情感词典自动构建方法综述

王 科 [1] 夏 睿 [1]

摘 要 情感词典作为判断词语和文本情感倾向的重要工具, 其自动构建方法已成为情感分析和观点挖掘领域的一项重要研究内容. 本文整理了现有的中、英文情感词典资源, 同时分别从知识库、语料库、以及两者结合的角度, 归纳现有英文和中文情感词典的构建方法, 分析了各种方法的优缺点, 并总结了情感词典构建中的若干难点问题. 之后, 我们回顾了情感词典性能评估方法及相关评测竞赛. 最后总结了情感词典构建任务的发展前景以及一些亟需解决的问题.

关键词 自然语言处理, 情感分析, 观点挖掘, 情感词典, 词典构建

引用格式 王科, 夏睿. 情感词典自动构建方法综述. 自动化学报, 2016, **42**(4): 495−511

DOI 10.16383/j.aas.2016.c150585

A Survey on Automatical Construction Methods of Sentiment Lexicons

WANG Ke[1] XIA Rui[1]

Abstract Sentiment lexicon is an important tool of identifying the sentiment polarity of words and texts. How to automatically construct sentiment lexicons has become a research topic in the field of sentiment analysis and opinion mining. We review the existing sentiment lexicon construction methods, for both English and Chinese languages, from the perspectives of lexicons, corpus, and the combination of the two. We analyze the advantages and disadvantages of each method and point out some special problems in sentiment lexicon construction. We furthermore summarize the evaluation methods and review several competitions related to sentiment lexicon construction. Finally, we discuss the prospect of sentiment lexicon construction, and present some problems that remain to be solved.

Key words Natural language processing, sentiment analysis, opinion mining, sentiment lexicon, lexicon construction

Citation Wang Ke, Xia Rui. A survey on automatical construction methods of sentiment lexicons. *Acta Automatica Sinica*, 2016, **42**(4): 495−511

随着 Web 2.0、社交媒体和电子商务的兴起, 互联网用户由单纯的信息接受者, 开始向信息发布的参与者转变, 互联网上出现了大量的用户评论文本, 这些文本信息通常表达了用户的主观观点和情感. 通过对这些包含用户情感的评论文本的挖掘与分析, 人们能快速有效地获取所需要的信息. 然而数据指数级的增长速度, 依靠人工进行评论分析需要耗费大量的人力物力, 在这种背景下, 计算机自动文本情感分析技术应运而生. 情感词典作为文本情感分析的重要工具, 同时作为情感分析的基础任务, 其构建问题也逐渐成为自然语言处理领域的研究热点之一.

收稿日期 2015-09-14 录用日期 2016-01-23
Manuscript received September 14, 2015; accepted January 23, 2016
国家自然科学基金 (61305090), 软件新技术国家重点实验室开放基金, 江苏省自然科学基金 (BK2012396) 资助
Supported by National Natural Science Foundation of China (61305090), Open Fund of the State Key Laboratory for Novel Software Technology, and Jiangsu Provincial Natural Science Foundation of China (BK2012396)
本文责任编委 张民
Recommended by Associate Editor ZHANG Min
1. 南京理工大学计算机科学与工程学院 南京 210094
1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094

1 概述

文本情感分析和观点挖掘 (Sentiment analysis and opinion mining) 是自然语言处理领域的一个重要研究方向. 它是利用计算机对主观文本中所包含的情感、态度, 进行分析、挖掘、总结和归纳的一项技术. 相对于客观文本, 主观文本包含了用户个人的想法或态度, 而客观文本更多的是对事物本身属性的客观描述, 并不反映用户的任何观点. 客观文本描述的是客观事实, 不存在用户情感, 故不需进行情感分析; 主观文本是用户群体对某产品或事件, 从不同的角度、不同的用户需求和用户体验去分析评价的结果, 这些评价信息具有主观能动性和多样性, 具有情感分析意义.

目前, 情感分析技术可以分为两类. 一类是基于机器学习的方法, 通过大量有标注、无标注的主观语料, 使用统计机器学习算法, 抽取特征, 再进行文本情感分析. 另一类是基于情感词典的方法, 根据情感词典所提供的词的情感倾向性, 从而进行不同粒度下的文本情感分析.

按照文本的粒度, 可以将情感分析任务归纳为词、短语、属性、句子、篇章等多个级别. 情感词典

表 1 常见的通用情感词典简介

Table 1 Common sentiment lexicon introduction

语言	词典名	说明
英文	SentiWordNet	英文中最为著名的一款情感词典, 它基于 WordNet, 为 WordNet 中每一个同义词集分别给出正、负和客观情感得分.
	General Inquirer	General Inquirer 被认为是最早的一款情感词库兼计算机情感分析程序, 其情绪词来源于《哈佛词典 (第 4 版)》和《拉斯韦尔词典》, 按照情感正负性对词汇进行分类.
	Opinion Lexicon	Bing Liu 发布的一款英文情感词典, 不仅包含情感词, 还包含了拼写错误、语法变形, 俚语以及社交媒体标记等信息.
中文	HowNet 情感词典	董振东和董强建立的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, 其中包括情感分析用词语集.
	DUTIR 情感词汇本体库	大连理工大学信息检索研究室整理和标注的一个中文本体资源. 该资源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.
	NTUSD	来源于台湾大学自然语言处理实验室的中文情感极性词典.

在不同粒度的情感分析任务中扮演着不同的角色, 比如:

1) 词/短语级别情感分析任务中, 对词或短语进行情感倾向判断的过程, 基本等价于情感词典的构建. 情感词典具有的词/短语分析能力, 能够快速地完成对小粒度文本的情感分析.

2) 句子/篇章级情感分析任务中, 句子的情感极性可以通过对褒贬情感词的得分累加或数量比较进行判断[1], 也可以构建统计机器学习方法, 利用情感词典作为分类器特征, 进行文档分类[2].

3) 属性级情感分析任务, 需要先找出属性词对应的情感词; 针对该情感词, 使用词/短语的方法判断其情感, 就是该属性对应的情感倾向. 此外, 有些词的情感极性依赖于所属领域和上下文, 我们将在第 5.2 节进行特别总结.

综上所述, 基于一份全面而精确的情感词典进行情感分析, 通常能获得较高的准确率. 情感词典除了能够准确判断词/短语的情感极性外, 还可有效地辅助属性级、句子/篇章级的情感分析任务. 多数主观文本均包含情感词, 情感词是情感分析的重要依据, 所以找出情感词, 正确判断其情感极性, 从而构造高准确率和覆盖率的情感词典, 具有至关重要的意义.

由于情感词典在情感文本分析中的重要地位, 其自动构建方法是近几年自然语言处理领域的一个研究热点. Liu 在文献 [3] 第七章对情感词典构建方法进行了综述. 与之相比, 本文具有下列三个特点: 一是本文从方法论的角度对每类方法进行了深入的总结和归类; 二是本文除了介绍英文词典构建方法之外, 还就每类方法特别总结了中文词典构建的研究现状; 三是除了相关方法描述之外, 本文对现有的一些中英文情感词典资源进行了整理, 并总结了词典性能的评估方法, 以及部分国内外著名的评测竞赛.

本文章节安排如下. 第 2 节总结了情感词典的现有资源. 第 3 节, 我们回顾了英文情感词典自动构建的现状. 第 4 节, 我们着重阐述了一些中文情感词典构建现状. 第 5 节总结了情感词典构建中的难点问题. 第 6 节介绍了词典性能评估方法和相关评测. 最后对情感词典构建工作做出了展望.

2 现有情感词典资源

目前大部分的通用情感词典, 是通过人工构建的. 人工构建方法主要是通过阅读大量相关语料或借助现有词典, 人工总结出具有情感倾向的词, 标注其情感极性或强度, 构成词典. 但是, 这种方法花费的代价很大. 表 1 总结了中英文常见的通用情感词典. 情感词典的人工构建工作, 国外起步较早, 其中比较著名的是 SentiWordNet[1]. 它根据 WordNet[2], 把释义一致的词合并在一起, 给与相应的正面得分和负面得分, 用户既可以精确定位情感得分, 也可以用一个词的正负平均得分表示该词的情感得分. General Inquirer (GI)[3] 被认为是最早的一个情感词库, 早期包含 1 915 个褒义词, 2 291 个贬义词, 它还给每个词语按情感极性、强度、词性等属性打上不同的标签, 便于在情感分析中能够灵活应用. Hu 等[1] 提供的一份情感词典 Opinion Lexicon[4] 在业内也被广泛使用, 该词典包含的词较多, 其中褒义词 2 006 个, 贬义词 4 783 个.

[1] http://sentiwordnet.isti.cnr.it/
[2] http://wordnet.princeton.edu/wordnet/download/
[3] http://www.wjh.harvard.edu/ inquirer/
[4] https://www.cs.uic.edu/ liub/FBS/sentiment-analysis.html#lexicon

表 2 基于知识库的构建方法概述

Table 2 Summary of the lexicon-based approach

方法	概述	参考文献
词关系扩展法	利用已知褒贬的种子词集, 在语义知识库中寻找同义词、反义词等词间关系, 进行扩展, 去噪后得到一份通用情感词典.	(Hu 等, 2004)[1], (Strapparava 等, 2004)[4], (Neviarouskaya 等, 2011)[5], (Kim 等, 2004)[6], (Blair-Goldensohn 等, 2008)[7]
迭代路径法	计算知识库中两个词通过同义词 (或其他关系) 迭代到彼此需要的次数, 判断两个词极性的相似性, 从而确定未知词的极性.	(Kamps 等, 2004)[8], (Hassan 等, 2011)[9], (柳位平等, 2009)[2]
释义扩展法	将同义词的释义也作为训练语料, 或寻找词和释义中词的关系.	(Andreevskaia 等, 2006)[10], (Baccianella 等, 2010)[11], (Esuli 等, 2007)[12]

规范的中文情感词典相对缺乏, 最早也是最普遍传播的是知网 (HowNet) 提供的情感分析用词语集[5], 其中包含了中英文褒贬的评价词、情感词, 中文褒贬词分别为 4 569 个、4 370 个. 除了情感词外, HowNet 还有类似 WordNet 的特点, 它构建了一个词与词之间的大型关系网络. 大连理工的情感词汇本体库[6]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. 它是在国外比较有影响的 Ekman 的 6 大类情感分类体系的基础上构建的, 并且在 Ekman 的基础上, 该词汇本体加入情感类别 “好”, 对正面情感进行了更细致的划分, 最终词汇本体中的情感共分为 7 大类 21 小类, 情感强度分为 1、3、5、7、9 五档, 9 表示强度最大, 含 11 229 个褒义词、10 783 个贬义词. 台湾大学构建的情感词典 NTUSD[7], 也是一个比较常用的情感词典, 它包含 2 810 褒义词, 8 276 个贬义词, 与 HowNet、情感词汇本体库一起构成了目前中文处理中最常用的三个开放情感词典.

