

# 互联网商品评论情感分析研究综述<sup>①</sup>

张紫琼, 叶 强, 李一军

(哈尔滨工业大学管理学院, 哈尔滨 150001)

**摘要:** 互联网评论情感分析是 Web 信息挖掘的一个新兴领域, 近年来受到计算机科学、经济学、管理学等相关学科的广泛关注。目前, 情感分析领域的研究主要集中在主观性内容识别、褒贬情感分类以及在线评论的经济价值挖掘等几个方面, 大部分研究借鉴文本挖掘、信息检索、机器学习、自然语言处理、统计学等方面的技术和方法, 也提出了一些针对评论情感分析的特定方法。对在线评论情感分析领域的研究现状与进展动态进行归纳和分析, 重点论述现有研究采用的主要方法和关键技术, 以及研究中存在的问题, 最后提出了未来的研究方向。

**关键词:** 情感分析; 在线评论; 主观性识别; 褒贬分类; 在线口碑效应

**中图分类号:** TP18 C931 **文献标识码:** A **文章编号:** 1007-9807(2010)06-0084-13

## 0 引 言

情感分析 (sentiment analysis), 又称评论挖掘或意见挖掘 (opinion mining), 是指通过自动分析某种商品评论的文本内容, 发现消费者对该商品的褒贬态度和意见<sup>[1-2]</sup>。利用对互联网上商品评论信息的挖掘与分析结果, 消费者可以了解人们对某种商品的态度倾向分布, 优化购买决策; 生产商和销售商可以了解消费者对其商品和服务的反馈信息, 以及消费者对自己和对竞争对手的评价, 从而改进产品改善服务, 赢得竞争优势。

随着信息技术及其应用在过去十几年中的快速发展, 互联网不仅对企业的业务流程带来了巨大的变革, 也对消费者的行为模式产生了深刻的影响。DoubleClick Inc<sup>[3]</sup>进行了一项针对美国服装业、计算机硬件设备业、运动与健身产品行业及旅游业网络客户的研究, 发现这些行业中都有近一半以上的消费者在做出购买决定前会在互联网上搜索相关商品的介绍及其他消费者对商品的评论信息, 互联网商品评论对于消费者的购买决策有着重要的影响。据统计, 截至 2009 年 9 月, 全球

上网人数已达 17.34 亿人<sup>[4]</sup>, 充分挖掘消费者的喜好偏爱对商家具有重要的意义。

作为 Web 信息挖掘的一个新兴领域, 在线评论情感分析涉及的研究内容十分广泛, 如自动识别互联网中传播的商品评论信息、识别被评论的商品属性、判断客户意见的褒贬态度以及挖掘在线评论与商品销售情况的关系等。姚天昉等<sup>[5]</sup>对情感分析的研究现状进行了总结, 主要内容包括: ①介绍情感分析的定义和研究的目的; ②从主题的认识、意见持有者的认识、情感描述的选择和情感的分析四个方面进行评述, 并介绍了一些成型的系统; ③讨论中文情感分析的研究现状。而本文分别从篇章、语句和词语等不同粒度上, 围绕情感分析技术的基础性工作, 主观性内容识别和褒贬情感分类两方面展开评述, 并对在线评论将如何影响消费者的购买行为以及如何影响商品销售情况的工作进行整理和分析。

针对在线评论挖掘开展的研究越来越多, 然而相关研究仍然处在探索阶段, 在方案设计和技术运用方面都比较分散, 缺乏统一的实验平台和

① 收稿日期: 2009-02-16; 修订日期: 2010-03-18。

基金项目: 国家自然科学基金资助项目 (70971033 70890082); 新世纪优秀人才支持计划资助项目 (NCET-08-0172)。

作者简介: 张紫琼 (1982-), 女, 黑龙江大庆人, 博士生, Email: ziqiong@hit.edu.cn

资源. 由于各种方法预期的目的、使用的数据集和评价方法均不完全相同, 很难进行研究结论之间的比较和评价.

## 1 主观性内容识别

情感分析处理的文本类型是主观性文本, 若要在评论情感分析系统中做到自动处理, 首先需要区别主、客观文本信息, 这是一个十分重要的预处理环节.

一般而言, 主观语言是指人们用来表达自己的立场、态度和感情的语言<sup>[6]</sup>, 主观语言存在褒义与贬义之分, 称为语言的情感倾向或极性. 本节主要介绍主观内容识别的相关工作, 在第二节介绍关于褒贬情感分类的研究. 下面从三个层次介绍主观性文本识别的研究现状: 词语, 句子, 篇章.

在词语水平上, 主观性识别研究的任务是确定主观词语. 现有研究<sup>[7-12]</sup>主要是针对英文主观性形容词、名词、动词和短语的自动识别. 叶强等<sup>[13]</sup>也提出了用于识别汉语主观短语的双词词类搭配. 然而, 现有方法的评价指标大多采用整个词表对主观句的识别正确率和召回率<sup>[7, 10-11, 13]</sup>, 对于主观词语本身的判定仍未提出普遍认可的评价标准. 研究表明, 无论是手工方式, 还是机器自动的方式, 判断词语的主客观性都难于判断词语的褒贬倾向<sup>[14-16]</sup>.

在句子水平上, 主观性识别研究主要涉及自动识别表达情感的主观性语句<sup>[12, 17-20]</sup>, 以及语句中的意见和意见来源等组成成分<sup>[21-22]</sup>, 机器学习方法是相关研究主要采用的方法. Yu等<sup>[17]</sup>发现 Bayes分类器对主观句能够达到较高的识别准确率和召回率, 而对客观句的识别准确率和召回率均较低. 可以看出: 主观语言中包含比较明显的表达情感的词汇, 某些词语的出现可以有效地识别主观句; 客观语言相对缺乏典型的词汇. 客观语言可能由于主观性词语的加入改变其客观性, 反之, 主观语言则不会因为客观词语的加入改变其主观性. 词语位置信息的引入使客观句的识别效果有所提高<sup>[18]</sup>. 文献[12, 19]进一步提出了语句主观性程度的计算方法.

