

# 基于指数损失间隔的多标记特征选择算法

李雨婷

(南京邮电大学 计算机学院、软件学院、网络空间安全学院,江苏 南京 210023)

**摘要:**在多标记学习的任务中,多标记学习的每个样本可被多个标签标记,比单标记学习的应用空间更广关注度更高,多标记学习可以利用关联性提高算法的性能。在多标记学习中,传统特征选择算法已不再适用,一方面,传统的特征选择算法可被用于单标记的评估标准。多标记学习使得多个标记被同时优化;而且在多标记学习中关联信息存在于不同标记间。因此,可设计一种能够处理多标记问题的特征选择算法,使标记之间的关联信息能够被提取和利用。通过设计最优的目标损失函数,提出了基于指数损失间隔的多标记特征选择算法。该算法可以通过样本相似性的方法,将特征空间和标记空间的信息融合在一起,独立于特定的分类算法或转换策略。优于其他特征选择算法的分类性能。在现实世界的数据集上验证了所提算法的正确性以及较好的性能。

**关键词:**多标记学习;特征选择;分类间隔;指数损失

中图分类号:TP181

文献标识码:A

文章编号:1673-629X(2020)04-0046-06

doi:10.3969/j.issn.1673-629X.2020.04.009

## Exponential Loss Margin Based Multi-label Feature Selection

LI Yu-ting

(School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

**Abstract:** In multi-label learning tasks, each sample can be associated with multiple labels at the same time, it has a wider application space than the single-label learning problem. Multi-label learning problem employs correlation information to improve the performance of the algorithm. In the multi-labeling process, the traditional feature selection algorithm is no longer applicable. On the one hand, they are generally designed to evaluate criteria for single markers. In the multi-label learning, it is necessary to optimize multiple tags at the same time; on the other hand, there is a certain amount of associated information between different tags in multi-label learning. Therefore, it is necessary to design a feature selection algorithm capable of handling multi-labeling problems, which is capable of extracting and utilizing association information between labels. An improved multi-label feature selection algorithm based on exponential loss margin is proposed, benefiting from the large margin based multi-label feature selection algorithm. The algorithm combines the information of feature space with mark space according to sample similarity. The correlation information is also independent of the specific classification algorithm or transformation strategy. The experiments on real world datasets demonstrate the correctness and high performance of the proposed algorithm.

**Key words:** multi-label learning; feature selection; margin; exponential loss

## 0 引言

多标记学习的概念来源于文档分类<sup>[1-2]</sup>过程中遇到的多语义问题,经过了十几年的发展,多标记学习也慢慢成为了国际机器学习领域的热点话题。多标记学习的研究成果也在实际问题中得到了较好的应用,逐渐在图像分割<sup>[3]</sup>、图像标注<sup>[4]</sup>、面部表情识别<sup>[5]</sup>、动作识别<sup>[6]</sup>、广告推荐<sup>[7]</sup>、文本分类和生物信息学<sup>[8]</sup>等领域扮演重要的角色。

在机器学习领域,机器学习问题通常可分为监督学习、无监督学习<sup>[9]</sup>。在传统的监督学习中,每个样本与一个标记相关,而在多标记学习中一个样本可以有多个标记。且标记之间存在一定的关联性,如何提取和利用标记间的关联性成为提高多标记学习算法性能的关键。为此,可采用基于指数损失间隔的多标记特征选择算法进行特征选择,该算法不仅不依赖于具体的转换策略或分类算法,也可以有效地利用标记间的

关联信息。

文中首先介绍了多标记学习和特征选择算法,以及一种多标记数据样本相似性的度量方法;然后提出一种基于指数损失间隔的多标记特征选择算法。该算法可使用样本标记的关联性使特征空间与标记空间的信息结合在一起,可得到标记间的关联信息,也独立于特定的分类算法或转换策略;最后通过现实世界数据集上的相关实验及结果对算法进行验证。

