
基于句法路径的情感评价单元识别^{*}

赵妍妍⁺, 秦兵, 车万翔, 刘挺

(哈尔滨工业大学 计算机科学与技术学院 信息检索研究室, 黑龙江 哈尔滨 150001)

Syntactic Path Based Appraisal Expression Recognition^{*}

ZHAO Yan-Yan⁺, QIN Bing, CHE Wan-Xiang, LIU Ting

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: Phn: +86-451-86413683 ext 800, E-mail: zyy@ir.hit.edu.cn, <http://ir.hit.edu.cn/~zyy/>

Abstract: An appraisal expression is a collocation of a polarity word and its related target, which expresses an evaluative stance towards a target. Recognizing appraisal expressions is a fundamental task for sentiment analysis. However, the existing research is far from enough. Different from the previous work such as hand-crafted pattern or rule based methods, this paper proposes a novel method that uses syntactic paths to automatically recognize the appraisal expressions. In detail, the syntactic paths are first automatically collected to describe the relationships between the polarity words and their corresponding targets, then an edit distance based syntactic path matching strategy is used to improve the performance of the exact syntactic path matching method. Experimental results on the camera and MP3 player domains show that (1) syntactic paths are helpful for appraisal expression recognition, and (2) edit distance based syntactic path matching method can further improve the performance of the appraisal expression recognition.

Key words: appraisal expression; sentiment analysis; syntactic path; polarity word; target

摘 要: 情感评价单元是评价对象及其相关评价词语的搭配.情感评价单元的识别是情感分析领域的一项基础任务.不同于已有的基于手工模板和规则的方法,本文提出了一种基于句法路径的情感评价单元自动识别方法.该方法自动获取句法路径来描述评价对象及其评价词语之间的修饰关系,并通过计算句法路径编辑距离来改进情感评价单元抽取的系统性能.本文的实验语料来自数码相机和 MP3 播放器两个典型的电子产品领域,实验结果表明:(1)句法路径能够有效描述评价对象及其评价词语之间的关系,对情感评价单元的识别有很大帮助,(2)基于编辑距离的句法路径改进策略能够进一步提高情感评价单元识别的系统性能.

关键词: 情感评价单元;情感分析;句法路径;评价词语;评价对象

中图法分类号: TP391 文献标识码: A

1 研究背景

随着 Web2.0 的蓬勃发展,越来越多的用户参与到互联网内容建设的过程中.互联网上产生了大量的带有情

^{*} Supported by the National Natural Science Foundation of China under Grant Nos. 60675034, 60803093 (国家自然科学基金) and the “863” National High-Tech Research and Development of China via grant 2008AA01Z144(863 计划探索类专题项目)

感色彩的评论信息,如:用户针对某一产品或某一事件的主观评论.为了更好的对它们进行加工汇总,并快速找出富有价值的信息,情感分析(sentiment analysis)应运而生.情感分析又称意见挖掘(opinion mining),是一项针对评论信息进行信息识别,挖掘和组织的研究任务.情感分析是一个多学科综合的研究领域,包含多项具有挑战性的研究任务,如:主客观识别^[1],情感分类^[2-4],评价词语识别^[2,4],评价对象识别^[5]等任务.

在以往的研究工作中,研究者们通常将评价词语的识别作为情感分析的一项基本研究任务^[6].具体的,首先利用各种方法识别出评价词语,继而基于评价词语完成情感分析的某些任务.例如在句子级的情感分类任务中,研究者们主要采用两种研究方案:基于情感句打分的方法^[4-5]和基于有指导的分类方法^[7-9].前者主要通过累加情感句中各个评价词的极性来确定情感句最终的情感倾向性.后者则将评价词作为分类的特征,联合其他特征训练分类器来完成情感句的分类.然而,实际上只考虑评价词语在情感分析中的应用是远远不够的,大部分的研究工作忽视了以下两个问题:

Pro1: 出现在情感句中的评价词语并不一定总能表现出一定的情感倾向性.只有当评价词语真正用于修饰评价对象的时候,才是有价值的评价词.例如:在情感句“*Well, this camera did a bad work.*”中,词语“*well*”和“*bad*”单独来看是分别带有褒义和贬义情感倾向性的评价词语.但由于“*well*”在该情感句中不修饰任何评价对象,因此不属于有价值的评价词语.相反,评价词语“*bad*”修饰评价对象“*work*”,是有价值的评价词语.

Pro2: 情感句的情感倾向性并不仅仅由情感句中的评价词语决定,而是由评价词语及其所修饰的评价对象共同决定的.例如:评价词语“*long*”属于上下文相关的评价词,即在不同的上下文中表现出不同的情感倾向性.当评价词语“*long*”修饰评价对象“*life*”时,情感句“*The battery life is very long.*”呈褒义;而当评价词语“*long*”修饰评价对象“*startup*”时,情感句“*The camera has long startup.*”呈贬义.

鉴于此,我们发现,评价词语及其所修饰的评价对象的搭配(如:“*bad-work*”, “*long-life*”, “*long-startup*”)在情感句的倾向性判断上起主宰作用.本文称搭配<评价词语,评价对象>为情感评价单元(Appraisal Expression).情感评价单元可以看作情感分析领域的一个原子任务,表示为借助评价词语对某一评价对象的情感评价.基于此,对于 Pro1,情感评价单元的识别可以帮助获取有价值的评价词语“*bad*”,自动忽略无价值的评价词语“*well*”;而对于 Pro2,则可以帮助上下文相关的评价词“*long*”找到所修饰的评价对象,继而二者共同决定评价词语的情感倾向性.因此,情感评价单元的识别可以有效帮助情感分析各项任务的实现,可以看作情感分析的一个基础研究任务.本文着重研究情感评价单元的识别,而它的情感倾向性识别(如:褒义/贬义)将在未来工作中展开.

