

中文口语理解弱监督训练方法

李艳玲^{1,2*}, 颜永红²

(1. 内蒙古师范大学 计算机与信息工程学院, 呼和浩特 010022; 2. 中国科学院语言声学内容与理解重点实验室(中国科学院声学研究所), 北京 100190)
(*通信作者电子邮箱 liyanling7871397@163.com)

摘要:标注数据的获取一直是有监督方法需要面临的一个难题,针对中文口语理解任务中的意图识别研究了结合主动学习和自训练、协同训练两种弱监督训练方法,提出在级联框架下,从关键语义概念识别中获取语义类特征子集和句子本身的字特征子集分别作为两个“视角”的特征进行协同训练。通过在中文口语语料上进行的实验表明:结合主动学习和自训练的方法与被动学习、主动学习相比较,可以最大限度地降低人工标注量;而协同训练在很少的初始标注数据的前提下,利用两个特征子集进行协同训练,最终使得单一字特征子集上的分类错误率平均下降了0.52%。

关键词:意图识别;口语理解;弱监督训练;协同训练;主动学习

中图分类号: TP391.1; TP18 **文献标志码:** A

Weakly-supervised training method about Chinese spoken language understanding

LI Yanling^{1,2*}, YAN Yonghong²

(1. College of Computer and Information Engineering, Inner Mongolia Normal University, Hohhot Nei Mongol 010022, China;
2. Key Laboratory of Speech Acoustics and Content Understanding, Chinese Academy of Sciences (Institute of Acoustics, Chinese Academy of Sciences), Beijing 100190, China)

Abstract: Annotated corpus acquisition is a difficult problem in supervised approach. Aiming at the intention recognition task of Chinese spoken language understanding, two weakly supervised training approaches were studied. One is combining active learning with self-training, the other is co-training. A new method of acquiring two independent feature sets as two views for co-training was proposed based on spoken language understanding data in cascade frame. The two feature sets were character features of sentence and semantic class features obtained from key semantic concept recognition task. The experimental results on Chinese spoken language corpus show that the method combining active learning with self-training can minimize manual annotation compared with passive learning and active learning. Furthermore, under the premise of a few initial annotation data, co-training based on two feature sets can make the classification error rate fall in an average of 0.52% with single character feature set.

Key words: intention recognition; spoken language understanding; weakly-supervised training; co-training; active learning

0 引言

口语对话系统(Spoken Dialog System, SDS)发展了几十年,至今仍旧存在一些问题,比如系统的鲁棒性问题、领域移植性问题等。在系统开发过程中,使用有监督方法首先面临的一个问题是数据标注。拥有大量标注数据是训练优质分类器的前提。然而,通常在口语对话系统开发初期,获得大量标注数据是非常困难的,这成为影响口语对话系统发展的主要瓶颈。然而,与获取大量标注数据相比,未标注数据的获取则相对容易很多,因此研究短时间内如何利用未标注数据进一步提高系统性能是一个非常意义的研究课题。换句话说,研究分类器的弱监督(weakly-supervised)训练^[1]对于缩短系统开发周期、减少开发成本具有重要的意义和价值。

口语理解方面的弱监督训练研究代表工作主要有:应用主动学习和半监督学习进行对话行为的弱监督训练^[2]。其思想为:利用少量标注样本进行有监督训练,再利用这个训练好的模型对大量无标注的样本进行分类,选择置信度较低的样本进行人工标注。基于支持向量机(Support Vector Machine, SVM)的弱监督口语理解方法使用主动学习和半监督学习相结合对意图识别进行弱监督训练,同时还对语义槽分类提出了一种利用文字上下文特征和语义槽上下文特征进行语义槽分类的弱监督训练。

本文使用级联的口语理解框架,即首先对用户的输入语句进行关键语义概念识别,然后再进行意图识别。采用这种框架的原因是:关键语义概念识别的结果可以为意图识别提供有效特征。而众多丰富的特征有利于挑选不同视角的特征

