

面向文本情感分析的中文情感词典构建方法

周咏梅¹ 杨佳能² 阳爱民²

(1. 广东外语外贸大学思科信息学院, 广东 广州 510006;

2. 广东外语外贸大学国际工商管理学院, 广东 广州 510006)

摘要: 提出了构建基于 HowNet 和 SentiWordNet 的中文情感词典方法。将词语自动分解为多个义元后计算其情感倾向强度, 并且使用词典校对方法对词语情感倾向强度进行优化。将所构建词典应用到文本情感分析任务中, 使用支持向量机构建文本情感分类器进行实验。实验结果表明, 该词典优于一般极性情感词典, 为情感分析研究提供了有效的词典资源。

关键词: 情感词典; 情感强度; 支持向量机; 情感分析; 中文文本

中图分类号: TP391. 1 **文献标志码:** A

A method on building Chinese sentiment lexicon for text sentiment analysis

ZHOU Yong-mei¹, YANG Jia-neng², YANG Ai-min¹

(1. Cisco School of Informatics, Guangdong University of Foreign Studies, Guangzhou 510006, China;

2. School of Management, Guangdong University of Foreign Studies, Guangzhou 510006, China)

Abstract: A method on building Chinese sentiment lexicon based on HowNet and SentiWordNet was proposed in which sentiment intensity of the word was automatically calculated by decomposing it into multiple semantic units and a lexicon proofreading technique was used to optimize the value of sentiment intensity of the word. The building lexicon was applied to the task of sentiment analysis, in which the support vector machine was used to build the sentiment classifier. The experiment results showed that the built sentiment lexicon was more effective than the general polar sentiment lexicon and provided an effective dictionary resource for the research of sentiment analysis.

Key words: sentiment lexicon; sentiment intensity; support vector machine; sentiment analysis; Chinese text

0 引言

随着 Web2.0 应用的普及,越来越多的用户倾向于在网上发表自己对各种产品或热点事件的观点。这些评论信息表达了用户的主观情感色彩,包含了正面情感和负面情感等。因此,互联网上产生了海量的包含用户主观情感色彩的评论信息,这些信息若采用人工的方式无法实现高效的收集和处

理,由此产生了文本情感分析(text sentiment analysis)技术。文本情感分析又称意见挖掘,是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程^[1]。

词语情感倾向性的判别是文本情感分析研究的基础工作。基于情感词典的文本情感分析是一种非常有效的方法,并得到了广泛的应用。因此,情感词典的构建方法研究是非常重要的课题。目前针对文本情感分析较为典型的研究方法是结合情感词典对

收稿日期: 2013-04-30 网络出版时间: 2013-11-07 12:25

网络出版地址: <http://www.cnki.net/kcms/detail/37.1391.T.20131107.1225.001.html>

基金项目: 国家社科基金资助项目(12BYY045); 教育部人文社会科学研究青年资助项目(10YJCZH247); 广东省科技计划资助项目(2010B031000014)

作者简介: 周咏梅(1971-),女,湖南永州人,教授,硕士,主要研究方向为文本情感分析。E-mail: yongmeizhou@163.com

文本进行特征抽取,然后结合机器学习实现对文本的情感倾向性进行分类。但是由于网络评论往往具有观点表达的隐晦、口语色彩浓重和非规范性等语言特点,研究者在机器学习的基础上引入语义规则分析对文本情感分类系统进行改进。这些改进使得系统分类性能得到了提高。但是没有从根本上解决中文文本在复杂语境中含义多样化的问题。徐琳宏等人构建的中文情感词汇本体库除了包括词语的极性外还包括词语情感倾向的强度值及少量的多义词。这些多义词包含两种以上的极性标注以及相应的情感倾向强度值。然而,情感词汇本体库的构建过程并没有对每个词的情感倾向进行多倾向性分析。这使得这类情感词典所包含的情感极性和相应的强度值具有片面性,不具有普遍适用性。针对这个问题,提出一种综合考虑词语在不同语义情况下的情感倾向权值的计算方法,从而构建同时具有正、负面情感倾向强度值的中文情感词典,并将该词典应用于文本情感特征提取。在文本情感倾向分类实验中取得了较好的分类效果,验证了该词典的有效性。