情感评价单元的概念由 Bloom^[6]等人首次提出,Wiebe 和 Wilson^[10]也曾提出过相似的概念“*Private State*”.早期的一部分研究者将这项任务分为两个步骤,首先获取情感句中的评价对象,继而选择距离评价对象窗口为 k 的评价词语^[5,7].但是由于这种方法经验性太强,导致系统性能有限.近期的一部分研究者将评价对象和评价词语的识别合并为一个独立的任务,发掘二者之间的关系,提出了基于模板^[11]或规则^[6,12-14]的方法来识别情感句中的情感评价单元.这种方法大大提高了识别的准确率,但由于模板或规则是手工构建的,召回率不高.

鉴于已有的情感评价单元识别方法存在的缺陷,本文提出了一种基于句法路径的方法自动识别情感句中的情感评价单元.本文提及的句法路径是指链接评价词语和评价对象的一种有向句法结构.事实证明,它可以很好的描述评价词语和评价对象之间的句法关系.具体的,本文首先提出了一种自动采集大量句法路径的方法,继而基于句法路径精确匹配算法来自动获取情感句中的情感评价单元.进一步的,本文还提出了一种基于编辑距离的句法路径匹配改进策略来提高系统的性能.实验证明,本文提出的基于句法路径的方法对于情感评价单元的识别是有效的,即句法路径能够较好的挖掘评价词语和评价对象之间的关系.利用本方法,在数码相机和 MP3 播放器两个电子产品领域的情感句测试集上, F -score 分别达到 82.87%和 83.22%,显著超过了以往的相关方法.

本文后续章节组织如下:第 2 章介绍相关工作;第 3 章介绍基于句法路径的情感评价单元识别方法;第 4 和第 5 章分别对本文的实验设置和实验结果进行了详细描述;第 6 章为结论及对未来工作的展望.

2 相关工作

本文的研究任务和 Hu^[5]等人以及 Kim^[7]等人的工作很类似,目的都在于获取评价词语及其所修饰的评价对象.他们首先通过某些方法获取情感句的评价对象.继而,Hu 等人选取距离评价对象最近的形容词作为其对

应的评价词语,而 Kim 等人在距离评价对象窗口为 k 的范围内选择评价词语.但是很显然,他们的方法经验性太强,很多时候评价词语和评价对象距离较远.例如:在情感句“*Well, the picture taken by Cannon camera seems to be good.*”中,真正修饰评价对象“*picture*”的评价词语并不是距其较近的“*well*”,而是距其较远的“*good*”.因此,这种方法具有很大的局限性.

另外,一些学者通过发掘评价词语和评价对象之间的关系来获取情感评价单元.Kobayashi^[11]等人定义了一个相似的评价单元三元组<evaluated subject, focused attribute, value>,其中“focused attribute”对应情感评价单元中的评价对象,“value”对应评价词语.他们将评价词语和评价对象之间的修饰关系用 8 个共现模板(如:<Attribute> of <Subject> is <Value>)来描述.然而,由于模板太过简单且修饰关系仅仅停留在词表面,在模板匹配的过程中,该方法产生了大量的候选评价词语和候选评价对象,需要人工筛选来完成情感评价单元的获取.

为了深入挖掘评价词语和评价对象之间的修饰关系,一部分学者尝试使用句法关系.Bloom^[6]等人利用 Stanford Parser 手工构建了 31 条句法规则(linkage specification)来描述评价词语与评价对象之间的关系(如:*target* \xrightarrow{nsbj} x \xleftarrow{dobj} y \xleftarrow{amod} *polarityword*).此外,Popescu^[12]等人利用 MINIPAR Parser 手工构建了 10 条依存句法抽取模板来获取情感评价单元.同时,国内的姚天昉等人基于依存句法分析总结出“上行路径”和“下行路径”的匹配规则^[13];后续研究利用“SBV”、“VOB”及“ADV”三种结构总结出 SBV 极性传递的一些规则^[14],用于情感评价单元的识别.可以看出,他们的工作融入了更多对评价对象和评价词语之间深层关系的挖掘,避免了单纯停留在词表面上.但由于匹配规则或模板的制定参与了过多的人工,覆盖率较低,且较难重现.

在情感分析的基础任务研究中,还有一些研究者致力于评价短语(“subjective expression”^[15] or “appraisal group”^[16])的识别和分类.评价短语具体指连续出现的一组评价词语,如:“*very good*”或“*not terribly funny*”.此外,Moilanen^[17]等人和 Choi^[18]等人提出了“组合语义单元”(compositional semantics)的概念.组合语义单元表现为一组词,通过相互作用来表达某种情感倾向性.如:在情感句“*[I did [not] have any [doubt] about it.]⁺*”中,词语“*not*”和“*doubt*”都隐含有否定的含义.双重否定表示肯定,因此情感句最终呈褒义.词语“*not*”和“*doubt*”的组合即“组合语义单元”.然而,以上的这几项工作虽然关注了含有某些潜在倾向性的评价词语或评价短语,较好的解决了前述的 Pro1,但是同样忽略了 Pro2 中评价词语及其评价对象的搭配对情感句倾向性等情感分析任务的影响.

3 基于句法路径的情感评价单元识别

3.1 句法路径

句法路径是指在句法树上链接任意两个节点之间的句法结构.在本文中,句法路径特指在短语结构句法树上链接评价词语和评价对象的有向句法结构.图 1 分别显示了 Sen1 和 Sen2 中的句法路径.

Sen1: The camera's image is perfect.

Sen2: The camera has very perfect image.

