# 基于语料资源的情感词典扩展方法研究

谢松县，刘博，王挺

（国防科学技术大学 电子科学与工程学院，湖南 长沙 410073 ）

摘要：为了对雷达天线扫描方式进行自动识别，改进开发了天线扫描方式模拟器，并分别研究了电子扫描和机械扫描的特征提取和识别方法。基于最大主瓣脉冲序列的特征参数实现电子扫描和机械扫描的区分，基于单个天线扫描周期脉冲序列的特征参数实现8种机械扫描方式的自动识别。仿真结果表明，该方法能够区分一维电扫、二维电扫和机械扫描，并且采用支持向量机决策树对机械扫描方式的识别正确率高于决策树方法。（小五）

关键词**：**天线扫描方式；自动识别；电子扫描；机械扫描；支持向量机决策树（小五）

中图分类号**：**TN95 文献标志码**：**A　　　文章编号： （小五）

**Automatic recognition method of radar antenna scan type（三号，加粗）**

LI Cheng1, WANG Xuesong2（小四）

(1.College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China；2.College of Science，National University of Defense Technology, Changsha 410073, China) （小五）

**Abstract:** In order to recognize radar antenna scan type automatically, an improved antenna scan pattern simulator is developed in this paper, and features extraction and automatic recognition methods for electronic scan type (EST) and mechanical scan type (MST) are studied respectively. EST is firstly distinguished from MST based on the characteristic parameters extracted from the maximum main beam pulse sequence; then 8 MSTs are automatically recognized based on the parameters extracted from the pulse sequence in a scan period. The simulation results show that it is able to distinguish between one-dimensional EST, two-dimensional EST and MSTs by this method. Moreover, the correct recognition ratio of MSTs by support vector machine decision tree in this paper is higher than that by decision tree. （小五）

**Keywords:** antenna scan type; automatic recognition; electronic scan type; mechanical scan type; support vector machine decision tree（小五）

人们在信息交流中总是很自然地对周围事物表达情感，意见和喜好。当人们接触到关于某个主题或实体的信息，通常都会形成对这些话题（实体）的个人倾向和看法。这些带有主观性内容的信息，无论是书面或者口头的，理解其表达的正确的观点（情感）是极其重要的。因此，意见和情感的理解是人际交往的一个重要方面。多年以来，为了理解人类行为，个体或者群体情感一直得以研究。对情感或观点的群体或者社会化分析迎合了对于某事件或实体对一群人所造成的影响或极化度量需求。社会情绪进行了研究，政治理解和预测选举结果，以及在市场营销，来预测某一个产品的成功，并建议他人。比如在政治领域，可以研究社会化情感来预测选举的结果；在商品市场，通过研究对某些商品社会化观点的可以预测畅销与否，并且向客户推荐。

在社交媒体出现以前，收集带有观点信息的数据费力且达到的规模比较小。社交网络出现后，大规模的用户开始频繁使用社交媒体进行信息交流，收集用户的观点数据就变得简单了。从这方面来说，社交媒体开启了用户之间交互新的可能。尤其是微博平台，可以允许用户实时分享评论和观点。比如Twitter，作为极度受欢迎的微博平台，吸引了上百万用户每天分享数以百万计的帖子。这种数量巨大的丰富的用户产生数据（User Generated Content，UGC）为研究用户行为提供了无尽的机会。

能够从文本中自动识别出观点和情感等主观性信息近年来吸引了学术界和工业界的广泛关注。人工对数量庞大的文本进行分析发现主观信息是不可行的，因此出现了所谓观点挖掘（或成为情感分析，opinion mining, sentiment analysis）研究，可以从自然语言文本中自动识别出用户的观点信息。采用计算方法进行观点情感分析会在不同维度上对情感进行度量。其中使用极性（porlarity）分类方法[1-4]将情感分析问题简化为分类问题，使用有监督或无监督方法将用户的的观点进行积极（positive）、消极（negative）以及中性（neutral）分类。极性分类相关的一类问题时主客观分类问题，就是将主观文本和客观文本分开，一般使用的是有监督的方法[5]。此外，观点强度分析也是研究的方向，比如，SentiStrength[6]，能够从句子层面计算出积极和消极的观点强度值。情绪也可以通过创建情绪词典来进行计算，Plutchik情绪轮盘[7]，提出了四对对立的情绪状态：joy-trust，sadness-anger，surprise-fear和anticipation-disgust。Mohammad等[8]根据Plutchik情绪类别标注了一系列的词发展出了NRC词情绪关联词典。以上的情感分析方法都是在句法层面，还有其他的基于语义知识库在语义层面进行情感分析的方法[9-11]。

