第45卷 第8期 西 安 交 通 大 学 学 报 Vol. 45 No. 8
2011年8月 JOURNAL OF XI′AN JIAOTONG UNIVERSITY Aug. 2011

构建新包空间的多示例学习方法

温超，耿国华，李展
（西北大学信息科学与技术学院，710069，西安）

摘要：针对已有神经网络方法采用示例决定标记从而导致多示例学习(MIL)中包结构信息丢失的问题，提出了一种新的RK_BP多示例学习方法.在示例空间，首先采用粗糙集对其进行属性约简；然后进行K均值聚类，利用聚类点构造新包空间；在新空间中，利用误差反向传播神经网络算法进行分类.在多个测试数据集上对算法进行测试，结果表明该算法可有效解决已有神经网络方法包结构信息丢失问题，明显提高分类性能.

关键词：多示例学习；反向传播算法；粗糙集；K均值聚类；新空间
中图分类号：TP18 **文献标志码**：A **文章编号**：0253-987X(2011)08-0062-05

Multiple Instance Learning Method Based on Building New Bag Space

WEN Chao, GENG Guohua, LI Zhan
(School of Information Science and Technology, Northwest University, Xi′an 710069, China)

Abstract: Aiming at bag structure information loss problem caused by single instance deciding bag label in multiple instance learning (MIL), a new MIL algorithm named RK_BP is proposed. Firstly, rough set method is adopted to reduce the redundant information in the instance feature space, then K means algorithm is applied to cluster and build a new bag space, and finally back propagation algorithm is used to classify bags in the new space. Experiments on data sets show that this algorithm deals well with the multiple instance problems and provides better classification results.

Keywords: multiple instance learning; back propagation; rough set; K means clustering; new space

Dietterich在1997年研究药物活性预测问题时，第一次提出了多示例学习(MIL)问题.他在研究中发现，分子存在多种低能形状，而药物实验只能辨别分子参与制药能力，无法确定何种具体形状.他将分子看成包，分子的低能形状作为包中的一个示例.在该问题中，如果包标记为负，则包内所有示例标记都为负；如果包标记为正，则包中必然存在正示例，但是不能确定是哪个或哪些.正包示例标记的不确定就形成了MIL学习的难点，也就是歧义性问题.

针对多示例学习问题，文献[1]提出了轴平行矩形(APR)学习算法.该算法对属性值进行合取，在属性空间中寻找合适的轴平行矩形.APR算法找出一个覆盖了每个正包中至少一个示例的最小矩形，根据测试包是否有示例在轴平行矩形内给定包标记.文献[2]通过定义多样性密度(DD)函数，利用示例独立同假设和贝叶斯公式，寻找属性空间中的正包最多点和负包最远点.示例出现在潜在目标处的概率为示例间距离函数，由于引入了属性相关度参数，因而利用梯度下降法获得最大密度点后，待判定包标记由最大密度点的超椭球体判别.APR算法[1]与DD算法[2]都隐含着要求示例组成空间中正包的潜在正示例具有空间聚集性.文献[3]将DD算法与

收稿日期：2011-03-03. 作者简介：温超(1978－)，男，博士生；耿国华(联系人)，女，教授，博士生导师. 基金项目：国家自然科学基金资助项目(60873094)；高等学校博士学科点专项科研基金资助项目(200806970014).
网络出版时间：2011-07-18 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20110718.1730.007.html

期望最大(EM)算法相结合提出了 EM-DD 算法. 文献[4]通过修改支持向量机(SVM)的目标函数，从而形成了 mi-SVM 和 MI-SVM 算法. 另外，文献[5]提出了一种基于结构的 MIL 算法，分析了包的结构特性，定义了包空间的统计特性形成统计核. 文献[6]认为在不同 MIL 应用领域中，示例对包标记的影响不同，在此基础上提出了 P 后验混合模型(PPMM)核. 文献[7]则将包直接映射到一个无向图，将示例看作包中的一个节点，核函数为节点相似性和边相似性累加之和. 文献[8]提出了 MiniMIL 算法，利用流行学习对 MIL 问题进行了研究.