到目前为止, 相关研究在主、客观句的划分标准上仍存在分歧, 使用的语料集存在差异, 因而获

得指标的数值不能完全作为分类效果比较的依据. 关于语句意见和意见源等成分识别的研究<sup>[21-22]</sup>还处在探索中, 准确地判别语言成分及其功能, 需要借鉴相关的语言学知识, 许多问题尚未得到很深入的研究.

在篇章水平上, 由于评论与非评论使用的词汇差别较大, 以词作为项, Naïve Bayes和 SVM能达到相当高的分类精度<sup>[17, 23]</sup>, 所以从篇章层次对主、客观文档进行分类的研究相对较少, 不是研究的重点.

此外, 文献[24]提出了一种自动计算“词义”主观性程度的方法. 文献[25]尝试利用现有英文主观词典及分析工具生成目标语言的主观词典与研究工具, 但是生成的资源在实际应用中的效果还有待验证.

## 2 褒贬情感倾向分类

褒贬情感分类 ( sentiment classification ) 是通过分析在线商品评论的文本内容, 自动将其判断为正面评价或负面评价, 从而挖掘消费者情感倾向分布的过程.

### 2.1 评论篇章情感分类

早期的褒贬情感分类研究始于篇章水平, 目标是挖掘评论文章对某种商品或服务所持的总体褒贬态度. 根据采用的技术, 相关工作可主要分为基于传统文本分类技术的方法、基于褒贬词汇统计的方法和基于回归模型与序列标注的方法.

将文本简化为 BOW ( bag of words ), 在特征处理和统计学习算法的基础上获得对文本内容的表示和类别的预测已经成为传统主题分类的标准模式. 评论情感分类与主题分类的一个较大区别是情感更多地体现为词语之间的语义关系, 许多词语在共现时才能表达一定的情感倾向. 因而, 基于传统文本分类技术的情感分类工作主要使用向量空间模型表示文本和机器学习算法预测类别, 选择什么语义单元作为特征是相关研究的焦点, 主要考虑的问题包括:

1) 权值的计算. 在文本分类和信息检索领域, 通常采用基于词语频率的方法计算项的权重, 而 Pang等<sup>[1]</sup>发现对于评论情感分类, 以词语是否出现作为特征在分类效果上优于词频特征, 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的出现,而不是出现的次数。

2) *n*-gram 项. 一些研究希望通过 *n*-gram 项表示被 BOW 忽略的词语之间的语义联系,但对于 *n*-gram 项的效果还存在争议. Ng 等<sup>[23]</sup>发现将 *bi*-gram 和 *tri*-gram 加入 *uni*-gram 项后能够提高 SVM 的分类性能,但如果分别单独使用 *uni*-gram、*bi*-gram 或 *tri*-gram 作为特征项,分类精度随着阶数的增加反而下降,Ng 等认为原因可能是高阶 *n*-gram 产生的数据稀疏问题导致了分类精度的下降. 而 Cui 等<sup>[26]</sup>得到的结论与文献 [1, 23] 相反,高阶 *n*-gram 项能够提高情感分类精度,并认为原因可能是采用的大规模语料消除了数据稀疏问题,使高阶 *n*-gram 项更有效地表达了语义信息。

3) 词性. 词性是能够标识语义信息的重要语法特征. Hatzivassiloglou 等<sup>[8]</sup>指出一些形容词的出现可以有效地识别主观性语言. 在许多情感识别和抽取工作中,特别是无监督方法,往往抽取文本中的形容词、名词、动词或副词作为潜在的情感表达单元<sup>[27-32]</sup>. Turney<sup>[2]</sup>提出五个包含形容词或副词的词性组合识别语言情感. 在评论篇章情感分类方面,文献 [1, 23] 把词性相关信息加入文本表示的特征项,使用机器学习算法判断评论篇章的褒贬倾向,但分类效果并不十分理想,词性信息的有效利用有待进一步研究。

4) 句法结构特征. 句法分析常常被应用于句子水平的情感分析,以识别句子主题、意见描述项和意见持有者等成分<sup>[31-35]</sup>. 有些研究<sup>[23, 36]</sup>也将句法结构特征作为特征项,对评论篇章进行情感分类. Gamon<sup>[36]</sup>在 *uni*-gram + *bi*-gram + *tri*-gram 项的基础上加入句法结构相关特征后, SVM 的分类精度有所提高. Ng 等<sup>[23]</sup>发现将根据依存关系抽取的特征加入 *bi*-gram + *tri*-gram 项不能明显改善分类效果,但如果仅有 *uni*-gram 项,使用依存关系特征有助于提高分类精度,并认为原因是依存关系特征与 *bi*-gram / *tri*-gram 可以获得相近的信息。

5) 否定结构. 否定的识别和表示对于情感分析问题非常重要. 对于显式的否定结构,相关研究主要采取两种处理方式,一是首先忽略否定词语表示文本,如果否定词语存在,则取相反的情感计算结果<sup>[31, 37]</sup>. 另一种方法是把否定结构编码到文本特征中, Das 等<sup>[38]</sup>提出将符号 “-n” 编码到靠近

否定词语的项,从而形成一个新的项,例如 “favorable-n”。

基于传统文本分类的方法是当前篇章级情感分类的主流方法,虽然评论情感分类的精度通常低于已知的主题分类,但以模式识别为出发点的机器学习算法也能取得较好的效果<sup>[1, 39]</sup>. 评论情感分类的效果不甚理想,是因为主观评论表达的语义信息往往需要复杂的语言知识才能处理. 然而主观性语言大量存在,其统计规律依然可以为特征提取和情感分类提供支持. 对语义信息的挖掘和利用将有助于机器学习技术在相关任务中的应用,如何恰当地表示这些语义信息以及自动地学习它们的统计模式,仍然是一个开放的问题。

