doi:10.3969/j.issn.1673-4785.2011.03.011

一种基于主动学习支持向量机哈萨克文文本分类方法

古丽娜孜[1,2],孙铁利[2],伊力亚尔[1],吴迪[2]
(1.伊犁师范学院 电子与信息工程学院,新疆 伊宁 835000; 2.东北师范大学 计算机科学与信息技术学院,吉林 长春 130117)

摘 要:将文本分类理论应用于哈萨克语中,给出基于支持向量机的哈萨克文文本分类系统的设计思想.从哈萨克语言学的角度对哈萨克文分析,提出哈萨克文词干提取的方法.在对支持向量机的理论分析基础上,提出主动学习算法对支持向量机进行训练,使用训练后的分类器对新的文本进行分类.实验结果表明,该方法在哈萨克文文本分类中能获得可接受的分类性能.

关键词:支持向量机;哈萨克文文本分类;主动学习
中图分类号:TP391.1 **文献标识码**:A **文章编号**:1673-4785(2011)03-0261-07

An approach to the text categorization of the Kazakh language based on an active learning support vector machine

GU Linazi[1,2], SUN Tieli[2], YI Liyaer[1], WU Di[2]
(1. School of Electronic and Information Engineering, Yili Normal University, Yining 835000, China; 2. School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China)

Abstract: In applying the theory of text categorization to the study to the Kazakh language, an approach to text categorization of Kazakh text based on a support vector machine system was introduced. In this paper, from the Kazakh linguistic angle, the method to extract word stems was analyzed. Based on analysis of the support vector machine, the proposed active learning algorithm was adopted for training. The trained classifier was used to classify new text. The experimental results show that this approach to Kazakh text classification has an acceptable classification performance.

Keywords: support vector machine; Kazakh text categorization; active learning

文本自动分类任务是对未知类别的文本文档进行自动处理,判别它们所属预定义类别集中的一个或多个类别.迄今为止,文本分类在那些被广泛使用的语言中得到了较好的研究和应用[1-2],但在哈萨克语中没有得到很好的发展.这是因为哈萨克语的文本分类技术研究起步比较晚,而且哈萨克语单词的自动切分处理有一定的难度.由于哈萨克语的语法体系与其他语言语法体系之间存在很大的差别,因此不能直接采用其他语言文本处理方面的研究成果,需要研究出适合哈萨克文自己的文本分类体系.实现分类任务需要一定的语料库,但目前为止在哈萨克语言中还没有一个公认的哈萨克文语料库,为此需考虑语料集的规范,笔者对公认的中文语料库中的部分文本进行翻译构建了本文分类器的语料库,同时选择最适合这种小样本分类任务的支持向量机分类方法来实现分类.

收稿日期:2010-10-08.
基金项目:教育部科技发展中心网络时代的科技论文快速共享研究项目(20090043110010);吉林省科技规划资助项目(20090503);吉林省教育厅"十一五"科研规划资助项目(2009587).
通信作者:孙铁利. E-mail:suntl@nenu.edu.cn.

1 文本特征提取

实现文本分类任务大致需要经过以下几个步骤:1)文本预处理,通过特征提取将文本转换成自动处理的文本模型;2)根据问题性质和研究结果建立分类器模型,并以相应数据集来训练分类器;3)测试新的文本样本.本文将就上述关键技术进行讨论,并给出研究方法和实验结果.

1.1 文本预处理

文本预处理主要是从文本中提取关键词来表示文本的处理过程. 在预处理过程中,要对文本进行分词处理,将连续的语句分隔为分散的有独立意义的词集,然后去除集合中的停用词,获得文本的关键词集合. 文本预处理方法是影响文本分类准确度的关键因素之一,对于不同语言,需要采用不同的预处理技术得到相关文本的词性信息. 例如,中文需要进行分词,英文则需要进行词干提取.

哈萨克文为拼音文字,属于阿尔泰语系突厥语族的克普恰克语支,中国境内通用的哈萨克文借用了阿拉伯语和部分波斯文字母. 哈萨克文共有33个字母,其中有24个辅音字母、9个元音字母,每个字母的位置有词首、词中、词末、独立4种变体,词与词之间有空格分开,所以哈萨克文不需要分词处理. 作为黏着语,哈萨克语的语法形式是通过在单词原形的后面或前面附加一定的附加成分来完成的. 这就造成在哈萨克语文本中,一个哈萨克语词对应多个字符串形式,因此一定要对其进行词干提取.

