

图像检索中结合文本信息的多示例原型选择及主动学习策略

李 净^{1,2*}, 郭洪禹¹

(1. 上海海洋大学 信息学院, 上海 201306; 2. 同济大学 电子与信息工程学院, 上海 201804)

(* 通信作者电子邮箱 j_li@shou.edu.cn)

摘 要: 针对基于区域的图像检索系统检索精度不高的问题, 提出结合文本信息的多示例原型选择算法和反馈标注机制。在示例原型选择时, 首先使用文本信息进行正例拓展, 然后通过估计负例分布进行最初示例选择, 最后通过示例更新和分类器学习的交替优化获得真的示例原型。相关反馈采用了多策略相结合的主动学习机制, 通过信息值控制主动学习策略的自动切换, 使系统能够自动选择当前最适合的主动学习策略。实验结果表明, 该方法有效且性能优于其他方法。

关键词: 多示例学习; 文本信息; 示例原型; 主动学习; 相关反馈

中图分类号: TP18; TP391 **文献标志码:** A

Multi-instance prototype selection and active learning combined with textual information in image retrieval

LI Jing^{1,2*}, GUO Hong-yu¹

(1. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China;

2. School of Electronics and Information, Tongji University, Shanghai 201804, China)

Abstract: For the poor precision of region-based image retrieval, Multi-Instance Learning (MIL) prototype selection algorithm and feedback mechanism with reference to textual information were proposed. In the process of instance prototype selection, textual information was used to extend the positive examples, and negative example distribution was used to select initial instances and the iterative optimization process of instance updating and classifier training were used to obtain the true instance prototypes. In the process of relevance feedback, active learning with the combined learning methods was adopted. The switch of active learning strategy was controlled by the information value in the feedback process. The experimental results show that this algorithm is feasible, and the performance is superior to other MIL algorithms.

Key words: Multi-Instance Learning (MIL); textual information; instance prototype; active learning; relevance feedback

0 引言

图像检索系统中, 图像底层视觉特征的抽取主要包括全图特征抽取和局部图像区域特征抽取两种方式。虽然基于全图的特征抽取方式非常简单, 但是研究表明: 图像特征的局部抽取方式更接近于人类的理解; 图像特定的某个局部区域或某几个局部区域能够更加有效地描述图像内容, 所以当前的图像检索系统大多采用基于局部区域的特征抽取方式。此时, 图像区域由一个视觉特征向量表示, 可以看作一个示例, 整个图像为一个示例集合, 可以看作一个包, 检索过程中用户对图像(包粒度)的标注将图像检索被映射为一个多示例学习问题(Multi-Instance Learning, MIL)。

MIL的概念最早由Dietterich等提出, 并用于药物活性预测问题。不同于传统的监督学习, MIL是在示例包有标签, 包中示例没有明确标签的情况下, 对包粒度训练集建模并预测未知包类标签的学习方式。到目前为止, MIL主要包括三类解决方法: 多样性密度(Diverse Density, DD)及改进算法^[1-2]、单示例框架下增加约束的MIL^[3-4]以及构造特征空间(示例原型)并通过特征映射将MIL转化为单示例学习的DD-SVM^[5]、MILES^[6]、EC-SVM^[7]和MILIS^[8]方法。其中第三类方法在图像检索中得到广泛的应用, 然而DD-SVM没有进行示例原型选择, 计算量大; MILES虽是DD-SVM的改进, 但

MILES包含了所有无关原型而导致维数灾难; EC-SVM和MILIS虽然进行了特征选择, 但却都没有有效利用图像的文本信息, 因此, 本文旨在充分利用文本信息, 设计一个结合文本信息的示例原型选择方法。

此外, 为了减少用户在相关反馈过程对图像的标注量, 将主动学习应用于MIL。多示例主动学习包括包层主动学习^[9]、示例层主动学习^[10]和混合层主动学习^[11]。虽然Liu等^[11]的研究表明包/示例混合层样本选取策略普遍优于另外两种策略, 但在实际的图像检索中用户无法准确标注图像区域示例, 因此, 本文旨在充分利用文本信息, 设计一个有效、自动切换的多示例主动学习策略。

1 图像检索系统流程

图像检索系统的流程如图1: 1) 用户给出查询样例, 系统随机返回 n 个图像; 2) 如果用户对返回的结果满意就结束, 否则如果是首轮系统只对随机返回的图像进行手工标注; 如果非首轮, 则进行多策略主动学习返回最富信息量的图像进行手工标注; 3) 进行结合文本的MIL学习并返回top- n 个图像, 转2)。

