

DOI:10.11992/tis.201607023
网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.tp.20170405.1519.002.html>

认知视角下的舆论观点句情感计算

张冬慧¹,程显毅²

(1.北京信息科技大学 计算中心,北京 100192; 2.南通大学 计算机科学与技术学院,江苏 南通 226019)

摘 要:针对目前观点分析方法局限于传统的文本分析技术,只能将舆论观点句分为肯定和否定两极或确定每一极的程度(粗粒度),不能进一步给出舆论观点句是积极的还是消极的程度的问题。本文从认知学角度研究细粒度语义情感计算框架。提出了一种舆情观点句的定量分析方法,该方法将对于某话题的文本集合作为输入,输出一个实数表示文本中所表达观点的能量。本文在 NLPPIR 共享平台上进行了相关实验,给出了粗粒度情感和细粒度情感对观点句识别的对比实验,实验表明,两种方法对观点句的识别性能相差不大;对非观点句细粒度方法好于粗粒度方法。
关键词:认知;情感计算;舆论观点句;观点的能量;消极;积极;语义;粗粒度;细粒度
中图分类号:TP391.1 **文献标志码:**A **文章编号:**1673-4785(2017)04-0498-06

中文引用格式:张冬慧,程显毅.认知视角下的舆论观点句情感计算[J]. 智能系统学报, 2017, 12(4): 498-503.
英文引用格式:ZHANG Donghui, CHENG Xianyi. Research on computation of affect in public opinion sentences from the cognition viewpoint[J]. CAAI transactions on intelligent systems, 2017, 12(4): 498-503.

Research on computation of affect in public opinion sentences
from the cognition viewpoint

ZHANG Donghui¹, CHENG Xianyi²

(1. Computing Center, Beijing Information Science & Technology University, Beijing 100192, China; 2. School of Computer Science and Technology, Nantong University, Nantong 226019, China)

Abstract:The current viewpoint analysis method is limited to the traditional text analysis technology, whereby a public opinion sentence can only be divided into positive and negative poles and the extent of each pole (coarse-grained) determined. It is difficult to determine whether a public opinion sentence is active or passive. In this paper, we discuss a computation framework for fine-grained semantic sentiments from the cognitive science viewpoint and propose a quantitative analysis method for public opinion sentences. This method takes the text collection of some topic as input and uses a real number to represent the energy of a viewpoint in the text. We conducted an experiment using the Natural Language Processing and Information Retrieval (NLPPIR) sharing platform and a contrasting experiment with respect to view recognition by comparing coarse-grained and fine-grained affects. The experimental results show that the two methods have the same recognition performance regarding sentence viewpoints. For no-opinion sentences, the fine-grained method performs better than the coarse-grained method.
Keywords: cognitive; sentiment computer; public opinion sentence; energy of view; active; negative; semantic; coarse-grained; fine granularity

当前,计算文本情感主要采用情感词加权、语义模式等技术^[1]。语义模式技术,首先提取句子情感特征,然后确定句子的情感色彩,但是语义模式方法对于没有明显情感词的句子,无法判断句子的情感倾向程度。

文献[2]从认知学角度出发,考虑到情绪和认知的关系,提出一种文本情感计算的认知模型,该模型核心是引入了情感图式,但情感图式的构造主要以手工标注为主,缺少推理功能。任巨伟等^[3]在文献[2]的基础上提出了一种新的情感图式构造方

收稿日期:2016-07-23. 网络出版日期:2017-04-05.
基金项目:国家自然科学基金项目(61340037).
通信作者:程显毅. E-mail:xycheng@ntu.edu.cn.

法,利用上下位关系并结合认知语境扩充图式,增加图式的推理功能。但文献[3]没有区分情感和情绪,不利于舆论观点句的情感计算。

在观点句的情感计算中,仅仅计算语句的肯定或否定还不够,需要知道肯定或否定的趋势,即是积极的还是消极的^[4]。

本文首先依据认知学原理,将传统的情感细分为:情感、情绪和感受3类。区分后的情感称为细粒度情感,区分前的情感称为粗粒度情感;然后再将细粒度情感细分为舆论观点和意见评价两个子类,最后关注舆论观点的情感计算。该方法输入的是某一话题的文本集合,输出是 $[-1,1]$ 的实数,其中正数表示积极的态度,负数表示消极的态度,绝对值越大表示态度越强烈。研究的动机有两个:1)观点态度识别在舆情分析中需求迫切;2)从细粒度语义角度进行情感计算。

