

# 一种基于示例非独立同分布的多示例多标签分类算法

陈彤彤<sup>1</sup> 丁昕苗<sup>2</sup> 柳婵娟<sup>1</sup> 邹海林<sup>1</sup> 周树森<sup>1</sup> 刘影<sup>1</sup>

(鲁东大学信息与电气工程学院 烟台 264025)<sup>1</sup> (山东工商学院信息与电子工程学院 烟台 264005)<sup>2</sup>

**摘要** 多示例多标签学习是一种新型的机器学习框架。在多示例多标签学习中,样本以包的形式存在,一个包由多个示例组成,并被标记多个标签。以往的多示例多标签学习研究中,通常认为包中的示例是独立同分布的,但这个假设在实际应用中是很难保证的。为了利用包中示例的相关性特征,提出了一种基于示例非独立同分布的多示例多标签分类算法。该算法首先通过建立相关性矩阵表示出包内示例的相关关系,每个多示例包由一个相关性矩阵表示;然后建立基于不同尺度的相关性矩阵的核函数;最后考虑到不同标签的预测对应不同的核函数,引入多核学习构造并训练针对不同标签预测的多核 SVM 分类器。图像和文本数据集上的实验结果表明,该算法大大提高了多标签分类的准确性。

**关键词** 多示例学习,多示例多标签学习,示例非独立同分布,多核学习

**中图分类号** TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.2.060

## Multi-instance Multi-label Learning Algorithm by Treating Instances as Non-independent Identically Distributed Samples

CHEN Tong-tong<sup>1</sup> DING Xin-miao<sup>2</sup> LIU Chan-juan<sup>1</sup> ZOU Hai-lin<sup>1</sup> ZHOU Shu-sen<sup>1</sup> LIU Ying<sup>1</sup>

(School of Information and Electrical Engineering, Ludong University, Yantai 264025, China)<sup>1</sup>

(School of Information and Electronic Engineering, Shandong Institute of Business and Technology, Yantai 264005, China)<sup>2</sup>

**Abstract** Multi-instance multi-label learning (MIML) is a new machine learning framework. In this framework, an object is represented as a bag which is decomposed of multiple instances and labeled with multiple labels. Previous studies on MIML typically treated instances in the bags are independently identically distributed. However, it is difficult to be guaranteed in real tasks. Considering correlation features of instance in a bag, a multi-instance multi-label learning algorithm by treating instances as non-independent identically distributed samples was proposed. Firstly, instance correlations are considered by establishing an affinity matrix. By this means each bag is represented with an affinity matrix. Then, kernel functions based on the affinity matrix in different scales are established. Finally, considering predictions of different kinds of labels are corresponding to different kernels, multiple kernel learning is introduced to construct and train the MKSVMs. Experimental results on image data set and text data set show that the proposed algorithm greatly improves the accuracy of the image multi-label classification compared with other methods.

**Keywords** Multi-instance learning, Multi-instance multi-label learning, Non-I. I. D. instances, Multiple kernel learning

## 1 引言

在传统的监督学习中,一个样本可以表示为一个示例,并被一个标签标记。这种模型在机器学习中得到广泛的应用,但是对于现实世界的复杂问题并不适用。多示例学习可以看作是监督学习的一种变形,因为其独特的性质,被认为是一种新的学习框架<sup>[2]</sup>,与监督学习、非监督学习和强化学习并列。1997年,Dieterich<sup>[1]</sup>等人在研究药物活性预测问题时首次提出了多示例学习的概念。多示例学习将每个样本定义为包,

包中包含多个示例,若包中至少有一个示例为正例,则该包被标记为正包;若包中所有示例为负例,则该包被标记为负包。在多示例学习中,包有标签,包内的示例没有标签。Dieterich等还提出了3种轴平行矩形学习算法用于求解多示例学习问题,即GFS elim-count APR、GFS elim-kde APR和Iterated-discrim APR算法。这类算法采用贪婪或迭代的搜索策略,在属性空间里寻找包含每个正包至少一个示例的最小轴平行矩形,预测未知包的类别时,根据其是否有示例在矩形内判定包的正负。多示例学习的应用领域逐渐扩展到了图

到稿日期:2015-06-12 返修日期:2015-09-12 本文受国家自然科学基金(61170161,61300155,61303086),山东省政府留学基金委,鲁东大学博士基金(LY2014033)资助。

陈彤彤(1989-),女,硕士生,主要研究方向为模式识别、机器学习,E-mail:dreamofctt@163.com;丁昕苗(1979-),女,博士,副教授,主要研究方向为网络多媒体信息理解与识别;柳婵娟(1980-),女,博士,副教授,主要研究方向为图像处理与模式识别、机器学习,E-mail:luckyjcj80@sina.com(通信作者);邹海林(1962-),男,博士,教授,主要研究方向为智能信息处理、模式识别、图像分析与处理、遥感遥测信息处理、地质体三维建模与信息可视化等;周树森(1984-),男,博士,副教授,主要研究方向为模式识别、机器学习、图像处理、自然语言处理等;刘影(1989-),女,硕士生,主要研究方向为图像分析与处理、模式识别。

