

# 基于 SVM 期望间隔的多标签分类的主动学习

刘端阳 邱卫杰

(浙江工业大学计算机科学与技术学院 杭州 310023)

**摘要** 分类是数据挖掘领域研究中的核心技术之一。得到一个性能良好的分类器需要大量的训练样本,而对样本进行标记是一个十分消耗资源的过程,对多标签样本进行标记就更加困难。为了尽可能降低标记样本的成本,需要找出最能代表类别信息的样本。在基于 SVM 的分类方法中,分类器间隔越大,分类的精度就会越差。提出了一种基于期望间隔的主动学习方法,即依据当前分类器,选择最快缩小分类间隔的样本。通过实验证明,基于期望间隔的学习策略比基于决策值以及基于后验概率的策略有着更好的学习效果。

**关键词** 多标签,后验概率,期望间隔,主动学习,支持向量机

**中图法分类号** TP3 **文献标识码** A

## Active Learning for Multi-label Classification Based on SVM's Expect Margin

LIU Duan-yang QIU Wei-jie

(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

**Abstract** Classification is one of the key techniques of data mining. It requires a large number of training samples to obtain a favorable classifier, but it is resource-consuming to create label for each sample, it is even more so for multi-label samples. In order to reduce costs, it should find the most informative samples which can represent the classes. The classifiers which are based on SVM, the larger margin, the classifier's accuracy will be poorer. This paper proposed an active learning method based on SVM's expect margin which relies on current classifier, select samples that can reduce classifier's margin fastest. The experimental results show that the method based on expect margin outperforms than other active learning strategy based on decision value and posterior probability strategy.

**Keywords** Multi-label, Posterior probability, Expect margin, Active learning, SVM

## 1 前言

在传统的分类问题中,都是假定一个样本只属于一个类标签。但是由于客观事物本身的复杂性,一个样本可以同时拥有多个标签。例如,在文档分类中,每个文档可能同时隶属于多个主题,如艺术和健康;在图像分类中,每幅图像可能同时属于多种场景,如森林和山。在所有的多标签学习框架中,每个样本与一个类标集合相关联,多标签学习的任务就是要为未知样本预测其标签集,且标签集的大小是未知的。

监督的学习方法在分类领域得到了广泛的应用,但要想在这种方法下得到一个比较满意的分类模型则需要大量的训练数据。而构造训练样本集则需要耗费领域专家大量的工作。同时,训练样本过多将使得学习过程变慢,甚至变得难以接受。而主动学习能有效地克服这两个瓶颈。它根据现有的分类模型,采用某种样本选择策略,迭代地选择一些最有价值的样本进行标记,能最快地改进现有模型的分类性能。

目前,关于 SVM 主动学习的多标签分类文章还相当匮乏<sup>[1]</sup>,主要关注点还停留在单标签的主动学习上。据了解,国内最新且最有价值的关于多标签的研究就是杨碧姗的文

献<sup>[1]</sup>,其先估计样本在每个标签上的后验概率并排序,接着利用逻辑回归预测标签的数目,然后近似地确定样本标签,以此来构造损失函数并作为样本的选择策略。而国内其他文章几乎都是针对多类的研究。比如袁勋等利用每个类别的后验概率构造样本置信度,并以此作为样本选择依据<sup>[2]</sup>。宋鑫颖等通过减少非支持向量来保证训练速度<sup>[3]</sup>。国外最早使用 SVM 方法来解决多标签主动学习的是 Xuchun Li,其利用在相邻两个循环中期望损失下降的程度作为分类器改进的标志,使用了最大平均损失值和最大损失值两种计算方法<sup>[4]</sup>。Klaus Brinker 利用的选择策略是依据所有二分类支持向量的输出绝对值的最小值,以期能够最大限度地约简版本空间<sup>[5]</sup>。由于不同的分类器间输出的值不具有直接可比性,因此 Mohan Singh 等提出了一种利用后验概率来选择样本的方法<sup>[6]</sup>。