基于极性词汇统计的方法一般首先计算词语的褒贬倾向性,通过对篇章中极性词语计数、或对其褒贬程度值求和或均值,获得文章的总体情感倾向<sup>[2, 37, 40]</sup>. 极性词语的选择可以借鉴基于传统文本分类方法中的特征选择. Turney<sup>[2]</sup>提出一种利用搜索引擎估计短语褒贬倾向的方法,进而使用篇章中短语极性的平均值代表整体的情感倾向,但分类精度低于有监督学习方法<sup>[41]</sup>. Dave 等<sup>[40]</sup>使用了信息检索中的多种技术进行特征抽取、特征加权,然后利用特征权重的累加计算产品评论的褒贬倾向,但分类效果没有明显优于传统的机器学习方法。

顺序回归模型适用于评分推测 (rating-inference)<sup>[42]</sup>以及与程度有关的排序问题<sup>[19]</sup>. 序列标注考虑到相邻语句之间<sup>[43]</sup>、语句和文档之间<sup>[44]</sup>的褒贬倾向具有相关性. Mao 等<sup>[43]</sup>将句子的情感倾向分类转化为一个情感流问题,利用条件随机场模型标注句子的倾向性,进而推测文档的情感倾向. McDonald 等<sup>[44]</sup>在一个全局模型中利用 Viterb 算法对文档和句子同时进行情感序列标注. 至今,情感序列标注在理论上的合理性,以及面对实际应用的可扩展性还需要深入验证,尚未得到广泛的应用。

文献 [45] 利用 WordNet 等辅助工具建立了文档的语义图结构,但这种文本表示方法在情感分类上未见明显优势,而且需要比较复杂的语言预处理。

## 2.2 词语的褒贬情感分类

随着互联网情感分析研究的迅速发展,作为

一项基础性工作, 词语的褒贬倾向分类研究也得到了广泛关注和快速发展. 相关工作涉及自动识别主观词语, 判断其褒贬倾向性, 进而建立褒贬词典或为语句的情感分析工作提供支持. 采用的方法主要包括基于语料挖掘的方法<sup>[46-51]</sup>和基于词典等资源扩展的方法<sup>[14, 52-56]</sup> (见表 1).

Hatzivassiloglou 等<sup>[46]</sup>认为文本中的连词可以指示所连接词语之间的语义关系, 通过生成形容词之间同义或反义的连接图, 应用聚类的方法将形容词划分为褒义和贬义两类, 虽然该方法的分类精度达到 90% 以上, 但是只能处理有限的由连词关联的形容词. Wilson 等<sup>[47]</sup>选取了包括褒贬词语、句法结构特征, 以及上下文信息等大量特征, 利用监督学习的方法判断短语在上下文中的主客观性和褒贬倾向. Turney<sup>[48]</sup>提出了一种借助搜索引擎数据库判断词语的极性及其强度的无监督方法, 其假设是词语与标准褒义 (贬义) 词语在数据

库中的共现率越高, 则该词语越倾向于褒义 (贬义). 文献 [49] 认为语言的情感倾向具有连贯性, 提出一个无监督方法, 利用词语之间和分句之间的连接指示情感的变化, 对词语的初始褒贬倾向进行修正. 文献 [50] 利用特定的网页格式和手工建立的语言规则从大量网页中识别抽取褒、贬语句, 进而根据短语在两类语句中出现的概率估计该短语的褒贬倾向.

除了从语料中挖掘词语的情感, WordNet 等词典资源在相关任务中也起到重要作用. WordNet 中词语之间的距离、词语的注解, 以及同义词和反义词集揭示的语义联系, 被用于测量词语的褒贬倾向<sup>[14, 52-54]</sup>, 其思想是利用 WordNet 提供的语义信息对种子词语进行扩展. 这类方法的主要问题在于, WordNet 按照同义词集合 (synonymy set) 组织信息, 而同义词语不一定具有相同的褒贬倾向, 这将导致对词语情感倾向的估计出现偏差.

表 1 词语褒贬情感分类的方法

Table 1 Analysis of methods for word and phrase sentiment classification

文献	对象 <sup>①</sup>	方法 <sup>②</sup>	资源 <sup>③</sup>	测试 比较集 <sup>④</sup>	效果 <sup>⑤</sup>	CR <sup>⑥</sup>	HD <sup>⑦</sup>
Hatzivassiloglou 等 <sup>[46]</sup>	adjective	supervised log-linear regression, clustering	1987 Wall Street Journal corpus	1336 manually labeled adjectives (HM)	90%		✓
Wilson 等 <sup>[47]</sup>	phrase	supervised AdaBoost MH	MQA	HM, General Inquirer lexicon (GI)	65.7%		✓
Turney <sup>[48]</sup>	adjective adverb noun verb	search engine-based Pointwise Mutual Information unsupervised	Web pages indexed by AltaVista	HM, GI	90%	✓	
Kanayana 等 <sup>[49]</sup>	phrase	unsupervised pattern matching context coherency	Web pages	200 manually labeled phrases	—		✓
Kaji 等 <sup>[50]</sup>	adjective phrase	unsupervised pattern matching $\chi^2$ and PMI statistical measure	Web pages	405 manually labeled phrases	85%		✓
Ghose 等 <sup>[51]</sup>	phrase	empirical analysis regression	Amazon reviews	—	—	✓	
Andrevskaia 等 <sup>[14]</sup>	adjective	gloss and lexical relation based word overlap measure	WordNet	GI	—	✓	
Kamps 等 <sup>[52]</sup>	adjective	semantic distance measure	WordNet	GI	68.19%	✓	
Esuli 等 <sup>[53]</sup>	all POS	gloss-based word representation supervised learning	WordNet	HM, 同文献 [48], 同文献 [52]	88.05%		✓
Takanura 等 <sup>[54]</sup>	adjective adverb noun verb auxiliary	gloss and lexical relation based word network spin model	WordNet	GI	91%		✓
Esuli 等 <sup>[55]</sup>	synset	gloss-based synset representation committee of classifiers	WordNet	—	—		✓
Esuli 等 <sup>[56]</sup>	synset	gloss-based synset network PageRank algorithm	WordNet	—	—	✓	

注: 表 1 中的数字角标表示的是: ①研究对象; ②模型方法; ③使用的语料或词典资源; ④主要采用测试或比较集; ⑤报道的分类效果 (精度或 F1 值, 测试条件不同, 结果仅供参考); ⑥ class ranking 方法 (输出词语极性关于类的相对相似性或排序); ⑦ hard decision 方法 (输出词语极性的类别标记). 由于测试集合和测试条件的差异, 指标的数值仅作为方法效果的参考, 不能完全作为方法效果间比较的依据.

