

基于情感词典扩展技术的网络舆情倾向性分析

杨超^{2,3},冯时²,王大玲^{1,2},杨楠²,于戈^{1,2}

¹(东北大学医学影像计算教育部重点实验室,辽宁沈阳110004)

²(东北大学信息科学与工程学院,辽宁沈阳110004)

³(河北省电力研究院,河北石家庄050021)

E-mail: yangchaoneucs@ sina. com

摘要:随着Web2.0时代的到来,网络已逐渐成为反映社会舆情的重要载体之一,网络舆情发现及网民的观点和倾向性挖掘也成为新的研究热点,但是目前尚无有效反应网民对热点事件或话题总体态度的舆情分析系统.本文针对网民关于话题评论简单、数目众多的特点,应用HowNet和NTUSD两种资源对现有情感词典进行扩展,建立了一个新的、具有倾向程度的情感词典.基于扩展的情感词典,开发了一个半自动化网络舆情分析系统.该系统能够为用户提供更加细致、准确的评论倾向性分析结果

关键词:观点挖掘;情感词典;倾向性分析;网络舆情

中图分类号: TP311

文献标识码: A

文章编号: 1000-1220(2010)04-0691-05

Analysis on Web Public Opinion Orientation Based on Extending Sentiment Lexicon

YANG Chao^{2,3}, FENG Shi², WANG Da-ling^{1,2}, YANG Nan², YU Ge^{1,2}

¹(Key Laboratory of Medical Image Computing, Northeastern University, Ministry of Education, Shenyang 110004, China)

²(School of Information Science and Engineering, Northeastern University, Shenyang 110004, China)

³(Hebei Electric Power Academe, Shijiazhuang 050021, China)

Abstract: As the Web 2.0, network becomes one of the important medium for reflecting public opinions, finding and mining public opinion orientation become an issue. But till now, no effective opinion monitoring system of reflecting the total orientation of the netizens on some hot events or topics has been proposed. In this paper, according to the characteristics of simplicity and large amount of opinions on some topics, an existing sentiment words lexicon is extended using HowNet and NTUSD, and a new sentiment lexicon with sentiment orientation extent is built. Based on the extended sentiment lexicon, a semi-automatic web public opinion analysis system is proposed, which can provide users more detail and precise opinion orientation analysis results.

Key words: opinion mining; sentiment lexicon; orientation analysis; web public opinions

1 引言

近年来,对于描述非事实的主观性文本处理方面的研究越来越多^[1,2],主要特点是分析和处理个人、群体等所表达的意见、情感和态度等.观点挖掘已经成为数据挖掘中的一个重要分支,它可以面向许多领域,如:市场分析、市场预测、民意调查以及信息监控等等.但是,目前大多数的研究工作主要面向商品的评论和反馈^[3,4],涉及网络舆情、特别是基于观点分析的网络舆情的研究相对较少.

随着Web2.0时代的到来,网络成了反映社会舆情的重要载体之一,越来越多的人通过博客、论坛以及网站留言板发表自己对热点事件的观点和看法.对于某个热点事件,如果将其有关的网络舆情信息加以汇总并且进行分析,就可以反应出对于这个事件民众所持有的态度和观点的倾向性.这种汇总的网络舆情,可以有效地帮助相关政府职能部门了解民意,进而做出及时的反馈.本文应用观点挖掘技术通过对新闻网页的评论进行收集并进行分析,将网民的评论汇总成肯定、

否定和中性三类,取得了良好的效果.

本文其余部分组织结构如下:第2部分简单介绍相关工作;第3部分着重介绍情感词典的扩展及面向网评的情感倾向性计算方法;第4部分介绍网络舆情倾向性分析系统的构建及相关功能;第5部分给出实验结果及分析;最后在第六部分对本文工作进行总结.

2 相关工作

分析评论中文本的情感倾向性是观点挖掘的主要任务之一,目前倾向性的分析主要针对词汇、句子和篇章三个层面进行分析.词汇的情感分析目前主要有三种方法,一种是基于WordNet^[5]和HowNet^[6]这样的知识库,首先选择两组具有明显正向和负向极性的词语作为种子词,对于一个情感倾向未知的词,计算这个词与两组种子词的相似度,与正向种子词组相似度高则判定为正面倾向,反之则判定为负面倾向.词汇倾向性分析的另一方法是无监督学习方法^[7],这种方法同

收稿日期:2008-12-17 收修修改稿日期:2009-03-03 基金项目:国家自然科学基金项目(60973019)资助. 作者简介:杨超,男,1984年生,硕士研究生,研究方向为文本挖掘;冯时,男,1981年生,博士研究生,研究方向为文本挖掘;王大玲,女,1962年生,博士,教授,博士生导师,研究方向为Web挖掘.