收稿日期:2015-01-23;修回日期:2015-03-19。

基金项目:国家自然科学基金资助项目(10925419, 90920302, 61072124, 11074275, 11161140319, 91120001, 61271426);中国科学院战略性先导科技专项(XDA06030100, XDA06030500);国家863计划项目(2012AA012503);中国科学院重点部署项目(KGZD-EW-103-2);内蒙古师范大学“十百千”人才培养工程项目;内蒙古自然科学基金面上项目(2012MS0930, 2013MS0912);内蒙古自治区高等学校科学研究项目(NJZY12032, NJZY028);内蒙古师范大学引进高层次人才科研启动经费项目(2014YJRC036)。

作者简介:李艳玲(1978-),女,内蒙古呼和浩特人,讲师,博士,CCF会员,主要研究方向:自然语言处理、机器学习、信号处理;颜永红(1967-),男,江苏无锡人,研究员,博士生导师,主要研究方向:语音信号处理、口语系统及多模系统、人机界面。

集进行协同训练。本文的弱监督训练主要针对意图识别任务。本文的突出贡献在于:1)提出使用级联框架为协同训练提供有效的语义类特征,使得协同训练可以应用于口语语料;2)比较和分析了几种弱监督训练方法,并将其应用到口语理解中以减少系统开发初期人工标注的工作量。

1 相关工作

弱监督训练是只有小部分标注样本,利用大规模未标注样本提升分类器性能的一种机器学习方法。主动学习(active learning)^[3]是一类利用小部分标注样本达到对未标记样本进行选择标注的机器学习方法。通常人们主要通过被动学习(passive learning)的方法对样本进行标注,即随机选择需要标注的训练样本,进而得到具有相似分布的训练集和测试集。主动学习方法可以分为两类:基于确定性(certainty-based)和基于委员会(committee-based)^[4]。这两类方法的区别在于采取不同的策略确定样本的信息量:前者主要通过分类器预测样本的置信度判断样本的信息量;而后者通过小规模样本得到的多个分类器对未标注样本类别的投票结果评判样本的信息量。

主动学习方法已经应用于很多领域,如文本聚类^[5]、观点分析^[6]、视频分类^[7]、词性标注^[4]、视觉场景理解^[8]等。

文献[9]较早地介绍了半监督学习,但未标注样本的价值早在 20 世纪 80 年代末就已经被一些学者意识到了。随着统计学习技术的不断发展,以及利用未标注样本这一需求的日渐强烈,半监督学习才在近年来逐渐成为一个研究热点。在文本挖掘领域,半监督学习已经被广泛应用,如文本分类^[10]、实体识别和关系抽取等。常用的半监督学习方法有:自训练、协同训练、半监督支持向量机、基于图的半监督学习等^[11]。

另外,将上述两种方法结合起来最早是由文献[12]提出来的,文中使用主动学习和半监督期望最大化(expectation Maximization, EM)算法结合进行文本分类。除此之外,将主动学习和半监督学习相结合的方法还应用到对话行为分类^[2]、图像分类^[13-14]等。但还未见到任何将协同训练应用于口语语料的研究。

2 本文的弱监督训练方法

2.1 本文的口语理解框架

口语理解任务可以分解为两个子任务:关键语义概念识别和意图识别。本文的口语理解框架使用级联形式,即先进行关键语义概念识别,再进行意图识别。而本文所描述的弱监督训练主要针对意图识别进行。这里的意图定义为:用户输入语句所属的领域。本文的领域主要分为:影视、音乐、网站、电视台和应用。

本文将意图识别定义为文本分类问题,利用支持向量机的方法进行解决。传统的文本分类使用文本中的字或者词作为特征,但由于口语数据稀疏,全部的字和词特征不具有很好的鉴别性,因此本文对字特征和词特征进行了筛选,即从训练语料中统计各类别频率较高的字词,经过人工筛选之后分别生成字和词两种特征词典。其中词的生成使用中国科学院计算技术研究所的分词工具 ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) (<http://www.iclcl.org/>)得到。另外,关键语义概念识别得到的语义概念类别可以为意图识别提供深层的语义类特征。深层特征可以更好地反映用户输入语句的语义,具有更强的分类能力;而且

