基于多词典融合的词汇语义倾向判别

刘清松,张仰森
(北京信息科技大学 人工智能实验室,北京 100192)

摘 要:文本情感倾向性判别是情感分析的重要组成部分,而精确的词汇情感计算是文本情感倾向性判别的基础。基于词汇知识库的情感词倾向判别计算引起了学者们的广泛关注与研究。文中融合国内知名的三大词典:HowNet、同义词词林、情感词汇本体,重新对基准词对做进一步的归纳与总结,从不同的角度构建三类等价情感倾向集合并提出两种处理集合的策略,建立了待定情感词与特定等价情感倾向集合的情感映射关系。实验结果表明:该方法获得的最高准确率可达91.62%,平均正确率85.31%,符合人们对词语的情感倾向认识。

关键词:HowNet;情感词汇本体;同义词词林;等价情感倾向集合
中图分类号:TP301 **文献标识码**:A **文章编号**:1673-629X(2015)05-0104-06
doi:10.3969/j.issn.1673-629X.2015.05.025

Lexical Semantic Tendency Determination Based on Multi-dictionary Strategy

LIU Qing-song, ZHANG Yang-sen
(Institute of Intelligent Information Processing, Beijing Information Science and Technology University, Beijing 100192, China)

Abstract: Text sentiment orientation analysis is an important part of textual emotion classification, and to determine the emotional bias for words precisely is the very foundation of text sentiment orientation analysis. Nowadays, dictionary-based methods of sentiment orientation computing for words have been widely concerned and researched. In this paper, summarize and conclude benchmark words considering knowledge from HowNet, Tongyici-Cilin and affective lexicon ontology, then build three emotion-equal sets from different perspective and propose two strategies of set processing, finally establish the mapping between the sentiment orientation of a word to be computed and a specific emotion-equal set. Experimental results has generally agreed with the sentiment orientation felt by human, and achieved maximum correctness 91.62% and 85.31% in average.

Key words: HowNet; affective lexicon ontology; Tongyici-Cilin; emotion-equal set

0 引 言

信息技术的飞速发展推动了互联网的普及与传播,网上可共享的信息资源与日俱增,对知识的处理要比知识的渴求更重要。由于情感分析技术[1]在人工智能、舆情分析等领域有着广阔的发展空间,尤其在微博情感倾向性分析[2]方向的应用,使其成为研究热点,技术也日益成熟。词汇语义倾向性判别是句子、段落、篇章、文本级情感分析的前提,而词语情感计算的常规思路是对词语倾向性的定性判别过度到定量分析。

词汇情感倾向性计算主要集中在基于词汇知识库与基于统计方法上。朱嫣岚[3]在HowNet知识库的基础上,提出基于语义相似度与语义相关场两种词汇语义倾向性计算方法,通过计算待定词与基准词的相似度与相关场,确定待定词的情感倾向,准确率可达80%以上;徐琳宏[4]在选取与知网中已标注情感倾向词语相似度高的词汇作为特征值,利用SVM分析文本倾向性,取得相当高的分类效果;Turney[5]利用选取的基准词对,采用PMI来度量词语与基准词对的相关联程度,进而确定词语情感倾向,准确率达82.8%;陈晓东[6]运用情感倾向点互信息(Semantic Orientation

收稿日期:2014-07-05 **修回日期**:2014-10-06 **网络出版时间**:2015-04-22
基金项目:国家自然科学基金资助项目(61370139);北京市属高等学校创新团队建设与教师职业发展计划项目(IDHT20130519);北京市教委专项基金(PXM2013_014224_000042,PXM2014_014224_000067)
作者简介:刘清松(1987-),男,硕士研究生,研究方向为自然语言处理;张仰森,博士,教授,研究方向为自然语言处理、人工智能。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20150422.1008.023.html

Pointwise Mutual Information,SO-PMI)算法对新浪微博进行语料测试,自动获取领域情感词,然后构建面向中文微博情感词典,最高准确率达 74.2%。

单一词典难以做到尽善尽美,文中融合三大词典,旨在挖掘待定情感词汇的语义构成与词词关系,表现在知网对词汇概念的阐述与词语之间的近义同义关系,尝试对词汇的情感倾向判别做到最准确。主要做的工作:选取规范的标准词对;获取三组与待定情感词汇等价的情感词语集合;提出两种获取情感词倾向策略并实现。

