dataset	MDDMproj	MDDMspc	PMU	Relief-ML	LDFS
Birds	0.099 6±0.023 8(2)	0.104 2±0.018 6(3)	0.106 5±0.022 5(4)	0.125 1±0.019 3(5)	0.081 7±0.016 7(1)
Emotions	0.174 3±0.032 2(3)	0.190 2±0.025 2(4)	0.191 8±0.020 1(5)	0.169 2±0.030 4(2)	0.166 8±0.029 9(1)
Flags	0.206 3±0.061 2(3)	0.204 6±0.051 2(2)	0.207 4±0.053 8(4)	0.229 9±0.060 5(5)	0.202 7±0.060 7(1)
Gnegative	0.106 7±0.012 9(1)	0.110 7±0.012 5(4)	0.108 6±0.011 0(2)	0.111 3±0.018 3(5)	0.109 2±0.016 3(3)
Gpositive	0.172 0±0.029 4(3)	0.188 4±0.025 7(5)	0.181 2±0.023 6(4)	0.171 8±0.031 2(2)	0.162 2±0.029 7(1)
Plant	0.217 8±0.021 8(4)	0.216 6±0.024 5(3)	0.216 1±0.020 5(2)	0.225 2±0.016 8(5)	0.215 7±0.021 9(1)
Virus	0.255±0.029 6(5)	0.247 0±0.042 8(4)	0.235 2±0.042 3(3)	0.228 3±0.039 9(2)	0.209 0±0.030 5(1)
Yeast	0.178 5±0.015 5(4)	0.179 7±0.012 4(5)	0.174 9±0.011 9(3)	0.174 4±0.012 3(2)	0.173 2±0.014 0(1)
Avg Rank	3.125	3.750	3.375	3.500	1.250

在每个表所示的评价指标中,使用了特定的符号表示指标值与分类性能的关系:“↑”表示值越大算法性能越佳,“↓”表示值越小算法性能越佳,并对每个数据集上的算法进行排序,其最佳算法性能指标使用黑体表示。此外,将各个算法在8个数据集上的平均排序列在最后,排序值越小,算法性能越高。

通过表2~6中的实验结果,可得实验的分析对比结果如下:

(1)由表2可知,算法LDFS在6个数据集上的汉明损失分类指标均优于其他4个算法,算法PMU和MDDMspc分别在数据集Birds与Flags上获得了汉明损失的最小值。在数据集Birds和Flags上,算法LDFS取得了汉明损失次优值。在8个数据集上的算法排序,算法LDFS的平均排序排在首位。

(2)由表3中的1-错误率可知,算法LDFS也在6个数据集上取得较优的表现,在其他的两个数据集Birds和Yeast上,算法LDFS的1-错误率取得了次优值。另外,在数据集Yeast上,算法LDFS与算法PMU的1-错误率相等。依据在8个数据集上的平均排序,算法LDFS在1-错误率上性能最优。

(3)由表4中的数据可知,算法LDFS在7个数据集上平均准确率的值都好过其他对比算法,仅在数据集Gnegative上次于算法MDDMproj,在其他多标记数据集上,算法LDFS的平均准确率要明显优于算法MDDMproj。

(4)对于表5中的覆盖率,算法LDFS、MDDMproj和Relief-ML分别在5、2和1个数据集上获得最

小值,覆盖率在所有数据集上的排序,依次为算法 LDFS、MDDMproj、Relief-ML、PMU 和 MDDMspc。

(5)由表 6 可知,算法 LDFS 在大部分数据集上取得排序损失的最小值,除了数据集 Gnegative,其排序损失的最优值由算法 MDDMproj 取得。在平均排序上,算法 LDFS 的表现最为出色。