3 情感词典自动构建方法

基于人工构建的情感词典虽然具备较好的通用性, 但是在实际使用中, 难以覆盖来自不同领域的情感词, 领域适应性较差. 同时, 人工情感词典构建需要耗费大量的人力物力. 因此, 学术界更多地聚焦于情感词典的自动构建工作, 这也是本文总结的重点. 情感词典自动构建方法, 主要有三类: 基于知识库的方法、基于语料库的方法、以及知识库和语料库相结合的方法.

3.1 基于语料库的方法

一些语种具有完备、开放的语义知识库 (比如英文的 WordNet), 通过挖掘其中词与词之间的关系 (如同义、反义、上位以及下位关系等), 就能构建出一部通用性较强的情感词典. 我们在表 2 中将这类方法又归为三类, 分别是: 词关系扩展法、迭代路径法和释义扩展法.

3.1.1 词关系扩展法

基于 WordNet 的同义、反义词关系扩展法的基本流程: 先分别人工构建少量褒义、贬义形容词集合, 然后在 WordNet 语义知识库中, 查找它们的同义词与反义词来扩大这两个集合, 将同义词放入种子词所在集合, 反义词放入另一集合. 通过循环迭代, 构建一个具有一定规模的形容词词库. 最后人工整合, 筛除错分的词, 构成情感词典. Hu 等[1] 在总结用户评论时, 采用了这个方法. 但是, 情感词不仅仅只有形容词, 一些名词或者动词 (比如 “beauty”), 甚至副词也可能包含情感倾向. Strapparava 等[4] 在形容词的基础上, 还加入了名词和一小部分动词和副词作为初始词典, 使得到的情感词典更加全面. Neviarouskaya 等[5] 还引入情感得分和情感权重, 使得篇章级文本情感值计算更加合理.

大型知识库中, 词与词之间的关系错综复杂, 一些词在经过若干次迭代之后, 可能会迭代到它的反义词, 比如 “good (好)” 在经过若干次同义词迭代之后, 有可能会得到极性完全相反的词 “bad (坏)” [good, sound, heavy, big, bad], 这给词典构建工作带来了一些困难. 为了排除这些词汇, Kim 等[6] 在词典构建过程中使用贝叶斯分类器, 分别计算该词属于正面和负面的概率, 从而确定其极性. Blair-Goldensohn 等[7] 添加了一个中性词集合, 再根据同义、反义关系, 结合图方法进行扩展, 若扩展得到的词在中性集中, 则不扩展, 从而提高了候选词集合的准确率.

3.1.2 迭代路径法

由于语义知识库中, 词间的网状关系, 任意两个词之间都可能存在着千丝万缕的联系. Kamps 等[8] 认为, 意思越相似的两个词, 它们通过同义词迭代所

[5] http://www.keenage.com/html/e_index.html
[6] http://ir.dlut.edu.cn/EmotionOntologyDownload
[7] http://www.datatang.com/data/44317

需的次数就越少. 他们使用两个词相互迭代所需的次数衡量两者的相似性, 并用于计算候选词的情感倾向:

$$SO(w) = \frac{d(t, w^-) - d(t, w^+)}{d(w^+, w^-)} \quad (1)$$

其中, $d(t, w)$ 是指未知极性的词 t 通过同义词迭代到已知极性的词 w 所需的最少次数, w^+ 和 w^- 分别代表褒义词和贬义词. 计算迭代次数的思想, 与下文介绍的点互信息类似, 都可以衡量两个词之间的相似性. 不同的是, 逐点互信息 (Pointwise mutual information, PMI) 是基于语料统计计算两个词之间的共现信息作为相似性度量. Hassan 等[9] 根据 WordNet 构建一幅词间关系图, 结合已知情感极性的种子词集 S, 先从任意单词 w_i (不属于 S) 开始, 按照一定转移概率移动, 直到遇到 w_k (属于 S). 反复多次, 分别计算 w_i 到褒义、贬义种子词的平均移动次数, 以次数少的确定为词 w_i 的情感极性.

3.1.3 释义扩展法

一些知识库还给出了词的释义. 若将知识库中的褒义词和贬义词视为两个类别, 那么这些词的释义便可以看成是一个二分类的已标注语料库. Andreevskaia 等[10] 同时使用 WordNet 中的词间关系和释义进行扩展. 先标注一部分种子词, 对其利用词关系进行扩展, 再遍历 WordNet 中的所有释义, 对释义中含有种子词的单词, 进行过滤消歧之后构成情感词典. Baccianella 等[11] 使用半监督机器学习, 先通过 WordNet 扩展初始标注的种子情感词集和客观词集. 然后使用词的释义作为训练集, 构造一个三类 (褒义、贬义、客观) 分类器, 来判断未知情感的释义, 以确定其对应的同义词集中所有词的极性, 最后使用随机游走 (Random-walk) 确定词的得分, 形成情感词典. Esuli 等[12] 认为同义词的释义通常会包含同样极性的其他词, 如果同义词集合的释义中包含另一个同义词集的词, 则认为这两个集合有联系.

综上所述, 基于知识库的方法不需要依赖于语料库, 仅依靠一个完备的语义知识库, 就能较快地得到情感词典, 并且该词典能覆盖大部分语料中的情感词, 通用性强, 在对精度要求不高的情况下, 该方法较为实用. 但是对于英语以外的大部分语言, 类似 WordNet 的语义知识库相对缺乏, 无法使用这类方法. 即便使用基于知识库的方法, 由于知识库内部、词语之间的复杂关系, 随着迭代次数增多, 准确率会下降, 需要辅以其他方法来确定词的极性. 其次, 基于知识库的方法通常只能获得一个通用的情感词典. 然而, 情感知识库存在领域适应问题 (我们将在第 5.1 节详细叙述), 同一词语在不同领域 (甚至不同主题) 下, 可能表达出不同的情感. 此时, 基于知识库的方法就不再适合.

3.2 基于语料库的方法

通用情感词典能满足大部分情感分析任务的需求. 然而, 为了解决某些特定领域的情感分析任务, 或者为了提高情感分析的精度, 需要使用领域情感词典. 领域情感词典是根据某领域大量语料构建的情感词库, 它具有领域特定、时效性高等特点. 由于领域众多, 且新词不断涌现, 这些词通常不能被通用词典及时收录. 同时, 由于各个领域的情感词有所差异, 特定领域的情感词典用于另一领域时, 词典评估的召回率通常较低, 所以鲜见针对某些特定领域的人工情感词典发布, 目前大多还是基于语料库进行领域情感词典构建.

语料相对于语义知识库而言, 其优点是容易获得且数量充裕. 基于语料库的方法能够从语料中自动学习得到一部情感词典, 可以节省大量的人力、物力, 同时, 在不同领域的语料上可以得到领域特定的情感词典, 更加具有实用意义. 我们总结了基于语料库的情感词典构建方法, 并将其大致分为两类: 连词关系法和词语共现法, 如表 3 所示.

表 3 基于语料库的情感词典方法概述

Table 3 Summary of the corpus-based approach

方法	概述	参考文献
连词关系法	通常, 文本中转折词的出现会伴随着情感极性的改变, 而一般的连词 (如并列、承接词) 则不会导致极性改变. 该类方法利用连词信息, 判断文本片段的情感, 从而判断片段中情感词的情感.	(Hatzivassiloglou 等, 1997)[13], (Kanayama 等, 2006)[14], (Huang 等, 2014)[15], (王科等, 2015)[16], (Xia 等, 2014)[17]
词语共现法	单词 A、B 若经常在一起出现, 则可认为它们之间存在一定相关性. 该类方法以此来判断两个情感词极性的相似程度.	(Huang 等, 2014)[15], (Church 等, 1990)[18], (Turney, 2001)[19], (Turney, 2002)[20], (Turney, 2003)[21], (Krestel 等, 2013)[22], (Tai 等, 2013)[23], (Wawer, 2012)[24], (Glavaš 等, 2012)[25], (Bollegala 等, 2011)[26], (Velikovich 等, 2010)[27], (阳爱民等, 2013)[28], (魏志生等, 2011)[29]

3.2.1 连词关系法

基于语料库构建情感词典的方法很多, 其中最经典的就是利用语句中的连词来判断前后词语的情感极性关系. Hatzivassiloglou 等[13] 详细总结了英语中的语言规则和连接模式, 并通过大量实验数据证明了连词前后词的极性关系, 之后基于语料库和情感种子词集, 识别形容词的情感指向. 具体地, 首先提取出连词连接的形容词, 标注其中高频词的极性, 根据形容词对在不同连词下出现的次数, 使用 log 线性回归模型来确定连词前后的两个词具有相同还是相反的情感极性, 接着使用聚类算法产生褒、贬两个词集, 最后基于以下目标函数来调整这两个集合:

$$\Phi(p)=\sum_{i=1}^{2}\left(\frac{1}{|C_i|}\sum_{x,y\in C_i,x\neq y}d(x,y)\right) \tag{2}$$

其中, $d(x,y)$ 表示词 x, y 的距离, 通常同义词间的距离较小, 反义词间的距离较大. $|C_i|$ 是聚类产生的词集的大小, Φ 是两个集合内词间距的总和, 值越小, 说明聚类效果越好. 移动某个词 (如从 C_1 到 C_2), 若 Φ 值减小, 说明集合间总距离减小, 移动的单词与 C_2 中的单词距离更接近, 则移动; 若变大, 则不动. 最后根据划分, 确定词集极性. Kanayama等[14]先利用规则模式 (比如 “I think + v”、“not + v”) 和句法分析的结果抽取情感词, 统计全语料中, 上下文情感的一致性准确率和密度, 设置阈值, 筛选情感词, 再针对句内和句间情感进行一致性判别, 认为连续出现的单词具有相同的极性, 只有遇到转折词的时候 (比如 “but”), 情感极性才会反转, 以此判断情感词的极性. Huang 等[15] 利用连词判断单词间的极性关系, 并结合单词形态上的否定形式 (如: “X” 和 “un$X$”), 构建情感极性约束矩阵, 再利用逐点互信息 (Pointwise mutual information, PMI), 判断单词的情感极性.

连接关系法依赖连词判断前后文本的情感极性变化, 以此判断其中情感词的极性变化, 故该类方法主要适用于评论等主观性较强, 且句子间有情感变化的语料, 如: 商品评论, 含有明显的针对商品属性的褒贬评价. 实验证明, 即使是最简单的连接关系法, 也能在领域语料上表现出比通用词典更好的性能[16]. 但是, 连接关系法在构建领域情感词典时, 需事先得到候选情感词集, 再针对候选词进行褒贬分类. 语料中通常有很多情感词, 上述方法通常采用形容词作为候选词集, 然而情感词典可能包括动词和名词, 甚至副词, 我们需要先把这些带修饰性的、有情感极性的词找出来, 再利用相应算法确定极性.

3.2.2 词语共现法

逐点互信息 (Pointwise mutual information, PMI)[18] 是信息论和自然语言处理中的一个基本概念, 它常被用来衡量两个词间的独立性. 其计算方式如式 (3) 所示:

$$pmi(x,y)=\log\frac{p(x,y)}{p(x)p(y)} \tag{3}$$

其中, $p(x,y)$ 表示词 x 和 y 一起出现的概率, $p(x)$ 表示词 x 出现的概率, $p(y)$ 表示词 y 出现的概率. $pmi(x,y)$ 表示词 x 和 y 共现程度, 值越大, 两者共现越频繁, 独立性越小, 两者关系越紧密.

