

# 图像语义分析的多示例学习算法综述

李大湘<sup>1,2</sup>, 赵小强<sup>1</sup>, 李娜<sup>1,2</sup>(1. 西安邮电大学 通信与信息工程学院, 西安 710121; 2. 陕西省  
法庭科学电子信息实验研究中心 图像处理实验室, 西安 710121)

**摘要:** 多示例学习(MIL)作为第4种机器学习框架,已在图像语义分析中得到了广泛应用. 首先介绍MIL的起源、特点、相关概念和数据集;然后以图像语义分析为应用背景,对相关MIL算法进行详细综述,按照算法采用的学习机制对其进行分类,并重点分析了各类算法提出的思路 and 主要特点;最后,对MIL未来的研究方向作了探讨.

**关键词:** 多示例学习; 图像分类; 图像检索; 图像语义分析

**中图分类号:** TP391

**文献标志码:** A

## Survey of MIL algorithms for image semantic analysis

LI Da-xiang<sup>1,2</sup>, ZHAO Xiao-qiang<sup>1</sup>, LI Na<sup>1,2</sup>

(1. School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 2. Lab of Image Processing, Crime Scene Investigation Unit of Shaanxi Province, Xi'an 710121, China. Correspondent: LI Da-xiang, E-mail: 35108809@qq.com)

**Abstract:** Multi-instance learning(MIL) has been recognized as the fourth machine learning framework, and has been widely used in the image semantic analysis. Firstly, the concepts such as development history, characteristics and many useful testing datasets of MIL techniques are reviewed. Then, many popular MIL algorithms are also introduced in detail by using real-world applications based on image semantic analysis. Meanwhile, based on their machine learning mechanisms, related MIL algorithms are divided into a variety of categories, which highlights the processes and dominant features of different MIL algorithms. Finally, the trends and possible outputs for further researches are discussed in details.

**Key words:** multi-instance learning; image categorization; image retrieval; image semantic analysis

## 0 引言

图像语义分析是以图像底层视觉特征为输入数据,分析图像中包含哪些主要对象或属于哪种场景类型,然后采用高层语义对图像的内容进行描述,从而实现图像的语义理解<sup>[1]</sup>. 其主要包括:语义图像检索<sup>[2]</sup>、语义图像分类<sup>[3]</sup>、图像标注<sup>[4]</sup>等研究内容.

由于“语义鸿沟”的存在,利用有监督学习框架建立图像的底层视觉特征与高层语义之间的联系时,需面临两个主要问题<sup>[5-6]</sup>: 1) 图像语义的表示问题,即研究如何表示图像所包含的各种高层语义; 2) 训练样本的标注问题,因为以手工标注的方式为有监督学习获得训练样本时不但费时费力,而且还带有主观偏差.

针对上述问题,文献[6]将图像视为包,将分割区域的底层视觉特征作为示例,在多示例学习(MIL)

的框架下对图像语义进行了分析. 较之传统有监督学习框架, MIL 的优势在于<sup>[7]</sup>: 1) MIL 的训练样本称为包,每个包中含有数量不等的多个示例,相对于单示例而言,多示例更能表达图像所包含的语义概念及相互关系; 2) 在训练样本手工标注过程中,标签只标注在包(图像)级而非示例(分割区域)级,简化了训练样本的标注过程. 因此, MIL 算法在图像语义分析领域得到了广泛应用,已成为机器学习界的一个研究热点<sup>[8-9]</sup>.

## 1 MIL的起源与特点

### 1.1 MIL的起源

1997年, Dietterich等<sup>[10]</sup>在药物活性预测的研究中首次正式提出了MIL的概念. 药物活性预测是指,通过对已知适于或不适于制药的分子进行分析来预

收稿日期: 2012-07-12; 修回日期: 2012-10-15.

基金项目: 国家自然科学基金项目(61003255); 陕西省教育厅自然科学基金项目(12JK0734, 12JK0504); 西安邮电大学青年基金项目(1090428).

作者简介: 李大湘(1974—),男,讲师,博士,从事机器学习与图像语义分析等研究; 赵小强(1978—),男,副教授,博士,从事图像处理等研究.

测某类新的分子是否可以用来制药. 其主要困难是: 每类分子都有很多种不同的低能形状, 在目前的生化技术条件下, 生物化学专家只知道哪些分子适于制药, 并不知道是哪一种具体形状起了决定性的作用. 如图 1 所示, 这是两个同类的分子, 因一个分子的某个化学键发生了旋转, 使分子形状相差很大. 若将适于制药的分子的所有低能形状都作为正例, 而将所有不适于制药的分子的所有低能形状都作为反例, 直接采用传统的有监督学习方法来训练分类器, 则会因为正样本中存在的噪声比例太高而无法获得理想的分类器.

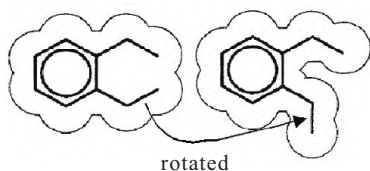


图 1 同一种分子的两种不同低能形状

针对上述问题, Dietterich 等<sup>[10]</sup>将分子视为包, 分子的每一种具体形状视为包中的一个示例, 当某个分子具有活性时 (适于制药), 它对应的包被标记为正包, 否则标记为负包, 由此提出了 MIL 框架. 在 MIL 训练过程中, 概念标号只分配给每个训练包, 至于包中示例的标号不必知道. 为了用一个特征向量对分子的形状进行描述, 首先按照相同的位置和方向对分子进行固定; 然后从分子的中心点均匀地放射出 162 条射线, 测量每条射线从中心点到分子表面之间截线段的长度, 用其作为分子形状描述的一个特征值, 还要加上 4 个固定氧原子位置的信息; 最后, 用一个 166 维的特征向量对分子的形状 (示例) 进行表示. MIL 的目标是通过分析已知的训练包学习得到一个分类器 (目标函数), 以尽可能正确地预测未知包的标号.