哈萨克文预处理工作中,除了词干提取以外,还要进行构形附加成分的切分. 这是由于构形附加成分与词干互相黏连,构形附加成分之间也互相黏连,而且构形附加成分往往可以表示一定意义,如果不将这些黏连在一起的构形附加成分切分开,就不能准确地表达整个单词的含义. 因此,对于哈萨克语词干提取方法的研究和应用其构形语素的分析需要并行处理. 本文完成了哈萨克文词干提取以及词性标注所需的哈萨克语词干表的构建. 该词干表收录了由新疆人民出版社出版的《哈萨克语详解词典》中的6万多个哈萨克语词干,形式如图1所示. 图2是附加成分表的一部分,其收录了438个哈萨克语附加成分. 整个附加成分可分为构形附加成分和构词附加成分两大类. 其中,构形附加成分为4种,即复数附加成分、领属性人称附加成分、格附加成分和谓语性人称. 构词附加成分同样分为4种,包括动词、形容词、数词和副词附加成分. 附加成分的详细分类有助于在附加成分切分阶段进行词法分析.

id	word	pos
1	شۇرىشتىڭكەرە	v
2	شۇلعىلىك	adj
3	شۇي	n
4	شۇيىر	v
5	شۇيىرىن	vc
6	شۇيىرلىۋ	va

图1 哈萨克语词干

Fig. 1 Kazakh text stem

名词、动词和形容词是哈萨克语中数量最多、变换最复杂的词类,是词切分中的难点. 本文在上述各类前期准备工作的基础上,给出了这3种词性的有限状态自动机,然后采用双向全切分和词法分析相结合的改进方法实现哈萨克语词干提取和构形附加成分的细切分. 对于词干表搜索,首次采用了改进的逐字母二分词典查询机制来提高词干提取的效率,对歧义词和未登陆词的切分采用概率统计的方法.

index	type	suffix	btype
215	adj	دىك	gc
201	adj	ۇلى	gc
228	adj	سىز	gc
227	adj	پەرەز	gc
226	adj	پا	gc

图2 哈萨克语附加成分

Fig. 2 Additional components in Kazakh text

在以上研究基础上,设计实现了哈萨克语自动词法分析中的附加成分的切分和词干提取程序,完成了哈萨克文文本的读取预处理. 处理结果如图3所示,上半窗体显示内容是待切分的原文,下半窗体显示内容是切分后的结果.

图3 哈萨克文词干切分结果示例

Fig. 3 Example of segmentation results of the Kazakh text stem

1.2 文本模型

文本属于一种非结构化的数据,无法被学习算法直接用于训练或分类. 通过简单而准确的方法,将文本表示成机器可处理的形式,是进行文本分类的基础. 向量空间模型(vector space model,VSM)是最经典的文本形式化表示方法,它将文本表示为包含特征项和特征项权重的向量. 在VSM中,用 d(document)表示文本,特征项是指出现在文本 d 中且可代表该文本内容的基本语言单位,用 t(term)表示,特征项权重是指词对文本的重要程度,用 w(weight)表示. VSM将文本映射为一个特征向量 $V(d)=((t_1,w_1),(t_2,w_2),\cdots,(t_i,w_i),\cdots,(t_n,w_n))$,其中 $t_i(i=1,2,\cdots,n)$ 为一些互不相同的特征

词,w_i 表示特征项 t_i 在文本中的权重,其计算公式[3]为

$$w_i = \frac{f(t,d)\log(n/n_t + 0.01)}{\sqrt{\sum_{t=d}[f(t,d)\log(n/n_t + 0.01)]^2}}.$$

式中:$f(t,d)$ 为特征项 t 在文本 d 中的词频;n 为训练文本总数;n_t 为训练文本集中包含特征项 t 的文本数.