它能把词聚集到相对较少的几类,可以减少分类模型的参数数目,能减轻数据稀疏问题。此时意图识别就有了这样两大类特征,即字/词特征、深层语义类特征。

2.2 结合主动学习和自训练

本文选用基于确定性的主动学习方法,即利用少量已经标注的样本训练一个初始分类器,然后用该分类器对未标注样本进行预测,选择置信度较低的样本进行人工标注,然后将标注好的样本放入训练集中,与原始种子数据一起重新训练分类器。这个过程迭代进行,直到标注完所有样本。半监督学习采用自训练方法,与基于确定性的主动学习类似,不同的是自训练方法选择置信度比较高的样本,使用机器预测的类别作为样本的类别,然后将标注好的样本放入到训练集中。

将主动学习和自训练结合起来的方法:在每轮迭代中,基于确定性主动学习总是选择置信度较低的未标记样本进行人工标注,而自训练总是选择置信度较高的样本使用机器标注加入到训练集中。这两种方法是互补的,因此两者可以结合起来。具体的算法如下。

算法 1 结合主动学习和自训练。

1) 给定已标注的小样本集 S_t 和样本池中未标注的大样本集 $S_u = s_1, s_2, \dots$, 使用 S_t 训练一个分类器;

2) 若 S_u 非空,则重复以下过程:

a) 从 S_u 取出 n 个样本到样本池 S_p 中;

b) 利用当前的分类器对 S_p 中的样本进行分类,并计算相应的置信度分数 $CS(s_i), i = 1, 2, \dots, n$;

c) 从 S_p 中选择数据集 $S_m = \{s_i: CS(s_i) \geq th\}$, th 为阈值;

d) 手工标注 S_p 中剩下的样本集 $S_k = S_p \setminus S_m$;

e) $S_t = S_t \cup S_k \cup S_m$;

f) 在新的训练集 S_t 上重新训练分类器。

该算法结合了主动学习和自训练算法。该算法与主动学习算法不同之处在于:后者忽略具有较高置信度的未标注样本,而前者利用了它们。另外,在该算法中,主动学习和自训练可以使用同一个阈值。如果手工标注的样本非常少,则可以以为自训练设定另外一个阈值以避免引入噪声数据。因为置信度分数较低的样本已经被主动学习选择,并且被手工标注,这样为自训练减少了噪声数据。

2.3 协同训练

协同训练(co-training)方法是由 A. Blum 等^[15]在 1998 年提出的。该方法是基于这样的假设:即每个样本存在两个具有相互独立“视角”的特征子集,用 $x = [x^{(1)}, x^{(2)}]$ 表示,并且每一个特征子集都可以训练一个足够好的分类器。如关键语义概念识别任务,第一类特征集可以用关键语义概念本身包含的词作为特征,第二类特征集可以用语义概念的上下文词汇作为特征。在网页分类任务,第一类特征集可以用文档中出现的词汇作为特征,第二类特征集可以用其他网页链接到当前网页的锚文本作为特征^[15]。使用少量样本在两个视角 $x^{(1)}$ 和 $x^{(2)}$ 上分别训练两个分类器,在协同训练过程中,每个分类器独立对未标记样本进行预测,并从预测结果中挑选出若干置信度较高的样本连同预测结果一起加入另一个分类器的训练集中,以便对方利用这些新标记的样本进行更新。协同训练过程迭代进行,直到满足结束条件为止。文献[16]对 co-training 进行了实验分析,表明 co-training 在特征条件独立的情况下具有较好的效果。本文利用协同训练进行意图分类,具体算法如下。

算法 2 协同训练。

给定已标注的样本集 $S_t = \{(x_i, y_i), i = 1, 2, \dots, l\}$ 和未

标注的样本集 $S_u = \{x_j, j = l+1, l+2, \dots, l+u\}$, 每个样本存在两个视角 $x_i = [x_i^{(1)}, x_i^{(2)}]$, 对 S_u 中的数据从两个视角提取特征, 得到 $S_u^{(1)}, S_u^{(2)}$ 。

1) 初始化两个训练样本集 $L_1 = \{(x_1^{(1)}, y_1), (x_2^{(1)}, y_2), \dots, (x_l^{(1)}, y_l)\}$, $L_2 = \{(x_1^{(2)}, y_1), (x_2^{(2)}, y_2), \dots, (x_l^{(2)}, y_l)\}$ 。