## 1 相关工作

### 1.1 多标记学习

用  $X \in \mathcal{R}^D$  来表示  $D$  维的样本空间,  $y = \{l^1, l^2, \dots, l^q, \dots, l^Q\}$  表示有  $Q$  个不同标记的集合,  $l^q$  表示第  $q$  个标记,多标记学习的主要任务是从训练数据集  $D_{\text{train}} = \{(x_i, y_i) \mid 1 \leq i \leq n\}$  学习来得到函数  $h: X \rightarrow 2^Y$ , 对于标记未知的样本  $x'_i \in D_{\text{test}}$ , 使得  $h(x'_i) = \tilde{y}_i$ 。其中,  $x'_i = (x'_i{}^1, x'_i{}^2, \dots, x'_i{}^d, \dots, x'_i{}^D)$  表示测试集中的第  $i$  个样本,  $\tilde{y}_i = (\tilde{y}_i^1, \tilde{y}_i^2, \dots, \tilde{y}_i^q, \dots, \tilde{y}_i^Q)$  表示预测出的各标记与样本  $x'_i$  的关系,  $\tilde{y}_i^q \in \{0, 1\}$ ,  $\tilde{y}_i^q = 1$  表示第  $q$  个标记与该样本  $i$  相关,  $\tilde{y}_i^q = 0$  表示第  $q$  个标记与该样本  $i$  无关,  $y_i$  表示测试集中的样本与标记的真实关系,  $\tilde{y}_i$  表示  $h(x'_i)$  的预测结果,若函数  $h(x'_i)$  性能很好,则应有  $y_i = \tilde{y}_i$ 。

目前,根据多标记学习算法是否利用标签之间的相关性,可以分为三个阶次:常见的一阶算法,如 binary relevance, 将候选标记个数为  $Q$  个的多标记问题转变成  $Q$  个独立的二分类问题,再对每个二分类问题独立求解。calibrated label ranking<sup>[4]</sup> 是一种二阶算法;此外,分类器链(classifier chain, CC)<sup>[10]</sup> 算法属于高阶算法,也可用来处理多标记问题。

### 1.2 特征选择

特征选择<sup>[11-12]</sup> 是一种数据预处理技术,常用于去除数据中的噪音和冗余特征。相比其他的降维方法,特征选择不改变数据的原有特征,使原始的数据特征被保留。

通常将特征选择方法分为3种模型:封装器模型、过滤器模型和嵌入式模型<sup>[13-14]</sup>。常用过滤器模型的特征选择方法有 ReliefF<sup>[15]</sup>、Fisher Score<sup>[16]</sup> 算法等。这种特征选择模型,独立于具体的分类器,所以当选择不同的分类器时也不会对算法的结果产生影响,因此被广泛应用。封装器模型直接针对给定的分类器进行优化,所以对于不同的分类学习算法,其分类性能也有不同的差异,封装器模型算法也因此受限于所选分类器的性能。而嵌入式模型算法独立于具体的分类学习算法,且其算法计算量比封装器模型小。

文中所提出的多标记特征选择算法属于过滤器模

型,该算法与具体的转换策略<sup>[14]</sup> 无关,可通过分类学习算法得到若干特征子空间,再根据分类性能来选择最优的特征子空间作为最终分类结果。

### 1.3 多标记样本的相似度

由于在多标记学习中,不同数据集的标记之间通常会有不同的关联性,这里通过构建概念结构图的方式去假设标记的结构。在传统的单标记问题中,样本间的相似度<sup>[17-18]</sup> 可通过式(1)计算:

$$s_{(i_1, i_2)} = \begin{cases} \frac{1}{n_l}, & \text{第 } i_1, i_2 \text{ 个样本均属于类别 } l \\ 0, & \text{第 } i_1, i_2 \text{ 个样本均不属于类别 } l \end{cases} \quad (1)$$