语言中，一个词的语义极性（semantic polarity）或倾向性（orientation）表示其对于它的语义组（semantic group）或词汇场（lexical field）规范式的偏离方向[12]。情感分析通常会使用一些标注了积极和消极分数的词汇的词典资源。近年来情感词典资源的构建方法研究已经得到越来越多的计算语言学和自然语言处理研究人员注意。Wilson等[13, 14]对一些英文单词进行了积极和消极类别的标注形成了OpinionFinder词典；Bradley等[1]标注了并发布了情感规范（affective norms）的英文单词词典ANEW，并且Nielsen等[3]在Twitter上应用并扩展了ANEW，形成AFINN词典。Esuli和Sebastiani[15]以及后来Baccianella[2]等在著名的英文词典Wordnet基础上采用自动计算的方式开发出了情感词典SentiWordnet。Thelwall等[6]设计实现了能对词语的情感强度进行估计的词典。Mohammad和Turney[8]根据Plutchik情绪轮分类方法使用情绪分值标注了一些词语形成NRC情绪词典。在2013和2014年举办的SemEval（Semantic Evaluation）评测中，NRC-Canada队利用NRC词典并扩展出两种新的词典，取得了最好成绩[16, 17]。为了克服以上语法层面建立的词典的上下文语境以及领域适应性问题，一些学者提出了基于概念（concept-based）构建情感词典[18]，其中SenticNet是使用常识知识库建立的公开可用的基于概念的情感词典[9]。

中文情感分析研究起步较晚，缺乏普遍认可的可靠的中文情感词典。目前研究使用主要有HowNet情感词典[19]，NTUSD情感词典[20]以及大连理工大学的情感词汇本体词库[21]。这些词典主要是以手工或半自动方式编辑而成。我们前期的工作提出了根据语义词典HowNet语义关系将英文情感词典转换为中文情感词典的方法，并构建了比较全面的情感词典[22]。

基于语义词典的情感词典构建方法是一种通用的情感词典构建方法。采用这种方法的优势在于可以获取大量的情感词语，基于词语的语义关系也易于进行情感倾向计算。但是，基于语义词典的情感词典构建方法受限于语义词典的规模和语义关系的定义，而且对于专业领域中不断涌现的新词语，采用这种方法难以对情感词典进行扩展。随着互联网应用的不断发展，网络上的信息资源不断涌现，研究如何利用语料等信息资源对情感词典进行自动扩展具有十分重要的意义。

本文提出的基于语料资源的无监督的情感词典扩展方法可以使用于无需标注的语料数据对中文情感词典进行扩充。

1 问题描述

通过对相关工作进行研究分析发现，基于语料资源的情感词语选择与倾向计算通常有两种实现思路。

基于语言特征的方法。例如，Hatzivassiloglou和Mckeown[9]等人采用并列或转折连词来判断新的情感词并计算其极性。

基于统计特征的方法。例如，Turney[39]采用点互信息等统计学方法从语料中抽取新的情感词并计算其极性。

基于以上对语料预处理技术和情感词语选择与倾向计算技术的相关分析，本文将基于语料资源扩展情感词典时需要解决的问题描述如下：

对语料进行预处理，将语料资源进行分词处理并标注词性，形成结构化的抽取记录，并根据已有情感词典对词语进行极性标注；

基于语言特征和统计特征相关知识设计情感词语选择和极性计算方法，计算并标注语料中的形容词类和副词类词语的极性值（积极性极值和消极性极值）；

生成情感词典，并将情感词典中的词语分为三类（积极、消极、中性）。

2 数据集及预处理

基于语料资源的情感词典扩展通常包括语料预处理、情感词选择与倾向计算、情感词典生成三个基本过程。本文选取的语料资源是谭松波博士提供的酒店评论、书籍评论和电子商品评论三个领域的语料文本各4000篇。

表4-2　实验使用的资源

|  |  |  |
| --- | --- | --- |
| **词典** | SentiHowNet | 基于第三章工作设计实现 |
| **语料** | Hotel语料 | 4000篇 |
| Book语料 | 4000篇 |
| NoteBook语料 | 4000篇 |