2) 以下步骤迭代 n 次:

a) 用 L_1 和 L_2 分别训练两个分类器 $f^{(1)}$ 和 $f^{(2)}$;

b) 用 $f^{(1)}$ 和 $f^{(2)}$ 分别对 $S_u^{(1)}$ 和 $S_u^{(2)}$ 中的样本进行分类;

c) 将 $f^{(1)}$ 分类得到的前 k 个置信度最大的样本 $(x^{(1)}, f^{(1)}(x))$ 放入集合 L_2 ;

将 $f^{(2)}$ 分类得到的前 k 个置信度最大的样本 $(x^{(2)}, f^{(2)}(x))$ 放入集合 L_1 ;

将上述选择的样本分别从 $S_u^{(1)}$ 和 $S_u^{(2)}$ 中删除。

2.4 几种方法的理论分析

本文使用的几种方法包括主动学习、半监督学习(自训练)、结合主动学习和自训练、协同训练。主动学习可以通过少量标注的样本选择提升分类器性能的样本进行人工标注。其优势是分类器的性能可以随着样本数量的增加而提升, 然后收敛到最佳值; 缺点是尽管样本是通过机器选择出来的, 然而避免不了人工标注。半监督学习(自训练)可以通过少量标注的样本自动对未标注样本进行标注。其优势是整个过程无需人工标注; 缺点是由于采用了机器自动标注的样本进行训练, 可能会混有噪声, 导致分类性能提升缓慢甚至可能下降。结合主动学习和自训练将两者的优势结合起来, 这种方法既减少了人工标注量, 又保证了机器标注结果的正确性。和单独使用两种方法比较, 该方法的收敛速度最快。协同训练方法和自训练类似, 主要是利用两个少量的已经标注好的训练集互相配合为对方提供标注样本。该方法的优点是整个过程无需人工标注数据, 缺点是只有当两个特征子集独立, 才能保证相互加入的样本可以提升各自分类器的性能, 从而达到自动标注数据的目的。与结合主动学习和自训练的方法比较, 协同训练更加便捷, 无需人工参与, 无需选择主动学习和自训练两部分加入训练集的样本个数。

3 实验

3.1 实验设置

本文所使用的数据由实验室内部收集, 其中训练数据 6000 句, 测试数据 2039 句。有监督训练是假定训练集中所有的样本都已经进行了人工标注, 而弱监督训练是假定训练集中只有少量的种子数据进行了人工标注, 希望通过这些种子数据对大部分未标注的样本进行标注, 以减少人工劳动, 并保持甚至提升分类器性能。对于输入的语句则根据领域词典中是否出现该字或词, 用二值特征表示, 总共提取了 64 个字特征, 119 个词特征, 7 个语义类特征。这里的实验评价指标使用意图分类的错误率(error rate)。意图分类性能如表 1 所示, 其中“字+语义类特征”得到的结果最好。

表 1 各个特征得到的错误率对比 %

特征	错误率	特征	错误率
字特征	1.52	字+语义类特征	0.20
词特征	1.08	词+语义类特征	0.49
语义类特征	1.52		

实验中需要使用分类的置信度分数作为衡量样本信息量的指标, 这里的置信度计算方法^[17]如下:

$$\text{confidence}(x_i) = \max_j p(c_j | x_i) \quad (1)$$

其中: x_i 表示样本, c_j 表示第 j 个类别, 置信度即选取样本在所有类别中的最大后验概率。本文的文本分类基于 LIBSVM (library for Support Vector Machines) (<http://www.csie.ntu.edu.tw/~cjlin/>) 开源工具包实现的。

3.2 结合主动学习和自训练的实验

主动学习和自训练选择数据的策略是不同的。主动学习每次选择分类器分类性能比较弱的样本进行人工标注, 加入到训练集中。而半监督学习每次选择分类性能比较强的样本直接使用机器标注加入到训练集中。主动学习平衡了分类器不擅长的那部分数据, 而半监督方法增加了分类器原本就擅长的样本, 两种方法都破坏了训练数据的分布, 但是两者结合起来却使训练数据的分布得到平衡, 因此这里结合了这两种方法, 实验结果如图 1 所示。其中: 被动学习每次随机选择 300 个样本进行人工标注; 主动学习采用基于动态样本池的方法, 样本池的大小选择为 300。首先以人工标注的 300 个样本作为种子数据, 从未标注的数据中随机选择 300 个样本填充样本池; 其次用当前的分类器对样本池中的数据进行分类, 选择置信度最小的 1/4 (即为 75 句) 数据进行人工标注; 然后加入到训练集中。丢弃样本池的其他数据。