1 相关研究工作

情感词是构成情感词典的重要组成部分。研究者们提出了诸多方法来提取和标注主观性文本的情感词,如 HATZIVASSILOPOULOS 采用机器学习的方法来计算形容词之间的关系强度实现对形容词的情感倾向进行预测^[2]; WIEBE J M 等人在少量已标注的种子词的基础上,根据其分布相似性对形容词进行聚类,从而实现对未标注词进行分类^[3]; RILOFF E 采用基于 Bootstrapping 的模式匹配算法实现名词抽取^[4]; BARONI M 等人在主观性形容词的情感权值排序中考虑网络信息多样性的因素^[5]; 以及 MOILANEN K 等人对语素进行情感标注等^[6]。实用的情感词典还应当包括词语的极性(正面和负面)以及相应的情感强度值。为了计算词语级别的情感强度值,TURNER P 等人提出了潜在语义分析(latent semantic analysis,LSA)技术^[7]、逐点互信息(pointwise mutual information,PMI)技术^[8-9]以及在 WordNet^[10]中引入同义词集结构关系的方法^[11-12]等。这些技术和方法都提高和改善情感词典构建的效率和准确性。意大利信息科技研究所的 SentiWordnet^[13-14]使用了已有的语言资源 WordNet 中词语的注释作为词语的特征,对注释文本进行分类,采用定量分析的方法来确定每个同义词集中对应词的正面情感、负面情感以及客观性权值。

中文情感词典构建主要的研究思路有基于语料统计以及语义词典等方法。基于语料的情感词典构建方法通过统计词语间的共现信息以及利用词语相似度来计算词语的语义倾向。通常,基于语料的情感词典构建方法通过词语间的共现信息计算不同词语之间的相似度,并利用词语相似度计算词语语义倾向。基于语义知识库的情感词典构建方法通常采用语义词典来计算词语相似度,常用的中文语义词典主要包括 HowNet^[15]和同义词辞林等。国内关于汉语词汇、短语的情感倾向研究已经取得了一些阶段性成果。朱嫣岚等人提出了基于语义相似度和语义相关场的两种词汇语义倾向性计算方法,通过计算目标词汇与 HowNet 中已标注褒贬性词汇间的相似度得到目标词汇的倾向性^[16]。李军等人采用合并去重的方法构建了中文褒贬义词典^[17-21]。柳位平等人采用种子词结合 HowNet 语义相似度计算的方法构建了中文基础情感词典^[22]。徐琳宏等人通过整理和标注构建了中文情感词汇本体库,从不同角度描述一个中文词汇或者短语,包括词语词性种类、情感类别、情感强度及极性等信息^[23]。台湾大学整理构建了中文情感词典(national taiwan university sentiment dictionary, NTUSD),包括正向评价词2 812个,负向评价词8 276个。张成功等人整理了包含基础情感词典及领域词典、网络词典及修饰词词典的综合词典^[24]。考虑情感词在不同语义环境的情感倾向,建立了基于 HowNet 和 SentiWordNet 的中文情感词典(Sentiment lexicon based on HowNet and SentiWordNet, SLHS)。SLHS 中包含各情感词的倾向和强度,为中英文文本情感分析提供了新的词典资源。

2 中文 SLHS 构建方法

2.1 构建方法概述

中文词语含义丰富,同一个词语在不同的语境中往往包含多个不同的意义。例如词语“好事”,作为名词时通常指的是“有益的事情”。但是作为动词时通常“形容一个人闲不住,喜欢生事端”。这个词语在不同的语境中具有明显的情感倾向差异,所以在情感词典的构建中有必要充分考虑这一因素。HowNet 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。在 HowNet 中,中文词语的多义性得到了充分的体现。WordNet 是一个在线的英语词汇数据库。名词、动

词、形容词和副词各自被组织成一个同义词的网络,每个同义词集合都代表一个基本的语义概念,并且这些集合之间也由各种关系连接。SentiWordNet 是以 WordNet 中的同义词集合为基础,对这些同义词集合进行褒义、贬义和客观性 3 种不同类别的情感标注并给出相应的情感强度值。

本研究将 HowNet 中中文词语在不同语义情况下所对应的英文翻译定义为义元。SLHS 的主要构建步骤包括:首先利用 HowNet 获取中文词语的对应的各项英文义元;其次使用 SentiWordNet 数据库检索每个英文义元所处的各个同义词集合;接着计算这些同义词集合的平均情感强度值得到每个义元的情感倾向性强度值;最后计算各项义元的平均情感强度值,即得到中文词语的情感倾向强度值。