1.3　特征处理

对训练样本集进行预处理所得到的关键词的集合构成了初始特征项(词)集合,简称特征集. 通常该集合中特征项数目过多是制约分类的重要因素,即使是一个小规模的样本集,经过预处理也会得到一定数量的特征词,其中有些特征词对文本内容和类别的贡献很小,及时消除这些特征词会有效地控制向量空间的维数. 因此,必须通过降维处理去除弱关联词,抽取强关联词构成用于学习的特征集.

特征就是区分类别的尺度,不同的模式分类问题有不同的特征选择方法,在文本分类中所用到的方法有文档频率(DF)、信息增益(IG)、互信息(MI)、X^2 统计量(CHI)、卡方统计量、期望交叉熵、文本证据权以及几率比等.

上述评判函数是目前用的比较多的特征抽取评估函数,它们各有各的优缺点,IG、MI、CHI 侧重于低频词,而 DF 侧重于高频词. 目前,研究者分别对这些方法做了不同的优化改进,达到了各自的理想效果,其中文本频率比值法 DFR(document frequency ratio)[4]以简单、快捷等优点克服了以上几种方法目前所存在的缺点,综合考虑了类内外文本频率,其计算公式为

$$f_{\mathrm{DFR}}(t,C_i) = \frac{(N - n_i) \times N(t,C_i)}{n_i \times N'(t,C_i)}.$$

式中:N 为训练集中的总文本数;n_i 是 C_i 类中的文本数;$N(t,C_i)$ 表示类别 C_i 中包含词 t 的文本数;而 $N'(t,C_i)$ 表示除去 C_i 以外的其他类别中包含词 t 的文本数.

2　基于主动学习支持向量机的文本分类

2.1　支持向量机

支持向量机(support vector machines, SVM)是由 Vapnik[5]提出的一种基于结构风险最小化原理的机器学习方法[6-7]. 在最简单的情形中,线性 SVM 通过学习得到一个超平面,该超平面以最大分类间隔将正样本集合与负样本集合分离开,此处的间隔(margin)是指超平面与距离它最近的正样本和负样本之间的距离.

SVM 解决分类问题时根据数据集的来源特征将分类问题分为线性可分状态和线性不可分状态. 针对训练样本集(1),线性可分问题可以用式(2)所示的 SVM 数学模型来解决.

$$S = \{(\boldsymbol{x}_i, y_i), i = 1,2,\cdots,r\},\quad \boldsymbol{x}_i \in \mathbf{R}^n, y_i \in \{+1, -1\}; \tag{1}$$

$$\min \phi(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2,$$

$$\text{s.t. } y_i(\boldsymbol{w}\boldsymbol{x}_i + b) - 1 \geqslant 0,\ i = 1,2,\cdots,l. \tag{2}$$

n 维空间中分界面为 $\boldsymbol{w}\boldsymbol{x} + b = 0$,能使到该分界面最近的 2 类样本之间的距离 $\frac{2}{\|\boldsymbol{w}\|}$ 最大,也就是 $\|\boldsymbol{w}\|$ 最小,该分类界面就称为最优分类界面. 从而最终可得到所求的最优分类函数为

$$f(\boldsymbol{x}) = \mathrm{sign}\left(\sum_{i=1}^{l} a_i y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b\right).$$

式中:对应 a_i 不为 0 的样本就是支持向量. 最优化问题的解 a_i 的每一个分量都与一个训练点相对应,显然以上算法所构造的划分超平面,仅仅依赖于那些相应于 a_i 不为零的训练点$(\boldsymbol{x}_i \cdot \boldsymbol{x})$,而与相应于 a_i 为零的那些训练点无关. 相应于 a_i 不为零的训练点$(\boldsymbol{x}_i \cdot \boldsymbol{x})$中的输入 $\boldsymbol{x}_i$ 为支持向量. 显然,只有支持向量对最终求得的划分超平面的法方向 $\boldsymbol{w}$ 有影响,而它与非支持向量无关,这种方法就是支持向量机.