1　已有神经网络方法存在的问题

MIL 问题可以描述为寻找一个分类函数 $f:\boldsymbol{X}\rightarrow\Omega$，$\boldsymbol{X}$ 表示包含所有示例的集合，$\Omega=\{+1,-1\}$ 是包的标记集合. 所找到的分类函数 f 需正确预测未知包 B_i 的标记 $f(B_i)$，$B_i\subset\boldsymbol{X}$，$i=1,\cdots,s$. 在这里未知包的标记满足下面条件：如果 B_i 存在示例为正，$f(B_i)=+1$；如果 B_i 所有示例为负，$f(B_i)=-1$.

针对 MIL 分类问题，已有的神经网络方法[9-10]都基于单示例决定包标记. 这种方法隐含着要求示例间是示例独立同假设，不能有相关性，而这种假设通常不能成立. 在 MIL 的多个应用领域中，示例间通常具有很强的共生性，例如图像分类中老虎图像通常和森林、草地等其他草食动物出现在同一幅图像中. 忽略 MIL 中示例的共生性，在独立同假设下采取单示例决定标记的方法进行分类，会丧失包本身的结构性，从而影响 MIL 的分类性能. 本文提出的算法可较好地解决该问题.

2　相关定义

定义 1　B 表示 MIL 问题中所有包组成的集合，则有 $B=\{\bigcup_{i=1\sim s}B_i\}$，其中 s 为包个数. 令 B_i^+，$i=1,\cdots,n$ 代表第 i 个正包，n 为正包个数，同理令 B_i^-，$i=1,\cdots,m$ 代表第 i 个负包，m 为负包个数，满足 $s=m+n$.

定义 2　$\boldsymbol{X}_{ij}$ 表示 MIL 问题中第 i 个包的第 j 个示例，则有 $B_i=\{\bigcup_{j=1\sim l}\boldsymbol{X}_{ij}\}$，其中 l 为包中的示例个数，并用 $\boldsymbol{X}_{ij}^+$、$\boldsymbol{X}_{ij}^-$ 分别表示正包示例和负包示例. 令示例集合 $\boldsymbol{X}$ 为建立在示例属性域(R)上的 D 维向量空间，则 $\boldsymbol{X}_{ij}$ 可视为属性空间 R^D 中的一点.

定义 3　x_{ijp} ($p=1,\cdots,D$)表示例 $\boldsymbol{X}_{ij}$ 的第 p 个属性，则有 $\boldsymbol{X}_{ij}=\{\bigcup_{p=1\sim D}x_{ijp}\}$，$\boldsymbol{X}_{ij}$ 为由属性 x_{ijp} 组成的 D 维向量，如为药物活性问题，则 D 为 166.

定义 4　在 $\boldsymbol{X}$ 空间进行 K 聚类，定义聚类中心为 $\boldsymbol{O}=\{\boldsymbol{O}_1,\cdots,\boldsymbol{O}_k\}$

3　RK_BP 多示例学习算法

基于对 MIL 数据中存在冗余信息特性[11]和 MIL 问题具有包结构特性[7]的认识，本文提出了 RK_BP 算法. 算法采用前向贪心粗糙集(RS)算法[12]对 MIL 特征数据进行属性约简，在属性约简的基础上，利用 K 均值聚类获取聚类点和构建新包空间，在该空间最后利用 BP[13]进行分类.

3.1　前向贪心 RS 算法

粗糙集理论[14]的核心基础是从近似空间导出的一对近似算子，即上近似算子和下近似算子(又称上、下近似集)，属性约简是其核心内容. 属性约简在保持特征表达能力不变的条件下，删除其中不相关或不重要的属性. 文献[11]的实验结果表明，MIL 数据中可能存在冗余数据信息，因而采用粗糙集处理应能提高 MIL 的分类性能. Pawlak 粗糙集定义在经典的等价关系和等价类基础上，只适合处理名义型变量，对于现实应用中广泛存在的数值型数据却不能直接处理. 针对该问题，文献[12]提出了一种邻域粗糙集模型，并基于依赖性函数评价条件属性对分类的贡献，构造了一个适合于数值型属性的前向贪心约简算法. 其算法思想为：以空集为起点，每次计算全部剩余属性的重要度，从中选择重要度最大的属性加入约简集，直到所有剩余属性的重要度为 0，即加入任何新属性，系统依赖性函数值不发生变化.