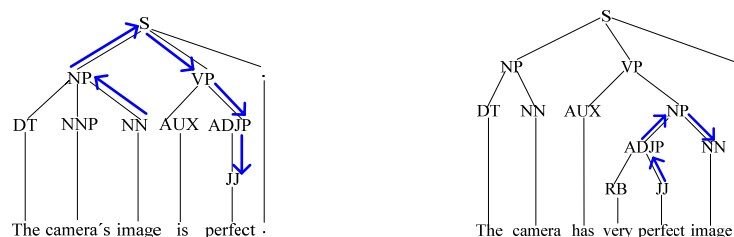


Fig.1 Examples of syntactic paths

图 1 句法路径示例

通过观察 Sen1 的句法树,可以发现评价对象“*image*”和评价词语“*perfect*”之间的句法路径描述了一种特殊的“主系表关系”,表示为 $NN \uparrow NP \uparrow S \downarrow VP \downarrow ADJP \downarrow JJ$.而在句子 Sen2 的句法树中,评价对象“*image*”和评价词语“*perfect*”之间的句法路径则描述了一种“修饰语-中心语”关系,表示为 $JJ \uparrow ADJP \uparrow NP \downarrow NN$.可以看出一方面,句法

路径不仅能够表示出评价对象及其评价词语之间的句法关系;另一方面,由于本文的句法路径采用了自动构建的方式,相比其他的手工构建的句法规则更为全面,即能覆盖更多的句法关系,召回率更高.基于此,本文首先自动采集有效的句法路径构建句法路径库,然后采用基于句法路径的匹配算法来识别情感句中的情感评价单元.

3.2 构建句法路径库

句法路径库的构建主要分为两步,即:句法路径的生成与句法路径的泛化.

在句法路径的生成步骤中,首先需要确定情感句中的评价词语和评价对象.其中,本文涉及到的评价词语来自于 HowNet (<http://www.keenage.com/>) 的英文评价词典,共包含 4,490 个评价词语.由于电子产品领域的评价对象一般呈现为名词(例如:“image”, “lens”)或代词(例如:“it”, “they”),因此本文设定情感句中词性为名词(NOUN)或代词(PRON)的词语为候选评价对象¹.详细的词性描述如下所示:

- NOUN: NN, NNS, NNP, NNPS
- PRON: PRP

基于此,若一个情感句中含有 n 个评价词语, m 个候选评价对象,则会产生 $n*m$ 个候选句法路径.为了从大量的候选句法路径中统计出真正有价值的句法路径,本文在数码相机领域构造了一个含有 20,000 个情感句的无人工标注的大规模语料库².我们为语料库中的每个情感句通过匹配评价词典找出评价词语,通过词性分析找出候选评价对象,进而自动获取其含有的所有候选句法路径.通过这种方法获取的候选句法路径是大量的且纷繁复杂的,他们中的一些由于细小的差别而成为两种句法路径,如路径“JJ↑ADJP↑NP↓NN”与路径“JJ↑ADJP↑NP↓NNS”.为了挑选出有代表性的句法路径,本文做了如下两步泛化.

Gen1: 如果一个句法路径含有一连串相同的句法成分标签(例如:VP,JJ,NNP 等都称为句法成分标签),用一个标签取代之.

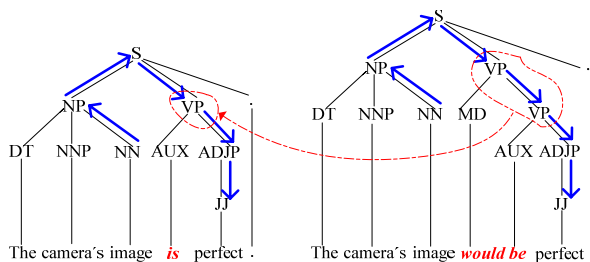


Fig.2 Gen1 between syntactic paths

图2 句法路径的第一种泛化示例

在某些情况下,一连串相同的句法成分标签和单独一个句法成分标签描述了相同的句法功能,可以使用单独的一个句法标签取而代之,完成泛化.例如,在图2中,右边的句法路径“NN↑NP↑S↓VP↓VP↓ADJP↓JJ”所描述的评价对象和评价词语间的句法关系完全可以被左边的句法路径“NN↑NP↑S↓VP↓ADJP↓JJ”而代替.这是因为左右两个情感句的句子结构是相似的,只有谓词所指示的句法成分标签不同,但这完全不影响左右两个情感句中评价词语和评价对象之间句法关系的相似性.因此,我们可以使用一个单独的VP↓来代表两个句子中的谓词.

Gen2: 如果某些句法成分标签表达了相似的含义,用一个规范化的标签将它们代替.

如图3所示,左右两个情感句的结构是完全相同的.但由于右边句子中形容词的比较级“better”和名词复数

¹ 也可以采用其他的评价对象获取方案,比如其他词性的词语(如,动词)或是短语也可作为评价对象.在这里,本文为简单起见,仅选取常用的名词和代词作为评价对象的代表来介绍情感评价单元的抽取方法.同样的抽取方法可完全适用于其他词性的词语(如,动词)或是短语作为评价对象的情况.因此,评价对象获取方案可以看作一个可替换模块,并不影响本文方法的执行.

² 这些情感句来自于一个著名的产品在线评价网站 epinions (<http://www.epinions.com/>).其中,每个情感句的平均长度在 30 个词左右,大部分情感句含有多个分句(平均 3 个左右,大多以“,”隔开).本文对这些情感句没有做任何处理,尽量反映用户评论的真实面貌,进而在此基础上进行无指导的句法路径统计.

“images”的使用,导致左右两个情感句中句法路径的不同.通过观察,我们发现左边句子的“JJ”和右边句子的“JJR”在句法树中的功能相同,而“NN”和“NNS”也是如此.因此,完全可以将“JJR”用归一化的标签“JJ”代替,将“NNS”用归一化的标签“NN”代替.那么,句法路径“JJR↑ADJP↑NP↓NNS”可被泛化为“JJ↑ADJP↑NP↓NN”.具体的句法成分及其泛化标签如表 1 所示.