对于非线性问题,支持向量机通过选择适当的非线性变换,将输入空间中的训练样本集映射到某个高维特征空间中,使得在目标高维空间中这些样本集线性可分,然后再构造一个最优分界面来逼近理想分类结果[8]. 为此,需要在式(2)中增加一个松弛变量 ξ_i 和惩罚因子 C,从而式(2)变为:

$$\min \phi(\boldsymbol{w},\xi) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{l}\xi_i;$$

$$\text{s.t. } y_i(\boldsymbol{w}\boldsymbol{x}_i + b) - 1 + \xi_i \geqslant 0,$$

$$\xi_i \geqslant 0, i = 1,2,\cdots,l.$$

式中:C 为某个指定的常数,控制对错分样本惩罚的程度,C 值越大,惩罚越重.

SVM 采用不同的核函数 $K(\boldsymbol{x},\boldsymbol{y})$,可以实现输入空间中的不同类型的非线性分类问题转化为线性分类情况,进而产生不同的支持向量算法,引进核函数以后的最优分类函数为

$$f(\boldsymbol{x}) = \mathrm{sign}\left(\sum_{i=1}^{l} a_i y_i K(\boldsymbol{x} \cdot \boldsymbol{y}) + b\right).$$

2.2　主动学习算法

根据对训练样本处理方式的不同,可将学习方法分为主动学习和被动学习 2 类. 主动学习在学习过程中可以根据学习机推进的情况,选择最有利于

分类器性能的样本来进一步训练分类器,这样它能有效地减少评价样本数量;而被动学习只是随机地选择训练样本,被动地接受这些样本的信息进行学习.引入主动学习目的主要是从减少评价样本所需的代价,最大的优点是通过仔细、合理地选择训练样本后,需要的实际训练样本数量将大大减少,评价样本所需的代价也就随之减少.

针对哈萨克文文本的预处理的复杂性和SVM方法只与支持向量有关这2个因素,对SVM算法进行了改进,用主动学习方法[9-10]处理SVM分类器的训练文本.为了更好地满足分类要求,文本分类模型采用多分类模式[11].

主动学习从形式上是一个循环反复的过程,应用SVM方法实现主动学习,采用何种算法有效地筛选训练样本,以便快速进入训练阶段是研究的关键.主动学习首先根据先验知识或者随机地从未带类别标注的所有候选样本集中选择少量样本并标注它们的类别,构造初始训练样本集,确保初始训练样本集中至少包含有一个正例样本和一个负例样本.利用初始训练样本集中这些带类别标注的样本训练一个分类器,在该分类器下,采用某种采样算法,从候选样本集中选择最有利分类器性能的样本,标注类别并加入到训练样本集中,重新训练分类器,再次选择最有利分类器性能的样本.重复以上过程,直到候选样本集为空或达到某种指标[12].

文献[13]提出一种新的多类学习模型,即决策有向无环图(directed acyclic graph,DAG).每条边都有方向、且不存在任何回路的图称为有向无环图,图中惟一没有入度的节点则是DAG的根.在分类任务中,可以引入此种数据结构构造SVM分类模型,即有向无环图SVM(DAGSVM).对于DAGSVM,输入一个样本,从根节点开始判决,一直访问到叶子结点就是要得到的结果类别,这样对于N类的问题,要进行$N-1$次判别. DAGSVM最大的优势在于能够准确定位结果类别,具有准确性较高的特点[14].如图4是具有4个类别的DAG.

有向无环图SVM算法的中心思想如下.

输入:未知类别信息文本.

输出:最优的结果类别.

算法:

首先对未知文本采用主动学习SVM进行类别的分类;

C_i排成队列,$i=1,2,\cdots,n$; //将所有的类别按任意顺序排成队列

取首个和最后一个类别对应的分类器对未知样本进行分类,根据分类结果,删除其中表现差的一个类别,由剩下的类别形成新的队列;

While(队列中所剩下的类别非单一类别)

{

取新队列的首个和最后一个类别对未知样本进行分类,保留较优的结果类别,删除次优的结果类别;

}

队列中所剩下的最后一个类别就是得到的最优类别.

算法结束.