本文研究的对语料进行预处理主要是将语料文本进行分词并进行词性标注。中文分词处理是对语料进行进一步处理的基础。本文在对语料进行分词处理时，采用的是中科院设计实现的ICTCLAS分词软件[23]。

3 方法设计

本章设计的情感词典扩展方法主要包括三个阶段：语料资源预处理、情感词语选择与倾向计算和生成情感词典。情感词典扩展方法设计框架如图4-1所示。

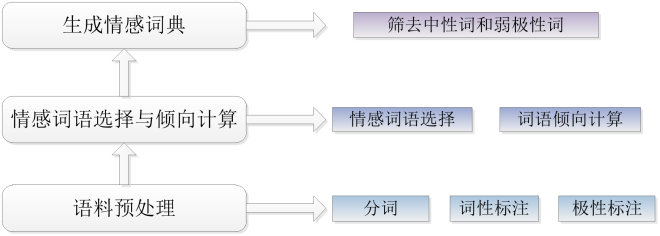


图4-1　情感词典扩展方法设计框架图

语料预处理：语料预处理主要是将语料文本进行分词处理、词性标注，根据分词和词性标注结果生成结构化的抽取记录，并基于已有的情感词典对抽取记录进行极性标注。

情感词语选择与倾向计算：为了设计高效的情感词语选择和倾向计算方法，本文对基于语言特征和基于统计特征的方法分别进行了设计实现和研究分析，设计实现了基于语言特征和统计特征混合的词语抽取和倾向计算方法。

生成情感词典：分析所有标注的情感词语的极性值，去除其中的中性词和弱极性词，生成情感词典。

4 基于语言特征的情感词典扩展

早期关于英文的一些研究[12]发现，由连词（如and或but）连接的两个形容词的极性往往存在一定的关联性，如“and”连接的形容词（如“nice and good”）极性相同，而“but”连接的形容词（如“nice but unnatural”）极性相反。基于语言特征的中文情感词典扩展吸收借鉴了这些研究成果。

4.1 连词选择

连词是用来连接词与词、词组与词组或句子与句子、表示某种逻辑关系的虚词。连词可以表示并列、承接、转折、因果、选择、假设、比较、让步等关系。

通常由并列连词或递进连词连接的两个形容词的极性是相同的，由转折连词连接的两个形容词的极性是相反的。本文主要研究基于表达并列、转折和递进三种关系的连词如何进行情感词典扩展。

* 并列关系连词：和、跟、与、既、同、及、况、况且、乃至、并、也、又；
* 转折关系连词：却、虽然、但是、然而、偏偏、只是、不过、至于、致、不料、岂知；
* 递进关系连词：不但、不仅、何况、并、且、而且。

4.2 基于连词的情感词倾向计算

基于连词的情感词语选取和倾向计算的流程如图4-4所示，主要包括获取待评价词语、连词结构分析和倾向计算三个过程。待评价词采用基于连词的方法能够计算出情感倾向值的选取为情感词语，其情感倾向可以根据计算得到的情感极性值判别。

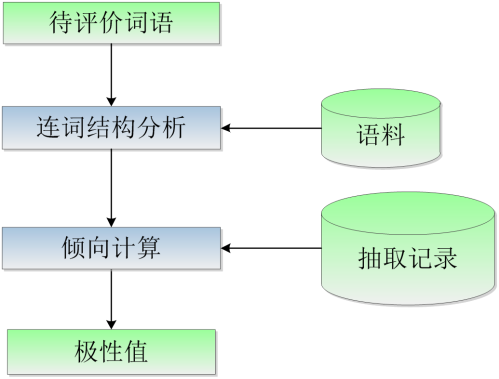


图4-4　基于连词的情感词语选取和倾向计算流程图

在进行情感倾向计算时，遵循以下原则：

位于连词同一侧的形容词或副词具有相同极性；

位于并列连词和递进连词两侧的形容词或副词具有相同极性；

位于转折连词两侧的形容词或副词具有相反极性。

例如，按照以上原则，“颠狂”的情感极性和“枯”相同，积极极性值为0.0，消极极性值为0.3235；“凄苦”的情感极性与“枯”相反，积极极性值为0.3235，消极极性值为0.0。

4.3 实验

经语料预处理得到待评价词语集合，经基于连词的方法进行情感词语选取和倾向计算得到情感词语集合和相应的倾向极值。实验中用于评测参考的人工标注的情感词语是基于人工标注和网络注释（百度百科等）等多种途径综合得到的。

