

# 基于句子级情感的中文网络评论的情感极性分类<sup>①</sup>

王洪伟<sup>1</sup>, 郑丽娟<sup>1</sup>, 尹 裴<sup>1</sup>, 何绍义<sup>2</sup>

(1. 同济大学经济与管理学院, 上海 200092;  
2. 加州州立大学圣马可斯分校商学院, 圣马可斯, 美国)

**摘要:** 为提高网络评论段落的情感极性分类准确率, 在考虑人们表达习惯和语料粒度的基础上, 提出一种基于句子情感的段落情感极性分类方法。该方法通过句子的情感极性和句子的情感极性贡献度来对段落进行情感分类, 采用传统分类方法预测句子的情感极性, 提出等权重、相关度、情感条件假设 3 种方法, 能够根据训练语料的统计数据动态地确定段落中不同位置句子的情感极性贡献度。最后, 以超过 2 个句子的手机和酒店网络评论为对象进行实验分析, 实验结果显示, 与传统方法相比, 考虑了人们表达习惯的相关度和情感条件假设方法显著提高了段落分类的准确率, 且具有一定的自适应性。

**关键词:** 情感极性; 句子情感; 段落情感; 中文网络评论; 情感极性贡献度

中图分类号: C931.6; H042 文献标识码: A 文章编号: 1007-9807(2013)09-0064-11

## 0 引言

近年来, 越来越多的用户愿意在线分享自己的观点。这些评论反映了用户对产品(服务)的看法。另一方面, 在电子商务环境下, 由于缺少线下体验, 用户更倾向于先看网络评论, 后做购买决策。然而, 由于分析手段薄弱, 面对海量的评论, 商家难以识别用户的情感倾向, 更无法根据用户反馈及时改进产品和调整价格, 从而影响商家的业绩。为此, 情感极性分析(sentiment polarity analysis)技术应运而生, 也就是对网评的文本进行分析, 判断用户的观点是“赞同”还是“反对”, 或识别情感趋向是“高兴”还是“伤悲”。

最初的研究关注面向句子的情感极性分析。然而, 用户的情感表露错综复杂, 常常发表混合观点评论, 既肯定某方面, 同时又在批评其它方面。混合观点评论具有句子多、信息量大、噪音多、情感表达复

杂等特点。如果将面向句子的方法直接应用在面向段落的(即整篇评论)情感极性分析中, 准确率将降低。为此, 一些学者主张按照产品(服务)特征, 将混合观点评论分割成若干个评论片段, 然后采用面向句子的方法对评论片段进行情感极性分析<sup>[1-3]</sup>。诚然, 这种做法可以细粒度地分析用户的情感倾向, 却无法给出用户对该产品(服务)的整体观点。

基于上述原因, 在考虑用户表达习惯和语料粒度的基础上, 本文提出一种基于句子情感的段落情感极性分类方法。该方法具有一定的自适应性, 可以给出用户对产品(服务)的整体观点, 且能显著提高段落的分类准确率。

## 1 相关研究综述

### 1.1 传统的情感分类方法

情感极性分类主要基于两种研究范式: 情感

① 收稿日期: 2011-06-27; 修訂日期: 2013-03-04.

基金项目: 国家自然科学基金资助项目(70971099; 71371144); 中央高校基本科研业务费专项资金资助项目(1200219198); 上海市哲学社会科学规划课题一般资助项目(2013BGL004); 上海市科技发展基金软科学研究博士生学位论文资助项目(12692193000).

作者简介: 王洪伟(1973—), 男, 辽宁大连人, 博士, 副教授, 博士生导师. Email: hwwang@tongji.edu.cn

词汇语义特性和统计自然语言处理.

### 1) 基于情感词汇语义特征的方法

这类方法是利用词汇的感情倾向来判断评论文本的情感极性. 首先计算或判断词汇或词组的褒贬倾向性, 再对段落中极性词汇或词组计数, 或对其褒贬程度值求和或求均值, 或结合句法分析等获得句子或段落的整体情感极性.

Turney<sup>[4]</sup>提出基于情感词组的 SO-PMI 语义分类方法. 该方法提取符合一定模式的形容词或副词双词词组作为情感词组, 并计算抽取出的词组与情感词“excellent”和“poor”的 PMI 值和 SO (semantic opinion orientation) 值, 最后通过计算评论中所有提取出的情感词组的平均 SO 值来区分情感极性. 熊德兰等<sup>[5]</sup>选取褒贬基准词, 根据词

汇相似度判断词汇的倾向性, 并结合句法分析结果和词汇语义倾向性衡量句子褒贬倾向性. 实验结果表明, 该方法的计算结果与人工判别结果比较接近. 闻彬等<sup>[6]</sup>提出基于语义理解的文本情感分类方法. 该方法首先提取符合一定规则的由形容词、名词、动词构成的 5 种情感词组, 并通过语义处理计算抽取出的情感词组的情感值, 最后采用统计的方法, 将所有词组的情感值进行累加从而得到文本的情感值, 并根据该值的大小判定文本的情感倾向性.

### 2) 基于统计自然语言的方法

这类方法的流程如图 1, 即经过预处理、文本表示(特征项选择、特征项降维、特征项权重设置)、分类器处理, 最终得到情感类别.