3.2　新包空间构建

在构建新包空间的过程中，需要考虑两方面问题. 一方面，样本可分性的维持，也就是 MIL 问题中原示例向量空间 $\boldsymbol{X}$ 中的包是可分的，在新构建空间 $\boldsymbol{Y}$ 仍要保持其可分性. 另一方面，包结构性的保持，在新构建空间 $\boldsymbol{Y}$ 中，包 B_i 生成的新向量 $\boldsymbol{Y}_i$ 应保留原包的结构信息.

在考虑以上两个要求的情况下，首先应对已有方法进行分析以便得到启示. 在 MIL 问题中示例的属性空间寻找特殊点的经典算法为 DD 算法. DD 算法定义最大多样性密度点为属性空间中某点 d，该点附近出现的正包数最多，负包示例出现在远处，则寻找该点的目标函数可定义为

$$\arg\max_{x}\Pr(x=d\mid B_1^+,\cdots,B_n^+,B_1^-,\cdots,B_m^-)\quad(1)$$

假定各包独立，则根据 Bayes 理论，可通过下式

确定 d

$$\arg\max_x \prod_i (\Pr(x=d \mid B_i^+)) \prod_i \Pr(x=d \mid B_i^-) \tag{2}$$

式(2)就是多样性密度的一般定义,不妨设 d 为多样性密度值最大的点.在实际使用时,需要对式(2)中的乘积项进行示例化.文献[9]采用了 noisy-or 模型

$$\Pr(x=d \mid B_i^+) = 1 - \prod_j (1 - \Pr(x=d \mid \boldsymbol{X}_{ij}^+)) \tag{3}$$

$$\Pr(x=d \mid B_i^-) = \prod_j (1 - \Pr(x=d \mid \boldsymbol{X}_{ij}^-)) \tag{4}$$

将示例出现在潜在目标点 x 处的概率定义为该示例与潜在目标之间的距离函数,即

$$\Pr(x=d \mid \boldsymbol{X}_{ij}) = \exp(-\| \boldsymbol{X}_{ij} - \boldsymbol{x} \|^2) \tag{5}$$

如果用权 s_p 来表示第 p 个属性的相关度,则式(5)中的距离计算式为

$$\| \boldsymbol{X}_{ij} - \boldsymbol{x} \|^2 = \sum_p s_p (x_{ijp} - x_p)^2 \tag{6}$$

文献[15-16]与 DD 算法相关,都采用获取多个正负包示例局部最优 d 点的方式,然后分别针对支持向量机模型进行了适应性修改.正包示例 d 点的特点在于该点邻域正包数相对较多,负包离得较远,而负包示例 d 点的特点在于该点邻域负包数相对较多,正包离得较远.这些 d 点体现了正包示例和负包示例有很好的聚集性.K 聚类是常用的寻找空间集聚点的方法.如果将正包示例和负包示例的向量空间看成两类空间,则聚类方法和 DD 算法应该能得到相同性能.

通过上面的分析,采用 K 均值聚类方法,对原示例的向量空间 $\boldsymbol{X}$ 获取聚类中心点 $\boldsymbol{O}=\{\boldsymbol{O}_1,\cdots,\boldsymbol{O}_k\}$.$K$ 均值聚类方法是一种划分式聚类算法[17],其核心思想是找出 K 个聚类中心$\{\boldsymbol{O}_1,\cdots,\boldsymbol{O}_k\}$,使得每一个数据点 $\boldsymbol{X}_i$ 和与其最近的聚类中心的平方距离和被最小化,平方和公式为

$$J = \sum_{r=1}^{k} \sum_{i=1}^{n} \| \boldsymbol{X}_i^{(r)} - \boldsymbol{O}_r \|$$