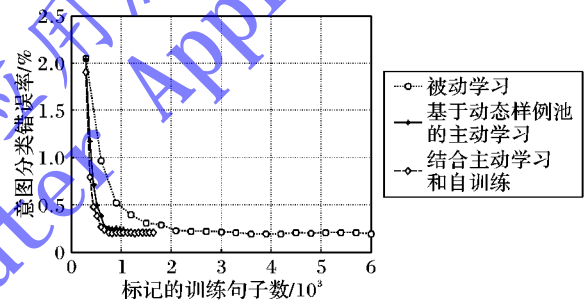


图 1 结合主动学习和自训练的意图分类曲线

结合主动学习和自训练的方法与基于动态样本池的方法类似, 唯一不同的是, 用种子数据训练的分类器对样本池中的数据进行分类后, 选择置信度大于阈值的样本池的 1/8 数据 (这里选取了 40) 进行机器标注加入到训练集中, 阈值设置为 0.9, 选择置信度最低的 35 句进行人工标注, 再加入到训练集中。实验结果显示: 达到错误率为 0.2%, 被动学习、主动学习以及结合主动学习和自训练 3 种方法需要人工标注的样本数分别为: 3300, 1125 和 510。由此可见, 结合主动学习和自训练的方法在达到同样性能的前提下, 人工标注样本是最少的。

3.3 协同训练实验

协同训练要求选取两个相互独立的特征子集, 而且每个单一的特征子集要求达到比较好的分类性能。根据表 1 的意图分类性能, 本文选择字特征和语义类特征作为两个特征子集。因为这两个特征集得到的分类性能相对较好, 而且又相互独立, 满足了协同训练的前提条件。

实验 1 协同训练的训练集选择初始训练种子数为 2000, 3000, 4000, 阈值设置为 0.9, 分别迭代 30 次, 重复进行了 10 次取平均得到的结果。图 2(a) 和 (b) 分别表示字特征和语义类特征在 3 种不同初始训练集下得到的协同训练的性能比较。从图 2 中的结果可以看到, 两种视角的训练集利用了对方的数据之后, 一个是性能没有明显的变化, 一个是性能有了下降。究其原因可能是训练数据中含有冗余数据, 因此考虑对数据去重后再进行协同训练。

实验 2 对 6000 个训练样本去重, 去重的步骤直接在

SVM 的特征文件上进行,得到了一个比较小的数据集,236 句,仍旧以字特征和语义类特征作为两个视角,初始种子数为 100,110,120,阈值设置为 0.9,迭代 9 次,实验重复 10 次取平均作为最终结果。图 3(a)、(b) 分别显示了字特征集和语义类特征集的协同训练性能,其中字特征集随着迭代次数的增加,分类错误率呈现下降的趋势,表明字特征集可以更好地利用新加入的训练样本,增强其泛化能力。而语义类特征在这 3 种不同种子个数的情况下,其分类的错误率有所上升。由于语义类特征集中的特征个数相对较少,因此容易过拟合,所以很难使其性能上升。以初始种子为 120 句为例,字特征子集的分类错误率在协同训练前后分别为 1.90% 和 1.41%,比单一使用字特征得到的结果 (1.52%) 分类错误率下降了 0.11%。表 2 中列出了初始种子数为 100,110 和 120 三种情况下,两个特征集合协同训练前后的性能对比。

