

基于浮动阈值分类器组合的多标签分类算法

张丹普^{1,2*}, 付忠良¹, 王莉莉^{1,2}, 李 昕^{1,2}

(1. 中国科学院 成都计算机应用研究所, 成都 610041; 2. 中国科学院大学, 北京 100049)

(* 通信作者电子邮箱 linda_zdp@126.com)

摘 要: 针对目标可以同时属于多个类别的多标签分类问题, 提出了一种基于浮动阈值分类器组合的多标签分类算法。首先, 分析探讨了基于浮动阈值分类器的 AdaBoost 算法(AdaBoost. FT)的原理及错误率估计, 证明了该算法能克服固定分段阈值分类器对分类边界附近点分类不稳定的缺点从而提高分类准确率; 然后, 采用二分类(BR)方法将该单标签学习算法应用于多标签分类问题, 得到基于浮动阈值分类器组合的多标签分类方法, 即多标签 AdaBoost. FT。实验结果表明, 所提算法的平均分类精度在 Emotions 数据集上比 AdaBoost. MH、ML-kNN、RankSVM 这 3 种算法分别提高约 4%、8%、11%; 在 Scene、Yeast 数据集上仅比 RankSVM 低约 3%、1%。由实验分析可知, 在不同类别标记之间基本没有关联关系或标签数目较少的数据集上, 该算法均能得到较好的分类效果。

关键词: 连续 AdaBoost; 浮动阈值; 极大似然原理; 多标签分类; 集成学习; 二分类方法

中图分类号: TP391.4; TP18 **文献标志码:** A

Multi-label classification algorithm based on floating threshold classifiers combination

ZHANG Danpu^{1,2*}, FU Zhongliang¹, WANG Lili^{1,2}, LI Xin^{1,2}

(1. Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu Sichuan 610041, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: To solve the multi-label classification problem that a target belongs to multiple classes, a new multi-label classification algorithm based on floating threshold classifiers combination was proposed. Firstly, the theory and error estimation of the AdaBoost algorithm with floating threshold (AdaBoost. FT) were analyzed and discussed, and it was proved that AdaBoost. FT algorithm could overcome the defect of unstability when the fixed segmentation threshold classifier was used to classify the points near classifying boundary, the classification accuracy of single-label classification algorithm was improved. And then, the Binary Relevance (BR) method was introduced to apply AdaBoost. FT algorithm into multi-label classification problem, and the multi-label classification algorithm based on floating threshold classifiers combination was presented, namely multi-label AdaBoost. FT. The experimental results show that the average precision of multi-label AdaBoost. FT outperforms the other three multi-label algorithms, AdaBoost. MH (multiclass, multi-label version of AdaBoost based on Hamming loss), ML-kNN (Multi-Label k-Nearest Neighbor), RankSVM (Ranking Support Vector Machine) about 4%, 8%, 11% respectively in Emotions dataset, and is just little worse than RankSVM about 3%, 1% respectively in Scene and Yeast datasets. The experimental analyses show that multi-label AdaBoost. FT can obtain the better classification results in the datasets which have small number of labels or whose different labels are irrelevant.

Key words: real AdaBoost; floating threshold; maximum likelihood principle; multi-label classification; ensemble learning; Binary Relevance (BR) method

0 引言

多标签学习的研究对于解决多义对象的学习问题有十分重要的意义, 现已逐渐成为研究的一个热点。目前已有大量的多标签学习算法涌现, 大致可以分为算法适应和问题转换两种方法^[1]。算法适应方法是通过直接改造现存的多标签学习算法, 使之能够适应多标签数据的处理, 该类方法代表性的学习算法有 ML-kNN (Multi-Label k-Nearest Neighbor)^[2]、RankSVM (Ranking Support Vector Machine)^[3]、AdaBoost. MH (multiclass, multi-label version of AdaBoost based on Hamming loss)^[4]、BoosTexter (A Boosting-based System for Text

Categorization)^[5]以及基于标签依赖的多标签学习(multi-label Learning by Exploiting Label Dependency, LEAD)^[6]等。问题转换的方法就是通过改造数据将多标签学习问题转化为其他已知的单标签学习问题进行求解, 该方法不受特定算法的限制, 目前已成熟的单标签分类算法有支持向量机、k近邻方法、贝叶斯方和提升方法等, 常用的问题转换方法有幂集法^[7]、一对多分解法^[8-9]、一对一分解法^[10]、分类器链法^[11]等。

连续 AdaBoost 算法是一种较成熟的单标签学习算法, 作为 AdaBoost 算法的扩展, 其可以输出连续的置信度, 能更好地刻画分类边界, 从而得到了广泛的应用和研究。然而, 不管

