

## 半监督学习和主动学习相结合的浅层语义分析

陈耀东, 王挺, 陈火旺

(国防科技大学计算机学院, 长沙 410073)

**摘 要:** 语义分析是基于内容的文本挖掘领域的重要技术和研究难点。有监督机器学习方法受限于标注语料的规模, 在小规模标注样本中难以获取较高性能。本文面向浅层语义分析任务, 采用一种新颖的半监督学习方法——直推式支持向量机, 并结合其训练特点提出了基于主动学习的样本优化策略。实验表明, 本文提出的浅层语义分析方法通过整合主动学习与半监督学习, 在小规模标注样本环境中取得了良好的学习效果。

**关键词:** 浅层语义分析, 直推式支持向量机, 样本优化, 主动学习

## Combining Semi-Supervised Learning and Active Learning for Shallow Semantic Parsing

Yaodong Chen, Ting Wang, Huowang Chen

(School of Computer, National University of Defense Technology, Changsha, China, 410073)

**Abstract:** Semantic analysis is one of the fundamental and key problems in the research of content-based Text Mining. Most of supervised machine learning methods had poor performance when work on limited tagged data. This paper investigated a novel semi-supervised learning algorithm--Transductive Support Vector Machine for shallow semantic parsing. An optimization strategy of selecting training instances, based on active learning, was integrated with TSVM. The experiment result shows that the method integrating TSVM and optimization strategy for shallow semantic parsing outperforms supervised methods on small tagged data.

**key words:** Shallow Semantic Parsing, Semi-Supervised Learning, Transductive SVM, Active Learning

### 1 引言

文本挖掘的任务之一是通过语义分析技术获取半结构或无结构文本的内容与意义。

传统语义分析方法以深层分析为指导,旨在获取句义的“全面”形式化表示,这种思想在处理复杂语言现象或者面对大规模文本应用时效率和准确率均得不到保证。近年来,浅层语义分析技术逐渐受到人们的关注,它仅面向与任务相关的部分内容,摒弃了处理全部语言信息的复杂性,因而能在真实环境中实现快速高效的分析。语义角色标注(Semantic Role Labeling),简称 SRL 是浅层语义分析的一种实现技术,其目标是识别句子谓词的论元并标注相应的句子成分。当前,语义角色标注主要采用有监督学习方法,在大规模语料环境中能获取较好的效果[1]。然而,标注语料的获取需要花费大量人力物力,且语料的构建通常与特定的语义理论和动词分类体系密切相关,从而导致语义资源在实际应用中既难以获取,也难具通用性。因此,研究基于小规模标注语料的高性能学习器对提高浅层语义分析的实际应用能力具有重要意义。

通常情形下,大规模未标注文本可以简单方便地获取。鉴于这一事实,本文引入了一种新颖的半监督学习方法——直推式支持向量机(Transductive Support Vector Machines,简称 TSVM)[2]。TSVM 仅以少量标注样本训练,利用待分类样本的内部信息提高学习器的局部分类性能。一般地,一个机器学习方法的语料选取是随机的,这是基于学习理论的一个基本假设,即训练样本与待处理文本独立同分布。然而,这一假设在半监督学习中通常难以成立,因为大量未标注文本可能来自于与标注样本的分布不同或完全未知的环境,另外这些未加工的自由文本多半带有噪声。那么,如何从海量自由文本中选取有利于学习模型的训练样本以及如何去除噪声,这是半监督学习方法值得研究的重要课题。为此,我们结合 TSVM 的特点提出了基于主动学习(Active Learning)的样本优化策略,对语料进行指导性选择。我们面向中文浅层语义分析任务,设计了整合主动学习和 TSVM 的浅层语义标注系统,实验比较了有监督学习方法(SVM),评估了主动学习策略对 TSVM 的性能影响。实验表明,与传统有监督学习方法相比,整合主动学习和 TSVM 的语义标注训练算法能在小规模标注样本的环境中实现高效的分类能力。