## 2 相关知识介绍

### 2.1 支持向量机简介

支持向量机(Support Vector Machine, SVM)是一种建立在统计学习理论基础上的分类技术。它是基于结构风险最小

到稿日期:2010-05-31 返修日期:2010-10-10

刘端阳(1975-),男,博士,副教授,主要研究方向为数据挖掘、分布式计算, E-mail: ldy@zjut.edu.cn;邱卫杰(1986-),男,硕士,主要研究方向为数据挖掘、人工智能。

化原则,根据有限样本信息在模型的复杂度和学习能力之间寻求最佳的折中。由于出色的泛化性能,它成为目前解决分类问题的主要工具。

在实际的分类中,多数情况下不能满足线性可分,而SVM能很好地解决这个问题。其基本原理是将在原始空间中线性不可分的输入向量  $X$  通过核函数映射到一个特征空间  $F$ ,并在这个空间中构造最优分类超平面,使其线性可分,该超平面使类别间的分类间隔最大。由于在特征空间中只需要进行内积运算,使得SVM对高维数据有很好的处理能力。在线性可分的情况下,SVM可通过计算最小化  $\|W\|^2/2$  来构造最优超平面。最优超平面为训练集中向量的线性组合:

$$w = \sum_{i=1}^l y_i a_i X_i \quad (1)$$

式中,当  $a_i=0$  时所对应的样本对构造分类器没有贡献;当  $a_i > 0$  时样本对分类器的构造起作用,也即是支持向量。由此,分类函数表示为:

$$f(X) = \text{sgn}[\sum_{i=1}^l a_i y_i K(X_i, X) + b] \quad (2)$$

式(2)用于对预测样本的分类,  $X$  为待分类样本。

## 2.2 多标签分类

多标签分类的任务就是自动地把样本划到一个或多个预先定义好的类别中。给定训练样本集  $X_1, \dots, X_l$  和  $n$  个类别  $(1, \dots, n)$ 。用一个标签集合  $Y_i = [y_1^i, \dots, y_n^i]$ ,  $y_j^i \in \{1, -1\}$  来表示一个样本属于哪些类别,当  $y_j^i = 1$  时说明其属于  $j$  类别。多标签分类可以由这个决策函数表示  $f: X \rightarrow Y^{[1]}$ 。目前,多标签分类都是通过转化成多类问题来计算求解,而SVM作为二分类方法,需要对此进行进一步转化。对于多分类问题,一般采用 one-versus-one 或 one-versus-all 方法。本文中采用 one-versus-one 方法,因为它有着良好的计算效率,并且有着非常优秀的分类效果<sup>[7]</sup>。此种方法需要构造  $n(n-1)/2$  个分类器。定义  $f^{ij}$  为对应  $i$  类和  $j$  类的分类函数,当  $f^{ij}(X) > 0$  时,就认为其属于  $i$  类,给予投票,否则就属于  $j$  类,最后根据竞争机制确定预测样本属于哪个类别。

## 2.3 后验概率

根据分类器输出的结果来确定样本的后验概率具有很大的价值,本文中的主动学习方法就是依赖于样本的后验概率估计。为了得到这些估计值,采用了经过改进后的概率估计方法。这个方法的基本思想就是用一个 sigmoid 函数把SVM的输出结果转化成后验概率的值<sup>[8]</sup>。假定  $f(X)$  是其中一个决策函数,则它的后验概率可以由下述公式近似得到:

$$P(y=1|X) = \frac{1}{1 + \exp(Af(X) + B)} \quad (3)$$

式中,  $A, B$  为估计值。为了得到合理的  $A, B$  值,通过以下的方法计算求得:

$$\min_{(A,B)} - \sum_{i=1}^l (t_i \log(P_i) + (1-t_i) \log(1-P_i)) \quad (4)$$

$$P_i = \frac{1}{1 + \exp(Af(X) + B)} \quad (5)$$

$$t_i = \begin{cases} \frac{N_p + 1}{N_p + 2}, & \text{if } y_i = 1 \\ \frac{1}{N_n + 2}, & \text{if } y_i = -1 \end{cases} \quad (6)$$