2 结合文本信息的MIL

结合文本信息的MIL算法为MILIS^[8]基础上的改进, 其

收稿日期: 2012-04-17; 修回日期: 2012-06-18。 基金项目: 国家863计划项目(2009AA11Z220)。

作者简介: 李净(1977-), 女, 山西阳泉人, 副教授, 博士, 主要研究方向: 图像检索、机器学习; 郭洪禹(1974-), 女, 黑龙江齐齐哈尔人, 讲师, 博士研究生, 主要研究方向: 模式识别、语音处理。

算法为:首先进行结合文本信息的示例原型选择,然后使用式(3)将图像库中的图像 B_i 映射到示例原型特征空间中,最后使用示例原型选择已经训练好的 SVM 进行结果分类。

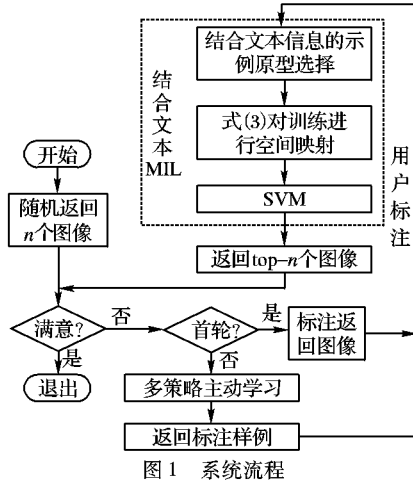


图1 系统流程

2.1 结合文本信息的示例原型选择

为叙述方便, B_i^+ 表示一个正例图像包; B_i^- 表示一个负例图像包; B_{ij}^+ 表示正例图像包 B_i^+ 中的第 j 个示例; B_{ij}^- 表示负例图像包 B_i^- 中的第 j 个示例。 $B = \{B_1^+, B_2^+, \dots, B_{n^+}^+, B_1^-, B_2^-, \dots, B_{n^-}^-\}$ 表示用户相关反馈后的图像包样本集合, 其中包括 n^+ 个正例包和 n^- 个负例包。 $Y = \{y_1, y_2, \dots, y_n\}$, 其中, $y_i \in \{1, -1\}$ 为反馈图像包对应的用户标注。令, $K^+ = \{K_1^+, K_2^+, \dots, K_r^+\}$ 和 $K^- = \{K_1^-, K_2^-, \dots, K_r^-\}$ 分别为用户反馈的相关或不相关文本关键字集合。

结合文本信息的示例原型选择是对 MILIS 的改进, 该算法为一迭代过程, 包括示例筛选、特征映射、分类其训练以及示例原型更新四个步骤, 其流程如图 2。

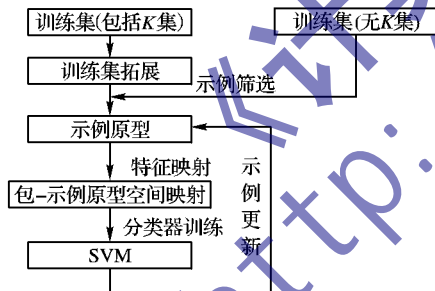


图2 示例原型选择流程

在示例筛选操作中如果训练集包括用户反馈的文本关键字, 选出最有可能为正例的 n' 个拓展图像包加入训练集; 如果训练集中不包括反馈文本关键字, 维持训练集不变。在训练集上, 示例筛选操作对负示例分布进行建模, 然后从每个正例包和拓展包中选择一个与负示例分布最不相似的示例作为一个原型示例, 此时示例原型的数目等于训练集中正例包(包括拓展的图像包)的数目。选出示例原型后通过式(3)将训练集中的包图像进行映射得到包级特征向量, 然后以此为训练样本对 SVM 进行训练, 最后现有的示例原型进行更新直到收敛, 然后删除小于阈值 θ 的示例原型。

1) 包图像特征表示。为有效使用基于特征映射的 SVM, 首先需要定义图像包 B_i 与示例原型示例 x_k 间的相似度函数为:

$$s(B_i, x_k) = \exp(-\lambda d'(B_i, x_k)) = \max_{B_{ij} \in B_i} \exp\left(-\frac{d(B_{ij}, x_k)^2}{\sigma^2}\right) \quad (1)$$

