

# 基于标签相关性的多标签分类 AdaBoost 算法

王莉莉<sup>1,2</sup>, 付忠良<sup>1,2</sup>

(1. 中国科学院 成都计算机应用研究所, 四川 成都 610041; 2. 中国科学院大学, 北京 100049)

**摘要:**在多标签分类问题中,标签之间往往是相关的,为了提高分类性能,利用标签之间的相关性,提出 AdaBoost. MLR 算法和标签相关性分析方法。AdaBoost. MLR 算法采用余弦相似度来计算标签相关性矩阵,利用标签相关性矩阵对原始标签矩阵进行补全转换为模糊标签矩阵,将标签空间划分为标签集、标签相关集和标签无关集,结合标签之间的相关性和弱分类器的分类情况,对样本权重进行调整。AdaBoost. MLR 算法也能解决多类别分类问题,在其标签相关性的计算中,根据已经训练的弱分类器得到的临时强分类器的分类结果,构造标签相似性矩阵。实验结果表明,文中提出的算法在实验数据集上优于现有的算法,尤其在标签相关性复杂的数据集上分类性能有显著提升。

**关键词:**标签相关性;多标签分类;多分类问题;AdaBoost 算法;分类器组合

中图分类号:TP391

文献标志码:A

## Multi-label AdaBoost Algorithm Based on Label Correlations

WANG Lili<sup>1,2</sup>, FU Zhongliang<sup>1,2</sup>

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

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

**Abstract:** In order to improve classification performance and exploit label correlations, AdaBoost. MLR algorithm was proposed. Cosine similarity was adopted to capture the complex correlations among labels in AdaBoost. MLR algorithm, a supplementary label matrix was incorporated, which augments the incomplete original label matrix by exploiting the label correlations, label space was divided into three parts of label set, relevant label set and irrelevant label set, weight-update rule was modified according to correlations among labels and the results of weak learner. AdaBoost. MLR algorithm was able to solve multi-class classification problem specially, label similarity matrix, instead of cosine similarity, was constructed by the classification results of temporary strong learner combined by previous trained weak learners. The experimental results illustrated that the proposed algorithm was superior to existing algorithms, and the classification performance was improved significantly on datasets had complex correlations among labels.

**Key words:** ensemble learning; label correlation; multi-label classification; multi-class classification; AdaBoost algorithm; classifier combination

在多标签分类问题的研究过程中,大多数算法可分为2类:问题转化方法和算法自适应方法。问题转化方法通常将多标签分类问题转化为二分类问题<sup>[1-2]</sup>和多分类问题<sup>[3]</sup>等。BR (binary relevance) 算法<sup>[4]</sup>,通过把多标签分类问题转化为一系列相互独立的二分类问题,其中每个标签对应一个二分类

器,该方法直观高效,但仅适用于标签密度低的数据集,然而在很多领域中标签是有树状的层次关联的, BR 算法没有考虑标签之间的关联性,局限性比较大。CC (classifier chains) 方法<sup>[5]</sup>克服了这一缺点,将基分类器串联起来,使得标签信息在分类器之间传递,但误差信息也会随着分类器链传播,造成分类

收稿日期:2015-10-23

基金项目:四川省科技支撑计划资助项目(2012GZ0106);中科院西部之光人才培养计划资助项目;四川省科技创新苗子工程资助项目(2015060)

作者简介:王莉莉(1987—),女,博士生.研究方向:机器学习;模式识别;图像处理. E-mail: wanglili8773@163.com

网络出版时间:2016-9-9 10:48:18 网络出版地址: <http://www.cnki.net/kcms/detail/51.1596.T.20160909.1048.002.html>

结果存在不确定性和不稳定性。LP 方法的思路是将每个多标签样本的所属标签联合起来创建新的标签,但新标签的数量会大大增加。RAkEL<sup>[6]</sup>算法是基于 LP 分类器的集成算法,考虑了标签之间的相互关系,避免了 LP 的缺陷。算法自适应方法是改编一些已知的算法来直接处理多标签分类问题<sup>[7]</sup>,MLkNN 算法<sup>[8]</sup>的基本思想是采用 k 近邻分类准则,统计近邻样本的类别标签信息,通过最大化后验概率的方式推理未知样本的标签集合。该方法简单、错误率低,但需要计算两两样本之间的距离,存储空间占用较大。RankSVM 算法<sup>[9]</sup>是一种用 SVM 思想解决多标签问题的算法,定义了一个 Ranking Loss 的代价函数以及相应的边界,在试图降低 Ranking Loss 的同时保持边界最优。这 2 种算法都需要占用较大的内存空间,不适用于样本空间大或复杂的数据集。BPMLL<sup>[10]</sup>是修改流行的 BP 反向传播算法来适应多标签数据的一个算法,主要是引入了新的误差公式来考虑多标签。