收稿日期: 2014-08-01; 修回日期: 2014-09-19。 基金项目: 四川省科技支撑计划项目(2011GZ0171; 2012GZ0106)。

作者简介: 张丹普(1986-), 女, 河南平顶山人, 博士研究生, 主要研究方向: 机器学习、模式识别; 付忠良(1967-), 男, 重庆合川人, 研究员, 博士生导师, 主要研究方向: 机器学习、模式识别; 王莉莉(1987-), 女, 河南周口人, 博士研究生, 主要研究方向: 机器学习、模式识别; 李昕(1985-), 男, 陕西汉中, 博士研究生, 主要研究方向: 图形图像处理、模式识别。

是 AdaBoost 算法还是连续 AdaBoost 算法,其得到的各个弱分类器的分段阈值都是确定的,待分类目标只能与固定的分段阈值比较。于是,当目标位于分段阈值附近时,目标值较小的变化都会导致输出两个完全不同的结果,从而导致分类结果的不稳定,这一现象显然不太合理。

文献[12-13]中,作者在连续 AdaBoost 算法的基础上,基于统计学中的极大似然原理,为解决算法对弱分类器的分段阈值敏感的问题,提出了基于浮动阈值分类器的 AdaBoost 算法(AdaBoost algorithm with floating threshold, AdaBoost. FT),即各个弱分类器的输出值不是根据分段阈值而是根据训练样本的分布而确定,该算法对各个弱分类器采取浮动分类阈值,能克服固定分段阈值分类器对分类边界附近点分类不稳定这一缺点。本文基于浮动阈值分类器的 AdaBoost 算法,研究浮动阈值分类器的训练和组合问题,通过对多个分类器的组合来提升分类精度,并采用二分类(Binary Relevance, BR)方法^[14]将该集成学习算法应用于多标签分类问题。问题转换的方法的关键是如何构造高效的单标签分类算法,从而在不考虑标签相关性的前提下,更加准确地达到分类预测的目的。因此,本文通过对数据进行分解来改造数据从而将问题转化,并将分类性能良好的基于浮动阈值分类器的 AdaBoost 单标签分类算法应用于多标签分类问题,有效地解决了多标签分类问题,并取得了良好的分类效果。

1 二分类 Real AdaBoost 算法

先考虑二分类问题,设训练样本集 $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, $y_i \in \{-1, +1\}$, $i = 1, 2, \dots, m$, $(x_i, y_i) \in S$ 简单记为 $x_i \in S$ 。连续 AdaBoost 算法流程如下。

1) 初始化权值: $\omega_i^1 = 1/m$ ($i = 1, 2, \dots, m$)。

2) DO FOR $t = 1, 2, \dots, T$

① 基于带权值 ω_i^t 的训练集 S 训练弱分类器,如下所示。

a) 对样本空间 S 进行一个 n_t 段划分: $S = S_1^t \cup S_2^t \cup \dots \cup S_{n_t}^t$, $i \neq j$ 时, $S_i^t \cap S_j^t = \emptyset$ 。

b) 计算 S_j^t 中 l 类样本权值之和: $p_l^{j,t} = \sum_{i: (x_i \in S_j^t, y_i = l)} \omega_i^t$ ($l = 1, -1$)。

c) 定义基分类器 $h_t(x): \forall x \in S$, 令 $h_t(x) = 0.5 \times \ln(p_l^{j,t}/p_{l'}^{j,t})$ ($j = 1, 2, \dots, n_t$)。

d) 选取 $h_t(x)$: 选取 $h_t(x)$ 使 $Z_t = 2 \sum_{j=1}^{n_t} \sqrt{p_l^{j,t} p_{l'}^{j,t}}$ 最小。

② 调整样本权值: $\omega_i^{t+1} = \frac{\omega_i^t}{Z_t} \exp(-y_i h_t(x_i))$ 。

3) 循环结束后强分类器为: $H(x) = \text{sgn}(f(x))$, 其中 $f(x) = \sum_{t=1}^T h_t(x)$ 。

具体使用时,为了避免零对数和零除数,一般需要引入平滑因子 δ ,即用 $h_t(x) = 0.5 \times \ln((p_l^{j,t} + \delta)/(p_{l'}^{j,t} + \delta))$ 代替 $h_t(x) = 0.5 \times \ln(p_l^{j,t}/p_{l'}^{j,t})$, 其中 δ 为一小正数。

连续 AdaBoost 算法的强分类器为各个带置信度输出的弱分类器之和,并有训练错误率估计:

$$\frac{1}{m} |\{i: H(x_i) \neq y_i\}| \leq \prod_{t=1}^T \left(2 \sum_{j=1}^{n_t} \sqrt{p_l^{j,t} p_{l'}^{j,t}} \right) \quad (1)$$