表4-3　各个领域语料的实验结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 待评价词语 | 机器标注 | 正确标注 |
| Hotel语料 | 103 | 51 | 45 |
| Book语料 | 166 | 166 | 112 |
| NoteBook语料 | 61 | 61 | 41 |

针对三个领域（Hotel、Book、NoteBook）的情感词典扩展实验结果如表4-3所示，其中待评价词语是指在各个语料中标注为形容词性和副词性的未被标注极性的词语，机器标注词语指的是采用本文设计算法标注的词语，正确标注词语指的是采用机器标注和对待评价词语采用人工标注极性相同的词语。

使用第二章给出的情感分析的评测标准，参考公式（2.1）、（2.2）和（2.3）对基于语言特征的情感词典扩展方法进行评测，结果如表4-4所示。对于三个语料，其召回率均达到67%以上，正确率达43%以上。其中对于Hotel语料，其召回率最低，为43.69%，其余语料召回率较高。经分析，Hotel语料中可以用于计算的连词结构的语句所占的比例小于其他语料。

表4-4　各个领域性能评测结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 正确率（P） | 召回率（R） | F值 |
| Hotel语料 | 43.69% | 88.24% | 58.44% |
| Book语料 | 67.47% | 67.47% | 67.47% |
| NoteBook语料 | 67.21% | 67.21% | 67.21% |

5 基于统计特征的情感词典扩展

本文设计实现的基于统计特征的情感词典扩展方法主要是采用基于上下文的方法进行情感词语的选取和倾向计算。

词语的上下文是词语在实际应用中的语言环境，它在自然语言处理中的价值体现在两个方面：一方面，在自然语言知识获取的过程中，上下文是知识获取的来源，在相应推理机制下，上下文本身就是知识；另一方面，在自然语言处理的应用问题解决过程中，上下文扮演着解决所需信息和资源提供者的重要角色。特别是在语料库语言学中，各种机器学习方法的引入使词语的上下文成为计算语言学知识获取和问题求解过程中最为重要的资源，在无监督学习方法中更是如此[41]。

通常情况下，上下文的选取时基于核心词左右一定范围进行的，这个固定的范围被称为“窗口”。选择合适的窗口，可以使得上下文的计算提供的信息量足够大，产生的噪声足够小。对于英文，文[42]指出核心词左右5个词的范围可以为词语搭配提供95%的信息，文[43]指出上下文±2是最好的选择，范围进一步扩大后提供的信息量不会有明显的增加且会带来不必要的计算开销。

有许多研究人员基于语料资源的统计特征开展情感分析研究工作。Wiebe[10]等人采用类似相似度分布的词聚类方法在大语料库中完成了形容词性的情感词语的获取。Turney[11]等人提出使用点互信息（Point Mutual Information）的方法来识别各种词性的情感词语。这些基于语料资源统计特征的情感词语获取和判别等技术研究为本文开展相关研究有着重要指导意义。文[44]中对有监督的基于上下文的统计特征的情感词语的识别和极性计算进行了探索。本文设计实现的基于统计特征的情感词典扩展方法是一种基于上下文的无监督的方法。本文的方法首先是对待评价词语，分析其上下文词语的词性，获取其特征向量；其次，根据其上下文特征向量实现情感词语选取和倾向计算方法。

5.1 统计特征选择

本文选取的统计特征为词语的上下文。我们选取的词语的上下文特征是与词语两边相邻的词语的词性。基于以上分析，我们给出特征向量和窗口的概念。

定义4-1　词语的特征向量和窗口。词语的特征向量是指由词语和其相邻上下文词语的词性组成的向量，具体形式见公式（4.1），其中，表示词语的词性，表示与相邻的词语的词性，表示与词语的相对距离，表示窗口，即特征向量中与词语相对距离的最大值。

 　　（4.1）

例如，4.3.4节示例中词语“颠狂”的在窗口为3（）时的特征向量为。

5.2 基于上下文的情感词倾向计算

基于上下文的情感词语选取和倾向计算的流程如图4-5所示，主要包括获取待评价词语、上下文分析、极性分类及倾向计算三个过程。待评价词采用基于上下文的方法能够计算出情感倾向值的选取为情感词语，其情感倾向可以根据计算得到的情感极性值判别。