## 1.2 MIL 的特点

与传统有监督机器学习框架相比较, MIL 所具有的独特性质如下<sup>[11-12]</sup>.

1) 在 MIL 框架中, 训练样本称作包, 每个包中含有若干个 (数量不等) 不同的示例; 而传统的有监督学习框架中的训练样本与特征向量具有一对一的关系, 通常也称其为单示例学习, 可视为 MIL 中的特例. 二者训练样本的形式如图 2 所示.

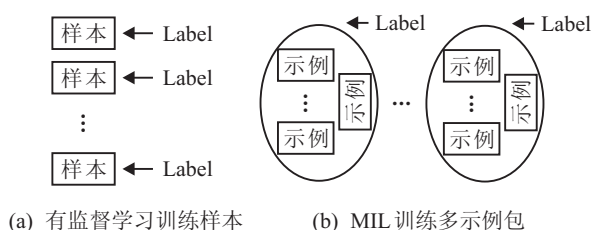


图 2 传统机器学习与 MIL 的训练样本

2) 在 MIL 训练过程中, 只知道包的标签, 包中的

示例标签是未知的; 而在有监督学习训练中, 每个训练样本的标号都是已知的.

3) 在 MIL 框架中, 一个包被标为“正”包, 当且仅当包中至少有一个示例是正确的, 否则将被标为“负”包, 当分类器训练出来之后, 测试包的标号由该包中的各个示例的最大输出值来决定; 而在传统的有监督学习中, 测试样本的标号由该对象本身的输出值来决定.

根据上述特点, MIL 框架与其他学习框架本质的区别是: 传统机器学习框架中的“对象: 描述: 类别”之间是“1: 1: 1”的关系, 即样本、特征与标号之间都是一对一的关系; 而在 MIL 框架中的“对象: 描述: 类别”之间是“1: N: 1”的关系, 即样本和特征之间是一对多的关系.

在 MIL 框架中, 因其训练样本标号具有独特的歧义性, 且是机器学习中的一个研究盲区, 故被认为是一种并列于强化学习、有监督学习与非监督学习的第 4 种学习框架<sup>[13]</sup>.

## 2 MIL 的相关概念

在 MIL 框架下进行图像语义分析, 主要涉及到的相关概念如下.

**定义 1 (MIL 建模)** MIL 建模是指研究如何将图像构造成多示例包的形式, 从而将图像语义分析问题转化成 MIL 问题进行求解. 设  $X$  表示一幅图像, 常用的方法是: 采用图像分割 (或分块) 的方法对图像  $X$  进行自动分割, 并提取每个分割区域的底层视觉特征 (例如: 颜色、纹理和形状特征等), 将整幅图像当作包, 分割区域的底层视觉特征当作包中的示例. 最后, 由  $X$  构造的多示例包记为  $X = \{x_1, x_2, \dots, x_n\}$ . 其中:  $n$  表示图像分割的区域数,  $x_i$  表示第  $i$  个区域的底层视觉特征向量.

**定义 2 (MIL 的数学定义)** 设  $\chi = R^d$  为一个  $d$  维的示例空间,  $\Omega = \{-1, +1\}$  表示二类标记空间. 数据集  $D = \{(X_1, y_1), (X_2, y_2), \dots, (X_m, y_m)\}$  表示  $m$  个已标注的多示例训练包. 其中: 包  $X_i (i = 1, 2, \dots, m)$  中含有  $n_i$  个示例  $\{x_{i1}, x_{i2}, \dots, x_{in_i}\}$ ,  $x_{ij} \in \chi (j = 1, 2, \dots, n_i)$ ;  $y_i$  表示与  $X_i$  相对应的标号. MIL 的任务是通过学习  $D$ , 得到一个分类函数  $f_{\text{MIL}}: 2^\chi \rightarrow \Omega$ , 使其能预测任意未标记包  $X$  的标号<sup>[10,14]</sup>.

根据多示例包中示例与包标签之间的关系, MIL 大致可分成如下两种模型.

**模型 1 (标准 MIL 模型)** Dietterich 等<sup>[10]</sup>总结了关于分子活性预测的研究工作, 提出了标准的 MIL 模型. 该模型认为每个示例都具有一个隐含的类别标号  $c \in \Omega = \{+1, -1\}$ , 一个包被标为正的充要条件是它

至少包含一个示例是正例。

令  $X = \{x_1, x_2, \dots, x_n\}$  表示一个包含  $n$  个示例  $x_i \in \chi$  ( $i = 1, 2, \dots, n$ ) 的包,  $c(x)$  表示一个示例级的分类函数, 则包的标号为

$$f_{\text{MIL}}(X) = \begin{cases} 1, & \exists c(x_i) = 1; \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

该模型特别适合于图像中的“对象语义”分析问题。这是因为, 对象语义属于简单语义, 图像中的单个区域即能表达。如图 3(a) 所示, 图像中第 2 和第 3 个区域便能表示“horse”这个对象语义。



(a) “horse”对象语义 (b) “beaches”场景语义

图3 图像的“对象语义”与“场景语义”样图

**模型2 (广义 MIL 模型)** 广义 MIL 模型最先是 Weidmann 等<sup>[15]</sup>提出的, 并且定义了 3 种广义 MIL 范式, 即: 基于出现的多示例学习、基于阈值的多示例学习和基于计数的多示例学习范式。在广义 MIL 模型中, 正包的标号不是由包中的单个示例来决定, 而是由包中的多个示例共同作用才能决定。