早期词语水平的研究是对词语本身褒贬倾向的判断,文献[46-48]将词语判断为具有褒义或贬义倾向,并给出了相应的强度,而文献[27]认为一个词语同时具备褒义和贬义倾向,分别给出了褒义和贬义程度值。E. Suli等<sup>[55-56]</sup>则指出褒贬性是词义的属性,因为词语的不同含义可能具有不同的褒贬态度,提出一种针对 WordNet同义词集 synset褒贬程度的测量方法。

此外,考虑到在线客户评论会影响商家要价,Ghose等<sup>[51]</sup>基于 Amazon.com交易数据,提出了一种新颖的根据商家价格推测客户评论所用词语褒贬性的方法。该方法针对某种商品,根据商家获得的价格溢价,计算客户评论所用词语的“经济价值”,从而判断客户评论词语的褒贬倾向和强度。

### 2.3 语句的情感分析

对句子水平的情感分析,一些研究<sup>[17, 37, 41, 43-44, 57]</sup>讨论了句子总体情感倾向的计算,从而实现了对在线评论信息的搜索和分类。但是在实际应用中,一篇评论经常涉及商品的多个属性,并对这些属性分别进行评价。因而,还有研究将句子作为意见表达和分析单元,主要讨论了商品属性与消费者意见对应关系的识别,以及意见的褒贬倾向分类,下面着重对这方面工作进行介绍。

Kim等<sup>[27]</sup>认为情感(opinion)可以由主题、意见持有者、情感描述项和褒贬倾向性四个部分来描述,即意见持有者针对主题表达了具有某种褒贬倾向的情感描述。语句的情感分析是指在语句文本中自动确定这些元素以及它们之间关系的过程。

语句中主题的形式有两种,一种是显式主题,它可以直接从文本中获得。另一种是隐式主题,需要依靠对当前语句的上下文进行指代关系分析才能获得。现有的主题识别方法主要指显式主题的识别。对于在线商品评论,主题通常是指商品的属性。Yi等<sup>[28]</sup>认为商品的属性包括:①商品名称;②商品的组成部分;③商品的特点和功能;④商品属性的特点和功能。

到目前为止,关于语句水平的情感分析还没有形成十分有效的解决方案。很多研究借助褒贬词典确定句子的情感描述项及褒贬程度。在各成分的对对应关系识别方面,一些研究<sup>[27, 29-30, 59-60]</sup>使用了词性标注、实体识别等文本分析工具,进而根据词语共现性和统计方法判断商品属性与情感描

述项的对应关系,还有一些研究<sup>[31-35]</sup>在句法和语义分析的基础上通过手工定义或机器学习获取一定的语言规则,进行各成分之间对应关系的识别。

根据商品属性是否已经确定,相关工作主要划分为三类:

1)已知商品的属性列表,自动抽取相应的情感描述部分并判断它的褒贬倾向。该技术可应用于面向特定产品领域的情感挖掘。因为商品的属性已经确定,相关研究大多使用自动获取或者手工定义的语言规则判断属性和情感描述项之间的对应关系。

Zhuang等<sup>[33]</sup>首先归纳出电影领域的相关属性和极性词语,然后从训练句子中得到属性和极性词语之间的最短依存路径,作为属性及其情感描述项的依存关系规则,用于挖掘二者之间的对应关系。姚天昉等<sup>[35]</sup>针对汉语汽车评论提出了利用领域本体识别句子的属性、利用主谓结构和定中结构识别属性及其情感描述项的对应关系,以及计算褒贬程度的方法。

2)已知初始商品属性(一般是商品名称或品牌),识别全部被评论的属性及对应的情感描述项。在应用中,如果以产品的名称作为查询条件,那么结合情感分析技术,搜索引擎就能够实现网络情感搜索,并把归纳好的搜索结果提供给消费者。

文献[58]是早期分析比较不同品牌在线口碑的研究。Morinaga等<sup>[58]</sup>预先建立了一个褒贬词典,根据商品名称到褒贬词语的距离确定褒、贬评价语句,然后利用字串的随机复杂度从褒、贬评价语句中抽取描述各品牌的典型词语作为该品牌的口碑。Kim等<sup>[27]</sup>使用实体识别工具提取距离主题最近的人名和机构名称作为意见持有者,并提出一种利用 WordNet计算词语褒、贬情感程度的方法,通过累计意见持有者附近的情感推测句子的褒贬倾向。Nasukawa等<sup>[31]</sup>建立了以谓词为中心的情感传递模式库,在句子属性附近,经过句法分析和模式匹配能够比较精确地描述情感在论元之间的传递关系。Popescu等<sup>[34]</sup>根据名词短语与商品名称的共现性确定商品的属性,在依存分析的基础上,根据手工定义的规则迭代标记元组〈word〉〈word, topic〉和〈word, topic, sentence〉的褒贬倾向。Yi等<sup>[28]</sup>基于主题相关文档和无关文档,采用特征提取的方法从主题相关文档中识别商品的属性。

3) 不限定商品的属性. 针对电子商务网站的消费者反馈或 blog 中的自由评论, 相关技术能够自动识别文本中出现的商品属性及消费者意见.