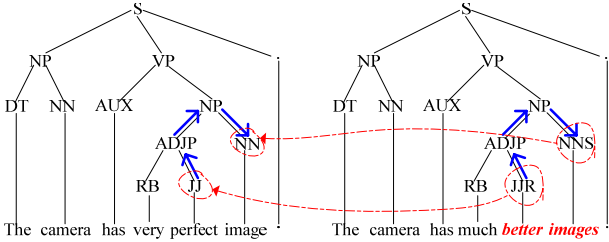


Fig.3 Gen2 between syntactic paths
图 3 句法路径的第二种泛化示例

Table 1 The generalization of similar syntactic constituents
表 1 相似句法成分标签的泛化

泛化标签	相似句法成分标签
JJ	JJR, JJS
NN	NNS, NNP, NNPS, CD
RB	RBR, RBS
VB	VBD, VBG, VBN, VBP, VBZ, VV
S	SBAR, SBARQ, SINU, SQ

在为所有的候选句法路径完成了两步泛化之后,本文统计出各种句法路径的出现频率,并通过阈值 th_{path} 的限定来选取较为频繁的候选句法路径构成句法路径库.这种通过考察频率来确定句法路径正确性的方法,基于下面的一个假设:

评价词语与其真正具有搭配关系的评价对象之间满足一定的句法关系,并且这些句法关系是有规律可循的、可总结的,而并非杂乱无章的.

正是基于这个假设,很多研究者^[6,12-14]使用人工总结句法关系的方法来进行情感评价单元的识别.同时,他们较为成功的研究工作也证明了该假设的正确性.文本提出的句法路径可以看作句法关系的一种表现形式,因此,同样基于这条假设,在较大规模的语料库(20,000 句情感句)中,评价词语及其真正有搭配关系的评价对象之间的句法路径也是有规律可言的,是可归纳总结的;换言之,这类正确的句法路径在大语料库中出现的次数将会较多,因此统计频率也会较高.相反,由于评价词语和那些并非真正存在搭配关系的评价对象之间的句法路径没有规律可言,句法路径形式也较为杂乱无章,因此这些错误的句法路径的频率统计值将较低.基于此,在大语料库上统计的这些句法路径(经过泛化处理后的)的频率,能够较为真实的反应句法路径的正确性和可用性.

Table 2 The selected frequent syntactic paths
表 2 高频句法路径示例

句法路径	出现频率(次数)	位置
JJ↑NP↓NN	26,223	Front
NN↑NP↑S↓VP↓ADJP↓JJ	17,528	Back
NN↑NP↑S↓VP↓NP↓JJ	5,413	Back
JJ↑NP↑PP↓NP↓NN	5,142	Front
.....

表 2 列出了一些最为频繁的句法路径示例.其中第二列给出了该句法路径在语料库中的出现次数,同时也代表了它的匹配优先级,即频率越高,优先级越高.当一个评价词语在寻找其所修饰的评价对象的时候,优先级高的句法路径将被优先选择.最后一列“位置”具体描述了评价词语在句法路径中的位置,“Front”代表了评价词语位于句法路径的最前面,即评价词语先于评价对象出现;而“Back”代表了评价词语位于句法路径的最后面.

3.3 基于句法路径匹配的情感评价单元识别

一个情感评价单元包括两个部分:评价词语和评价对象.本文首先通过情感评价词典(详见 3.2)获取情感句中的评价词语.为了找到与这些评价词语相搭配的评价对象,本文采取了一种句法路径精确匹配(exact matching)的方法.假定要获取情感句中某一评价词语 $w_{polarity}$ 所修饰的评价对象 w_{target} .首先我们将情感句进行句法分析,获取句法树;继而,统计 $w_{polarity}$ 和句中所有候选评价对象(名词或代词)所产生的候选句法路径,并进行两步泛化;将泛化后的候选句法路径逐条和句法路径库中的标准句法路径进行精确匹配,匹配优先级高的候选句法路径所链接的评价对象为最终的 w_{target} .当然,如果这种精确匹配的方法无法成功匹配上句法路径库中的任一条句法路径,则 $w_{polarity}$ 在该情感句中不视为有价值的评价词语.

虽然这种句法路径精确匹配的方法可以准确识别出情感句中的情感评价单元,但是仍然存在如下两点不足,导致召回率不高.

Lim1: 句法路径库仅包含有限条句法路径,并不能涵盖所有种类的句法关系实例.因此,句法路径精确匹配算法会丢失某些潜在的情感评价单元.

Lim2: 句法分析器的分析结果并不完全准确³,存在一些错误.精确匹配算法对这些错误的句法分析结果比较敏感,因此不能避免(或减少)由于错误的句法分析结果而导致丢失的情感评价单元.

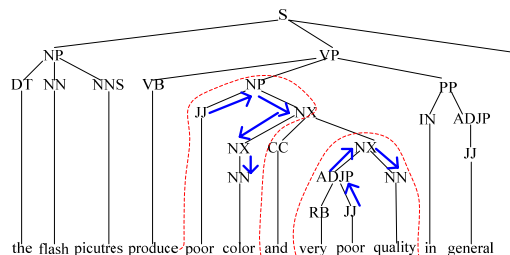


Fig.4 Syntactic analysis result containing two limitations

图 4 包含两点不足的句法分析结果

图4给出了这两点不足的实例。左边虚线括起的部分反映了 Lim2。我们可以发现,评价词语“poor”和评价对象“color”之间的句法路径为 JJ↑NP↓NX↓NN,由于错误的句法分析结果导致该句法路径是错误的,即并不存在于句法路径库中。右边虚线括起的部分反映了 Lim1。评价词语“poor”和评价对象“quality”之间的句法路径表示为 JJ↑ADJP↑NX↓NN,虽然该路径是正确的,但由于出现频率较低,因此也不在句法路径库中。因此,基于句法路径精确匹配的方法将会丢掉该情感句中的两个情感评价单元“poor-color”和“poor-quality”。