定义 5 设$\{\boldsymbol{O}_1,\cdots,\boldsymbol{O}_k\}$是空间 $\boldsymbol{X}$ 的聚类中心,包 B_i 在新包空间的坐标定义为包中示例到聚类中心最小的欧氏距离

$$\boldsymbol{Y}_i = \min_{i=1\sim k\, j=1\sim m} \| \boldsymbol{X}_j - \boldsymbol{O}_i \|_2$$

定义 6 令集合$\boldsymbol{Y}=\{\bigcup_{i=1\sim k} \boldsymbol{Y}_i\}$,$\boldsymbol{Y}_i$ 表示 MIL 问题中的包 B_i 在新包空间中的坐标,则空间 $\boldsymbol{Y}$ 的维数为 K.

图 1 表示了构造新包空间的方法.假定 $\boldsymbol{X}$ 为平面中实数域 R 上的 2 维向量空间,选择 K 均值聚类的参数 K 为 2.这里有 4 个包,包中示例个数为 3、2、4、2.包中示例数据为 $B_1\{\boldsymbol{X}_{11}(0.5,0.5),\boldsymbol{X}_{12}(0.3,0.4),\boldsymbol{X}_{13}(0.2,0.9)\}$、$B_2\{\boldsymbol{X}_{21}(0.1,0.1),\boldsymbol{X}_{22}(0.7,0.8)\}$、$B_3\{\boldsymbol{X}_{31}(-1.5,-0.5),\boldsymbol{X}_{32}(-0.3,-1.4),\boldsymbol{X}_{33}(-0.2,-1.9),\boldsymbol{X}_{34}(-1.1,-0.9)\}$、$B_4\{\boldsymbol{X}_{41}(-0.1,-1.1),\boldsymbol{X}_{42}(-1.7,-0.8)\}$.聚类的中心为 $\boldsymbol{O}=\{\boldsymbol{O}_1(0.36,0.54),\boldsymbol{O}_2(-0.8167,-1.1)\}$.$B_4$ 包有两个示例 $\boldsymbol{X}_{41}$ 和 $\boldsymbol{X}_{42}$.计算 B_4 在新生成包空间中的坐标

$$\| \boldsymbol{X}_{41} - \boldsymbol{O}_1 \|_2 =$$

$$((x_{411}-o_{11})^2 + (x_{412}-o_{12})^2)^{1/2} =$$

$$((-0.1-0.36)^2 + (-1.1-0.54)^2)^{1/2} = 1.703\ 3$$

与此类似,得到 $\| \boldsymbol{X}_{41} - \boldsymbol{O}_2 \|_2 = 0.716\ 71$、$\| \boldsymbol{X}_{42} - \boldsymbol{O}_1 \|_2 = 2.457\ 5$ 和 $\| \boldsymbol{X}_{42} - \boldsymbol{O}_2 \|_2 = 0.932\ 9$,则得到 B_4 在新空间的坐标为(1.703 3, 0.716 71).

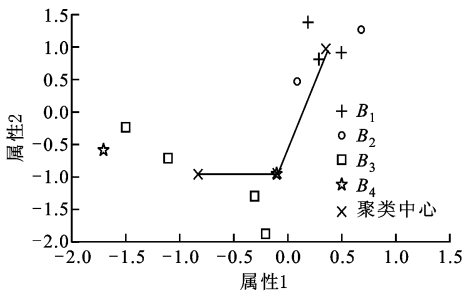

图 1 创建新包空间的方法

很显然,这种构造包空间的方法很好地保留了包的结构信息.下面,我们证明在这个新构造的包空间中包仍然可分.

定理 1 假定在 MIL 问题向量空间 $\boldsymbol{X}$ 中包可被函数 f 正确分类,则在新构造包空间 $\boldsymbol{Y}$ 中,包仍然可分.

证明 MIL 的核心问题在于其歧义性,也就是正包示例的不确定性.APR 和 DD 算法的成功,表明 MIL 问题的正包示例具有强聚集性.