文献[11]提出一种直接的多分类 AdaBoost 算法,弱分类器直接给每个样本分配  $K$  个可能标签中的一个,要求弱分类错误率小于  $1/K$ ,  $K$  为类别个数,但当类别个数很大时,对弱分类器分类性能要求过高。AdaBoost. M2 算法<sup>[12]</sup>在 AdaBoost. M1 的基础上,采用伪损失代替错分率来衡量分类精度,每一迭代回合的弱分类器不只是注重错分样本,还注重错分的类别,使弱分类器也将注意力放在难以辨别的标签上,并且对弱分类器的分类精度要求有所降低,只要伪损失稍比随机猜测好就可以,但是该算法计算比较复杂,成本和时间也相应增加。AdaBoost. MH<sup>[13]</sup>是一种直接将  $K$  类问题转化为  $K$  个两类问题的算法,该算法在现有的多类分类算法中准确率较高,弱分类器的精度要求可以较容易地得到满足。然而这些算法都没有考虑标签之间的相关性。

文献[14]基于 AdaBoost 思想,给予每个样本 2 个不同的权重和不同的权重调整因子,从一定程度上考虑了标签之间的相关性。文献[15]针对多标签 AdaBoost 算法提出了 2 种改进思路,一种思路是修改算法的样本权重调整策略;另一种思路是在训练弱分类器时兼顾后续待学习的弱分类器对学习错误的影响。虽然这 2 种改进算法在弱分类器的迭代训练中兼顾了弱分类器对各个标签的分类性能,但固定的权值调整因子容易造成不同类别的分类性能差异很大。

多标签分类问题的特征是标签伴随出现,多分

类问题是多标签分类问题的特例,但每个样本只有一个标签,因此,对这两种分类问题的标签相关性进行考虑时,需要采用不同的方法进行分析。对于多标签分类问题,采用余弦相似性计算标签相关性矩阵,并利用标签相关性矩阵对原始标签进行补全,不仅能考虑 1 阶标签相关性,还能考虑更复杂的高阶标签相关性;对于多分类问题,如果分类器对于某两个标签错分率比较高,则可认为这两个标签之间具有相似性。通过标签相关性分析后,将标签空间划分为标签集、标签相关集和标签无关集,并结合标签之间相关性和弱分类器的分类情况动态给出样本权重调整因子,对 AdaBoost. MH 算法进行改进。

## 1 标签相关性分析

### 1.1 多标签分类问题描述

相比于单标签分类,多标签分类问题更符合真实世界。多标签分类问题描述如下。

**定义 1** 设一个包含  $m$  个样本的数据集  $X = \{(x_1, Y_1), \dots, (x_m, Y_m)\}$ , 候选标签集合  $Y = \{1, 2, \dots, K\}$ , 样本  $x_i$  所对应的标签集合  $Y_i \subseteq Y$ , 对于样本  $x_i$ , 对应的标签向量为  $(Y_i(1), \dots, Y_i(l), \dots, Y_i(K))$ , 若  $l \in Y_i$ , 则  $Y_i(l) = 1$ , 否则,  $Y_i(l) = 0$ 。

### 1.2 余弦相似度分析标签相关性

对于样本  $x_i$ , 集合  $Y$  被划分为标签集  $Y_i$  和非标签集  $N_i$ , 事实上, 在非标签集合  $N_i$  中存在一些标签与标签集  $Y_i$  有很大的相关性。例如, “大海” 标签可能伴随“沙滩” 标签出现, 认为是 1 阶相关的。一幅图像包含标签“狮子” 和“草原”, 那么该图片包含标签“非洲” 的可能性就会比较大, 认为是高阶相关的。因此, 标签相关程度可以由余弦相似性度量。

在采用余弦相似度来计算标签相关性矩阵时, 令  $\mathbf{y}(l) = (Y_1(l), \dots, Y_m(l))$ , 假设标签  $l_1 \in Y_i, l_2 \notin Y_i$ , 那么  $l_1$  和  $l_2$  的余弦相似度为  $\text{sim}(l_1, l_2) = \frac{\mathbf{y}(l_1) \cdot \mathbf{y}(l_2)}{\|\mathbf{y}(l_1)\| \|\mathbf{y}(l_2)\|}$ , 计算方式如图 1 所示。

$$L = \begin{matrix} & \mathbf{y}(1) & \mathbf{y}(2) & \dots & \mathbf{y}(K) \\ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} & \dots & \dots & \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \end{matrix} \longrightarrow \text{sim}(l_1, l_2) = \frac{\mathbf{y}(l_1) \cdot \mathbf{y}(l_2)}{\|\mathbf{y}(l_1)\| \|\mathbf{y}(l_2)\|}$$