2.2 SLHS 构建过程

词语集合的收集是情感词典构建的基本工作。收集了 HowNet 情感词、NTUSD、褒贬义词词典和杨鼎等构建的情感词典,采用合并去重的方法取得词语集,并将其作为 SLHS 的基础情感词语集。该词语集合共包含 24 130 个词,其中正面情感词 10 269 个,负面情感词 13 861 个。SLHS 具体构建过程如图 1 所示。

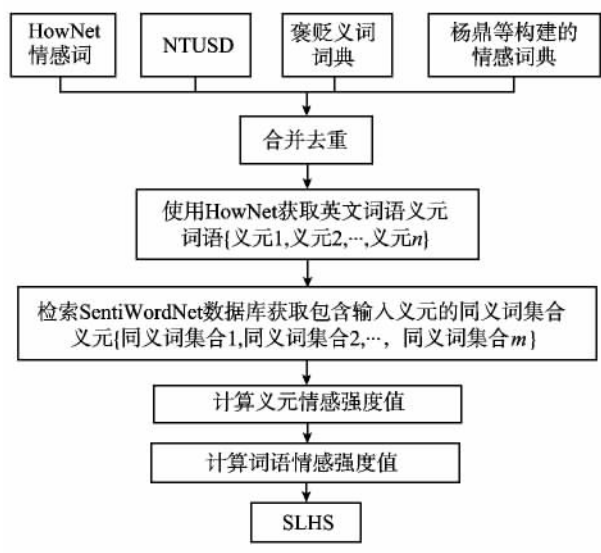


图 1 SLHS 构建过程

Fig.1 Building procedures of SLHS

2.2.1 利用 HowNet 获取中文词语的对应的各项英文义元

在 HowNet 中,词语往往包含多个义元,用以表示在不同语境中的词语所反映的不同意义。定义一个中文词语义元集合为 W ,则 $W = \{M_1, M_2, \dots, M_N\}$,其中 N 表示词语所包含的义元数量, $M_n (n = 1, 2, \dots, N)$ 表示词语的第 n 个义元。通过编程调用 HowNet 的搜索 API 可得到词语的义元集合,例如词语“秀丽”的义元集合为 $\{\text{beautiful}, \text{comely}, \text{handsome}, \text{pretty}\}$ 。如果 HowNet 中不存在所输入的词语,则采用有道词典提供的 API 获取该词语的相应的英文解释作为该词语的义元。但是在词典构建过程中仍然存在少数词语无法通过以上方式获取义元集合,此时本文采用人工干预的方式生成词语的相应义元集合。

设 W_Pos 为词语的正面情感强度值, W_Neg 为词语的负面情感强度值,则

$$W_Pos = \frac{1}{N} \sum_{n=1}^N M_Pos_n, \quad (1)$$

$$W_Neg = \frac{1}{N} \sum_{n=1}^N M_Neg_n, \quad (2)$$

其中 M_Pos_n 为第 $S(M)$ 个义元的正面情感强度值, M_Neg_n 为第 $S(M)$ 个义元的负面情感强度值。

2.2.2 使用 SentiWordNet 数据库检索得出包含输入义元的各个同义词集合

同义词集合是 SentiWordNet 的基本组成单元,一个同义词集合由意义相近的多个词语构成,用以表示一种语意概念。同时, SentiWordNet 中每个同义词集都有与之相对应的情感倾向强度值标注,如表 1 所示, Pos 表示正面情感强度值, Neg 表示负面情感强度值。一个英文词语可能同时存在于多个同义词集合中,表示其含有多种不同语意概念。定义包含义元 M 的同义词集的集合为 $S(M)$,则 $S(M) = \{s_1, s_2, \dots, s_k\}$,其中 K 表示同义词集的数量, $s_k (k = 1, 2, \dots, K)$ 为第 k 个包含输入义元 M 的同义词集。编程使用 SentiWordNet 数据库对输入义元进行检索即可得到相应的同义词集合正、负情感强度值。