像、文本和视频的分类检索中。2006年,戴宏斌<sup>[3]</sup>等将多示例学习框架引入到基于内容的图像检索中,通过图像的局部颜色和纹理特征生成多示例包,使用多示例学习算法进行训练和预测,更好地处理图像的歧义性。李大湘<sup>[4]</sup>等将多示例学习引入到基于对象的图像检索中,以分割区域的视觉特征作为包中示例,提取视觉语义构造投影空间,投影约简后利用直推式支持向量机实现半监督学习。文献[5]利用多样性密度函数,结合量子粒子群优化(QPSO)算法,提出了QPSO-MIL算法,用以解决图像标注问题。为解决正包中的某些负示例会带来冗余和噪声的问题,Chen<sup>[6]</sup>等通过构建投影空间,将包映射成空间中的点,将多示例学习问题转化为监督学习问题,采用支持向量机训练分类器,提出了MILES(Multiple-Instance Learning via Embedded Instance Selection)算法,通过特征选择剔除噪声特征,提高了包特征的判别性。Fu<sup>[7]</sup>等在MILES的基础上又提出了MILIS(Multiple-Instance Learning with Instance Selection)算法,通过高斯核密度估计函数构建负包示例的分布模型,在正包中挑选似然值最小的示例,在负包中挑选似然值最大的示例,并交替进行示例更新和分类学习,进一步提高了分类效果。张钢<sup>[8]</sup>等人提出了一种基于多示例核的半监督学习算法,同时增加了平滑性惩罚项,以利用未知标签的包。Gärtner<sup>[9]</sup>等将多示例包看作一组特征向量,并直接用于构造多示例核函数MI-Kernel,进而解决多示例问题。Li<sup>[10]</sup>等利用稀疏 $\epsilon$ 图表示每个包中示例间的上下文结构,提出了一种基于分层稀疏表示的多示例学习算法。Liu<sup>[11]</sup>等提出了主示例检测问题,并通过引入示例间的关系构造了一个投票框架,利用该框架实现了标签的预测。丁昕苗<sup>[12]</sup>等将视频场景作为包,镜头作为包内示例,采用多示例学习进行恐怖视频识别,并考虑到视频的上下文信息和示例包的统计特性,利用联合稀疏表示将多个视角融合到一个分类框架中。Feng<sup>[13]</sup>等则引入分层稀疏表示,提出了一种半监督多示例学习算法。上述算法主要用于预测多示例包的单个标签,但对于多示例包的多类标签的预测问题并不适用。

随着人们对事物多重语义理解的深入,周志华<sup>[14,15]</sup>等首次提出了多示例多标签学习框架,其成为了当前一个新的研究热点。该框架将多示例多标签问题转化为多示例单标签问题和单示例多标签问题,进而转化为传统的监督学习问题加以解决。他们也提出了MIML\_BOOST、MIML\_SVM方法学习框架,并将其成功应用到了图像场景分类问题中。其中,MIML\_BOOST算法是先将多示例多标签问题转化为多示例单标签问题,再利用MIBOOSTING算法解决多示例问题;MIML\_SVM算法则是先将多示例多标签问题转化为单示例多标签问题,再利用MLSVM算法解决多标签问题。Li<sup>[16]</sup>等提出了KISAR算法,该算法首先将聚类中心看作主要代表性示例,然后用将每个包映射到代表性示例的方式实现多示例包向单示例样本的转化,进而通过构造凸优化问题训练得到面向所有标签的分类器。Briggs<sup>[21]</sup>为了改善算法的排序损失,提出了RankLossSIM(Rank-Loss Support Instance Machines for multi-instance multi-label instance annotation)。此后,Zhou<sup>[17]</sup>等还提出了带有软标签的多示例多标签学习算法MIMLwel(Multi-Instance Multi-Label Learning with Weak Label)。为了高效地实现大数据集的多标签预测,Huang<sup>[18]</sup>

等提出了MIMLfast算法,该算法首先构建一个所有标签共用的低维子空间,然后利用随机梯度下降法训练标签特定的线性模型,从而优化排序损失。

以上这些算法尽管在解决多示例多标签问题上取得了较好的效果,但都以包中的示例是独立同分布为前提,在实际应用中,示例的独立同分布假设是很难保证的;同时忽略了包内示例的相关性特征,这必然造成特征信息的损失;另外,解决多示例多标签问题的经典算法MIMLBOOST和MIMLSVM存在一定的缺陷。在利用MIBOOSTING算法解决多示例问题时,给包的所有示例标记该包的标签,未考虑到正包中包含大量的负示例,因而会造成较大的误差;MIMLSVM算法则用两个包内示例与示例之间的最小Hausdorff距离来表示包与包之间的距离,当一个正包的负示例与负包的示例很相似时,该距离的表示方法会降低正包和负包的判别性,影响分类效果。

为了解决上述问题,本文提出了一种基于示例非独立同分布的多示例多标签分类算法Non-I. I. D. MIML(Multi-instance Multi-label Learning algorithm by treating instances as Non-Independent Identically Distributed samples),算法框图如图1所示。NON-I. I. D. MIML首先通过建立相关性矩阵描述包内示例的相关性特征,然后通过核函数实现高维映射,将相关性矩阵投影为高维空间的向量来表示一个包的特征,使包的特征表示更全面;并且该特征将多示例包在高维空间转变成了单示例样本,回避了包内示例的标记问题,同时在计算两包之间的距离时回避了用Hausdorff距离来表示包距离产生的偏差。针对不同类别标签的预测对应的核函数有所不同,NON-I. I. D. MIML又引入了多核融合的方法,以不同尺度的相关性矩阵和不同的参数构造多个基本核函数,通过将基本核函数进行凸组合实现了核函数的多样化,凸组合时不同的系数组合会得到不同的核函数。最后,通过训练得到针对不同标签的多核分类器。