其中  $n_l$  表示属于类别  $l$  的样本个数。在多标记的学习中,同一个样本有可能属于多个标记,此时不能通过式(1)明确得到两个样本所属的类别。在这里,利用式(2)<sup>[19]</sup> 可以计算两个多标记样本间的相似性  $s_{(i_1, i_2)}$ , 其中  $n_q$  为与第  $q$  个标记相关的样本个数,式(2)也可看成是 Jaccard<sup>[20]</sup> 的另一种表现形式。数据集中不同标记出现次数不相同,通过改变权重  $n_q$  表示不同标记对样本相似度的影响。

$$s_{(i_1, i_2)} = \frac{\sum_{q=1}^Q n_q \cdot [y_{i_1}^q = 1 \wedge y_{i_2}^q = 1]}{\sum_{q=1}^Q n_q \cdot [y_{i_1}^q = 1 \vee y_{i_2}^q = 1]} \quad (2)$$

如图1中的上半部分,表示含有ABC三种标记的结构图,图中的三角形、星形、菱形、正方形表示不同的样本,用不同的形状来描述样本在特征  $R^1$  上的不同取值情况。与单标记问题不同,多标记问题中同一个样本可以同时和多个标记相关联,例如图1上半部分所示,特征  $R^1$  取值为星形的样本,既与标记A相关联,又与标记B相关联。将每种标记当成一个类别,多标记问题变成了多类问题,将图1上半部分的重叠区域进行分解,如图1下半部分所示。

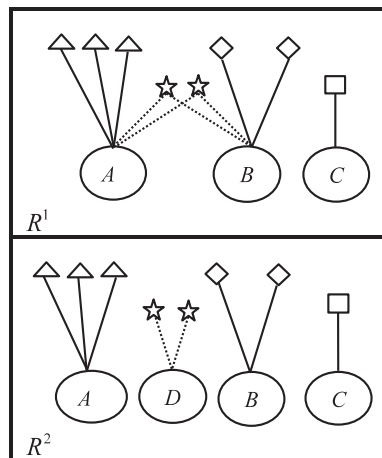


图1 多标记问题的分类

当原始问题被转换为二分类问题,如图1下半部分所示,定义了新的类别  $D$ , 变成了与  $A$  和  $B$  完全不

同的类别,根据式(1)可知它们之间的相似度为0,但真实的样本间相似度并不是0,因此文中所使用的方法对描述样本间相似度的关系更有效。

## 2 基于指数损失间隔的多标记特征选择

图2是通过式(2)来构建的简单无向图,边距越短,两个样本之间的相似度就越大。实心圆形表示样本  $x_i$ ,空心圆形表示与样本  $x_i$  带有相同标记,且两样本间的相似度大于阈值  $s_{\min}$  的样本,七角星表示与样本  $x_i$  的标记不同,且样本间相似度小于阈值  $s_{\min}$  的样本。若给出样本的相似度以及阈值  $s_{\min}$ ,可在图2的标记空间(a)中找出样本  $x_i$  的近邻样本,可通过近邻样本来预测样本的标记。

图2(b)所示的是特征空间,通过特征选择可以使得样本间的距离发生变化,箭头的方向表示通过特征选择使样本间的距离发生变化的方向。图2(b)中由七角星到实心圆形的距离可能比空心圆形到实心圆形的距离更近,意味着样本在特征空间和标记空间上分布不一致。

特征空间和标记空间的样本分布不一致的特性,可能会导致样本的分类错误。如果能找到如图2(c)所示的特征子空间,使得样本(特征子空间的样本)在特征空间上的分布与标记空间上的分布一致,就可用这部分特征来寻找第  $i$  个样本的近邻,从而使分类效果更好。

特征子空间的不同会对样本间的距离产生不同的影响,从而影响损失函数值,找到合适的特征子空间,可以使损失函数最小,所以特征选择的评价准则可以选择损失函数来衡量。由此,提出了基于指数损失分类间隔的多标记特征选择算法。给定样本  $(x_i, y_i)$  及相似度  $s_{\min}$ ,可定义:

$$\text{sim}(i) = \{ (x_{i'}, y_{i'}) \mid s_{(i,i')} \geq s_{\min}, 1 \leq i' \leq n \text{ 且 } i' \neq i \} \quad (3)$$

$$\text{dissim}(i) = \{ (x_{i'}, y_{i'}) \mid s_{(i,i')} < s_{\min}, 1 \leq i' \leq n \text{ 且 } i' \neq i \} \quad (4)$$