式中,  $N_p, N_n$  是该分类器在训练过程中分别属于正类和负类的个数。

## 3 主动学习

### 3.1 基于水池的主动学习

在基于水池的主动学习中,水池中的样本均没有标记过,且假定样本是独立的,依据某个潜在的分布函数  $F(x)$ 。假定存在一个多标签样本水池 Pool,从中选择部分样本并进行标记得  $L^0$  和剩下的未标记的样本集  $U^0$  ( $L^0 \ll U^0$ )。主动学习的主要目的就是 from  $U^0$  中挑选出最富含分类信息的样本,把这些数据交于领域专家进行标记,然后加入训练集,重新训练得到新的分类器,以期能最快地改进当前分类器的分类性能。这个过程可以有如下描述:在初始状态时,由  $L^0$  训练得到多标签分类器集  $f^0$ ,根据某个样本选择策略选择最有价值的样本。被选择的样本经过标记后加入到  $L^0$  后得到一个新的标记样本集  $L^1$ ,同时从未标记样本集  $U^0$  中除去被选择的样本并得到新的未标记样本集  $U^1$ ,然后循环迭代,直到满足某个特定的停止条件<sup>[4]</sup>。

### 3.2 期望分类间隔

根据 Simon Tong 的理论<sup>[9]</sup>,由训练样本集  $L$  和核函数  $K$ ,可以得到在特征空间中有一系列的分割平面,称这一系列一致性假设为版本空间。版本空间中的假设对于任何样本都能正确分类,用公式表示为:

$$v = \{w \in w \mid |w| = 1, y_i (w \cdot \Phi(X_i)) > 0, i=1, \dots, n\} \quad (7)$$

式中,  $w$  定义为参数空间。在版本空间中,SVM得到的超平面能最大化特征空间中的分类间隔,也即

$$\max_{w \in F} \min_i \{y_i (w \cdot \Phi(X_i))\} \quad (8)$$

定义  $Area_L$  为在当前标记样本集下,版本空间  $v$  占超球  $\|W\|=1$  所得到的表面积。为了最快地降低版本空间的大小,Simon Tong 每次尽量选择那些能平分版本空间的样本,也即:

$$\min_{X \in U} |Area_{L+(X,+)} - Area_{L+(X,-)}| \quad (9)$$

但上述方法具有较慢的收敛速度,于是 Jingrui He 提出了平均版本空间的方法,使每次选择的样本能最小化版本空间的大小,并且成功地运用到了多类领域<sup>[10]</sup>。对此,可以将此思想应用到多标签的分类中。计算版本空间的大小不切实际,所以版本空间的大小用分类间隔来近似。但还是有两个明显的瓶颈:1)虽然经过了近似,但计算量仍旧很大。2)只能使用 RBF 核函数,或者要求所有样本满足  $\|\Phi(X_i)\| = \lambda$ 。对于多标签的情况,仅仅计算量就使上述方法变得难以接受。因为上述方法要对每个未知样本分别标记为正类和负类,然后加入到训练集中训练后才能得到近似的版本空间大小。若使用 one-versus-all 的分解方法,需要构建  $n$  个分类器,而如果用 one-versus-one 的方法,将需要多达  $n(n-1)/2$  个,且需要进行两次二次寻优才能得到平均版本空间,这个计算量是难以接受的,所以它寻找一种近似的方法。

在对多标签问题的分析中,Brinker<sup>[5]</sup>提出了一种新的约简版本空间的方法,亦即将约简后的版本空间大小与原本版本空间大小的比值作为选择样本的标准:

$$\min_{X \in U} \frac{Area_{L+(X,v)}}{Area_L} \approx \frac{1 + y^j f_j^i(X)}{2} \quad (10)$$