1 感觉、感受、情绪和情感

1.1 感觉

感觉指人脑对直接作用于感觉器官的客观事物的某些属性的反应。感觉是认识活动的开端,是人认识客观事物的第1步。其要点是:1)客观事物直接作用于感觉器官;2)是对客观事物个别属性的反应。如我的身高属于我这个人的个别属性。我衣服的颜色属于衣服的个别属性。颜色这个个别属性直接作用于你的感觉器官,就是眼睛,具体说是视网膜上的视细胞,接收信息的器官叫做感受器。感受器接受信息后,经传入神经,传输到大脑中枢,与大脑中枢已有的颜色的概念结合,得出这是红色。这就是感觉^[5]。

常用描述感觉的词包括:冷、热、亮、暗、臭、香、痛、痒等。

1.2 感受

感觉产生的同时,还有一定的感受。你觉得我的衣服怎么样?漂亮还是难看?这就是你产生的感受。所有的人产生一样的感受吗?有人就觉得这件衣服不好看不适合我,有人就觉得我穿红颜色很好看。如果我们的感觉器官神经系统没有问题,所产生感觉的结果基本会相同,感受却可能不同^[6]。

在感觉的基础上加上了主观判断而产生的心理现象称为感受。对同样的客观事物,不同个体可能会产生不同的感受,即使是同一个体,在不同的时候对同一事物也会产生不同的感受^[7]。当客观事物直接作用于个体的感觉器官时,首先产生的是感觉,之后才有了感受,一般情况下感觉大体是相

同的,感受却大不相同。比如春天看到田野到处是盛开的油菜花,有人觉得漂亮,想多看几眼,有人没觉得有什么特别,这就是感受。

常用描述感受的词包括:尊重、敬仰、欣赏、鄙视、舒服、满意、幸福、难过、安全、快乐等。

1.3 情绪

伴随着感受又会出现高兴或厌烦的体验,这时又产生了另一种心理现象,就是情绪,是感受的外部表现。情绪没有对错,不同的感受就会有不同的情绪。感受是内在的,通过情绪表现出来,让人能够觉察到。人的面部表情,说话的音调、节奏,肢体,还有走路的步伐等都渗透着情绪。如:当人愤怒的时候眉毛会立起来,恐惧时会大喊。于是通过人的外部的表情表现,可以推测人的内在感受。面对客观事物时,首先产生的是感觉,然后是感受,最后发生的是情绪。但通常情况下是很难区分的。如:我看到老公醉醺醺的回来,脸立即拉下来了,直接表达了情绪。其实是先看到他走路歪歪扭扭,然后听到他说话结结巴巴,又闻到浓浓的酒味,这些都是感觉。心里想,又喝醉了,几乎每天都是这样,我非常不满,这是感受。我很生气,甚至愤怒,这是情绪。外部表现是拉着脸,瞪着他或不理他。往往直接暴露的就是情绪。三者是密切联系在一起的,甚至有时感到同时发生。而情绪的发生又会影响感官的工作状态。如:面对突发事件时,有人呆若木鸡,有人急中生智。因此说情绪是能量,当情绪为我们服务,有助于我们达成目标时,情绪转化的能量分为正能量和负能量。表1列举了部分描述情绪的词^[8]。

表 1 部分描述情绪的词
Table 1 A part of describes emotional words

类别	描述词
挫折	倒霉、无奈、泄气、无助、失望、颓丧
生气	震怒、恼怒、大怒、气愤、激怒、发狂、
憎恨	轻视、轻蔑、痛恨、厌恶、可恶、讥讽、排拒
难过	悲伤、伤心、沮丧、悲哀
快乐	狂喜、高兴、欣喜、兴奋
激动	悲喜交集、义愤填膺、百感交集、情不自禁
害怕	恐慌、可怕、不安、心乱、慌乱、惊恐、焦虑
惊讶	惊喜、震惊、叹为观止、吃惊、目瞪口呆
忧郁	闷闷、烦躁、郁闷、忧愁、苦闷、抑郁、沮丧
惭愧	可惜、尴尬、遗憾、羞耻、愧疚、羞愧、后悔
紧张	彷徨、着急、焦急
寂寞	无聊、孤独、孤单、寂寞、空虚