通过 K 均值算法,找到聚类中心 $\boldsymbol{O}_i$,$i=1,\cdots,K$.假定 $\boldsymbol{O}_m$ 为正包示例的聚类中心,则可以通过包中示例和 $\boldsymbol{O}_m$ 的最小距离和某个阈值比较判断.如果存在某个 $\delta>0$,使得

$\min \| \boldsymbol{X}_{ij} - \boldsymbol{O}_m \|_2 \leqslant \delta$,if $f\exists \boldsymbol{X}_{ij} \in B_i$,$B_i$ 是正包

$\min \| \boldsymbol{X}_{ij} - \boldsymbol{O}_m \|_2 > \delta$, if $f\forall \boldsymbol{X}_{ij} \in B_i$,$B_i$ 是负包

显然,本文所构造的新包空间 $\boldsymbol{Y}$ 满足上面的条件,命题得证.

3.3 RK_BP 算法步骤

综上所述，基于新包空间和 BP 算法步骤总结如下：

(1)按照定义 2，将 MIL 问题中所有训练包组成示例集合 $\boldsymbol{X}$；

(2)采用 3.1 节前向贪心 RS 算法，针对示例集合 $\boldsymbol{X}$ 进行属性简约，获取简约集合 $\widetilde{\boldsymbol{X}}$；

(3)采用 K 均值聚类方法，按照定义 6 获取所有训练包新向量集合 $\boldsymbol{Y}$ 和测试包新向量集合 $\boldsymbol{Z}$；

(4)利用集合 $\boldsymbol{Y}$ 训练 BP 神经网络模型，并用该模型对集合 $\boldsymbol{Z}$ 进行预测.

4 实验结果与分析

实验中选择具有单隐层 BP 神经网络，其中输入层神经元个数为聚类值 K，隐含层神经元个数为 $2K$，输出层神经元个数为 1. 当输出值为[0.5，1]时是正包、[0，0.5]时是负包. 学习速率为 0.01. 由于新建的包空间依赖于 K 值的选择，因此首先对 K 值对算法的影响进行了实验，在此基础上分别对文献[1,4,15]所使用的数据库进行了对比实验.

4.1 K 值影响分析

K 值的选择对算法影响巨大，因此首先对 Musk1 数据和 Elephant 图像测试不同 K 值对算法分类性能的影响. 将 K 值设定为从 2 变化到 10，变化步长为 1. 图 2 给出了 K 值对测试数据库分类效果的影响.

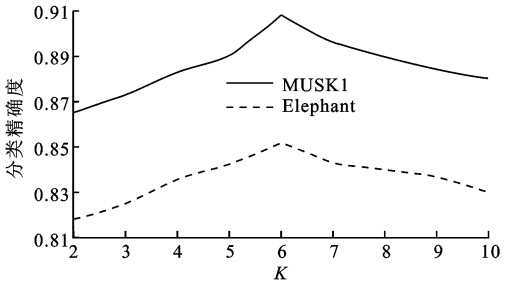

图 2　K 值对算法分类效果的影响

从图 2 可以看到，在 $K=6$ 时 Musk1 数据和 Elephant 图像都取得了最好的性能，因此在后面实验中，我们都选择 $K=6$ 进行聚类.

4.2 基准测试数据库

选择测试库 Musk、Elephant、Fox 和 Tiger 进行基准测试. Musk 数据库是由 Dietterich 等在研究分子活性问题时提出的，它由 Musk1 和 Musk2 组成，可在加州大学尔湾分校(UCI)机器学习知识库提供的地址处下载. Musk1 包括 47 个正包、45 个负包，其中正示例数为 207，负示例数为 269. 每个包中示例数从 2 到 40 不等. Musk2 包含 39 个正包、63 个负包，其中“实际”正示例数为 1 017，负示例数为 5 581. 包中示例数从 1 到 1 044. 表 1 列出了 Musk 数据库的详细信息.

表 1　Musk 数据库的详细信息

数据库	维数	包数			示例数	包中示例数		
		总数	正包	负包		最少	最多	平均
Musk1	166	92	47	45	476	2	40	5.17
Musk2	166	102	39	63	6 598	1	1 044	64.69

文献[4]从 Corel 数据库选择 Elephant、Fox 和 Tiger 三类动物图像，利用 Blobworld 系统分割. 一个图像特征由颜色、纹理和形状组成的 230 维向量表示，每类中有 100 个正包和 100 个负包组成. 表 2 列出了 Elephant、Fox 和 Tiger 图像数据的详细信息.