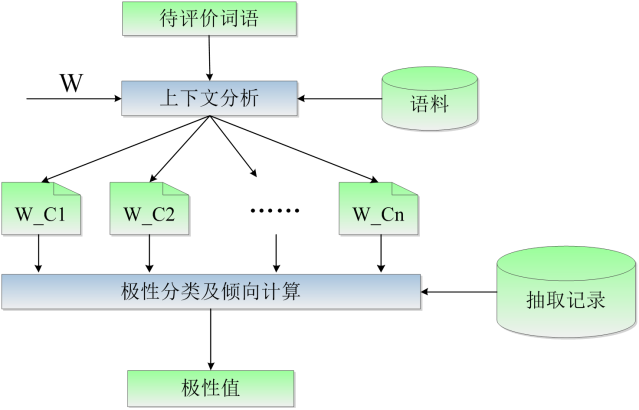
以Book语料中编号（ID）为15609的词语“亮丽”为例对图4-5所示流程进行具体说明。

图4-5　基于上下文的情感词语选取和倾向计算流程图

上下文分析主要是根据设定窗口的值获取待评价词语的特征向量，并从语料资源中获取所有与待评价词语特征相同的词语。例如，设定，则“亮丽”的特征向量为，获取与其特征向量相同的所有词语如表4-5所示。

极性分类及倾向计算

基于上下文特征和抽取记录，对分析得到的与“亮丽”的特征向量相同的词语进行极性分类，并根据分类结果计算待评价词语的极性。对所得词语列表分类是将其分为积极性、消极性和中性三类，待评价词语的极性与记录数最多的一类相同。

例如，经过极性分类，表4-5中的词语最终被分为两类，其中积极倾向词语有14个，消极倾向的词语有6个。故词语“亮丽”的情感极性为积极倾向，其极性值由积极倾向的词语计算得到，其计算方法参考公式（3.1）。经过计算，“亮丽”的积极极性值为0.4297，消极极性值0.0421。

表4-5　特征向量与“亮丽”的相同的词语列表

|  |  |  |  |
| --- | --- | --- | --- |
| W\_C | 词语上下文 | W\_C | 词语上下文 |
| 大 | 没有多大的可读性 | 弱小 | 再贫困弱小的人 |
| 大 | 没有多大的意义 | 贤惠 | 又温柔贤惠的妻子 |
| 简单 | 太浅显简单的东西 | 实 | 真真实实的人 |
| 大 | 没有多大的意义 | 有趣 | 又生动有趣的绘画 |
| 真实 | 比较客观真实的角度 | 直观 | 没有明显直观的效益 |
| 新鲜 | 更多新鲜的元素 | 奇妙 | 更多奇妙的东西 |
| 严格 | 太多严格的界限 | 好 | 更多好的作品 |
| 石破天惊 | 颇多石破天惊之语 | 肮脏 | 太多肮脏的东西 |
| 柔软 | 最脆弱柔软的心脏 | 多 | 反正好多的故事 |
| 粗俗 | 原本低级粗俗的俚语 | 曲折 | 明明惊险曲折的战争 |

5.3 实验

经语料预处理得到待评价词语集合，经基于上下文的方法进行情感词语选取和倾向计算得到情感词语集合和相应的倾向极值。实验中用于评测的参考的人工标注的情感词语是基于人工标注和网络注释（百度百科等）等多种途径综合得到的。

针对三个领域（Hotel、Book、NoteBook）的情感词典扩展实验结果如表4-6所示，其中待评价词语是指在各个语料中标注为形容词性和副词性的未被标注极性的词语，机器标注词语指的是采用本文设计算法标注的词语，正确标注词语指的是采用机器标注和对待评价词语采用人工标注极性相同的词语。

使用第二章给出的情感分析的评测标准，参考公式（2.1）、（2.2）和（2.3）对基于语言特征的情感词典扩展方法进行评测，结果如图4-6、图4-7和图4-8所示。对于三个语料，当窗口时，准确率最高，分别为67.65%、72.89%和72.13%；当窗口时，召回率有所上升，准确率略有下降；当窗口时，召回率最高，准确率和值下降较多。