由上可知,在 8 个数据集和 5 种评价指标,算法 LDFS 在 77.5% 的情况下取得了较明显优势,在其他的 9 种情况中获得了 8 次次优值。而算法 MDDMproj、MDDMspc、PMU 和 Relief-ML 分别在 4、1、2、2 种情况下取得较优情况。且算法 LDFS 均能在至少 5 个数据集上要优于其他算法。由此可见,相对于其他的多标记特征选择算法,算法 LDFS 的优势性能更明显且稳定。

为了进一步研究算法的有效性,选取数据集 Gpositive,在 5 个评价指标下分析了各个算法的分类性能随特征子集的数目的变化趋势,实验的具体情况如图 2 所示。

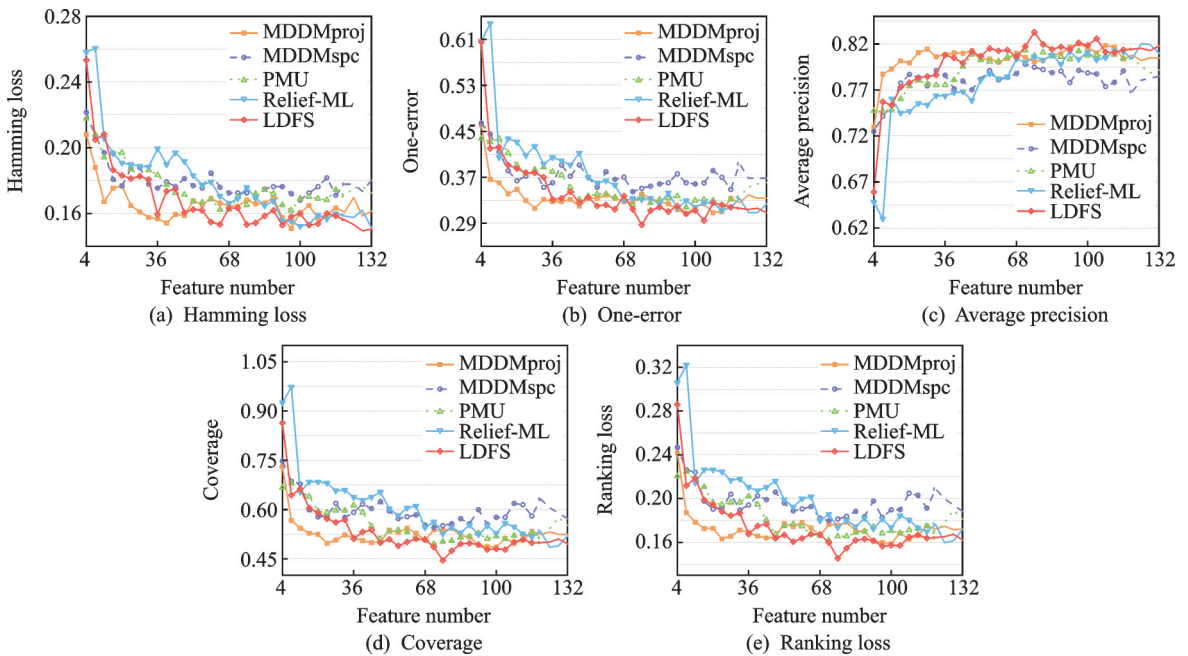


图 2 Gpositive 数据集分类效果随特征数量变化情况

Fig.2 Variation of the classification performance with the number of feature for datasets Gpositive

在图 2 中,使用颜色不同的曲线代表各个算法。由图 2 可知,对 5 个评价指标而言,当特征子集的数量变化时,各个算法的分类效果也随之变化。从总体趋势上来看,随着特征子集的增大,不同算法的分类效果都在不断优化并趋于稳定。对比其他多标记特征选择算法,算法 LDFS 在特征数量相对较少时优势并不明显,这是由于算法考虑的是特征之间的辨别能力排序,在特征数量较少时特征之间的冗余性对分类性能的影响较大。随着特征数量的增长,算法 LDFS 与其他算法的分类性能之间的差距逐渐缩小并实现超越。当特征子集规模在 76 左右时,算法 LDFS 在 5 个评价指标上取得较明显优势,进一步验证了算法 LDFS 能够提高学习模型分类性能。