$$\text{nearhit}(i) = \min_{(x_{i'}, y_{i'}) \in \text{sim}(i)} \|x_i - x_{i'}\|, \quad (x_{i'}, y_{i'}) \in \text{sim}(i) \quad (5)$$

$$\text{nearmiss}(i) = \min_{(x_{i'}, y_{i'}) \in \text{dissim}(i)} \|x_i - x_{i'}\|, \quad (x_{i'}, y_{i'}) \in \text{dissim}(i) \quad (6)$$

$$\text{margin}(i) = \frac{1}{2} \left( \|x_i - x_{\text{nearhit}(i)}\|^2 - \|x_i - x_{\text{nearmiss}(i)}\|^2 \right) \quad (7)$$

可以按照式(2)和阈值  $s_{\min}$ ,对数据集进行划分,划分为与样本  $x_i$  具有相同标记的样本集合  $\text{sim}(i)$  和具有不同标记的样本集合  $\text{dissim}(i)$ ,通过特征空间中的欧氏距离,分别在两集合中找出与样本  $x_i$  具有相

同类别的最近邻  $\text{nearhit}(i)$  和不同类别的最近邻  $\text{nearmiss}(i)$ ,再通过  $\text{nearmiss}(i)$ 、 $\text{nearhit}(i)$  来计算分类间隔  $\text{margin}(i)$ 。在此基础上,用  $\text{nb}(i)$  ( $\text{near neighbor}$ ) 表示  $x_i$  的目标近邻样本,并定义损失函数  $\text{Loss}_{x_i}(w)$ :

$$\text{Loss}_{x_i}(w) = \sum_{(x_j, y_j) \in \text{nb}(i)} \|wx_i - wx_j'\|^2 + \lambda \sum_{(x_j, y_j) \in \text{dissim}(i)} \delta(i', j) \quad (8)$$

损失函数的第一项表示的是惩罚与样本  $x_i$  相距较远的样本,而并不是指所有的与  $x_i$  具有相同标记的样本,第二项表示的是惩罚与样本  $x_i$  的相距较近且与  $x_i$  标记不同的样本。 $w$  为特征的权重向量,调节参数  $\lambda > 0$ ,  $\delta(i', j)$  见式(9),指数损失见式(10):

$$\delta(i', j) = (1 - s_{(i,j)}) \cdot \text{EXP}(x_i) \quad (9)$$

$$\text{EXP}(x_i) = \exp(\max(0, \text{margin}(i) + \frac{1}{2} (\|wx_i - wx_j'\|^2 - \|wx_i - wx_j\|^2))) \quad (10)$$

在这里关注的是特殊区域里被惩罚的样本且与样本  $x_i$  标记不相似的样本。该区域的样本与样本  $x_i$  的距离小于样本  $x_i$  到其任何目标近邻的距离加上  $\text{margin}$ ,对此时的样本进行惩罚。若训练集中的样本都有着损失低、分类间隔大的特点,就可以使得算法具有更好的推广性能。

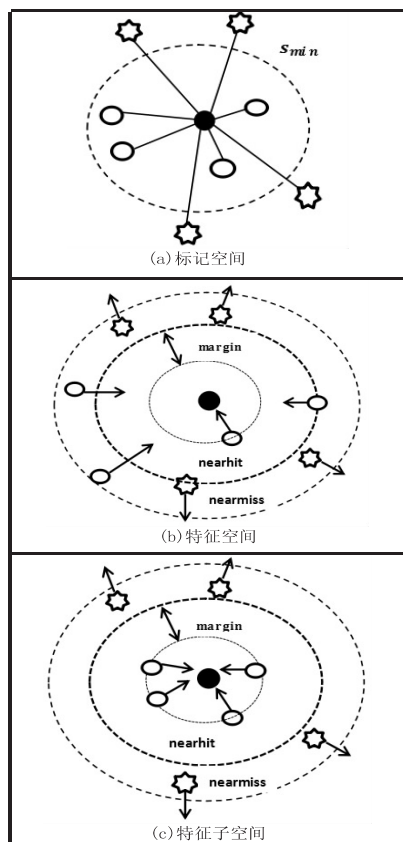


图2 比较不同空间中的图

为使得损失函数最小,可优化函数式(11),求得

一组最优特征权重  $w^*$ , 使  $\text{nb}(i)$  中的样本到  $x_i$  的距离均小于  $\text{dissim}(i)$  中的样本到样本  $x_i$  的距离, 使标记空间中的样本分布和特征空间一致。  $w^*$  可以通过梯度下降法求得。具体见算法 1。