仅有 PMI, 只能用来判断两个词的共现程度, 还不足以用来判断一个词的极性. Turney[20] 使用词与正面、负面种子词的紧密程度, 来判断一个词的情感倾向 (Sentiment orientation, SO). 其计算方式如式 (4) 所示:

$$SO(w)=pmi(w,w^+)-pmi(w,w^-) \tag{4}$$

其中, w 是待确定极性的情感词, w^+ 和 w^- 分别表示正面和负面种子词. 若 SO 值大于阈值, 说明词跟正面词更紧密, 则其为正面词的概率比较大, 反之则为负面词的概率较大, 以此来确定词的极性.

除了 PMI, 情感倾向的计算还可以借助其他统计模型得到. Turney 等[21] 采用一个词与其邻近词的情感趋于一致的思想, 同时结合了潜在语义分析 (Latent semantic analysis, LSA), 挖掘文档中潜在的信息. 通过计算单词的 SO-LSA, 从大量语料中构建词典. 其计算方法如式 (5) 所示:

$$\begin{aligned}SO\text{-}LSA(w)=&\sum_{w^+\in Pwords}lsa(w,w^+)-\\&\sum_{w^-\in Nwords}lsa(w,w^-)\end{aligned} \tag{5}$$

其中 w 为待确定极性的词, w^+ 和 w^- 分别表示正面和负面种子词. $Pwords$ 和 $Nwords$ 分别表示正面词集和负面词集. Turney[20] 基于 AltaVista 搜索引擎的 NEAR 操作, 分别统计每个词与 “excellect” 和 “poor” 的共现数量 (间隔不超过 10), 并基于以下公式计算情感倾向. 此时情感倾向 SO 的计算公式为

$$\begin{aligned}&SO(w)=\\&\log\left(\frac{hits(w\,NEAR\text{ “excellent”})hits(\text{“poor”})}{hits(w\,NEAR\text{ “poor”})hits(\text{“excellent”})}\right)\end{aligned} \tag{6}$$

其中, $hits$ 即搜索引擎返回的共现数量. 值得提起的是, 这里的 SO 与上文我们给出的 PMI 和 SO 的计算方法[20] 是等价的.

后续不少学者对该方法加以利用, 提出了其他词共现信息 (或相似性) 计算方法. Krestel 等[22] 首先利用 LDA (Latent dirichlet allocation) 将语料分为几个主题, 用评论的星数确定评论的情感倾向, 之后将主题模型与 sigmoid 函数引入到 PMI 的计算中, 得到情感词的情感得分来判断极性. Tai 等[23] 使用了二阶共现点互信息 (Second order co-occurrence PMI, SOC-PMI) 来判断短文本语料中词的共现关系. SOC-PMI 是针对短文本的一种处理方法, 假定词 A、B 一起出现, 词 A、C 一起出现, 则通过 SOC-PMI, 我们可以得到词 B、C 的相似关系. Wawer[24] 基于 Turney 的思想, 认为褒贬种子词的选取对结果有较大的影响, 于是采用了自动生成的方式, 通过使用搜索引擎, 依据部分固定模式检索, 获取语料库, 从中获得候选词, 对其构建词频矩阵后进行 SVD 操作, 得到词间潜在的关联, 获得褒贬种子词, 用于 SO-PMI 计算. Bollegala 等[26] 利用本领域标注评论和目标领域未标注的语料, 先根据词性选取候选词; 对每个候选词, 使用极性特征表示: 首先将与候选词一起出现的所有词的极性标注为评论整体的极性, 并用词性代替词, 构成其特征. 接着根据候选词特征, 使用 PMI, 并计算候选词与已知情感词的相关性来判断其极性, 构成情感词典. Velikovich 等[27] 利用大量网页构建词典, 先根据频率和互信息筛选部分短句, 之后利用 N 元文法构建特征向量, 计算向量间的余弦值来衡量词间的相似性. 有些网页中的某些表格会指明这一行或一列单词的褒贬, Kaji 等[30] 基于这一思想从大量网页中获得正面和负面的词.

PMI 是衡量词间相关性的一种有效方法. 通过 PMI 值, 我们能够间接得获取单词间极性的相似性; 在基于图的方法 (本文第 3.3.1 节) 中, PMI 通常被用于构造词相关性矩阵, 再利用相似矩阵推导.

词共现法考虑的是词的相关性, 通用性较强, 适用于大部分语料, 包括新闻语料等非主观语料. 相比连词关系法, 词语共现法可以不指定候选情感词集. 但是共现法过分依赖于统计信息, 只考虑词语的共现情况, 而缺少对复杂语言现象 (如极性转移问题) 的建模, 使得结果会存在一定偏差. 如 “东西不错, 就是贵”, 未考虑转折关系, 使得 “不错” 和 “贵” 的极性会被认为是一致的. 而同样地, “他很和善, 一点都不挑剔 ”, 由于未考虑否定关系, 也会对 “和善”、“挑剔 ” 的情感关系作出错误的判断.

3.3 知识库与语料库结合的方法

基于语料库的方法能利用大规模语料, 无监督地获得领域特定的情感知识库, 但是与基于知识库的方法相比, 在准确率和通用性上尚有一定的欠缺. 所以, 目前很多方法将知识库和语料库结合起来. 知识库和语料库都提供了词与词之间的关系, 知识库主要提供词间标准的语义关系 (同义、反义、上位、下位等), 而语料库则主要提供两个词在语料中的关系, 包括位置信息、共现信息、情感保持、情感反转等. 利用现有知识库作为先验知识, 提供精确的种子词集, 并结合语料库中推导、约束信息, 得到其他未知情感词的极性, 构成一个更为完善的领域情感知识库. 我们将这一类方法归纳为几种常用的方法, 如表 4 所示.

3.3.1 关系图半监督法

在标注资源不足的情况下, 基于图的方法常用于情感词典构建. 通常, 该方法将词看作节点, 将词间的相似度作为两个连接节点的边的权重, 从部分已知极性的词开始, 推导未知极性的词的情感倾向. Peng 等[31] 先利用 WordNet 中的同义词、反义词, 对种子词进行扩展; 然后提取语料库中用 “and” 和 “but” 连接的形容词; 最后根据同义词关系和语料中 “and” 两种关系构建一张关系图, 根据反义词关系和语料中 “but” 关系构建一个限制矩阵, 使用限制的非负矩阵分解 (Constrained symmetric non-negative matrix factorization, CSNMF) 算法判断极性, 构成情感词典. Tai 等[23] 使用了类似的方法, 首先进行一些基础的自然语言预处理 (如词性标注、

表 4 知识库与语料库结合的构建方法

Table 4 Summary of the combined approach of lexicon and corpus

方法	概述	参考文献
关系图半监督法	以词与词之间的相似关系构建词间关系图, 利用已知极性的情感词, 结合图算法, 如标签传播算法, 推测其他情感词的极性.	(Esuli 等, 2007)[12], (Huang 等, 2014a)[15], (Tai 等, 2013)[23], (Glavaš 等, 2012)[25], (Peng 等, 2011)[31], (Rao 等, 2009)[32], (Xu 等, 2010)[33], (李荣军等, 2010)[34], (李寿山等, 2013)[35]
自举半监督法	为克服标注语料不足的问题, 先利用少量标注词确定文本片段的极性, 再结合抽取结果, 继续判断未知情感的文本片段.	(Volkova 等, 2013)[36], (Zhang 等, 2014)[37], (Weichselbraun 等, 2011)[38], (Gao 等, 2013)[39]
深度表示法	根据上下文, 训练词向量, 使得语义相近的词在向量空间上距离较近, 以此来判断词的情感极性.	(Tang 等, 2014a)[40], (梁军等, 2014)[41], (杨阳等, 2014)[42], (Tang 等, 2014b)[43]

词干化), 然后利用 WordNet 构建词与词之间的关系, 使用依存分析器[8], 获取语料库中单词的连接关系构造关系矩阵, 统计在一个窗口范围内词与词之间共现信息, 并计算二阶共现点互信息, 接着, 结合 WordNet、连接关系, 以及 SOC-PMI 构建一个相似度矩阵, 最后利用标签传播算法, 来判断未知极性情感词的情感. 同样使用标签传播方法的还有文献 [15], 它们先使用依存关系和现有通用情感词典, 提取语料中情感词, 同时提取标签样本中的高频情感词构造种子词集. 之后使用 PMI 构建相关性关系图, 同时抽取语料中的极性约束关系, 包括连词 "and" 和 "but" 等, 以及词在形态上的翻转, 如 "X" 和 "un$X$" 对文本情感造成的影响, 用于定义约束关系矩阵, 以推导出更多词间的相关关系. 最后使用标签传播算法, 得到其他词的情感倾向. Glavaš 等[25] 还使用了潜在语义分析, 获得语义层的相似性, 来帮助构建相似性矩阵, 同时使用 PMI、随机索引 (Random indexing, RI)、随机游走算法获得单词相关性, 之后根据构建的相似性矩阵, 结合种子词集, 使用 PageRank 判断其他词的情感倾向. Esuli 等[12] 通过释义, 构建同义词间关系后, 以同义词集为点, 以集合间关系为边, 使用 PageRank 算法, 分别得到褒、贬词的倾向性排名, 最后构成情感词典. Rao 等[32] 将三种基于图的半监督算法 (Mincut、randomized mincuts、label propagation) 做了对比, 分别利用这三种算法, 对未知情感极性的词进行标记, 然后利用已有资源和一定数量的种子词, 确定未知词的情感极性.

关系图半监督法同时利用知识库和语料库中的词关系构建相关矩阵. 该方法同时考虑了语料中转折和共现对词间关系的影响, 再加上知识库资源的约束, 使得结果更加严谨, 适用于大部分语料; 然而, 更多的约束条件使得后续算法迭代速度变得缓慢. 此外, 半监督算法的运行结果, 还依赖于初始标注种子词的质量. 如何选取合适的种子词, 是该类算法的一个重要问题.

3.3.2 自举半监督法

自举法 (Bootstrapping) 也是一类半监督机器学习算法, 其原理是利用少量标注样本构建分类器, 对未标注样本进行预测, 并将置信度较高的样本添加到标注集中, 训练出一个相对完善的分类器. Volkova 等[36] 利用 Bootstrapping 思想, 使用具有较强主观性的词作为初始词典, 将语料中包含一个主观词以上的句子视为主观句. 考虑否定词的情况下, 若一条语料中均为褒义/贬义, 则将该条看成褒义/贬义. 利用上一次迭代得到的词典, 计算词属于褒义、贬义的概率, 并选取置信度较高的添加到词典中, 用于下次判断. Zhang 等[37] 以少量带标签的文本和通用情感词典作为输入, 使用 Bootstrapping 算法, 对无标签文本分类, 然后按照转折词将带标签语料分成多个情感片段, 使得情感片段内情感一致, 再利用依存分析得到情感片段的依存关系, 并获取候选情感词, 最后利用整个情感片段的极性确定其中候选情感词的极性, 同时将其用于下一个情感片段的判断. Weichselbraun 等[38] 首先利用初始情感词典, 结合否定词, 计算文档的情感得分; 然后设置得分阈值选取部分文档构成语料库; 之后利用贝叶斯公式, 分别计算单词属于正面、负面的概率, 选取概率排名高的前 m 个单词添加到情感词典中 (已经在词典中的, 或者词频低于 n 的则不添加). 反复多次, 最后形成一个较为完整的情感词典. Gao 等[39] 认为两种语言的情感信息能用于相互提高分类器的学习. 借助机器翻译将英语和其他语言一一对应, 以构建两种语言间的关系 R. 利用 co-training 思想, 分别使用英语和其他语言的标注数据集训练分类器 C_E 和 C_T, 并使用 C_E、C_T 去预测各自数据集中未知极性的词, 选取置信度较高的 n 个, 使用关系 R, 最终选取置信度最高的 m 个词, 添加标注数据集中, 继续训练分类器.