例如: 在基于出现的 MIL 范式中, 认为存在  $r$  个示例级的概念  $C = \{c_1, c_2, \dots, c_r\}$  ( $c_i: \chi \rightarrow \Omega$ ) 来共同决定包的标号。设  $\Delta(X, c_i)$  表示多示例包  $X$  中与概念  $c_i$  对应的示例个数, 则包  $X$  的标号为<sup>[14-15]</sup>

$$f_{\text{PB-MIL}}(X) = \begin{cases} 1, & \forall c_i \in C: \Delta(X, c_i) \geq 1; \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

显然, 若包  $X$  被标为“正”, 则对于  $C$  中每一个概念  $c_i: i = 1, 2, \dots, r$ , 在  $X$  中至少存在一个示例与它相对应。

该范式特别适于图像的“场景语义”分析问题。如图 3(b) 所示, 对于“beaches”场景图像, 图像中的单个区域无法表达它所属的场景类型, 只有当图像中至少有一个区域表示“sea”、“sky”、“sands”等概念时, 该图像才可被预测为“beaches”场景。

### 3 MIL 的测试数据集

为了客观评估各种 MIL 算法的性能, 常用标准测试集如下。

#### 3.1 Musk 数据集

Dietterich 等<sup>[10]</sup>在药物活性预测工作中建立了 Musk 数据集 (<http://www1.cs.columbia.edu/andrews/mil/datasets.html>)。该数据集是从实际的麝香分子数据中生成的, 包括 Musk 1 和 Musk 2 两个子集, 它们成

为衡量 MIL 算法性能好坏的标准测试集, 相关细节信息如表 1 所示。通过数据对比可见, 对 Musk 2 的学习难度远大于 Musk 1。

表1 Musk 数据集的相关细节信息

数据集	Musk 1	Musk 2	数据集	Musk 1	Musk 2
正包数量	47	39	最少示例数	2	1
负包数量	45	63	最多示例数	40	1044
总示例数	476	6598	平均示例数	5.2	64.7

#### 3.2 COREL 2k 数据集

Chen 等<sup>[16]</sup>在用 MIL 进行图像分类时, 建立了 COREL 2k 图像集, 且已成为基于 MIL 的图像检索或分类算法的标准测试集。该数据集从 COREL 图像库中选取 20 类图像, 即: Africa、beach、building、buses、dinosaurs、elephants、flowers、mountains、food、dogs、horses、lizards、fashion models、sunset scenes、cars、waterfalls、antique furniture、battle ships、skiing 和 desserts, 每一类包含 100 幅彩色图像, 共 2 000 幅图像。

该图像集中的每幅图像均被预分割成约 5 个区域, 且每个区域的一个 9 维的底层视觉特征向量(颜色、纹理和形状)已经被提取, 细节可参阅文献[16], 所有图像与特征数据集可下载于 <http://cs.olemiss.edu/~ychen/ddsvm.html>。

#### 3.3 SIVAL 数据集

Rahmani 等<sup>[17]</sup>在进行基于对象的图像检索时, 构造了 SIVAL (spatially independent, variable area and lighting) 数据集, 主要用于半监督 MIL 算法的对比实验。该图像集是一个人为设计的对象图像检索标准测试集, 包含 25 类不同对象, 每类 60 幅图像, 分别放置在 10 种不同场景中, 且放置位置和光照条件分为 6 种不同情况, 构成一个 1 500 幅图像的数据库。每幅图像被预先分割成约 30 个区域, 每个区域 30 维的颜色、纹理和近邻的底层视觉特征也被提取出来, 细节参阅文献[17], 所有图像和数据可从 <http://www.cs.wustl.edu/accio> 下载。

除上述主要测试集外, 还有来自 COREL 的 5 种场景 (waterfalls、fields、mountains、sunsets 和 lakes)<sup>[18]</sup>, 3 种动物 (elephant、fox 和 tiger)<sup>[19]</sup> 图像集, 以及 Zhou 等<sup>[20]</sup>建立的 2 000 幅多示例多标签图像集 (分 20 类, 每类 100 幅) 等。

### 4 MIL 的主要算法与分析

根据 MIL 算法的发展脉络与学习机制, 对各类代表性 MIL 算法进行综述。

#### 4.1 轴平行矩形 (APR) 算法

Dietterich 等<sup>[10]</sup>在提出 MIL 概念的同时, 基于贪婪属性选择 (GFS) 与迭代鉴别提出了 3 种轴平行矩形

(APR)算法, 即: GFS elim-count APR, GFS elim-kde APR 和 Iterated-discrim APR 算法. 其共同思想是: 采用贪婪或迭代的搜索策略, 对示例向量的属性值进行合取, 在示例空间中寻找合适的轴平行矩形, 使其包含尽可能多的正示例, 且尽量不含负示例, 以此作为最终的目标概念区. APR 算法作为后续 MIL 的研究基础, 在图像检索或分类问题中, 已成为衡量其他 MIL 算法性能的对比基准而参与各种对比实验<sup>[17,19-20]</sup>.

## 4.2 多样性密度 (DD) 算法

多样性密度 (DD) 算法是 Maron 等<sup>[18]</sup>提出的, 且用于图像分类问题. 其思路是: 在特征空间中找到一个概念点, 使每个正包中至少有一个示例离该点足够近, 而所有负包中的示例均远离该点; 然后, 度量新包中的每一个示例与概念点之间的距离来判断该示例的标签.