$$\text{Loss}_w(w) = \sum_{i=1}^n \text{Loss}_{x_i}(w)$$

(11)

$$\frac{\partial \text{Loss}_w(w)}{\partial w^d} = 2w^d \sum_{(x_i, y_i) \in \text{nb}(i)} \|x_i^d - x_i^d\|^2 + \lambda \sum_{(x_i, y_i) \in \text{dissim}(i)} \frac{\partial \delta(i', j)}{\partial w^d}$$

(12)

其中,

$$\frac{\partial \delta(i', j)}{\partial w^d} = \begin{cases} 0, & \text{if } \text{margin}(i) + \|x_i - x_i\|^2 < \|x_i - x_j\|^2 \\ \text{diff}(w^d), & \text{otherwise} \end{cases}$$

(13)

$$\text{diff}(w^d) = 2w^d \sum_{(x_i, y_i) \in \text{dissim}(i)} (1 - s_{(i, j)}) (\|x_i^d - x_i^d\|^2 - \|x_i^d - x_j^d\|^2) \cdot (\exp(\text{margin}(i) + \|x_i - x_i\|^2 - \|x_i - x_j\|^2))$$

(14)

算法 1: GMBA\_exp。

输入: 梯度下降算法的步长  $\beta_i$ , 近邻数  $k$ , 训练数据集  $D_{\text{train}}$ , 阈值  $s_{\min}$ , 调谐参数  $\lambda$ ;

输出: 特征权重  $w$ 。

1. 初始化  $w = 1$
2. for  $i = 1$  to  $n$  ;
3. 从训练数据中得到  $\text{sim}(i)$ ,  $\text{dissim}(i)$ ,  $\text{nearhit}(i)$ ,  $\text{nearmiss}(i)$ , 计算  $\text{margin}(i)$
4. 依据  $k$  求  $\text{nb}(i)$
5. for  $d = 1$  to  $D$  ;
6.  $\nabla^d = 2w^d \sum_{(x_i, y_i)} \|x_i^d - x_i^d\|^2 + \lambda \sum_{(x_i, y_i) \in \text{dissim}(i)} \frac{\partial \delta(i', j)}{\partial w^d}$
7. end
8.  $w = w - (\beta_i \nabla) / \|\nabla\|$ , 其中  $\beta_i$  为步长
9. end
10. 输出  $w$

GMBA<sup>[19]</sup>算法与 GMBA\_exp 算法, 主要差别在于损失函数的形式上, GMBA<sup>[19]</sup>算法采用 hinge 损失<sup>[19]</sup>函数的形式。

3 实 验

在算法的有效性验证方面, 先在数据集上将 SPEC (spectral feature selection)<sup>[14]</sup>、GMBA<sup>[19]</sup> (运用 hinge 损失算法)、GMBA\_exp、多标记 ML\_FScore

(MLFS)<sup>[21]</sup>算法和 ML\_ReliefF (MLRF)<sup>[21]</sup>算法进行比较, MLFS、MLRF 对应的是传统单标记问题中经典特征选择算法 Fisher Score<sup>[16]</sup>、Relief<sup>[15]</sup>的拓展。SPEC (spectral feature selection)<sup>[14]</sup>谱特征选择, 它不仅可以提取图谱信息, 也可以提取类标间的结构信息。五种特征选择算法均属于过滤器模型。