图 1 情感分类的基本流程

Fig. 1 The process of sentiment classification

①特征项选择. 选取适当的语义单元作为特征项, 它是影响情感分类效果的重要因素. 特征项既要真实地反映文档的情感信息, 也要对不同文档有较强的区分能力, 可以是词、词的组合、 $N$  元组等. 徐军等<sup>[7]</sup>利用朴素贝叶斯和最大熵方法对新闻语料进行情感分类. 实验结果显示, 选择形容词和名词作为特征项时, 具有较高的分类准确率, 且分类性能明显好于只选择形容词. Zhang 等<sup>[8]</sup>选取  $N$ -gram 作为情感特征项, 针对餐饮评论进行研究, 结果显示, Bigram 的分类效果好于 Unigram 和 Trigram. 王洪伟等<sup>[9]</sup>选取词、词的组合、 $N$ -gram 作为情感文本的候选特征项, 以手机在线评论为对象进行实验分析, 结果表明, 中文在线评论的情感分类中, 将形容词、副词和动词一同作为特征项, 分类准确率最高.

②特征项降维. 剔除特征项集中不能有效反映类别信息的特征, 从而缩短训练时间, 提高分类准确率. 常用的特征项降维方法有: 文档频率法 (document frequency, DF)、信息增益法 (information gain, IG)、统计量法 (chi-square statistic, CHI)、互信息法 (mutual information, MI) 等. 基本

思想是计算每个特征项的某种统计度量值, 再设定一个阈值, 把度量值小于该阈值的特征项过滤掉, 留下能对文档进行有效表征的特征. 唐慧丰等<sup>[10]</sup>采用了 MI、IG、CHI 和 DF4 种特征降维方法, 在不同的特征数量和不同规模的训练集情况下进行了实验, 实验结果表明采用 IG 法的情感极性分类效果较好. Yao 等<sup>[11]</sup>对 DF、MI 和 CHI 进行比较, 实验结果显示, DF 方法的分类效果较好, 同时发现 MI 方法不适用于情感特征项降维.

③特征项权重计算. 常用方法有布尔权重、绝对词频 (TF)、倒排文档频度 (IDF)、词频 – 逆文档频率 (TF-IDF)、TFC、ITC、熵权重、TF-IWF 等. Pang 等<sup>[12]</sup>采用布尔权重法进行实验, 情感极性分类准确率达到 82.9%, 优于其他权重设置法. 这是因为语言的褒贬倾向主要取决于正面或负面词语在语言中是否出现, 而不是出现的次数. 通常, 带有情感倾向的特征项出现次数并不重要, 关键在于是否出现, 在哪个类别中出现.

④分类器的选择. 常用的分类器包括: 支持向量机 (support vector machines, SVM)、最大熵 (maximum entropy, ME)、朴素贝叶斯 (Naïve

Bayes, NB) 等. Pang 等<sup>[12]</sup> 人工标记电影评论中常有的特征情感词, 以特征项在文本中出现的频率作为分类特征, 采用 NB、ME 和 SVM 3 种分类器进行对比实验, 结果证明 SVM 分类效果最好. Xia 和 Peng<sup>[13]</sup> 以虚拟社区中的旅馆评论为语料库, 使用 SVM 进行情感分类, 实验结果显示, 随语料库内评论数量的增加, SVM 分类准确性有所提高. Ye 等<sup>[14]</sup> 以旅游博客上的评论作为语料库, 对朴素贝叶斯和 SVM 的分类效果进行比较, 实验结果显示 SVM 优于朴素贝叶斯. 基于上述分析, 本文选择 SVM 作为实验的情感极性分类器.

已有研究显示, 语义方法和统计方法各有其优缺点. 语义方法不需要大量人工标注的训练语料, 但分类准确率却低于统计方法. Ye 等<sup>[15]</sup> 比较 SVM 和语义方法的分类效果, 结果证明 SVM 方法优于基于语义的方法.

### 1.2 基于句子情感的网络评论极性分类

用户常常发表混合观点评论, 这类评论(即段落级)句子多、信息量大、情感倾向多样化. 如果将现有的方法直接应用在面向段落的混合观点评论的情感极性分析中, 准确率将受到影响. 刘勰<sup>[16]</sup> 采用统计的方法, 分别对句子级和段落级两类评论进行实验, 结果显示, 传统方法下, 句子级评论的分类准确率较高, 且远高于段落级的评论.

因此, 有学者尝试依据句子的情感倾向计算段落的情感极性. Mao 和 Lebanon<sup>[17]</sup> 使用连续条件随机场模型对句子进行情感极性分类, 将段落级的评论表示成一系列的句子情感流, 并采用句子情感流对段落级情感极性进行预测. Zhang 等<sup>[18]</sup> 针对中文评论, 提出基于规则的、从句子级转换到段落级的情感极性分类方法. 首先根据情感词和句法结构判断句子的情感极性, 然后根据句子的位置等 5 个特征对句子的重要性进行测量, 最后整合句子的情感极性来预测段落的情感. Sunil<sup>[19]</sup> 采用 BOS(bag of sentence) 的观点对段落进行分类. 首先对句子进行情感极性分类, 然后使用句子的位置等特征对句子打分, 根据分数判断句子对段落的作用.

### 1.3 研究评述

总体而言, 通过句子对更大粒度的段落情感极性进行预测的方法尚处于探索阶段. 一些研究

虽然认识到句子位置会影响句子的重要性, 但只是通过简单的线性函数来量化影响程度. 已有研究中, 对其它可能影响情感分类结果的因素也鲜有考虑, 如人们的表达习惯. 针对以上问题, 本文提出一种基于句子情感的段落情感极性分类方法, 该方法考虑了人们的表达习惯, 根据评论者在段首、段中和段尾的情感流露方式的不同, 通过训练语料的统计数据定义句子的重要程度, 进而通过句子情感预测段落情感.