其中: $d(B_{ij}, x_k)$ 为示例间加权欧氏距离(2.2 节将详细介绍权

值学习); σ 反映示例点影响范围, 在 MILE 中, σ 值从一个通过交叉验证预定义候选值集合中选择。这种方法不适用于小样本集合, 这里通过训练包中示例分布来估计 σ 值。不同示例原型影响的范围不同, 则其对应的 σ 也不同(σ_k 对应 x_k)。通过计算 d_k^+ (d_k^-) 作为原型示例 x_k 与 n^+ (n^-) 个正(反)包的平均最好距离来估计 σ_k , $\sigma_k = (d_k^+ + d_k^-)/2$, d_k^+ 和 d_k^- 为:

$$\begin{cases} d_k^+ = \sum_{i=1}^{n^+} \frac{\min_j d(x_k, B_{ij})}{n^+} \\ d_k^- = \sum_{i=1}^{n^-} \frac{\min_j d(x_k, B_{ij})}{n^-} \end{cases} \quad (2)$$

相似度函数式(1)采用的指数形式加强了邻居点示例点的作用, 却减弱了其他点的作用。如果 x 为真的正示例, 正包与 x 的相似值(式(1))很高, 而负包与 x 的相似值较低。使用包到示例相似度函数, 图像包 B_i 被映射到示例原型空间, 其向量记为 S_i 。

$$S_i = [s(B_i, x_1), s(B_i, x_2), \dots, s(B_i, x_n)] \quad (3)$$

2) 示例原型筛选。如果训练集中包含反馈的相关文本关键字集 K^+ 和 K^- , 采用伪反馈方式拓展训练集的正例包。具体步骤如下所示。

① 计算数据库中各图像的条件概率:

$$P(I_i | K^+, K^-) = \frac{P(I_i, K^+, K^-)}{P(K^+, K^-)} = \frac{P(K^+, K^- | I_i) P(I_i)}{P(K^+, K^-)} \quad (4)$$

假设数据库中所有图像的概率 $P(I_i)$ 相等, 对相同的 K^+ , K^- 集 $P(K^+, K^-)$ 为一个常量, 则式(4)简化为求 $P(I_i | K^+, K^-)$ 即可。假设文本关键字集 $K = K^+ \cup K^-$ 中各个关键字相互独立, 则: $P(K^+, K^- | I_i) = P(K^+ | I_i) P(K^- | I_i)$ 。通过 Noisy-Or 模型对 $P(K_i | I_i)$ 进行估计:

$$P(K^+ | I_i) = 1 - \prod_{j \in K^+} [1 - F_{i, K_j}^K] \quad (5)$$

$$P(K^- | I_i) = \prod_{j \in K^-} [1 - F_{i, K_j}^K] \quad (6)$$

其中 F_{i, K_j}^K 为图像 I_i 文本特征向量中关键字 K_j 维的值。

② 将数据库中的图像按 $P(I_i | K^+, K^-)$ 降序排列, 将前 n' 个图像作为拓展正例集加入到训练集中, 参与示例原型的选择。

示例原型筛选是从拓展训练集的所有示例中选出一个有效的示例原型子集, 即选择真正的正示例和负示例组成示例原型子空间来计算有判别特征图。由于多示例学习假设, 正例包中至少包含一个正例, 而负例包中所有示例都为负例, 所以示例原型选择中最重要的问题在于寻找真正的正示例。虽然基于 DD 的选择方法可以直接定位正示例, 但是 DD 方法计算量太大。负例包是明确的、无二义性, 所以对负例包中的示例分布进行建模对定位真的正示例比基于 DD 的方法更有效。鉴于负示例具有一般分布, 采用与 MILIS 相同的基于高斯核的核密度估计(Kernel Density Estimation, KDE)对负例包中的所有示例建模。基于高斯核的核密度估计为:

$$f(x) = \frac{1}{C \sum_{i=1}^{n^-} n_i} \sum_{i=1}^{n^-} \sum_{j=1}^{n_i} \exp\left(-\frac{\|x - B_{ij}^-\|^2}{\sigma^2}\right) \quad (7)$$

其中: C 为归一化因子, C 在示例空间中为一个常数, 本系统忽略此常数。 σ 为尺度参数, 用于控制对训练集示例的影响范围。 $f(x)$ 为示例点 x 的密度值, 它是一个归一化的概率密度函数, 也表示示例点 x 与这些负示例间的相似度。

对负例包中的所有示例建模完成后, 对正例包集以及拓

展集中的所有示例使用式(7)进行概率密度估计,然后对每个正包集及拓展集中的每个图像分别选择一个概率密度值最小的示例作为初始示例原型(共 $n' + n^+$ 个示例原型)。