Real AdaBoost 算法的基本思想是,不断训练新的弱分类

器,然后组合它们形成强分类器从而不断降低训练错误率。通过理论分析易证明,如果各个弱分类器相互独立,则无论 AdaBoost 算法还是连续 AdaBoost 算法,都是与 Bayes 统计推断等价的。

2 二分类 AdaBoost. FT 算法

根据前面的分析,弱分类器 $h_t(x)$ 实际上对应于样本空间的一个 n_t 段划分,当目标位于划分段 S_j^t 时,根据该段内 1 类和 -1 类样本发生的概率 $p_l^{j,t}$ 与 $p_{l'}^{j,t}$,弱分类器 $h_t(x)$ 将输出 $0.5 \times \ln(p_l^{j,t}/p_{l'}^{j,t})$ 。显然,位于同一划分段内的目标,弱分类器输出值是一样的。当目标位于分段阈值附近时,目标值较小的变化将导致弱分类器输出两个不同的值,即当目标位于弱分类器的分段阈值附近时,分类结果是不稳定的。

分析上述不稳定问题的原因在于弱分类器 $h_t(x)$ 的输出值由目标位于弱分类器的划分段 S_j^t 内的两类样本的发生概率 $p_l^{j,t}$ 与 $p_{l'}^{j,t}$ 决定,当 $i \neq j$ 时, $S_i^t \cap S_j^t = \emptyset$, 忽视了阈值附近点的边界情况。为了克服这种不合理现象,一种可行办法就是使得弱分类器 $h_t(x)$ 的输出值改为由该目标值为中心的一个区域段内的两类样本的发生概率来决定,此时,当 $i \neq j$ 时,允许 $S_i^t \cap S_j^t \neq \emptyset$, $h_t(x)$ 的输出值仍然为 $h_t(x_i) = 0.5 \times \ln(p_l^{i,t}/p_{l'}^{i,t})$ 。根据统计学中的极大似然思想,已经发生的事件应该具有大的概率,如果用区域中间值来体现具有大的概率,上述做法便体现了极大似然思想。

根据前面的分析,于是可得到二分类浮动阈值分类器的 AdaBoost 算法,即二分类 AdaBoost. FT 算法流程如下。

1) 初始化权值: $\omega_i^1 = 1/m$ ($i = 1, 2, \dots, m$)。

2) DO FOR $t = 1, 2, \dots, T$

① 基于带权值 ω_i^t 的训练集 S 训练弱分类器,如下所示。

a) $\forall x_i \in S$, 定义一个区间 S_i^t ($i = 1, 2, \dots, m$)。

b) 定义 $h_t(x): \forall x_i \in S$, 令 $h_t(x_i) = 0.5 \times \ln(p_l^{i,t}/p_{l'}^{i,t})$,

$$p_l^{i,t} = \sum_{i: (x_i \in S_i^t, y_i = l)} \omega_i^t \quad (l = 1, -1)。$$

c) 选取 $h_t(x)$: 选取 $h_t(x)$ 使 $Z_t = \sum_{i=1}^m \omega_i^t (\sqrt{p_l^{i,t}/p_{l'}^{i,t}})^{-y_i}$ 最小。

② 调整样本权值: $\omega_i^{t+1} = \frac{\omega_i^t}{Z_t} \exp(-y_i h_t(x_i))$ 。

3) 循环结束后强分类器为: $H(x) = \text{sgn}(f(x))$, 其中 $f(x) = \sum_{t=1}^T h_t(x)$ 。

由第 1 章分析可推出,连续 AdaBoost 比 AdaBoost 有更小的训练错误率:

$$\prod_{t=1}^T \left(2 \sum_{j=1}^{n_t} \sqrt{p_l^{j,t} p_{l'}^{j,t}} \right) \leq 2^T \prod_{t=1}^T \sqrt{(1 - \varepsilon_t) \varepsilon_t} \quad (2)$$

需要指出的是,对连续 AdaBoost 算法,采取最小化 Z_t 的弱分类器选取策略,实际上就是最小化样本权值调整后的归一化因子。可以验证:

$$\sum_{i=1}^m \omega_i^t \exp(-y_i h_t(x_i)) = 2 \sum_{j=1}^{n_t} \sqrt{p_l^{j,t} p_{l'}^{j,t}} = Z_t \quad (3)$$

实际上是最小化 $\sum_{i=1}^m \omega_i^t \exp(-y_i h_t(x_i))$ 。

AdaBoost. FT 算法的实质也是最小化权值调整后的归一化因子 Z_t , 即可推出式(4):

$$\sum_{i=1}^m \omega_i^t \exp(-y_i h_i(x_i)) = \sum_{i=1}^m \omega_i^t (\sqrt{p_i^{j,1}/p_i^{j,-1}})^{-y_i} = Z_i \quad (4)$$