图 1 标签之间余弦相似性

Fig. 1 Cosine similarity among labels  
利用标签相关性矩阵  $\text{sim}$  将原始标签矩阵  $L$  转

换为模糊标签矩阵  $\hat{Y}$ , 模糊标签矩阵通过如下方式构造:

$$\hat{Y} = L \times sim \quad (1)$$

通过模糊标签矩阵, 可以将样本  $x_i$  的标签空间划分为标签集  $Y_i$ 、标签相关集  $R_i$  和标签无关集  $U_i$ , 划分方法如下: 对于标签  $l \notin Y_i$ , 如果  $\hat{y}(i, l) > thresh$ , 则令  $l \in R_i$ , 否则, 令  $l \in U_i$ 。

### 1.3 多类别分类问题描述

多类别问题是二分类问题的扩展, 每个样本对应多个类别中的一个标签。多标签分类是多类别问题的扩展, 每个样本对应多个类别中的某几个标签。大多数多标签分类算法都能解决多类别问题, 那么引入标签相关性的 MLR 算法能否解决多类别分类问题? 对多类别分类来说, 样本只能包含一个标签, 不能采用余弦相似性来度量标签之间的关系, 但有些标签之间存在很强的“混淆”度, 比如“日出”场景和“夕阳”场景具有很大的相似性, 分类器很容易错分。本文根据分类器分类结果, 将“混淆”程度高的 2 个标签认为是标签相关的。

### 1.4 分类器分类结果分析标签相关性

首先调用分类器并统计分类结果  $C = \{C(l_1, l_2)\}_{K \times K}$ ,  $l_1 \neq l_2$ , 其中,  $C(l_1, l_2)$  表示分类器把标签为  $l_1$  的样本错分为  $l_2$  的个数; 然后计算分类器把标签为  $l_1$  的样本错分为  $l_2$  的概率:  $P(l_1, l_2) = C(l_1, l_2) / \text{sum}(C(l_1, :))$ ; 最后, 将样本  $x_i$  的标签空间划分为标签集  $Y_i$ 、标签相关集  $R_i$  和标签无关集  $U_i$ , 划分方法如下: 对于标签  $l_1 \in Y_i, l_2 \notin Y_i$ , 如果  $P(l_1, l_2) > thresh$ , 则令  $l_2 \in R_i$ , 否则令  $l_2 \in U_i$ 。

## 2 基于标签相关性的多标签分类 AdaBoost 算法——AdaBoost. MLR

### 2.1 AdaBoost. MLR 算法流程

输入: 训练集  $X = \{(x_1, Y_1), \dots, (x_m, Y_m)\}$ , 样本权重  $D$ , 弱分类器  $h: X \times Y \rightarrow \mathbb{R}$ , 迭代次数  $T$ 。

初始化:  $D_1(i, l) = 1/(mK)$ , 其中,  $i = 1, \dots, m$ ,  $l = 1, \dots, K$ 。

For  $t = 1, \dots, T$

1) 根据样本分布  $D_t$ , 训练弱分类器  $h_t: X \times Y \rightarrow \mathbb{R}$ 。

2) 根据第 1 节所述方法对每个样本  $x_i$ , 将样本空间划分为  $Y_i, R_i$  和  $U_i$ 。

3) 对权值调整因子  $\beta_t$  赋值: 如果  $l \in R_i$ , 令  $\beta_t(i, l) = c_2$ ; 否则,  $\beta_t(i, l) = c_1$ 。其中,  $c_1, c_2 > 0$ 。

4) 计算弱分类器权重  $\alpha_t$ 。

5) 更新权重: 如果  $l \in Y_i$ ,

$$D_{t+1}(i, l) = \frac{D_t(i, l) \exp(-\alpha_t h_t(x_i, l) \beta_t(i, l))}{Z_t},$$

否则,

$$D_{t+1}(i, l) = \frac{D_t(i, l) \exp(\alpha_t h_t(x_i, l) \beta_t(i, l))}{Z_t},$$

其中,  $Z_t = \sum_{i, l} D_t(i, l) \exp(-\alpha_t Y_i(l) h_t(x_i, l) \beta_t(i, l))$ 。

输出: 强分类器

$$H(x, l) = \begin{cases} 1, & \text{if } \sum_{i=1}^T \alpha_i h_i(x, l) > 0; \\ 0, & \text{else.} \end{cases}$$