该方法具有如下优点:1) 多示例学习中,负示例点比正示例点要多得多,用负示例点进行概率密度建模比用正示例点建模更准确;2) 为减少计算量,对真的正示例点的密度估计可以选择该示例点的 K 个近邻负示例点来进行,后面的实验证明其有效性(目前有许多加速近邻搜索的方法^[12]);3) 伪反馈实际是一个基于文本关键字的查询,只通过文本特征对图像进行度量,这样可以使那些视觉信息不相关但文本信息相关的图像可以参与原型示例点的选择,拓展了示例原型的数目。

3) 分类器学习。将上步选出的示例原型记为 b_i^* , $i = 1, 2, \dots, n^+, \dots, n^+ + n'$, 使用式(3)将未拓展训练集中的所有图像包样本进行映射得到一个新的训练集合 $S = \{S_1, S_2, \dots, S_{n^+ + n'}\}$ 和 $Y = \{y_1, y_2, \dots, y_n\}$, 对 L^2 范数线性 SVM 训练后得到分类器的权值向量 W 。 L^2 范数线性 SVM 的求解为:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_i (1 - y_i W^T S_i)^2 \quad (8)$$

4) 示例原型更新。初始示例原型的选择不基于负示例的密度估计,没有结合分类器进行综合考虑。示例更新就是从正例包及拓展正例包中重新选择示例以提高 SVM 的分类性能,即进一步最小化式(8)。式(8)的第一项是常量,对式(8)的优化相当于选择那些可以减小式子第二项的示例原型。定义一个标志向量 $F = \{F_1, F_2, \dots, F_{n^+ + n'}\}$, 其中 $F_i = j$ 表示第 i 个正例包中的第 j 个示例被选为示例原型; $g(B_i, F) = [s(B_i, B_{1F_1}), s(B_i, B_{2F_2}), \dots, s(B_i, B_{nF_n})]$, $n = n^+ + n'$ 表示图像包 B_i 到示例原型空间的映射向量。示例更新如下:

$$\begin{cases} F_1^{t+1} = \operatorname{argmin}_j \sum_i (1 - y_i W^T g(B_i, \{j, F_2^t, \dots, F_n^t\}))^2 \\ \dots \\ F_i^{t+1} = \operatorname{argmin}_j \sum_i (1 - y_i W^T g(B_i, \{F_1^t, F_{i-1}^t, j, F_{i+1}^t, \dots\}))^2 \\ \dots \\ F_n^{t+1} = \operatorname{argmin}_j \sum_i (1 - y_i W^T g(B_i, \{F_1^t, F_2^t, \dots, F_n^t\}))^2 \end{cases} \quad (9)$$

其中: F_i^t 和 F_i^{t+1} 分别代表第 i 个正例包中第 t 轮和第 $t+1$ 轮所对应的示例序号。示例更新过程简单,有效,因为:1) 虽然每个 F_i 的更新简单地将其余 F 固定,但是每个示例 F_i 的更新都进一步缩小了分类器的错误上界。2) 算法对示例原型的更新顺序没有要求。由于对分类正确的示例包先更新不可能减少分类器的错误上界,所以先对错分类的正例包进行示例更新。3) 示例更新时,可以通过增量计算方式加快计算速度。更新 F_i 时,包到示例原型空间的映射只有第 i 个特征维发生变化,所以只需要用第 i 个特征的变化值乘以对应权值去更新分类器的值即可。

上述示例原型选择和分类器训练是以两步优化的方式交叉进行。分类器被训练好时,固定 SVM 的权值进行示例更新。示例更新后,各视觉分量权值和包-示例空间的特征映射被更新,分类器再从更新的特征映射中重新学习。系统性能随迭代次数增加而增加,后续实验将迭代次数固定为 3,系统已经取得了不错的效果。

2.2 基于反馈的示例间距离定义

2.1 节的包到示例映射由示例原型与示例间的距离 $d(B_{ij}, x_k)$ 决定。如果示例间距离能够将正例类与其他类别分开,分类器的性能就越好。这里,通过相关反馈定义一个权值

动态调整的距离公式 $d(B_{ij}, x_k)$ 。设图像示例的视觉特征维度为 V , 加权欧氏距离为:

$$d(B_{ij}, x_k) = \sqrt{\sum_{p=1}^V w_p' \cdot (B_{ij,p} - x_{k,p})^2} \quad (10)$$