## 2 基于句子情感的段落情感极性分类原理

### 2.1 段落情感极性值的计算方法

给定网络评论集合  $C = \{c_1, c_2, \dots, c_m\}$ , 每一条评论  $c_i \in C$  都以段落的形式存在. 因此,  $c_i$  可以表示成句子的有限序列, 即  $c_i = < s_{i1}, s_{i2}, \dots, s_{in} >$ , 其中  $s_{ij}$  是一个完整的句子. 不论是段落还是句子, 统称为语料. 本文只考虑段首(first)、段中(middle) 和段尾(end)3 种情况, 即  $s_{i1}$  为段首,  $s_{i2}, \dots, s_{i(n-1)}$  为段中,  $s_{in}$  为段尾. 情感极性通常分为正向和负向两类, 表明评论者肯定或否定的态度,  $C_{pos}$  和  $C_{neg}$  分别表示正类和负类的语料集合.

按照人们的表达习惯, 段落中每个句子对整篇段落情感极性的贡献度, 按其所处的不同位置而各不相同. 如段首的句子常具有总起作用, 段尾的句子常具有总结作用, 所以相比段落中间的句子, 处于段首或段尾的句子对整篇段落情感极性的贡献度更大. 基于上述考虑, 本文构造式(1) 来计算某一段落的情感极性值, 并通过情感极性值来判断段落的情感极性类别.

$$T(c_i) = \sum_{j=1}^n T(s_{ij}) w_{ij} \quad (1)$$

$T(c_i)$  是第  $i$  篇评论  $c_i$  (即段落  $c_i$ ) 的情感极性值.  $s_{ij}$  是  $c_i$  中第  $j$  个句子,  $T(s_{ij})$  是  $s_{ij}$  的情感极性值. 当  $s_{ij} \in C_{pos}$ , 则  $T(s_{ij}) = 1$ ; 当  $s_{ij} \in C_{neg}$ , 则  $T(s_{ij}) = -1$ .  $w_{ij}$  是  $s_{ij}$  对  $c_i$  的情感极性贡献度. 经式(1) 计算, 如果  $T(c_i) > 0$ ,  $c_i$  属于正类; 如果  $T(c_i) < 0$ ,  $c_i$  属于负类.

本文认为, 句子情感极性贡献度是由句子在段落中出现的位置决定的, 也就是处于段首、段中

或段尾的句子具有不同的情感极性贡献度, 分别表示为  $w_F$ 、 $w_M$  和  $w_E$ . 于是, 式(1)可以转变为

$$T(c_i) = \begin{cases} \frac{T(s_{il})(w_F + w_E)}{2}, & \text{若 } n = 1 \\ T(s_{il})w_F + T(s_{i2})w_E, & \text{若 } n = 2 \\ T(s_{il})w_F + \sum_{j=2}^{n-1} T(s_{ij})w_M + \\ T(s_{in})w_E, & \text{若 } n \geq 3 \end{cases} \quad (2)$$

## 2.2 情感极性贡献度的确定方法

根据文本属性和统计原理, 本文提出 3 种确定情感极性贡献度的方法. 通常, 用户的背景和评论的对象都会影响情感表达方式, 从而导致处于不同位置的句子在段落中的情感贡献度难以确定. 而本文提出的相关度方法和情感条件假设方法, 是根据训练语料自动地动态确定情感极性贡献度, 具有一定的自适应性.

### 1) 等权重方法

假设  $w_F = w_M = w_E$ , 段落情感极性值等于段落中每个句子的情感极性值之和. 这种方法操作简单, 不需要统计训练语料各位置的句子和段落情感的相关度. 但由于没有考虑人们的表达习惯, 这种方法会影响段落情感极性分类的准确性.

### 2) 相关度方法

给定网络评论  $c = \langle s_1, s_2, \dots, s_n \rangle$ , 根据句子和段落在情感极性上相同的可能性, 来确定句子  $s_i$  对  $c$  的情感极性贡献度  $w_i$ , 如下式所示

$$w_i = P((c \in C_{\text{pos}} \wedge s_i \in C_{\text{pos}}) \vee (c \in C_{\text{neg}} \wedge s_i \in C_{\text{neg}})) \quad (3)$$

此方法将段落和句子放在一起统计, 通过段落和句子情感极性的相似概率来衡量情感极性贡献度, 考虑了人们的表达习惯, 且较为简单. 但该方法忽视了正类段落和负类段落中句子情感极性贡献度的差异, 例如, 同样位于段首, 句子在正类段落中和负类段落中的情感相关度可能并不一致. 这种现象也是由人们表达习惯所造成的.

### 3) 情感条件假设方法

为了反映相同位置的句子对不同情感极性的段落情感极性贡献度的差异, 本文提出情感条件假设方法. 以句子的极性为已知条件, 该句子对段

落的情感极性贡献度为条件概率, 也就是  $w_F$ 、 $w_M$  和  $w_E$  分别取决于段首、段中和段尾的句子极性. 假设某个情感极性的句子本身只为该情感极性的段落做贡献(即正类句子只为正类段落做贡献, 负类句子只为负类段落做贡献), 因此情感贡献率即为在该句子极性条件下, 段落与句子极性相同的概率. 给定网络评论  $c = \langle s_1, s_2, \dots, s_n \rangle$ , 情感极性贡献度的计算如公式为

$$w_i = \begin{cases} P(c \in C_{\text{pos}} \mid s_i \in C_{\text{pos}}), & \text{若 } s_i \in C_{\text{pos}} \\ P(c \in C_{\text{neg}} \mid s_i \in C_{\text{neg}}), & \text{若 } s_i \in C_{\text{neg}} \end{cases} \quad (4)$$

其中  $i$  为段首、段中或段尾. 该方法将句子极性作为条件, 将条件概率的思想融入了情感极性贡献度计算中, 具有理论上的优良性.