### 2.2 训练误差有界性验证

引理 令  $f'(x_i, l) = \sum_{i=1}^T \alpha_i h_i(x_i, l) \beta_i(i, l)$ ,

$$H(x_i, l) = \begin{cases} 1, & \text{if } \sum_{i=1}^T \alpha_i h_i(x_i, l) \beta_i(i, l) > 0 \\ 0, & \text{else} \end{cases}, \text{那么对}$$

于  $\forall x_i \in X$ , 有  $H'(x_i, l) = Y_i(l) \Rightarrow H(x_i, l) = Y_i(l)$ 。

证明: 若  $H'(x_i, l) = Y_i(l)$  成立, 根据上文假设, 存在 2 种情况:

1)  $H'(x_i, l) = 1$ , 且  $Y_i(l) = 1$ ;

2)  $H'(x_i, l) = 0$ , 且  $Y_i(l) = 0$ 。

以下对 2 种情况分别给予证明。

情况 1) 由于  $H'(x_i, l) = 1$ , 且  $Y_i(l) = 1$ , 则有

$\sum_{i=1}^T \alpha_i h_i(x, l) \beta_i(i, l) > 0$ , 此时  $l \in Y_i$ , 则根据算法中权值调整因子规则有  $\beta_i(i, l) = c_1$ , 所以有  $\sum_{i=1}^T \alpha_i h_i(x, l) > 0$ , 因此有  $H(x_i, l) = Y_i(l)$ 。

情况 2) 由于  $H'(x_i, l) = 0$ , 且  $Y_i(l) = 0$ , 则有

$\sum_{i=1}^T \alpha_i h_i(x, l) \beta_i(i, l) < 0$ , 此时  $l \notin Y_i$ , 若  $l \in R_i$ , 则  $\beta_i(i, l) = c_2$ , 否则,  $\beta_i(i, l) = c_1$ , 所以有  $\sum_{i=1}^T \alpha_i h_i(x, l) < 0$ , 因此有  $H(x_i, l) = Y_i(l)$ 。

综上所述,  $\forall x_i \in X$ , 有

$$H'(x_i, l) = Y_i(l) \Rightarrow H(x_i, l) = Y_i(l) \quad (2)$$

定理

$$\varepsilon = \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \llbracket H(x_i, l) \neq Y_i(l) \rrbracket \leq d \prod_i Z_i.$$

其中,  $\llbracket \pi \rrbracket = \begin{cases} 1, & \text{if } \pi \text{ 成立;} \\ 0, & \text{else.} \end{cases}$

证明:由引理可知,

$$\varepsilon = \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \llbracket H(x_i, l) \neq Y_i(l) \rrbracket \leq \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \llbracket H'(x_i, l) \neq Y_i(l) \rrbracket \quad (3)$$

若  $H'(x_i, l) \neq Y_i(l)$  成立,则存在 2 种情况:

1)  $H'(x_i, l) = 1$ , 且  $Y_i(l) = 0$ ;

2)  $H'(x_i, l) = 0$ , 且  $Y_i(l) = 1$ 。

以下对 2 种情况分别给予证明。

情况 1) 由  $Y_i(l) = 0$ , 知  $l \notin Y_i$ 。由权值更新公式得:

$$D_T(i, l) = \frac{D_1(i, l) \exp(\sum_t \alpha_t h_t(x_i, l) \beta_t(i, l))}{\prod_t Z_t} \quad (4)$$

所以,

$$\exp(\sum_t \alpha_t h_t(x_i, l) \beta_t(i, l)) = \frac{D_T(i, l) \prod_t Z_t}{D_1(i, l)} \quad (5)$$

由  $H'(x_i, l) = 1$ , 知  $\sum_{i=1}^T \alpha_t h_t(x, l) \beta_t(i, l) > 0$ ,

所以,

$$\varepsilon \leq \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \exp(\sum_t \alpha_t h_t(x_i, l) \beta_t(i, l)) = \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \frac{D_T(i, l) \prod_t Z_t}{D_1(i, l)} \quad (6)$$

情况 2) 由  $Y_i(l) = 1$  知,  $l \in Y_i$ 。由权值更新公式得:

$$D_T(i, l) = \frac{D_1(i, l) \exp(\sum_t -\alpha_t h_t(x_i, l) \beta_t(i, l))}{\prod_t Z_t} \quad (7)$$

所以,

$$\exp(\sum_t -\alpha_t h_t(x_i, l) \beta_t(i, l)) = \frac{D_T(i, l) \prod_t Z_t}{D_1(i, l)} \quad (8)$$

由  $H'(x_i, l) = 0$ , 知  $\sum_{i=1}^T \alpha_t h_t(x, l) \beta_t(i, l) < 0$ ,

所以,

$$\varepsilon \leq \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \exp(\sum_t -\alpha_t h_t(x_i, l) \beta_t(i, l)) = \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \frac{D_T(i, l) \prod_t Z_t}{D_1(i, l)} \quad (9)$$