1.4 情感

1.4.1 情绪与情感的区别

从它们产生的基础和特征表现上来看,情绪与情感有所区别^[9]。

1)情绪。情绪出现较早,取决于人的生理需要,是人和动物共有的特征。其特征是:情境性、暂时性、冲动性、明显的外部表现。

2)情感。情感出现晚些,取决于人的社会需要,是只有人才有的特征。情感的主要特征是:深刻性、稳定性。情绪常随着环境的变化而变化。情感多指内心的体验,不轻易表现流露出来。

3)情绪与情感的联系。情绪和情感尽管有所不同,但却是密不可分。情感是在多次情绪体验的基础上形成的,并通过情绪表现出来;反过来,情绪的表现受到情感的制约。所以,情绪是情感的基础,情感是情绪的深化。

1.4.2 描述情感的词

表 2 列出了部分描述情感的词语^[10]。

表 2 部分描述情感的词

Table 2 A part of describes sentiment words

类别	描述词
主张	声称、断言、要求
支持	提倡、拥护、倡议、支持、赞成、同意
反对	否认、与……相背、与……矛盾、反驳
好处	有利于、积极、有益于、值得
弊端	损失、不利、有害、消极、短处、缺陷
重要	基本、本质、内在、精华、必不可少
普遍	广泛、流行、通用
争论	争议、异议、余地
假设	认为、推向、猜想、料想、预报、预言
趋势	走向、带来、引起、导致、结果是
探索	探测、探讨、检查、调查、确定、鉴定

2 舆论观点句

2.1 主观情感与客观情感

情感是人的本性,但它是主观和客观对立统一的产物。客观事物是情感产生的源泉,人的主观需要是情感产生的内在原因。情感作为信息加工、调节主体行为的一种方式,成为主体反映客体的特殊形式,对人的思维方式发生影响。它影响着思维的波动性、非逻辑性、选择性和指向性。它可以转化为个体的思维动机。所以,情感与“主观性”有关也与“客观性”有关。

所谓主观性情感就是自己由自己的认识、理解、看法所引发的心理反应。就是没有什么道理可言,自然而然地就产生这种感情了,不受自我控制^[11]。

被观察事物的性质和规律不随观察者的意愿而改变的情感称为客观情感。如“股票涨了”就是客观情感句^[12]。

2.2 舆论观点句

网络的匿名性、开放性、平等性、交互性等特点不可避免地会出现一些不和谐“杂音”。因为所有网民都可以自由发表自己的信息或观点,为了使发布的信息得到跟帖或采纳,就会突出感性色彩以产生感染力和影响力。因此,在网络空间,不加证实随意发布信息,或有意制造假消息是不可避免的,以至于造成网络欺骗,一些网民哗众取宠、使得网上出现非理性的语言人身攻击、谩骂等。所以研究舆论观点句判别方法成为网络信息安全迫切需要解决的问题。

由于我们对情感加以限制,排除了感觉、感受和情绪的词语,所以舆论观点句就是包含主观情感词语的句子。

3 舆论观点句情感计算

3.1 主观特征词本体构建

人之所以比机器聪明,一个很重要的因素就是人有解决问题的领域常识或语境知识(本体),在解决问题中有着特殊的作用。

本文使用的领域本体包括:褒贬词典^[13-14]、Hownet 情感词典^[15]、台湾大学 NTUSD 情感词典^[9]等词典资源。因为粗粒度情感词大部分和舆论观点的判别关系不大,首先去除感受词和情绪词,然后进行筛选和整理,最后确定 426 个静态情感词,其中正向情感词 287 个,负向情感词 139 个,在此基础上,增加程度副词、否定词、连词、动词、代词、叹词、疑问词、敏感词、网络用语集,扩展后的主观特征词汇本体并不是封闭的,在系统工作时,用 GATE^[16-17]向本体中动态添加新的主观特征词。

主观特征词汇本体中概念层次关系如图 1 所示(片段)。矩形框代表本体类别,圆框代表本体实例,每个实例附带的数字是该特征在主观性判别所起的作用,我们称之为主观性权重,取值在 $[-1,1]$ 。取负表示有利于持反对观点的主观句判别,取正表示有利于持支持观点的主观句判别。本体共收录 18 个否定词、219 个程度副词、68 个连词、21 个情感动词、17 个叹词、25 个疑问词、210 个敏感词、185 个网络流行词。

3.2 主观句判别算法

- 1) 对语料自动分词和词性标注算法^[8];
- 2) 对词性标注后的句子,提取全部连续的 2-gram 模式;
- 3) 根据 2) 提取出的 2-gram 模式计算 2-gram 模式主观性权重 $\text{Onto}(t)$ 。
- 4) 根据式(1)计算句子的主观性权重 $w(S)$:

$$w(S) = \frac{o}{n} \sqrt{\frac{\prod_{t \in 2_pos} \text{Onto}(t)}{1}}$$

(1)