样需要先确定两组等量具有明显倾向性的种子词,一组是褒义种子词,一组是贬义种子词。对于一个新词,根据它和两组种子词的紧密程度对其倾向性进行推断,紧密程度的判断是根据词语在语料库中的共现频率,称为点态互信息量,将词语与褒义种子词的点态互信息量之和减去与各贬义种子词的互信息量之和,结果的正负即表示词语的倾向性,而且结果的大小还表示了倾向性的强度。这个方法的点态互信息量也可以通过使用搜索引擎来计算,其概率可以通过搜索引擎返回的 Hits 值占搜索引擎总的索引页面数的比例来计算,因此无需语料库。

在英文的应用系统的研究上,Bing Liu 等学者研究并开发了 Opinion Observer^[8],主要针对商品评论做了更深入的分析,突破了仅仅给出篇章总体倾向性的研究,研究了从同一类商品的多个评论中抽取主题的算法,对子主题倾向性分析,综合多个语篇的分析得出总结性的结果,具有比较实际的商用价值。这也是商品评论比较特别之处,同一类商品的子主题比较容易确定,比如数码相机的评论一般包含多个主要部件或属性的评论,尺寸大小、照片质量、电池寿命、相机重量等。

在汉语的应用系统的研究上,姚天昉等学者研究并开发了一个用于汉语汽车评论的观点挖掘系统^[4],该系统在电子公告板、门户网站的各大论坛上挖掘并概括顾客们对各种汽车品牌的各种不同性能指标或重要部件的评论和意见,并且判断这些意见的褒贬性以及强度,最后总结并得出可视化的结果。

观点挖掘和情感分析的研究工作还有很多,其中较成熟的研究成果大多面向商品评价,这里不再一一赘述。

3 面向网评的情感词典及情感倾向性计算

3.1 网民评论的特点

通过对新浪、搜狐等门户网站关于热点话题留言板的分析可见,绝大多数网民发表的留言简短,但倾向性明显。如:

(1) 多才多艺才能真正体现一个人的综合才华,有才艺的人比死读书的人更有前途,更受别人爱戴,对社会繁荣更为重要。一定要加分!

(2) 很不公平,考试考的是学习,而不是艺术院校招生。如果你认为孩子有特长,去考艺术院校呀!

这是两个网民在搜狐辩论区上关于“你认为特长生考试加分,这样公平吗?”中的留言^[9],代表了两种不同的观点。很明显,第一个表达肯定的观点,第二个是表达否定的观点。本文观察到网民在留言中,绝大部分的观点性很明确,情感词汇运用广泛,我们如果做到分词准确,并且情感词汇表建立的准确,能够取得很好的效果。

3.2 ICTCLAS 标注方法

本文采用的分词及词性标注工具是中国科学院计算技术研究所开发的汉语词法分析系统 ICTCLAS^[10],其主要功能包括中文分词,词性标注,命名实体识别,新词识别,同时支持用户词典。ICTCLAS2008 分词速度单机 996KB/s,分词精度 98.45%,API 不超过 200KB,各种词典数据压缩后不到 3M。我们针对网络上的新兴词语做了网络词汇词典,用以增加分

词的准确性。

针对 3.1 小节提到的两条评论,ICTCLAS 的分词标注结果如下:

(1) 多才多艺/vl 才能/n 真正/d 体现/v 一个/mq 人/n 的/udel 综合/vn 才华/n,/wd 有/vyou 才艺/ng 的/udel 人/n 比/p 死/v 读书/vi 的/udel 人/n 更/d 有/vyou 前途/n,/wd 更/d 受/v 别人/rr 爱戴/v./wj 对/p 社会/n 繁荣/an 更为/d 重要/a./wj 一定/d 要/v 加分/v.