某个视觉分量越能够很好区分正、反例包,该分量对应的权值 w_p' 越大。这里采用 Fisher 标准 J_p ^[13] 和判别比率 δ_p 进行判别式分析,其定义为:

$$J_p = \frac{(\mu_p^+ - \mu_p^-)^2}{(\sigma_p^+)^2 + (\sigma_p^-)^2} \quad (11)$$

$$\delta_p = 1 - \frac{|LN_p|}{|NI_p|} \quad (12)$$

其中: μ_p^+ (μ_p^-)、 $(\sigma_p^+)^2$ ($(\sigma_p^-)^2$) 分别表示正(负)示例分布的均值和方差; $|NI_p|$ 表示负示例的数目; $|LN_p|$ 表示第 p 维视觉特征落入正示例第 p 维视觉特征范围的负示例数。将判别比率 δ_p 和 J_p 合并,视觉分量权值定义为:

$$w_p' = \frac{Q_p}{\sum_p Q_p}; Q_p = \frac{\delta_p J_p}{\sigma_p^+} \quad (13)$$

注意:1) 视觉权值学习依赖于反馈样本以及上一轮示例原型;2) 视觉权值学习后进行本轮的示例原型选择;3) 检索初期还没进行示例原型选择,设各视觉分量初始权值为特征数的倒数。

多示例学习只能获得包标识,无法直接计算正、负示例分布的均值、方差。正示例分布可以通过原型示例来构造;负示例分布可以用两种方式获得:1) 选择线性 SVM 中的负支持向量中的所有示例作为负示例;2) 将负包中的所有示例作为负示例。由于线性 SVM 不一定能够很好地选择负支持向量,因此采用第二种负示例构造方法。直观而言,由于负包中所有示例是负的,所以使用所有负示例将能提供一个好的类分布估计。后续实验证明了方法的有效性。

2.3 主动学习策略

在示例标签可以被准确标注的情况下,混合层样本选择策略优于其他两种样本选择策略。但在交互式的图像检索系统中,用户无法对大量的图像区域示例进行准确、快速标注,混合层样本查询策略也不适用于此系统。因此探索文本关键字标注和图像包标注的相关反馈模式是有效的,因为:1) 对用户而言,文本信息是比图像区域示例更直观的信息,用户能够快速标注此信息;2) 对某个文本关键字的标注可以获取与之相关图像集合,它方便探索数据空间中的示例聚类;3) 包图像不仅能被用户方便标注,还能有效更新分类超平面。

1) 包图像主动学习策略。最不确定包选择是最传统的主动学习策略,在本文使用的 SVM 分类器中,与决策平面的距离是对这种不确定性的最好估计。最不确定图像包被定义为: $\operatorname{Crit}^{\text{UB}}(B_i) = \operatorname{argmin}_i |f(B_i)|$ 在数据分布已知时该策略非常有效。

系统初期示例原型不具有代表性,此时通过选择被 SVM 分类为正的包图像来探索数据分布非常必要,此策略被称为最相似策略(MRS)。在使用 SVM 进行分类时,最可能为正的包图像定义为:

$$\operatorname{Crit}^{\text{PB}}(B_i) = \operatorname{argmax}_i f(B_i)$$

为保证新选图像有效,选择的图像不应该与已有训练包图像学习的示例原型相似,即使用 $\operatorname{Crit}^{\text{PB}}$ 策略迭代地选择最可能为正的包图像直到被选图像不与示例原型相似。此主动学习策略为:

$$\operatorname{Crit}^{\text{MRS}}(B_i) = \operatorname{argmax}_i f(B_i)$$

s. t. $f(B_i) > t_i$

$$\max_i s(B_i, x_i) < t_2 \quad (14)$$

参数 t_1, t_2 的选择依赖于: 1) 正类与负类间的不平衡度。两者间越不平衡, 正例包的选择就越重要, 两个参数也应该越松弛。2) 示例密度。示例分布越密选择与之相似的示例的几率就越大, 参数 t_2 应该越低。

2) 基于文本关键字的主动学习策略。文本关键字在示例原型选择时用于拓展正例集, 丰富示例原型选择范围。显然, 通过选取最相关的文本关键字有利于正例集拓展。通过贝叶斯公式计算文本关键字 K_j 的相关程度:

$$P(K_j | (I^+, I^-)) = \frac{P((I^+, I^-) | K_j) * P(K_j)}{P(I^+ I^-)} = \frac{\prod_{i=1}^{n^+} P(I_i^+ | K_j) \cdot \prod_{w=1}^{n^-} P(I_w^- | K_j) \cdot P(K_j)}{P(I^+ I^-)} = \frac{\left\{ \prod_{i=1}^{n^+} F_{i, K_j}^K \cdot \prod_{w=1}^{n^-} (1 - F_{w, K_j}^K) \right\} \cdot P(K_j)}{P(I^+ I^-)} \quad (15)$$