表4-6　各个领域语料在不同窗口下的实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 窗口大小 | 领域语料 | 待评价词语 | 机器标注 | 正确标注 |
|  | Hotel语料 | 103 | 102 | 69 |
| Book语料 | 166 | 166 | 121 |
| NoteBook语料 | 61 | 61 | 44 |
|  | Hotel语料 | 103 | 94 | 69 |
| Book语料 | 166 | 145 | 110 |
| NoteBook语料 | 61 | 52 | 39 |
|  | Hotel语料 | 103 | 54 | 45 |
| Book语料 | 166 | 78 | 56 |
| NoteBook语料 | 61 | 29 | 17 |

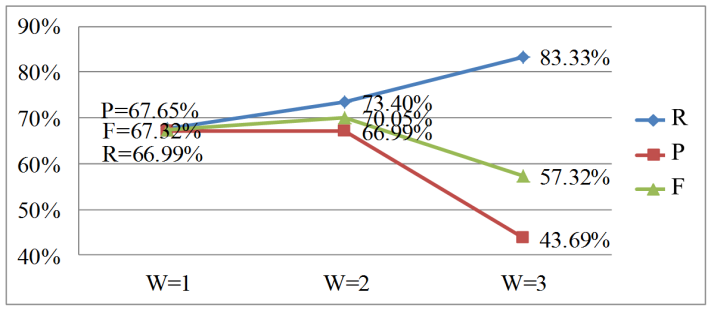
通过对评测结果进行分析，本文发现在设计基于统计特征的情感词典扩展方法时，采用窗口进行情感词语选择，采用窗口进行情感词语倾向计算，可以获得较好的性能。

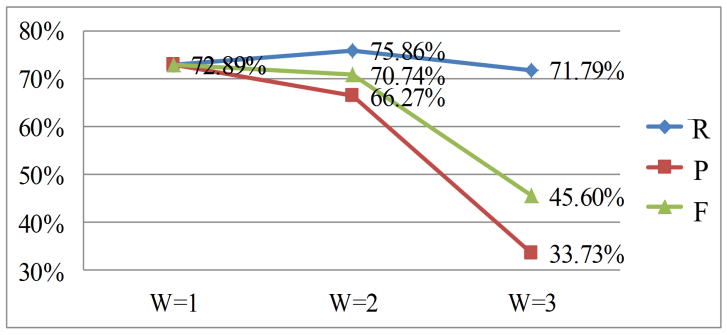
图4-6　Hotel语料评测结果

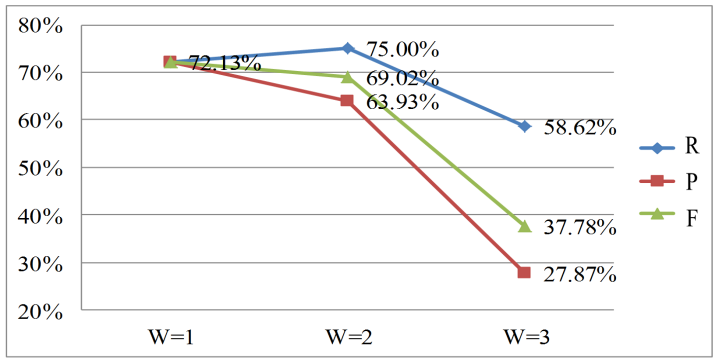
图4-7　Book语料评测结果

图4-8　NoteBook语料评测结果

6 基于混合特征的情感词典扩展

本文对设计实现的基于语言特征的情感词典扩展方法和基于统计特征的情感词典扩展方法的实验结果进行仔细分析发现，采用语言特征无法进行选择和情感倾向计算的词语，可以采用统计特征进行处理；同样的，采用统计特征无法进行选择和情感倾向计算的词语，可以采用语言特征进行处理；两种方法可以相互补充。综合本文4.4节和4.5节的工作，本文提出基于混合特征的情感词典扩展方法。

6.1 基于混合特征的情感词倾向计算

基于混合特征的情感词语选取和倾向计算的流程如图4-9所示，主要包括情感词语选取和词语情感倾向计算两个过程。待评价词语采用基于混合特征的方法能够计算出情感倾向值的选取为情感词语，其情感倾向可以根据计算得到的情感极性值判别。