(2) 很/d 不/d 公平/a,/wd 考试/vi 考/v 的/udel 是/vshi 学习/v,/wd 而/cc 不/d 是/vshi 艺术/n 院校/n 招生/vn./wj 如果/c 你/rr 认为/v 孩子/n 有/vyou 特长/n,/wd 去/vf 考/v 艺术/n 院校/n 呀/y! /wt

3.3 对情感词典的扩展

中文词汇的倾向性是研究中文文本情感分类问题的基础。本文分别应用两种方法对中文情感词典进行扩展。扩展后的情感词典中,除了反映词汇的倾向性以外,还反映了该词汇的情感程度数值,这为下一步的分类工作奠定了基础。

3.3.1 基于 HowNet 的情感词典扩展

HowNet (中文名称为知网) 是一个以汉语和英语的词语所代表的概念为描述对象,以揭示概念与概念之间以及概念所具有的属性之间的关系为基本内容的常识知识库^[6]。HowNet 不是一部义类词典,HowNet 是把概念与概念之间的关系以及概念的属性与属性之间的关系形成一个网状的知识系统。HowNet 认为所有的概念都可以分解成各种各样的义原,并且可以利用一个有限的义原集合来组成一个无限的概念集合,并描述概念之间的关系以及属性与属性之间的关系。刘群等学者利用 HowNet 中义原树中的距离,计算义原之间的相似程度,如公式(1),其中 α 是正的可变参数, $\text{dist}(p_1, p_2)$ 表示义原树中的距离^[11]。同时,作者选取两个词之间的最大义原相似度作为两个词的相似度。

$$\text{sim}(p_1, p_2) = \frac{\alpha}{\alpha + \text{dist}(p_1, p_2)} \quad (1)$$

刘群等的词汇相似度计算方法在区分词的情感倾向上还不甚理想。比如利用该方法计算“美丽”与“贼眉鼠眼”的相似度为 0.81,会比“美丽”与“优雅”的相似度 0.78 还要高。江敏等^[12]提出了一个新的义原相似度计算方法,考虑了义原在义原树中的层次,认为两个词虽然距离相同,但是如果位于的层次越深,相似度越大,具体如公式(2)所示。

$$\text{sim}(p_1, p_2) = \alpha \times (\text{depth}(p_1) + \text{depth}(p_2)) / [\alpha \times \alpha (\text{depth}(p_1) + \text{depth}(p_2)) + \text{dist}(p_1, p_2) + |\text{depth}(p_1) - \text{depth}(p_2)|] \quad (2)$$

同时,有以下规定:

(1) 如果两个义原是对义或反义关系,相似度为-1;

(2) 如果两个义原路径中存在对义或反义,相似度为-1 * $\text{sim}(p_1, p_2)$,其中 $\text{sim}(p_1, p_2)$ 为把距离最近的一对对义或反义关系的义原节点看作同一个节点,再应用前述公式(1)计算得出的结果。这种方法考虑了 HowNet 中的反义和对义关系,对倾向性的判断有很大的帮助。知网中的概念可由独立、关系和符号三种义原描述,利用义原间的相似度就可以计

算概念间的相似度。

有了计算词汇之间的相似度的方法,对于任意一个词即可根据该词与两组基准词的语义关系紧密程度,通过计算得到语义倾向性度量值。这两组基准词是人工选取的,每个词都有明确的倾向性,一组表达正面情感,一组表达负面情感。公式(3)的结果可作为一个倾向性度量,其中 $seed_1$ 表示正面基准词组, $seed_2$ 表示负面基准词组,两组基准词的个数分别为 n, m :

$$\text{orientation}(\text{word}) = \frac{1}{n} \sum_{i=1}^n \text{sim}(\text{word}, \text{seed}_{1i}) - \frac{1}{m} \sum_{i=1}^m \text{sim}(\text{word}, \text{seed}_{2i}) \quad (3)$$

如果结果为正,则表示该词为褒义,如果结果为负,则表示该词为贬义。朱妍岚等通过实验证明倾向性判断的准确性与基准词的选取有很大关系,实验结果表明了基准词数量越多,倾向性判断准确率越高^[13]。通过选择 40 对基准词,实验表明该方法可以达到 99% 的准确率。同时,公式(3)的结果还反映一个词表达情感的程度,如“肤浅”的计算结果为 -0.67,而情感更鲜明的词“腐朽”的计算结果为 -1。我们利用 HowNet 提供的情感分析专用词典,根据公式(3)计算每个词的情感倾向和情感强度,对现有情感词典扩展后,构建了一个新的情感词典 HWSD。HWSD 中的每一项由一个三元组组成,即 $\text{Item} \langle \text{Word}, \text{Polarity}, \text{Strength} \rangle$,其中 Word 表示词汇名称, Polarity 表示该词的倾向性, Strength 表示该词的情感强度,其数值用 $\text{orientation}(\text{Word})$ 的绝对值表示。HWSD 总共收录 8341 个中文词汇,其中褒义词 4095 个,贬义词 4246 个。