综上所述,相对于其他多标记特征选择算法,本文中的算法通过标记增强能够获得比多标记数据中更加丰富的标记分布信息,并利用在标记上的标记差异性和在特征上的模糊相对辨识关系,这些信息能够对特征辨识能力的刻画更加合理,使得算法能够在特征数量较少的情况下获得稳定且较优的分类性能。

## 5 结束语

在传统的多标记学习中,示例的相关标记假定为同等重要;然而,对于现实生活中的示例而言,其相关标记之间的相对重要度是不同的。为此,将示例关于标记的模糊相似性融入到原多标记数据的标记关系上,提出了基于模糊相似性的标记增强算法,将逻辑标记转化为标记分布。在此基础上,结合示例对之间的辨识度和标记差异性,提出了基于标记增强和模糊辨识度的标记分布特征选择算法。算法通过对模糊辨识关系的拓展,提出了模糊相对辨识关系以衡量不同特征对于示例对的辨识能力;然后,结合模糊相对辨识关系和标记差异性,并定义了对特征辨识能力的衡量指标。依据此衡量指标获得特征排序集合。在多个数据集上的对比实验表明,相较于其他多标记特征选择算法,算法能获得优势较明显的分类性能,并验证了算法的可行性和稳定性。在未来的研究中,将研究标记分布信息的差异性,构造基于不同搜索策略的特征选择方法,扩展其应用领域。

### 参考文献:

- [1] ZHANG Minling, ZHOU Zhihua. A review on multi-label learning algorithms[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2014, 26(8): 1819-1837.
- [2] BURKHARDT S, KRAMER S. Online multi-label dependency topic models for text classification[J]. *Machine Learning*, 2018, 107(5): 859-886.
- [3] JING Xiaoyuan, WU Fei, LI Zhiqiang, et al. Multi-label dictionary learning for image annotation[J]. *IEEE Transactions on Image Processing*, 2016, 25(6): 2712-2725.
- [4] 徐计, 王国胤, 于洪. 基于粒计算的大数据处理[J]. *计算机学报*, 2015, 38(8): 1497-1517.  
XU Ji, WANG Guoyin, YU Hong. Review of big data processing based on granular computing[J]. *Chinese Journal of Computers*, 2015, 38(8): 1497-1517.
- [5] FAN Xiaodong, ZHAO Weida, WANG Changzhong, et al. Attribute reduction based on max-decision neighborhood rough set model[J]. *Knowledge-Based Systems*, 2018, 151: 16-23.
- [6] DAI Jianhua, HU Qinghua, HU Hu, et al. Neighbor inconsistent pair selection for attribute reduction by rough set approach[J]. *IEEE Transactions on Fuzzy Systems*, 2018, 26(2): 937-950.
- [7] DUBOIS D, PRADE H. Rough fuzzy sets and fuzzy rough sets[J]. *International Journal of General Systems*, 1990, 17: 191-209.
- [8] WANG Changzhong, QI Yali, SHAO Mingwen, et al. A fitting model for feature selection with fuzzy rough sets[J]. *IEEE Transactions on Fuzzy Systems*, 2017, 25(4): 741-753.
- [9] 陈俞, 赵素云, 李雪峰, 等. 基于随机抽样的模糊粗糙约简[J]. *软件学报*, 2017, 28(11): 2825-2835.  
CHEN Yu, ZHAO Suyun, LI Xuefeng, et al. Fuzzy rough reduction based on random sampling[J]. *Journal of Software*, 2017, 28(11): 2825-2835.
- [10] ZHAO Hong, WANG Ping, HU Qinghua, et al. Fuzzy rough set based feature selection for large-scale hierarchical classification[J]. *IEEE Transactions on Fuzzy Systems*, 2019, 27(10): 1891-1903.
- [11] 耿新, 徐宁. 标记分布学习与标记增强[J]. *中国科学:信息科学*, 2018, 48(5): 521-530.  
GENG Xin, XU Ning. Label distribution learning and label enhancement [J]. *Scientia Sinica(Informationis)*, 2018, 48(5): 521-530.
- [12] XU Ning, TAO An, GENG Xin, et al. Label enhancement for label distribution learning[C]//*Proc of Int Joint Conf on Artificial Intelligence*. Stockholm: IJCAI, 2018: 2926-2932.
- [13] SPOLAOR N, CHERMAN E A, MONARD M C, et al. Relief for multi-label feature selection[C]//*Proc of 2013 Brazilian Conf on Intelligent Systems*. Fortaleza: IEEE, 2013: 6-11.
- [14] ZHANG Yin, ZHOU Zhihua. Multilabel dimensionality reduction via dependence maximization[J]. *ACM Transactions on Knowledge Discovery From Data*, 2010, 4(3): 1-21.
- [15] LEE J, KIM D. Feature selection for multi-label classification using multivariate mutual information[J]. *Pattern Recognition*