因此,理论上 AdaBoost、FT 算法与连续 AdaBoost 算法具有相同的错误率估计。但是,为了提高分类的稳定性和准确性,AdaBoost、FT 算法在上述基础上作了如下改进:

1) 主要体现在弱分类器的划分上。连续 AdaBoost 算法的 $h_i(x)$ 对应一个 n_i 段划分,即 $S = S_1^i \cup S_2^i \cup \dots \cup S_{n_i}^i$, $i \neq j$ 时, $S_i^i \cap S_j^j = \emptyset$ 。而浮动阈值分类器的 AdaBoost 算法 $h_i(x)$ 对应的不是一个划分了。 $h_i(x)$ 输出值有类似公式 $h_i(x_i) = 0.5 \times \ln(p_i^{j,1}/p_i^{j,-1})$,但 $p_i^{j,1}$ 与 $p_i^{j,-1}$ 变成包含 x_i (一般以 x_i 为中心) 的样本空间的一个子集 R_i 的正类样本累积权值与反类样本累积权值。假设每一个 x_i 得到样本空间 S 的一个子集 $S_i^i = S_1^i \cup S_2^i \cup \dots \cup S_m^i$,但 $i \neq j$ 时,允许 $S_i^i \cap S_j^j \neq \emptyset$,即 $R_i = \bigcup_k S_k$ 。

2) 上述算法能够根据 $h_i(x)$ 输出值近似得到样本空间的分布,而不只是简单的正反类样本在划分段内的分布差异。因此,算法实际上是把连续 AdaBoost 算法只能针对弱分类器的组合,扩展到了针对样本分布的组合。

算法应用的两点说明:

1) 确定目标中心区域。目标中心区域为以该目标为中心,区域宽度为给定的样本值变化幅度的一半。当目标靠近两端时,目标中心区域可选取包含该目标对应半宽区域。

2) 目标中心区域的宽度可以动态更新。以该目标值为中心,逐步向两边扩展,直至两类样本发生概率之差值最大,同时限制区域宽度不超过样本值变化幅度的一半。

3 多标签 AdaBoost、FT 算法

根据前面介绍的浮动阈值分类器用于二分类问题的分析可知,采用浮动阈值分类器模型进行训练学习的过程中需要保存每个阶段的权重值,将其应用于单标签学习问题,存储和计算的工作量相对较少,如果采用基于整体优化的方法将浮动阈值分类器模型应用于多标签学习,明显会使得分类效率低下,考虑到单标签分类问题是多标签分类问题的一个特例,因此,可以通过改造数据集,将多标签数据集分解成多个单标签数据集,进而采用单标签分类算法进行训练学习,最终达到多标签分类目的。本文采用基于问题转换方法,即 BR 方法^[14],基本思想是将多标签学习问题转化为多个独立的单标签二分类问题,其中每个二分类问题对应于标签空间的一个类别标记,于是得到基于浮动阈值分类器组合的多标签分类算法,即多标签 AdaBoost、FT 算法。

设 X 为示例空间,多标签样本集 $S = \{(x_1, Y_1), (x_2, Y_2), \dots, (x_m, Y_m)\}$, $Y_i \subseteq L$ 且 $|Y_i| \geq 2$, $L = \{y_1, y_2, \dots, y_k\}$ ($i = 1, 2, \dots, m$),其中 Y_i 为隶属于 x_i 的相关标记集合,对于第 j 个类别 y_j ($1 \leq j \leq K$) 而言,构造与该类别对应的二类训练集:

$$D_j = \{(x_i, \varphi(Y_i, y_j)) \mid 1 \leq i \leq m\} \quad (5)$$

其中:

$$\varphi(Y_i, y_j) = \begin{cases} +1, & y_j \in Y_i \\ -1, & \text{其他} \end{cases}$$

基于此,BR 方法采用基于二分类浮动阈值分类器的 AdaBoost 算法训练二分类器 $h_j: X \rightarrow \{y_j, \neg y_j\}$,对于任一多标签样本 (x_i, Y_i) ,示例 x_i 将参与 K 个二分类器的学习,其中:

对于相关标签 $y_j \in Y_i$ 而言, x_i 在构造二分类器时对应于正例;对于无关标签 $y_j \in \neg Y_i$ 而言, x_i 在构造二分类器时对应于反例。