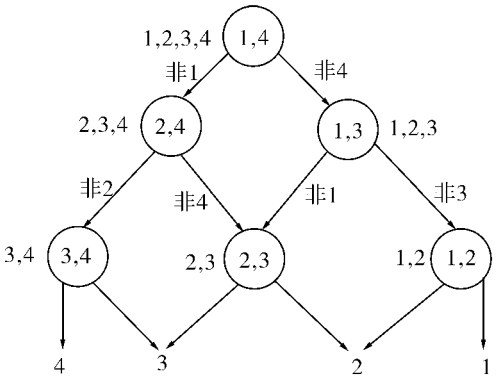


图4 4个单分类器用DAGSVM融合

Fig.4 Four single sorters fused with DAGSVM

3 实验结果

通常分类器所选的训练文本集和测试文本集的质量是最直接影响分类精度的因素之一.一般要选择公认的、通用的语料集,而且数据集中所选类别应是典型的、含有明显类别信息的文本类别,并且所选文本应该是客观存在的各个类别中的实际文本.但是,对于哈萨克文文本分类器来说,目前还没有公认的标准语料集,本文所构建的语料集尽管没有达到上述标准,但作为初期研究哈萨克文文本分类处理尚有研究意义.通过人工翻译等方法,笔者收集了一部分哈萨克语文本的内容,并做了人工分类.实验的训练集中共有5个类别,分别是交通、体育、医药、艺术、政治,其中交通包含8篇文章、体育包含12篇文章、医药包含10篇文章、艺术包含10篇文章、政治包含10篇文章.

3.1 词统计及文档的向量化表示

图5为每类文档词频统计结果,即分别在交通类、体育类、医药类、政治类、艺术类文档里词的总出现次数.

图6为词权重计算结果,词前数字表示各个词的权重,词的权重表明该词在判别文档类别所属过程中的重要程度.

图 5　词频统计结果

Fig. 5　Statistical results of term frequency

图 6　词权重计算结果

Fig. 6　Term weight computed results

图 7 为生成的文档向量. 它是文档的特征向量表示,冒号前的数字表示特征词的编号,冒号后的数字表示生成的文档的特征向量.

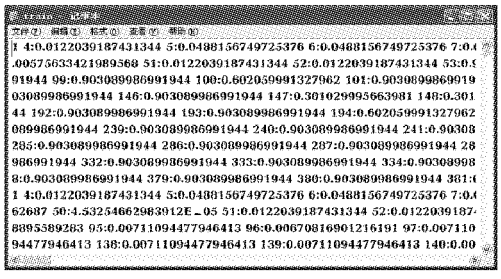


图 7　文本向量文件

Fig. 7　Text vector files

3.2　分类器的模型表示

分类器在整个文本分类系统中处于核心地位,本文采用了台湾大学 LibSVM 的开源代码. LibSVM 在给出源代码的同时还提供了 Windows 操作系统下的可执行文件,包括:进行支持向量机训练的 svmtrain. exe、根据已有 SVM 模型对数据集进行预测的 svmpredict. exe 以及对训练数据与测试数据进行简单缩放的 svmscale. exe. 它们都可以直接在 DOS 环境中使用,也便于研究者根据需要进行改进(譬如设计符合自己特定问题需要的核函数等).

通过执行 svmtrain. exe 实现对训练数据集的训练,可获得如图 8 所示的 SVM 模型文件.

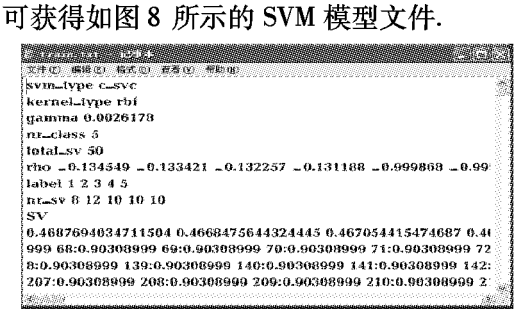


图 8　分类器模型文件

Fig. 8　Classifier model file

根据训练获得的模型,可对数据集进行预测. 经 svmpredict 分类后,在存放结果的文件中会出现一列

数字,这些数值就对应着输入文件中每个样本的分类类别.经过系统分类的结果势必与理想情况有一定的差距,衡量差距的方法就是不断地调试参数直到满足一定的要求为止.最后,还要通过 svmscale.exe 对训练数据与测试数据进行简单缩放.

通过与文本实际类别对比统计,可得如图9所示的哈萨克文文本分类器的性能.