Lu 等<sup>[29]</sup>、Hu 等<sup>[30, 59]</sup>提出在线评论按形式可分为三类: ①评论分别列出产品的优点和缺点; ②评论列出优点和缺点, 同时进行自由评论; ③无固定格式的自由评论. 针对形式一和形式二中优、缺点比较简短的特点, 文献 [29] 在词性标注的基础上手工对商品属性进行标记, 然后使用关联规则挖掘的方法得到关于属性与词类 (或词语) 的关联规则. 对于形式二和形式三中的自由评论, 文献 [30, 59] 使用频繁项集挖掘的方法从名词和名词短语中挖掘频繁属性, 然后提取频繁属性附近的形容词作为其情感描述项, 再利用这些情感描述项识别不频繁的属性, 而李实等<sup>[60]</sup>基于该方法对汉语评论的产品特征进行了挖掘. Kin 等<sup>[32]</sup>借助极性词语所在的 FrameNetII 语义框架, 采用 ME 机器学习方法从框架元素 (frame element) 中识别句子的主题和意见持有者. 文献 [61] 提出将机器翻译技术用于语句情感单元 (情感倾向, 谓词, 论元) 的抽取.

句子粒度的情感分析研究是在线商品评论挖掘的一个研究重点. 至今在商品属性与消费者意见的对应关系识别方面还没有提出十分有效的解决办法. 方法中有些假设来自研究者的主观直觉<sup>[27]</sup>. 一些研究<sup>[27, 29-30, 59]</sup>仅使用了词性标注等文本分析手段, 这些方法的精度不高; 利用语言规则的方法<sup>[31-35]</sup>可能产生比较高的精度, 但它需要大

量手工工作并且语言现象的覆盖面有限, 领域的可移植性较差. 且仅有为数不多的研究考虑到词语的上下文极性<sup>[34, 61]</sup>和对隐式属性的识别<sup>[29]</sup>.

### 3 互联网商品评论的经济价值挖掘

前两部分介绍的研究均为关于情感分析技术的讨论, 虽然相关技术可以实现和改进对在线评论的自动处理能力, 却无法回答在线商品评论对企业和消费者将产生怎样的影响, 以及如何有效管理和利用这些在线评论等问题.

最近, 这些问题引起了包括自然语言处理和经济管理领域学者的极大兴趣, 相关研究主要从实证分析角度, 对在线评论的褒贬情感将如何影响消费者的购买行为以及如何进一步影响相关产品销售情况的机制进行探索并建立理论模型.

按照获取网络口碑方式的不同, 现有工作可分为两类: 一是利用某些电子商务网站上的客户评分作为该商品的口碑, 例如, Amazon 和 Yahoo! 鼓励消费者在给出文字评论的同时, 用星号数目表示对该商品的评价. 相关研究大多基于实证数据建立回归模型或结构方程模型, 考察客户评分对于商品销售相关指标的影响 (见表 2). 还有研究从情感分析技术出发, 直接从评论文本中挖掘消费者的情感倾向, 进而讨论了消费者情感倾向与商品销售情况的关系, 研究对象不再限于少数电子商务网站提供的客户评分, 互联网上的任何文本信息都可能成为消费者意见的来源 (见表 3).

表 2 在线消费者评分、评论数量和销售情况的关系

Table 2 Relationships between the amount/orientation of customer ratings and product sales

文献	商品类型	正面评论	负面评论	评论数量较多	评论数量较少
Godes 等 <sup>[62]</sup>	电视节目	—	—	无影响	无影响
Chen 等 <sup>[63]</sup>	书籍	无影响	无影响	↑ 销售	↓ 销售
Li 等 <sup>[64]</sup>	电影	无影响	无影响	↑ 销售	↓ 销售
Duan 等 <sup>[65]</sup>	电影	无影响	无影响	↑ 销售	↓ 销售
Duan 等 <sup>[66]</sup>	电影	(通过影响评论数量) ↑ 销售	(通过影响评论数量) ↓ 销售	↑ 销售	↓ 销售
Chevalier 等 <sup>[67]</sup>	书籍	↑ 销售	↓ 销售	↑ 销售	↓ 销售
Dellarocas 等 <sup>[68]</sup>	电影	↑ 销售	↓ 销售	↑ 销售	↓ 销售
Zhang 等 <sup>[69]</sup>	餐馆	↑ 销售	↓ 销售	↑ 销售	↓ 销售
Ye 等 <sup>[70-71]</sup>	酒店	↑ 销售	↓ 销售	—	—
Clemons 等 <sup>[72]</sup>	啤酒	↑ 销售	↓ 销售	无影响	无影响

表3 从在线评论文本中挖掘消费者情感的实证研究

Table 3 A survey of empirical studies that mine customer sentiment from online review text

文献	商品类型	在线评论褒贬分类方法	实证分析方法	主要结论
Das等 <sup>[38]</sup>	股票	分类器组合	linear regression	股票总指数与前一天在线股评所反映的“投资情绪”显著正相关,但对于单个公司,这种关系不成立。
Anweiler等 <sup>[73]</sup>	股票	Naïve Bayes	linear regression	讨论了在线股评数量、股评反映的“投资情绪”与股票交易量、波动率等指标的关系,一些变量之间的显著相关性表明在线股评对股市的影响力。
Mishne等 <sup>[74]</sup>	电影	Window	pearson's correlation	电影首映之前的网络好评数量与票房的成功具有较强的相关性,但这种相关性不足以预测票房。
Liu等 <sup>[75]</sup>	电影	潜在语义分析(LSA)	autoregression	首先建立票房收入的自回归模型,加入在线评论褒贬性变量以后可以提高模型对第二期票房收入的预测作用。
Ghose等 <sup>[76]</sup>	音频/视频播放设备、数码相机	LingPipe	linear regression	在线评论的褒贬性与音频/视频播放设备的销售排名显著正相关,但对数码相机的销售没有显著影响。
Pavlou等 <sup>[77]</sup>	ipod、DVD、CD等十种商品	手工标注	partial least square	借助在线文字评论可以比评分更有效地发现可靠的商家,评论的褒贬性可以显著影响消费者对商家的感知信任和所支付的价格溢价。

关于在线商品评论对企业和消费者影响的研究刚刚起步,并开始活跃起来,所涉及的商品包括书籍、电影、股票、电子产品等。虽然现有工作得到的结论并不完全一致:在线评论对某种商品销售影响的分析模型不一定适用于其他商品,即使针对同种商品,也存在相反的研究结论。但通过对这些实证研究的总体分析,可以发现在线商品评论对商品的销售具有一定影响,而目前的研究还不足以揭示这种影响关系及其作用机理,要深入地研究在线评论与商品销售情况的关系,还需要考虑以下问题:

1)除在线口碑外,商品销售还受商品质量、价格等多种因素的影响,不同的模型、计量分析方法以及数据,所得出的实证结果都会产生差异,因此对于不同情况,重要的是模型的适用性。

2)作为一种电子口碑,在线评论与商品销售是一种相互作用的关系<sup>[62]</sup>,而现有研究主要考虑在线评论对于销售的影响,对两者相互影响的深层机制的研究还略显不足。

3)每种甚至每个商品(如电影)都有各自的目标客户群,不同特征目标客户的购买决策受在线评论影响的程度不同,甚至不受在线评论的影

响,因而体现的商品评论与销售的关系不尽相同。进一步的研究还应该借鉴市场营销学的相关理论和方法。

4)从评论文本中挖掘消费者情感的研究使用不同方法计算评论的褒贬性,准确的文本情感计算是实证分析的前提和保证,应引起研究者的关注。

5)有些研究所用的数据较少,一些结论尚需更大量的数据进行验证。

## 4 存在的问题和未来研究展望

本文综述了目前互联网评论情感分析领域的三个主要研究方向,分别阐述了这三个研究方向的基本内容、常用方法和研究进展。虽然相关研究刚刚起步,但互联网评论情感分析已经成为一个新兴的研究领域。许多相关领域的研究成果都可以借鉴,包括文本挖掘、自然语言处理、Web挖掘、机器学习等方法和技术。然而,由于相关研究工作开展的时间较短,以及语义分析和理解的复杂性,该领域研究总体上还处于探索阶段,存在许多亟待研究和解决的问题。主要包括:

1) 缺乏实验语料和平台. 目前还没有公开发布并具有广泛影响的实验平台, 用于实验的公共语料库较少, 特别是经过加工的基准语料相对缺乏. 现有英文情感分析研究使用较多的是 MPQA 新闻语料<sup>②</sup>和 Pang 等建立的电影评论语料库<sup>③</sup>, 这些语料库所针对的任务和对语料加工的程度都比较有限. 中文公开语料大多处于建设中, 因为缺乏开放的实验平台和统一的评价标准, 方法的有效性很难得到验证.

2) 对各语义成分的识别精度不高. 目前关于情感分析技术的研究相对较为粗糙, 大多借鉴了文本分类、信息检索等领域的研究成果, 直接沿用了相关方法或在其基础上进行修改. 对篇章和语句整体情感倾向的计算, 将文本编码为 BOW, 在特征处理的基础上采用机器学习方法已经取得了较好的效果. 然而, 对语句中各语义成分及其对应关系的识别, 还没有形成有效的解决方案. 具体问题有: 主题和情感描述项识别的精度和鲁棒性不高; 情感的上下文极性研究不足; 隐式主题的识别研究欠缺; 涉及的主题和情感之间关系主要是一对一的简单关系等<sup>[5]</sup>.

3) 规范化方面需要大量复杂而细致的工作. 现有研究在主观性定义、语料标注规范、方法效果评价等很多方面都没有达成共识, 很难对各种方法的实验结果进行客观的比较和评价. 这些问题的解决需要借鉴语言学知识界定语言情感相关的基本概念, 分析语言规律, 准确地判别语言成分及其功能, 进而才能探讨语料标注的规范和方法.

总之, 国际上在评论情感挖掘方面的研究很活跃, 并开发了一些相关的工具, 国内在情感挖掘方面的研究刚刚起步. 针对情感分析领域的研究现状, 以下几个方面是未来研究中值得关注的方向:

1) 情感资源的建设和评价体系的建立. 语料库的建设是相关工作顺利开展的基础, 语料标注是语料库建设的关键技术, 应进一步研究情感标注体系, 考虑如何从包括词汇层(如词性标注)、句法层(如语法分析)和语义层(如属性和情感的对应关系)的多个分析层次体现主观性文本的语

言现象, 形成较细颗粒度的公开语料库. 也应进一步丰富词典资源, 一方面是静态层面的研究, 即词语在静态词典中所表现出的褒义、贬义等情感属性, 这些属性可以直接在语法信息词典中描述; 另一方面是描述词语的情感信息在进入句子框架下发生的偏移现象, 即词语褒贬倾向的动态句法(框架)研究. 统一的评价体系是对方法进行客观评价的前提, 也是不可忽视的基础性工作, 对相关研究的开展具有重要的意义.

2) 语句中各语义成分及其对应关系识别. 情感分析的最终目标是挖掘消费者对商品属性、功能等的态度或评价, 因而在句子粒度上识别各语义成分及其相互关系是在线商品评论挖掘的一个研究重点, 但以目前的自然语言处理现状, 要自动理解文本的情感, 给出较完整的语法和语义分析, 还有相当的困难. 可在现有句法分析研究(如依存分析等)的基础上, 结合情感知识, 获得针对情感分析问题的方法或模型.

3) 系统研究的开展. 目前, 相关工作比较零散, 而情感分析领域的问题之间具有紧密的联系, 系统性地研究非常重要. 例如, 词语的褒贬倾向往往依赖于上下文, 准确地理解这些词语的情感需要以一定的句法分析为基础, 同时, 情感词语也是句子语义成分分析(如属性和情感之间关系识别)的线索. 一方面应该着手建立初始的情感词典资源, 为较细粒度的句子情感分析提供支持, 另一方面也应依据句子水平的分析结果, 对词语在具体上下文中的褒贬性进行判断.