## 3 基于句子情感的段落情感极性分类实验设计

### 3.1 分类的步骤

基于句子情感的段落情感极性分类的流程如图 2 所示分 3 部分: ① 将段落划分为句子, 并按照传统情感极性分类流程, 用分类器预测句子情感极性值; ② 采用统计方法计算段落中不同位置的句子的情感极性贡献度; ③ 由句子情感极性值和情感极性贡献度计算段落的情感极性值, 并根据该值对段落进行极性分类.

### 3.2 语料库

针对服务型和产品型两类评论, 选择超过 2 个句子的评论作为段落级语料, 如表 1 所示. 这样的语料比较符合用户在线评论的习惯, 但随着语料长度增加, 情感表达也变复杂, 有可能一篇段落级评论语料中包含褒贬两种情感. 段落级语料示例.

#### 1) 酒店评论语料库——服务型评论

选取携程网([www.ctrip.com](http://www.ctrip.com)) 的酒店评论作为语料库, 正负语料各 2 000 篇, 按 3:1 的比例划分, 其中 1 500 篇为训练语料, 其余 500 篇为测试语料. 经过剔除重复及破损的语料, 修正分类错误等预处理操作, 酒店评论语料库的训练集包括 1 056 篇正类段落和 1 220 篇负类段落, 测试集包括 330 篇正类段落和 284 篇负类段落, 有效语料总

数为 2 890 篇.

## 2) 手机评论语料库——产品型评论

选取淘宝网 ([www.taobao.com](http://www.taobao.com)) 的手机评论作为语料库, 正负语料各 2 000 篇, 按 3:1 的比例划分, 其中 1 500 篇为训练语料, 其余 500 篇为测

试语料. 经过剔除重复及破损的语料, 修正分类错误等预处理操作, 手机评论语料库获得 1 017 篇正类段落和 1 152 篇负类段落作为训练集, 获得 268 篇正类段落和 332 篇负类段落作为测试集, 有效总语料数为 2 769 篇.

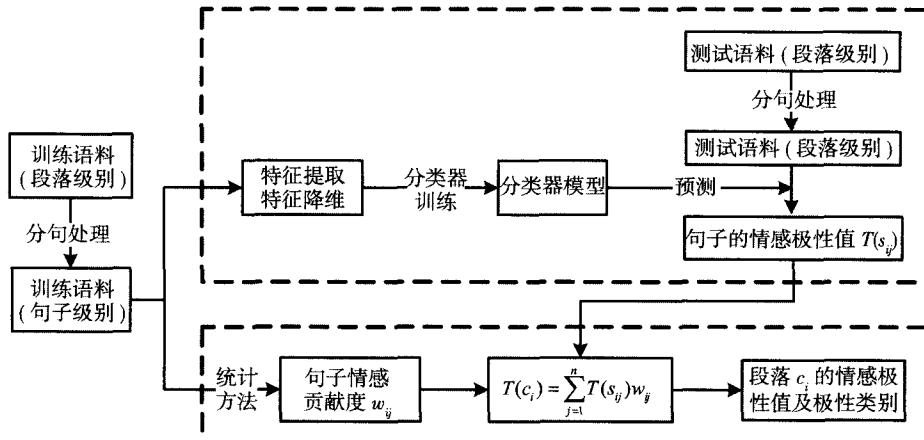


图 2 基于句子情感的段落情感极性分类流程

Fig. 2 The process of sentence-to-paragraph sentiment polarity classification

表 1 段落级语料示例

Table 1 An example of paragraph sentiment classification

评论类型	语料实例	总体情感极性	句子数量	两重情感
酒店评论	过了好久才想起来评价, 记得离火车站超级近, 不过方便的同时必然会觉得比较吵. 韩日旅游团住这里的很多, 前台服务冷淡. 两个人住标准间, 只给一张房卡, 还很挑衅的(地)看我. 气的(得)没心情.	负类	3	否
	酒店的设施是很不错的, 房间大, 设施新, 床和被子都是很新的. 洗手间也是干、湿分离. 服务员态度也是很好, 出入都会打招呼. 就是餐厅大堂中, 适合 2~3 人吃饭的桌子太少了. 还有就是房价在衢州应该属于偏高了.	正类	5	是
手机评论	手机不是正品, 触屏超级不灵, 发信息非常困难. 而且没有导航键, 超级难用, 我无语了.	负类	2	否
	优点太多啦, 首先触屏我很喜欢, 画面的质感特别好, 特别的方便. 然后手机的功能很强大, 单词库里的单词挺多的, 还可以当电子词典用. 当然最有特色的还是粉色的后盖了, 颜色很好, 看着很舒心.	正类	3	否

### 3.3 句子情感极性的确定

首先, 对训练集和测试集的语料进行分句处理, 在此基础上, 为每个段落标注出段首句、段中句和段尾句以方便计算段落的情感极性值.

其次, 按照传统情感极性分类方法对句子进行分类. 通常采用统计机器学习的相关技术确定情感极性, 该部分研究已在文献 [9,11] 中完成.

根据文献 [9,11] 中的研究结论, 为了提高句子级情感分类准确率, 本文选取情感较强烈的名词、形容词、动词、副词作为特征项, 选用 IG 特征抽取方法抽取 300 个特征值来表示文本向量, 并选用分类效果较好的 SVM 进行分类. 正类和负类句子分别以 1 和 -1 来标识. 段落、分句和特征项选择如表 2 所示.