3.3.2 基于 NTUSD 的情感词典扩展

National Taiwan University Sentiment Dictionary (NTUSD) 是台湾大学总结整理的中文情感词典,包括简体中文和繁体中文两个版本,每个版本包含正向评价词 2812 个,负向评价词 8276 个^[14]。本文选择简体中文版的 NTUSD 作为情感分类的基础词汇集,然而 NTUSD 只提供了词汇列表,并没有提供这些词汇所反映情感的强度。可以观察到这样的现象,当人们读到一个新的中文词汇时,可以根据由组成该词的每个汉字来猜测该词的意思。基于此,我们提出这样的假设,一个中文词的情感倾向是组成该词的每个汉字倾向性的函数。利用 Ku 等提出的方法^[15],我们可以根据在 NTUSD 中的出现频率来计算一个汉字的倾向性,如公式(4)和(5)所示。该公式不仅可以反映出一个汉字的情感倾向性,还能反映这个倾向性的强度。

$$P_{ci} = \frac{fp_{ci} / \sum_{j=1}^n fp_{cj}}{fp_{ci} / \sum_{j=1}^n fp_{cj} + fn_{ci} / \sum_{j=1}^m fn_{cj}} \quad (4)$$

$$N_{ci} = \frac{fn_{ci} / \sum_{j=1}^m fn_{cj}}{fp_{ci} / \sum_{j=1}^n fp_{cj} + fn_{ci} / \sum_{j=1}^m fn_{cj}} \quad (5)$$

其中 fp_{ci} 和 fn_{ci} 分别表示一个汉字在 NTUSD 褒义词集和贬义词集中出现的频率, n 和 m 分别表示在 NTUSD 褒义词集和贬义词集中不同汉字的个数。由于在 NTUSD 中,表达负向情感的词要比正向情感的词多很多,所以在统计汉字频率

时,可能造成褒义词集中汉字的出现频率要比贬义词集中的出现频率小。因此,公式(4)和(5)对汉字的出现频率进行了归一化处理,从而保证计算结果的统一性。 P_{ci} 和 N_{ci} 分别表示一个汉字的正向倾向性和负向倾向性,两者之差反映了该汉字的整体情感倾向性以及强度,如公式(6)所示。

$$S_{ci} = (P_{ci} - N_{ci}) \quad (6)$$

设一个中文词汇 w 由 p 个汉字组成,即 c_1, c_2, \dots, c_p , 这样 w 的情感倾向 S_w 可以由公式(7)计算。

$$S_w = \frac{1}{p} \times \sum_{j=1}^p S_{c_j} \quad (7)$$

根据公式(6)和(7),一个中文词汇的情感倾向 S_w 取值范围在 -1 和 1 之间, S_w 大于 0 表示该词表达正向情感, S_w 小于 0 表示该词表达负向情感, S_w 的绝对值越大,表示情感越强烈。利用公式(7),我们可以计算 NTUSD 中每个词汇的倾向性。类似 HWSD,对 NTUSD 中的每一个词汇,我们用一个三元组表示,即 $\text{Item} \langle \text{Word}, \text{Polarity}, \text{Strength} \rangle$ 。其中 Word 表示词汇名称, Polarity 表示该词的倾向性, Strength 表示该词的情感强度,其具体结果由 S_w 的绝对值表示。

3.4 修饰词词典的构建

我们发现在网民评论中,有大量的副词修饰情感词汇,比如在 3.1 小节提到的“更为重要”、“很不公平”。“更为”修饰“重要”,表达的情感明显比“重要”要强烈;“不”修饰“公平”,表达了否定的意向,用“很”修饰“不公平”,表达的情感明显比“不公平”强烈。为了更加准确地计算评论的倾向性,我们构建了修饰词词典。