表 2　Elephant、Fox 和 Tiger 图像数据的详细信息

图像数据	维数	包数			示例数	包中示例数		
		总数	正包	负包		最少	最多	平均
Elephant	230	20	100	100	1 391	2	13	6.96
Fox	230	200	100	100	1 320	2	13	6.60
Tiger	230	200	100	100	1 220	1	13	6.10

通过 10 交叉验证进行实验，每 5 次作为一个分组，共进行 10 次分组实验. 将本文方法和文献[2-7,9]算法进行对比，表 3 给出了算法分类精确度之间的比较.

表 3　基准库分类结果比较

算法	分类精确度/%				
	Musk1	Musk2	Elephant	Fox	Tiger
RK_BP	91.3±3.6	90.6 ±3.2	85.3 ±2.1	70.0±2.6	85.8±1.9
MIGraph	90.0±3.8	90.0 ±2.7	85.1±2.8	61.2±1.7	81.9±1.5
miGraph	88.9±3.3	90.3±2.6	86.8±0.7	61.6±2.8	86.0±1.6
MI-Kernel	88.0±3.1	89.3 ±1.5	843.0±1.6	60.3±1.9	842±1.0
MI-SVM	77.9	84.3	81.4	59.4	84.0
mi-SVM	87.4	83.6	82.0	58.2	78.9
PPMM	95.6	81.2	82.4	60.3	82.4
DD	88.0	84.0			
EM-DD	84.8	84.9	78.3	56.1	72.1
BP_MIP	83.8±9.2				

注：加下划线的数据为各个测试库上得到的最好分类结果.

表3表明,我们所提出的RK_BP算法性能较优.相对于BP_MIP,在数据Musk1上算法分类精确度得到了很大提高,精确度从83%提高到91%.在Musk1、Musk2、Elephant和Tiger基准数据上,BP_MIP和miGraph性能差在2%以内,而且在Fox图像数据上,RK_BP算法取得了最好的性能.RK_BP算法性能较优的主要原因是:利用粗糙集对示例进行属性约简,可有效解决MIL中存在的数据冗余问题;融合K均值方法,提取多个概念点,并构建新包空间,可获取更好的包结构表达方式,从而大大提高了BP算法的性能,且具有很强的鲁棒性.其他方法或基于示例独立同假设条件,或基于单示例决定包标记假设,都未能充分考虑包结构问题,从而导致最终分类精度较低.

4.3 图像数据库

图像分类是MIL学习最成功的应用领域之一.本文选用文献[15]所提供的图像集,图像来自Corel库图像集中1k-Corel和2k-Corel,分别选取其中10类和20类图像,每一类包含100幅图像,将整个图像视为包,图像分割区域对应的视觉特征当作包中的示例.

本文采用和文献[12]相同的方法获取图像特征信息,使用基于颜色和空间关系的K均值方法分割图像,在每个区域获取9维特征向量(3维LUV颜色特征、3维Daubechies小波变换纹理特征和一、二、三阶归一化惯性因子所形成3维形状特征).实验中,将1k-Corel和2k-Corel图像集中每类随机等分成两份,一份训练,另一份测试实验进行5次,每次随机等分,采取一对一的方式进行分类测试.表4给出了算法分类精确度之间的比较.

表4 图像分类结果的比较

算法	分类精确度/%					
	1k-Corel			2k-Corel		
	平均值	分类1	分类2	平均值	分类1	分类2
RK_BP	82.6	79.3	85.9	719	65.3	72.5
MIGraph	83.9	81.2	85.7	72.1	71.0	73.2
miGraph	82.4	80.2	82.6	70.5	68.7	72.3
MI-Kernel	81.8	80.1	83.6	72.0	71.2	72.8
MI-SVM	74.7	74.1	75.3	54.6	53.1	56.1

从表4可以看到,本文提出的RK_BP算法取得了很好的分类精确度,与算法MIGraph性能差在2%以内.与其他算法一样,该算法也在分类Beach和Mountains时误分率最大,这主要是因为这两类有相同语义和视觉相似区域.