1　知识库简介与词语相似度计算

1.1　HowNet、同义词词林、情感词汇本体介绍

1.1.1　HowNet 的知识原理

依据董振东[7]的观点,知网(英文名称为 HowNet)是一个以汉英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。《知网》中的“概念(也称为义项)”是对词语的一种语义上的描述,每个词语可表示为一个或多个概念,概念是用一种崭新的规范描述体系—KDML(Knowledge Database Mark-up Language)来体现。而概念中包含许多“义原”,义原可认为是在知网体系中不易于分割、最基本的最小意义单位,用它来表达概念与概念之间、概念的属性之间的关系。义原可分为基本义原、次级义原、语法义原、关系义原。知网的哲学思想与其根本特性决定了知网是以概念为描述对象的知识系统。概念集的组成结构见表 1。

表 1　2012 年知网知识库概念结构示意图

义项属性	对应值	中文解释
NO. =	020209	记录的编号
W_C =	表扬	中文词语
G_C =	verb [biao3 yang2]	中文词语词性
S_C =	PlusFeeling|正面情感	情感倾向性
E_C =		中文词语举例
W_E =	commend	英文单词
G_E =	verb[3commendverb-0vt,sobj,ofnpa,endprep44]	英文单词词性
S_E =	PlusFeeling|正面情感	英文情感倾向
E_E =		英文单词举例
DEF =	{praise|夸奖}	概念定义
RMK =		不常用词

1.1.2　同义词词林介绍

同义词词林[8]是梅家驹[9]等人在 1983 年编制而成,这部词典不仅包括词语的同义词,同时也有一定数量的同类词,即广义的相关词。后期哈尔滨工业大学信息检索实验室利用多种词语相关资源对其进行扩展,剔除 14 706 个罕用词和非常用词,完成一部具有汉语大词表的《同义词词林(扩展版)》,最终共收录 77 343 条词语。

词林按照树状层次结构把所有词条组织在一起,词汇可分成大、中、小三类,而小类中的词汇又可根据词义的远近和相关性分成词群(或称为段落),词群中的词语又进一步分为行(原子词群),同一行中词义相近或相关性很强。使扩展版的词林在原有的基础上深加了两层,形成具有五层树状层次结构的词典。其编码方式的第八位标记有三种:“=”代表“相等”、“同义”;“#”代表“不等”“同类”;“@”代表“独立”“自我封闭”。

1.1.3　情感词汇本体库

中文情感词汇本体库是大连理工大学信息检索研究室在林鸿飞教授[10]的指导下,经全体成员的努力整理和标注汇总而成的一个中文本体资源。描述了中文词汇的词性种类、情感类别、情感强度及极性等信息。希望在情感计算方向,为中文文本情感分析和倾向性判别提供便捷而可靠的辅助资源。词汇本体的情感分类为 7 大类 21 小类,收录情感词总共 27 466 条。每个情感词可对应多个情感,情感分类用于刻画主要情感倾向,辅助情感为该情感词在具有主要情感分类的同时包含其他情感分类倾向。“词性种类”共 7 类,分别为名词(noun)、动词(verb)、形容词(adj)、副词(adv)、网络词语(nw)、成语(idiom)、介词短语(prep);“强度”分为 1,3,5,7,9 五档,9 表示强度最大,1 为强度最小。“极性”中 0 代表中性,1 代表褒义,2 代表贬义,3 代表褒贬两性。

1.2　基于 HowNet 词语相似度计算

在知网体系中,词语相似度计算可由义项相似度计算得到,而义项的知识表示是通过义原及其义原之间关系来呈现的,因此义原相似度的计算是词语相似度的基础与前提。

刘群[11]认为词语相似度是主观性很强的一个概念,并没有客观的定量标准来衡量词词之间的相似程度。它涉及到词语的词法、语义、句法、语用等方面,其中语义方面影响最大。刘群的思想与知网中对概念的语义阐述不谋而合,实验得出的结论令人信服,结果趋于合理。随着知网本体库的不断更新,知识体系的完善,人们对知网的研究也更加深入。葛斌[12]等在实验的基础上,认为知网义原相似度计算不仅要考虑到树的深度,而且要兼顾树的密度因素对义原节点权重的影响,而义项(概念)相似度计算是通过 KDML 描述的结构来进行的,最终义项(概念)间的相似度计算公式如下:

$$\text{simConcept}(a,b)=\sum_{i=1}^{4}\beta_i\prod_{j=1}^{i}\text{sim}_j(a,b) \quad (1)$$