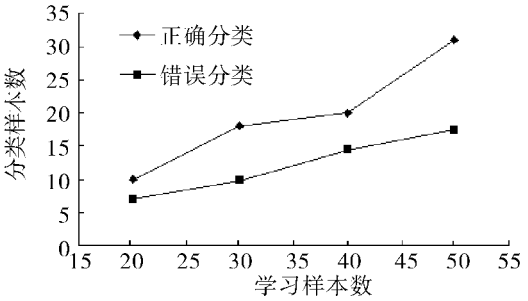


图9 分类正确率随学习样本数变化曲线

Fig.9 Classification accuracy changes with the number of training samples

实验结果表明,随着样本数的增加,分类器的区分度也明显提高.当然,所增加的样本质量也是直接影响区分度的因素,所以保证样本质量才能得到较理想的效果.在文献[15]的实验中还表明:选择不同的采样数 n,可以得到近似的结果.

进一步的研究表明:n 越小,分类效果越好,但需要学习的次数就越多,计算的复杂度就越大.从分类正确率和计算复杂度综合考虑,需要选择一个恰当的 n 值.

4 结束语

本文提出了一种基于主动学习支持向量机的哈萨克文文本分类方法.首先,实现哈萨克文词干解析以完成哈萨克文文本的读取预处理,然后用文本频率比值法 DFR 进行特征降维,用向量空间模型 VSM 表示文本,最后采用支持向量机对文本进行训练和分类,给出了哈萨克文文本分类方法的实验结果与分析.这也初步实现了基于支持向量机的哈萨克文文本分类工作,达到了预期研究目标.

根据实验结果,与其他性能较好的分类软件相比尚有一定的差距.究其原因,主要有以下几点.1)人工创建的语料库的质量和数量会直接影响最终结果.从数量上说,本文使用的语料相对较小,因此影响了语料的质量.通常,语料质量的提高可以通过增加文本的数量来弥补,降低类别代表性不足的缺点,更能提供足够的经验来强化典型和淡化例外.同时,测试语料也会一定程度上影响最终结果的正确性.2)对哈萨克语单词的切分处理.系统错误的切分单词会导致特征选择的误差,影响文本内容的表示,所提出的哈萨克文词干提取规则还不够完善,仍需要继续做详细的划分规则.

今后的研究中,在做好基础性工作的前提下,深入研究哈萨克文文本分类及其优化技术,利用现有语料库,结合其他新的文本分类方法构建混合型分类器以提高文本分类效果.

参考文献:

[1]LV lin, LIU Yushu. Research of English text classification methods based on semantic meaning[C]//2005 International Conference on Information and Communication Technologies. Karachi, Pakistan, 2005: 689-700.

[2]陈立伟,井志强,葛秘蕾.基于特征项扩展的中文文本分类方法[J].应用科技, 2010, 37(3): 1-5.
CHEN Liwei, JING Zhiqiang, GE Milei. A Chinese text classification method based on feature expansion[J]. Applied Science and Technology, 2010, 37(3): 1-5.

[3]张彰,樊孝忠.一种改进的基于 SVM 的文本分类算法[J].计算机工程与设计, 2006, 27(21): 4078-4080.
ZHANG Zhang, FAN Xiaozhong. Improved VSM based on Chinese text categorization[J]. Computer Engineering and Design, 2006, 27(21): 4078-4080.

[4]JOACHIMS T. Text categorization with support vector machines: learning with many relevant features[C]//Proceedings of the 10th European Conference on Machine Learning (ECML-98). Chemnitz, Germany: Springer Verlag, 1998: 137-142.

[5]VAPNIK V N. Statistical learning theory[M]. New York, USS: John Wiley & Sons Inc, 1998: 375-570.

[6]SEBASTIANI F. Machine learning in automated text categorization[J]. ACM Computing Surveys, 2002, 34(1): 1-47.

[7]TONG S, KOLLER D. Support vector machine active learning with applications to text classification[J]. Journal of Machine Learning Research, 2001, 2(1): 45-66.

[8]BURGES J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.

[9]朱红斌,蔡郁.基于主动学习支持向量机的文本分类[J].计算机工程应用, 2009, 45(2): 134-136.
ZHU Hongbin, CAI Yu. Text categorization based on active learning support vector machines[J]. Computer Engineering and Applications, 2009, 45(2): 134-136.