自举半监督法是一种较为实用的方法, 仅使用少量的标注信息, 便可以扩展得到其他词的情感倾向, 适用于句型结构相似的语料. 相比关系图半监督方法, 在没有语义知识库的情况下, 该方法也能使用, 但对语料中包含的信息捕捉不够全面, 如其中的并列转折关系, 共现关系等. 此类方法还有两个较为重要问题值得关注. 第一, 初始标注数据的选择. 由于算法是通过不断迭代来获得其他词的情感倾向的, 所以若初始标注数据与语料关联不大, 或不是语料中具有代表性的情感词, 则会使得在判断相关句子或词的情感时, 置信度偏低, 影响判断质量. 第二, 迭代过程中新添加的情感词的质量控制. 由于添加的情感词会用于判断其他句子或词的情感倾向, 若其中有较多错分的情感词, 则会影响后续的判断, 使获得的情感词的准确率偏低.

3.3.3 深度表示法

近年来, 随着神经网络和深度学习不断发展和成功应用, "词向量" (Word embedding)[44] 成了自然语言处理领域 (包括情感词典构建) 的一个热门话题. Tang 等[40] 将情感词典的构建视为对词/短语的情感分类任务. 他们使用 Mikolov 等[45] 提出的 skip-gram 模型, 依据大规模文本来训练词向量, 并用词向量均值来表示句子, 使用三元组 "⟨ 短语, 短

[8] http://nlp.stanford.edu/software/lex-parser.shtml#Download

语所在句子, 短语极性 〉" 作为分类器输入. 分类器的训练集, 是通过 Urban 词典扩展种子词库后获得, 最后使用 Softmax regression 分类器进行短语级别的文本分类, 从而根据词向量判断其极性. 梁军等[41] 提出了一种基于递归自编码器 (Recursive autoencoder, RAE) 的情感极性转移模型, 该模型先将文本转为低维实数向量 (即词向量), 以建立文本表示矩阵, 然后将其作为输入, 训练时, 将文本的否定考虑其中, 使用 LBFGS 算法多次迭代生成最终的情感分析模型, 来判断单词和短语的情感倾向. 杨阳等[42] 使用 word2vec 训练词向量, 并使用大连理工情感词典本体作为种子词, 从语料中选取与种子词的余弦相似度大于 0.8 的词作为备选后, 再对备选集中每个词计算与种子词的余弦相似度, 用于更新词分别属于褒贬的概率, 最后判断单词的情感倾向. 现有的一些词向量训练方法得到的结果, 很可能出现两个向量代表的词, 语义相近但极性完全相反, 比如 "好" 和 "坏", Tang 等[43] 提出了一种新的方法, 利用大量带有弱监督 (Weakly-supervised) 的 tweets 语料, 在传统的四层神经网络结构 (look up、linear、hTanh、linear) C&W 的基础上做了改进, 将情感信息融入到词向量中, 提出了三种改进模型: 1) 增加一个 softmax 层, 使得在输出词向量之前, 先进行一次情感分类, 如果是褒义的, 则结果为 [1, 0], 贬义则为 [0, 1]. 2) 作者认为上述的结果太过苛刻, 对于诸如 [0.7, 0.3] 这样的结果, 我们也可以认为是褒义的, 所以去掉了 softmax 层, 仅依靠输出的褒贬概率, 来判断词的情感倾向. 3) 前两个模型都考虑了句子的情感极性, 但忽略了词的语境, 作者在第三个模型中将语境的相关损失函数考虑其中.

深度学习是自然语言处理领域的近期研究热点, 在相似度计算和语义表示方面取得了突破性进展, 具有广阔的应用前景. 但是, 基于深度学习的文本隐式表示如何与基于规则的文本显式表示很好地结合, 是深度学习与情感词典构建等自然语言处理任务中值得关注的一个问题. 另外, 一些深度学习方法对语料数据比较敏感, 结论只适用于当前所用语料, 而不具有通用性.

4 中文情感词典构建方法

目前, 中文的情感词典构建方法研究相对较少. 我们对这些方法进行了梳理, 同样按照基于知识库、基于语料库、知识库与语料库相结合三类分别进行总结, 将这些工作对应归纳在表 2～表 4 当中.

4.1 基于知识库的方法

在中文情感词典构建方面, 由于完备的汉语语义知识库相对欠缺, 纯粹依靠知识库方法的研究不是很多. 柳位平等[2] 利用 HowNet 进行扩展, 先挑选出一部分常用的情感词构成基础情感词语集, 然后采用词语义元距离计算相似度, 得出每一个词的情感倾向值, 最终构成一个基础情感词典. 杨超等[46] 同时利用 HowNet 和 NTUSD 两种资源, 分别采用计算相似度和统计汉字词频的方式, 判断词的情感倾向性. 周咏梅等[47] 使用跨语言方法, 先获取 HowNet 的英文义元, 然后将义元与 SentiWordNet 对应, 计算这些义元的平均情感强度, 最终得到对应中文的情感强度.

4.2 基于语料库的方法

李勇敢等[48] 是在中文依存句法分析的基础上, 对依存分析的结果进行剪枝和归并, 剪枝主要是删除冗余信息 (比如助词及介词标签 DEC、DEG、P 等), 归并主要是将 "不" 和 "佳"、"酒店" 和 "服务" 这样的词合并为一个, 方便处理, 再利用一些依存规则进行情感词的抽取和极性判断. 共现关系法也适用于构造中文情感词典. Turney 的 PMI-IR 算法中, 为了确定已知词和待确定词的紧密程度, 利用搜索引擎, 对这些词进行检索, 检索到的信息再计算 PMI 值, 从而找出与已知词最相似的词, 构成同义词词典. 阳爱民等[28] 借用 Turney 的思想, 利用种子词与其他词的百度搜索返回结果, 计算词的 SO-PMI, 来判断词的情感极性. 魏志生等[29] 通过计算所有形容词、副词与类别的 MI 值, 取 MI 值最大的 10 % 作为种子词; 再计算种子词与各个类别的 PMI 值来确定种子词的情感倾向. 之后再计算候选词和种子词的 SO-PMI 来确定词的情感极性. 殷春霞等[49] 认为语料中的相同评论对象 (如: 电影、新闻事件等) 使用的同一个情感词的情感倾向, 通常在该评论对象的所有上下文中均是一致的. 因此可以假设: 只要语料足够充分, 通过词汇间的上下文关系便可以计算任意两个情感词间的情感倾向关系. 对于其中可能出现的关系冲突的情况, 作者对两个情感词在语料中所有关系 (转折、非转折、不存在关系) 进行了统计分析, 利用复杂网络确定两者关系.

4.3 知识库和语料库相结合的方法

在中文评论语料相对缺乏的情况下, 李寿山等[35]利用英文种子词典, 借助机器翻译把原评论和对应的翻译评论作为一篇文档, 计算其他词与种子词的 PMI; 然后利用词之间的 PMI 值, 构建连接矩阵, 借助标签传播算法将英文的情感词极性传播到中文词上, 克服中文在现有资源上的一些劣势, 从而构建情感词典. He 等[50] 使用英语的词典资源及机器翻译技术, 进行跨语言情感分类. 作者认为, 在发表评论的时候, 会先考虑文档的整体情感, 然后再考虑用词, 于是使用 LDA 来对文档的

三个主题 (褒义, 贬义, 中性) 建模, 并根据单词的概率分布, 对其进行情感分类. Xu 等[33] 基于人民日报 1997～2004、哈工大同义词词林、现代汉语词典, 结合拥有公共字的词相似度比较大的思想, 构建四个相似度矩阵; 接着挑选并标注一部分种子词, 用于迭代推导未知词的情感极性, 并人工纠正迭代过程中产生的错误. 王昌厚等[51] 使用基于模式的 Bootstrapping 方法, 找出种子词所在的上下文模式, 提取该模式, 接着用提取到的模式抽取新的情感词, 然后循环该过程. 比如, "总体还不错吧", 其中"不错" 是一个情感词, 于是抽取出的模式就是 "很 + instance + 吧", 用这个模式能继续抽取其他情感词. 李荣军等[34] 提出了基于 PageRank 模型情感词极性判断方法, 利用 HowNet 语义相似度构造相似矩阵, 然后使用 PageRank 算法进行迭代计算. 在开始迭代前, 考虑到一些待确定情感词连接权重较低的节点的 "投票" 可能不可靠, 也为保证迭代收敛, 对种子词相似矩阵、待确定词相似矩阵和极性矩阵做了一些处理. 王科等[16] 利用语料中的连接关系, 依据转折词和否定词对文本情感极性产生的影响, 将单词划分成两个集合, 并利用通用情感词典中情感表达明确的词对判断结果进行纠错. 对部分有歧义的情感词, 将其与所描述的对象相结合起来, 作为一个情感词.

4.4 中文情感词典构建尚存的问题

中文情感词典构建相对于英文而言, 存在着无可争议的差距, 这些差距的原因, 主要有如下几个方面: 首先, 中文知识库和语料库资源缺乏. 可使用的知识库与语料库资源数量和质量, 是影响情感词典构建的主要因素. 英语情感分析的研究不但起步较早, 且具有完善的语义知识库 (如: WordNet), 为英文情感词典构建工作带来了巨大的便利. 除了知识库资源, 英语还有大量公开的且得到业内普遍认可的标注、无标注语料库资源, 然而在中文领域却尚缺类似 WordNet 的语义知识库, 目前也并没有形成太多公认的语料库. 因此, 中文情感词典构建工作与英文相关工作相比, 还存在较大差距.

其次, 中文语言分析工具不够成熟. 中文情感词典构建首先需要进行中文分词, 然而现有的分词系统在开放语料上的性能还不够成熟, 很多新词、未登录词不能够正确识别. 除了分词外, 其他语言分析工具, 如: 词性标注、句法分析等, 中文与英文的准确率也存在一定差距. 这些是导致情感词典质量下降的一个重要因素.

此外, 汉语博大精深, 对于同一句话也可能会有不同的理解, 比如: "中国队大败美国队", 由于可能存在中文介词省略的用法, 即 "中国队大败 (于) 美国队", 使得句子的情感完全相反; "冬天: 能穿多少穿多少"、"夏天: 能穿多少穿多少". 两句中的 "多少", 前者正确的分词应该是 "多少", 表示尽可能多, 而后者的正确分词应该是 "多/少", 表示尽可能少. 汉语语言的复杂性也是影响中文情感词典构建性能的一个原因.