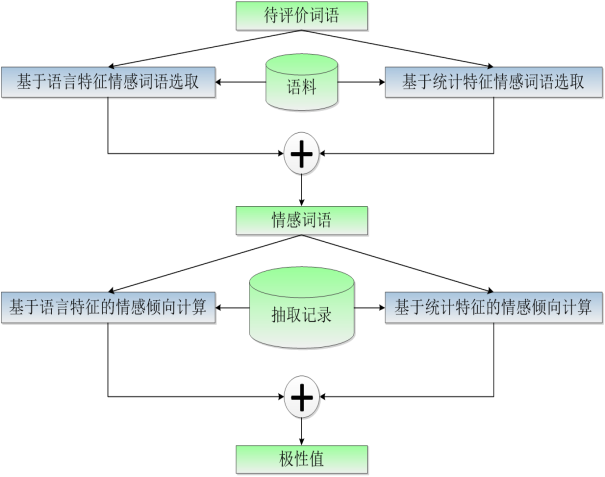


图4-9　基于混合特征的情感词语选取和倾向计算流程图

将选取的情感词语集合分别采用两种方法进行倾向计算，在将两种方法计算的倾向值合成时，遵循以下原则：

优先采用基于统计特征的方法计算出的情感倾向值作为待评价词语的情感极性值。

当采用基于统计特征的方法进行计算时，优先设置窗口大小为2，其次为1。

当采用基于统计特征的方法无法对待评价词语进行情感计算时，采用基于语言特征的方法进行计算。

6.2 实验

针对三个领域（Hotel、Book、NoteBook）的情感词典扩展实验结果如表4-7所示。

表4-7　各个领域的实验结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 待评价词语 | 机器标注 | 正确标注 |
| Hotel语料 | 103 | 102 | 77 |
| Book语料 | 166 | 166 | 128 |
| NoteBook语料 | 61 | 61 | 48 |

使用第二章给出的情感分析的评测标准，参考公式（2.1）、（2.2）和（2.3）对基于语言特征的情感词典扩展方法进行评测，结果如表4-9所示。

表4-8　各个领域的评测结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 正确率（P） | 召回率（R） | F值 |
| Hotel语料 | 75.49% | 74.76% | 75.12% |
| Book语料 | 77.11% | 77.11% | 77.11% |
| NoteBook语料 | 78.69% | 78.69% | 78.69% |

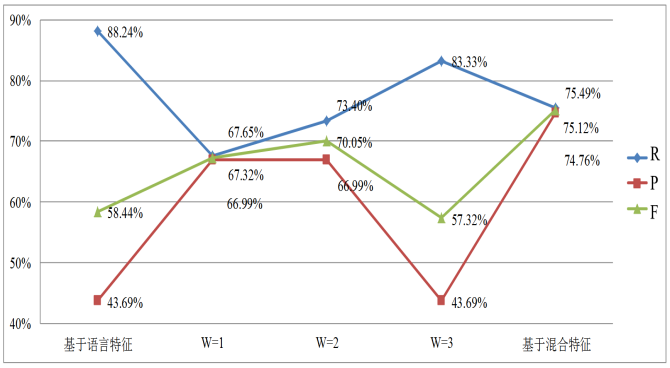
基于语言特征的情感词典扩展、基于统计特征的情感词典扩展和基于混合特征的情感词典扩展的实验评测结果对比情况如图4-10、图4-11和图4-12所示，通过分析发现，基于混合特征的情感词典扩展方法的评测性能是在各个领域预料中均是最优的。因此，可以依据基于混合特征的情感词语选取和倾向计算方法的实验结果扩展情感词典。