文献[16]中列举了副词类型及每种类型副词主要包括的词汇,我们选取其中有语气强弱之分的副词和否定副词构建修饰词词典。因为此种副词的个数不多,我们人工对此类词汇进行语气强弱的标注,用一个二元组表示为 $\text{Item} \langle \text{Adv}, \text{Modality} \rangle$,其中 Adv 表示词汇名称, Modality 表示该词的语气强度,一个副词的语气强度 M_{adv} 取值范围在 -1 或 (0, 2) 之间。-1 是否定副词的阈值,其它词汇的强度在 0 到 2 之间,越接近 0 说明该词表达的强度越弱,越靠近 2 说明该词表达的强度越强烈。如:“不”的强度为 -1,“也许”的强度设置为 0。

4,“极其”的强度设置为 1.9。

3.5 网络评论倾向性计算

因为网民的评论一般都很短,对分词标注好的评论,根据构建好的两种情感词典和修饰词词典,能够快捷、准确地计算评论的倾向性。

本文以每一句话为一分割(包含逗号分割),将每条评论分成 n 个句子 $Sen_1, Sen_2, \dots, Sen_n$ 。提取每一句话中的情感词,如果情感词 W_i 前面出现的两个词中有一个或两个副词,并且副词 Adv_h, Adv_e 在修饰表中。那么每个情感词最终表达的倾向性以及强度,如公式(8)所示。

$$O_{w_i} = M_{advh} \times M_{adve} \times S_{w_i} \quad (8)$$

句子 Sen_m 中包括 k 个情感词,记为 $W_1, W_2, \dots, W_k, Sen_m$ 最终表达的倾向性以及强度,如公式(9)所示。

$$O_{Sen_m} = \sum_{i=1}^k O_{w_i} \quad (9)$$

那么含有 n 个句子的评论 C_i , 所表达的倾向性, 如公式 (10) 所示。

$$O_{c_i} = \sum_{l=1}^n O_{s_{m_l}} \quad (10)$$

我们将 O_{c_i} 得分在 $[-0.1, 0.1]$ 之间的评论算作中性评论, 大于 0.1 的评论为正向评论, 小于 -0.1 的评论记为负向评论。

4 网络舆情倾向性分析系统的设计与实现

根据前面提到的问题, 我们设计了基于词典技术的评论挖掘系统。图 1 显示了系统的整体架构。

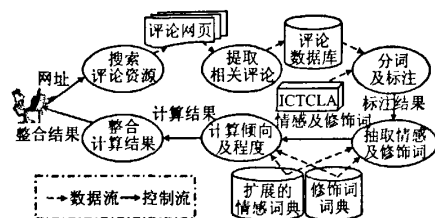


图 1 网络舆情倾向性分析系统整体架构

Fig. 1 Architecture of web public polarity analysis system

该系统中, 由用户输入相关网址 (如网易、搜狐等), 系统搜索带有评论的资源 (如留言板、相关评论等), 获取相关网页, 从中抽取相关的评论并存储。基于存储的这些评论, 采用 3.2 节的方法进行评论和标注, 基于扩展的情感词典从中抽取情感词及其修饰词 (副词), 这里的扩展情感词典是采用 3.3 节的方法对现有的情感词典扩展后建立的, 而修饰词词典则是采用 3.4 节方法建立的。应用 3.5 节的计算方法计算情感词的倾向性及倾向程度, 再对所有评论的情感倾向计算结果进行整合, 汇总出观点、倾向及程度。

基于上述过程, 系统分为如下几个模块:

(1) 网页抓取模块。输入一条门户网站的新闻网页地址后, 爬虫分析出该新闻网页的留言板地址, 开始顺序抓取留言板下每个页面的信息, 并保存到本地;

(2) 网页分析处理模块。根据网页的结构, 分析、处理得到每个评论信息, 并保存每个评论的网民 ID、IP 地址、评论时间、评论内容到数据库中;

(3) 去冗余模块。根据 IP 地址、评论时间和内容, 去除同一网民同一评论在留言板上多次出现的现象;

(4) 分词与标注模块。应用 ICTCLAS 对数据库中每一条评论进行分词和标注;

(5) 计算模块。提取评论中的情感词汇, 通过情感词汇表和修饰词词典, 计算得到词语的倾向性, 进而计算句子的倾向性, 得到该条评论的倾向性, 迭代直到所有评论分析完毕;