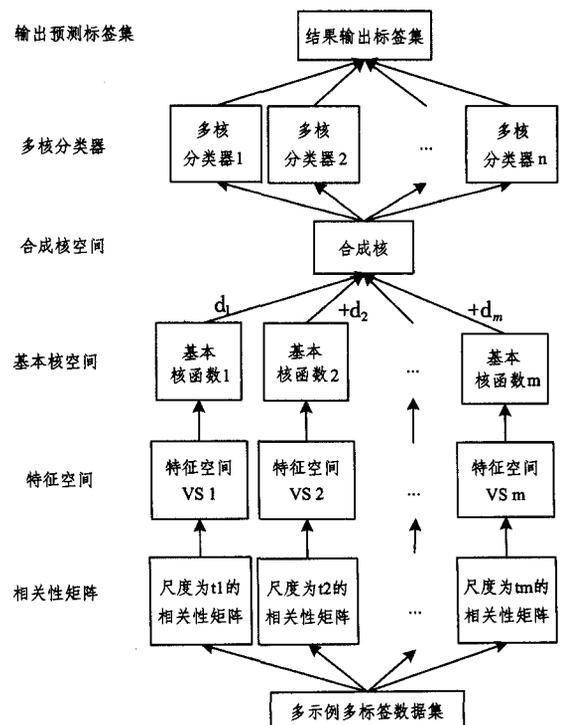


图1 NON-I. I. D. MIML 算法框图

本文第2节详细介绍了 NON-I. I. D. MIML 算法, 主要包括包内示例相关性图的构造、基于相关性矩阵的核函数以及多核融合的实现 3 方面的内容; 第3节通过在图像数据集上的实验验证了 NON-I. I. D. MIML 算法在图像场景分类上的应用效果; 最后总结全文。

## 2 基于示例非独立同分布的多示例多标签分类算法

设有示例空间  $\mathcal{X}$ 、类标签集  $Y$ , 多示例多标签学习定义如下:

给定一个数据集  $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_m, Y_m)\}$ , 其中包  $X_i \subseteq \mathcal{X}$  为一组示例  $\{x_{i1}, x_{i2}, \dots, x_{in_i}\}$ ,  $x \in \mathcal{X} (j=1, 2, \dots, n_i)$ 。而  $Y_i \subseteq Y$  为  $X_i$  的标签集  $\{y_{i1}, y_{i2}, \dots, y_{il_i}\}$ ,  $y_k \in Y (k=1, 2, \dots, l_i)$ 。  $n_i$  为包  $X_i$  的示例个数,  $l_i$  为  $Y_i$  中包含的标签个数。图 2 为多示例多标签学习框架。多示例多标签学习的目标就是根据这样一个数据集训练学得一个预测函数  $f: 2^{\mathcal{X}} \rightarrow 2^Y$ , 从而利用该预测函数预测未知包的标签集。

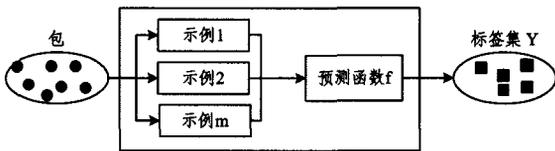


图 2 多示例多标签学习框架

### 2.1 包内示例相关性矩阵的构造

一般的多示例学习研究中, 示例之间被看作是相互独立的, 但是在实际任务中, 包中的示例之间存在着某些特征上的联系, 因此示例的独立性假设是不合适的。在现实世界里, 许多实体内部存在着固有的结构, 这种结构关系可以很好地作为特征来表示实体。如图 3 所示, 该图可经过分割得到多个固定大小的区域块, 将该图看作一个包, 分割后的每一个区域块对应一个示例, 图中标出的 4 个图像块对应包内 4 个示例, 显然, 4 个示例间从颜色特征上看是相似的, 因而可以将这种相似性作为描述图像特征的重要信息。传统的多示例学习研究中, 示例之间通常被看作是相互独立的, 这必然给多示例学习的预测带来一定的信息损失。因此, NON-I. I. D. MIML 算法将包内示例非独立同分布的特征引入到多示例多标签分类中。

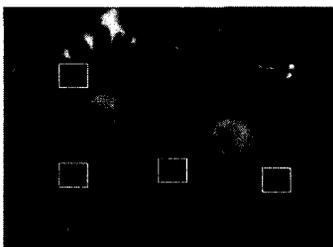


图 3 标有 4 个示例的图像举例

鉴于 Zhou 等<sup>[19]</sup>在 2009 年提出的 MIGraph 和 miGraph 的成功应用, NON-I. I. D. MIML 也采用构建相关性矩阵的方法对包内示例的相互关系进行建模。对于包  $X_i \subseteq \mathcal{X}$ , 建立一个  $n_i \times n_i$  的相关性矩阵  $W^i$ 。矩阵的行号和列号分别对应包内的两个示例, 对应图的两个顶点。如果示例  $x_a$  和  $x_b$  之间的高斯距离小于预设阈值  $\delta$ , 则  $W^i$  的第  $a$  行  $b$  列的元素  $w_{ab}^i = 1$ , 否则  $w_{ab}^i = 0$ 。这样每个包就对应生成了一个相关性矩阵,