在测试阶段,对于未见示例 x ,采用如下方式预测其类别标签集合 Y :

$$Y = \{y_j \mid h_j(x) > 0, 1 \leq j \leq K\} \quad (6)$$

多标签 AdaBoost、FT 算法除了具有浮动阈值的特性之外,也结合了 BR 方法的特点,对应标签数目具有线性复杂度。该算法每次仅仅考虑单个标签的类别标记,认为不同的类别标记之间相互独立,互不影响。因此,该方法的优点是,其不仅适合于那些不同类别标记之间基本没有关联关系的数据集,同样适合于类别标记之间关联关系不断变化的场合。但是在数据集的转换过程中,该方法完全忽略了训练数据集中类别标记之间的关联关系,得到的预测标签集可能会包含过多或者过少的类别标记,这是其不足之处。

4 实验与分析

4.1 实验方法和数据

实验使用 Matlab(R2013a) 开发平台,为了验证算法的有效性,实验数据集采用单标签数据集和多标签数据集,单标签数据集来源于 UCI 数据集的 Ionosphere、Pima、WDBC、Sonar (见表 1),多标签数据集来源于 Mulan (<http://mulan.sourceforge.net/datasets.html>) 的 Scene、Yeast、Emotions (见表 2)。

表 1 单标签实验数据集

数据名称	正类样本数	负类样本数	属性个数
Ionosphere	225	126	34
Pima	268	500	8
WDBC	357	212	32
Sonar	97	111	60

表 2 多标签实验数据集

数据名称	样本容量	属性	类别	标签基数
Scene	2407	294	6	1.074
Yeast	2417	103	14	4.237
Emotions	593	72	6	1.869

实验时对数据进行 6:4 比例随机划分得到训练数据集和测试数据集,训练 20 个弱分类器,随机实验 10 次统计平均测试错误。弱分类器基于单个属性(特征)来构造,实验中将样本空间划分为 N 段(N 可取任意数值,本实验中取 8,16,32),样本值幅度变化比率为 W (W 可取 0 到 1 任意数值,本实验中取 0.1,0.3,0.5),即样本值的变化幅度为 $N \times W$,由分析知,区域宽度为样本值变化幅度的一半,可表示为 $\theta = \text{int}(N \times W/2)$ 。下文的 AB、AB_R、AB_FT、M_FT 分别表示 AdaBoost 算法、连续 AdaBoost 算法、二分类 AdaBoost、FT 算法、多标签 AdaBoost、FT 算法。

4.2 实验结果与分析

1) 将二分类 AdaBoost、FT 算法在 UCI 的 4 个标准数据集 Ionosphere、Pima、WDBC、Sonar 上进行实验,针对不同的划分段数 N 和不同的幅度变化比率 W 分别给出了测试错误率,并将测试结果与 AdaBoost、连续 AdaBoost 的性能进行了对比(具体见表 3)。表 3 的对比实验中性能最优的结果由加黑的数字显示。由表 3 可直观得出二分类 AdaBoost、FT 算法的性能整体上优于传统的 AdaBoost 算法和连续 AdaBoost 算法。

最好情况时,在 Ionosphere 数据集上的测试错误率分别可下降 8.21% 和 10.71%, 情况较差的 WDBC 数据集上, 测试错误率分别下降 1.08% 和 0.79%。从表 3 还可得到, 当分段数较多 ($N = 32$), 幅度变化率较小 ($W = 10\%$) 时, 样本空间中的样本划分更加细致, 分类效果整体较好, 从而验证了浮动阈值分类器能够克服分类阈值附近点分类不稳定的缺陷。

表 3 几种算法在单标签数据集上测试错误率对比

Algorithm	Pima	WDBC	Ionosphere	Sonar
AB	0.263 4	0.050 7	0.161 4	0.265 1
AB_R	0.258 0	0.047 8	0.186 4	0.254 8
AB_FT($N=8, W=0.1$)	0.246 6	0.057 9	0.140 0	0.285 2
AB_FT($N=8, W=0.3$)	0.238 8	0.043 0	0.090 0	0.232 5
AB_FT($N=8, W=0.5$)	0.265 5	0.053 5	0.116 4	0.275 9
AB_FT($N=16, W=0.1$)	0.238 1	0.046 5	0.093 6	0.230 1
AB_FT($N=16, W=0.3$)	0.239 7	0.044 3	0.093 6	0.215 7
AB_FT($N=16, W=0.5$)	0.257 0	0.042 1	0.115 0	0.279 5
AB_FT($N=32, W=0.1$)	0.235 2	0.039 9	0.079 3	0.245 8
AB_FT($N=32, W=0.3$)	0.235 5	0.043 9	0.082 9	0.208 4
AB_FT($N=32, W=0.5$)	0.255 0	0.039 9	0.114 3	0.257 8