表2 段落分句和特征项选择

Table 2 An example of sentence sentiment classification

段落	分句	句子位置	句子情感极性	名词	形容词	副词	动词
地点不错,从客运站坐TAXI到饭店大约7.5块,柜台小姐先生服务很好!房间更是优,干净整洁舒服,一晚355的价格,算很划算.不过早餐有点贵!	地点不错,从客运站坐TAXI到饭店大约7.5块,柜台小姐先生服务很好!	句首	正向	地点	不错	从	坐
				客运站	很好	大约	到
				TAXI			
				饭店			
				7.5块			
				柜台小姐先生			
	房间更是优,干净整洁舒服,一晚355的价格,算很划算.	句中	正向	服务			
				房间	更是优		算
				一晚	干净		
				价格	整洁		
				舒服			
	不过早餐有点贵!	句尾	负向	早餐	有点贵		

### 3.4 句子情感极性贡献度的确定

利用训练语料的统计数据,计算位于段落不同位置的句子的情感极性贡献度.为了表述清楚,构造了段落情感极性标识符,将各种情感极性贡献度计算公式和训练语料的统计数据联系起来,如表3所示.

表3 段落情感极性标识符

Table 3 Identifier of sentence position and sentiment polarity

句子位置	句子情感极性	正类段落情感极性标识符	负类段落情感极性标识符
段首	正类	FPP	FPN
	负类	FNP	FNN
段中	正类	MPP	MPN
	负类	MNP	MNN
段尾	正类	EPP	EPN
	负类	ENP	ENN

段落情感极性标识符的第1个字母表示句子位置,第2个字母为句子极性,第3个字母表示段落极性.例如,MPP表示正类段落里位于段中的正类句子数量.因为每个段落必有段首句,所以正类的训练段落数为 $L_{FPP} + L_{FNP}$ (或 $L_{EPP} + L_{ENP}$ ),负类的段落语料数为 $L_{FPN} + L_{FNN}$ (或 $L_{EPN} + L_{ENN}$ ).

#### 1) 相关度方法

基于表3标识符和式(3),段首、段中和段尾句子的情感极性贡献度可表示为

$$\begin{aligned} w_F &= P((c \in C_{pos} \wedge s_1 \in C_{pos}) \vee \\ &\quad (c \in C_{neg} \wedge s_1 \in C_{neg})) \\ &= \frac{L_{FPP} + L_{FNN}}{L_{FPP} + L_{FNP} + L_{FPN} + L_{FNN}} \end{aligned} \quad (5)$$

$$\begin{aligned} w_M &= P((c \in C_{pos} \wedge (s_2, \dots, s_{n-1}) \in C_{pos}) \vee \\ &\quad (c \in C_{neg} \wedge (s_2, \dots, s_{n-1}) \in C_{neg})) \\ &= \frac{L_{MPP} + L_{MNN}}{L_{MPP} + L_{MNP} + L_{MPN} + L_{MNN}} \end{aligned} \quad (6)$$

$$\begin{aligned} w_E &= P((c \in C_{pos} \wedge s_n \in C_{pos}) \vee \\ &\quad (c \in C_{neg} \wedge s_n \in C_{neg})) \\ &= \frac{L_{EPP} + L_{ENN}}{L_{EPP} + L_{ENP} + L_{EPN} + L_{ENN}} \end{aligned} \quad (7)$$

#### 2) 情感条件假设方法

情感条件假设方法下的贡献度计算较为复杂,会随着句子极性不同而变化.如果该句子为正类,基于表3标识符和式(4),段首、段中和段尾的情感极性贡献度可表示为

$$w_F^{(pos)} = \frac{L_{FPP}}{L_{FPP} + L_{FPN}} \quad (8)$$

$$w_M^{(pos)} = \frac{L_{MPP}}{L_{MPP} + L_{MPN}} \quad (9)$$

$$w_E^{(pos)} = \frac{L_{EPP}}{L_{EPP} + L_{EPN}} \quad (10)$$

如果该句子为负类,可表示为

$$w_F^{(neg)} = \frac{L_{FNN}}{L_{FNP} + L_{FNN}} \quad (11)$$

$$w_M^{(\text{neg})} = \frac{L_{MNP}}{L_{MNP} + L_{MNN}} \quad (12)$$

$$w_E^{(\text{neg})} = \frac{L_{ENN}}{L_{ENN} + L_{ENP}} \quad (13)$$

最后根据句子情感极性和句子情感极性贡献度,由公式(2)计算出段落的情感极性值,得分大于0为正类,小于0为负类.将段落分类结果与人工分类结果进行对比,计算段落的分类准确率.其中,分类准确率  $P = (A + D)/(A + B + C + D)$ , 正类准确率  $P_p = A/(A + B)$ , 负类准确率  $P_n = D/(C + D)$ .  $A, B, C, D$  的含义如表4所示.

表4 分类准确率

Table 4 Accuracy of classification

	实际为肯定的评论数	实际为否定的评论数
标注为肯定的评论数	$A$	$B$
标注为否定的评论数	$C$	$D$

## 4 实验结果分析

### 4.1 句子情感极性贡献度的计算结果

按照上述算法,对训练语料(酒店评论和手

机评论)的段落情感和相应位置的句子情感的关系进行统计,统计结果如表5所示.

表5 段落和句子相关实验数据

Table 5 Labeling results of sentence-to-paragraph level reviews in training corpus

训练语料	句子位置	句子情感极性	情感极性为正的段落数	情感极性为负的段落数
酒店	段首	正类	952	76
		负类	104	1 144
	段中	正类	1 094	120
		负类	414	1 176
	段尾	正类	790	55
		负类	266	1 165
手机	段首	正类	909	84
		负类	108	1 068
	段中	正类	2 325	258
		负类	810	2 823
	段尾	正类	885	96
		负类	132	1 056

按照相关度方法和情感条件假设方法的公式,计算情感极性贡献度,结果如表6.