并将其作为包的特征。通过改变阈值  $\delta$  的大小, 可以获取不同尺度的示例相关性特征。

### 2.2 基于相关性矩阵的核函数

作为描述包特征的相关性矩阵, 其无法直接用于模式分类。因此需要通过高维投影将相关性矩阵转换为高维空间的向量来表示一个包的特征, 并基于文献<sup>[19]</sup>在该空间定义核函数, 如式(1)所示:

$$k_g(X_i, X_j) = \frac{\sum_{a=1}^{n_i} \sum_{b=1}^{n_j} W_{ia}^i W_{jb}^j k(x_{ia}, x_{jb})}{\sum_{a=1}^{n_i} W_{ia}^i \sum_{b=1}^{n_j} W_{jb}^j} \quad (1)$$

其中,  $W_{ia}^i = 1 / \sum_{u=1}^{n_i} w_{iu}^i$ ,  $W_{jb}^j = 1 / \sum_{v=1}^{n_j} w_{jv}^j$ ,  $w_{iu}^i$  表示包  $i$  的相关性矩阵中第  $a$  行  $u$  列的元素值,  $k(x_{ia}, x_{jb})$  利用高斯核函数求得,  $k(x_{ia}, x_{jb}) = \exp(-\gamma \|x_{ia} - x_{jb}\|^2)$ 。得到核函数之后, 则可以利用基于核的 SVM 分类器进行学习分类。

### 2.3 多核融合的实现

由于要解决的是多标签问题, 而针对每一种标签, 其样本分布情况存在差异性, 如果采用单个核函数, 也就是数据集采用同一种高维映射方式, 便有了局限性。因此, NON-I. I. D. MIML 引入多核学习来解决基于示例相关性的多示例多标签问题。

多核学习算法通过组合多个核函数达到组合多个特征空间的目的, 使得原数据集映射到多个不同的特征空间, 经过组合使算法的灵活性和准确性更高。最简单的多核模型就是考虑将多个基本核函数进行凸组合<sup>[20]</sup>:

$$K(x, x') = \sum_{m=1}^M d_m K_m(x, x'), d_m \geq 0, \sum_{m=1}^M d_m = 1 \quad (2)$$

其中,  $K_m(x, x')$  为基本核函数,  $M$  为基本核函数的总个数,  $d_m$  为权系数。基本核函数  $K_m(x, x')$  可以是对应不同参数的传统核函数(比如高斯核函数)。多核函数线性组合算法的基本流程<sup>[19]</sup>如图 4 所示, 凸组合是线性组合的一种特殊形式。在多核学习框架中, 数据通过核在特征空间的表示问题转化为权系数  $d_m$  的选择问题。

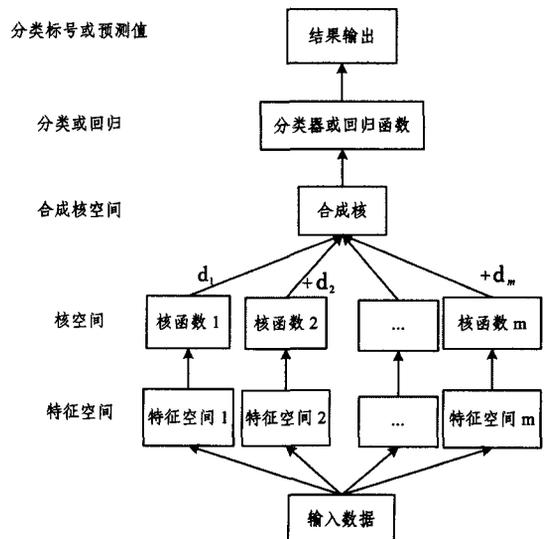


图 4 多核函数线性组合算法流程

由于不同类别标签的预测对应的相关性特征的尺度有所不同, NON-I. I. D. MIML 算法以不同尺度的相关性矩阵, 针

对核函数式(1),采用不同的  $\gamma$  参数构造多个不同的基本核函数  $k_{gm}(X_i, X_j)$ ,然后将基本核函数进行凸组合,得到多核融合后的核函数,如式(3)所示:

$$K(X_i, X_j) = \sum_{m=1}^M d_m k_{gm}(X_i, X_j), d_m \geq 0, \sum_{m=1}^M d_m = 1 \quad (3)$$

对每个  $y \in Y$ ,以不同尺度的相关性矩阵对应的不同参数核矩阵作为基本核,采用 SimpleMKL 算法<sup>[20]</sup>训练得到一个 MKSVM 分类器  $h_y$ ,求得基本核函数的权系数  $d_m$ ,这样就实现了多核融合。最后,利用学习好的 MKSVM 分类器对未知包的标签集进行预测。表 1 详细描述了算法的流程。该算法通过多核融合的方式提升了算法的灵活性,更适用于多标签分类问题,同时也能较好地应对样本集数据异构等多种复杂的情况。