2) 将二分类 AdaBoost. FT 算法推广到多标签分类, 并在 Yeast、Emotions、Scene 三个多标签数据集上进行验证, 表 4~6 分别给出了多标签 AdaBoost. FT 算法在 3 个数据集上 5 项度量指标 (Hamming Loss、Ranking Loss、Coverage、One Error、Average Precision) 的测试结果, 取不同参数值的对比实验中性能最优的结果由加黑的数字显示。分析知, 该算法在 3 个数据集上的整体性能稳定, 当分类效果较好时, 仍然需要较多的划分段落和较小的幅度变化率。该算法有效克服了分类算法对分类阈值敏感的缺点, 达到了良好的分类效果。

表 4 几种算法在 Scene 数据集上的实验结果

Algorithm	Hamming Loss	Ranking Loss	Coverage	One Error	Average Precision
M_FT($N=8, W=0.1$)	0.122 9	0.108 3	0.632 4	0.310 1	0.814 7
M_FT($N=8, W=0.3$)	0.114 8	0.107 3	0.617 2	0.297 4	0.820 3
M_FT($N=8, W=0.5$)	0.136 3	0.133 2	0.753 7	0.361 6	0.780 4
M_FT($N=16, W=0.1$)	0.109 2	0.092 2	0.548 3	0.267 1	0.840 6
M_FT($N=16, W=0.3$)	0.109 4	0.096 2	0.561 6	0.277 5	0.834 0
M_FT($N=16, W=0.5$)	0.129 9	0.117 5	0.673 3	0.328 3	0.802 0
M_FT($N=32, W=0.1$)	0.105 3	0.087 2	0.520 5	0.259 4	0.846 7
M_FT($N=32, W=0.3$)	0.111 5	0.098 3	0.573 0	0.286 2	0.829 2
M_FT($N=32, W=0.5$)	0.124 7	0.116 3	0.667 3	0.328 3	0.802 6

表 5 几种算法在 Yeast 数据集上的实验结果

Algorithm	Hamming Loss	Ranking Loss	Coverage	One Error	Average Precision
M_FT($N=8, W=0.1$)	0.225 9	0.196 9	6.917 1	0.286 9	0.721 5
M_FT($N=8, W=0.3$)	0.207 1	0.178 6	6.578 9	0.240 1	0.749 5
M_FT($N=8, W=0.5$)	0.207 3	0.179 3	6.514 8	0.236 0	0.745 8
M_FT($N=16, W=0.1$)	0.209 2	0.178 5	6.627 1	0.244 9	0.748 2
M_FT($N=16, W=0.3$)	0.203 8	0.173 6	6.492 5	0.233 7	0.755 5
M_FT($N=16, W=0.5$)	0.208 1	0.180 1	6.500 3	0.245 3	0.741 4
M_FT($N=32, W=0.1$)	0.214 5	0.184 4	6.666 4	0.263 9	0.738 5
M_FT($N=32, W=0.3$)	0.201 6	0.162 5	6.434 3	0.233 5	0.753 8
M_FT($N=32, W=0.5$)	0.207 9	0.181 1	6.521 4	0.241 4	0.742 9

3) 表 7~9 显示了多标签 AdaBoost. FT 算法与现有常用的多标签分类算法 AdaBoost. MH、ML- k NN($k=7$)、RankSVM

的性能比较, 分别在 3 个数据集上针对 5 项度量指标进行测试结果对比。参数选取时, 当 $k=7$ 时 ML- k NN 算法性能最优, RankSVM 算法在 3 个数据集上分别选取不同的核函数可取得最好的分类性能。加黑的数字显示了表中 4 种算法的对比实验中性能最优的结果。

表 6 几种算法在 Emotions 数据集上的实验结果

Algorithm	Hamming Loss	Ranking Loss	Coverage	One Error	Average Precision
M_FT($N=8, W=0.1$)	0.259 8	0.212 4	2.050 6	0.357 0	0.744 4
M_FT($N=8, W=0.3$)	0.219 7	0.179 6	1.886 1	0.308 0	0.779 3
M_FT($N=8, W=0.5$)	0.228 8	0.194 5	1.938 4	0.326 6	0.767 6
M_FT($N=16, W=0.1$)	0.226 6	0.188 2	1.922 4	0.313 1	0.774 2
M_FT($N=16, W=0.3$)	0.216 9	0.173 8	1.833 8	0.297 9	0.784 9
M_FT($N=16, W=0.5$)	0.225 0	0.192 0	1.951 1	0.321 5	0.767 7
M_FT($N=32, W=0.1$)	0.232 9	0.187 2	1.912 2	0.336 7	0.764 6
M_FT($N=32, W=0.3$)	0.210 2	0.172 5	1.838 8	0.297 0	0.787 7
M_FT($N=32, W=0.5$)	0.222 6	0.191 0	1.959 5	0.304 6	0.774 5