本文的组织结构如下:第一节介绍本文工作的背景与动机;第二节给出浅层语义分析的定义、方法和资源;第三节首先介绍了 TSVM 和主动学习策略的基本原理及工作过程,然后描述了整合主动学习与半监督学习的浅层语义分析方法;第四节给出相关的实验数据与分析;第五节总结全文并提出下一步工作。

## 2 浅层语义分析

浅层语义分析是基于浅层分析思想对传统语义分析技术的改进。浅层语义分析首见于 CoNLL-2004[3],该会议提出语义角色标注的共享任务,其目标是基于大规模标注语料库开发具有识别动词框架并标注框架内语义角色的机器学习系统。下面给出 SRL 的定义:

• 定义 给定语义角色集  $R$ , 一个句子  $S$  和句子子串  $c_i(c_1, \dots, c_n \in S)$ , SRL 表示为:

对句子  $S$  的某个谓词  $p$ , 找出映射:  $c \rightarrow R \cup \text{NONE}$

SRL 通常可以分解成两个子任务——识别(Identification)与分类(Classification), 其中: Identification 表示为  $c \rightarrow \{\text{NONE}, \text{ARG}\}$  的二值分类函数,  $c$  代表任意的句子子串, 若  $c$  可用角色集  $R$  中的元素标注, 则赋值 ARG, 否则赋值 NONE, 识别过程主要是抽取可能标注为角色的候选成分; Classification 表示为  $c \rightarrow R$  的多值分类函数, 这里  $c$  表示候选标注的句子子串, 分类过程是给每个候选成分指派相应的语义角色, 是一个多值分类问题。

许多统计模型在 SRL 中已经获得了广泛的应用,如支持向量机(SVM)、最大熵(ME)、条件随机域(CRF)以及核方法(Kernel-based)[1]。文献[1]对比了 CoNLL-05[1]和

CoNLL-04[3]两次 SRL 系统评测的结果, 总结得出标注语料库的扩充是提升 SRL 系统性能的重要因素之一。然而, 标注语料的获取需要花费大量人力物力, 且语料的构建通常与某种语义理论相关, 从而导致在实际应用中, 大规模标注语义资源既难以获取, 也难以具有通用性, 这一点在中文语义处理上更为突出。对此, 我们会提出如下问题: 上述基于有监督学习方法的 SRL 系统能否在小规模标注样本环境中同样获取高性能? 面对大量易得的自由文本, 能否开发出基于未标注语料的 SRL 学习模型? 本文将引入一种新颖的半监督学习方法——直推式支持向量机, 它仅以少量标注样本训练, 利用待分类样本的内部信息有针对性地提高分类效果。当前 SRL 研究主要用到两个标注语料库, Propbank 和 FrameNet。Propbank 基于 Levin Class[4]的动词分类规范对 Penn Treebank 进行标注。Propbank 定义了 6 种核心角色与十余种附加角色, 所有谓词均使用这些角色标注句子成分。FrameNet 依据框架语义学思想, 语义角色与特定谓词相关, 每个语义角色代表该谓词所处场景的一个参与者 participant。本文的实验面向中文 SRL, 使用 Chinese Propbank 1.0 语料库<sup>1</sup>, 该语料以 Chinese Treebank 5 为标注对象, 共标注 37, 183 个谓词命题, 涉及 4, 865 个动词和 5, 298 个动词框架。

### 3 整合主动学习与半监督学习的浅层语义分析系统

#### 3.1 直推式支持向量机

当前机器学习的模式有两种: 归纳式学习 (Inductive Learning) 和直推式学习 (Transductive Learning)。归纳学习通过一批假定来自于整个问题空间的标注样本来训练一个全局模型 (global model), 然后用该模型处理新样本, 期望达到全局泛化错误最小化。SVM、ME 和 Kernel-based 方法均基于此学习模式。然而, 某些应用环境可能不具备归纳学习模式的条件, 这将导致传统学习器性能的下降。例如学习用的标注样本数量不足以支持学习器在假设空间内搜索到目标函数。与归纳学习不同, 直推式学习使用待分类样本的内在信息获取与问题相关的局部模型 (local model), 节省了构建全局模型的开销。TSVM 是直推式学习的一种实现方法, 它以 SVM 为基础, 使用待分类样本的内部特征加大 SVM 的分类超平面间隔, 使分类器有针对性的提高分类效果。