为了改进句法路径精确匹配的不足,本文提出一种基于编辑距离的句法路径匹配算法.通过观察,我们发现句法路径库中的路径“JJ↑NP↓NN”和路径“JJ↑ADJP↑NP↓NN”可以分别近似的描述图 4 中的两个情感评价单元所在的句法路径“JJ↑NP↓NX↓NN”和“JJ↑ADJP↑NX↓NN”.因此,本文将这类句法路径 $path$ (如:JJ↑NP↓NX↓NN 或 JJ↑ADJP↑NX↓NN)和句法路径库中与之最相近的句法路径 $path'$ (如: JJ↑NP↓NN 或 JJ↑ADJP↑NP↓NN)进行编辑距离的计算来获取他们之间的相似度,如图 5 所示.如果 $path$ 与 $path'$ 之间的编辑距离小于阈值 th_{ed} ,则本文近似认为 $path$ 所链接的评价词语和评价对象组成了最终的情感评价单元.

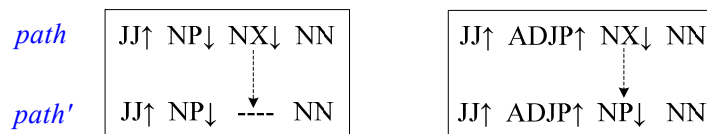


Fig.5 Examples of edit distances between syntactic paths

图 5 句法路径之间的编辑距离计算示例

例如:世界上著名的 Charniak Parser 的准确率/召回率在 90%左右.

编辑距离计算表现为一个动态规划算法^[19].两个句法路径 $path_1$ 和 $path_2$ 之间的编辑距离被定义为将 $path_1$ 转换为 $path_2$ 的最小操作数,包括插入、删除和替换三种操作.考虑到句法路径的方向性(“↑”和“↓”),本文将句法路径中的句法成分和其后的方向指示作为基本的计算单元(如:图 5 中的“JJ↑”).基于此,我们可以通过计算两个句法路径的编辑距离来考察二者的相似程度.按照此算法,图 5 左侧图的句法路径“JJ↑NP↓NX↓NN”与“JJ↑NP↓NN”之间的编辑距离为 1,表现为图中用虚线指示的一个删除操作;相似的,右侧图中的句法路径“JJ↑ADJP↑NX↓NN”和“JJ↑ADJP↑NP↓NN”之间的编辑距离为 1,表现为虚线指示的一个替换操作.

4 实验设置

4.1 实验数据及评价方法

由于国内外没有公开发布的情感评价单元测试语料,本文首先通过人工标注的方法构建了一个测试集.该测试集主要包括来自 epinion 的数码相机和 MP3 播放器两个典型的电子产品领域的英文情感句各 600 句.将其交由两名标注者手工标注其中的情感评价单元.本文使用 $Kappa$ ^[20]值来计算两名标注者标注结果的一致性.结果显示这两个领域标注结果的 $Kappa$ 值均大于 0.8,说明两名标注者的一致性很高,语料可用.

Table 3 Statistical analysis of the sentiment sentence test data

表 3 情感句测试集的统计结果

统计项目	数码相机	MP3 播放器
所含情感句总数	600	600
含有情感评价单元的情感句总数	553	466
所含情感评价单元总数	894	703
所含评价词语总数	979	765
情感评价单元个数/句	1.49	1.17
评价词语个数/句	1.63	1.28

本文对这两个领域手工标注过的测试集进行了详细的统计分析,如表 3 所示.针对数码相机领域的测试集而言,我们发现 90%以上的情感句中含有情感评价单元.这说明情感评价单元是情感分析领域的一个重要的基本单元.因此,情感评价单元的正确识别有助于情感分析相关任务的解决.此外,该测试集情感句中不存在 979 个评价词语,但是仅构成了 894 个情感评价单元.这说明约有 100 个左右的评价词语在其所在的情感句语境中并不是真正有价值的评价词语.这种现象与本文初始提及的想法相吻合,即出现在情感句中的评价词语并不一定总能表现出一定的情感倾向性.只有当其真正用于修饰评价对象的时候,才是有价值的评价词语.因此,在情感分析任务中,不加分析而直接使用情感句中的评价词语是粗糙的.同时,这也可以进一步证明情感评价单元识别的重要性.同样,从表 3 可以看出,在 MP3 播放器测试集上的统计结果与数码相机领域是类似的.

本文使用信息检索领域标准的评价准则:准确率 P ,召回率 R 和调和评价值 F 来对情感句中的情感评价单元进行评价.具体定义为: $P = |A \cap B| / |A|$; $R = |A \cap B| / |B|$; $F = 2 * P * R / (P + R)$.其中, A 表示系统识别出的情感评价单元集合, B 表示人工标注的情感评价单元集合.

4.2 基线实验

本文选取了两种已有的情感评价单元识别方法作为基线实验.

Nearest:在很多情感分析系统中,人们近似的认为评价词语修饰距其最近的评价对象^[5],因此本文拟定一个基线系统 Nearest.具体的,我们首先通过使用评价词典(详见 Section 3.2 提及的 HowNet 英文评价词典,同一个评价词词典)识别出情感句中所有的评价词语;继而距离评价词语最近的名词或代词(具体的名词和代词词性表示见 Section 3.2)被挑选作为其对应的评价对象;二者的组合作为最终的情感评价单元.

Rule_based:本文重现了 Bloom 等人的方法作为基于规则/模板的基线系统代表.由于 Bloom 等人提出情感评价单元的概念,且其手工构建的 31 条句法规则^[6]较其他研究者的工作更为细致,因此本文仅考虑其为基线系统进行对比实验.首先,本文使用同一评价词词典识别出情感句中所有的评价词语;继而使用 Bloom 等人的 31 条句法规则为情感句中的评价词语找出其所对应的评价对象,二者的组合作为最终的情感评价单元.