式中: $w(S)$ 为句子 S 的主观性权重; o 为句子 S 的主观性 2-gram 模式个数, n 为 S 的 2-gram 模式个数,式(1)分子是句子主观性系数,分母是对在句子出现的本体实例权重之积的归一化。

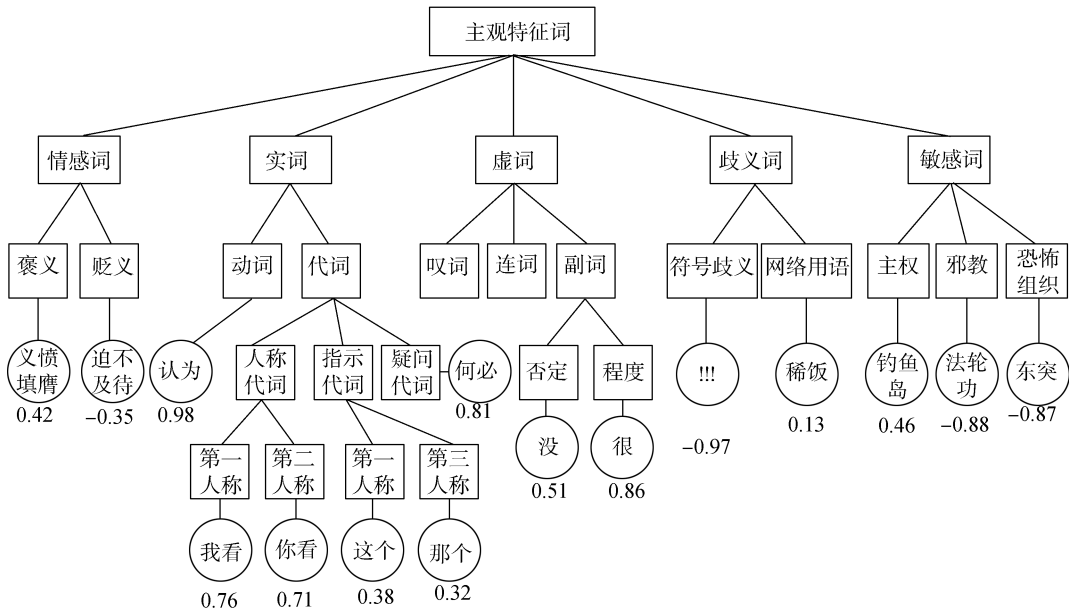


图 1 主观特征词汇本体片段
Fig.1 The fragment of subjective feature words ontology

如果 $w(S) > \alpha$, 则识别句子 S 为主观性句子, 阈值 α 是一个经验常数。

3.3 实验结果分析

- 1) 实验数据
- 为了推动主观性判别的研究,北京理工大学张华平博士曾在 NLPPIR 共享平台分享了微博语料^[18],此外 NLP&CC 2012 评测集等^[19],极大地方便了主观性判别的研究分析。这些语料评价的力度较粗,不适合本文的实验。所以,本文的实验语料,来自采集 2016 年 3 月 10 日~6 月 10 日期间新浪微博热门话题数据集,共包含 28 个热门话题,其中每个话题约有 1 000 条左右的微博。采用投票策略对训练语料进行标注,首先由 4 人分别对同一个句子进行标注,然后由第 5 个人对标注结果进行核查,剔除意见分歧较大、主观性不易确定的句子。为使训练语料更加平衡,本文最终收集的训练语料共包含非观点句 500 条,观点句 480 条,其中正能量观点句 250 条,负能量观点句 230 条。对语料 2/3 作训练,1/3 作测试。
- 2) 观点句分类性能分析

表 3 给出了粗粒度情感和细粒度情感对观点句识别的对比实验,粗粒度情感的观点句识别采用文献^[20]抽取主观模式的方法。

结果表明,两种方法对观点句的识别性能相差不大,并且准确率和召回率不尽人意,可能的原因是本体的构造还有待改进,也可能是人为对语料的标注有问题。对非观点句细粒度方法好于粗粒度方法,原因是有一些非观点句包含了观点句的模式。

表 3 粗粒度情感和细粒度情感对观点句识别的对比实验
Table 3 Experiments of view recognition to compare coarse-grained affection with fine-grained affection