设  $B_i^+$  表示第  $i$  个正包,  $B_{i,j}^+$  表示第  $i$  个正包的第  $j$  个示例,  $B_{i,j,k}^+$  表示第  $i$  个正包的第  $j$  个示例的第  $k$  个属性值; 类似地, 设  $B_i^-$  表示第  $i$  个负包,  $B_{i,j}^-$  表示第  $i$  个负包的第  $j$  个示例,  $B_{i,j,k}^-$  表示第  $i$  个负包的第  $j$  个示例的第  $k$  个属性值. 令  $h$  代表 DD 值最大的点, 通过最大化  $\Pr(x = h|B_1^+, \dots, B_{l^+}^+, B_1^-, \dots, B_{l^-}^-)$  目标函数来确定目标点  $h$ . 设包之间是相互独立的, 由 Bayes 理论, 可通过下式寻找  $h$  点<sup>[22,29]</sup>:

$$\operatorname{argmax}_h \left( \prod_i l^+ P(h|B_i^+) \prod_i l^- P(h|B_i^-) \right). \quad (3)$$

式 (3) 称为多样性密度函数, 即 DD 函数. 在具体问题求解时, noisy-or 模型被用来量化式 (3) 中的乘积项, 即

$$\Pr(x = h|B_i^+) = 1 - \prod_j (1 - \Pr(x = h|B_{i,j}^+)),$$

$$\Pr(x = h|B_i^-) = \prod_j (1 - \Pr(x = h|B_{i,j}^-)). \quad (4)$$

此时, 根据目标概念点  $h$  与示例  $B_{i,j}$  之间的距离,  $B_{i,j}$  隶属于概念点  $h$  的条件概率定义为如下形式:

$$\Pr(x = h|B_{i,j}) = \exp(-\|B_{i,j} - x\|^2), \quad (5)$$

其中  $\|B_{i,j} - x\|^2 = \sum_k w_k^2 (B_{i,j,k} - x_k)^2$ . DD 算法的优点是: 在确定多样性密度最大点的同时, 还可得到示例属性的相关度, 即得到一个概念点位置  $h$  的同时, 还可以得到一组反映属性相关度的权值  $w_k$ . 但因 DD 函数是一个高阶非线性连续函数, 其存在多个局部极大值点, 为了得到最优解, 需把所有正包中的每个示例都作为搜索起点, 利用梯度下降法进行穷举搜索, 其训练时间开销相当大.

针对 DD 算法的不足, Zhang 等<sup>[21]</sup>结合期望最大

化 (EM) 方法的思想, 提出了一种改进型 DD 算法, 即 EM-DD 算法. 对比实验结果表明, 预测精度与 DD 算法相当, 但训练效率得到了大幅度提高; 此外, 文献 [22] 将弱分类器 boosting 的过程视为一个梯度下降的过程, 提出了一种 EM-DD-boosting 算法; 文献 [23] 则对 DD 函数进行了改进, 并结合量子粒子群优化 (QPSO) 算法, 提出了 QPSO-MIL 算法, 用于图像标注问题.

## 4.3 基于 KNN 的惰性 MIL 算法

Wang 等<sup>[24]</sup>对  $k$ -近邻 ( $k$ NN) 算法进行了扩展, 使用修正的 Hausdorff 距离来度量多示例包之间的相似度, 提出两种惰性 MIL 算法, 即 Bayesian- $k$ NN 和 Citation- $k$ NN. 在 Bayesian- $k$ NN 算法中使用 Bayes 理论对邻包进行分析, 从而获取新包的概念标记; 而 Citation- $k$ NN 算法则借用了科学文献“引用与参考”的概念, 要同时考虑新包的 cite 和 reference 的近邻关系. 该算法的优点是: 分类前不需要训练分类器, 且易于采用反馈的方式将错分样本加入训练集, 以提高下一次的分类精度; 缺点是: 由于需要保存所有的训练数据, 存储开销比较大.

此外, 专门针对场景图像分类问题, Li 等<sup>[25]</sup>采用改进的推土机距离<sup>[26]</sup>来度量多示例包之间的相似性, 提出了一种新的惰性 MIL 算法, 用于图像分类和检索问题.

## 4.4 基于 SVM 的 MIL 算法

支持向量机 (SVM)<sup>[27]</sup>作为一种经典的机器学习方法已成功地用于求解 MIL 问题, 且已成为 MIL 中非常重要的一类方法.

### 4.4.1 mi-SVM 和 MI-SVM 算法

Andrews 等<sup>[19]</sup>对 SVM 进行扩展, 将 MIL 的约束引入 SVM 的目标函数中, 提出了 mi-SVM 和 MI-SVM 两种基于 SVM 的 MIL 算法. 其中: mi-SVM 称为基于示例的 SVM 算法, MI-SVM 称为包的 SVM 算法. 在 MIL 问题中, 最难处理的问题是正包中示例的标签, 因为仅仅知道其中存在至少一个真正的正示例, 但并不知道哪个或者哪几个示例真是正的. MI-SVM 算法是从每个正包中选出一个最正的示例, 认为是正样本, 把这些样本和其他负包中的样本放在一起用 SVM 进行训练分类器; 而 mi-SVM 算法是给每个正包中的样本赋予一个伪标签, 然后用 SVM 训练分类器. 其中: mi-SVM 算法的目标函数为

$$\min_{\{y_i\}} \min_{\{w, b, \xi\}} \frac{1}{2} \|w\|^2 + C \sum_i \xi_i;$$

$$\text{s.t. } \forall i: y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, y_i \in \{-1, +1\},$$

$$\sum_{i \in I} \frac{y_i + 1}{2} \geq 1, \forall I, Y_I = 1, \\ \text{and } y_i = -1, \forall I, Y_I = -1. \quad (6)$$