随后 Brinker 用  $\min_{X \in U} |f(X)|$  作为选择策略,每次选择最靠近超平面的样本。纯粹地利用  $\min_{X \in U} |f(X)|$  作为选择策略不是

一个有效的方法,因为分类器不确定所选择的样本属于正类还是负类,而且处在超平面的不同侧对提升分类器性能有着不同的影响。由 SVM 理论可知,它是在特征空间中最大化分类间隔。当前分类器的确定依据已标记的样本集。在主动学习中,初始状态下的标记样本数量十分有限,不能包含整个类别的信息,所以分类间隔被扩大了。因而主动学习的目的就是依据当前样本集,选择那些能最快地缩小当前分类器间隔的样本。

基于以上分析,本文提出了基于期望间隔的选择策略。根据已标记样本集,定义当前分类器的间隔为  $Margin$ 。对于水池中的每一个样本,假设所选择样本为正类,则新分类器的间隔可以估计为:

$$Margin^+ \approx \frac{Margin}{2} + f(X) \quad (11)$$

反之当样本为负类时,期望间隔为:

$$Margin^- \approx \frac{Margin}{2} - f(X) \quad (12)$$

根据图 1,可以清楚地看到以上分析过程,当选择样本为正类的时候,超平面可由  $S1$  近似成  $S2$ ,于是间隔变成了式(11)所求解的值。反之亦然。

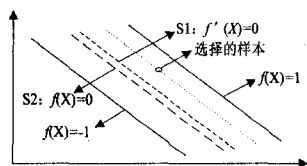


图 1

对于软间隔的问题,我们定义:

$$Margin^+ \approx \frac{Margin}{2} - \xi + f(X) \quad (13)$$

$$Margin^- \approx \frac{Margin}{2} - \xi - f(X) \quad (14)$$

式中, $\xi$ 为松弛变量。但是必须指出,上述方法是基于一种假设,就是所得到的新的分割超平面和原超平面的变化不是很大,也即与原超平面的夹角较小。而这种假设被认为是可取的,因为最初分类器是由能包含类别信息的样本训练而成,

有一定的分类效果。期望分类间隔就可以通过以下公式得到:

$$E_M = Margin^+ * P(y=1|X) + Margin^- * P(y=-1|X) \quad (15)$$

基于以上分析,本文提出了一种选择策略,最小期望间隔为:

$$\min_{i=1, \dots, l} \left( \min_{j=1, \dots, m(n-1)/2} E_M \right) \quad (16)$$

#### 4 实验结果及分析

SVM 训练程序是来自台湾大学林智仁教授等开发的 libsvm<sup>[11]</sup>。实验数据同样来自林智仁教授处理的多标签基准数据(scene 数据集)。使用基准数据集是因为它是机器学习领域权威的数据集,能够被研究者方便地共享,使本文的实验结果可以得到广泛的比较。该数据集共有 6 个标签,294 个特征向量,1211 个训练样本,1196 个测试样本。普遍使用的查全率和查准率等方法虽然能衡量分类效果,但是这些评价方法都不是单一数值,通过它们无法直接地看出分类效果随新样本的加入而变化的情况,所以本文采用正确分类的样本数量来衡量分类效果<sup>[2]</sup>。

本次实验的 SVM 训练参数均是 libsvm 提供的默认参数。选择了两种常用的核函数,分别为 Radial basis 和 Sigmoid。本文将对比 3 种主动学习的方法:

Min\_Valu: 最小决策值;

Min\_Prob: 最小概率;

Min\_E<sub>M</sub>: 最小期望间隔。

为了分析本文提出的方法,实验用基于 Radial basis 核函数的分类器来仔细对比本文提出的主动学习方法,同时用 Sigmoid 函数来验证对其他核函数的适用性。实验选择相同的初始训练样本集,所以不同条件下实验的第一次分类精度都是相同的,在实验结果中不予以表示。在选择额定的样本数目(200 个)后,主动学习即停止。本文共进行了 4 次实验,实验结果见表 1,表中 S 为每次迭代选择的样本数。