5 情感词典构建的难点问题

在情感词典构建过程中, 有很多难点问题需要特别对待, 部分难点问题如表 5 所示.

5.1 情感词典领域适应问题

公开的通用词典准确率是毋庸置疑的, 对这些资源加以改造和利用, 构造领域知识库, 能降低噪音影响, 提高准确率. Choi 等[52] 尝试把一个通用的词典运用到特定领域的情感分类. 他们使用整数线性规划 (Integer linear programming), 利用表达式级的情感来改进通用情感词典, 并用改进后的词典来提升表达式级文本的情感判断性能. Huang 等[15] 使用半监督学习的方法, 结合通用情感词典, 判断领域内单词的情感极性. 除了通用情感词典之外, 现有的很多词典, 都是领域词典, 结合对应的语料库, 利用词与文档间的关系, 能构建跨领域的情感词典. Du 等[53] 利用标记的文档和其对应的情感词典, 生成另一个未标记领域的情感词典. 虽然两个领域分布不同, 但是也可以找出它们的共同部分. 对于未标记文档, 他们的两个基本假设是: 如果一个文档中包含许多正面 (负面) 词, 那么这个文档就是正面 (负面) 的, 并且一个单词如果出现在许多正面 (负面) 文档中, 那么这个单词就是正面 (负面) 的, 它们是相互增强的关系, 之后使用信息瓶颈方法 (Information bottleneck method) 来构建领域词典. Li 等[54] 通过选取一部分两个领域共同的情感词作为种子词, 并通过句法分析, 获得主题词与情感词之间的关系, 之后再根据两者构建领域种子词. 使用一种基于 Bootstrapping 的方法来扩展种子词集, 先用原领域词典构建分类器对目标领域情感词分类, 选取置信度较高的 k 个新词, 用于构建主题–情感词关系图, 并计算目标领域单词的情感得分, 多次迭代获得所有情感词的情感得分.

5.2 属性 – 情感词对构建问题

不管是基于语料库的方法还是语料库与知识库结合的方法, 虽然能够找到特定领域的情感词和它们的情感指向, 但在实 际应用 中还是不够 的. Ding

表 5 情感词典构建中的难点问题

Table 5 Difficult problems in the construction of sentiment lexicon

方法	概述	参考文献
情感词典领域适应问题	领域 A 的语料结合通用词典, 构建领域 A 的情感词典; 或领域 A 的语料结合领域 B 的语料与领域 B 的词典, 构建领域 A 的词典.	(Huang 等, 2014a)[15], (Choi 等, 2009)[52], (Du 等, 2010)[53], (Li 等, 2012)[54]
属性–情感词对构建问题	一般情感词和属性词都是成对出现的, 利用这一点, 我们能够找出情感词; 有些情感词针对不同的属性, 其情感极性不一定相同, 结合属性词能克服这一点.	(Ding 等, 2008)[55], (Lek 等, 2012)[56], (Qiu 等, 2009)[57], (Balahur 等, 2010)[58]
情感词消歧问题	一些情感词包含多种释义, 在判断这些情感词的极性时, 需要先确定其含义, 才能确定其极性.	(Dragut 等, 2010)[59], (Wu 等, 2010)[60], (谢松县等, 2014)[61]
含蓄情感词问题	部分词不直接带有情感色彩, 但是在表达时, 结合上下文便会表现出一定的情感色彩, 比如 “山”, 在描述床板时, 可能是在表达床板有凸起而显得凹凸不平.	(Feng 等, 2011)[62], (Zhang 等, 2011)[63], (Balahur 等, 2011)[64]
新情感词问题	所谓新情感词, 主要针对网络上时常会出现的一些新兴词, 这些新词可能是现有词的另类含义, 也可能是由网友自己创造. 发现并识别其情感加入情感词典中.	(Brody 等, 2011)[65], (Huang 等, 2014b)[66], (张清亮等, 2011)[67]
情感词情感强度问题	情感强度是情感词在其所在极性上变现出的程度值, 是情感词的一个重要属性, 利用情感强度, 能较为精确地衡量句子或文章的情感极性.	(Kim 等)[6], (Williams 等, 2009)[68], (Esuli 等, 2006)[69], (Kumar 等, 2012)[70], (Lu 等, 2010)[71], (柳位平等, 2009)[2], (Gatti 等, 2012)[72]

等[55] 指出, 许多单词在同一个领域的不同文本中, 可能会有不同的情感指向. 如相机领域, “长” 在描述电池续航时间和聚焦时间上的情感极性是相反的. 要解决这个问题, 最好的方法就是将情感词的情感倾向与属性对应. Lek 等[56] 首先构造一个属性词和情感词提取器, 经依存分析之后, 根据英文语法的常用组合模式来同时提取属性词及其对应情感; 对提取的结果先进行属性集聚类, 再利用 WordNet 对情感词和属性词的近义词进行合并, 最后对情感词分类, 构成情感词典. Qiu 等[57] 同时扩展属性词和情感词. 他们使用句法依存关系, 抽取情感词之间、特征之间、以及情感词和特征之间的关系, 扩展属性词库和情感词库. 最后, 他们根据转折词和否定词判断情感, 并作为特征用于分类, 提高准确率. Xia 等[17] 对情感词所在句子的内部特征作了进一步分析, 增加了评价词语的修饰副词和情感词等, 并使用句子之间的连词规则, 通过贝叶斯分类器对属性依赖词进行极性判断. Balahur 等[58] 使用三种策略的多数投票结果作为属性–情感词对的极性, 三种策略分别是基于上下文的有监督学习、基于网络查询的最高点击, 以及基于规则的方法. 这里的网络查询形如: ⟨ 属性 + 评价词语 + and + 预定义褒贬义词 ⟩, 例如: “价格高并且好”, “价格高并且差”, 取返回结果数最多的类别作为属性–情感词对的极性.

5.3 情感词消歧问题

同一个单词通常具有多种释义, 这些释义可能会具有不同的情感极性; 此外, 语言有语境和语气上的问题, 不同语境或者不同语气下的同一句话, 可能会导致相反结果. 比如: “我为你骄傲!”, “他一有点成就, 就会骄傲”. 这里的 “骄傲” 显然具有相反的含义. Dragut 等[59] 使用 WordNet 把已知情感极性的单词作为输入, 产生同义词集合的情感指向. 在产生过程中, 针对同一单词的不同释义的情感极性可能会有所不同的情况, 以使用频率多的释义的极性来表示这个单词的情感极性, 然后推导出的同义词集合的情感指向能进一步用来推导其他单词的极性. Wu 等[60] 提出了根据现有知识对情感词消歧的方法. 针对中文中的 “大、小、多、少、高、低” 等 14 个高频属性依赖词, 通过对属性词和修饰形容词的副词 (“有点”、“那么”) 的判断, 使用固定模式来判断词的情感. 对于无法用模式判断的属性词 X, 作者通过 HowNet、百度搜索结果、PMI 值, 将 X 与已知的属性词关联起来, 判断情感倾向. 中文在构建情感词典时, 通常会引入已有资源帮助消歧. 谢松县等[61] 将中文翻译与 SentiWordNet 结合使用, 对单词的所有释义进行加权计算, 获得唯一的褒贬得分, 从而达到消歧的目的.

5.4 含蓄情感词问题

含蓄情感词是指这样一些词, 它们在通常情况下是中性或客观的, 但是在某些特定上下文中, 它们会表现出一定的情感. Feng 等[62] 研究了含蓄字典的问题. 情感词典一般是直接或者间接表达情感, 而含蓄字典一般和情感的特殊极性有关, 比如 “award” 和 “promotion” 一般是正面的, 而

“war”、“cancer” 一般是负面的. 这些词也属于主观词, 需要加入到情感词典中. Zhang 等[63] 提出了一个确定名词和名词词组特征的方法, 同时也暗示了特定领域的情感. 这些单独的名词和名词词组没有情感, 但是在某一领域的文本中, 也许表达了一些情感. 比如 “山谷” 和 “山”, 一般没情感, 但是在描述床垫的时候, 一般是贬义的. Balahur 等[64] 基于常识构建行为反应链库, 针对不同的反应行为链, 来判断主体的情感.

5.5 新情感词问题

随着网络的迅速发展, 生活中出现了越来越多的网络词, 这些词有些是旧词新用 (比如: “灌水”), 有些是借鉴了一些方言 (比如: “给力”), 还有些是自造的 (比如: “喜大普奔”). 处理好它们的分词和词性标注工作, 就能用现有的情感判断方法识别. Huang 等[66] 使用种子词和模板结合的方法发现情感新词, 模板为种子词的上下文语境, 如 “太 * 了”、“都 * 的” 等, 利用模板迭代寻找新的情感词, 迭代过程中仅更新情感词前的副词和后面的助词. 在这之前, 作者首先衡量了模板的有效性, 之后根据有效性, 每次迭代计算置信度, 选取置信度较高的词添加到种子词中. 张清亮等[67] 利用 PMI-IR 方法, 结合种子词, 来判断网络词的情感极性. 而在英文中, 词的改变通常会使用一些夸张的手法, 比如 “cooooooooool” 形容非常冷. Brody 等[65] 通过对词长的研究, 来判断对情感的影响.

5.6 情感词情感强度问题

情感强度是情感词典的一个属性. 它是情感词表现出的程度值. 如描述建筑时, “好看” 表现出的情感强度要小于 “辉煌”. 情感强度对情感词典的现实应用具有重要的价值. 值得一提的是, 词语共现法 (本文第 3.2.2 节) 涉及到的 SO-PMI 及相关方法, 均可以得到单词的情感强度. Williams 等[68] 认为, 在语义关系图中, 词的路径权值, 与两个词的公共释义数相关, 公共释义越少, 则权值越大. 再分别计算词到一对褒贬种子词的加权距离, 以此来表示词的极性强度, 如式 (7) 所示:

$$SO_{(w^+,w^-)}(w)=\frac{G(w,w^-)-G(w,w^+)}{\max(G(w,w^-),G(w,w^+))} \quad (7)$$

其中, $G(w^+,w^-)$ 表示关系图中词 w^+ 到 w^- 的最短距离. Kim 等[6] 利用 WordNet 中的同义词关系来计算单词的情感强度. 对于一个未知极性的词, 通过查找其同义词与种子词的极性关系, 来近似计算单词的褒贬强度, 以绝对值大小确定该词的情感极性和情感强度, 计算方法如式 (8) 所示:

$$\begin{aligned}&\arg\max_c p(c|w)=\arg\max_c p(c)p(w|c)=\\&\arg\max_c p(syn_1,syn_2,\cdots,syn_n|c)=\\&\arg\max_c p(c)\prod_{k=1}^{m}p(f_k|c)^{count(f_k,synset(w))}\end{aligned} \quad (8)$$