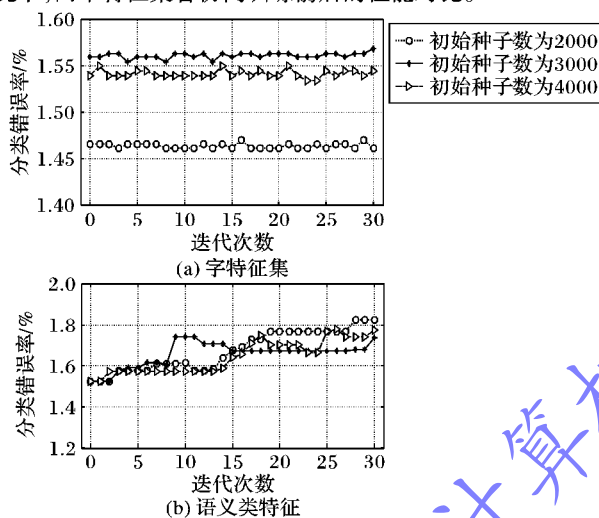


图2 协同训练测试结果

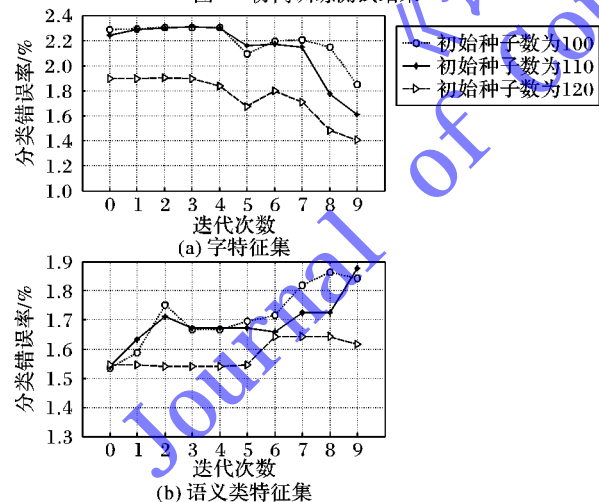


图3 去重后的协同训练测试结果

表2 去重后协同训练前后的最终分类错误率对比 %

特征集	协同训练前			协同训练后		
	初始种子 100	初始种子 110	初始种子 120	初始种子 100	初始种子 110	初始种子 120
字特征	2.29	2.24	1.90	1.85	1.61	1.41
语义类特征	1.54	1.54	1.55	1.84	1.88	1.62

3.4 实验结果分析

由以上两种弱监督训练结果,可以分析得到:结合主动学习和自训练方法因其对样本的选择比较严格,与主动学习和

被动学习相比较,其收敛速度都快。协同训练通过两个视角的特征子集互相提供标注数据,整个标注过程无需人工干预。这两种方法相比较而言,结合主动学习和自训练方法更加可靠,结果相对稳定,而协同训练则依赖于两个视角的特征子集的独立性,以及两个特征子集中特征的个数,只有在两个视角的特征子集独立而且两个特征子集中的特征个数相差不多的情况下,协同训练才会互相帮助对方提升分类性能。本实验中字特征个数是 64,语义类特征个数是 7,两者相差较大,导致字特征可以利用语义类特征子集标注的样本提升分类性能,而语义类特征不太容易利用字特征子集标注的样本提升性能。

4 结语

本文主要从中文口语理解的角度进行了弱监督训练的研究,以希望解决口语理解过程中人工标注数据的繁重劳动。本文首先介绍了口语理解的级联框架(即先进行命名关键语义概念识别,再进行意图识别),然后对意图识别进行了两种弱监督训练:一种是结合主动学习和自训练的方式;一种是协同训练的方式。其中结合主动学习和自训练的方法比被动学习和主动学习需要的人工标注数据最少,较好地减少了人工标注。而协同训练的前提是两个视角的特征集必须相互独立,而级联方式的关键语义概念识别恰好可以给意图识别提供语义类特征,采用字特征集和语义类特征集作为两个特征子集进行协同训练,通过实验在两个特征子集得到的结果显示:字特征集能够更好地利用新加入的训练样本,提高分类性能,而语义类特征则容易过拟合,导致性能下降。本文得到的结论是:主动学习和自训练相结合的方法应用范围更加广泛,不受分类器的限制,在对话系统开发初期可以利用该分类器降低人工标注的工作量。而协同训练方法的限制条件比较苛刻:首先两个独立的特征子集的寻找是很困难的,其次要求每个特征子集单独可以获得比较好的分类性能也是不容易做到的,因此在实际应用中,使用协同训练减少人工标注不太容易实现。但是如果满足了前提条件,则协同训练进行自动标注数据的效果还是比较好的,同时还要避免某个特征子集的分类器出现过拟合现象。