其中,β_i 为调节参数,并满足 $\beta_1+\beta_2+\beta_3+\beta_4=1$,且 $\beta_1\geqslant\beta_2\geqslant\beta_3\geqslant\beta_4>0$;$\text{sim}_j(a,b)$ 为义原相似度。

2 词汇语义倾向性算法

2.1 获取等价情感倾向词集合

等价情感倾向词集合:与特定词语在情感信息方面相等价(极相似)的情感词的汇总集合。而情感信息包含情感极性、情感类别、情感强度等。成员集合主要来源于情感词汇本体和同义词词林。

2.1.1 基于 HowNet 义项限制条件的等价情感倾向词集合的获取

词与词之间的语义相似程度越强,情感倾向也越接近。而词语相似度计算最终会回归到对义项相似度的计算,进一步会涉及到 KDML 对义项的结构化描述。以“DEF=”开始,在四类义原中,基本义原是最重要的,也是区分义项间是否接近程度最高的实体义原,获取等价情感倾向词集合必不可少。文中采取两种思路来进行对等价情感倾向词集合的合理提取。

第一种:如果两个义项之间的情感倾向性相似或等价,与相应义原数目紧密相关:

(1)基本义原相等,次级义原集合相等;

(2)基本义原相等,关系义原集合相等;

(3)基本义原相等,符号义原集合相等。

利用知网知识,从情感词汇本体中,抽取所有满足前提 1 的情感词集合,记 WordList1;满足前提 2 的情感词集合,记 WordList2;满足前提 3 的情感词集合,记 WordList3。若三个情感词集合存在交集词汇,则认为它们与待定情感词语情感倾向越接近。所有获取的等价情感倾向词集合可总记为 SWL0(Semantic Word List Zero)。

第二种:义项之间的语义相似程度,取决于义原的相等义原数目和义原种类数目。因此可限制条件来获取等价情感倾向词集合:基本义原必须相等,即义原所在的树状层次结构中,处于同一级;次级义原与关系义原至少一半相等;若义项存在符号义原,则须至少保持一个相等,否则不必考虑此条件。由此得到等价情感倾向词集合 SWL1。

2.1.2 近义/同义词集合的获取

基本思路:首先从同义词词林中获取与其等价的词集合,即待定情感词在同义词词林中的编码:第八位必须为“=”而非“#”“@”,记为 WordList4;因知网近义词词库出现的词都会在知网词汇表中出现,所以它们都具有语义信息或者有潜在的情感倾向,抽取所有与待定情感词具有近义关系的词集合,记为 WordList5。而获得的所有集合元素必须存在情感词汇本体中,最终所得词集合记为 SWL2。公式如下:

$$\text{SWL2}=\text{WordList4}\cup\text{WordList5}\cap\text{Lexicon} \quad (2)$$

其中,Lexicon 表示为情感词汇本体库中的词集合。

2.2 基准词对的选取

基准词对:具有明显情感倾向的、感情强烈、代表性强的情感词汇组成的集合。基准词集合的选取须满足条件:情感强度大、代表性强、覆盖面广、不出现语义相同等因素。而情感词汇本体共分为 21 小类,从各个类中平均挑选 4~5 个情感强度≥9 的情感词且不存在辅助情感情况;进一步查看选取小类词汇在 Google 搜索所返回的结果数,选择返回结果大的词汇,说明此词常用即满足代表性强;然后通过知网知识库融合或替换语义相同或相似的词汇;最后再有目的地结合王素格[13-14]等提出的基准词集合,因为实践已证明他们所选取的基准词集合的正确性与合适性。选取的基准词对集合见图 1。

2.3 词汇语义情感倾向计算

词汇语义情感倾向计算结果的数值范围一般为[-1,1]。由于缺少语料库的支持,同时相比较在相同测试集的情况下,计算方法的优劣。对词汇语义情感倾向的计算主要有两种策略:

(1)对等价情感倾向词集合的归一化处理;

(2)等价情感倾向词集合元素与基准词的相似度计算。

2.3.1 等价情感倾向词集合 SWL0 的归一化处理

SWL0 包括 WordList1、WordList2、WordList3 集合。分两步走:

Step1:单独对三个集合进行归一化处理,例如对于在 WordList1 中的每个词语的情感强度记为 EI_j,此集合中的词语总数记为 Len_1,则 WordList1 归一化处理后的结果为:

褒义词集合									
中奖	吉祥	诚信	光艳	深情	甜蜜	朋友	奇迹	拥戴	精英
天堂	恬淡	公道	必然	真理	推举	祝福	完美	俊杰	贤惠
不错	先进	美丽	不凡	优胜	嘉奖	圣人	佳节	美好	成功
弘扬	慧心	最好	杰作	极品	毅然	经典	文明	广博	快乐
贬义词集合									
报仇	勃然	不幸	凄惨	心碎	震惊	有害	遗憾	车祸	失败
骗人	魔鬼	痛恨	烦躁	忧愁	悔恨	绝望	事故	郁闷	邪恶
陷阱	凶蛮	卑劣	奸诈	抱怨	最差	猥琐	禽兽	缺点	肮脏
罪恶	虚伪	龌龊	咆哮	灾难	崩溃	怀疑	疯狂	暴力	浪费

图 1　40 组褒贬基准词集合

$$E\_orientation(WordList1)=\frac{\sum_{j=1}^{Len_1}EI_j}{Len_1\times 9}\quad(3)$$

其中,正向极性的情感强度记为正, EI_j 表现为“+”;反之负向极性的情感强度记为负, EI_j 表现为“-”。同理可获得 E_orientation(WordList2)、E_orientation(WordList3)。

Step2:对三个集合归一化处理后的结果再进行带有权重的归一化处理。由于次级义原的重要性不小于关系义原和符号义原,即 WordList1 集合相对而言比其他两组重要,体现在调节参数大小的不平等。公式如下:

$$\begin{aligned}orientation0(word)=&\alpha E\_orientation(WordList1)+\\&\beta E\_orientation(WordList2)+\\&\gamma E\_orientation(WordList3)\end{aligned}\quad(4)$$

其中,word 为待定情感词, $\alpha+\beta+\gamma=1$, $\alpha\geqslant\beta>0$, $\alpha\geqslant\gamma>0$。

2.3.2　等价情感倾向词集合 SWL0 与基准词对情感相似度计算

词集合与基准词对的相似度计算最终会转换到词与基准词对元素的相似度计算,文献[14]引入语义相似度最大值,使得准确率上升许多。公式如下:

$$\begin{aligned}orientation(W)=&\left(\frac{1}{\alpha}\sum_{i=1}^{k}Simlarity(key-p_i,w)+\right.\\&\left.\frac{1}{\beta}\max_{i=1,2,\cdots,k}Simlarity(key-p_i,w)\right)-\\&\left(\frac{1}{\alpha}\sum_{j=1}^{k}Simlarity(key-n_j,w)+\right.\\&\left.\frac{1}{\beta}\max_{j=1,2,\cdots,k}Simlarity(key-n_j,w)\right)\end{aligned}\quad(5)$$

其中, k 表示 k 对基准词,基准词有褒贬之分, $key-p_i$ 表示褒义基准词, $key-n_j$ 表示贬义基准词; $Simlarity(key-p_i,w)$ 是词与词之间的语义相似度值; α、β 为可调节参数。

为了计算集合与基准词对的相似度,首先对三集合单独计算其情感相似度,例如:对 WordList1 计算公式如下所示:

$$orientationSet(WordList1)=\frac{\sum_{i=1}^{Len1}orientation(W_i)}{Len_1}\quad(6)$$

其中, Len_1 为 WordList1 的词语总数目。同理可得 OrientationSet(WordList2)、OrientationSet(WordList3)。

进行归一化处理,把 OrientationSet(WordList1)、OrientationSet(WordList2)、OrientationSet(WordList3)依次代入到公式(4)中的 E_orientation(WordList1)、E_orientation(WordList2)、E_orientation(WordList3)。得到结果为:

$$\begin{aligned}orientation1(word)=&\alpha orientationSet(WordList1)+\\&\beta orientationSet(WordList2)+\\&\gamma orientationSet(WordList3)\end{aligned}\quad(7)$$

其中,word 为待定情感词, $\alpha+\beta+\gamma=1$, $\alpha\geqslant\beta>0$, $\alpha\geqslant\gamma>0$。

2.3.3　等价情感倾向词集合 SWL1、SWL2 的归一化处理

对 SWL1 与 SWL2 集合进行统一处理,原因如下:

(1)集合中的词语元素出现相等(重合)情况;

(2)来源于同一知识库—情感词汇本体,都构成一个且唯一一个集合;

(3)集合选取的基本思路一致—从情感词汇本体知识库中,选取语义上近义或同义的词语。

由于集合元素的无序性,决定元素(词语)之间关系重要程度的平等性。对集合进行归一化处理即利用公式(3)可得出结果 orientation2(word)、orientation4(word)。