表 1 实验结果

选择的 样本总数	Radial basis, S=2			Radial basis, S=20			Sigmoid, S=5			Sigmoid, S=20		
	Min_Valu	Min_Prob	Min_E <sub>M</sub>	Min_Valu	Min_Prob	Min_E <sub>M</sub>	Min_Valu	Min_Prob	Min_E <sub>M</sub>	Min_Valu	Min_Prob	Min_E <sub>M</sub>
ΣS=20	199	362	215	195	181	246	199	168	338	195	173	236
ΣS=40	199	338	269	197	274	354	199	226	283	195	149	242
ΣS=60	199	320	463	323	454	439	195	231	299	195	195	268
ΣS=80	199	390	366	330	410	563	196	276	273	195	341	261
ΣS=100	270	440	402	348	418	514	208	267	371	195	249	328
ΣS=120	341	458	507	363	404	449	209	211	375	195	386	246
ΣS=140	389	454	502	383	428	536	203	212	362	220	248	310
ΣS=160	413	429	575	398	473	548	207	242	381	221	231	422
ΣS=180	423	534	661	374	491	528	236	353	417	237	288	452
ΣS=200	428	571	588	403	477	531	278	301	418	295	437	528

从表 1 可知,本文提出的算法的预测正确率在经过前几次的迭代后一直都是优于基于决策值和后验概率的主动学习方法,而与每次迭代中选择的样本数量值无关。同时发现基于后验概率的方法优于基于决策值的方法。从表中的后两组实验可以发现,基于期望间隔的方法能适用于其他核函数,消除了平均版本空间中对核函数或者样本的限制。实验结果证明本文提出的方法能更好地评价未标记样本,并从中选择最

富含分类信息的样本,有效地提高分类器的性能。

**结束语** 本文试图解决分类中多标签的主动学习问题。分类的目标就是最快地减小分类器的间隔。我们提出了一种新的多标签分类的主动学习算法,通过计算分类器的期望间隔作为样本选择的标准。一系列实验证明,本文提出的方法优于基于决策值和后验概率的选择策略,能有效地提高分类

(下转第 266 页)

$(x \rightarrow y) < \tilde{i}$  矛盾,故  $\max\{\tilde{\mu}(x \rightarrow y), [0.5, 0.5]\} \geq \min\{\tilde{\mu}(x), \tilde{\mu}(y)\}$ 。

若  $\max\{\tilde{\mu}(O), [0.5, 0.5]\} < \tilde{\mu}(I)$ , 则存在  $\tilde{i} \in D(0.5, 1]$ , 使得  $\max\{\tilde{\mu}(O), [0.5, 0.5]\} < \tilde{i} < \tilde{\mu}(I)$ , 因此  $\tilde{\mu}(O) < \tilde{i}$ , 即  $O \notin U(\tilde{\mu}; \tilde{i})$ , 这与  $U(\tilde{\mu}; \tilde{i})$  是格蕴涵子代数矛盾。因此  $\max\{\tilde{\mu}(O), [0.5, 0.5]\} \geq \tilde{\mu}(I)$ 。

反之, 设  $\forall x, y \in L, x, y \in U(\tilde{\mu}; \tilde{i})$ , 则  $\tilde{\mu}(x) \geq \tilde{i}, \tilde{\mu}(y) \geq \tilde{i}$ 。因为

$$\max\{\tilde{\mu}(x \rightarrow y), [0.5, 0.5]\} \geq \min\{\tilde{\mu}(x), \tilde{\mu}(y)\} \geq \tilde{i}$$