表 1 NON-I. I. D. MIML 算法

输入: 训练集 $\Omega = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_m, Y_m)\}$ ; 标签集 $Y$ ; 预测包 $X^*$ ; 两个长度为 $L$ 的参数数组 $\text{threshold}, \text{gam}$
步骤:
(1) 将训练集中的样本 $(X_u, Y_u) (u=1, 2, \dots, m)$ 的标签去掉,组成一个无标签的数据集 $\Gamma = \{X_u   u=1, 2, \dots, m\}$ , $X_u$ 为一组示例 $\{x_{u1}, x_{u2}, \dots, x_{un}\}$ 。
(2) 重复进行如下操作 $L$ 次:
2a) 对数据集 $\Gamma$ 中的每个包 $X_u$ 建立不同尺度的相关性矩阵 $W^u$ , 其中使用的阈值 $\delta = \text{threshold}(l) (l=1, 2, \dots, L)$ 。
2b) 根据图形核函数 $k_g$ 求不同尺度的相关性矩阵对应的核矩阵 $k_{\text{train}}$ , 核函数系数 $\gamma = \text{gam}(l) (l=1, 2, \dots, L)$ 。
$k_g(X_i, X_j) = \frac{\sum_{a=1}^{n_i} \sum_{b=1}^{n_j} W_{ia} W_{jb} k(x_{ia}, x_{jb})}{\sum_{a=1}^{n_i} W_{ia} \sum_{b=1}^{n_j} W_{jb}}$
(3) 对每个 $y \in Y$ ,以不同尺度相关性矩阵对应的核矩阵 $k_{\text{train}}$ 作为基本核,依据训练集的该类标签值,采用 SimpleMKL 算法训练得到一个 MKSVM 分类器 $h_y$ 。
(4) 根据 $Y^* = \{\arg \max_{y \in Y} h_y(k(X^*))\} \cup \{y   h_y(k(X^*)) \geq 0, y \in Y\}$ , 利用训练得到的所有 MKSVM 分类器求预测包 $X^*$ 的标签集, $k(X^*)$ 为 $X^*$ 的核矩阵。
输出: 预测包 $X^*$ 的标签集 $Y^*$

### 3 实验结果与分析

#### 3.1 实验数据集及评价标准

为验证 NON-I. I. D. MIML 算法的性能,在 MSRC v2<sup>[18]</sup> 和 Letter-Frost<sup>[21]</sup> 数据集上进行了实验。MSRC v2 数据集包含 23 类的 591 张图片,其中有很多图像同时属于多类。通过分割,每张图片用包含多个对应不同区域的特征向量表示。Letter-Frost 文本数据集由 144 个单词组成,这些单词来源于诗“The Road Not Taken”,每个单词对应一个包,包中的示例对应组成该单词的字母。包的标签即为包含的字母类别。每个字母由一个 16 维的特征向量表示。

在传统的监督学习中,每个对象只有一个标签,通常只用 accuracy 来评价性能就可以了。但对于多标签学习问题,需要对一个对象预测多类标签,仅用 accuracy 已经不具有说服力了。因此通常用 5 个指标来评价多标签学习性能,分别是 hamming loss、one-error、coverage、ranking loss、average precision。其中前 4 个指标的值越小,性能越好;average precision 的值越大,性能越好。此外还有两个新的多标签评价指标,分别为 average recall、average F1<sup>[15]</sup>。average recall 计算了被预测出的合适标签的平均分数,average F1 描述了平均精度与平均召回率之间的权衡,两种评价指标越大,表示该算法的性能越好。本文综合采用这 7 种评价指标(hamming loss、one-error、coverage、ranking loss、average precision、average recall、average F1)来评价算法的性能,通过对比 7 种指标来分析算法的优劣。

#### 3.2 实验结果及分析

MSRC v2 图像数据集上的实验随机选取 2/3 的样本作为训练集,剩余的样本用作测试。Letter-Frost 数据集上的实验随机选取 96 个样本作为训练集,剩余 48 个作为测试集。实验重复 30 次,最后求得算法分类性能指标的平均值及标准偏差。对于 MSRC v2,构图阈值取为  $\text{threshold} = [0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2]$ ,对应的不同高斯核函数参数  $\text{gam} = [0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2]$ ;而对于 Letter-Frost 数据集,实验选取构图阈值为  $\text{threshold} = [0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2]$ ,对应的不同高斯核函数参数为  $\text{gam} = [0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2]$ 。为了验证算法的有效性,实验中将本文提出的 NON-I. I. D. MIML 算法与多种多示例多标签算法的分类性能指标进行了比较,包括 MIMLBOOST<sup>[15]</sup>、MIMLSVM<sup>[15]</sup>、MIMLSVM-mi<sup>[15]</sup>、MIMLNN<sup>[15]</sup>、MIMLfast<sup>[18]</sup> 以及 KISAR<sup>[16]</sup>。其中, MIMLSVMmi 算法是在 MIMLBOOST 算法的基础上利用 MI-SVM<sup>[23]</sup> 方法代替 MIBOOSTING 方法做出的改进, MIMLNN 算法是在 MIMLSVM 算法的基础上利用两层神经网络<sup>[24]</sup> 代替 MLSVM 方法做出的改进。另外,还与经典的多标签学习算法 MLkNN<sup>[25,26]</sup> 进行了对比。MIMLfast 算法的步长  $\gamma_i = \frac{\gamma_0}{(1 + \eta \gamma_0 t)}$ ,  $\gamma_0 = 0.005, \eta = 10^{-5}$ , 范数的上界设为 1。MIMLBOOST 中 boosting 轮数设为 50。MIMLSVM 的参数  $k$  设为训练集的 20%。MLkNN 中最近邻数设为 10。