2.3.4　等价情感倾向词集合 SWL1、SWL2 与基准词

情感相似度计算

既然集合元素的重要度是相同的,元素(词语)对计算集合与基准词对的情感相似度的贡献也是相同的。结合式(5)、(6)先计算单个词语的情感倾向,然后在集合中计算算术平均值作为结果 orientationSet(SWL1)。

$$\text{orientation3(word)} = A \times \text{orientation(word)} + B \times \text{orientationSet(SWL1)} \quad (8)$$

其中,orientation(word)为 word 代入公式(5)的 W 参数所构成的实参函数; A,B 为调节参数且满足 $A + B = 1$。

公式(8)表示 A,B 的大小可影响待定情感词的情感倾向与 SWL1 集合的情感倾向的偏向程度。调节参数的存在使实验得出的结果与真实的情感倾向不会出现很多的偏薄。

同理可得 orientation5(word)。

3 实 验

3.1 实验数据

对于测试集的选取,考虑到获得的基准词对集合与知识库、网络有关,测试集可从两方面收集。

测试集 1:知网知识库提供了准确的中文情感倾向词汇集,总共有 9 568 条情感词,正向词汇共计4 880 条,负向词汇 4 688 条。之所以选取知网的情感知识库,是由于知网的不断更新,知识库表现为数据信息的丰富、符号方面的简化、易于理解的语言、义原层次结构的严谨等等,受到国内外专家学者的一致认可,权威性不可忽视。

测试集 2:根据第六届中文倾向性分析评测(COAE2014)会议所提供的文本语料与爬取中文微博所得到的语料。依据主观认知,人为来收集词汇,共计捡取 6 670 词条。其中褒义词汇 3 550 条,贬义词汇 3 120条。

3.2 实验设计与结果分析

根据词汇语义情感倾向计算的设计思路,每个测试集都会有六个结果,由于每个待确定情感词通过计算而得出的语义倾向 O_i 值都处于[-1,1]范围内,可通过公式(9)来确定其褒贬倾向性。

$$\text{judge(word)} = \begin{cases} 1, 0.1 \leqslant O_i \leqslant 1 \\ 0, -0.1 \leqslant O_i \leqslant 0.1 \\ -1, -1 \leqslant O_i \leqslant -0.1 \end{cases} \quad (9)$$

其中, judge(word) 为词语的情感倾向极性,1 代表褒义,0 代表中性,-1 代表贬义。

在测试集 1 与测试集 2 的基础上,为了验证该思路的有效性,通过计算得出的测试结果如表 2、3 所示。

表 2 测试集 1 的词语倾向极性判别结果

测试集 1	语义倾向值	正向正确率/%	负向正确率/%	总正确率/%
SWL0	orientation0(word)	83.25	80.14	81.73
	orientation1(word)	89.10	82.62	85.92
SWL1	orientation2(word)	87.32	88.15	87.73
	orientation3(word)	91.62	90.21	90.93
SWL2	orientation4(word)	82.42	80.20	81.33
	orientation5(word)	85.11	83.65	84.39

表 3 测试集 2 的词语倾向极性判别结果

测试集 2	语义倾向值	正向正确率/%	负向正确率/%	总正确率/%
SWL0	orientation0(word)	79.12	80.55	80.02
	orientation1(word)	81.66	82. 00	81.83
SWL1	orientation2(word)	82.36	80.18	81.27
	orientation3(word)	85.87	86.01	85.93
SWL2	orientation4(word)	70.15	76.44	73.30
	orientation5(word)	78.52	80.62	79.57

由表 2 可知:

(1)由于三个等价情感倾向词集合获取的方式不同,得到整体语义倾向的正确率也不尽相同。SWL1 的整体正确率(正向、负向及总体)比 SWL0、SWL2 高出 5 ~ 6 个百分点,因为 SWL2 直接获取的是词,SWL0、SWL1 首先获取的是在知网词语义项中最能体现词语语义值的义原,而义原最具体,涵义最明确,然后得到与之相对应的词,也就解释了 SWL0 的整体测试结果比 SWL2 的要好。而 SWL1 中的每个元素都能满足概念意义相近的义原,使得获取词的语义信息不