综上所述,有:

$$\varepsilon = \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \llbracket H(x_i, l) \neq Y_i(l) \rrbracket \leq d \prod_t Z_t \quad (10)$$

其中,  $d = \frac{1}{mK} \sum_{i=1}^m \sum_{l=1}^K \frac{D_T(i, l)}{D_1(i, l)}$ 。

### 2.3 如何选择 $\alpha_t$

根据定理,在每轮迭代训练中,如果  $Z_t$  达到最小值,则训练误差可以认为达到最小值,根据文献[13]中  $\alpha_t$  的证明,假设  $h_t: X \times Y \rightarrow [-1, 1]$ , 且  $\beta_t(i, l) \in [0, 1]$ , 由于

$$Z_t = \sum_{i,l} D_t(i, l) \exp(-\alpha_t u_t(x_i, l)) \leq \frac{1+r_t}{2} e^{-\alpha_t} + \frac{1-r_t}{2} e^{\alpha_t},$$

其中:

$$u_t(x_i) = \begin{cases} h_t(x_i, l) \beta_t(i, l), & \text{if } l \in Y_i; \\ -h_t(x_i, l) \beta_t(i, l), & \text{else.} \end{cases}$$

$$r_t = \sum_{i,l} D_t(i, l) u_t(x_i, l), \text{ 令 } \alpha_t = \frac{1}{2} \ln\left(\frac{1+r_t}{1-r_t}\right),$$

由文献[13]证明可知  $Z_t \leq \sqrt{1-r_t^2} \leq 1$ 。

所以在选择弱分类器时,选择能使得  $r_t = \sum_i D_t(i, l) u_t(x_i)$  最大的弱分类器  $h_t$ , 且随着弱分类器个数的增加,训练误差会逐渐趋向于 0。

## 3 实验与分析

在对 AdaBoost. MLR (MLR)) 算法的有效性以及与主流算法 (AdaBoost. MH (MH)、MLkNN、Rank-SVM (RSVM)) 的对比分析中,采用了 3 个多标签分类数据集和 3 个多类别分类 UCI 数据集。

### 3.1 多标签分类数据集实验

MLR 算法在对多标签分类问题标签之间相关性的分析中,采用余弦相似性来度量,具体见第 1.1 节。实验数据集 Image、Scene 和 Emotions 来自于 Mulan, 见表 1。

表 1 多标签数据集

Tab. 1 Multi-label data sets

数据集	样本容量	属性个数	标签个数	标签基数	标签密度
Emotions	593	72	6	1.869	0.311
Image	2 000	294	5	1.236	0.247
Scene	2 407	294	6	1.074	0.179

对于 Image 数据集,其包含 5 个标签 desert、mountains、sea、sunset、trees。其中, mountains 和 sea 经常同时出现,其相关度为 0.22; sea 和 sunset 经常同时出现,其相关度为 0.34; 其余几乎可以认为标签独立。

在对 Scene 数据集的标签相关性分析后发现,标签 field 和标签 mountain 的相关性比较大。包含 mountain 标签的图像大约有 16% 均包含 field 标签,其余标签如 beach、sunset、fallfoliage、urban 几乎可以认为是标签独立的。

Emotions 数据集标签之间的关系比较复杂,其包含 6 个标签 amazed、happy、relaxing、quiet、sad、angry。经过分析发现,amazed 标签会伴随 happy 标签或者 angry 标签;relaxing 标签、quiet 标签和 sad 标签经常两两同时出现;而 amazed 标签和 quiet 标签不同时出现。如图 2 所示,样本标签经过相关性矩阵转换后被分为标签集  $Y_i$ 、标签相关集  $R_i$  和标签无关集  $U_i$ 。

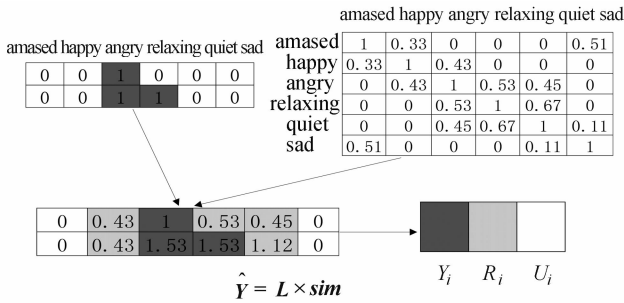


图2 标签转换流程

Fig.2 Label transformation process

包含 relaxing 标签和 quiet 标签的样本,其标签向量为(0,0,1,1,0,0),经过转换得到模糊标签向量(0,0.43,1.53,1.53,1.12,0),说明该样本包含 happy 标签的置信度为 0.43,包含 sad 标签的置信度为 1.12。样本空间被划分为标签集 {relaxing, quiet}、标签相关集 {sad, happy}、标签无关集 {angry}。说明 MLR 算法不仅能考虑 1 阶标签相关性,还能考虑更复杂的高阶标签相关性。