表 1 SentiWordNet 中的同义词集示例
Table 1 Examples of synsets in SentiWordNet

同义词集和注解	Pos	Neg
{pride#1,plume#2,congratulate#3}: be proud of; "He prides himself on making it into law school"	0.250	0.500
{pride#1,pridefulness#1}: a feeling of self-respect and personal worth	0.625	0.000
{pride#2}: satisfaction with your (or another's) achievements; "he takes pride in his son's success"	0.375	0.250
{pride#3}: the trait of being spurred on by a dislike of falling below your standards	0.500	0.250
{pride#4}: a group of lions	0.000	0.000
{pride#5,superbia#1}: unreasonable and inordinate self-esteem (personified as one of the deadly sins)	0.625	0.250

2.2.3 计算这些同义词集的平均情感强度值得到每个义元的情感倾向强度值

经过上一步骤,已经可以取得同义词集的正、负情感倾向强度值,由式(3)、(4)可计算出义元的情感强度值。

$$M_Pos = \frac{1}{K} \sum_{k=1}^K s_Pos_k, \quad (3)$$

$$M_Neg = \frac{1}{K} \sum_{k=1}^K s_Neg_k, \quad (4)$$

其中 s_Pos_k 为第 k 个同义词集的正面情感强度值, s_Neg_k 为第 k 个同义词集的负面情感强度值。

2.2.4 计算各项义元的平均情感强度值得到中文词语的情感倾向强度值

将式(3)和(4)分别代入式(1)和(2)得

$$W_Pos = \sum_{n=1}^N \sum_{k=1}^K \frac{s_Pos_k}{NK_n}, \quad (5)$$

$$W_Neg = \sum_{n=1}^N \sum_{k=1}^K \frac{s_Neg_k}{NK_n}, \quad (6)$$

其中 K_n 表示输入词语的第 n 个义元所包含的同义词集数量,将从 SentiWordNet 中获取的相应的同义词集的正、负情感倾向值代入式(5)和(6)即可计算出给定中文词语的正、负情感倾向强度值。在完成全部词语的计算后,将词语的情感倾向强度值进行归一化处理,即得到 SLHS 中各词语的情感强度值。

2.3 SLHS 结构分析

在一般的极性词典中,词语的情感倾向没有区分情感强度,将词语当成具有单一情感成分的个体。从表2中的数据可知,和极性词典不同,SLHS 同时给出了词语的正面情感倾向值和负面情感倾向值。SLHS 中词语情感倾向权值的计算过程体现了词语情感分解的过程,将词语分解为表示为多个义元,以求更加准确地描述词语的复合情感成分。

表2 SLHS 中情感词示例

Table 2 Examples of emotional words in SLHS

情感词	义元	Pos	Neg
绝佳	{excellent}	1.000	0.000
精彩	{splendid, splendidly, wonderful, wonderfully}	0.599	0.063
尊贵	{honourable, respectable, respected}	0.708	0.042
悲伤	{heart sore, lamenting, sad, sorrowful, woeful}	0.026	0.729
腐臭	{decaying, stinking, smelly}	0.000	0.750

情感词典是文本情感分析系统中非常重要的资源。情感词典被广泛应用于文本分词、情感词提取和文本情感特征提取等任务中。本研究所构建的中文情感词典科学地解决了词语情感强度值的问题,较为准确地度量了中文词语的正、负情感词。

3 实验及结果分析

3.1 实验介绍

3.1.1 实验概述

为了验证 SLHS 应用于文本情感分类的效果,结合情感词典对文本内容进行特征提取,采用支持向量机(support vector machines, SVM)方法对微博文本进行情感分类,具体流程如图2所示。在实验中分别使用 SLHS 和 NTUSD 对文本特征进行抽取,进行对比分析,最后将输入文本的情感倾向分类为正面情感和负面情感。

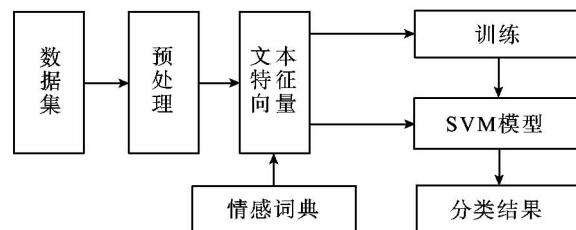


图2 实验方法流程图

Fig. 2 The flow chart of the experimental method

3.1.2 数据准备

实验数据来源于 NLP&CC2012 的中文微博情感分析评测的样例数据集。该数据使用 XML 文件存储,包含 20 个话题,共 2 173 条数据,其中正面情感文本数据 407 条,负面情感文本数据 1 766 条。本文随机抽取三分之二的的数据用于模型训练,剩下数据用于测试。测试数据共 724 条,其中正面文本数据 130 条,负面文本数据 594 条。