表6 各方法下的情感极性贡献度

Table 6 Contribution of sentence-to-paragraph calculated by different algorithms

句子位置	句子情感极性	情感极性贡献度(%) (相关度方法)	情感极性贡献度(%) (情感条件假设方法)
段首	正类	$W_F = 92.1$	$W_F^{(\text{pos})} = 92.6$
	负类		$W_F^{(\text{neg})} = 91.7$
段中	正类	$W_M = 81.0$	$W_M^{(\text{pos})} = 90.1$
	负类		$W_M^{(\text{neg})} = 74.0$
段尾	正类	$W_E = 85.9$	$W_E^{(\text{pos})} = 93.5$
	负类		$W_E^{(\text{neg})} = 81.4$
段首	正类	$W_F = 91.2$	$W_F^{(\text{pos})} = 91.5$
	负类		$W_F^{(\text{neg})} = 90.8$
段中	正类	$W_M = 82.8$	$W_M^{(\text{pos})} = 90.0$
	负类		$W_M^{(\text{neg})} = 77.7$
段尾	正类	$W_E = 89.5$	$W_E^{(\text{pos})} = 90.2$
	负类		$W_E^{(\text{neg})} = 88.9$

从表5可以得到以下结论:

1) 相关度方法的结果表明,不同位置的句子对段落的情感贡献度不同,其中段首句和段落的情感相关度最高,达90%以上,而段尾句次之,段中句最低.这与本文假设段首和段尾会比较重要

的想法相符合,并表明段首相对更为重要.这也一定程度上反映出,人们在表达中往往先在段首直抒情感,定下整篇评论的情感基点,而在结尾又会适时进行总结.但是,这种句尾总结并不是必然的,表现为段尾和段中的情感极性贡献度差距并

不是那么巨大。

2) 情感条件假设方法的结果表明,不管在段落的哪个位置,句子极性为正类条件下的情感极性贡献度都高于句子极性为负类的情感极性贡献度。其中,当句子极性为正类时,段首和段尾的情感极性贡献度还是比段中的稍高,但是三者之间的差异非常小,几乎可以忽略不计;而当句子极性为负类时,三者之间的差异就较为明显:段首的贡

献度明显较高,达到 90% 以上,段尾次之,段中最低,只有 75% 左右。这说明在负类句子下,表达习惯有较大的影响。

#### 4.2 基于句子情感的段落情感极性分类结果

选用 614 篇(正类 330 篇,负类 284 篇)酒店评论和 600 篇(正类 268 篇,负类 332 篇)手机段落评论作为测试集,并采用分类准确率表示分类的效果,统计结果见表 7.

表 7 分类结果统计

Table 7 Sentiment classification of sentence and paragraph

段落编号		段落极性 (人工标注)	SVM 分类器预测			等权重方法		相关度方法		情感条件假设方法	
			段首极性	段中极性	段尾极性	段落情感极性值	段落极性实际	段落情感极性值	段落极性	段落情感极性值	段落极性
酒店	Pos1	1	-1		1	0	?	-0.062	-1	0.018	1
	Neg1	-1	1	-1	-1	-1	-1	-0.748	-1	-0.628	-1
手机	Pos2	1	1	-1, -1	1	0	?	0.151	1	0.263	1
	Neg2	-1	-1	1	-1	-1	-1	-0.979	-1	-0.897	-1

由表 7 看出,不同情感极性贡献度方法下的段落情感极性分类结果存在差异。在等权重方法下,容易出现情感得分为 0 的情况(比如段首 1,段中 -2,段尾 1),造成无法判断段落情感极性,在表中用?进行标注。表 8 显示 3 种情感极性贡献度方法下的分类准确率。

表 8 不同情感极性贡献度计算方法下的情感极性分类准确率

Table 8 Accuracies of paragraph sentiment classification with different algorithms

评论类型	方法	正类准确率 $P_p$ (%)	负类准确率 $P_n$ (%)	准确率 P (%)
酒店评论	等权重方法	81.8	81.6	81.7
	相关度方法	90.4	86.9	88.8
	情感条件假设方法	93.3	86.3	90.1
手机评论	等权重方法	80.8	80.2	80.5
	相关度方法	87.5	85.2	86.2
	情感条件假设方法	86.5	88.6	87.7

由表 8 看出,相关度方法和情感条件假设方法的准确率明显高于等权重方法。这说明考虑表达习惯的情感极性值方法能显著提高分类效果,同时也印证:人们在表达意见的时候,的确存在较为一致的表达习惯。

#### 4.3 实验结果比较

为了验证基于句子情感的段落情感极性分类的效果,通过 4 组实验与基于传统方法的情感极性分类进行对比,实验结果如图 3 所示。

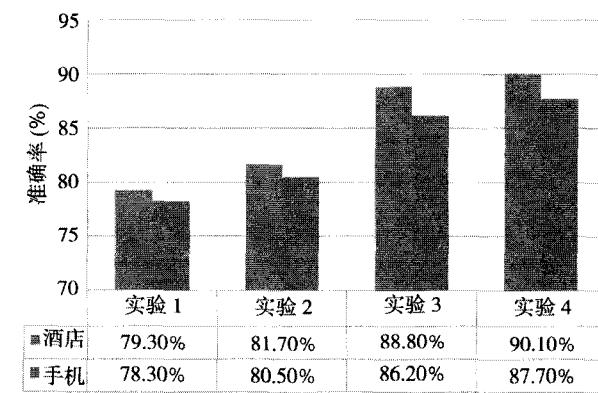


图 3 不同方法下的段落情感分类结果比较

Fig. 3 Comparison of baseline and our approaches

**实验 1** 采用传统情感极性分类方法。用 SVM 分类器判断段落级语料,用 IC 算法抽取特征

300 个进行分类实验;

**实验 2** 采用基于句子极性的段落情感得分方法对段落语料进行分类, 其中情感极性贡献度采用等权重方法;

**实验 3** 采用基于句子极性的段落情感得分方法对段落语料进行分类, 其中情感极性贡献度采用相关度方法;

**实验 4** 采用基于句子极性的段落情感得分方法对段落语料进行分类, 其中情感极性贡献度采用情感条件假设方法.

#### 4.4 实验结论

##### 1) 语料粒度对分类效果的影响

对比实验 1 和实验 2 可知, 单纯将段落分类微粒化, 将句子极性简单相“加”的方法虽然能稍微提高分类准确率, 但提高的幅度不大. 可见, 将语料粒度作为分类依据能提高分类效果, 但并不显著.