MI-SVM算法的目标函数为

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_I \xi_i; \\ \text{s.t. } \forall I: Y_I(\langle w, x_i \rangle + b) \geq 1 - \xi_I, \xi_I \geq 0. \quad (7)$$

式(6)和(7)中:  $w, b$ 表示SVM的分类超平面参数;  $\xi_i$ 表示SVM的松弛系数;  $x_i, y_i$ 分别表示包中的示例和示例标签;  $I, Y_I$ 分别表示包与包的标签. 从以上两个目标函数可知, mi-SVM方法的最大化margin是基于包中单个样本的, 在原来的SVM基础上引入了对包的标签的约束, 即正包中应含有一个以上的正样本, 而负包中应全是负样本. MI-SVM方法的margin是基于单个包的, 它认为包的软标签等于包内正概率最大的样本.

该类方法的优点是: 直接对SVM的目标函数进行修改, 用于求解MIL问题, 具有很好的泛化能力和小样本学习能力. 其缺点是: 待求解的目标函数是一个非凸的优化问题, 无法直接求其代数解, 而要采用启发式的迭代算法进行逼近, 训练效率较低, 且容易受局部极值的影响. 对此, Gehler等<sup>[28]</sup>采用确定性退火的方法求解这两个优化问题; Bunescu等<sup>[29]</sup>通过在优化目标函数中引入新的松弛变量, 再利用CCCP (concave-convex solver) 进行求解, 实验证明效果更好.

#### 4.4.2 基于核的方法

传统的SVM学习方法中核是定义在两个单样本之间的, 而Gartner等<sup>[30]</sup>直接在多示例包之间定义了核, 用来度量包之间的相似性, 直接把MIL问题转化成SVM学习问题进行求解. 其提出了如下多示例核:

##### 1) 集核

$$k_{\text{set}}(B, B') = \sum_{x \in B, x' \in B'} k(x, x'); \quad (8)$$

##### 2) 统计核

$$k_{\text{stat}}(B, B') = k(s(B), s(B')). \quad (9)$$

其中:  $B, B'$ 表示多示例包;  $x, x'$ 表示示例;  $s(B)$ 是反映包的性质的一个统计量, 如: 均值、中值、最大值、最小值等. 例如, 在最小最大核里, 有

$$s(B) = (\min_{x \in B} x_1, \dots, \min_{x \in B} x_m, \max_{x \in B} x_1, \dots, \max_{x \in B} x_m), \quad (10)$$

其中  $x_i (i = 1, 2, \dots, m)$  表示示例  $x$  的第  $i$  个分量值.

该类方法的优点是: 直接在多示例包之间定义核函数, 用SVM求解MIL问题, 具有很好的泛化能力和小样本学习能力. 其缺点是: 核函数在本质上体现的是多示例包之间的相似性, 这种相似性无法反映MIL

框架中正包中至少包含一个正示例, 负包中全是负示例的性质. 针对这些问题, Kwok等<sup>[31]</sup>提出了边缘核的概念, Zhou等<sup>[32]</sup>提出了图核, 它们更能从包的内部结构上反映多示例包之间的相似度. 图像分类实验表明, 这类方法的分类精度更高.

#### 4.5 基于元数据的MIL算法

该类算法的思想是: 将多示例包转化成单个样本, 即用某个特征向量来表征每个多示例包, 从而将MIL转化成标准的单示例学习(有监督学习)问题进行求解. 其中最经典的有Chen等提出的DD-SVM<sup>[16]</sup>和MILES (multiple-instance learning via embedded instance selection)<sup>[33]</sup>算法, 均用于图像分类问题. 在DD-SVM算法中, 首先基于DD函数采用拟牛顿搜索法获得一组概念点  $\{(x_k^*, w_k^*) : k = 1, 2, \dots, n\}$ , 用来构造一个目标空间. 对于任意的多示例包  $B_i = \{x_{ij} : j = 1, 2, \dots, N_i\}$ ,  $x_{ij}$ 表示第  $j$  个示例,  $N_i$ 表示示例数量, 其投影特征为

$$\varphi(B_i) = \begin{bmatrix} \min_{j=1, \dots, N_i} \|x_{i,j} - x_1^*\|_{w_1^*} \\ \min_{j=1, \dots, N_i} \|x_{i,j} - x_2^*\|_{w_2^*} \\ \vdots \\ \min_{j=1, \dots, N_i} \|x_{i,j} - x_n^*\|_{w_n^*} \end{bmatrix}. \quad (11)$$

利用式(11), 将  $B_i$  转化成单个样本, 于是MIL问题被转化成一个传统的有监督学习问题, 从而利用SVM进行求解. 由于DD-SVM要用DD方法寻找概念点, 其训练效率极低, Chen等<sup>[33]</sup>又提出了改进型的MILES算法, 该算法简单地利用训练包中的所有示例作为概念点, 用来构造目标空间. 对于任意多示例包  $B_i$ , 其投影特征定义为

$$m(B_i) = [s(x^1, B_i), s(x^2, B_i), \dots, s(x^n, B_i)]^T. \quad (12)$$

其中:  $s(x^k, B_i) = \max_j \exp(-\|x_{ij} - x^k\|/\delta^2)$ ,  $x^k$ 表示第  $k$  个概念点,  $\delta$ 是一个预先确定的尺度参数. 在MILES算法中, 因为通过式(12)获得包的投影特征的维数很高, 存在很多与分类无关的冗余信息, 所以Chen<sup>[33]</sup>采用了1范数SVM进行学习, 即在SVM学习的同时, 具有特征选择的功能, 但1范数SVM的学习与分类能力往往不如2范数SVM<sup>[34]</sup>.