3.1.3 文本特征提取

在提取文本特征之前,先对文本进行预处理。文本预处理工作包括微博文本分词、文本情感特征选择和特征权值计算等。

文本分词是文本处理的基础工作,采用 ICTCLAS 对输入文本进行细粒度分词及词性标注。在完成分词和词性标注之后对分词所得词语进行词性统计,包括名词、形容词、叹词、拟声词等出现的次数。在完成词性统计后,将分词结果和否定词表、转折词表进行匹配,从而确定文本出现否定词和转折词的数量。所用否定词和转折词如表3所示。

表3 常用否定词和转折词列表

Table 3 Word list of the commonly used negative words and turning words

词语类别	常用词
否定词	不、弗、毋、勿、未、否、非、无、莫
转折词	却、但是、可是、然而、而、偏偏、只是、不过、至于、致、不料、岂知

结合 SLHS 中词语同时存在正面和负面情感强度值的情况,在选取文本情感特征时主要考虑了词语的正面和负面情感强度特征、词性计数、符号计数、否定词计数和转折词计数等,见表4。

表4 文本情感特征
Table 4 Sentiment features of text

类型	特征内容	描述
情感词典	情感词正面强度值、负面强度值、情感词数量	SLHS 情感词: 24 130 个,其中正面情感词 10 269 个,负面情感词 13 861 个
标点符号	是否存在问号、感叹号	存在用 1 表示否则为 0
词性统计	名词数量、动词数量、形容词数量、连词数量、助词数量、叹词数量、语气词数量、拟声词数量	使用中科院计算所汉语词性标记集作为分类标准
否定词统计	否定词数量	根据否定词表进行统计
转折词统计	转折词数量	根据转折词表进行统计

在选定文本情感特征后,将各个特征转化为数量表示即可获得文本的特征向量。需要指出的是,由于 NTUSD 属于极性情感词典,没有具体的情感强度值。所以本文在使用 NTUSD 进行情感词强度提取时,将该词典的正面情感词的正面强度值设置为 1,负面强度值设置为 0,负面情感词反之。

3.1.4 SVM 分类器

采用 SVM 构建文本情感分类器。SVM 是一种监督式学习的方法,属于一般化线性分类器。它能够同时最小化经验误差与最大化几何边缘区,因此 SVM 也被称为最大边缘区分类器。本文使用的 SVM 工具是台湾大学林智仁等人开发设计的 SVM 模式识别与回归的软件包^[25]。

3.2 实验结果分析

3.2.1 SLHS 与其他情感词典比较实验

由于 SLHS 使用无监督方法自动构建,其中的部分情感词强度值与真实情感可能存在误差。因此有必要对 SLHS 进行修正,然后验证其有效性。使用 NTUSD 和清华大学褒贬义词典^[21]对 SLHS 进行校正。由于引入比较的两个情感词典都属于极性词典,所以在实验时要对 SLHS 中的词语添加极性标记。将 SLHS 中词语的正负情感强度值进行比较,当正面情感强度值大于负面情感强度值时将改词语标记为褒义词,反之则为贬义词。然后分别遍历 NTUSD 和褒贬义词典,将其词语逐个与 SLHS 中的

词语进行比较,实验结果如表5所示。从实验数据可以看出相对于人工干预标注的极性词典,SLHS 的情感强度值准确性还有一定的差距。

表5 SLHS 比较分析实验结果数据
Table 5 SLHS comparative analysis result data

输入词典	褒义词数量	贬义词数量	准确率/%
NTUSD	2 647	7 741	77
褒贬义词典	5 565	4 467	72

3.2.2 SLHS 情感倾向性强度值校正

接下来,本研究将对 SLHS 的情感倾向性强度值进行校正调整,以提高 SLHS 词语情感强度的准确度。词典校正实验步骤包括:首先按照上文所述方法将 SLHS 中的词语进行极性标注;