假设各关键字被选则的概率相同, 对不同 K_j , 式(15) 的变化部分为:

$$\prod_{i=1}^{n^+} F_{i, K_j}^K \cdot \prod_{w=1}^{n^-} (1 - F_{w, K_j}^K)$$

则最相似关键字策略为:

$$Crit^{MPK}(K_j) = \operatorname{argmax}_j \left(\prod_{i=1}^{n^+} F_{i, K_j}^K \cdot \prod_{w=1}^{n^-} (1 - F_{w, K_j}^K) \right) \quad (16)$$

2.4 多策略相结合的反饋

不同检索阶段各主动学习策略适用性不同, 如: 检索初期 MRS 优于 UB, 但随反馈的进行, UB 可能又优于 MRS。在图像检索系统中, 结合使用各种主动学习策略(系统动态使用更优秀的学习策略)是值得探索的一种方式。

1) 包图像策略的切换。

检索初期我们首先使用 MRS 策略选取样例, 当式(14) 中两个约束条件中任一个无法满足时, 使用随机策略选择包图像; 当示例原型空间稳定时, 使用最不确定包策略修正分类超平面。交互式在线系统中, 通过计算策略的信息值控制随即策略到 UB 策略的切换。信息值大表示当前策略有效, 系统仍然保持当前策略; 信息值小表示当前策略不在有效, 需要切换到另一学习策略。

示例原型系统中更多关注正例包, 通过新的正例包与训练集间的差异来衡量信息值。如果新的正例包中的示例与当前示例原型不同, 就认为此示例包是富含信息的。式(17) 计算第 t 轮时正例包 B_i^+ 与当前示例原型的差异, 式(18) 计算所有正例包的差异平均。

$$(D_i^+)_t = \max_k s((B_i^+)_t, x_k) \quad (17)$$

$$(D^+)_t = \frac{\sum_{i=l_t^+}^{l_{t+1}^+} (D_i^+)_t}{l_t^+ - l_{t-1}^+} \quad (18)$$

其中: l_{t-1}^+ 和 l_t^+ 分别表示在 $t-1$ 轮和 t 轮时正包的数目。 $(D^+)_t$ 小表明新的正例包与当前示例原型不同, 也表明上一轮的学习策略有效, 本轮依然使用此策略, 反之进行策略切换。使用 $(D^+)_t$ 构造策略保持概率 p_t 为:

$$p_t = \max[\min(p_{t-1} \exp(-\gamma(D^+)_t), 1 - \varepsilon), \varepsilon] \quad (19)$$

2) 包策略与文本关键字策略的混合反馈。

文本关键字用于拓展正例训练集, 因此在各阶段采用 MPK 策略与相关包策略结合的方式。由于很难衡量文本关键字策略与包策略的关系, 本文采用同时标注方式, 即, 两策

略产生的查询结果同时返回给用户标注。特别地, 在 MPK + MRS 情况下, 最相似包图像和最相似关键字集被同时返回, 系统检索返回的图像与主动学习标注图像相同, 此时样例标注过程和系统检索过程统一。其他两种方式下, 系统检索返回的图像和主动学习标注图像是不同, 系统通过将标注图像与结果图像分区显示将样例标注过程和系统检索过程统一。

3 实验分析

本文的实验平台为: 双核 2.8 GHz 处理器, 2 GB 内存, Win XP 和 VC++。由于与经典的 MILES 进行实验对比, 本文采用与 MILES 相同的图像分割和图像特征提取方式对 Corel5K 进行离线处理; 采用如下方法学习图像的文本关键字特征: 首先选择一小部分图片作为训练集(每个主题随机选取 20 张图片, 共 1000 张图片), 然后采用 Tang 等^[14] 提出的方法对关键字繁衍, 归一化之后构建了图像文本关键字特征(共 374 维)。