##### 2) 表达习惯对分类效果的影响

对比实验 1、实验 3 可知, 考虑评论者在段首、段中和段尾的情感表达习惯能够显著提高段落级语料的情感分类准确率.

对比实验 1、实验 4 可知, 考虑评论者在段首、段中和段尾的情感表达习惯的同时, 考虑评论者正类情感和负类情感的流露方式, 可以进一步提高分类准确率, 分类准确率从 84% 提高到接近 95%, 基本满足现实商务系统的应用要求.

对比实验 2、实验 3、实验 4 可知, 相对于语料粒度, 情感表达习惯在段落评论中起了更重要的作用, 将其作为分类依据能明显提高分类效果.

##### 3) 分类的自适应性

用户的背景和评论的对象都有可能影响情感表达方式, 从而使不同位置句子的重要性产生变化. 实验 3 和实验 4, 采用相关度方法和情感条件假设方法, 根据训练语料自动地动态确定情感极性贡献度, 不仅显著提高了分类效果, 且具有一定的自适应性.

#### 参 考 文 献:

- [1] Miao Q L, Li Q D, Zeng D. Fine-grained opinion mining by integrating multiple review sources [J]. Journal of the American

## 5 结束语

本文在考虑人们表达习惯和区分语料粒度的前提下, 提出了基于句子情感的段落情感极性分类方法. 该方法通过句子的情感极性和对段落的贡献度来对段落进行分类. 实验结果显示, 该方法可以给出用户对该产品(服务)的整体观点, 且显著提高段落分类的准确率.

今后将在以下几个方面进一步探讨.

1) 基于句子情感判断段落情感极性分类中, 假定段落的每个句子非负即正, 而忽略了“中立”句, 甚至客观性句子. 因此将句子分为正类、中立、负类, 分别赋值 1、0、-1, 再计算段落情感极性值, 将是今后考虑的研究内容.

2) 基于句子情感的段落情感极性分类中, 以 0 作为正负类分界点. 尤其当句子采用正类、中立、负类时, 3 个分类间的分界点的确定则更为重要. 理论上应以能使训练语料的分类准确率最高的分界点作为临界, 例如将不同位置的句子极性作为向量特征, 再采用诸如 SVM 的机器算法来训练分类模型, 最后实现自动对段落语料的分类, 这也是今后可以考虑的研究内容.

3) 中文网络评论中存在的一些连接词, 也可能对句子的权重产生影响, 例如: 起总结作用的“总之”、“总而言之”等, 起转折作用的“不过”、“但是”等. 因此, 这些连接词如何影响句子的重要性, 将是今后可以考虑的研究内容.

4) 为了提高情感分析的应用价值, 除了单纯地识别情感极性, 网络评论情感分析还需与其他文本挖掘技术结合, 挖掘出比单独的褒或贬的情感极性更有价值的信息. 其中, 情感极性和情感对象的关系抽取是个应用价值非常广泛的课题. 在段落级的情感极性分类中, 有可能两句话是对同一个情感对象的评论, 因此可以考虑不单纯按照句号对段落进行分句, 而是先提取情感对象, 再按照不同的情感对象对段落进行分句.