表 2 和表 3 分别给出了在两个实验数据集上的实验对比结果,最优结果用粗体表示。

表 2 MSRC v2 数据集上的实验结果

算法名称	评价标准						
	hamming loss ↓	one-error ↓	coverage ↓	ranking loss ↓	average precision ↑	average recall ↑	average F1 ↑
NON-I. I. D. MIML	<b>0.070 ± 0.003</b>	0.210 ± 0.021	4.656 ± 0.270	<b>0.085 ± 0.008</b>	0.770 ± 0.015	0.658 ± 0.019	<b>0.710 ± 0.015</b>
MIMLBOOST	0.113 ± 0.004	0.526 ± 0.035	11.144 ± 0.125	0.361 ± 0.025	0.368 ± 0.017	0.046 ± 0.009	0.082 ± 0.015
MIMLSVM	0.081 ± 0.004	0.303 ± 0.030	5.508 ± 0.449	0.115 ± 0.012	0.704 ± 0.019	0.435 ± 0.026	0.537 ± 0.024
MIMLSVMmi	0.083 ± 0.004	0.331 ± 0.026	7.665 ± 0.399	0.167 ± 0.011	0.627 ± 0.021	0.364 ± 0.029	0.460 ± 0.027
MIMLNN	0.073 ± 0.002	0.245 ± 0.019	5.256 ± 0.178	0.099 ± 0.005	0.743 ± 0.007	0.466 ± 0.012	0.573 ± 0.010
MIMLfast	0.105 ± 0.006	0.392 ± 0.041	6.973 ± 0.368	0.156 ± 0.012	0.601 ± 0.023	0.361 ± 0.034	0.451 ± 0.032
KISAR	0.077 ± 0.003	<b>0.084 ± 0.017</b>	13.348 ± 0.558	0.565 ± 0.026	0.433 ± 0.018	0.295 ± 0.020	0.351 ± 0.020
MLkNN	0.088 ± 0.003	0.347 ± 0.031	6.133 ± 0.286	0.141 ± 0.011	0.648 ± 0.020	0.306 ± 0.027	0.415 ± 0.026

表3 Letter-Frost 数据集上的实验结果

算法名称	评价标准						
	hamming loss ↓	one-error ↓	coverage ↓	ranking loss ↓	average precision ↑	average recall ↑	average F1 ↑
NON-I. I. D. MIML	0.103±0.013	0.150±0.071	9.713±1.359	0.130±0.025	0.742±0.044	0.594±0.057	0.659±0.049
MIMLBOOST	0.143±0.006	0.417±0.083	15.171±0.563	0.389±0.006	0.353±0.019	0.020±0.012	0.037±0.022
MIMLSVM	0.132±0.008	0.324±0.066	12.152±0.506	0.229±0.017	0.401±0.024	0.251±0.015	0.348±0.016
MIMLSVMmi	0.187±0.003	0.174±0.052	9.931±0.799	0.180±0.029	0.691±0.030	0.544±0.034	0.608±0.030
MIMLNN	0.112±0.004	0.248±0.035	9.980±0.410	0.169±0.006	0.655±0.010	0.447±0.015	0.531±0.011
MIMLfast	0.150±0.003	0.623±0.042	11.367±0.568	0.211±0.017	0.460±0.026	0.121±0.022	0.191±0.030
KISAR	0.114±0.006	0.419±0.107	13.846±0.990	0.247±0.032	0.515±0.055	0.239±0.026	0.327±0.033
MLkNN	0.129±0.008	0.443±0.062	10.183±0.868	0.211±0.019	0.536±0.036	0.146±0.040	0.227±0.050

由表2看出, NON-I. I. D. MIML 算法在 hamming loss、coverage、ranking loss、average precision、average recall 和 average F1 上优于其他 7 种算法。MIMLNN 算法的性能和 NON-I. I. D. MIML 的性能相差不多。同时, 通过比较整体性能的标准偏差看出, MIMLNN 的最小, NON-I. I. D. MIML 次之, 可见 NON-I. I. D. MIML 算法的稳定性是值得的。由算法在 Letter-Frost 数据集上的结果看出, NON-I. I. D. MIML 算法的性能明显好于其他 7 种算法。由此可见, 本文所提算法大大提高了图像分类和文本分类的准确性。NON-I. I. D. MIML 算法之所以能够取得最好的分类效果, 一方面, 由于包内示例间的相关性特征作为包的整体特征能够更全面地表示包; 另一方面, 多核学习的引入使得算法更适用于多标签问题, 针对不同类标签的预测, 训练出最适合的核函数和分类器, 灵活性提高。同时, 算法还存在一定的不足: 一方面, 算法通过构建相关性矩阵来表示示例间的相关性, 每个包对应一个相关性矩阵, 算法复杂度增加; 另一方面, 针对每类标签都要训练一个 MKSVM 分类器, 又提高了算法的复杂度。因此, 如何降低算法的复杂度还有待解决。