(6) 图形用户界面模块。提供交互信息, 并且显示最终的分析结果。

5 实验及分析

我们爬取了 SOHU 网站上有关“你认为特长生考试加分, 这样公平吗?”^[9] 的前 224 条网民的评论。经过人工分析,

其中 9 条属于冗余的评论, 17 条评论与此话题无关, 73 条支持加分的评论, 111 条反对加分的评论, 还有 14 条中性的评论。

我们应用图 1 所示的网络舆情倾向性分析系统, 分别单独使用 HowNet、NTUSD 构建的情感词典分析处理网民的评论, 接下来将两种词典进行合并, 抽取两个表中倾向性相反的相同词汇, 进行人工标注, 再汇总成一个词典, 再次分析处理网民的评论。这样我们自动得到三种不同倾向性的评论数目。3 种方法自动获取的不同倾向性评论数目见表 1, 准确率见图 2。

表 1 不同方法获取的评论数目及准确数目

Table 1 Reviewer and precision number of different methods

方法类别	HowNet		NTUSD		Both		手工标注
	自动识别数	准确数	自动识别数	准确数	自动识别数	准确数	
支持	83	62	58	38	88	68	73
反对	122	85	108	61	118	96	111
中性	10	2	49	2	9	3	14
合计	215	149	215	101	215	167	198

基于 HowNet 的情感词典褒义词、贬义词数目相当并且众多, 而基于 NTUSD 的情感词典褒义词数目偏少, 我们可以发现基于 HowNet 的准确率要比基于 NTUSD 的准确率高一

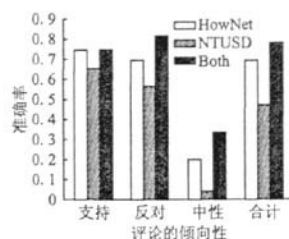


图 2 不同方法准确率比较

Fig. 2 Comparison of precision among different methods

些。当我们把两个词典汇总到一起之后, 准确率再次得到提升。这说明评论的准确性和情感词汇表中词汇的数目有很大关系, 即情感词典词数越大, 得到的结果越准确。

6 结论及进一步的工作

传统的机器学习方法需要人工标注训练出分类器, 工作量大。本文针对网民评论内容较短、所有情感词汇广泛的特点, 应用词典技术初步设计并实现了一个网络舆情系统, 通过实验表明该系统达到了较理想的效果, 并且分类的速度比机器学习方法快。但是其中还有很多不完善之处, 包括:

(1) 如何过滤掉与话题不相关的评论;

(2) 用 HowNet 构建的词汇表明明显比 NTUSD 构建的词汇表得出的结果要准确, 两者合并之后更加的准确, 经分析这与词汇表大小有很大关系, 那么如何建立更加准确的情感词汇表;

(3) 数据库中保存了 IP 地址和评论时间, 可以进一步分

析网民的地域和评论时间与该话题之间的联系,是否能挖掘到更为有趣的信息.

解决上述问题,提高系统分析的质量,是我们进一步的研究工作.

References:


- [1] Kim S, Hovy E. Determining the sentiment of opinions[C]. In: Proceedings of COLING-04, the Conference on Computational Linguistics (COLING-2004). Geneva, Switzerland: 2004, 1367-1373.
- [2] Hu M, Liu B. Mining opinion features in customer reviews[C]. In: Proceedings of Nineteenth National Conference on Artificial Intelligence (AAAI-2004). San Jose, USA: 2004, 755-760.
- [3] Gamon M, Aue A, Oliver S, et al. Pulse: mining customer opinions from free text[C]. In: Proceedings of IDA-05, the 6th International Symposium on Intelligent Data Analysis. Lecture Notes in Computer Science, Springer-Verlag. Madrid, Spain: 2005, 121-132.
- [4] Yao Tian-fang, Nie Qing-yang, Li Jian-chao, et al. An opinion mining system for Chinese automobile reviews[C]. Proceedings of 25th Chinese Information Academy. 2006, 260-281.
- [5] WordNet[EB/OL]. <http://wordnet.princeton.edu/>. 2006.
- [6] HowNet[EB/OL]. <http://www.keenage.com>. 2007.
- [7] Turney P. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews[C]. In Proceeding of ACL 2002, 2002, 417-424.
- [8] Liu B, Hu M, Cheng J. Opinion observer: analyzing and comparing opinions on the web [C]. In: Proceedings of WWW05, the 14th International Conference on World Wide Web, Chiba, Japan: 2005, 342-351.
- [9] <http://comment2.news.sohu.com/viewcomments.action?id=10000102625>[EB/OL]. 2006.
- [10] ICTCLAS. ICTCLAS's home page [EB/OL]. <http://ictclas.org/>. 2006.
- [11] Liu Qun, Li Jian-su. Word similarity computing based on how-net [C]. The 3rd Chinese Lexical Semantics Workshop, 2002.
- [12] Jiang Min, Xiao Shi-bin, Wang Hong-wei, et al. An improved word similarity computing method based on hownet[C]. Journal of Chinese Information Processing, 2008, 22(5): 84-89.
- [13] Zhu Yan-lan, Min Jin, Zhou Ya-qian, et al. Semantic orientation computing based on hownet[J]. Journal of Chinese Information Processing. 2006, 20(1): 14-20.
- [14] <http://nlg18.csie.ntu.edu.tw:8080/opinion/>, 2007.
- [15] Ku L, Lo Y, Chen H. Using opinion scores of words for sentence-level opinion extraction [C]. In Proceeding of the 6th NTCIR Workshop Meeting on Evaluation of Information Access Technologies, 2007, 316-322.
- [16] <http://www.yyxx.sdu.edu.cn/chinese/new/content/5/02/main03-05-02.htm>[EB/OL]. 2006.