给定一组已标注样本  $(x_1, y_1), \dots, (x_n, y_n) \in S_{train}$ ,  $y_1, \dots, y_n \in \{-1, +1\}$  和另一组未标注样本  $x_1^*, \dots, x_k^* \in S_{test}$ ,  $y_1^*, \dots, y_k^*$  为  $x_1^*, \dots, x_k^*$  的待分类值。TSVM 可表述为如下最优化问题[2]:

$$\begin{aligned} & \text{Minimize over } (y_1^*, \dots, y_k^*, w, b, \xi_1, \dots, \xi_n, \xi_1^*, \dots, \xi_k^*) \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + C^* \sum_{j=1}^k \xi_j^* \\ & \text{subject to: } \forall_{i=1}^n: y_i [w \cdot x_i + b] \geq 1 - \xi_i; \forall_{j=1}^k: y_j^* [w \cdot x_j^* + b] \geq 1 - \xi_j^*; \forall_{i=1}^n: \xi_i \geq 0; \forall_{j=1}^k: \xi_j^* \geq 0 \end{aligned}$$

其中参数  $C^*$  称为待分类样本在训练过程中的影响因子,  $C^* \xi_i^*$  称为待分类样本  $j$  在目标函数中的影响项。下面我们简单地介绍 TSVM 的工作过程, 具体的描述和证明参见文献[2]: 首先指定参数  $C$  和  $C^*$ , 对标注样本进行一次归纳式学习, 得到一个初始分类器, 并指定待分类样本中的正标注样本数  $N$ ; 接着初始分类器对待分类样本分类, 对输出值最大的  $N$  个待分类样本暂时赋正值, 其余的赋负值。指定一个临时影响因子  $C_{temp}^*$  然后对所有样本重新训练得到一个分类器, 按一定的规则交换一对标签值不同的待分类样本的标签符号, 使优化问题  $\text{Minimize over } (y_1^*, \dots, y_k^*, w, b, \xi_1, \dots, \xi_n, \xi_1^*, \dots, \xi_k^*)$  的目标函数值获得最大下降。

<sup>1</sup> <http://www.cis.upenn.edu/~chinese/cpb/>

这一步反复执行直到找不出符合交换条件的样本对为止, 算法均匀地增加临时影响因子  $C_{temp}^*$  值并反复交换标签值, 当  $C_{temp}^*$  大于  $C^*$  时算法结束, 输出所有未标注样本的标签值。

### 3.2 样本选取策略

主动学习[5]是指一个机器学习方法从候选样本集中按某种策略动态地选择样本进行训练的过程。从上一节我们了解到, TSVM 的训练中用到了未标注样本, 但是我们并不能据此得出语料越多, 学习性能就一定越好的结论。事实上, 半监督学习方法使用的未标注样本在实际应用中可能来自于不同环境, 语料的分布特征复杂且未知的, 也极可能带有噪声。主动学习可利用现有知识主动地选取最有可能解决问题的样本, 有效减少了所需评估样本的数量。当前成熟的主动学习策略较多, 如基于样本的不确定性方法、询问专家委员会的方法、版本空间和边缘的方法、基于统计的方法等。文献[6]提出一个基于 SVM 的主动学习算法, 其基本思想是: 首先从候选未标注样本集  $U$  中取任意个样本(至少包含正负例各一个)正确标注类别, 构造训练集  $T$ , 以  $T$  训练 SVM 分类器  $C$ ; 然后对  $U$  中所有样本使用  $C$  标注, 选择一个离分类超平面最近的未标注样本标注, 并加入训练集  $T$ ; 反复迭代第二步直至超过某一域值或达到退出条件。该算法实际上是采用版本空间和边缘方法选取样本标注。我们认为边缘方法与不确定性方法在效果上具有极大的相似性, 使用 SVM 分类值可以使两者达到统一。也即, 距离分类超平面近的样本所得到的分类绝对值越小, 不确定性越大, 反之亦然(分类值的绝对值衡量了分类正确的可能性, 也即确定性)。对于 TSVM 的待分类样本, 我们认为确定性强的其分布特征更为显著, 是噪声的可能性也越小。基于此, 我们提出针对 TSVM 的两种样本选取策略:

**主动学习策略 1** 对于待标注样本集, 每次迭代选取不确定性最强的一个样本标注

**主动学习策略 2** 对于待分类样本集, 每次迭代选取确定性强的若干样本

我们认为, 选取距离分类超平面近的样本进行标注类似于寻找支持向量, 这些样本对分类超平面的确定具有决定影响; 策略 2 实际是选择距离超平面远的待分类样本, 因为它们较为显著的分布特征可以加大对边缘间隔的影响, 而且距离越近, 是噪声的可能也越大。我们将在第五节通过实验评估两个学习策略的正确性。

### 3.3 整合主动学习与半监督学习的训练算法

本节介绍如何整合主动学习策略与 TSVM。基于 TSVM 的工作原理, 我们将 TSVM 的训练过程分解成两部分: SVM(即步骤 1、2)训练和 TSVM 迭代训练(即步骤 3、4)。我们对文献[6]的主动学习算法加以修改, 增加待分类样本集的主动学习过程, 并在迭代中使用分解后的 TSVM 分类器替换原先的 SVM。算法 1 展示了基本实现。算法首先从待标注样本集中选取任意个样本进行标注(至少包含正负例各一个), 然后训练 SVM 分类器并使用 SVM 分类待分类样本集。接着使用主动学习策略 2 从待分类样本集中抽取所有确定性样本, 与前面任取的标注样本组成训练集训练 TSVM。下一步, 训练得到的 TSVM 对剩下的待标注集与全部待分类集重新分类, 然后调用主动学习策略 1 和 2 构建新的训练集, 此过程迭代指定步后结束。我们以  $\pm 0.5$  作为样本确定性的阈值, 即样本通过分类器得到的分类值在  $\pm 0.5$  之间, 我们认为该样本是不确定的, 反之则认为是确定的。

**算法 1** 主动学习与半监督学习相结合的学习算法

输入  $S_{tag}$ : 待标注样本集, 且  $x_1, \dots, x_n \in S_{tag}$ ;  $S_{test}$ : 待分类样本集,  $x_1^*, \dots, x_m^* \in S_{test}$ ,  $y_1^*, \dots, y_m^*$  为  $x_1^*, \dots, x_m^*$  的待分类值;  $C$ : 分解后的 TSVM 学习器;  $d$ : 指定  $S_{tag}$  中需要标注的样本数 ( $d \leq n$ );  $i$ : 指定初始训练 SVM 的标注样本数

输出  $S_{test}$  的分类值  $y_1^*, \dots, y_m^*$

Step 1: 从候选样本集  $S_{tag}$  中任选  $i$  个样本标注 (至少正负例各一个), 构造初始训练样本集  $T$ , 以  $T$  为训练集训练一个 SVM 分类器

Step 2: 使用 SVM 分类器对  $S_{test}$  中的样本分类

Step 3: 从  $S_{test}$  中选取所有确定性超过某个阈值的样本, 构建  $T^*$

Step 4: 循环, 从 1 到  $d$ , 执行以下动作

4.1 以  $T$  作为标注样本集,  $T^*$  作为待分类样本集训练 TSVM 分类器

4.2 使用 TSVM 对  $S_{tag}-T$  类和  $S_{test}$  进行分类

4.3 调用主动学习策略 1, 选择  $S_{tag}-T$  中最不确定的样本加入  $T$ ;