未来的工作方向打算进一步探索关键语义概念识别的弱监督训练,进一步减少口语理解过程中两个子任务的人工标注。

参考文献:

- [1] TORRESANI L. Weakly supervised learning [M]// Computer Vision: A Reference Guide. Berlin: Springer, 2014: 883-885.
- [2] TUR G, HAKKANI-TUR D, SCHAPIRE R. Combining active and semi-supervised learning for spoken language understanding [J]. Speech Communication, 2005, 45(2): 171-186.
- [3] LIU K, QIAN X, WANG Z. Survey on active learning algorithms [J]. Computer Engineering and Applications, 2012, 48(34): 1-4. (刘康, 钱旭, 王自强. 主动学习算法综述[J]. 计算机工程与应用, 2012, 48(34): 1-4.)
- [4] TSUTAKA T, SHINODA K. Acoustic model training using committee-based active and semi-supervised learning for speech recognition [C]// APSIPA ASC 2012: 2012 Asia-Pacific Signal & Information Processing Association Annual Summit and Conference. Piscataway: IEEE, 2012: 1-4.
- [5] ZHAO W, MA H, LI Z, et al. Efficiently active learning for semi-supervised document clustering [J]. Journal of Software, 2012, 23(6): 1486-1499. (赵卫中, 马慧芳, 李志清, 等. 一种结合主动学习的半监督文档聚类算法[J]. 软件学报, 2012, 23(6): 1486-1499.)

(下转第 1974 页)

595. (何鹏, 阎兴顺, 侍洪波. 一种快速自适应蜂群算法及其应用[J]. 华东理工大学学报: 自然科学版, 2013, 39(5): 588-595.)
- [5] SABAT S L, UDGATA S K, ABRAHAM A. Artificial bee colony algorithm for small signal model parameter extraction of MESFET [J]. Engineering Applications of Artificial Intelligence, 2010, 23(5): 689-694.
- [6] KANG F, LI J, MA Z. Rosenbrock artificial bee colony algorithm for accurate global optimization of numerical functions [J]. Information Sciences, 2011, 181(16): 3508-3531.
- [7] WANG H, LIU J, WANG Q. Modified artificial bee colony algorithm for numerical function optimization [J]. Computer Engineering and Applications, 2012, 48(19): 36-39. (王慧颖, 刘建军, 王全洲. 改进的人工蜂群算法在函数优化问题中的应用[J]. 计算机工程与应用, 2012, 48(19): 36-39.)
- [8] YAN X, ZHU Y, ZOU W, *et al.* A new approach for data clustering using hybrid artificial bee colony algorithm [J]. Neurocomputing, 2012, 97(15): 241-250.
- [9] KAO Y, ZAHARA E, KAO I. A hybridized approach to data clustering [J]. Expert Systems with Applications, 2008, 34(3): 1754-1762.
- [10] KARABOGA D, AKAY B, OZTURK C. Artificial Bee Colony (ABC) optimization algorithm for training feed-forward neural network [C]// MDAI07: Proceedings of the 4th International Conference on Modeling Decisions for Artificial Intelligence. Berlin: Springer, 2007: 318-319.
- [11] GAO F, QI Y, YIN Q, *et al.* Online synchronization of uncertain chaotic systems by artificial bee colony algorithm in a non-Lyapunov way [C]// CISE 2010: Proceedings of the 2010 International Conference on Computational Intelligence and Software Engineering. Piscataway: IEEE, 2010: 1-4.
- [12] ZHANG Y, WU L. Face pose estimation by chaotic artificial bee colony [J]. International Journal of Digital Content Technology and Its Applications, 2011, 5(2): 55-63.
- [13] KARABOGA D, AKAY B. A comparative study of artificial bee colony algorithm [J]. Applied Mathematics and Computation, 2009, 241(1): 108-132.
- [14] KARABOGA D, BASTURK B. On the performance of Artificial Bee Colony (ABC) algorithm [J]. Applied Soft Computing, 2008, 8(1): 687-697.
- [15] ALIZADEGAN A, ASADY B, AHMADPOUR M. Two modified versions of artificial bee colony algorithm [J]. Applied Mathematics and Computation, 2013, 225(1): 601-609.
- [16] GAO W, LIU S. A modified artificial bee colony algorithm [J]. Computers and Operations Research, 2012, 39(3): 687-697.
- [17] LI B, ZENG J. Comparison and analysis of the selection mechanism in the artificial bee colony algorithm [C]// HIS 2009: Proceedings of the 2009 International Conference on Hybrid Intelligence System. Piscataway: IEEE, 2009: 411-416.
- [18] XIANG W, MA S. Artificial bee colony based on reverse selection of roulette [J]. Application Research of Computers, 2013, 30(1): 86-89. (向万里, 万寿峰. 基于轮盘赌反向选择机制的蜂群优化算法[J]. 计算机应用研究, 2013, 30(1): 86-89.)
- [19] DING H, FENG Q. Artificial bee colony algorithm based on Boltzmann selection policy [J]. Computer Engineering and Applications, 2009, 45(31): 53-55. (丁海军, 冯庆娟. 基于 Boltzmann 选择策略的人工蜂群算法[J]. 计算机工程与应用, 2009, 45(31): 53-55.)
- [20] BANHARNSAKUN A, ACHALAKUL T, SIRINAOVAKUL B. The best-so-far selection in artificial bee colony algorithm [J]. Applied Soft Computing, 2011, 11(2): 2888-2901.