[10]刘宏,屠铁清,黄上腾.一种新的支持向量机主动学习策略及其在文本分类中的应用[J].计算机科学, 2003, 30(6): 110-112.
LIU Hong, TU Zhiqing, HUANG Shangteng. A new support vector machines active learning approach and its appli-

cation in text classification[J]. Computer Science, 2003, 30(6): 110-112.

[11]刘志刚,李德仁,秦前清,等. 支持向量机在多类分类问题中的推广[J]. 计算机工程与应用, 2004, 40(7): 10-13, 65.
LIU Zhigang, LI Deren, QIN Qianqing, et al. An analytical overview of methods for multi-category with support vector machines[J]. Computer Engineering and Applications, 2004, 40(7): 10-13, 65.

[12]张健沛,徐华. 支持向量机(SVM)主动学习方法研究与应用[J]. 计算机应用, 2004(1): 1-3.
ZHANG Jianpei, XU Hua. Study and application of active learning with SVM[J]. Computer Applications, 2004(1): 1-3.

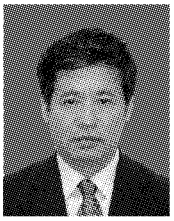
[13]PLATT J C, CRISTIANINI N, SHAWE-TAYLOR J. Large margin DAGs for multiclass classification[C]//Proceedings of Neural Information Processing Systems. Cambridge, USA: MIT Press, 2000: 547-553.

[14]贺慧,王俊义. 主动支持向量机的研究及其在蒙文文本分类中的应用[J]. 内蒙古大学学报:自然科学版, 2006, 37(5): 560-563.
HE Hui, WANG Junyi. Study of active learning support vector machine and its application on Mongolian text classification[J]. Acta Sientiarum Naturalium Universitatis NeiMongol, 2006, 37(5): 560-563.

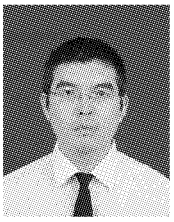
[15]SCHOHN G, COHN D. Less is more: active learning with support vector machines[C]//Proceedings of the Seventeenth International Conference on Machine Learning (ICML-2000). Stanford, USA, 2000: 839-846.

作者简介:

古丽娜孜,女,1972 年出生,讲师,主要研究方向为数据挖掘、文本分类等,发表学术论文 10 余篇.



孙铁利,男,1956 年出生,教授,博士生导师,伊犁师范学院兼职教授. 主要研究方向为智能用户接口、智能信息挖掘. 近年来承担国家级、省部级科研项目 8 项. 发表学术论文 100 余篇,出版专著及教材 10 余部.



伊力亚尔,男,1978 年出生,讲师,主要研究方向为计算机应用、自然语言信息处理,发表学术论文 10 余篇.

2011 第 6 届智能系统与知识工程国际会议(ISKE 2011)
2011 International Conference on Intelligent Systems and Knowledge Engineering (ISKE 2011)

The 2011 International Conference on Intelligent Systems and Knowledge Engineering (ISKE2011) is the sixth in a series of ISKE conferences. ISKE2011 follows the successful ISKE2006 in Shanghai, ISKE2007 in Chengdu, ISKE2008 in Xiamen, China, ISKE2009 in Hasselt, Belgium, ISKE2010 in Hangzhou, China, and will be held on Dec. 15-17, 2011 in Shanghai. Technically co-sponsored by California State University, Southwest Jiaotong University, Belgian Nuclear Research Centre (SCK CEN), Springer, and organized by Shanghai Jiao Tong University, ISKE 2011 emphasizes current practice, experience and promising new ideas in the broad area of intelligent systems and knowledge engineering.

Important Dates

Full paper submission: Aug. 31, 2011

Acceptance notification: Sept. 20, 2011

Final registration: Sept. 30, 2011

Final papers submissions: Sept. 30, 2011

Contact Us

E-mail: iske2011@cs.sjtu.edu.cn (for English language contact); iske2011@163.com (for Chinese language contact)

Secretary: Jing Zhou (周静)

Telephone: +86-18918042182