- Letters, 2013, 34(3): 349-357.
- [16] XU Haotian, XU Lingyu. Multi-label feature selection algorithm based on label pairwise ranking comparison transformation [C]//Proc of 2017 Int Joint Conf on Neural Networks. AK: IEEE, 2017: 1210-1217.
- [17] TEISSEYRE P. Feature ranking for multi-label classification using Markov networks[J]. Neurocomputing, 2016, 205: 439-454.
- [18] XU Suping, YANG Xibei, YU Hualong, et al. Multi-label learning with label-specific feature reduction[J]. Knowledge-Based Systems, 2016, 104: 52-61.
- [19] LIN Yaojin, HU Qinghua, LIU Jinghua, et al. Streaming feature selection for multi-label learning based on fuzzy mutual information[J]. IEEE Transactions on Fuzzy Systems, 2017, 25(6): 1491-1507.
- [20] HE Zhouzhou, LI Xi, ZHANG Zhongfei, et al. Data-dependent label distribution learning for age estimation[J]. IEEE Transactions on Image Processing, 2017, 26(8): 3846-3858.
- [21] LIU Anan, SHI Yingdi, JING Peiguang, et al. Structured low-rank inverse covariance estimation for visual sentiment distribution prediction[J]. Signal Processing, 2018, 152: 206-216.
- [22] ZHAO Peng, ZHOU Zhihua. Label distribution learning by optimal transport[C]//Proc of the 32nd AAAI Conf on Artificial Intelligence. LA: AAAI, 2018: 4506-4513.
- [23] DAI Jianhua, HU Hu, WU Weizhi. Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets[J]. IEEE Transactions on Fuzzy Systems, 2018, 26(4): 2174-2187.
- [24] CHEN Degang, ZHANG Lei, ZHAO Suyun, et al. A novel algorithm for finding reducts with fuzzy rough sets[J]. IEEE Transactions on Fuzzy Systems, 2012, 20(2): 385-389.
- [25] ZHANG Minling, ZHOU Zhihua. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40 (7): 2038-2048.
- [26] KASHEF S, NEZAMABADI-POUR H, NIKPOUR B. Multilabel feature selection: A comprehensive review and guiding experiments[J]. Data Mining & Knowledge Discovery, 2018, 8(10): e1240.

## 作者简介:



熊传镇(1998-),男,硕士研究生,研究方向:粒计算、知识发现,E-mail:xcz740@163.com。



钱文彬(1984-),通信作者,男,博士,副教授,研究方向:知识发现、粒计算、机器学习,E-mail:qianwenbin1027@126.com。



王映龙(1970-),男,博士,教授,研究方向:数据挖掘、知识发现,E-mail:wangylx@126.com。

(编辑:刘彦东)