3.1 示例原型选择

示例原型选择实验将迭代次数固定为 3, 示例删除阈值 θ 取值为 2, KDE 中的近邻数为 20, 尺度参数 σ 取为 0.04。将对数据集随机划分, 40% 为训练数据, 60% 用于测试数据, 同时每个分类随机标注 20 个文本关键字。表 1 为系统在 5 次不同划分情况下, MILES、EC-SVM、MIPS(本文方法)以及 MIPS-1(MIPS-1: 不考虑文本关键字时的示例原型选择)的平均正确率及训练各分类器所需的计算时间。实验表明即使在不使用文本关键字时, 本文方法在系统性能上也优于其他方法。MIPS 计算时间只需 0.081 min, 方法的计算时间与 EC-SVM 相当能够满足系统实时性要求。

表 1 EC-SVM、MILES 以及 MIPS-1 性能对比

算法	系统平均精度/%	计算时间/min
MIPS	64.1 ± 1.2	0.081
MIPS-1	62.3 ± 1.8	0.072
MILES	58.8 ± 1.4	0.725
EC-SVM	60.8 ± 2.0	0.063

图 3 为 MIPS、MIPS-1 和 MILES 的系统性能。实验表明对 MILES 进行示例原型删减时, 系统性能没有改变; MIPS 和 MIPS-1 的 θ 取值为 0 ~ 3.5 时系统性能稳定, 并在取值为 2 是达到最大。后续实验 θ 取为 2。

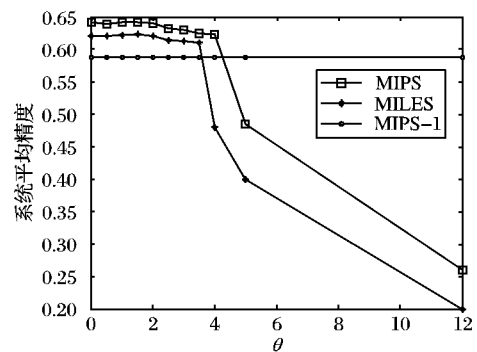


图 3 示例删减阈值 θ 对系统性能的影响

3.2 示例原型特征权值的学习

针对 2.2 节提出的 NSI(SVM 负支持向量中的所有示例作为负示例)和 NAI(负包中的所有示例为负示例)两种示例原型特征权值学习方法, 图 4 给出实验结果。实验表明, 三种权值学习方法都优于不使用权值调整的方法; NSI 和 NAI 两种负示例估计方法在开始阶段性能基本相当, 在训练集超过

25 时 MIPS(NAI)表现出更优越的性能。由于 NAI 使用所有负示例来学习负分布,NAI 比 NSI 能更准确估计负示例分布,所以 NAI 的优越特性显而易见。

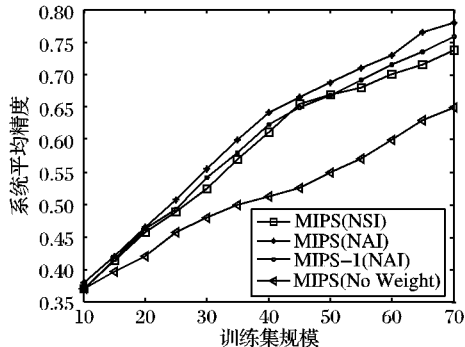


图4 三个示例原型特征权值学习方法的对比

3.3 主动学习

为验证系统的实际使用效果,组织 25 名学生对不同主动学习策略进行验证。学生参与者对每个分类随机选择 50 个图像实例进行查询,用户对反馈结果进行标注,当对结果满意时就停止,系统对每个用户每轮不同反馈数据下的系统精度、响应时间以及达到满意时反馈轮次都进行记录。

图5为系统返回 20 个图像时主动学习策略的对比实验,实验要求系统在每轮标注时提供 4 个图像和 4 个文本关键字。图中的 MIPS(Mix)为上节提出的组合策略,MIPS-1(Mix)为不考虑文本关键字的组合策略。通过一定范围的验证,组合策略中的参数 t_1, t_2, γ 和 p_i 分别选 0, 0.04, 500, 0.9 最优。对比实验表明:1) 组合策略在检索整个过程优于其他策略;2) MRS, RandB(随机策略)在不同情况下表现出不同特性,如,检索早期阶段 MRS 能有效获取信息,但随检索的进行 MRS 选择的查询可能与已经标注过的样例等价,此时 MRS 不如 RandB 策略。