4.4 调用主动学习策略 2, 选择  $S_{test}$  中所有确定性超过某个阈值的样本, 构建  $T^*$

Step 5: 使用 TSVM 分类器对  $S_{test}$  分类, 输出  $y_1^*, \dots, y_m^*$ , 算法结束 ■

## 4 实验与讨论

### 4.1 实验说明

本节我们评估基于整合算法所构建的 SRL 系统的学习性能。实验使用 SVM-light<sup>2</sup> 工具集。SVM-light 包含许多改善 SVM 模型性能和缩减训练时间的优化程序, 并提供了 TSVM 的近似实现。特征集是构造浅层语义分析系统的重要部分。我们参照文献[7] 处理中文 SRL 系统的特征集, 它们是已被证明为最具影响的 9 种核心特征, 列表 1 列出该特征集, 具体描述参见文献[7]。我们没有考虑文献[7]提到的组合特征和 verb class 特征, 这是因为前者实际上已经包含在上述特征集中, 而后者, 我们无法获悉它所依据的替换规则 (Alternations)。因此, 本文下面的实验将与文献[7]中未使用 verb class 特征的实验数据比较, 以保证两者在模型、语料和特征集上一致。对于分类器的设置, 我们建立了 15 个二值分类器对应 SRL 的 15 种主要角色 (Chinese Propbank 定义了 22 种语义角色类型, 其中 ARG5, ARGM-FRQ, ARGM-ASP, ARGM-PRD, ARGM-CRD, ARGM-T, ARGM-DGR 的出现频率极小, 仅占总频次的 0.02%, 予以忽略), 每个分类器使用算法 1 训练, 所有实验均采用微平均计算总体标注结果。我们首先在相同条件下构建与文献[7]相同的标注系统作为 baseline, baseline 的 F1 值为 89.97%, 性能与文献[7]描述的结果基本相当(90.3%)。

为评估算法 1 的各项学习性能并测试主动学习策略的正确性。我们将 Chinese Propbank 中的 propositions 转化为数值向量, 按不同比例划成待标注样本集和待分类样本集, 然后分别指定不同的标注样本数 (即算法 1 的  $d$  值) 和待分类样本数评估 TSVM 的学习性能。同时, 我们假设采用与主动学习策略 1 和 2 相逆的方式选取样本, 以评估主动学习策略的正确性。实验所使用的学习模型共有五种, 如表 2。

<sup>2</sup> <http://svmlight.joachims.org/>

表 1 浅层语义分析特征集

Tab.1 the features of Semantic Role Labeling

特征名	解释
Predicate	谓词词条
Subcat-Frame	动词的扩展规则
Path	角色对应的句法短语到谓词短语的路径
Position	角色对应的句法短语处于谓词短语的前后位置
Phrase Type	角色对应的句法短语类型
Phrase type of the sibling to the left	角色对应句法短语的左边第一个兄弟结点
Head Word and its Part Of Speech	角色对应短语的中心词与该词词性
First and last word of the constituent in focus	角色对应短语的首词与尾词
Syntactic Frame	与谓词相关的名词短语组成的框架

表 2 五种学习模型说明

Tab.2 the presentation of the five learning models

模型	说明
模型一	随机选取d个标注样本训练baseline, 即经典的受监督学习方法 (SVM)
模型二	随机选取d个标注样本和所有待分类样本训练 TSVM, 不使用主动学习策略
模型三	按算法1的框架训练TSVM, 但仅使用主动学习策略1选取d个样本标注
模型四	按算法1的框架训练TSVM, 使用与主动学习策略1 相反的方式选取d个标注样本 (即每次迭代选取确定性最高的样本标注), 对待分类样本集使用与主动学习策略2相反的方式 (即每次迭代选取确定性低的所有样本重新训练)
模型五	按算法1的框架训练TSVM, 使用主动学习策略1选取d个样本标注, 对所有待分类样本使用主动学习策略2