5 实验与分析

在实验中,本文选取 Charniak 短语句法分析器^[21]对情感句进行句法分析.此外,为了选取最优的 th_{path} ,本文观察集合 $th_{path}=\{10, 30, 50, 70, 90, 100, 150, 200\}$ 在基于句法路径精确匹配的情感评价单元识别中的性能.发现当 $th_{path}=70$ 的时候,即取前 70 条最频繁出现的句法路径时,系统性能达到最优.

5.1 对基于句法路径精确匹配的情感评价单元识别方法的评价

表 4 列出了本文提出的基于句法路径精确匹配算法与两个基线系统 Nearest 和 Rule_based 的对比实验结果.很显然,在两个领域的测试集上,我们的方法在各个指标上的性能都显著超出了这两个基线系统.

Table 4 Comparison between our appraisal expression recognition approach and two baselines
表 4 情感评价单元识别中各种方法的对比实验结果

方法类别	Camera			MP3 player		
	<i>P</i> (%)	<i>R</i> (%)	<i>F</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F</i> (%)
Nearest	65.80	65.44	65.62	60.65	66.00	63.22
Rule_based (Bloom)	85.16	66.11	74.43	84.88	63.87	72.89
Exact path matching (our)	86.19	78.19	81.99	85.53	76.53	80.78

分析表 4 中的结果,我们发现,对于每个领域,基线实验 Nearest 的准确率均较低.这是由于 Nearest 方法获取评价对象的过程太过经验性,缺乏对情感句内部深层信息的深入挖掘.相反,该方法默认每个出现的评价词语都是有价值的,而且为每个出现的评价词语强制匹配一个评价对象.如表 5 所示的情感句实例 1 和实例 2,其中第二列中标蓝的词语代表 Nearest 方法识别出的评价对象,标红的代表评价词语,箭头代表修饰关系.我们可以看出,由于 Nearest 方法的局限性,对于情感句 1 中的评价词语“friendly”,该方法错误的找到距其最近的名词“user”作为评价对象;而对于情感句 2 中的评价词语“well”,该方法也错误的为其强制匹配一个评价对象“camera”.相比而言,本文提出的基于句法路径的方法则可以很好的解决这些问题.借助句法路径“NN↑NP↑S↓VP↓ADJP↓JJ”为情感句 1 的评价词语“friendly”找出正确的评价对象“interface”,并借助句法路径“JJ↑NP↓NN”为情感句 2 找出标准情感评价单元“bad-work”,并舍弃掉错误的情感评价单元“well-camera”.

Table 5 Comparison between sentiment sentence examples
表 5 情感句实例对比

编号	情感句	标准情感评价单元	所用句法路径
1	The interface of this camera is very user friendly .	friendly-interface	NN↑NP↑S↓VP↓ADJP↓JJ
2	Well , the camera did a bad work .	bad-work	JJ↑NP↓NN
3	Very satisfied with this camera's picture.	satisfied-picture	JJ↑ADJP↓PP↓NP↓NN
4	The only drawback is the speed between shots.	drawback-speed	NN↑NP↑S↓VP↓NP↓NN

同样从表 4 中,我们发现对于每个领域,基线实验 Rule_based 的实验效果均要好于 Nearest.这间接反映出深入挖掘评价词语和评价对象之间的关系要比单纯考虑他们在情感句中的位置有价值的多;同时也说明使用句法关系挖掘二者之间的关系是有效的.尽管如此,虽然 Rule_based 的准确率 *P* 较高,但召回率 *R* 却很低.相比而言,在这两个领域的测试集上,本文提出的基于句法路径的方法均可取得较好的系统性能,特别是对于召回率 *R*.这充分证明了本文的句法路径的有效性.具体的,句法路径足够的全面和准确,能够较好的挖掘评价词语和评价对象之间的句法关系,且能够在兼顾准确率的同时提高情感评价单元识别的召回率;其次,传统的评价词语和评价对象之间的关系获取往往靠手工获取,而本文的句法路径的获取方法则完成了二者关系的一种自动获取.如表 5 所示,由于基线实验 Rule_based 方法中手工整理的句法关系种类的局限性,无法识别情感句 3,4 中的情感评价单元.而本文的基于句法路径的方法则可以通过统计大量的句法路径,自动的发掘出更多的评价对象和评价词语之间的句法关系,继而提高情感评价单元识别的召回率.如分别使用句法路径“JJ↑ADJP↓PP↓NP↓NN”和“NN↑NP↑S↓VP↓NP↓NN”,能够准确找出情感句 3,4 中的情感评价单元“satisfied-picture”,“drawback-speed”.

综上所述,通过对表 4 和表 5 中的数据和实例的深入剖析,足以证明本文通过考察句法路径的频率来选取的

这些高频的句法路径在大部分测试实例中发挥了积极的作用,也间接印证了这种句法路径获取方法是可行的。此外,虽然实验中用到的句法路径是在数码相机领域获取的,但表4显示,这些路径同样适用于MP3播放器领域,并取得了较好的实验效果。这足以证明本文的句法路径具有领域普适性,可以很容易的移植到其他领域中。尽管如此,与数码相机领域的三种方法的实验结果相比,MP3播放器领域的结果均略低一些。分析结果表明,主要原因为来自于实验测试集语料的差别,MP3播放器领域的语料书写格式更为随意,处理起来更为复杂一些。

通过对表4中数码相机和MP3播放器两个领域的实验结果的观察,我们发现基于句法路径精确匹配的情感评价单元识别方法的召回率 R 也并不高,分别为78.19%和76.53%。为了进一步提高召回率,避免由于句法路径精确匹配而带来的 $Lim1$ 和 $Lim2$,本文采用了基于编辑距离的句法路径匹配算法来改进系统的性能。