针对上述问题, Li等<sup>[34]</sup>提出了基于“视觉空间投影”与“模糊潜在语义分析”<sup>[35]</sup>的转化方法, 将MIL转化成有监督学习. 此外, 还有YARDS<sup>[36]</sup>、BARTMIP<sup>[37]</sup>、MICCLLR<sup>[38]</sup>和CCE<sup>[39]</sup>等方法, 目前已成为一类非常有效的MIL算法, 均用于图像检索或分类问题.

#### 4.6 半监督的MIL算法

在训练图像语义分析分类器时, 为了利用大量未标记图像来提高分类器的性能, Rahmani等<sup>[40]</sup>把基于



图的半监督学习与 MIL 相结合, 提出一种基于图的半监督 MIL 算法, 即 MISSL 算法, 用于对象图像检索问题. 该方法首先计算包之间的相似性, 建立基于包的相似矩阵, 对每个包引入“能量”的概念, 认为半监督学习即是一个正包中的标签向未标注包的传递过程. 其存在的问题是: 能量函数定义过于经验化, 缺乏理论上的推导, 包与包之间的相似性度量也不精确, 常导致最后结果出现偏差.

除此之外, 还有 Zhou 等<sup>[41]</sup>提出的 missSVM 算法、Wang 等<sup>[42]</sup>提出的 GMIL 算法和 Zhang 等<sup>[43]</sup>提出的基于直推式 SVM 的 UP-SSMIL 算法等, 均已用于基于对象的图像检索问题.

#### 4.7 其他 MIL 算法

除了上面提到的几类 MIL 算法之外, 还有一些算法不能归入其中, 在此只作简单介绍. Zhang 等<sup>[44]</sup>通过设计一个新颖的多示例误差函数, 成功地把人工神经网络应用于多示例学习, 设计了多示例神经网络算法; Ruffo 等<sup>[45]</sup>提出了基于 C4.5 决断树的 MIL 算法; Deselaers 等<sup>[46]</sup>将包视为条件随机场 (CRF) 的节点, 将示例视为节点的状态, 提出了基于 CRF 的 MIL 算法.

此外, 还有基于概念学习的 MIL 算法<sup>[47]</sup>、基于流形的算法<sup>[48]</sup>和基于集成学习的 MIL 算法<sup>[49-50]</sup>等. 这些 MIL 算法的基本思想都是用已有的机器学习方法解决多示例的问题<sup>[14,51]</sup>.

### 5 MIL 的总结与展望

本文以图像语义分析为应用背景, 对机器学习领域中具有代表性的 MIL 算法作了比较全面的综述, 并分析了各类 MIL 算法的主要特点. MIL 框架的提出, 不但丰富了机器学习的相关理论, 而且很多实际问题非常适于采用 MIL 求解, 具有广阔的应用前景. 近年来, 虽然很多学者致力于这方面的研究工作, 并发表了很多研究成果, 但仍有如下几种问题尚待进一步研究.

1) 包内示例数目对分类性能的影响. 例如: Musk 2 较之 Musk 1 数据集, 包中示例的数量差别很大, 而实验结果表明<sup>[10,18]</sup>, 在 Musk 1 数据集上预测精度稍高一些. 因此, 关于包中示例的数量对 MIL 算法的性能有怎样的影响, 值得进一步研究.

2) 正示例数量的不平衡问题. 根据标准 MIL 模型, 即当一个包标为正包时, 其中至少包含一个真正的正示例. 对于 MIL 算法而言, 因为事先并不知道正包中正示例的个数, 通常当正包中真正为正的示例个数太少时, 会影响预测精度, 所以, MIL 中数据的不平衡性对算法性能的影响程度值得进一步研究.

3) 多示例多标签学习. 多示例多标签的概念是 Zhou 等<sup>[20]</sup>首次提出的, 因为在很多实际问题 (如图像检索) 中, 每幅图像的不同局部区域可能对应不同的语义概念, 即每幅图像可能对应着多个不同的语义标签. 虽然 Zhou 等提出了多种多示例多标签学习算法<sup>[52-54]</sup>, 但其中研究的空間还很大.

4) MIL 中的特征选择和降维. 在机器学习中, 特征选择和降维是一个重点研究内容, 其往往对算法的性能影响较大. 但是对于 MIL 而言, 如何进行特征选择和降维仍没有系统深入的研究. 因此, 探讨特征选择和降维对 MIL 的影响, 是一个很有意义的研究方向<sup>[55]</sup>.

5) 新的 MIL 应用领域. 最初, MIL 主要用于分子活性预测、图像分类、图像检索、对象检测、文本分析、网页推荐和股票预测等<sup>[51]</sup>, 最近, MIL 被用于运动目标跟踪<sup>[56-57]</sup>和异常行为检测<sup>[58]</sup>等领域, 因此, 探索新的 MIL 应用领域将具有十分重要的意义.