**结束语** 考虑到多示例多标签问题中包内示例间存在相关性特征的问题, 本文提出了一种基于示例非独立同分布的多示例多标签分类算法 NON-I. I. D. MIML。该算法通过对每个包构建相关性矩阵得到了示例相关性特征, 同时考虑到不同标签的预测对应的高维投影不同, 将多核学习引入到分类器的构造中, 使算法更适于多标签预测, 提高了算法的灵活性。在不同类型及大小的两个数据集上的仿真实验表明, NON-I. I. D. MIML 算法的性能优于现有的其他多示例多标签算法, 大大提升了多示例多标签分类算法在图像分类及文本分类上的准确性。由于算法要建立不同尺度的相关性矩阵, 并针对不同类别的标签训练不同的 MKSVM 分类器, 算法的复杂度必然有所增加, 如何降低算法复杂度是后续研究工作的重点之一。对于示例相关性较小的样本, 利用该算法如何提高性能也是需要后续研究的问题。另外, 标签之间也是存在一定关系的, 将标签之间的关系融合到该算法中能否提高算法性能还有待研究。

## 参 考 文 献

[1] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles [J]. Artificial Intelligence, 1997, 89(1/2): 31-71

[2] Zhang Min-ling, Zhou Zhi-hua. A multi-instance regression algorithm based on neural network [J]. Journal of Software, 2003, 14(7): 1238-1242 (in Chinese)

张敏灵, 周志华. 基于神经网络的多示例回归算法 [J]. 软件学

报, 2003, 14(7): 1238-1242

[3] Dai Hong-bin, Zhang Min-ling, Zhou Zhi-hua. A Multi-instance learning based approach to image retrieval [J]. Pattern Recognition and Artificial Intelligence, 2006, 19(2): 179-185 (in Chinese)

戴宏斌, 张敏灵, 周志华. 一种基于多示例学习的图像检索方法 [J]. 模式识别与人工智能, 2006, 19(2): 179-185

[4] Li Da-xiang, Peng Jin-ye, Li Zhan. Object-based image retrieval using semi-supervised multi-instance learning algorithm [J]. Control and Decision, 2010, 25(7): 981-986 (in Chinese)

李大湘, 彭进业, 李展. 基于半监督多示例学习的对象图像检索 [J]. 控制与决策, 2010, 25(7): 981-986

[5] Li Da-xiang, Peng Jin-ye, Bu Qi-rong. QPSO-based multi-instance learning for image annotation [J]. Computer Science, 2010, 37(6): 278-282 (in Chinese)

李大湘, 彭进业, 卜起荣. 基于 QPSO-MIL 算法的图像标注 [J]. 计算机科学, 2010, 37(6): 278-282

[6] Chen Yi-xin, Bi Jin-bo, Wang J Z. MILES: Multiple-instance learning via embedded instance selection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 1931-1947

[7] Fu Zhou-yu, Robles-Kelly A, Zhou Jun. MILIS: multiple instance learning with instance selection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 958-977

[8] Zhang Gang, Yin Jian, Cheng Liang-lun, et al. Semi-supervised multi-instance kernel [J]. Computer Science, 2011, 38(9): 220-223 (in Chinese)

张钢, 印鉴, 程良伦, 等. 半监督多示例核 [J]. 计算机科学, 2011, 38(9): 220-223

[9] Gärtner T, Flach P A, Kowalczyk A, et al. Multi-instance kernels [C] // Proceedings of the 19th International Conference on Machine Learning, 2002, Sydney, Australia, 2002: 179-186

[10] Li Bing, Xiong Wei-hua, Hu Wei-ming. Context-aware multi-instance learning based on hierarchical sparse representation [C] // IEEE 11th International Conference on Data Mining, 2011. Vancouver, Canada; IEEE, 2011: 370-377

[11] Liu Guo-qing, Wu Jian-xin, Zhou Zhi-hua. Key instance detection in multi-instance learning [C] // Proceedings of the 4th Asian Conference on Machine Learning, 2012. Singapore, 2012: 253-268

[12] Ding Xin-miao, Li Bing, Hu Wei-ming, et al. Horror video scene recognition based on multi-view joint sparse coding [J]. Acta Electronica Sinica, 2014, 42(2): 301-305 (in Chinese)