## 4.2 实验结果及讨论

表 3-A 为固定 5000 个待分类样本条件下, 五个模型的学习性能在不同 d 值(即标注样本数)下的影响。模型一与模型二对比了有监督学习模型(SVM)与半监督学习模型(TSVM)在小规模标注样本环境中的性能, 我们看到当标注样本数小于 400 时, TSVM 的性能高出 SVM 大约 4%~6%, 这说明 TSVM 有效地利用了待分类数据的分布信息从而快速提高分类能力; 当标注样本数增大到一定程度(超过 800), SVM 的性能与 TSVM 趋近, 这说明 SVM 在标注语料充足环境中具有强泛化能力。模型二与模型五的对比评估了使用主动学习策略后对 TSVM 影响。与前面 TSVM 与 SVM 的比较相反, 当标注样本小于 200 时, 采用主动学习后性能提升不大, 但随着标注样本数量的增大, 主动学习对 TSVM 性能的影响增加(大约 1%~2%), 这说明主动学习可以在大规模标注样本集上发挥作用。模型三、四、五对不同的主动学习策略进行了对比。可以看出, 同时使用两个主动学习策略要比仅使用一个策略性能稍高(约 1%)。而若采用相反的样本选取策略(即模型四), 其性能比有监督模型(模型一)还要低(仅 40%), 这从另一个角度说明选择标注样本的好坏对有监督学习方法的性能有很大影响。

表 3 学习模型的性能比较

Fig.2 The performance estimation of five statistical models

表3-A 五种学习模型在不同标注样本数下的学习性能  
能固定5000个待分类样本

标注样本数	模型一	模型二	模型三	模型四	模型五
10	32.76	34.96	34.96	42.33	44.87
20	52.20	42.04	34.96	42.07	41.75
40	62.58	68.16	65.65	41.06	66.09
80	74.19	78.33	77.34	42.78	78.47
100	71.83	78.41	78.40	43.49	79.01
200	74.17	78.63	79.31	40.41	80.10
400	78.39	78.99	79.28	40.58	80.12
800	80.48	80.98	79.99	41.14	81.40
1000	78.76	79.21	80.33	40.28	81.56
2000	78.47	80.04	81.78	41.61	82.00
5000	79.40	80.00	81.90	41.67	82.32

表3-B 模型一、二和五在不同规模待分类样本下的学习性能, 固定200个标注样本

待分类样本数	模型一	模型二	模型五
100	72.37	77.00	77.10
200	72.45	77.23	77.34
400	72.90	77.68	78.00
800	73.05	78.00	78.10
1000	73.40	77.99	78.15
2000	73.00	78.78	79.01
4000	73.10	79.00	79.50
5000	73.80	79.49	80.09
8000	73.81	80.21	80.30
10000	73.70	80.22	80.40

为比较各模型在不同规模未标注样本下的学习性能,我们固定从待标注样本集中选取 200 个(即  $d=200$ )进行标注,然后观察待分类样本规模对三种学习模型(模型一、二、五)学习性能的影响。各模型的 F1 值见表 3-B。有监督学习模型(模型一)的 F1 值近似平坦,因为其学习性能仅与标注样本数有关,而与未标注样本数无关(实验中由于待分类样本的不同,分类效果略有波动);模型二与模型五的性能随着待分类样本规模的增大而稳步上升,分别从 77%和 77.10%提高到了 80.22%和 80.40%。我们对比模型二与模型五发现,尽管主动学习对 TSVM 有性能提升,但效果不明显(不到 1%)。我们研究模型五对待分类样本的分类值,发现其中不确定性样本(即分类值在  $\pm 0.5$  之间)的比例很小。这说明实验中的待分类样本分布特征显著,本身就具有较强的区分性,也不存在噪声,因而主动学习策略 2 几乎不必要对它们选择。我们认为当前的实验环境未能很好的模拟实际应用,主动学习可期待在实际应用发挥更明显的作用,这也为我们的下一步工作提出了要求。