#### 参考文献 (References)

- [1] 高隽, 谢昭, 张骏, 等. 图像语义分析与理解综述[J]. 模式识别与人工智能, 2010, 23(2): 191-202.  
(Gao J, Xie Z, Zhang J, et al. Image semantic analysis and understanding: A review[J]. Pattern Recognition and Artificial Intelligence, 2010, 23(2): 191-202.)
- [2] Manuel G, Miguel A V. An endmember-based distance for content based hyperspectral image retrieval[J]. Pattern Recognition, 2012, 45(9): 3472-3489.
- [3] Qin Jianzhao, Yung Nelson H C. Feature fusion within local region using localized maximum-margin learning for scene categorization[J]. Pattern Recognition, 2012, 45(4): 1671-1683.
- [4] Wu Fei, Tian Qi, Zhuang Yueting. Image annotation by input-output structural grouping sparsity[J]. IEEE Trans on Image Processing, 2012, 21(6): 3066-3079.
- [5] Yang Jun, Jiang Yu gang, Alexander G Hauptmann, et al. Evaluating bag-of-visual-words representations in scene classification[C]. Proc of the Int Workshop on Multimedia Information Retrieval. Bavaria, 2007: 197-206.
- [6] Zhang Dan, Wang Fei, Shi Zhenwei, et al. Interactive localized content based image retrieval with multiple-instance active learning[J]. Pattern Recognition, 2010, 43(3): 478-484.
- [7] Li D X, Fan J L, Wang D W, et al. Latent topic based multi-instance learning method for localized content-based image retrieval[J]. Computers & Mathematics with Applications, 2012, 64(4): 500-510.

- [8] Zhu Shunle, Tan Xiaoqi. A novel automatic image annotation method based on multi-instance learning[J]. *Procedia Engineering*, 2011, 15(8): 3439-3444.
- [9] Jeremy Bolton, Paul Gader, Hichem Frigui, et al. Random set framework for multiple instance learning[J]. *Information Sciences*, 2011, 181(10): 2061-2070.
- [10] Dietterich T G, Lathrop R H, Tomas L P. Solving the multiple instance problem with axis-parallel rectangles[J]. *Artificial Intelligence*, 1997, 89(12): 31-71.
- [11] 王春燕. 基于多示例学习的图像检索方法研究[D]. 北京: 北京林业大学信息学院, 2010.  
(Wang C Y. Research on image retrieval method based on multi-instance learning[D]. Beijing: School of Information, Beijing Forestry University, 2010.)
- [12] 黄波. 基于支持向量机的多示例学习研究与应用[D]. 武汉: 中国地质大学计算机学院, 2009.  
(Huang B. Research and application on multi-instance learning using support vector machine[D]. Wuhan: School of Computer, China University of Geosciences, 2009.)
- [13] Zhou Z H. Multi-instance learning: A survey[R]. Nanjing: Department of Computer Science & Technology, Nanjing University, 2004.
- [14] James Foulds, Eibe Frank. A review of multi-instance learning assumptions[J]. *Knowledge Engineering Review*, 2010, 25(1): 1-25.
- [15] Weidmann N. Two-level classification for generalized multi-instance data[D]. Freiburg: University of Freiburg, 2003.
- [16] Chen Yixin, Wang Jame Ze. Image categorization by learning and reasoning with regions[J]. *J of Machine Learning Research*, 2004, 5(8): 913-939.
- [17] Rahmani R, Goldman S A, Zhang H, et al. Localized content based-image retrieval[C]. *Proc of ACM Workshop on Multimedia Image Retrieval*. November, 2005: 227-236.
- [18] Maron Oded, Ratan Aparna Lakshmi. Multiple-instance learning for natural scene classification[C]. *Proc of the 15th Int Conf on Machine Learning*. Madison, 1998: 341-349.
- [19] Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple-instance learning[C]. *Proc of Advances in Neural Information Processing Systems*. Vancouver: MIT Press, 2002: 561-568.
- [20] Zhou Z H, Zhang M L. Multi-instance multi-label learning with application to scene classification[C]. *Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2007: 1609-1616.
- [21] Zhang Q, Goldman S. EM-DD: An improved multiple instance learning technique[C]. *Advances in Neural Information Processing System*. Vancouver, 2001: 1073-1080.
- [22] Mason L, Baxter J, Bartlett P, et al. Boosting algorithms as gradient descent in function space[C]. *Proc of Advances in Neural Information Processing Systems*. Denver Mariott City Center: IEEE Press, 1999: 132-138.
- [23] 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.  
(Li D X, Peng J Y, Bu Q R. QPSO-based multi-instance learning for image annotation[J]. *Computer Science*, 2010, 37(6): 278-282.)
- [24] Wang Jun, Zucker Jean-Daniel. Solving the multiple-instance problem: A lazy learning approach[C]. *Proc of the 17th Int Conf on Machine Learning*. San Francisco, 2000: 1119-1125.
- [25] 李大湘, 彭进业, 贺静芳. 基于 EMD-kNN 多示例学习算法的图像分类[J]. *光电子·激光*, 2010, 21(2): 302-306.  
(Li D X, Peng J Y, He J F. Image categorization based on EMD-kNN multi-instance learning algorithm[J]. *J of Optoelectronicso Laser*, 2010, 21(2): 302-306.)
- [26] Yossi R, Carlo T, Leonidas J G. The earthmover's distance as a metric for image retrieval[J]. *Int J of Computer Vision*, 2000, 40(2): 99-121.
- [27] Chang C C, Lin C J. LIBSVM: A library for support vector machines[EB/OL]. (2012-06-12). <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [28] Peter V Gehler, Olivier Chapelle. Deterministic annealing for multiple-instance learning[C]. *Proc of the 11th Int Conf on Artificial Intelligence and Statistics*. San Juan, 2007: 123-130.
- [29] Bunescu R C, Mooney R J. Multiple instance learning for sparse positive bags[C]. *Proc of the 24th Int Conf on Machine Learning*. New York, 2007: 105-112.
- [30] Gartner T, Peter A F, Adam K, et al. Multi-instance kernels[C]. *Proc of the 19th Int Conf on Machine Learning*. Sydney, 2002: 179-186.
- [31] James T K, Cheung P M. Marginalized multi-instance kernels[C]. *Proc of the 20th Int Joint Conf on Artificial Intelligence*. Hyderabad, 2007: 901-906.
- [32] Zhou Z H, Sun Y Y, Li Y F. Multi-instance learning by treating instances as non-i.i.d. samples[C]. *Proc of the 26th Annual Int Conf on Machine Learning*. Montreal, 2009: 1249-1256.
- [33] Chen Yixin, Bi Jinbo, Wang James Ze. MILES: Multiple-instance learning via embedded instance selection[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 1931-1947.
- [34] 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.  
(Li D X, Peng J Y, Li Z. Object-based image retrieval using semi-supervised MIL algorithm[J]. *Control and Decision*, 2010, 25(7): 981-986.)