3.1 数据集及参数

实验运行环境是 Python 2.7, 转化策略选用 BR<sup>[13]</sup>, KD 树、K 近邻算法从 Scikit-learn<sup>[22]</sup>中组成基分类器。MLRF 和 GMBA<sup>[19]</sup>、GMBA\_exp 中的近邻数均设置为 3, GMBA<sup>[19]</sup>、GMBA\_exp 中  $s_{\min}$  和  $\lambda$  均设置为 1,  $\beta_i = 0.9$ 。从 MULAN<sup>[23]</sup>上下载了四组来自不同领域的数据集, 数据集的具体信息见表 1。实验时, 将样本方差为 0 的特征进行人工去除, 并对数值型的特征进行归一化, 所有实验结果均由 10 折交叉验证获得。

3.2 评价指标

在 BR<sup>[13]</sup>转化策略下, 将 GMBA\_exp 与过滤器模型算法和封装器模型算法进行比较, 算法性能的评价指标选<sup>[24]</sup>用 F1\_Macro (↑) 及汉明损失 (↓), 用 ↑ 表示指标值越大越好, 用 ↓ 表示指标值越小越好, 具体实验结果见图 3 ~ 图 6 (由于篇幅限制, 在这里只显示 birds、emotions 数据集的实验结果)。水平线表示不用特征选择算法时分类的结果。横轴表示所选的特征个数占特征总数的百分比, 纵轴为指标值。

3.3 实验结果分析

通过两个不同的多标记评价指标, 在四个多标记数据集上进行实验。实验结果显示, 随着所选特征的增加, 评价指标并不是单调增加或者单调减少, 说明数据集中有的特征对分类性能的提高有负面影响。在多数情况下, 文中提出的 GMBA\_exp 算法相比其他算法具有更好的分类性能。此外, 在与过滤器模型算法、封装器模型算法的对比实验中可以看出, 在相同的 BR<sup>[13]</sup>转化策略下, GMBA\_exp 算法的分类性能最优。

表 1 数据集描述

数据集	领域	样本数	特征个数	标记数	维度
Birds	audio	645	258	19	258/19
emotions	music	593	72	6	72/6
genebase	biology	662	1 186	27	1 186/27
medical	text	978	1 449	45	1 449/45

4 结束语

实验结果表明, GMBA\_exp 算法优于其他算法。GMBA\_exp 和 GMBA<sup>[19]</sup>算法都致力于寻找最优的特征子空间, 使得样本在特征空间上的分布与标记空间



上的分布一致,在大多数情况下,GMBA\_exp 算法的效果要优于 GMBA<sup>[19]</sup>。综上,文中提出的基于指数损失间隔的多标记特征选择算法能更好地描述标记间的相关性,提高多标记的分类性能。

与过滤器模型算法比较的实验结果如图 3 和图 4 所示。

与封装器模型算法比较的实验结果如图 5 和图 6 所示。

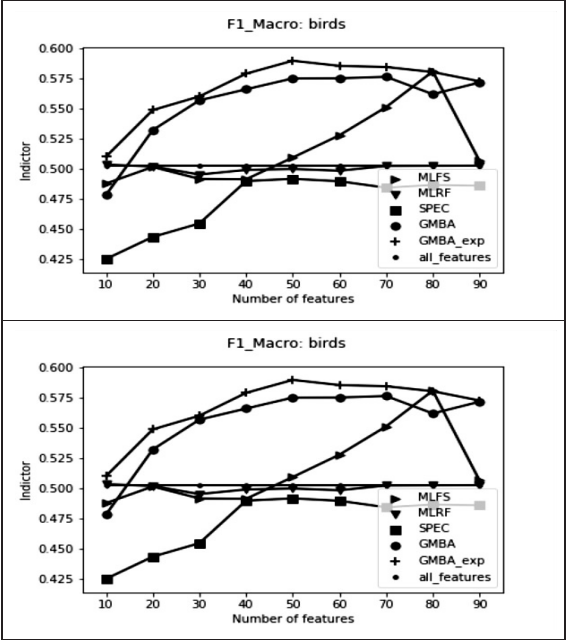


图 3 BR 转化策略下 F1\_Macro( ↑ )(1)