4) 在线商品评论经济价值的研究. 在线评论对消费者购买行为有着巨大的影响, 而相关的理论研究才刚刚起步, 仅有一些学者使用线性回归模型等简单计量经济模型对此进行了初步探索. 由于缺少对在线评论以外影响商品销售因素的控制, 如商品物理属性、市场需求环境和目标客户群特征等, 现有研究获得的结论并不一致, 规律性的发现还不多见. 此外, 对客户评论褒贬性的计算方法以及使用数据的不同, 也会导致研究结论存在一些差异. 因此, 对在线商品评论与销售情况的影响关系及其作用机制的研究尚需进一步深入, 企

② <http://www.cs.pitt.edu/mpqa>

③ <http://www.cs.cornell.edu/people/pabo/movie-review-data/>



业应如何管理和利用在线商品评论的研究仍略显不足,这些都是值得关注的课题。

### 参 考 文 献:

- [1] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques[C] // Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), Morristown, NJ, USA: Association for Computational Linguistics, 2002: 79–86.
- [2] TumeY 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.
- [3] DoubleClick. Search before the Purchase—Understanding Buyer Search Activity as it Builds to Online Purchase[R]. 2005 ([http://www.doubleclick.com/insight/pdfs/searchpurchase\\_0502.pdf](http://www.doubleclick.com/insight/pdfs/searchpurchase_0502.pdf)).
- [4] Minivatts Marketing Group. Internet Usage Statistics—The Internet Big Picture World Internet Users and Population Stats [R]. 2009 (<http://www.internetworldstats.com/stats.htm>).
- [5] 姚天昉, 程希文, 徐飞玉, 等. 文本意见挖掘综述[J]. 中文信息学报, 2008, 22(3): 71–80.  
Yao Tianfang, Cheng Xiwen, Xu Feiyu, et al. A survey of opinion mining for texts[J]. Journal of Chinese Information Processing, 2008, 22(3): 71–80 (in Chinese).
- [6] Lyons J. Semantics[M]. Cambridge: Cambridge University Press, 1977.
- [7] Wiebe J. Learning subjective adjectives from corpora[C] // Proceedings of 17th National Conference on Artificial Intelligence, Menlo Park, California, USA: AAAI Press, 2000: 735–740.
- [8] Hatzivassiloglou V, Wiebe J. Effects of adjective orientation and gradability on sentence subjectivity[C] // Proceedings of the 18th International Conference on Computational Linguistics (COLING), Morristown, NJ, USA: Association for Computational Linguistics, 2000: 299–305.
- [9] Grefenstette G, Qu Y, Evans D A, et al. Validating the coverage of lexical resources for affect analysis and automatically classifying new words along semantic axes[C] // Proceedings of AAAI 2004 Spring Symposium on Expressing Attitude and Affect in Text: Theories and Applications (AAAI), Menlo Park, California, USA: AAAI Press, 2004: 93–107.
- [10] Riloff E, Wiebe J, Wilson T. Learning subjective nouns using extraction pattern bootstrapping[C] // Proceedings of the 7th Conference on Computational Natural Language Learning (CoNLL), Morristown, NJ, USA: Association for Computational Linguistics, 2003: 25–32.
- [11] Riloff E, Wiebe J. Learning extraction patterns for subjective expressions[C] // Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP), Morristown, NJ, USA: Association for Computational Linguistics, 2003: 105–112.
- [12] Kim S-M, Hovy E. Automatic detection of opinion bearing words and sentences[C] // Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), New York: Springer, 2005: 61–66.
- [13] 叶强, 张紫琼, 罗振雄. 面向互联网评论情感分析的中文主观性自动判别方法研究[J]. 信息系统学报, 2007, 1(1): 79–91.  
Ye Qiang, Zhang Ziqiong, Luo Zhenxiong. Automatically measuring subjectivity of Chinese sentences for sentiment analysis to reviews on the internet[J]. China Journal of Information Systems, 2007, 1(1): 79–91 (in Chinese).
- [14] Andreevskaia A, Beigler S. Mining WorldNet for fuzzy sentiment: Sentiment tag extraction from wordnet glosses[C] // Proceedings of 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Trento, Italy: The European Association of Computational Linguistics, 2006: 209–216.
- [15] Esuli A, Sebastiani F. Determining term subjectivity and term orientation for opinion mining[C] // Proceedings of 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Trento, Italy: The European Association of Computational Linguistics, 2006: 193–200.
- [16] Takamura H, Inui T, Okumura M. Latent variable models for semantic orientations of phrases[C] // Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Trento, Italy: The European Association of Computational Linguistics, 2006: 193–200.

- Association of Computational Linguistics 2006: 201–208
- [17] Yu H, Hatzivassiloglou V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences[C] // Proceedings of 2003 Conference on Empirical Methods for Natural Language Processing (EMNLP), Morristown, NJ, USA: Association for Computational Linguistics, 2003: 129–136
- [18] Hiroshima N, Yamada S, Furuse O, et al. Searching for sentences expressing opinions by using declaratively subjective clues[C] // Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text, Morristown, NJ, USA: Association for Computational Linguistics, 2006: 39–46
- [19] Wilson T, Wiebe J, Hwa J. Just how mad are you? Finding strong and weak opinion clauses[C] // Proceedings of the 19th National Conference on Artificial Intelligence (AAAI), Menlo Park, California, USA: AAAI Press, 2004: 761–769
- [20] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summation based on minimum cuts[C] // Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL), Morristown, NJ, USA: Association for Computational Linguistics, 2004: 271–278
- [21] Kobayashi N, Iida R, Inui K, et al. Opinion mining on the web by extracting subject aspect evaluation relations[C] // Proceedings of 2006 AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs (AAAI), Menlo Park, California, USA: AAAI Press, 2006: 86–91
- [22] Bethard S, Yu H, Thomson A, et al. Automatic extraction of opinion propositions and their holders[C] // Proceedings of AAAI 2004 Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications (AAAI), Menlo Park, California, USA: AAAI Press, 2004: 20–27
- [23] Ng V, Dasgupta S, Arifin SM N. Examining the role of linguistic knowledge sources in the automatic identification and classification of review[s] [C] // Proceedings of the COLING/ACL Main Conference Poster Sessions, Morristown, NJ, USA: Association for Computational Linguistics, 2006: 611–618
- [24] Wiebe J, Mihalcea R. Word sense and subjectivity[C] // Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL), Morristown, NJ, USA: Association for Computational Linguistics, 2006: 1065–1072
- [25] Mihalcea R, Banea C, Wiebe J. Learning multilingual subjective language via cross-lingual projections[C] // Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Morristown, NJ, USA: Association for Computational Linguistics, 2007: 976–983
- [26] Cui H, Mittal V, Datar M. Comparative experiments on sentiment classification for online product review[s] [C] // Proceedings of the 21st National Conference on Artificial Intelligence (AAAI), Menlo Park, California, USA: AAAI Press, 2006: 1265–1270
- [27] Kim S-M, Hovy E. Determining the sentiment of opinions[C] // Proceedings of the 20th International Conference on Computational Linguistics (COLING), Morristown, NJ, USA: Association for Computational Linguistics, 2004: 1367–1373
- [28] Yijun Nasukawa T, Bunescu R C, et al. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques[C] // Proceedings of the IEEE International Conference of the IEEE Conference on Data Mining (ICDM), Washington, DC, USA: IEEE Computer Society, 2003: 427–434
- [29] Liu B, Hu M, Cheng J. Opinion observer: Analyzing and comparing opinions on the web[C] // Proceedings of the 14th International World Wide Web Conference (WWW), New York, NY, USA: ACM, 2005: 342–351
- [30] Hu M, Liu B. Mining and summarizing customer review[s] [C] // Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), New York, NY, USA: ACM, 2004: 168–177
- [31] Nasukawa T, Yijun J. Sentiment analysis: Capturing favorability using natural language processing[C] // Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP), New York, NY, USA: ACM, 2003: 70–77
- [32] Kim S-M, Hovy E. Extracting opinions, opinion holders, and topics expressed in online news media text[C] // Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text, Morristown, NJ, USA: Association for Computational Linguistics, 2006: 1–8
- [33] Zhuang L, Jing F, Zhu X-Y, et al. Movie review mining and summarization[C] // Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM), New York, NY, USA: ACM, 2006: 43–50