- [35] Li D X, Peng J Y, Li Zhan, et al. LSA based multi-instance learning algorithm for image retrieval[J]. Signal Processing, 2011, 91(8): 1993-2000.
- [36] Foulds J. Learning instance weights in multi-instance learning[D]. Hamilton: Department of Computer Science, University of Waikato, 2008.
- [37] Zhang M L, Zhou Z H. Multi-instance clustering with applications to multi-instance prediction[J]. Applied Intelligence, 2009, 31(1): 47-68.
- [38] Yasser E M, Vasant Honavar. MICLLR: Multiple-instance learning using class conditional log likelihood ratio[J]. Discovery Science, 2009, 58(8): 80-91.
- [39] Li W J, Yeung D Y. Localized content-based image retrieval through evidence region identification[C]. 2009 IEEE Conf on Computer Vision and Pattern Recognition. Miami, 2009: 1666-1673.
- [40] Rouhollah Rahmani, Sally A Goldman. MISSL: Multiple-instance semi-supervised learning[C]. Proc of the 23rd Int Conf on Machine Learning. Pittsburgh: Carnegie Mellon University Press, 2006: 705-712.
- [41] Zhou Z H, Xu J M. On the relation between multi-instance learning and semi-supervised learning[C]. Proc of the 24th Int Conf on Machine Learning. Corvallis, 2007: 1167-1174.
- [42] Zhou Z H, Xu J M. On the relation between multi-instance learning and semi-supervised learning[C]. Proc of the 24th Int Conf on Machine Learning. Corvallis, 2007: 1167-1174.
- [43] Zhang Dan, Shi Zhenwei, Song Yangqiu, et al. Localized content-based image retrieval using semi-supervised multiple instance learning[C]. Asian Conf on Computer Vision. Tokyo, 2007: 180-188.
- [44] 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.  
(Zhang M L, Zhou Z H. A multi-instance regression algorithm based on neural network[J]. J of Software, 2003, 14(7): 1238-1242.)
- [45] Ruffo G. Learning single and multiple instance decision tree for computer security applications[D]. Torino: Department of Computer Science, University of Turin, 2000.
- [46] Deselaers T, Ferrari V. A conditional random field for multiple-instance learning[C]. Proc of the 27th Int Conf on Machine Learning. CA: AAAI Press, 2010: 287-294.
- [47] Chevalyre Y, Zucker J. A framework for learning rules from multiple instance data[C]. Proc of the 12th European Conf on Machine Learning. Freiburg: IEEE Press, 2001: 49-60.
- [48] 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 D C, Zhou Z H. A manifold learning-based multi-instance regression algorithm[J]. Chinese J of Computers, 2006, 29(11): 1948-1955.)
- [49] Zhou Z H, Zhang M L. Ensembles of multi-instance learners[C]. Proc of the 14th European Conf on Machine Learning. Berlin, 2003: 492-502.
- [50] Kang X P, Li D Y, Wang S G. A multi-instance ensemble learning model based on concept lattice[J]. Knowledge-based Systems, 2011, 24(6): 1203-1213.
- [51] 蔡自兴, 李枚毅. 多示例学习及其研究现状[J]. 控制与决策, 2004, 19(6): 607-611.  
(Cai Z X, Li M Y. Multi-instance learning and its current research status[J]. Control and Decision, 2004, 19(6): 607-611.)
- [52] Wang W, Zhou Z H. Learnability of multi-instance multi-label learning[J]. Chinese Science Bulletin, 2012, 57(19): 2488-2491.
- [53] Zhou Z H, Zhang M L, Huang S J, et al. Multi-instance multi-label learning[J]. Artificial Intelligence, 2012, 176(1): 2291-2320.
- [54] Li Y X, Ji S, Kumar S, et al. Drosophila gene expression pattern annotation through multi-instance multi-label learning[J]. IEEE Trans on Computational Biology and Bioinformatics, 2012, 9(1): 98-112.
- [55] Amelia Zafra, Mykola Pechenizkiy, Sebastián Ventura. HyDR-MI: A hybrid algorithm to reduce dimensionality in multiple instance learning[J]. Information Sciences, 2011, 24(6): 1-20.
- [56] 周秋红. 基于多示例学习的运动目标跟踪算法研究[D]. 大连: 大连理工大学电子与信息工程学院, 2011.  
(Zhou Q H. Tracking of moving object based on multi-instance learning[D]. Dalian: School of Electronic and Information Engineering, Dalian University of Technology, 2011.)
- [57] Ali S. Human action recognition in videos using kinematic features and multiple instance learning[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2010, 32(2): 28-303.
- [58] 崔永艳. 基于多示例学习的异常行为检测方法研究[D]. 南京: 南京大学计算机科学与技术系, 2011.  
(Cui Y Y. Study on abnormal event detection based on the multi-instance learning[D]. Nanjing: Department of Computer Science and Technology, Nanjing University, 2011.)