在对 MLR 算法和其他算法的对比实验中,采用 5 种评价方式,包括 Hamming Loss (HL)、One-error (One-err)、Coverage (C)、Ranking Loss (RL)、Average Precision (Ave\_pre),具体计算方式参考文献 [10],其中 Average Precision 越高越好,其余评价方式均越低越好。实验采取十折交叉验证方法,实验 20 次取平均值,弱分类器选择 stump 决策树。训练 60 个弱分类器进行集成学习,实验结果见表 2。

表2 多标签分类实验对比结果

Tab.2 Comparison results of multi-label classification

算法	MLR	MH	MLkNN	RSVM	
HL	0.214 6	0.232 5	0.259 7	0.263 1	
One-err	0.295 4	0.318 4	0.381 4	0.410 7	
Emotions	C	1.851 0	1.922 9	2.342 3	2.355 2
	RL	0.178 7	0.193 0	0.268 3	0.267 6
	Ave_pre	0.782 0	0.768 2	0.702 9	0.698 4
HL	0.090 6	0.110 3	0.094 0	0.097 9	
One-err	0.259 1	0.285 1	0.241 6	0.259 0	
Scene	C	0.527 7	0.577 2	0.493 7	0.491 7
	RL	0.092 8	0.101 5	0.082 6	0.077 6
	Ave_pre	0.844 8	0.834 5	0.860 7	0.848 1
HL	0.178 4	0.196 7	0.187 7	0.181 0	
One-err	0.338 6	0.370 7	0.337 8	0.388 9	
Image	C	0.977 6	1.052 4	1.018 0	0.978 3
	RL	0.178 4	0.196 5	0.188 2	0.194 7
	Ave_pre	0.782 7	0.761 2	0.786 5	0.761 6

数据表明,MLR 算法在 5 个评价标准上的分类性能均优于 MH 算法。在 Emotions 数据集上,Ranking Loss 下降了 7.4% 左右;Image 数据集上的 Hamming Loss 降低了 9.3% 左右;Scene 数据集上 Hamming Loss 降低了 17.8% 左右。相比于 MLkNN 算法和 RSVM 算法,MLR 算法在 Image 数据集和 Scene 数据集上提升不明显,是因为这 2 个数据集标签之间相关性比较简单,大多数标签之间是无关的。Emotions 数据集标签之间的相关性比较复杂,MLR 由于考虑标签之间相关性,算法发挥了优势,在各个指标上的提升都很显著。

图 3 为 MLR 和 MH 在 Emotions 数据集上的实验对比分析图。可以看出,MLR 算法各指标性能随着弱分类器个数的增加而越来越好,说明该算法是有效的。

### 3.2 多类别分类数据集实验

在对多类别分类数据集进行对比实验中,选择 3 个标准 UCI 实验数据集,包括 Usps、Mnist 和 Pen-digits 数据集,具体数据集见表 3。

弱分类器选择 BP 神经网络算法,训练 30 个弱分类器。在每轮弱分类器的迭代训练过程中,先根据以往训练所得弱分类器组成的临时强分类器的分类结果,统计各个类别之间被“混淆”的程度,也就是“分错”的概率,判断标签之间相似性,见第 1.4 节。实验结果见表 4。