- Society for Information Science and Technology, 2010, 61(11) : 2288 – 2299.
- [2] Somprasertsri G, Lalitrojwong P. Mining feature-opinion in online customer reviews for opinion summarization [J]. Journal of Universal Computer Science, 2010, (16) : 938 – 955.
- [3] 李实, 叶强, 李一军, 等. 中文网络客户评论的产品特征挖掘方法研究 [J]. 管理科学学报, 2009, 12(2) : 142 – 152.  
Li Shi, Ye Qiang, Li Yijun, et al. Mining features of products from Chinese customer online reviews [J]. Journal of Management Sciences in China, 2009, 12(2) : 142 – 152. (in Chinese)
- [4] Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews [C]// Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Morristown, NJ, USA: Association for Computational Linguistics, 2002 : 417 – 424.
- [5] 熊德兰, 程菊明, 田胜利. 基于 HowNet 的句子褒贬倾向性研究 [J]. 计算机工程与应用, 2008, (22) : 143 – 145.  
Xiong Delan, Cheng Juming, Tian Shengli. Sentence orientation research based on HowNet [J]. Computer Engineering and Applications, 2008, (22) : 143 – 145. (in Chinese)
- [6] 闻彬, 何婷婷, 罗乐, 等. 基于语义理解的文本情感分类方法研究 [J]. 计算机科学, 2010, (06) : 261 – 264.  
Wen Bin, He Tingting, Luo Le, et al. Text sentiment classification research based on semantic comprehension [J]. Computer Science, 2010, (06) : 261 – 264. (in Chinese)
- [7] 徐军, 丁宇新, 王晓龙. 使用机器学习方法进行新闻的情感自动分类 [J]. 中文信息学报, 2007, 21 : 95 – 100.  
Xu Jun, Ding Yuxin, Wang Xiaolong. Sentiment classification for Chinese news using machine learning methods [J]. Journal of Chinese Information Processing, 2007, 21 : 95 – 100. (in Chinese)
- [8] Zhang Z Q, Ye Q, Zhang Z L, et al. Sentiment classification of internet restaurant reviews written in cantonese [J]. Expert Systems with Applications, 2011, 38(6) : 7674 – 7682.
- [9] 王洪伟, 郑丽娟, 刘仲英, 等. 中文网络评论的情感特征项选择研究 [J]. 信息系统学报, 2012, (10) : 76 – 86.  
Wang Hongwei, Zheng Lijuan, Liu Zhongying, et al. Sentiment feature selection from Chinese online reviews [J]. China Journal of Information Systems, 2012, (10) : 76 – 86. (in Chinese)
- [10] 唐慧丰, 谭松波, 程学旗. 基于监督学习的中文情感分类技术比较研究 [J]. 中文信息学报, 2007, 21(6) : 55 – 94.  
Tang Hufeng, Tan Songbo, Cheng Xueqi. Research on sentiment classification of Chinese reviews based on supervised machine learning techniques [J]. Journal of Chinese Information Processing, 2007, 21(6) : 55 – 94. (in Chinese)
- [11] Yao J N, Wang H W, Yin P. Sentiment feature identification from Chinese online reviews [J]. Communications in Computer and Information Science, 2011, 201 CCIS : 315 – 322.
- [12] Pang B, Lee L, Vaithyanathan S. Sentiment classification using machine learning techniques [C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Philadelphia, US, 2002 : 79 – 86.
- [13] Xia H S, Peng L Y. SVM-based comments classification and mining of virtual community: For case of sentiment classification of hotel reviews [C]// Proceedings of the International Symposium on Intelligent Information Systems and Applications (IISA'09), 2009, 10 : 507 – 511.
- [14] Ye Q, Zhang Z Q, Law R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches [J]. Expert Systems with Applications, 2009, 36(3) : 6527 – 6535.
- [15] Ye Q, Lin B, Li Y J. Sentiment classification for Chinese reviews: A comparison between SVM and semantic approaches. machine learning and cybernetics [C]// Proceedings of 2005 International Conference, 2005, 4 : 2341 – 2346.
- [16] 刘勰. 基于统计自然语言的中文评论情感极性分类研究 [D]. 上海: 同济大学, 2011.  
Liu Xie. Sentiment polarity classification on Chinese reviews based on statistic natural language [D]. shanghai: Tongji University, 2011. (in Chinese)
- [17] Mao Y, Lebanon G. Isotonic conditional random fields and local sentiment flow [C]// Proceedings of the 20th Annual Conference on Neural Information Processing Systems Conference (NIPS), Cambridge, MA: MIT Press, 2006, 961 – 968.
- [18] Zhang C L, Zeng D, Li J X. Sentiment analysis of Chinese documents: From sentence to document level [J]. Journal of the American Society for Information Science and Technology, 2009, 60(12) : 2474 – 2487.
- [19] Sunil K. Sentiment classification using language models and sentence position information [R]. <http://nlp.stanford.edu/courses/cs224n/2010/reports/sukhanal.pdf>

## Classification of sentimental polarity for Chinese online reviews based on sentence level sentiment

WANG Hong-wei<sup>1</sup>, ZHENG Li-juan<sup>1</sup>, YIN Pei<sup>1</sup>, HE Shao-yi<sup>2</sup>

1. School of Economics and Management, Tongji University, Shanghai 200092, China;

2. College of Business Administration, California State University, San Marcos, USA

**Abstract:** With the boost of online reviews, sentiment polarity classification rises in response to the requirement of retrieving consumers' positive or negative opinions on certain products. The primary goal of this research is to improve the accuracy of sentiment polarity classification at the level of paragraphs for Chinese online reviews. With a view to the ways of expression and the grain of corpus, this paper presents a method to predict the sentiment polarity of Chinese online reviews in paragraphs based on sentence level sentiment analysis. Firstly, traditional classification methods are applied to predict the sentiment polarity of sentence. Then, three different algorithms i. e., the equal weight, correlation degree and assumption of sentiment condition, are employed to calculate the contribution that each sentence lying in the different positions of paragraph makes to the sentiment polarity of paragraph. Finally an experiment has been made based on hotel and mobile phone online reviews with lengths beyond two sentences. The result shows that the accuracy of sentiment polarity classification at the level of paragraph is remarkably increased by the method proposed in this paper, by taking correlation degree of expression and assumption of sentiment condition into consideration.

**Key words:** sentiment polarity; sentiment polarity of sentence; sentiment polarity of paragraph; Chinese online review; contribution to sentiment polarity

(上接第 52 页)

[20] Hansen P, Mladenović N, José A, et al. Variable neighborhood search: Methods and applications [J]. Operations Research, 2010, 175 (1): 367 – 407.

## Synthesized model for demand forecasting and its integration with inventory decision

NI Dong-mei, ZHAO Qiu-hong, LI Hai-bin

School of Economics and Management, Beihang University, Beijing 100191, China

**Abstract:** Accurate forecast is helpful for the enterprises to make decisions, including the production plan, the pricing and promotion decisions, etc., so as to reduce the inventory cost and improve the service qualities. In this paper, by analyzing the factors influencing the demand for fast moving consumer goods (FMCG), a synthesized forecasting model, which involves both the time series and the multi-regression methods, is established. The synthesized forecasting model is further integrated with the inventory decision, with the purpose of minimizing the overall logistics costs. To solve the integrated model where multiple parameters are involved, a variable neighborhood search (VNS) based algorithm is developed. To evaluate both the synthesized forecasting model and the integrated model of demand forecasting and inventory decision, computational studies are conducted based on some real data. The computational results show an outperformance of the synthesized forecasting model regarding forecasting accuracy, and an outperformance of the integrated model of demand forecasting and inventory decision when the logistics costs are minimized.

**Key words:** fast moving consumer goods (FMCG); demand forecast; time series analysis; multi regression model; inventory decision