丁昕苗, 李兵, 胡卫明, 等. 基于多视角融合稀疏表示的恐怖视频识别 [J]. 电子学报, 2014, 42(2): 301-305

- [13] Feng Song-he, Xiong Wei-hua, Li Bing, et al. Hierarchical sparse representation based multi-instance semi-supervised learning with application to image categorization [J]. *Signal Processing*, 2014, 94(1): 595-607
- [14] Zhou Zhi-hua, Zhang Min-ling. Multi-instance multi-label learning with application to scene classification[C]// *Advances in Neural Information Processing Systems 19*, 2007. Cambridge, United Kingdom; MIT Press, 2007: 1609-1616
- [15] Zhou Zhi-hua, Zhang Min-ling, Huang Sheng-jun, et al. Multi-instance multi-label learning [J]. *Artificial Intelligence*, 2012, 176(1): 2291-2320
- [16] Li Yu-feng, Hu Ju-huan, Jiang Yuan, et al. Towards discovering what patterns trigger what labels[C]// *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, 2012. Toronto, Canada, 2012: 1012-1018
- [17] Yang Shu-jun, Jiang Yuan, Zhou Zhi-hua. Multi-instance multi-label learning with weak label[C]// *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2013. Beijing, China; AAAI Press, 2013: 1862-1868
- [18] Huang Sheng-jun, Gao Wei, Zhou Zhi-hua. Fast multi-instance multi-label learning[C]// *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014. Quebec City, Canada, 2014: 1868-1874
- [19] Zhou Zhi-hua, Sun Yu-yin, Li Yu-feng. Multi-instance learning by treating instances as non-I. I. D. samples[C]// *Proceedings of the 26th International Conference on Machine Learning*, 2009. Montreal, Canada, 2009: 1249-1256
- [20] Rakotomamonjy A, Bach F R, Canu S, et al. SimpleMKL [J]. *Journal of Machine Learning*, 2008, 9(11): 2491-2521
- [21] Briggs F, Fern X, Raich R. Rank-loss support instance machines for miml instance annotation[C]// *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012. Beijing, China, 2012: 534-542
- [22] Maron O, Ratan A L. Multiple-instance learning for natural scene classification[C]// *Proceedings of the 15th International Conference on Machine Learning*, 1998. San Francisco, America, 1998: 341-349
- [23] Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple-instance learning[C]// *Advances in Neural Information Processing Systems 15*, 2003. Cambridge, United Kingdom; MIT Press, 2003: 561-568
- [24] Zhang Min-ling, Zhou Zhi-hua. Multi-label learning by instance differentiation[C]// *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*. Vancouver, Canada; AAAI Press, 2007: 669-674
- [25] Zhang Min-ling, Zhou Zhi-hua. Multi-label learning by instance differentiation[J]. *Pattern Recognition*, 2007, 40(7): 2038-2048
- [26] Hao Hong, Ji Hua, Zhang Hua-xiang, et al. Multi-label scene classification based on I2C distance and label dependency [J]. *Computer Science*, 2014, 41(1): 88-90 (in Chinese)  
郝虹, 计华, 张化祥, 等. 基于 I2C 距离和标记相关性的多标记场景分类 [J]. *计算机科学*, 2014, 41(1): 88-90

(上接第 253 页)

- [16] Lin Hong-kang, Li Yu-ying. P-sets and its P-separation theorems [J]. *An International Journal Advances in Systems Science and Applications*, 2010, 10(2): 209-215
- [17] Lin Rong, Shi Kai-quan. Function P-Sets and Attribute Control of Information Laws [J]. *Computer Science*, 2012, 39(7): 225-228 (in Chinese)  
林蓉, 史开泉. 函数 P-集合与信息规律的属性控制[J]. *计算机科学*, 2012, 39(7): 225-228
- [18] Wang Yang, Geng Hong-qin, Shi Kai-quan. The mining of dynamic information based on P-sets and its application [J]. *An International Journal Advances in Systems Science and Applications*, 2010, 10(2): 234-240
- [19] Xu Feng-sheng, Yu Xiu-qing, Shi Kai-quan. Embedding camouflage of Inverse P-information and its Separation-discovery by Inverse Reasoning[J]. *Computer Science*, 2013, 40(8): 200-203 (in Chinese)  
徐凤生, 于秀清, 史开泉. 逆 P-信息嵌入隐藏与它的逆 P-集推理分离-发现[J]. *计算机科学*, 2013, 40(8): 200-203
- [20] Fan Cheng-xian, Huang Shun-Liang. Inverse P - reasoning discovery identification of inverse P-information [J]. *International Journal of Digital Content Technology and its Application*, 2012, 6(20): 735-744
- [21] Wu Song-li, Chen Gui-you, Shi Kai-quan. Internal Inverse P-Information Intelligent Fusion and its Attribute Disjunctive Character Application[J]. *Computer Science*, 2014, 41(1): 262-266 (in Chinese)  
吴松丽, 陈桂友, 史开泉. 内逆 P-信息智能融合与它的属性析取特征-应用[J]. *计算机科学*, 2014, 41(1): 262-266
- [22] Lin Hong-kang, Fan Cheng-xian. Embedding-camouflage of inverse P-information and application[J]. *International Journal of Convergence Information Technology*, 2012, 7(20): 471-480
- [23] Tang Ji-hua, Chen Bao-hui, Zhang Ling, et al. Function inverse P-Sets and the dynamic Separation of Inverse P-Sets Information Laws[J]. *Journal of Shandong University (Natural Science)*, 2013, 48(8): 104-110 (in Chinese)  
汤积华, 陈保会, 张凌, 等. 函数逆 P-集合与逆 P-信息规律动态分离[J]. *山东大学学报(理学版)*, 2013, 48(8): 104-110
- [24] Zhang Jing-xiao, Xu Feng-sheng. Application of Function Internal Inverse P-Sets to QSPR Research[J]. *Journal of Shandong University (Natural Science)*, 2013, 48(8): 92-96 (in Chinese)  
张景晓, 徐凤生. 函数内逆 P-集合在 QSPR 研究中的应用[J]. *山东大学学报(理学版)*, 2013, 48(8): 92-96
- [25] Lin Rong, Shi Kai-quan. Internal Packet-information Camouflage and True-False Information Identification[J]. *Journal of Systems Engineering and Electronics*, 2013, 35(4): 796-801 (in Chinese)  
林蓉, 史开泉. 内 P-信息伪装与真伪信息辨识[J]. *系统工程与电子技术*, 2013, 35(4): 796-801
- [26] Zhang Jing-xiao, Liu Hin-hua. Function P-sets and Discovery of Structure-physical Properties Law in Analogues System [J]. *Shandong University (Natural Science)*, 2012, 47(8): 98-102 (in Chinese)  
张景晓, 刘新华. 函数 P-集合与同物系物质结构-物性规律发现[J]. *山东大学学报(理学版)*, 2012, 47(8): 98-102