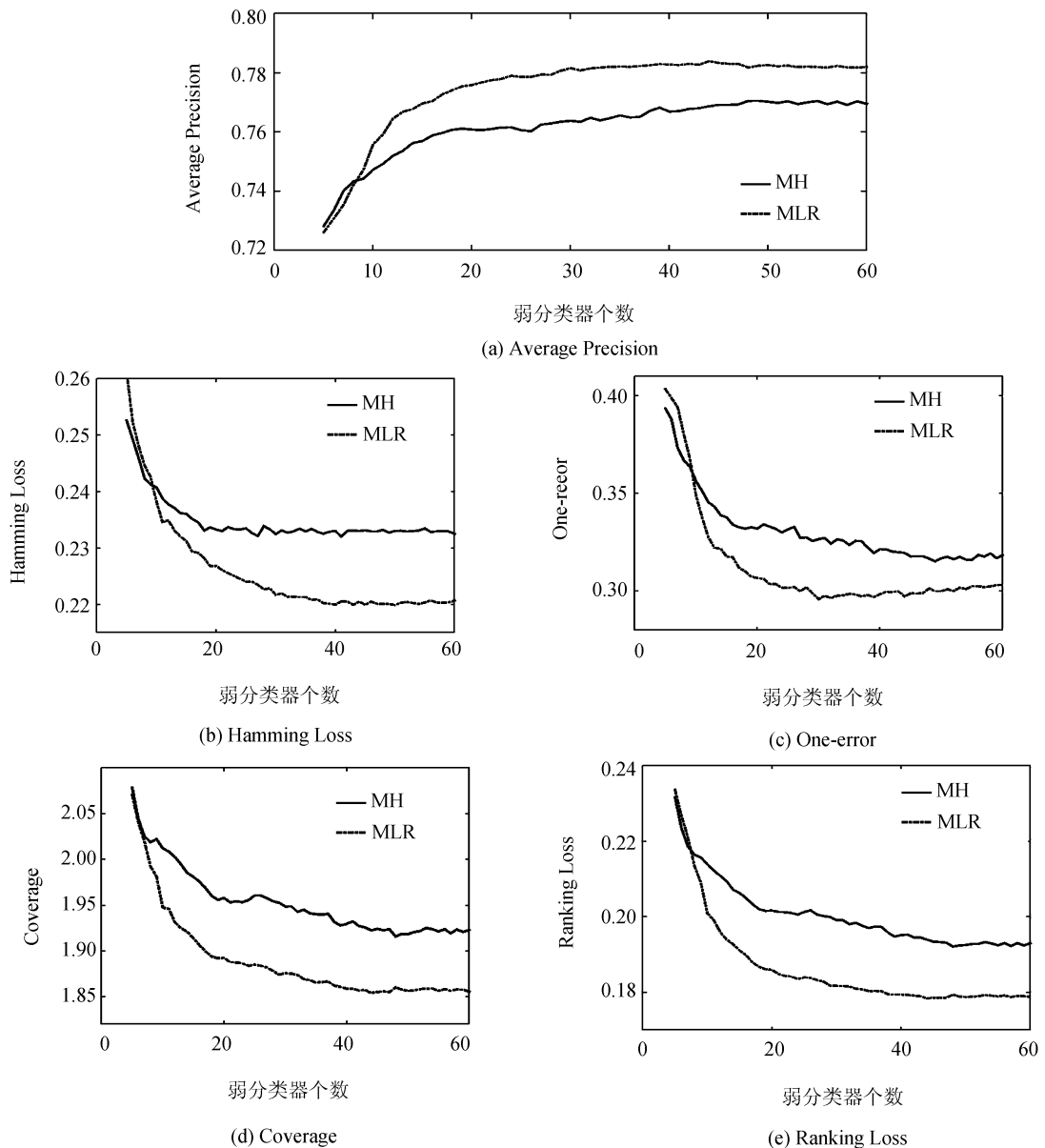


图 3 Emotions 数据集性能对比

Fig. 3 Comparison in Emotions data

表 3 多类别数据集

Tab. 3 Multi-class data sets

数据集	训练集	测试集	属性	标签
Usps	1 000	1 000	256	10
Mnist	60 000	10 000	784	10
Pendigits	7 494	3 498	16	10

表 4 多类别分类实验对比结果

Tab. 4 Comparison results of multi-class classification

算法	MLR	MH	MLkNN	RSVM
Usps	0.058 0	0.064 0	0.081 0	0.077 0
Mnist	0.091 9	0.101 7	NaN	NaN
Pendigits	0.020 1	0.032 1	0.026 6	0.035 4

可以看出,相比于其他 3 种算法,MLR 算法在 3

个数据集上分类效果最好。由于 Mnist 数据集样本空间太大,而 MLkNN 算法需要两两样本之间的距离,所占用的内存空间过大,RSVM 在大数据集上收敛过慢,因此,这 2 个算法未在该数据集上实验。

相比于 MH 算法,详细的数字识别错误率如表 5 所示。其中,第 1~10 行分别为数字 0~9 的错误率。

数据表明,本文算法在 3 个数据集上的总体识别错误率都有一定程度的降低,尤其是针对易“混淆”的类别,例如,Usps 数据集的数字 5 和数字 8 降低了 12.5% 左右;Mnist 数据集的数字 2 和数字 3 降低了 15% 左右,数字 5 降低了 9% 左右;Pendigits 数据集的数字 4 降低了 11.6% 左右。整体来说,MLR

算法性能在一些易“混淆”的类别上相较于 MH 算法有显著提升。这说明,在权值迭代过程中,考虑前面已经训练的弱分类器集成的临时强分类器来确定标签相似性矩阵,能够将分类器集中到易错分的样本和易错分的标签上,从而提高易“混淆”类别的分类性能,从而对整体性能有所提升。