其中, f_k 为词 w 属于 c 类情感的同义词, $count(f_k, synset(w))$ 为 f_k 在 w 的同义词集中出现的次数. Esuli 等[69] 认为情感相近的词具有相似的释义, 通过对每一个同义词集的释义使用 TF-IDF 技术, 来对同义词集进行向量化表示, 之后将 K 分别取 0、2、4、6 时得到的同义词集向量 (K 为同义词集通过 WordNet 迭代扩充的次数), 并使用支持向量机和朴素贝叶斯分类器分别对词进行极性判断, 最后结合这 8 个结果对单词的情感极性打分. Baccianella 等[11] 用词与同义词集释义的包含关系建立关系图后, 以两个随机游走算法分别求单词的褒贬得分, 由于得分普遍偏低, 对其进行指数加权后, 进行两者对比, 确定词的最终情感倾向和情感得分. Kumar 等[70] 使用 Hatzivassiloglou 等[13] 的思想, 依据语料库中的连接和转折关系, 判断词与种子词的相似度, 并用种子词的情感强度与两词相似度的乘积作为该词的情感强度值. Lu 等[71] 从语料库中抽取形容词间的并列关系, 构建关系网络, 然后用搜索引擎查找相连接的单词对, 按返回数来计算两个单词节点间的权重, 之后使用传播算法迭代更新, 得到每个词的强度. 柳位平等[2] 在构建情感词典时, 通过知网的语义相似度计算方法, 计算两个词的语义距离, 用情感词与种子词语义关联的紧密程度作为其情感强度. Gatti 等[72] 利用 SentiWordNet 来计算单词的情感强度, 分别提出了八种计算方式: 1) 随机取值. 2) 随机取词的一种释义, 在 SentiWordNet 中查找其褒/贬情感得分, 作为该词的情感强度. 3) 选取词的第一个释义对应的褒/贬情感得分作为该词的情感强度. 4) 分别使用词的平均褒/贬情感得分作为该词的情感强度. 5) 分别使用词非零褒/贬情感得分的平均值作为该词的情感强度. 6) 分别选取词最大的褒/贬得分作为其情感强度, 若两者相同, 则求最大值对应释义编号与释义总数的商, 商较大的释义对应的情感为单词的情感. 7) 和 8) 均按照使用频率, 对词的情感得分乘以调和系数的均值作为其褒/贬情感. Schneider 等[73] 利用线性最优理论, 结合 WordNet 等现有词典中词及其褒义、贬义和中性释义的数量, 来获得词分别属于这三类的概率, 取概率最大的作为词的极性, 值为词的情感得分.

表 6 相关测评竞赛

Table 6 Related evaluation contest

评测名称	任务编号	评测内容
TREC 2008	3	观点词的识别和极性判断.
SemEval 2010	18	对语料中部分极性依赖上下文的形容词进行消歧.
SemEval 2014	4.2	判断属性词对应的情感 (褒义、贬义、中性、褒贬兼具).
SemEval 2015	12.1	提取领域情感词并判断极性.
COAE 2008	1、2	分别是情感词的识别和褒贬分析.
COAE 2009	1	情感词的识别及分类.
COAE 2011	1	领域观点词的抽取与极性判别.
COAE 2014	3	给定大规模的微博句子集, 要求自动发现新的词语, 以及每个词语的情感倾向性.

6 情感词典性能评估

我们将情感词典的性能评估方式分为直接评估方法和间接评估方法. 直接评估通过生成词典与标准词典比较得到; 间接评估则将情感词典应用到情感分析任务中, 通过情感分析结果来评价词典的性能.

6.1 直接评估方法

直接评估方法主要是直接对词典本身进行评估, 其中一种方法是取词典中一定比例 (如: 50/100/200) 或者全部的词, 人工判断或与通用情感词典对比情感词的极性是否正确, 以这些词的准确率来衡量情感词典的性能. 当词典的排序按置信度或情感强度排序时, 排名靠前的情感词的准确率对词典性能的评估尤为重要.

另一种直接评估的方法是将情感词典与经过人工标注的情感词典进行比较, 计算精确率 (Presicion)、召回率 (Recall) 和 F_1 值:

$$P = \frac{right\_hit}{all\_hit} \tag{9}$$

$$R = \frac{right\_hit}{all\_related} \tag{10}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{11}$$

其中, *right_hit* 为被正确检索到的数目, *all_hit* 为被检索到的总数, *all_related* 为应该检索到的数目.

相关工作中, 文献 $[5, 8, 31-32, 56, 62, 74-75]$ 等通过与通用情感词典 (如: GI) 或人工标注结果的对比, 来评估它们构建的情感词典的性能. 但是, 在进行人工标注时, 尤其涉及到情感强度的标注时, 由于标注人员的主观性, 通常需要 kappa 统计量进行标注一致性检测, 来使标注结果更加可靠.

6.2 间接评估方法

词典性能的好坏, 还可以结合它在情感分析中的应用情况来进行评估. 性能好的词典, 能够提高文本情感分析的各项指标结果. 间接评估情感词典的方法有很多, 在属性级情感分析任务中, 一般做法是, 找出属性及其对应的情感词, 结合上下文语境和情感词典来判断情感词的极性, 与标准数据对比, 计算精确率、召回率和 F_1 值, 从而评价不同情感词典的性能[55].

情感词典是无监督句子/篇章级情感分析任务的主要依据. Kaji 等[30] 利用生成的情感词典, 对文本进行句子级情感分类, 并与其他方法构建的词典进行性能对比. Choi 等[52] 采用条件随机场 (Conditional random fields, CRF) 方法, 在句子级的情感分类任务上, 对比了是否使用词典对结果产生的影响. 对于监督学习方法, 柳位平等[2] 采用监督学习的方法来衡量词典的优劣, 即以得到的情感词作为特征进行文本分类, 通过比较分类准确率的高低来衡量词典的好坏. 杨鼎等[76] 对比了卡方检验和情感词典作为文本特征, 通过支持向量机进行文本情感分类的性能, 认为情感词典的效果相比卡方检验的结果有所提升. 王科等[16] 通过与 COAE 2008 任务三提供的标准答案对比, 判断情感词典在语料上的精确率、召回率和 F_1 值.

6.3 国内外相关评测竞赛

自 2006 年起, 国内外相关组织逐渐发起了一系列情感分析测评比赛, 如 TREC (Text retrieval conference)、SemEval (Semantic evaluation)、中文倾向性分析评测 (Chinese opinion analysis evaluation) 等. 我们在表 6 中总结了这些情感分析评测中与情感词典构建相关的任务. TREC 在 2006 年首次引入了博客检索任务, 更多的研究者致力于该任务的研究. 情感信息检索要求检索到的文档必须同时满足两个准则: 主题相关和具有情感倾向[77]. SemEval 是国际权威的语义分析相关评测, 由早先成立的词义消歧测评 Senseval 发展而来, 目前已经涉及文本语义相似度计算、语义分析、空间角色标注等多方面的任务. COAE 是由中国中文信息学会

信息检索专业委员会组织的中文情感分析测评竞赛, 涉及不同环境、不同粒度的情感分析任务.

在 TREC 2008 的情感分析任务中, Lee 等[78] 以平均精确率 (Mean average precision, MAP) 0.4052 的成绩取得了较好的成绩, 他们在构建情感词典时, 使用亚马逊带星级的的评论语料库, 并结合 SentiWordNet 来判断单词的情感倾向, 将 4～5 星的认为褒义, 1～2 星的为贬义, 之后使用 EM 算法来估计 $p(pos|w)$ 和 $p(neg|w)$, 判断词 w 的情感倾向. Xu 等[79] 在 SemEval 2010 中, 通过 HowNet 计算属性词与已知词的相似度, 来扩展部分属性词, 构建 〈 属性, 情感词 〉 种子对, 若属性依赖情感在种子对中, 则直接判断; 若属性依赖情感所在句子与已知情感词相邻, 则结合转折否定规则, 判断其情感, 同时更新种子词集. 当依赖属性的情感词独立出现时, 通过对句子中已知情感词的得分进行累加, 判断句子情感, 从而判断情感词极性; 若仍无法确定, 则句子 S_j 的情感 $P(S_j)$ 用上下文加权后计算. 最后获得了宏平均 0.953, 微平均 0.936 的成绩. 在 SemEval 2014 的情感词典分析任务中, Toh 等[80] 使用 Opinion Lexicon 作为种子词库, 并手动添加了部分餐馆领域的种子词. 为扩大覆盖率, 作者又通过同义/反义关系进行扩充; 之后使用双向传播算法扩展情感词和属性词. 由于部分情感词依赖于它所描述的属性, 作者分别构建了通用情感词库和具有属性依赖的情感词库. 对于属性依赖的情感词, 使用通用情感词典, 结合上下文语境来确定其极性. 在给定训练语料 (笔记本/餐馆) 的情况下, F_1 值分别为: 73.78 和 83.98. SemEval 2015 中, Saias 等[81] 也在上述两个领域的语料中, 以准确率 0.7934 和 0.7869 胜出. 他们的方法是, 通过提取句子主干, 包括否定关系、转折关系、动词、形容词词根等, 再利用通用情感词典发现情感词进行监督学习, 来判断属性对应的情感, 尤其是具有属性依赖的情感词的情感. 在 COAE 竞赛中, 刘军等[82] 直接使用 HowNet 作为情感词典, 利用句法分析工具, 找出与情感词相关的依存关系, 并结合语言规则, 分析情感词所处的上下文, 深入分析情感词的情感极性. 获得了 P@100 (前 100 个词的准确率, 下同) 0.88, P@1000 为 0.925 的成绩. 在 COAE 2009 中, 徐戈等[83] 人工标注一些情感色彩鲜明的词作为种子词, 之后通过多个语料, 分别求词间的相似度并融合得到矩阵 W. 之后结合奇异值分解技术, 获得单词的评分. 最终 F_1 得分为 0.1675. 在 COAE 2011 的情感分析任务中, 徐睿峰等[84] 使用现有的几个情感知识库 (褒义词、贬义词词典, HowNet, NTUSD 等), 对其进行合并构成情感词典, 初步实验后发现金融领域情感词覆盖率低. 于是通过一系列方法来扩展情感词典, 如: 用并列连词关系、高频共现关系、固定模式 “ 副词 + 评价词” 等. 最后在电子、影视、金融三个领域的上, 平均结果为: F_1 : 0.1476, P@1000 : 0.5713. 由于微博中情感新词较多, 分词时往往会被错误切分, 廖健等[85] 等利用互信息, 对分词结果进行重组, 并根据词性组合规则, 如 “名词 + 形容词”, 来发现情感新词. 之后借用外部情感词典, 根据新词与词典中词的 PMI 值, 获取新词的情感强度. 在 COAE 2014 中, 得到 F_1 值为 0.166.

7 总结与展望

本文对情感词典自动构建方法进行了综述. 从所需资源的角度, 我们将情感词典自动构建方法归纳为三大类: 基于知识库的方法、基于语料库的方法和基于知识库和语料库相结合的方法; 从方法论的角度, 我们将每一大类方法又分成若干小类. 按照这样的体系, 本文对中英文情感词典自动构建主要文献进行了详细回顾和总结, 阐述了每一类方法的优点与缺陷, 分析和归纳了情感词典构建任务中的若干难点问题. 此外, 本文还对现有情感知识库、情感词典评估方法、情感词典构建相关评测竞赛进行了总结.