且  $\tilde{i} \in D(0.5, 1]$ , 故必有  $\tilde{\mu}(x \rightarrow y) \geq \tilde{i}$ , 即  $x \rightarrow y \in U(\tilde{\mu}; \tilde{i})$ 。

同理可证  $O \in U(\tilde{\mu}; \tilde{i})$ 。因此  $U(\tilde{\mu}; \tilde{i}) (\neq \Phi), \tilde{i} \in D(0.5, 1]$  是  $\mathcal{L}$  的格蕴涵子代数。

在例 1 中我们给出了一个区间值  $(\in, \in \vee q)$ -模糊格蕴涵子代数, 而不是区间值  $(\in, \in)$ -模糊格蕴涵子代数。下面定理给出区间值  $(\in, \in \vee q)$ -模糊格蕴涵子代数构成区间值  $(\in, \in)$ -模糊格蕴涵子代数的条件。

**定理 9**  $\mathcal{L}$  的区间值模糊集  $\tilde{\mu}$  是区间值  $(\in, \in \vee q)$ -模糊格蕴涵子代数且对于  $\forall x \in L$ , 有  $\tilde{\mu}(x) < [0.5, 0.5]$ , 则  $\tilde{\mu}$  是区间值  $(\in, \in)$ -模糊格蕴涵子代数。

证明: 设  $I_s \in \tilde{\mu}$ , 则  $\tilde{\mu}(I) \geq \tilde{s}$ 。由于  $\tilde{\mu}$  是区间值  $(\in, \in \vee q)$ -模糊格蕴涵子代数且对于  $\forall x \in L$ , 有  $\tilde{\mu}(x) < [0.5, 0.5]$ , 所以  $\tilde{\mu}(O) \geq \min\{\tilde{\mu}(I), [0.5, 0.5]\} = \tilde{\mu}(I) \geq \tilde{s}$ 。即  $O_s \in \tilde{\mu}$ 。

设  $x, y \in \tilde{\mu}, \tilde{s}, \tilde{i} \in D(0, 1]$ , 则  $\tilde{\mu}(x) \geq \tilde{s}, \tilde{\mu}(y) \geq \tilde{i}$ 。根据区间值  $(\in, \in \vee q)$ -模糊格蕴涵子代数, 所以  $\tilde{\mu}(x \rightarrow y) \geq \min\{\tilde{\mu}(x), \tilde{\mu}(y), [0.5, 0.5]\} \geq \min\{\tilde{s}, \tilde{i}\}$ 。故有  $(x \rightarrow y)_{\min\{\tilde{s}, \tilde{i}\}} \in \tilde{\mu}$ 。

因此  $\tilde{\mu}$  是区间值  $(\in, \in)$ -模糊格蕴涵子代数。

**结束语** 为了给具有模糊性和不可比较性的不确定性信息处理提供可靠的逻辑基础, 以及建立真值域在格上的逻辑系统, 徐扬建立了格蕴涵代数, 随后建立了模糊格蕴涵代数。

(上接第 232 页)

精度和速度, 可以明显地降低标记样本所付出的费用。

## 参考文献

- [1] Yang Bisan, Jiao-Tao, Wang Teng-jiao, et al. Effective Multi-Label Active Learning for Text Classification[C]//KDD'09: Proceedings of the 15<sup>th</sup> ACM SIGKDD international conference on Knowledge discovery and data mining. Paris, 2009: 917-926
- [2] 袁勋, 吴秀清, 洪日昌, 等. 基于主动学习 SVM 分类器的视频分类[J]. 中国科学技术大学学报, 2009, 39(5): 473-478
- [3] 宋鑫颖, 周志远. 一种基于 SVM 的主动学习文本分类方法[J]. 计算机科学, 2006, 33(11): 288-290
- [4] Li Xu-chun, Wang Lei, Sung E. Multi-Label SVM Active Learning for Image Classification[C]//International Conference on Image Processing. Lion, 2004: 2207-2210
- [5] Brinker K. On Active Learning in Multi-label Classification[M]//Myra Spiliopoulou, Rudolf Kruse, Christian Borgelt, et al. "From Data and Information Analysis to Knowledge Engineering" of Book Series "Studies in Classification, Data Analysis, and Knowledge Organization". Berlin Heidelberg: Springer, 2006: 206-213
- [6] Singh M, Curran E, Cunningham P. Active Learning for Multi-