附中文参考文献:

- [4] 姚天防, 聂青阳, 李建超, 等. 一个用于汉语汽车评论的意见挖掘系统[C]. 中国中文信息学会二十五周年学术会议论文集, 北京: 清华大学出版社, 2006, 260-281.
- [11] 刘 群, 李素建. 基于《知网》的词汇语义相似度的计算[C]. 第三届汉语词汇语义学研讨会, 台北, 2002.
- [12] 江 敏, 肖诗斌, 王弘蔚, 等. 一种改进的基于《知网》的词语语义相似度计算[J]. 中文信息学报, 2008, 22(5): 84-89.
- [13] 朱嫣岚, 闵 锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报, 2006, 20(1): 14-20.

作者: 杨超, 冯时, 王大玲, 杨楠, 于戈, YANG Chao, FENG Shi, WANG Da-ling, YANG Nan, YU Ge

作者单位: 杨超, YANG Chao(东北大学, 信息科学与工程学院, 辽宁, 沈阳, 110004; 河北省电力研究院, 河北, 石家庄, 050021), 冯时, 杨楠, FENG Shi, YANG Nan(东北大学, 信息科学与工程学院, 辽宁, 沈阳, 110004), 王大玲, 于戈, WANG Da-ling, YU Ge(东北大学, 医学影像计算教育部重点实验室, 辽宁, 沈阳, 110004; 东北大学, 信息科学与工程学院, 辽宁, 沈阳, 110004)

刊名: 小型微型计算机系统 

英文刊名: JOURNAL OF CHINESE COMPUTER SYSTEMS

年, 卷(期): 2010, 31(4)

被引用次数: 13次

参考文献(20条)

1. Kim S. Hovy E [Determining the sentiment of opinions](#)[外文会议] 2004
2. Hu M. Liu B [Mining opinion features in customer reviews](#)[外文会议] 2004
3. Gamon M. Aue A. Oliver S [Pulse: mining customer opinions from free text](#) 2005
4. Yao Tian-fang. Nie Qing-yang. Li Jian-chao [An opinion mining system for Chinese automobile reviews](#) 2006
5. WordNet 2006
6. HowNet 2007
7. Turney P [Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews](#)[外文会议] 2002
8. Liu B. Hu M. Cheng J [Opinion observer: analyzing and comparing opinions on the web](#) 2005
9. [查看详情](#) 2006
10. ICTCLAS [ICTCLAS's home page](#) 2006
11. Liu Qun. Li Jian-su [Word similarity computing based on how net](#) 2002
12. Jiang Min. Xiao Shi-bin. Wang Hong-wei [An improved word similarity computing method based on hownet](#) [期刊论文]-Journal of Chinese Information Processing 2008(05)
13. Zhu Yan-lan. Min Jin. Zhou Ya-qian [Semantic orientation computing based on hownet](#)[期刊论文]-Journal of Chinese Information Processing 2006(01)
14. [查看详情](#) 2007
15. Ku L. Lo Y. Chen H [Using opinion scores of words for sentence level opinion extraction](#) 2007
16. [查看详情](#) 2006
17. 姚天昉. 聂青阳. 李建超 [一个用于汉语汽车评论的意见挖掘系统](#)[会议论文] 2006
18. 刘群. 李素建 [基于《知网》的词汇语义相似度的计算](#) 2002
19. 江敏. 肖诗斌. 王弘蔚 [一种改进的基于<知网>的词语语义相似度计算](#)[期刊论文]-中文信息学报 2008(05)
20. 朱嫣岚. 闵锦. 周雅倩 [基于HowNet的词汇语义倾向计算](#)[期刊论文]-中文信息学报 2006(01)