作为情感分析的一个基础任务, 尽管情感词典自动构建已经得到了广泛关注和深入研究. 但是仍然存在很多问题, 亟待在未来工作中进一步解决和完善. 对此, 我们作出以下几点展望:

1) 情感词典构建难点问题的突破

我们在第 5 节已经总结了情感词典构建的若干难点问题. 到目前为止, 大部分问题并没有得到完善的解决. 因此, 期望在情感词典构建研究上取得更好的进展, 需要在这些难点问题的解决方法上有所突破.

2) 自然语言处理基础资源的完善

英文情感词典自动构建工作相对成熟, 其重要原因在于英文具有较为完善的词典资源和相对成熟的自然语言分析工具. 我们在第 4.4 节曾总结了中文情感词典构建工作面临的几点困难, 中文自然语言处理基础资源不完善是其中一个重要因素. 因此, 在未来工作中, 以期提高中文情感词典自动构建的性能, 需要着力于建立完善的中文词典资源和提高中文自然语言分析工具的性能.

3) 深度学习等新技术方法的应用

深度学习是近几年机器学习、人工智能领域迅速发展的研究热点. 在自然语言处理领域, 基于深度学习的文本表示技术也得到了长足的进展和广泛的应用. 如何进一步利用深度学习等新兴技术方法, 更加丰富地挖掘词汇的情感信息, 更加深刻地度量情感词的相似性也是情感词典自动构建工作一个值得

关注的方向.

References

1 Hu M Q, Liu B. Mining and summarizing customer reviews. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2004. 168−177

2 Liu Wei-Ping, Zhu Yan-Hui, Li Chun-Liang, Xiang Hua-Zheng, Wen Zhi-Qiang. Research on building Chinese basic semantic lexicon. *Journal of Computer Applications*, 2009, **29**(10): 2875−2877
(柳位平, 朱艳辉, 栗春亮, 向华政, 文志强. 中文基础情感词词典构建方法研究. 计算机应用, 2009, **29**(10): 2875−2877)

3 Liu B. *Sentiment Analysis and Opinion Mining*. San Rafael, CA: Morgan & Claypool Publishers, 2012.

4 Strapparava C, Valitutti A. WordNet-affect: an affective extension of wordNet. In: Proceedings of the 2004 International Conference on Language Resources and Evaluation. Lisbon: LREC, 2004. 1083−1086

5 Neviarouskaya A, Prendinger H, Ishizuka M. SentiFul: a lexicon for sentiment analysis. *IEEE Transactions on Affective Computing*, 2011, **2**(1): 22−36

6 Kim S M, Hovy E. Determining the sentiment of opinions. In: Proceedings of the 20th International Conference on Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004. 1367−1377

7 Blair-Goldensohn S, Hannan K, McDonald R, Neylon T, Reis G, Reynar J. Building a sentiment summarizer for local service reviews. In: Proceedings of the WWW2008 Workshop: NLP in the Information Explosion Era. Beijing, China: NLPIX, 2008. 200−207

8 Kamps J, Marx M, Mokken R J, De Rijke M. Using wordnet to measure semantic orientations of adjectives. In: Proceedings of the 4th International Conference on Language Resources and Evaluation. Paris: European Language Resources Association, 2004. 1115−1118

9 Hassan A, Abu-Jbara A, Jha R, Radev D. Identifying the semantic orientation of foreign words. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. 592−597

10 Andreevskaia A, Bergler S. Mining WordNet for a fuzzy sentiment: sentiment tag extraction from wordNet glosses. In: Proceedings of the 2006 Conference of the European Chapter of the Association for Computational Linguistics. Budapest: EACL, 2006. 209−216

11 Baccianella S, Esuli A, Sebastiani F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the 2010 International Conference on Language Resources and Evaluation. Malta: LREC, 2010. 2200−2204

12 Esuli A, Sebastiani F. PageRanking wordNet synsets: an application to opinion mining. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics. Prague: Association for Computational Linguistics, 2007. 424−431

13 Hatzivassiloglou V, McKeown K R. Predicting the semantic orientation of adjectives. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 1997. 174−181

14 Kanayama H, Nasukawa T. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006. 355−363

15 Huang S, Niu Z D, Shi C Y. Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems*, 2014, **56**: 191−200

16 Wang Ke, Xia Rui. An approach to Chinese sentiment lexicon construction based on conjunction relation. In: Proceedings of the 14th China National Conference on Computational Linguistics. Guangzhou, China: CCL, 2015.
(王科, 夏睿. 一种基于连接关系的情感词典构建方法. 见: 第十四届全国计算语言学学术会议. 广州: 中国中文信息学会, 2015.)

17 Xia Y Q, Cambria E, Hussain A, Zhao H. Word polarity disambiguation using Bayesian model and opinion-level features. *Cognitive Computation*, 2014, **7**(3): 369−380

18 Church K W, Hanks P. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 1990, **16**(1): 22−29

19 Turney P D. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In: Proceedings of the 12th European Conference on Machine Learning. Berlin Heidelberg: Springer, 2001. 491−502

20 Turney P D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002. 417−424

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

22 Krestel R, Siersdorfer S. Generating contextualized sentiment lexica based on latent topics and user ratings. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. New York, NY: ACM, 2013. 129−138

23 Tai Y J, Kao H Y. Automatic domain-specific sentiment lexicon generation with label propagation. In: Proceedings of the 2013 International Conference on Information Integration and Web-based Applications & Services. New York, NY: ACM, 2013. 191−200

24 Wawer A. Mining co-occurrence matrices for SO-PMI paradigm word candidates. In: Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2012. 74−80

25 Glavaš G, Šnajder J, Baši′c B D. Experiments on hybrid corpus-based sentiment lexicon acquisitionIn: Proceedings of the 2012 Workshop on Innovative Hybrid Approaches to the Processing of Textual Data. Stroudsburg, PA, USA: Association for Computational Linguistics, 2012. 1−9

26 Bollegala D, Weir D, Carroll J. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. 132−141

27 Velikovich L, Blair-Goldensohn S, Hannan K, McDonald R. The viability of web-derived polarity lexicons. In: Proceedings of the 2010 North American Chapter of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. 777−785

28 Yang Ai-Ming, Lin Jiang-Hao, Zhou Yong-Mei. Method on building Chinese text sentiment lexicon. *Journal of Frontiers of Computer Science and Technology*, 2013, **7**(11): 1033−1039
(阳爱民, 林江豪, 周咏梅. 中文文本情感词典构建方法. 计算机科学与探索, 2013, **7**(11): 1033−1039)

29 Wei Zhi-Sheng, Ji Yang-Sheng, Luo Chun-Yong, Chen Jia-Jun. Generative sentiment classification model affiliating domain-specific sentiment lexicons. *Journal of Frontiers of Computer Science and Technology*, 2011, **5**(12): 1105−1113
(魏志生, 吉阳生, 罗春勇, 陈家骏. 加入领域先验知识的产生式情感分类模型. 计算机科学与探索, 2011, **5**(12): 1105−1113)

30 Kaji N, Kitsuregawa M. Building lexicon for sentiment analysis from massive collection of HTML documents. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Prague: Association for Computational Linguistics, 2007. 1075−1083

31 Peng W, Park D H. Generate adjective sentiment dictionary for social media sentiment analysis using constrained nonnegative matrix factorization. In: Proceedings of the 15th International AAAI Conference on Weblogs and Social Media. Menlo Park, CA: AAAI Press, 2011. 273−280

32 Rao D, Ravichandran D. Semi-supervised polarity lexicon induction. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009. 675−682

33 Xu G, Meng X F, Wang H F. Build Chinese emotion lexicons using a graph-based algorithm and multiple resources. In: Proceedings of the 23rd International Conference on Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. 1209−1217

34 Li Rong-Jun, Wang Xiao-Jie, Zhou Yan-Quan. Semantic orientation computing using PageRank model. *Journal of Beijing University of Posts and Telecommunications*, 2010, **33**(5): 141−144
(李荣军, 王小捷, 周延泉. PageRank 模型在中文情感词极性判别中的应用. 北京邮电大学学报, 2010, **33**(5): 141−144)

35 Li Shou-Shan, Li Yi-Wei, Huang Ju-Ren, Su Yan. Construction of Chinese sentiment lexicon using bilingual information and label propagation algorithm. *Journal of Chinese Information Processing*, 2013, **27**(6): 75−81
(李寿山, 李逸薇, 黄居仁, 苏艳. 基于双语信息和标签传播算法的中文情感词典构建方法. 中文信息学报, 2013, **27**(6): 75−81)

36 Volkova S, Wilson T, Yarowsky D. Exploring sentiment in social media: bootstrapping subjectivity clues from multilingual twitter streams. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria: Association for Computational Linguistics, 2013. 505−510

37 Zhang Z, Singh M P. ReNew: a semi-supervised framework for generating domain-specific lexicons and sentiment analysis. In: Proceedings of the 52nd Annual Meeting on Association for Computational Linguistics. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014. 542−551

38 Weichselbraun A, Gindl S, Scharl A. Using games with a purpose and bootstrapping to create domain-specific sentiment lexicons. In: Proceedings of the 20th ACM international conference on Information and knowledge management. New York, NY, USA: ACM, 2011. 1053−1060

39 Gao D H, Wei F R, Li W J, Liu X H, Zhou M. Co-training based bilingual sentiment lexicon learning. In: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2013. 26−28

40 Tang D Y, Wei F R, Qin B, Zhou M, Liu T. Building large-scale twitter-specific sentiment lexicon: a representation learning approach. In: Proceedings of the 25th International Conference on Computational Linguistics. Dublin, Ireland: COLING, 2014. 172−182

41 Liang Jun, Chai Yu-Mei, Yuan Hui-Bin, Zan Hong-Ying, Liu Min. Deep learning for Chinese micro-blog sentiment analysis. *Journal of Chinese Information Processing*, 2014, **28**(5): 155−61
(梁军, 柴玉梅, 原慧斌, 昝红英, 刘铭. 基于深度学习的微博情感分析. 中文信息学报, 2014, **28**(5): 155−61)

42 Yang Yang, Liu Long-Fei, Wei Xian-Hui, Lin Hong-Fei. New methods for extracting emotional words based on distributed representations of words. *Journal of Shandong University (Natural Science)*, 2014, **49**(11): 51−58
(杨阳, 刘龙飞, 魏现辉, 林鸿飞. 基于词向量的情感新词发现方法. 山东大学学报 (理学版), 2014, **49**(11): 51−58)

43 Tang D Y, Wei F R, Yang N, Zhou M, Liu T, Qin B. Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014. 1555−1565

44 Collobret R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, 2011, **12**(1): 2493−2537

45 Mikolov T, Sutskever I, Chen K, Corrado G S, Dean J. Distributed representations of words and phrases and their compositionality. In: Proceedings of the 2013 Advances in Neural Information Processing Systems. Nanjing: NIPS, 2013: 3111−3119