然后遍历 NTUSD 和褒贬义词典,将其情感词语极性与 SLHS 进行比较;若极性不同,则将 SLHS 中该词语的正、负情感强度值互换。

在完成词典校正后,SLHS 的情感倾向强度值的准确度将得到很大程度的提高。通过词典校正,SLHS 能够学习输入词典的词语情感知识,在调整权值准确性的同时保留了词语情感倾向的内部差异性。

3.2.3 SVM 类型及核函数选择

在完成数据集的情感特征提取后,选择不同类型的 SVM 以及核函数进行实验。使用五折交叉验证法,将数据集平均分为 5 份,轮流使用其中 4 份作为训练集,1 份作为测试集进行实验。实验结果如表6所示,从分类准确率数据可知,在选择 C-SVC 类型的 SVM 并且使用 RBF 核函数时获得了较好的分类效果。

表6 不同类型 SVM 分类准确率
Table 6 Classification accuracy of different types of SVM %

one-class		C-SVC	
RBF	sigmoid	RBF	sigmoid
51.17	48.27	83.06	70.55

3.2.4 SLHS 和 NTUSD 对比实验

分别使用 SLHS 和 NTUSD 对数据集进行特征向量提取时,生成两组不同的实验数据。抽取数据集的三分之二作为训练数据,剩下三分之一作为测试数据。实验时设置 SVM 类型为 C-SVC 并且使用 RBF 核函数。

由表7所示数据可知,使用 SLHS 进行文本情感特征提取时取得了较优的实验结果,验证了该情感词典的有效性。

表 7 微博文本情感分类结果数据

Table 7 Sentiment classification accuracy of micro blog texts %

情感词典	准确率	召回率	F-Measure
SLHS	83.84	57.54	68.24
NTUSD	83.56	54.70	66.12

4 结 束 语

本研究提出基于 HowNet 和 SentiWordNet 的情感词典构建方法,将中文词语进行义元分解,通过 SentiWordNet 计算义元的情感倾向值分别得到词语的正面、负面情感倾向值。在使用该方法构建中文情感词典 SLHS 后,提出词典校正方法对词语的情感倾向值进行修正,并将其应用到微博文本情感分析任务中,通过实验验证了该情感词典的有效性。

在词典的构建过程中,存在部分较为生僻的中文词语无法自动进行义元分解的问题。同时,词语情感强度值的计算也使用了较为简单的均值法。进一步的研究工作应考虑寻找更为合理的义元分解和情感强度值计算方法。

参考文献:

- [1] 赵妍妍,秦兵,刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.
ZHAO Yanyan, QIN Bing, LIU Ting. Sentiment analysis [J]. Journal of Software, 2010, 21(8): 1834-1848.
- [2] HATZIVASSILOPOULOS V, MCKEOWN K R. Predicting the semantic orientation of adjectives[C]//Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics. Madrid: Association for Computational Linguistics, 1997: 174-181.
- [3] WIEBE J M. Learning subjective adjectives from corpora [C]//Proceedings of the National Conference on Artificial Intelligence. Austin: AAAI Press, 2000: 735-740.
- [4] RILOFF E, WIEBE J M, WILSON T. Learning subjective nouns using extraction pattern bootstrapping [C]//Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. Edmonton, Canada: Association for Computational Linguistics, 2003: 25-32.
- [5] BARONI M, VEGNADUZZO S. Identifying subjective adjectives through Web-based mutual information [C]//Proceedings of KONVENS 2004. Vienna: University of Vienna, 2004: 17-24.
- [6] MOILANEN K, PULMAN S. The good, the bad, and the unknown: morphosyllabic sentiment tagging of unseen words [C]//Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers. Stroudsburg: Association for Computational Linguistics, 2008: 109-112.
- [7] TURNEY P, LITTMAN M L. Measuring praise and criticism: inference of semantic orientation from association [J]. ACM Trans Information Systems, 2003, 21(4): 315-346.
- [8] YANG A M, LIN J H, ZHOU Y M, et al. Research on building a Chinese sentiment lexicon based on SO-PMI [J]. Applied Mechanics and Materials, 2013, 263: 1688-1693.
- [9] READ J. Recognising affect in text using pointwise-mutual information[D]. Brighton: University of Sussex, UK, 2004.
- [10] MILLER G A, BECKWITH R, FELLBAUM C, et al. Introduction to wordnet: an on-line lexical database [J]. International Journal of Lexicography, 1990, 3(4): 235-244.
- [11] KAMPS J, MARX M, MOKKEN R J, et al. Words with attitude [C]//Proceedings of the First International Conference on Global WordNet, CIIL. Mysore, India: ACL-SIGLEX, 2002: 332-341.
- [12] KIM S M, HOVY E. Determining the sentiment of opinions [C]//Proceedings of the 20th International Conference on Computational Linguistics. Switzerland: Association for Computational Linguistics, 2004: 1367-1373.
- [13] ESULI A, SEBASTIANI F. Sentiwordnet: a publicly available lexical resource for opinion mining [C]//Proceedings of LREC. Genoa-Italy: LREC, 2006: 417-422.
- [14] BACCIANELLA S, ESULI A, SEBASTIANI F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining [C]//Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC'10). Valletta, Malta: European Language Resources Association, 2010: 2201-2204.
- [15] 董振东. 语义关系的表达和知识系统的建造 [J]. 语言文字应用, 1998, 3(3): 76-82.
DONG Zhendong. Semantic expression and the construction of knowledge systems [J]. Applied Linguistics, 1998, 3(3): 76-82.
- [16] 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.
ZHU Yanlan, MIN Jin, ZHOU Yaqian, et al. Semantic orientation computing based on HowNet [J]. Journal of Chinese Information Processing, 2006, 20(1): 14-20.
- [17] 郭先珍. 常用褒贬义词语详解词典 [M]. 北京: 商务印书馆, 1999.
- [18] 王国璋. 汉语褒贬义词语用法词典 [M]. 北京: 华语教学出版社, 2001.

- [19] 史继林. 褒义词词典[M]. 成都: 四川辞书出版社, 2005.
- [20] 史继林. 贬义词词典[M]. 成都: 四川辞书出版社, 2005.
- [21] LI J, SUN M. Experimental study on sentiment classification of Chinese review using machine learning techniques [C]//Natural Language Processing and Knowledge Engineering. Beijing: IEEE, 2007: 393-400.
- [22] 柳位平, 朱艳辉, 栗春亮, 等. 中文基础情感词词典构建方法研究[J]. 计算机应用, 2009, 29(11): 2882-2884.
- LIU Weiping, ZHU Yanhui, LI Chunliang, et al. Research on building Chinese basic semantic lexicon [J]. Journal of Computer Applications, 2009, 29(11): 2882-2884.
- [23] 徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造[J]. 情报学报, 2008, 27(2): 180-185.
- XU Linhong, LIN Hongfei, PAN Yu, et al. Constructing the affective lexicon ontology [J]. Journal of The China Society For Scientific and Technical Information, 2008, 27(2): 180-185.
- [24] 张成功, 刘培玉, 朱振方, 等. 一种基于极性词典的情感分析方法[J]. 山东大学学报: 理学版, 2012, 3: 47-50.
- ZHANG Chenggong, LIU Peiyu, ZHU Zhenfang, et al. A sentiment analysis method based on a polarity lexicon [J]. Journal of Shangdong University: Natural Science, 2012, 3: 47-50.
- [25] CHANG C C, LIN C J. LIBSVM: a library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.
- (编辑: 陈燕)

(上接第20页)

- [17] 何传江, 李梦, 詹毅. 用于图像分割的自适应距离保持水平集演化[J]. 软件学报, 2008, 19(12): 3161-3169.
- HE Chuanjiang, LI Meng, ZHAN Yi. Adaptive distance preserving level set evolution for image segmentation [J]. Journal of Software, 2008, 19(12): 3161-3169.
- [18] ZHANG K H, SONG H H, ZHANG L. Active contours driven by local image fitting energy [J]. Pattern Recognition, 2010, 43(4): 1199-1206.
- [19] LI C M, GAO C, GORE J, et al. Implicit active contours driven by local binary fitting energy [C]//Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minnesota: IEEE CS, 2007: 1-7.
- [20] LIU B, CHENG H D, HUANG J, et al. Probability density difference-based active contour for ultrasound image segmentation [J]. Pattern Recognition, 2010, 43(1): 2028-2042.
- (编辑: 陈燕)