本文读者也读过(5条)

1. 郑魁. 疏学明. 袁宏永. ZHENG Kui. SHU Xue-ming. YUAN Hong-yong [网络舆情热点信息自动发现方法](#)[期刊论文]-计算机工程2010, 36(3)
2. 郑魁. 疏学明. 袁宏永. 金思魁. Zheng Kui. Shu Xueming. Yuan Hongyong. Jin Sikui [突发事件网络舆情信息分类](#)

方法研究[期刊论文]-[计算机应用与软件](#)2010, 27(5)

3. [吴焕政](#), [吴渝](#), [WU Huan-zhen](#), [WU Yu](#) BBS网络舆情定量分析研究[期刊论文]-[广西师范大学学报\(自然科学版\)](#) 2010, 28(3)
4. [柳位平](#), [朱艳辉](#), [栗春亮](#), [向华政](#), [文志强](#), [LIU Wei-ping](#), [ZHU Yan-hui](#), [LI Chun-liang](#), [XIANG Hua-zheng](#), [WEN Zhi-qiang](#) 中文基础情感词典构建方法研究[期刊论文]-[计算机应用](#)2009, 29(10)
5. [刘磊](#), [Liu Lei](#) 网络舆情分析系统研究[期刊论文]-[情报探索](#)2010(10)

引证文献(13条)

1. [方媛](#), [李海军](#) Web舆情监测系统设计中BBS站点数据的采集与分析[期刊论文]-[自动化与仪器仪表](#)

2013(5)

2. [王征](#), [徐培文](#) 基于互联网用户心理挖掘的网站深翻系统[期刊论文]-[计算机科学](#) 2012(11)
3. [陈越](#), [李超零](#), [于洋](#), [黄惠新](#) 基于TOPSIS方法的网络舆情威胁评估模型[期刊论文]-[情报杂志](#) 2012(3)
4. [王晓东](#), [李永波](#), [郑颖](#) 基于模板匹配的网络评论倾向性分析[期刊论文]-[计算机工程与应用](#) 2012(32)
5. [余正炜](#), [钱松荣](#) 基于神经网络的文本倾向性分析系统的研究[期刊论文]-[微型电脑应用](#) 2011(12)
6. [丁菊玲](#), [勒中坚](#) 基于观点树的网络舆情危机预警方法[期刊论文]-[计算机应用研究](#) 2011(9)
7. [刘建波](#), [杨峰](#) 面向舆情分析的短文本频繁模式聚类算法[期刊论文]-[北京电子科技学院学报](#) 2010(4)
8. [陈波](#), [于泠](#), [刘君亭](#), [褚为民](#) 泛在媒体环境下的网络舆情传播控制模型[期刊论文]-[系统工程理论与实践](#) 2011(11)
9. [王祖辉](#), [姜维](#) 基于粗糙集的在线评论情感分析模型[期刊论文]-[计算机工程](#) 2012(16)
10. [方媛](#), [李海军](#) Web舆情监测系统设计中BBS站点数据的采集与分析[期刊论文]-[自动化与仪器仪表](#)

2013(5)

11. [郭建忠](#), [成毅](#), [傅文棋](#), [杨海鹏](#) 突发事件网络舆情可视化研究[期刊论文]-[地矿测绘](#) 2012(4)
12. [张玉峰](#), [何超](#) 基于Web挖掘的网络舆情智能分析研究[期刊论文]-[情报理论与实践](#) 2011(4)
13. [傅向华](#), [刘国](#), [郭岩岩](#), [郭武彪](#) 中文博客多方面话题情感分析研究[期刊论文]-[中文信息学报](#) 2013(1)

本文链接: http://d.wanfangdata.com.cn/Periodical_xwxjsjxt201004024.aspx