5.2 对基于编辑距离的句法路径匹配算法的评价

基于编辑距离的句法路径匹配算法将句法路径精确匹配算法灵活的转化为一种模糊匹配,更进一步的使句法路径在情感评价单元识别中发挥作用。在该策略的实现过程中,本文需要筛选出最优的编辑距离参数 th_{ed} 。图6描述了随着参数 th_{ed} 的变化,情感评价单元识别系统分别在两个领域上的性能变化曲线。

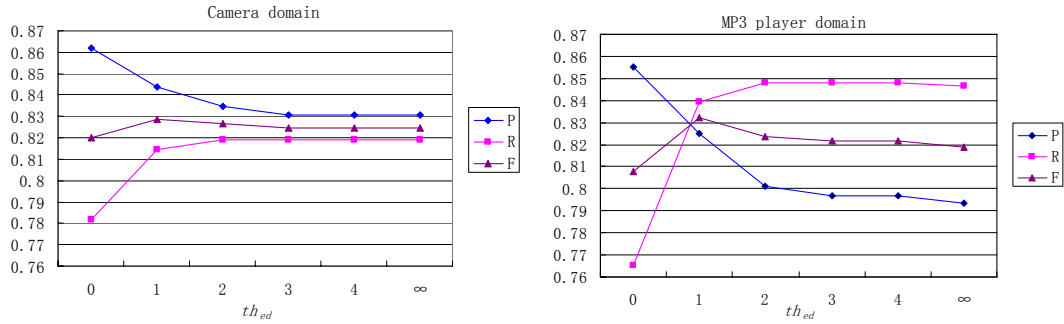


Fig.6 The performance curves changing with th_{ed} for appraisal expression recognition

图6 随参数 th_{ed} 变化而产生的情感评价单元识别系统的性能变化曲线

观察图6中两个领域的的数据曲线,我们可以看出表示准确率 P 的曲线值在持续下降,而表示召回率 R 的曲线值在不断上升。当 $th_{ed}=1$ 时,情感评价单元识别系统在这两个领域中的性能均达到最优, F -score分别达到了82.87%和83.22%。同时,召回率得到了显著的提升,较精确匹配的召回率分别提高了约3%和7%。为了测试该改进方案的系统性能显著性,本文使用了 χ^2 显著性检验。检验结果表明, $th_{ed}=1$ 时的情感评价单元识别系统的系统性能较句法路径精确匹配方案的性能提高是显著的(数码相机领域: $p<0.05$,MP3播放器领域: $p<0.01$)。

我们还可以发现,当 $th_{ed}=0$ 时,系统即为基于句法路径精确匹配的情感评价单元识别系统。相对最优的 $th_{ed}=1$ 而言,虽然它的准确率 P 相对高一些,但由于匹配过于严格,将导致召回率 R 的大大降低。当 $th_{ed} \rightarrow \infty$ 时,系统转变为,为情感句中的每个评价词语都必须选出一个评价对象,无论所计算出的句法路径编辑距离是多少。观察此阈值的系统性能,我们发现召回率 R 得到了少许提高。但是由于情感句中出现的某些评价词语并不是有价值的评价词语,而该方法强行为其匹配一个评价对象,导致准确率 P 的大幅下降。尽管如此,观察图6中的数据曲线,我们发现 th_{ed} 取任意值,基于编辑距离的句法路径匹配算法的性能都要优于基于句法路径精确匹配的方法。这也进一步证实了本文的基于编辑距离的句法路径匹配改进策略有助于情感评价单元的识别。

5.3 句法路径的数量分布对情感评价单元识别的影响

如前面所述,句法路径的使用大大提高了情感评价单元识别的系统性能。但事实上,并不是句法路径的数量越多,系统的性能就越好。以数码相机领域为例,图7描述了随着句法路径数量阈值 th_{path} 的变化,系统性能的变化曲线($th_{ed}=1$)。其中,句法路径按出现频率即匹配优先级排序,如“10”代表取最频繁的10条句法路径进行实验。

从图7中可以看出,当 th_{path} 在1到10之间变化时, F -score提高得非常迅速,即本文选取的前10条最频繁的句法路径即可以产生出较高的 F -score,为78.63%。这同时证实了这前10条句法路径在实际情况中也是最频繁使用的,可以覆盖大多数的情感评价单元的类型。当 th_{path} 从10变化到70时, F -score的增长速度越来越慢,且

当 $th_{path}=70$ 时,系统性能达到最优,为 82.87%。我们发现虽然使用的句法路径数量陡然增多了,但这些句法路径只产生了大约 4% 的 $F-score$ 增长率。其实这种现象不难解释。情感评价单元中的评价对象和评价词语之间的关系是有规律可循的,且它们频繁使用的几种关系数量有限。因此,前 10 条最频繁被使用的句法路径反应了二者之间最频繁的几种句法关系,自然可以以较少的句法路径的数量来识别大部分的情感评价单元。剩下的 60 次频繁的句法路径则用来解决评价对象及评价词语二者句法关系不常见的评价单元的识别,进一步的改进系统的性能。当 th_{path} 在 70 到 200 之间变化时,由于一些噪声句法路径的引入,系统性能开始缓慢下降。

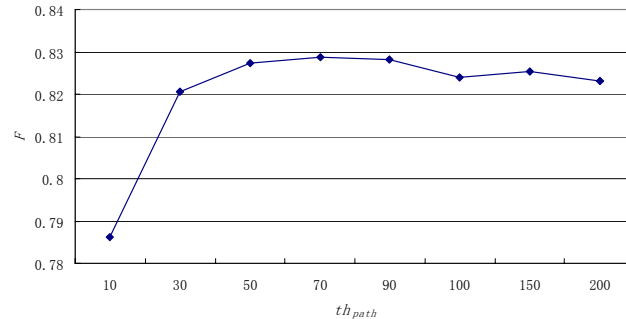


Fig.7 The performance curve changing with th_{path} for appraisal expression recognition ($th_{ed}=1$) on camera domain