(上接第1968页)

- [6] VIJAYALAKSHMI T, THUTHI SARABAI D. Aspect based topic and opinion mining [J]. International Journal of Computer Trends and Technology, 2014, 15(4): 168-173.
- [7] da SILVA A T, FALCÃO A X, MACALHÃES L P. Active learning paradigms for CBIR systems based on optimum-path forest classification [J]. Pattern Recognition, 2011, 44(12): 2971-2978.
- [8] YAO T. Active learning based visual scene understanding [D]. Hangzhou: Zhejiang University, 2011. (姚拓中. 结合主动学习的视觉场景理解[D]. 杭州: 浙江大学, 2011.)
- [9] SHAHSHAHANI B, LANDGREBE D. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon [J]. IEEE Transactions on Geoscience and Remote Sensing, 1994, 32(5): 1087-1095.
- [10] NIGAM K, McCALLUM A K, THRUN S, *et al.* Text classification from labeled and unlabeled documents using EM [J]. Machine Learning, 2000, 39(2/3): 103-134.
- [11] SUBRAMANYA A, TALUKDAR P P. Graph-based semi-supervised learning [J]. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2014, 8(4): 1-125.
- [12] McCALLUMZY A K, NIGAMY K. Employing EM and pool-based active learning for text classification [C]// Proceedings of the Fifteenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers, 1998: 350-358.
- [13] CHEN R, CAO Y, SUN H. Multi-class image classification with active learning and semi-supervised learning [J]. Acta Automatica Sinica, 2011, 37(8): 954-962. (陈荣, 曹永锋, 孙洪. 基于主动学习和半监督学习的多类图像分类[J]. 自动化学报, 2011, 37(8): 954-962.)
- [14] SUN F, XU M, JIANG X. Robust multi-label image classification with semi-supervised learning and active learning [C]// MMM 2015: Proceedings of the 21st International Conference on MultiMedia Modeling, LNCS 8936. Berlin: Springer, 2015: 512-523.
- [15] BLUM A, MITCHELL T. Combining labeled and unlabeled data with co-training [C]// COLT'98: Proceedings of the Eleventh Annual Conference on Computational Learning Theory. New York: ACM, 1998: 92-100.
- [16] DARNSTÄDT M, SIMON H U, SZÖRÉNYI B. Supervised learning and co-training [C]// ALT 2011: Proceedings of 22nd International Conference on Algorithmic Learning Theory, LNCS 6925. Berlin: Springer, 2014: 425-439.
- [17] WU W. Robust spoken language understanding across domains and languages [D]. Shanghai: Shanghai Jiao Tong University, 2007. (吴尉林. 可移植的稳健口语理解方法研究[D]. 上海: 上海交通大学, 2007.)