表5 详细错误率

Tab.5 Detailed classification error

数据集	Usps		Mnist		Pendigits	
	MH	MLR	MH	MLR	MH	MLR
0	0.017 0	0.011 4	0.032 7	0.031 8	0.014 0	0.013 7
1	0.009 9	0.009 9	0.024 7	0.029 1	0.046 3	0.043 7
2	0.061 7	0.076 4	0.154 1	0.131 5	0.007 4	0.007 6
3	0.075 9	0.078 6	0.110 9	0.094 1	0.005 3	0.006 2
4	0.051 9	0.039 0	0.081 5	0.110 0	0.018 1	0.016 0
5	0.290 9	0.254 5	0.180 5	0.163 7	0.036 4	0.034 2
6	0.048 4	0.032 3	0.069 5	0.068 9	0.008 6	0.008 9
7	0.065 4	0.065 4	0.095 3	0.100 2	0.025 0	0.024 2
8	0.133 3	0.116 7	0.154 0	0.143 7	0.012 1	0.012 3
9	0.090 0	0.100 0	0.138 8	0.129 8	0.028 2	0.027 9

## 4 结 论

MLR 算法采用余弦相似度模拟标签相关性,利用标签相关性把原始标签转换为模糊标签,这种转换方式不仅能考虑两两标签相关性,还能考虑多个标签之间的相关性。同样,MLR 算法也能解决多类别分类问题,通过判断 2 个标签是否容易“混淆”,将“混淆”程度高的两个标签认为是标签相似的,模拟标签之间的相关性。理论分析和对比实验表明,在保证学习错误率随着弱分类器增加而逐渐减少的前提下,MLR 算法在大部分数据集上能表现出更好的性能,尤其在标签关系复杂的数据集上的性能提升更为显著。

### 参考文献:

[1] Huang Shengjun, Zhou Zhihua. Multi-label learning by exploiting label correlations locally [C]//Proceedings of the 26th AAAI Conference on Artificial Intelligence, AAAI'12. Toronto: AAAI Press, 2012: 949 - 959.

[2] Li Jiayang. 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.]

[3] Tsoumakas G, Katakis I. Multi-label classification: An o-

view [J]. Greece: Aristotle University of Thessaloniki, 2006.

- [4] Boutell M R, Luo J, Shen X, et al. Learning multi-label scene classification [J]. Pattern Recognition, 2004, 37 (9): 1757 - 1771.
- [5] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification [J]. Machine Learning, 2011, 85 (3): 333 - 359.
- [6] Tsoumakas G, Vlahavas I. Random k-labelsets: An ensemble method for multilabel classification [M]//Machine learning: ECML 2007. Berlin: Springer, 2007: 406 - 417.
- [7] Yin Hui, Xu Jianhua, Xu Hua. A multi-label classification algorithm based on LS-SVM [J]. Journal of Nanjing Normal University (Engineer and Technology Edition), 2010, 10(2): 68 - 73. [殷会, 许建华, 许花. 基于 LS-SVM 的多标签分类算法 [J]. 南京师范大学学报 (工程技术版), 2010, 10(2): 68 - 73.]
- [8] Zhang Minling, Zhou Zhihua. ML-kNN: A lazy learning approach to multi-label learning [J]. Pattern Recognition, 2007, 40(7): 2038 - 2048.
- [9] Elisseeff A, Weston J. A kernel method for multi-labeled classification [C]//Proceedings of Advances in Neural Information Processing Systems Cambridge: MIT Press, 2002: 681 - 687.
- [10] Tsoumakas G, Katakis I, Vlahavas I. Random k-labelsets for multilabel classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(7): 1079 - 1089.
- [11] Zhu J, Zou H, Rosset S, et al. Multi-class adaboost [J]. Statistics and Its Interface, 2009, 2(3): 349 - 360.
- [12] Freund Y, Schapire R E. Experiments with a new boosting algorithm [C]//Proceedings of International Conference on Machine Learning, 1996: 148 - 156.
- [13] Schapire R E, Singer Y. Improved boosting algorithms using confidence-rated predictions [J]. Machine Learning, 1999, 37(3): 297 - 336.
- [14] Fu Zhongliang, Wang Lili, Zhang Danpu. An improved multi-label classification ensemble learning algorithm [M]//Pattern Recognition. Berlin: Springer, 2014: 243 - 252.
- [15] Fu Zhongliang, Zhang Danpu, Wang Lili. Improvement on AdaBoost for multi-label classification [J]. Journal of Sichuan University (Engineering Science Edition), 2015, 47 (5): 103 - 109. [付忠良, 张丹普, 王莉莉. 多标签 AdaBoost 算法的改进算法 [J]. 四川大学学报 (工程科学版), 2015, 47(5): 103 - 109.]