方法	准确率		召回率	
	非观点句	观点句	非观点句	观点句
粗粒度	98.8	87.5	76.2	84.1
细粒度	100	87.8	98.5	82.5

- 3) 观点能量值计算性能分析
- 根据式(1),每个句子都映射为 $[-1,1]$ 的一个实数。实验的目的验证通过定量的方法来对文本

分类,定量的准则不同可实现深层次语义分类。由于还没有发现文本定量分类的报道,所以无法实现对比实验,图 2 的实验结果也只是个实验型的,希望能为文本分类提供一种新的途径。

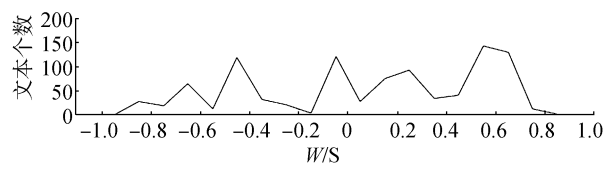


图 2 文本能量值分布图

Fig.2 Chart of text energy distribution

图 2 横轴为能量值,能量值为 0 表示该语句识别为非观点句,能量值为负表示该语句识别为负能量观点句,能量值为正表示该语句识别为正能量观点句。纵轴为对应能量值语句个数。将其转换为混合矩阵(表 4)。

从表 4 可知,正能量识别准确率为 32.3%,正能量识别召回率 54.4%,负能量识别准确率为 57.0%,负能量识别召回率 73.9%,非观点句识别准确率为 68.3%,非观点句识别召回率 16.8%。

表 4 细粒度情感对观点句识别的混合矩阵
Table 4 The mixed matrix of fine-grained emotion to view recognition

正能量观点句	负能量观点句	非观点句
136	88	26
47	170	13
376	40	84

4 结束语

针对现有的主观句判别多采用基于统计和关键词的方法,导致结果不准确的问题,从认知角度出发,将情感分为细粒度情感、感受和情绪,在此基础上构建了主观特征词汇本体,给出句子级深度语义情感分类方法,下一步的研究就是改进能量值计算模型,提高深度语义分类性能。

参考文献:

[1] MCKEOWN G, VALSTAR M F, COWIE R, et al. The SEMAINE corpus of emotionally colored character interactions[C]// Proceedings of IEEE International Conference on Multimedia and Expo, ICME 2010. IEEE Computer Society, 2010: 1079-1084.

[2] 徐琳宏,林鸿飞.认知视角下的文本情感计算[J]. 计算机科学, 2010, 37(12): 182-185.

XU Linhong, LIN Hongfei. Text affective computing from

cognitive perspective[J]. Computer science, 2010, 37(12): 182-185.

[3] 任巨伟,杨亮,林鸿飞.情感图式构造及其在文本情感计算中的应用[J].江西师范大学学报:自然科学版, 2013, 37(2): 130-136.

REN Juwei, YANG Liang, LIN Hongfei. The construction of affective schemata and its application in text affective computing[J]. Journal of Jiangxi normal university: natural science, 2013, 37(2): 130-136.

[4] PETRANTONAKIS P C, HADJILEONTIADIS L J. A novel emotion elicitation index using frontal brain asymmetry for enhanced eeg-based emotion recognition[J]. IEEE transactions on information technology in biomedicine, 2011, 15(5): 737-746.

[5] 宋静静. 中文短文本情感倾向性分析研究[D].重庆:重庆理工大学, 2013.

SONG Jingjing. Research on Chinese short-text sentiment analysis [D]. Chongqing: Chongqing University of Technology, 2013.

[6] 程显毅,刘颖.基于知识图的观点句识别算法研究[J]. 计算机科学, 2015, 42(6): 123-129.

CHENG Xianyi, LIU Ying. Research on algorithm of perspective sentence identification based on knowledge map [J]. Computer science, 2015, 42(6): 123-129.

[7] 蔡艳婧,程晓红,程显毅.网络敏感信息动态特征的抽取方法[J].常州大学学报, 2014, 16(4): 80-86.

CAI Yanjing, CHENG Xiaohong, CHENG Xianyi. Research on algorithm of network sensitive information features extracting [J]. Journal of changzhou university: natural science edition, 2014, 16(4): 80-86.

[8] 王志良.人工心理与人工情感[J].智能系统学报, 2006, 1(1): 38-44.

WANG Zhiliang. Artificial psychology and artificial emotion [J]. CAAI transactions on intelligent systems, 2006, 1(1): 38-44.