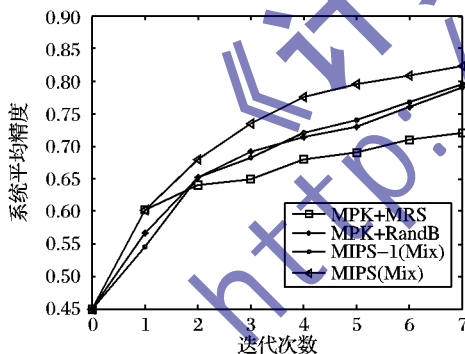


图5 主动学习策略的对比(Top-20)

本文还将 Corel5k 的所有区域示例进行人工标注,对 MIPS(Mix) 和 Liu 等^[11]提出的混合粒度标注的多示例主动学习算法进行实验验证(图6)。样例标注数设为 8,其中,MIPS(Mix) 为 4 个文本关键字 + 4 个包图像;Mix(Liu) 的标注包图像 + 示例共 8 个。虽然 MIPS(Mix) 比 Liu 的方法稍差,但本文方法更具操作性。

4 结语

本文提出结合文本信息的示例原型选择及相关反馈中的主动学习策略。该示例原型选择方法利用图像的文本信息进行正例拓展,通过估计负示例分布和示例更新进行示例筛选和降维。此外,探讨了三种主动学习策略的切换策略,提出一个新颖的反馈机制。通过实验表明本文的方法无论在性能上还是响应时间上都优于 MILES 和 EC-SVM。

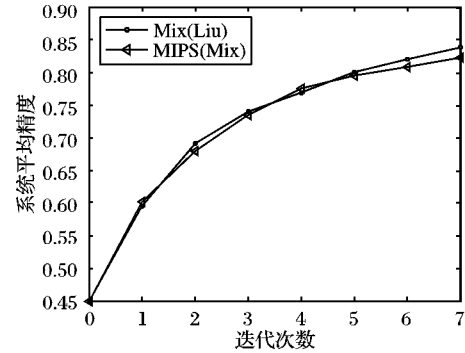


图6 主动学习策略的对比(Top-20)

参考文献:

- [1] RAHMANI R, GOLDMAN S A, ZHANG H. Localized content based image retrieval[C]// MIR'05: Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval. New York: ACM, 2005: 227-236.
- [2] ZHANG Q, GOLDMAN S A, YU W, *et al.* Content-based image retrieval using multiple-instance learning[C]// ICML'02: Proceedings of the Nineteenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 2002: 682-689.
- [3] BUNESCU R C, MOONEY R J. Multiple instance learning for sparse positive bags[C]// ICML'07: Proceedings of the 24th International Conference on Machine Learning. New York: ACM, 2007: 105-112.
- [4] YANG C, DONG M, HUA J. Region-based image annotation using asymmetrical support vector machine-based multiple instance learning[C]// CVPR'06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2006, 2: 2057-2063.
- [5] CHEN Y, WANG J Z. Image categorization by learning and reasoning with regions[J]. Journal of Machine Learning Research, 2004, 12(5): 913-939.
- [6] CHEN Y, BI J, 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] LI W-J, YEUNG D Y. Localized content-based image retrieval through evidence region identification[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE, 2009: 1666-1673.
- [8] FU Z, ROBLES-KELLY A, ZHOU J. MILIS: multiple instance learning with instance selection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 958-77.
- [9] ZHANG D, WANG F, SHI Z, *et al.* Interactive localized content based image retrieval with multiple-instance active learning[J]. Pattern Recognition, 2010, 43(2): 478-484.
- [10] SETTLES B, CRAVEN M, RAY S. Multiple instance active learning[EB/OL]. [2012-01-01]. http://www.books.nips.cc/papers/files/nips20/NIPS2007_1072.pdf.
- [11] LIU DONG, HUA XIAN-SHENG, YANG LIN-JUN, *et al.* Multiple-instance active learning for image categorization[C]// MMM'09: Proceedings of the 15th International Multimedia Modeling Conference on Advances in Multimedia Modeling. Berlin: Springer-Verlag, 2009: 239-249.
- [12] ARYA S, MOUNT D M, NETANYAHU N S, *et al.* An optimal algorithm for approximate nearest neighbor searching fixed dimensions[J]. Journal of the ACM, 1998, 45(6): 891-923.
- [13] DUDA R O, HART P E, STORK D G. Pattern Classification[M]. 2nd ed. New York: John Wiley and Sons, 2001.
- [14] TANG HANGHANG, HE JINGRUI, LI MINGJING, *et al.* Manifold-ranking-based keyword propagation for image retrieval[J]. Journal on Applied Signal Processing, 2006, 2006(1): 1-10.