46 Yang Chao, Feng Shi, Wang Da-Ling, Yang Nan, Yu Ge. Analysis on web public opinion orientation based on extending sentiment lexicon. *Journal of Chinese Computer Systems*, 2010, **31**(4): 691−695
(杨超, 冯时, 王大玲, 杨楠, 于戈. 基于情感词典扩展技术的网络舆情倾向性分析. 小型微型计算机系统, 2010, **31**(4): 691−695)

47 Zhou Yong-Mei, Yang Jia-Neng, Yang Ai-Ming. A method on building Chinese sentiment lexicon for text sentiment analysis. *Journal of Shandong University (Engineering Science)*, 2013, **43**(6): 27−33
(周咏梅, 杨佳能, 阳爱民. 面向文本情感分析的中文情感词典构建方法. 山东大学学报 (工学版), 2013, **43**(6): 27−33)

48 Li Yong-Gan, Zhou Xue-Guang, Sun Yan, Zhang Huan-Guo. The study of construction for emotion thesaurus based on dependency parsing combined with rules and statistics methods. *Journal of Wuhan University (Natural Science Edition)*, 2013, **59**(5): 491−498
(李勇敢, 周学广, 孙艳, 张焕国. 结合依存关联分析和规则统计分析的情感词库构建方法. 武汉大学学报 (理学版), 2013, **59**(5): 491−498)

49 Yin Chun-Xia, Peng Qin-Ke. Identifying word sentiment orientation for free comments via complex network. *Acta Automatica Sinica*, 2012, **38**(3): 389−398
(殷春霞, 彭勤科. 利用复杂网络为自由评论鉴定词汇情感倾向性. 自动化学报, 2012, **38**(3): 389−398)

50 He Y L, Alani H, Zhou D Y. Exploring English lexicon knowledge for Chinese sentiment analysis. In: Proceedings of CIPS-SIGHAN Joint Conference on Chinese Language Processing. Beijing, China: ORO, 2010. 91−104

51 Wang Chang-Hou, Wang Fei. Extracting sentiment words using pattern based Bootstrapping method. *Computer Engineering and Applications*, 2014, **50**(1): 127−129
(王昌厚, 王菲. 使用基于模式的 Bootstrapping 方法抽取情感词. 计算机工程与应用, 2014, **50** (1): 127−129)

52 Choi Y, Cardie C. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009. 590−598

53 Du W F, Tan S B, Cheng X Q, Yun X C. Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon. In: Proceedings of the 3rd ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2010. 111−120

54 Li F T, Pan S J, Jin O, Yang Q, Zhu X Y. Cross-domain co-extraction of sentiment and topic lexicons. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2012. 410−419

55 Ding X, Liu B, Yu P S. A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2008. 231−240

56 Lek H H, Poo D C C. Sentix: an aspect and domain sensitive sentiment lexicon. In: Proceedings of the 2012 IEEE 24th International Conference on Tools with Artificial Intelligence. Washington, DC, USA: IEEE Computer Society, 2012. 261−268

57 Qiu G, Liu B, Bu J J, Chen C. Expanding domain sentiment lexicon through double propagation. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2009. 1199−1204

58 Balahur A, Montoyo A. OpAL: applying opinion mining techniques for the disambiguation of sentiment ambiguous adjectives in SemEval-2 task 18. In: Proceedings of the 5th International Workshop on Semantic Evaluation. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. 444−447

59 Dragut E C, Yu C, Sistla P, Meng W Y. Construction of a sentimental word dictionary. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2010. 1761−1764

60 Wu Y F, Wen M M. Disambiguating dynamic sentiment ambiguous adjectives. In: Proceedings of the 23rd International Conference on Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. 1191−1199

61 Xie Song-Xian, Liu Bo, Wang Ting. Applying semantic relations to construct construct sentiment lexicon automaticlly. *Journal of National University of Defense Technology*, 2014, **36**(3): 111−115
(谢松县, 刘博, 王挺. 应用语义关系自动构建情感词典. 国防科技大学学报, 2014, **36**(3): 111−115)

62 Feng S, Bose R, Choi Y. Learning general connotation of words using graph-based algorithms. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, UK: Association for Computational Linguistics, 2011. 1092−1103

63 Zhang L, Liu B. Identifying noun product features that imply opinions. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. 575−580

64 Balahur A, Hermida J M, Montoyo A. Detecting implicit expressions of sentiment in text based on commonsense knowledge. In: Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. 53−60

65 Brody S, Diakopoulos N. Cooooooooooooooollllllllllll-lll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. 562−570

66 Huang M L, Ye B R, Wang Y C, Chen H Q, Cheng J J, Zhu X Y. New word detection for sentiment analysis. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014. 531−541

67 Zhang Qing-Liang, Xu Jian. Research on automatic extraction of web sentiment words. *Journal of Library and Information Technology*, 2011, **27**(10): 24−28
(张清亮, 徐健. 网络情感词自动识别方法研究. 现代图书情报技术, 2011, **27**(10): 24−28)

68 Williams G K, Anand S S. Predicting the polarity strength of adjectives using wordnet. In: Proceedings of the Third International ICWSM Conference. Menlo Park, CA: AAAI Press, 2009. 346−349

69 Esuli A, Sebastiani F. Sentiwordnet: a publicly available lexical resource for opinion mining. In: Proceedings of the 2006 Language Resources and Evaluation. Genoa, Italy: LREC, 2006. 417−422

70 Kumar A, Sebastian T M. Sentiment analysis on twitter. *International Journal of Computer Science Issues*, 2012, **9**(4): 372−378

71 Lu Y, Kong X F, Quan X J, Liu W Y, Xu Y L. Exploring the sentiment strength of user reviews. *Web-Age Information Management*. Berlin Heidelberg: Springer, 2010. 471−482

72 Gatti L, Guerini M. Assessing sentiment strength in words prior polarities. In: Proceedings of the 23th International Conference on Computational Linguistics. Mumbai: CSCL, 2012. 361−370

73 Schneider A, Dragut E. Towards debugging sentiment lexicons. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China: Association for Computational Linguistics, 2015. 1024−1034

74 Mohammad S, Dunne C, Dorr B. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009. 599−608

75 Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the 2005 Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005. 347−354

76 Yang Ding, Yang Ai-Min. Classification approach of Chinese texts sentiment based onsemantic lexicon and naive Bayesian. *Application Research of Computers*, 2010, **27**(10): 3737−3739
(杨鼎, 阳爱民. 一种基于情感词典和朴素贝叶斯的中文文本情感分类方法. 计算机应用研究, 2010, **27**(10): 3737−3739)

77 Zhao Yan-Yan, Qin Bing, Liu Ting. Sentiment analysis. *Journal of Software*, 2010, **21**(8): 1834−1848
(赵妍妍, 秦兵, 刘挺. 文本情感分析. 软件学报, 2010, **21**(8): 1834−1848)

78 Lee Y, Na S H, Kim J, Nam S H, Jng H Y, Lee J H. KLE at TREC 2008 blog track: blog post and feed retrieval. In: Proceedings of 2008 Text REtrieval Conference. Pohang, South Korea: Pohang University of Science and Technology (South Korea), 2008.

79 Xu R F, Xu J, Kit C. $HITSZ_CITYU$: Combine collocation, context words and neighboring sentence sentiment in sentiment adjectives disambiguation. In: Proceedings of the 5th International Workshop on Semantic Evaluation. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. 448−451

80 Toh Z Q, Wang W T. DLIREC: aspect term extraction and term polarity classification system. In: Proceedings of the 8th International Workshop on Semantic Evaluation. Dublin, Ireland: IWSE, 2014. 235−240

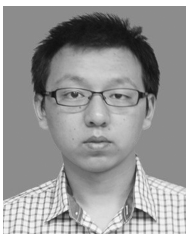
81 Saias J, Ramalho R R. Sentiue: target and aspect based sentiment analysis in SemEval-2015 task 12. In: Proceedings of the 9th International Workshop on Semantic Evaluation. Denver, Colorado: Association for Computational Linguistics, 2015. 767−771

82 Liu Jun, Liu Quan-Sheng, Chen Mo-Sha, Song Hong-Yan, Huang Gao-Hui, Zhang Xiao-Jun, Yao Tian-Fang. Analysis on the evaluation results of the first Chinese orientation analysis evaluation. In: Proceedings of the 1st Conference on Chinese Opinion Analysis Evaluation. Beijing, China: COAE, 2008. 125−141
(刘军, 刘全升, 陈漠沙, 宋鸿彦, 黄高辉, 张潇君, 姚天昉. 第一届中文倾向性分析评测结果浅析. 见: 第一届中文倾向性分析评测研讨会论文集. 北京: 中国中文信息学会, 2008. 125−141)

83 Xu Ge, Meng Xin-Fan, Wang Hou-Feng. Emotion ranking based on multi-modality learning. In: Proceedings of the 2nd Conference on Chinese Opinion Analysis Evaluation. Shanghai, China: COAE, 2009. 24−29
(徐戈, 蒙新泛, 王厚峰. 基于多模态学习的情感评级. 见: 第二届中文倾向性分析评测研讨会论文集. 上海: 中国中文信息学会, 2009. 24−29)

84 Xu Rui-Feng, Wang Ya-Wei, Xu Jun, Zhang Yue, Zheng Hai-Qing, Gui Lin, Ye Lu. Chinese opinion analysis based on multi knowledge integration and multi classifier voting. In: Proceedings of the 3rd Conference on Chinese Opinion Analysis Evaluation. Jí′nan, China: COAE, 2011. 77−87
(徐睿峰, 王亚伟, 徐军, 张玥, 郑海清, 桂林, 叶璐. 基于多知识源融合和多分类器表决的中文观点分析. 见: 第三届中文倾向性分析评测会议 (COAE 2011) 论文集. 济南: 中国中文信息学会, 2011. 77−87)

85 Liao Jian, Wang Su-Ge, Li De-Yu, Chen Xin. Using word-formation rules and mutual information for new sentiment word identification in microblogs. In: Proceedings of the 6th Conference on Chinese Opinion Analysis Evaluation. Kunming, China: COAE, 2014. 90−96
(廖健, 王素格, 李德玉, 陈鑫. 基于构词规则与互信息的微博情感新词发现与判定. 见: 第六届中文倾向性分析评测会议论文集. 昆明: 中国中文信息学会, 2014. 90−96)



王 科 南京理工大学计算机学院硕士研究生. 主要研究方向为自然语言处理和文本挖掘.
E-mail: wangkk998@gmail.com
(**WANG Ke** Master student at the School of Computer Science and Engineering, Nanjing University of Science and Technology. His research interest covers natural language processing and text mining.)

夏 睿 南京理工大学计算机学院副教授. 2011 年获得中国科学院自动化研究所博士学位. 主要研究方向为自然语言处理, 机器学习, 文本挖掘. 本文通信作者. E-mail: rxia@njust.edu.cn
(**XIA Rui** Associate professor at the School of Computer Science and Engineering, Nanjing University of Science and Technology. He received his Ph. D. degree from the Institute of Automation, Chinese Academy of Sciences in 2011. His research interest covers natural language processing, machine learning, and text mining. Corresponding author of this paper.)