5 结 论

本文提出了一种新的RK_BP学习算法.该算法针对已有神经网络算法在解决MIL问题时分类性能不佳以及丢失包结构信息问题,提出了利用前向贪心RS算法属性约简,然后利用K均值聚类创建新包空间,最后使用BP进行分类.通过在多个测试数据库上的实验,表明该方法可较好解决以上问题.

RK_BP算法作为一种构造新空间的学习方法,它的不足之处在于BP性能和模型有密切联系,且与K均值聚类一样,构造空间过程依赖于初始值的选取,因此如何获取更优模型和合理选择初始值参数提高分类性能是一个值得进一步研究的课题.

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] MARON O, LOZANO-PÉREZ T. Advances in neural information processing systems 10: a framework for multiple-instance learning [M]. Cambridge, MA, USA: MIT Press, 1998:570-576.

[3] ZHANG Qi, GOLDMAN S A. Advances in neural Information processing systems 14: EM-DD: an improved multiple-instance learning technique [M]. Cambridge, MA, USA: MIT Press, 2002: 1073-1080.

[4] ANDREWS S, HOFMANN T, TSOCHANTARIDIS I. Multiple instance learning with generalized support vector machines [C]// Proceedings of the 18th National Conference on Artificial Intelligence. Edmonton, Canada: AAAI Press, 2002: 943-944.

[5] GARTNER T, FLACH P A, KOWALCZYK A, et al. Multi-instance kernels [C]// Proceedings of the 19th International Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufmam Publishers Inc, 2002: 179-186.

[6] WANG Huayan, YANG Qiang, ZHA Hongbin. Adaptive p-posterior mixture-model kernels for multiple instance learning [C]// Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland: ACM Press, 2008: 1136-1143.

[7] ZHOU Zhihua, SUN Yuyin, LI Yufeng. Multi-instance

learning by treating instances as non-I. I. D. samples [C] // Proceedings of the 26th International Conference on Machine Learning. Quebec, Canada: [n. s.], 2009: 1249-1256.

[8] 詹德川，周志华. 基于流形学习的多示例回归算法 [J]. 计算机学报，2006，29(11)：1948-1955.
ZHAN Dechuan. ZHOU Zhihua. A manifold learning-based multi-instance regression algorithm [J]. Chinese Journal of Computers, 2006, 29(11): 1948-1955.

[9] ZHOU Zhihua, ZHANG Minling. Neural networks for multi-instance learning[C] // Proceedings of the International Conference on Intelligent Information Technology. Beijing, China: People's Post and Telecommunications Publishing House, 2002: 455-459.

[10] ZHANG Minling, ZHOU Zhihua. Adapting RBF neural networks to multi-instance learning[J]. Neural Processing Letters, 2006, 23(1): 1-26.

[11] ZHANG Minling, ZHOU Zhihua. Improve multi-instance neural networks through feature selection [J]. Neural Processing Letters, 2004, 19(1): 1-10.

[12] HU Qinhua, YU Daren, XIE Zongxia. Numerical attribute reduction based on neighborhood granulation and rough approximation [J]. Journal of Software, 2008, 19(3): 640-649.

[13] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning internal representations by error propagation [C] // Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol 1 Foundations. Cambridge, MA, USA: MIT Press, 1986: 318-362.

[14] PAWLAK Z. Rough set [J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356.

[15] CHEN Yixin, WANG J Z. Image categorization by learning and reasoning with regions [J]. Journal of Machine Learning Research, 2004, 5(8): 913-939.

[16] KWOK J T, CHEUNG P M. Marginalized multi-instance kernels [C] // Proceedings of the 20th International Joint Conference on Artificial Intelligence. Hydrabad, India: AAAI Press, 2007: 901-906.

[17] MACQUEEN J. Some methods for classification and analysis of multivariate observations [C] // Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, CA, USA: University of California Press, 1967: 281-297.

（编辑　杜秀杰　武红江）