## 5 总结

浅层分析是近十年来计算语言学在方法学上的重大突破。语义角色标注以浅层分析为指导思想,摆脱了分析全部语言成分和关系的复杂性,实现了高鲁棒性和强应用性,从而为基于内容的文本挖掘提供了新的解决途径。当前大多数语义角色标注系统都是建立大规模标注语料的基础上,在小样本环境中的性能难以得到保证。本文阐述了一种新的半监督学习方法——TSVM,它仅需要少量标注样本,利用待分类数据的内部信息提高对它们的分类能力。针对 TSVM 的特点,我们设计了两种基于主动学习的样本选取策略分别对待标注集和待分类集进行优化。在此基础上,我们面向中文 SRL 任务,开发了一个整合主动学习与 TSVM 的角色标注训练模型。实验表明,整合训练模型在小规模标注样本中具备比有监督学习模型更高的性能,验证了主动学习策略的正确性和一定程度上的有效性。事实上,整合主动学习与半监督学习的训练模型不仅适用于 SRL 任务,还可以方便快速地移植到其它标注语料受限的应用环境中。下一步,我们将从各种不同的实际环境收集更多的自由文本,进一步评估整合训练算法在 SRL 和其它领域的应用能力。

## 参考文献

- [1] Xavier Carreras, Lluís M'arquez. Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling. In Proceedings of CoNLL-2005, 2005.
- [2] T.Joachims. Transductive inference for text classification using support vector machines. In Proc. of ICML-99, 1999 pages 200-209, Bled, Slovenia, Jun.
- [3] X Carreras, L M'arques. Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling. In Proceedings of CoNLL-2004, Boston, 2004.
- [4] Levin, Beth. English Verb Class and Alternations: A Preliminary Investigation. Chicago: University of Chicago Press. 1993.
- [5] AP Engelbrecht et al. Incremental Learning Using Sensitivity Analysis. IJCNN'99, International Joint Conference on Neural Networks. 1999(2):1350-1355.
- [6] 张健沛, 徐华. 支持向量机主动学习方法研究与应用. 计算机应用, 2004, Vol. 24.
- [7] Nianwen Xue, Martha Palmer. Automatic Semantic Role Labeling for Chinese Verbs. In The Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland. 2005.

作者: 陈耀东, 王挺, 陈火旺  
作者单位: 国防科技大学计算机学院, 长沙 410073  
被引用次数: 1次

## 本文读者也读过(8条)

1. 胡正平, 高文涛. HU Zheng-ping, GAO Wen-tao 基于改进加权压缩近邻与最近边界规则SVM训练样本约减选择算法[期刊论文]-燕山大学学报2010, 34(5)
2. 刘石, 李合, 王啸吟, 张路, 谢冰 基于语法与语义分析的代码搜索结果优化[会议论文]-2008
3. 蔡健平, 王琳琳, 林世平 基于机器学习的词语和句子极性分析[会议论文]-2007
4. 李国正, 杨杰, 孔安生, 陈念貽 基于聚类算法的选择性神经网络集成[期刊论文]-复旦学报(自然科学版)2004, 43(5)
5. 刘晓坤 基于遗传算法的混合学习和集成学习研究[学位论文]2003
6. 陈冰, 张化祥, CHEN Bing, ZHANG Hua-xiang 集成学习的多分类器动态组合方法[期刊论文]-计算机工程2008, 34(24)
7. 倪勇, 吴汶芪, 李君, NI Yong, WU Wen-qi, LI Jun 基于小样本的集成学习研究[期刊论文]-机电工程2009, 26(12)
8. 谭泗桥, 袁哲明, 柏连阳, 熊洁仪, TAN Si-qiao, YUAN Zhe-ming, BAI Lian-yang, XIONG Jie-yi 基于支持向量机回归与K-最近邻法的组合预测用于除草剂QSAR建模[期刊论文]-农药学报2007, 9(4)

## 引证文献(1条)

1. 徐萍 如何提高小学语文教学有效性[期刊论文]-未来英才 2013(24)

引用本文格式: 陈耀东, 王挺, 陈火旺 半监督学习和主动学习相结合的浅层语义分析[会议论文] 2007