图 7 在数码相机领域中随参数 th_{path} 变化而产生的情感评价单元识别系统的性能变化曲线($th_{ed}=1$)

6 结论与展望

本文提出了一种基于句法路径的情感评价单元识别方法。具体的,本方法自动获取句法路径来描述评价对象及其评价词语之间的修饰关系,并使用基于编辑距离的句法路径匹配算法来改进情感评价单元的系统性能。在数码相机和 MP3 播放器两个领域的英文语料上的测试结果表明,本文的方法是有效的。本文的主要贡献可概括如下:首先,本文提出使用句法路径挖掘评价词语和评价单元之间的关系,与以往的工作不同,本文通过统计自动获得大量的句法路径,因此句法路径更全面,更详细;其次,本文尝试使用句法路径匹配来获取情感评价单元,并取得了较好的实验效果;再次,本文采用基于编辑距离的句法路径匹配算法改进了情感评价单元识别的性能,避免了因句法路径有限和句法分析错误而产生的情感评价单元漏识别和错识别问题。

在今后的工作中,我们会尝试通过调整本文编辑距离算法中的三种操作代价,来进一步提高系统的性能。例如,同为名词类的句法成分“NN↑”与“NP↑”的替换代价,理应比“NN↑”与动词类的句法成分“VP↑”的替换代价小。此外,我们也会考虑将本文提出的方法推广到评价对象为动词或短语的情感评价单元抽取中去。

致谢 在此,我们向对本研究工作提供帮助的老师和同学表示感谢。特别地,我们要感谢高雷雷,刘鸿宇等同学在实验数据标注上的工作,感谢赵世奇同学对本文文章初稿的审阅及提出的宝贵意见。

References:

- [1] Riloff E, Wiebe J. Learning extraction patterns for subjective expressions. In: Proceedings of EMNLP-2003. Morristown: ACL, 2003. 105–112.
- [2] Hatzivassiloglou V, McKeown K.R. Predicting the semantic orientation of adjectives. In: Proceedings of EACL-1997. Morristown: ACL, 1997. 174–181.
- [3] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of EMNLP-2002. Hanover: Now Publishers Inc., 2002. 79–86.
- [4] Kim S.M, Hovy E. Automatic detection of opinion bearing words and sentences. In: Proceedings of IJCNLP-2005. New York: Springer, 2005. 61–66.
- [5] Hu MQ, Liu B. Mining and summarizing customer reviews. In: Proceedings of KDD-2004. New York: ACM, 2004. 168–177.

-
- [6] Bloom K, Garg N, Argamon S. Extracting appraisal expressions. In: HLT-NAACL 2007. Association for Computational Linguistics, 2007. 308–315.
- [7] Kim S.M, Hovy E. Determining the sentiment of opinions. In: Proceedings of Coling-2004. Morristown, NJ, USA: Association for Computational Linguistics, 2004. 1367–1373.
- [8] Gamon M. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In: Proceedings of the 20th international conference on Computational Linguistics. Morristown: ACL, 2004. 841–847.
- [9] Zhao J, Liu K, Wang G. Adding redundant features for crfs-based sentence sentiment classification. In: Proceedings of Empirical Methods in Natural Language Processing. Morristown, NJ, USA: Association for Computational Linguistics, 2008. 117–126.
- [10] Wilson T, Wiebe J. Annotating attributions and private states. In: Proceedings of ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky. Morristown, NJ, USA: Association for Computational Linguistics, 2005. 53–60.
- [11] Kobayashi N, Inui K, Matsumoto Y, Tateishi K, Fukushima T. Collecting evaluative expressions for opinion extraction. In: Proceedings of the International Joint Conference on Natural Language Processing. New York: Springer, 2004. 584–589.
- [12] Popescu A.M, Etzioni O. Extracting product features and opinions from reviews. In: hltmnlp2005. Morristown, NJ, USA: Association for Computational Linguistics, 2005. 339–346.
- [13] Yao TF, Nie QY, Li JC. An Opinion Mining System for Chinese Automobile Reviews. In: Cao YQ, Sun MS, eds. Frontiers of Chinese Information Processing. Beijing: Tsinghua University Press, 2006. 260–281 (in Chinese with English abstract).
- [14] Yao TF, Lou DC. Research on semantic orientation analysis for topics in Chinese sentences. Journal of Chinese information processing, 2007, 21(5):73–79 (in Chinese with English abstract).
- [15] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of HLT/EMNLP-2005. Morristown, NJ, USA: Association for Computational Linguistics, 2005. 347–354.
- [16] Casey Whitelaw, Navendu Garg, and Shlomo Argamon. 2005. Using appraisal taxonomies for sentiment analysis. In In Proceedings of CIKM-05. New York, NY, USA: ACM, 2005. 625–631.
- [17] Moilanen K, Pulman S. Sentiment composition. In: Proceedings of the Recent Advances in Natural Language Processing International Conference. 2007. 378–382.
- [18] Choi Y, Cardie C. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In: Proceedings of the Empirical Methods in Natural Language Processing. Morristown: ACL, 2008. 793–801.
- [19] Wagner R.A, Fischer M.J. The string-to-string correction problem. Journal of the ACM, 1974, 21(1):168–173.
- [20] Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 1960, 20(1):37–46.
- [21] Charniak E. A maximum-entropy-inspired parser. In: Proceedings of NAACL. Providence: Brown University, 2000. 132–139.

附中文参考文献:

- [13] 姚天昉, 聂青阳, 李建超等. 一个用于汉语汽车评论的意见挖掘系统. 曹右琦, 孙茂松. 见: 中国中文信息学会成立二十五周年学术年会论文集. 2006, 260–281.
- [14] 姚天昉, 姜德成. 汉语语句主题语义倾向性分析方法的研究. 中文信息学报, 2007, 21(5): 73–79.