[9] 齐鑫.网络民意对我国政府决策的影响[D].沈阳:东北大学, 2010.

QI Xin. The influence of net citizen on government decision-making in China[D]. Shenyang: Northeastern University, 2010.

[10] 史杨.网络舆情与公共政策议程的设置[J].云南电大学报, 2011, 13(3): 55-59.

SHI Yang. Setting of internet public opinion and public policy agenda [J]. Journal of Yunnan RTV university, 2011, 13(3): 55-59.

[11] 孙浩博,侯军岐.论我国互联网种业发展[J].价值工程, 2016, 37(9): 327-329.

SUN Haobo, HOU Junqi. On the development of China's seed industry based on internet [J]. Value engineering,

- 2016, 37(9): 327-329.
- [12] 高云棋. 基于主题模型的舆情分析子系统研究与设计[D]. 成都:电子科技大学, 2013.
- GAO Yunqi. Research and development of opinion mining sub-system based on topic model[D]. Chengdu: University of Electronic Science and Technology of China, 2013.
- [13] 史继林, 朱英贵. 褒义词词典[M]. 成都: 四川辞书出版社, 2006: 23-28.
- SHI Jilin, ZHU Yinggui. The commendatory word dictionary[M]. ChengDu: Sichuan publishers of Lexicographical, 2006: 23-28.
- [14] 杨玲, 朱英贵. 贬义词词典[M]. 成都: 四川辞书出版社, 2006: 15-35.
- YANG Ling, Zhu Yinggui. Derogatory term dictionary[M]. Sichuan publishers of Lexicographical, 2006: 15-35.
- [15] 董振东. 知网的情感词典[EB/OL]. (2007-10-22) [2013-12-20]. http://www.keenage.com/html/c_bulletin_2007.htm.
- DONG Zhendong. Hownet dictionary [EB/OL]. (2007-10-22) [2013-12-20]. http://www.keenage.com/html/c_bulletin_2007.htm
- [16] rongzhe. NTUSD[EB/OL]. 台湾: 台北大学, (2013-8-2) [2014.11-21]. <http://www.datatang.com/data/11837.rongzhe.NTUSD> [EB/OL]. Taiwan: Taipei university, (2013-8-2) [2014.11-21]. <http://www.datatang.com/data/11837>.
- [17] 王晓东, 王娟, 张征. 基于情感词汇本体的主观性句子倾向性计算[J]. 计算机应用, 2012, 32(6): 1678-1681.
- WANG Xiaodong, WANG Juan, ZHANG Zheng. Computation on orientation for subjective sentence based on sentiment words ontology[J]. Journal of computer applications, 2012, 32(6): 1678-1681.
- [18] 张华平. NLPiR 微博关注关系语料库 1000 万条[EB/OL]. (2013-10-23) [2014-11-2]. <http://www.datatang.com/data/14350>.
- ZHANG Huaping. NLPiR Weibo focused on relationship between corpus of 10 million, [EB/OL]. (2013-10-23) [2014-11-2]. <http://www.datatang.com/data/14350>.
- [19] 中国计算机学会. 中文微博情感分析评测结果[EB/OL]. 北京: 北京大学. (2012-11-2) [2014-9-22]. http://tcci.ccf.org.cn/conference/2012/pages/page04_evares.html.
- CCF. Chinese microblog sentiment analysis evaluation results[EB/OL]. Peiking: Peiking university. (2012-11-2) [2014-9-22]. http://tcci.ccf.org.cn/conference/2012/pages/page04_evares.html.
- [20] 林慧恩, 林世平. 中文情感倾向分析中主观句子抽取方法的研究[C]//全国第20届计算机技术与应用学术会议(CACIS·2009)暨全国第1届安全关键技术与应用学术会议论文集(上册). 上海, 2009: 379-383.
- LIN Huien, Lin Shiping. Research on extracting subjective sentence in chinese sentiment orientation analysis [C]//Conference CACIS · 2009. Shanghai, China, 2009: 379-383.

作者简介:



张冬慧, 女, 1969年生, 博士, 主要研究方向为自然语言处理、计算机网络教育应用、知识工程。参与出版教材2部, 发表学术论文5篇。



程显毅, 男, 1956年生, 教授, 博士, 主要研究方向为知识工程、大数据应用、自然语言处理。主持国家自然科学基金2项、江苏省重点科技攻关项目1项、省部级项目6项。获省优秀教学成果一等奖1项, 二等奖1项。出版专著5部, 教材3部, 发表学术论文100余篇。