为了推动这一重要逻辑代数的发展, 我们必须弄清楚其结构。而代数的子结构是一重要的研究内容。本文提出了广义形式的模糊蕴涵子代数——区间值  $(\alpha, \beta)$ -模糊格蕴涵子代数, 研究了其性质, 得到了它的等价刻画。

## 参考文献

- [1] 徐扬. 格蕴涵代数[J]. 西南交通大学学报, 1993, 28(1): 20-27
- [2] Xu Y, Ruan D, Qin K Y, et al. Lattice-valued logic-an alternative approach to treat fuzziness and incomparability[M]. Berlin: Springer-Verlag, 2003
- [3] 徐扬, 秦克云. 模糊格蕴涵代数[J]. 西南交通大学学报, 1995, 30(2): 121-127
- [4] 徐扬, 秦克云. 格蕴 H 涵代数和格蕴涵代数类[J]. 河北煤炭工程学院学报, 1992(3): 139-143
- [5] 赖家俊, 徐扬. 赋范格 H 蕴涵代数和模糊格 H 蕴涵代[J]. 计算机科学, 2008, 35(11): 126-159
- [6] 赖家俊, 徐扬. 基于语言真值格值命题逻辑系统  $L_{\text{vpl}}$  的推理规则[J]. 计算机科学, 2008, 35(9): 230-232
- [7] 杨丽, 徐扬. 基于矩阵蕴涵的格值模糊概念格构造方法[J]. 计算机科学, 2009, 36(8): 264-267
- [8] 袁学海, 李红兴, 孙凯彪. 直觉模糊集和区间值模糊集的截集、分解定理和表示定理[J]. 中国科学(F 辑), 2009, 39(9): 933-945
- [9] Pu P M, Liu Y M. Fuzzy topology I, Neighborhood structure of a fuzzy point and Moore-Smith convergence[J]. J. Math. Anal. Appl., 1980(76): 571-599
- [10] Ma X L, Zhan J M, Shum K P. Interval valued  $(\in, \in \vee q)$ -fuzzy filters of MTL-algebras[J]. J. Math. Research. Exposition, 2010, 30(2): 121-127
- [11] Dudek W A, Shabir M, Ali M I.  $(\alpha, \beta)$ -fuzzy ideals of hemirings[J]. Comp. Math. Appl., 2009(58): 310-321

label Image Annotation[C]//The 19th Irish Conference on Artificial Intelligence and Cognitive Science. Cork, Ireland, 2008: 173-182

- [7] Joshi A J, Porikli F, Papanikolopoulos N. Multi-Class Active Learning for Image Classification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, 2009: 2372-2379
- [8] Lin Hsuan-tien, Lin Chih-jen, Weng R C. A Note on Platt's Probabilistic Outputs for Support Vector Machines[J]. Journal of Machine Learning Research, 2007, 68(3): 267-276
- [9] Tong S, Koller D. Support Vector Machine Active Learning with Applications to Text Classification[J]. Journal of Machine Learning Research, 2001: 45-66
- [10] He Jing-rui, Li Ming-jing, Zhang Hong-jiang, et al. Mean Version Space: a New Active Learning Method for Content-Based Image Retrieval[C]//Proceeding of the ACM SIGMM International Workshop on Multimedia Information Retrieval(MIR) at the International Multimedia Conference. 2004: 15-22
- [11] Chang Chih-chung, Lin Chi-jen. LIBSVM: a Library for support vector machine[OL]. <http://www.csie.ntu.tw/~cjlin/libsvm>, 2010-01-10