表 7 几种算法在 Scene 数据集上算法性能比较

Algorithm	Hamming Loss	Ranking Loss	Coverage	One Error	Average Precision
M_FT($N=32, W=0.1$)	0.105 3	0.087 2	0.520 5	0.259 4	0.846 7
AdaBoost. MH	0.124 6	0.128 7	0.713 9	0.348 1	0.789 9
ML- k NN($k=7$)	0.274 8	0.272 2	2.334 6	0.405 5	0.697 2
RankSVM(RBF)	0.095 5	0.076 9	0.487 0	0.205 6	0.875 5

表 8 几种算法在 Yeast 数据集上算法性能比较

Algorithm	Hamming Loss	Ranking Loss	Coverage	One Error	Average Precision
M_FT($N=32, W=0.3$)	0.201 6	0.162 5	6.434 3	0.233 5	0.753 8
AdaBoost. MH	0.185 9	0.214 8	6.686 9	0.256 3	0.737 1
ML- k NN($k=7$)	0.197 3	0.168 4	6.302 7	0.239 0	0.760 6
RankSVM(Poly)	0.196 2	0.225 7	6.717 0	0.179 3	0.763 5

表 9 几种算法在 Emotions 数据集上算法性能比较

Algorithm	Hamming Loss	Ranking Loss	Coverage	One Error	Average Precision
M_FT($N=32, W=0.3$)	0.210 2	0.172 5	1.838 8	0.297 0	0.787 7
AdaBoost. MH	0.211 9	0.240 8	2.011 6	0.358 2	0.748 8
ML- k NN($k=7$)	0.273 6	0.271 4	2.355 1	0.398 7	0.700 0
RankSVM(Linear)	0.287 6	0.306 2	2.379 7	0.459 9	0.679 1

从表 7~9 中结果可直观得到, 本文算法在各项指标上均有较好的性能体现, 并且测试结果相对稳定, 特别是在 Emotions 上的各项指标性能都明显优于其他 3 种算法, 在 Scene 上的分类性能略差于 RankSVM 取 RBF(Radial Basis Function) 核函数时的情况, 但明显优于另两种算法。尽管在 Yeast 数据集上的表现略差, 但多数指标结果依然存在优势。例如本文算法的 Average Precision 结果在 Emotions 数据集上比 AdaBoost. MH、ML- k NN、RankSVM 这 3 种算法分别提高约 4%、8%、11%; 在 Scene、Yeast 数据集上仅比 RankSVM 低约 3%、1%。其主要原因是, 多标签 AdaBoost. FT 算法采用的是 BR 方法, 认为不同类别标记之间是相互独立的, 忽视了标签之间的相关性, 当标签数量较多时, 标签之间相关的概率较大, 分类效果必然会受到影响, 这也是该算法存在的不足之处。

5 结语

针对多标签分类问题,本文提出了一种基于问题转换的多标签分类算法。该算法运用BR方法的思想,将多标签分类问题转化成多个单标签问题,采用基于浮动阈值分类器的单标签算法,通过解决算法对弱分类器的固定分段阈值敏感的问题,使分类边界附近点分类更加准确。因此,多标签AdaBoost.FT算法在不考虑标签相关性的前提下,能够有效地达到分类预测的目的。实验中将本文算法与现有多标签分类算法进行对比,实验结果表明,该算法在3个数据集上的分类性能稳定,准确性相对较高。在今后的工作中,将结合标签间的相关性^[15]进行深入研究,使得多标签分类更加准确。

参考文献:

- [1] TSOU MAKAS G, KATAKIS I. Multi-label classification: an overview [J]. *International Journal of Data Warehousing and Mining*, 2007, 3(3): 1-13.
- [2] ZHANG M, ZHOU Z. A k -nearest neighbor based algorithm for multi-label classification [C]// *Proceedings of the 2005 IEEE International Conference on Granular Computing*. Piscataway: IEEE, 2005: 718-721.
- [3] ELISSEEFF A, WESTON J. A kernel method for multi-labelled classification [C]// *NIPS 2001: Proceedings of the 2001 Neural Information Processing Systems: Natural and Synthetic*. Cambridge: MIT Press, 2001: 681-687.
- [4] SCHAPIRE R E, SINGER Y. Improved boosting algorithms using confidence-rated predictions [J]. *Machine Learning*, 1999, 37(3): 297-336.
- [5] SCHAPIRE R E, SINGER Y. BoosTexter: a boosting based system for text categorization [J]. *Machine Learning*, 2000, 39(2/3): 135-168.
- [6] ZHANG M, ZHANG K. Multi-label learning by exploiting label dependency [C]// *KDD10: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2010: 999-1007.
- [7] TROHIDIS K, TSOU MAKAS G, KALLIRIS G, *et al.* Multi-label classification of music into emotions [C]// *ISMIR 2008: Proceedings of the 9th International Conference on Music Information Retrieval*. Philadelphia: [s. n.], 2008: 325-330.
- [8] BOUTELL M R, LUO J, SHEN X, *et al.* Learning multi-label scene classification [J]. *Pattern Recognition*, 2004, 37(9): 1757-1771.
- [9] LI J. A fast multi-label classification algorithm based on double label support vector machine [D]. Nanjing: Nanjing Normal University, 2010. (李佳阳. 基于双标签支持向量机的快速多标签分类算法[D]. 南京: 南京师范大学, 2010.)
- [10] WAN S. A fast multi-label classification algorithm based on binary and triple class support vector machines [D]. Nanjing: Nanjing Normal University, 2008. (万书鹏. 基于两类和三类支持向量机的快速多标签分类算法[D]. 南京: 南京师范大学, 2008.)
- [11] READ J, PFAHRINGER B, HOLMES G, *et al.* Classifier chains for multi-label classification [J]. *Machine Learning*, 2011, 85(3): 333-359.
- [12] FU Z, ZHANG D, ZHAO X, *et al.* AdaBoost algorithm with floating threshold [C]// *ACAI 2012: Proceedings of the 2012 International Conference on Automatic Control and Artificial Intelligence*. Stevenage: IET, 2012: 349-354.
- [13] FU Z, ZHAO X, YAO Y, *et al.* A ensemble algorithm based on floating threshold classifiers: China, CN102163239A [P]. 2011-08-24. (付忠良, 赵向辉, 姚宇, 等. 一种基于浮动分类阈值的分类器集成方法: 中国, CN102163239A [P]. 2011-08-24.)
- [14] BOUTELL M R, LUO J, SHEN X, *et al.* Learning multi-label scene classification [J]. *Pattern Recognition*, 2004, 37(9): 1757-1771.
- [15] ZHANG Y, YEUNG D Y. Multilabel relationship learning [J]. *ACM Transactions on Knowledge Discovery from Data*, 2013, 7(2): 1-31.
- [8] FOX C R, CLEMEN R T. Subjective probability assessment in decision analysis: partition dependence and bias toward the ignorance prior [J]. *Management Science*, 2005, 51(9): 1417-1432.
- [9] SPEIRS-BRIDGE A, FIDLER F, McBRIDE M, *et al.* Reducing overconfidence in the interval judgments of experts [J]. *Risk Analysis*, 2010, 30(3): 512-523.
- [10] DEMPSTER A P. Upper and lower probabilities induced by a multiple valued mapping [J]. *The Annals of Mathematical Statistics*, 1967, 38(2): 325-339.
- [11] SHAFER G. A mathematical theory of evidence [M]. Princeton: Princeton University Press, 1976: 1-314.
- [12] BOUKHRIS I, ELOUEDI Z, BENFERHAT S M. Dealing with external actions in belief causal networks [J]. *International Journal of Approximate Reasoning*, 2013, 54(8): 978-999.
- [13] YANG C, LI H. An evidence reasoning model with its application to expert opinions combination [J]. *Systems Engineering—Theory and Practice*, 2001, 21(4): 43-48. (杨春, 李怀祖. 一个证据推理模型及其在专家意见综合中的应用[J]. *系统工程理论与实践*, 2001, 21(4): 43-48.)
- [14] BEYNON M. DS/AHP method: a mathematical analysis, including an understanding of uncertainty [J]. *European Journal of Operational Research*, 2002, 140(1): 148-164.
- [15] NIELSEN S H, NIELSEN T D. Adapting Bayesian network structures to non-stationary domains [J]. *International Journal of Approximate Reasoning*, 2008, 49(2): 379-397.
- [16] YAO S, GUO Y, HUANG W. An improved method of aggregation in DS/AHP for multi-criteria group decision-making based on distance measure [J]. *Control and Decision*, 2010, 25(6): 894-898. (姚爽, 郭亚军, 黄玮强. 基于证据距离的改进 DS/AHP 多属性群决策方法[J]. *控制与决策*, 2010, 25(6): 894-898.)
- [17] JU Y, WANG A. Emergency alternative evaluation under group decision makers: a method of incorporating DS/AHP with extended TOPSIS [J]. *Expert Systems with Applications*, 2012, 39(1): 1315-1323.
- [18] WU H, WANG W, YANG F. Structure learning method of Bayesian network with prior information [J]. *Systems Engineering and Electronics*, 2012, 34(12): 2585-2591. (吴红, 王维平, 杨峰. 融合先验信息的贝叶斯网络结构学习方法[J]. *系统工程与电子技术*, 2012, 34(12): 2585-2591.)

(上接第146页)