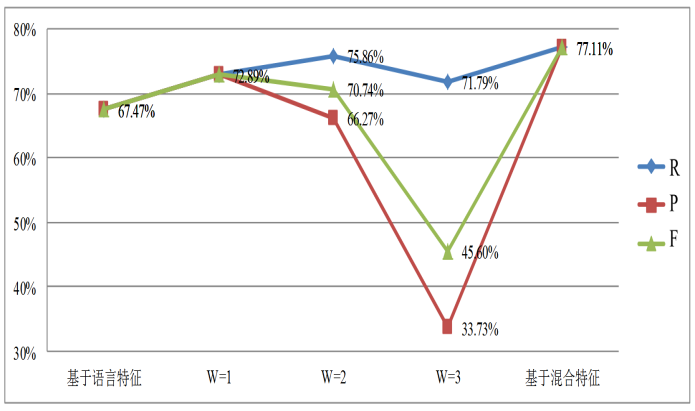
图4-10　Hotel语料评测结果综合比较

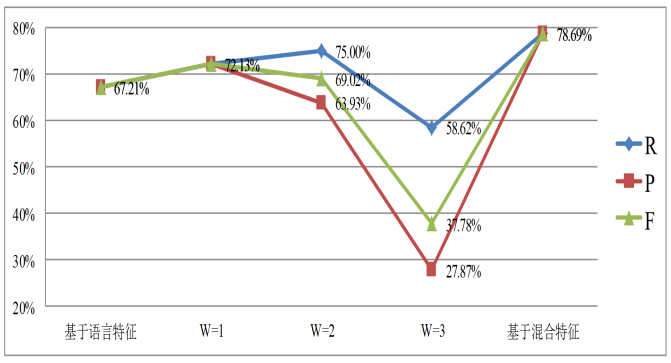
图4-11　Book语料评测结果综合比较

图4-12　NoteBook语料评测结果综合比较

7 结论

本章详细讨论了基于语料资源的情感词典扩展问题描述和方法设计；对基于语言特征的情感词典扩展和基于统计特征的情感词典扩展的关键技术分别进行了研究和实现；并设计了基于混合特征的无监督的情感词典扩展方法。本章还详细分析了每个扩展方法的实验结果，并基于实验结果对情感词典SentiHowNet进行了扩展。

**参考文献(References)**

[1] Bradley M M, Lang P J. Affective norms for English words (ANEW): Instruction manual and affective ratings[R]. Citeseer, 1999.

[2] Baccianella S, Esuli A, Sebastiani F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.[C] 2010: 2200-2204.

[3] Nielsen F A R. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs[J]. arXiv preprint arXiv:1103.2903. 2011

[4] Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision[J]. CS224N Project Report, Stanford. 2009: 1-12.

[5] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis[C] Stroudsburg, PA, USA: Association for Computational Linguistics, 2005: 347-354.

[6] Thelwall M, Buckley K, Paltoglou G. Sentiment strength detection for the social web[J]. Journal of the American Society for Information Science and Technology. 2012, 63(1): 163-173.

[7] Plutchik R. The nature of emotions[J]. American Scientist. 2001, 89(4): 344-350.

[8] Mohammad S M, Turney P D. Crowdsourcing a word--emotion association lexicon[J]. Computational Intelligence. 2013, 29(3): 436-465.

[9] Cambria E, Olsher D, Rajagopal D. SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis[C] //Twenty-Eighth AAAI Conference on Artificial Intelligence

. 2014.

[10] Grassi M, Cambria E, Hussain A, et al. Sentic web: A new paradigm for managing social media affective information[J]. Cognitive Computation. 2011, 3(3): 480-489.

[11] Olsher D J. Full spectrum opinion mining: Integrating domain, syntactic and lexical knowledge[C] 2012: 693-700.

[12] Hatzivassiloglou V, Mckeown K R. Predicting the semantic orientation of adjectives[C] 1997: 174-181.

[13] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis[C] 2005: 347-354.

[14] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis[J]. Computational linguistics. 2009, 35(3): 399-433.

[15] Esuli A, Sebastiani F. Sentiwordnet: A publicly available lexical resource for opinion mining[C] 2006: 417-422.

[16] Mohammad S M, Kiritchenko S, Zhu X. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets[C] Atlanta, Georgia, USA: 2013.

[17] Kiritchenko S, Zhu X, Cherry C, et al. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews[C] 2014.

[18] Tsai A C, Wu C, Tsai R T, et al. Building a concept-level sentiment dictionary based on commonsense knowledge[J]. IEEE Intelligent Systems. 2013, 28(2): 22-30.

[19] 知网HowNet评价词词典[EB/OL]. 2013/2013-07-25[2013-08-15]. http://www.keenage.com/html/c\_index.html.

[20] Ku L W, Chen H H. Mining opinions from the Web: Beyond relevance retrieval[J]. Journal of the American Society for Information Science and Technology. 2007, 58(12): 1838-1850. Wiley Online Library.

[21] 情感词汇本体库[EB/OL]. 2013/2013-07-30[2013-08-15]. http://ir.dlut.edu.cn/EmotionOntologyDownload.aspx.

[22] 谢松县，刘博，王挺. 应用语义关系自动构建情感词典[J]. 国防科技大学学报. 2014, 36(3): 111-115.

[23] 张华平. NLPIR/ICTCLAS2014分词系统[EB/OL]. 2014/2014-06-18http://ictclas.nlpir.org/newsdownloads?DocId=389#.