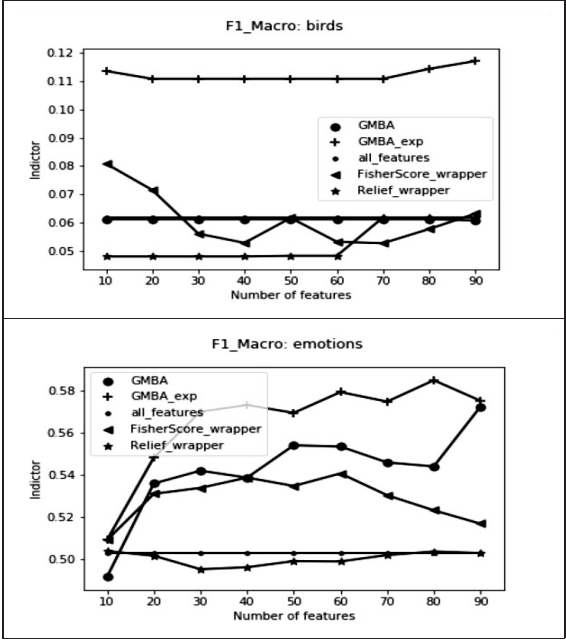


图 5 BR 转化策略下 F1\_Macro( ↑ )(2)

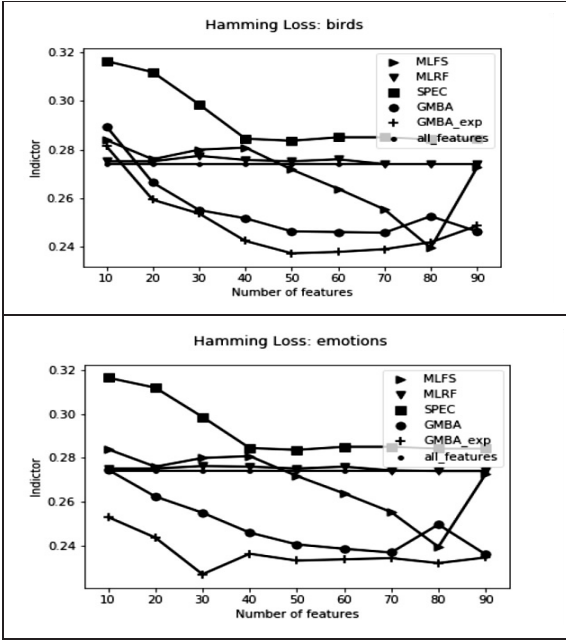


图 4 BR 转化策略下 Hamming loss( ↓ )(1)

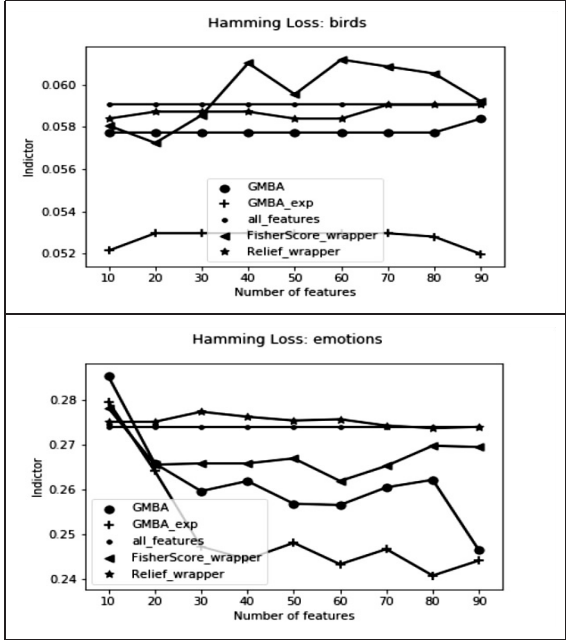


图 6 BR 转化策略下 Hamming loss( ↓ )(2)

参考文献:

[1] NAM J, KIM J, MENCÍA E L, et al. Large-scale multi-label text classification—revisiting neural networks[C]//Joint European conference on machine learning and knowledge discovery in databases. Berlin; Springer, 2014; 437–452.

[2] NGUYEN V A, BOYD-GRABER J, RESNIK P, et al. Learning a concept hierarchy from multi-labeled documents [C]//Proceedings of the 27th international conference on neural information processing systems – volume 2. Cavtat Dubrovnik, Croatia; MIT Press, 2014; 3671–3679.