会与待定情感词偏薄太远,所得到的语义倾向值比 SWL2 高。

(2)通过等价情感词集合与基准词情感相似度计算的整体结果优于仅仅通过归一化处理方法得到的数据。一方面,基准词对选取的合理性与调节参数的参与,都有助于正确率的提高;另一方面,基于知网的词汇语义相似度算法的改进,对于基准词相似度算法的高正确率奠定了基础,归一化处理的结果好坏依赖于情感词汇本体词语与情感强度的切合程度,由于情感强度都是正整奇数而没有浮点小数的参与,难以精确计算出待定情感词的情感倾向。

从表 3 可知:SWL1 的整体正确率依然比其他两个要高,SWL0 的结果次之,与测试集 1 的结果相符。由于人为因素的参与,褒贬词语的选择只取决于人的主观判别,同时也给计算带来判断词语倾向性上的误差,导致个别结果的正确率很低。

另外,关于词汇语义情感倾向计算中调节参数的选择:最终的综合公式一般选取 $\alpha = 0.4$、$\beta = 0.34$、$\gamma = 0.26$,选择的权重相差不是太大,是由于从两个方面获得的等价情感倾向词集合元素之间语义相似程度接近。公式(8)参数选取原则:$0 < A \leqslant 0.5$、$0.5 \leqslant B < 1$。因为实验发现公式(8)中,后者的计算结果整体上优于前者。

4　结束语

词汇语义倾向性判别是文本情感分析领域的发掘点与延伸点。文中在提取等价情感倾向词集合的基础上,分别运用集合的归一化处理和与基准词对相似度计算两种方式来计算待确定情感词的情感倾向性,进而来比较方法的优劣。实验数据表明,后者的计算结果优于前者,以限制知网义项条件获取的等价情感倾向词集合 SWL1 最符合待确定情感词的语义倾向。而知网知识库的不断更新,相信会吸引越来越多的学者来挖掘知网更多的潜在价值信息。

文中依然存在许多改进或思考的空间:如知识库的持续更新与利用,改进的知网词语相似度计算是否可以与相关度计算、与 PMI(点互信息)结合来进一步改善词词之间的关联程度;可构建新的同义词/近义词词典,来加强 SWL2 的质与量,实验发现其整体的计算结果,不太理想;尝试以义原为基础挖掘语义特征,利用机器学习算法进行分类,或与知识库相结合提高正确率。

参考文献:

[1] 周立柱,贺宇凯,王建勇. 情感分析研究综述[J]. 计算机应用,2008,28(11):2725-2728.

[2] 曹海涛. 基于 PAD 模型的中文微博情感分析研究[D]. 大连:大连理工大学,2013.

[3] 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.

[4] 徐琳宏,林鸿飞,杨志豪. 基于语义理解的文本倾向性识别机制[J]. 中文信息学报,2007,21(1):96-100.

[5] Turney P D, Littman M L. Measuring praise and criticism: inference of semantic orientation from association[J]. ACM Transactions on Information Systems, 2003, 21(4): 315-346.

[6] 陈晓东. 基于情感词典的中文微博情感倾向分析研究[D]. 武汉:华中科技大学,2012.

[7] 董振东,董　强,郝长伶. 知网的理论发现[J]. 中文信息学报,2007,21(4):3-9.

[8] 田久乐,赵　蔚. 基于同义词词林的词语相似度计算方法[J]. 吉林大学学报:信息科学版,2010,28(6):602-608.

[9] 梅家驹. 同义词词林[M]. 上海:上海辞书出版社,1993:106-108.

[10] 陈建美,林鸿飞,杨志豪. 基于语法的情感词汇自动获取[J]. 智能系统学报,2009,4(2):100-106.

[11] 刘　群,李素建. 基于《知网》的词汇语义相似度计算[C]//第三届汉语词汇语义研讨会. 出版地不详:出版者不详,2002:59-76.

[12] 葛　斌,李芳芳,郭丝路,等. 基于知网的词汇语义相似度计算方法研究[J]. 计算机应用研究,2010,27(9):3329-3333.

[13] 王素格,李德玉,魏英杰,等. 基于同义词的词汇情感倾向判别方法[J]. 中文信息学报,2009,23(5):68-74.

[14] 杨昱昺,吴贤伟. 改进的基于知网词汇语义褒贬倾向性计算[J]. 计算机工程与应用,2009,45(21):91-93.

基于多词典融合的词汇语义倾向判别

作者： 刘清松， 张仰森， LIU Qing-song， ZHANG Yang-sen
作者单位： 北京信息科技大学 人工智能实验室,北京,100192
刊名： 计算机技术与发展 ISTIC
英文刊名： Computer Technology and Development
年，卷(期)： 2015(5)