[3] SUNDARAMOORTHY G, HONG B W. Fast label: easy and efficient solution of joint multi-label and estimation problems [C]//2014 IEEE conference on computer vision and pattern recognition ( CVPR ). Columbus, OH, USA; IEEE, 2014; 3126–3133.

[4] CABRAL R S, DE LA TORRE F, COSTEIRA J P, et al. Matrix completion for weakly-supervised multi-label image

- classification[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2015, 37(1): 121–135.
- [5] ZHAO K, CHU W S, DE LA TORRE F, et al. Joint patch and multi-label learning for facial action unit detection [C]//2015 IEEE conference on computer vision and pattern recognition (CVPR). Boston, MA, USA; IEEE, 2015: 2207–2216.
- [6] KHAMIS S, DAVIS L S. Walking and talking: a bilinear approach to multi-label action recognition [C]//2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW). Boston, MA, USA; IEEE, 2015: 1–8.
- [7] GENTILE C, ORABONA F. On multilabel classification and ranking with partial feedback [C]//Proceedings of the 25th international conference on neural information processing systems—volume 1. Lake Tahoe, Nevada, USA; Curran Associates Inc., 2012: 1151–1159.
- [8] CHEN Z, CHEN M, WEINBERGER K Q, et al. Marginalized denoising for link prediction and multi-label learning [C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2015: 1707–1713.
- [9] DUDA R O, HART P E, STORK D G. Pattern classification [R]. New York, NY: Wiley, 2001.
- [10] READ J, PFAHRINGER B, HOLMES G, et al. Classifier chains for multi-label classification [J]. Machine Learning, 2011, 85(3): 333–359.
- [11] SPOLAÔR N, MONARD M C, LEE H D. Feature selection for multi-label learning [C]//Proceedings of the 24th international conference on artificial intelligence. Buenos Aires, Argentina: AAAI Press, 2015: 4401–4402.
- [12] LIU H, ZHANG S, WU X. MLSLR: multilabel learning via sparse logistic regression [J]. Information Sciences, 2014, 281: 310–320.
- [13] ALELYANI S, TANG J, LIU H. Feature selection for clustering: a review [M]. Tiruchirappalli: Data Clustering, Chapman and Hall/CRC, 2013: 29–60.
- [14] 严鹏, 李云. 基于图谱的多标记特征选择算法 [J]. 计算机科学与探索, 2016, 10(4): 543–553.
- [15] SPOLAÔR N, CHERMAN E A, MONARD M C, et al. ReliefF for multi-label feature selection [C]//2013 Brazilian conference on intelligent systems. Fortaleza, Brazil: IEEE, 2013: 6–11.
- [16] DUDA R O, HART P E, STORK D G. Pattern classification [M]. Hoboken, USA: John Wiley & Sons, 2012.
- [17] ZHAO Z, LIU H. Spectral feature selection for supervised and unsupervised learning [C]//Proceedings of the 24th international conference on machine learning. Ho Chi Minh City, Vietnam: ACM, 2007: 1151–1157.
- [18] ZHAO Z, LIU H. Spectral feature selection for data mining [R]. New York: Chapman & Hall/CRC, 2011.
- [19] YAN Peng, LI Yun. Graph-margin based multi-label feature selection [C]//Joint European conference on machine learning and knowledge discovery in databases. Berlin: Springer, 2016: 540–555.
- [20] SULAIMAN N H, MOHAMAD D. A Jaccard-based similarity measure for soft sets [C]//2012 IEEE symposium on humanities, science and engineering research. Kuala Lumpur, Malaysia: IEEE, 2012: 659–663.
- [21] HUANG H. Multi-label relief and f-statistic feature selections for image annotation [C]//IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE, 2012: 2352–2359.
- [22] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Mining multi-label data [M]//Data mining and knowledge discovery handbook. Boston: Springer, 2009: 667–685.
- [23] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: machine learning in python [J]. Journal of Machine Learning Research, 2011, 12(10): 2825–2830.
- [24] 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.