- [34] Popescu A-M, Etzioni O. Extracting product features and opinions from reviews[C] // Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP), Morristown, NJ, USA: Association for Computational Linguistics, 2005: 339–346.
- [35] 姚天昉, 娄德成. 汉语语句主题语义倾向分析方法的研究[J]. 中文信息学报, 2007, 21(5): 73–79.  
Yao Tianfang, Lou Decheng. Research on semantic orientation analysis for topics in Chinese sentence[J]. Journal of Chinese Information Processing, 2007, 21(5): 73–79 (in Chinese).
- [36] Gamon M. Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis[C] // Proceedings of the 20th International Conference on Computational Linguistics (COLING), Morristown, NJ, USA: Association for Computational Linguistics, 2004: 841–847.
- [37] Ku L W, Liang Y T, Chen H H. Opinion extraction, summarization and tracking in news and blog corpora[C] // AAAI 2006 Symposium on Computational Approaches to Analyzing Weblogs (AAAI), Menlo Park, California, USA: AAAI Press, 2006: 100–107.
- [38] Das S R, Chen M Y. Yahoo! for Amazon: Sentiment extraction from small talk on the web[J]. Management Science, 2007, 53(9): 1375–1388.
- [39] Ye Q, Zhang Z, 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.
- [40] Dave K, Lawrence S, Pennock D. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews[C] // Proceedings of the 12th International World Wide Web Conference (WWW), New York, NY, USA: ACM, 2003: 19–528.
- [41] Chaovalit P, Zhou L. Movie review mining: A comparison between supervised and unsupervised classification approaches[C] // Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS), Washington, DC, USA: IEEE Computer Society, 2005: 1–9.
- [42] Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales[C] // Proceedings of 43rd Annual Meeting of the Association for Computational Linguistics (ACL), Morristown, NJ, USA: Association for Computational Linguistics, 2005: 115–124.
- [43] Mao Y, Lebanon G. Isotonic conditional random fields and local sentiment flow[M] // Advances in Neural Information Processing Systems 19. Cambridge, MA, USA: MIT Press, 2007.
- [44] McDonald R, Hannan K, Neylon T, et al. Structured models for fine-to-coarse sentiment analysis[C] // Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL), Morristown, NJ, USA: Association for Computational Linguistics, 2007: 432–439.
- [45] Devitt A, Ahmad K. Sentiment polarity identification in financial news: A cohesion-based approach[C] // Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL), Morristown, NJ, USA: Association for Computational Linguistics, 2007: 984–991.
- [46] Hatzivassiloglou V, McKeown K. Predicting the semantic orientation of adjectives[C] // Proceedings of the 8th Conference on European Chapter of the Association for Computational Linguistics (ACL/EACL), Morristown, NJ, USA: Association for Computational Linguistics, 1997: 174–181.
- [47] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis[C] // Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP), Morristown, NJ, USA: Association for Computational Linguistics, 2005: 347–354.
- [48] Turney P. Measuring praise and criticism: Inference of semantic orientation from association[J]. ACM Transactions on Information Systems, 2003, 21(4): 315–346.
- [49] Kanayama H, Nasukawa T. Fully automatic lexicon expansion for domain-oriented sentiment analysis[C] // Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP), Morristown, NJ, USA: Association for Computational Linguistics, 2006: 355–363.
- [50] Kaji N, Kitsuregawa M. Building lexicon for sentiment analysis from massive collection of HTML documents[C] // Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language

- Learning (EMNLP/CoNLL), Morristown, NJ USA: Association for Computational Linguistics, 2007: 1075–1083.
- [51] Ghose A, Ipeirotis P G, Sundarajan A. Opinion mining using econometrics: A case study on reputation systems [C] // Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Morristown, NJ USA: Association for Computational Linguistics, 2007: 416–423.
- [52] Kamps J M a x M, M okken R J et al Using WordNet to measure semantic orientation of adjectives [C] // Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC), Paris, France: European Language Resources Association, 2004: 1115–1118.
- [53] Esuli A, Sebastiani F. Determining the semantic orientation of terms through gloss classification [C] // Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM), New York, NY, USA: ACM, 2005: 617–624.
- [54] Takanura H, Inui T, Okumura M. Extracting semantic orientation of words using spin model [C] // Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL), Morristown, NJ USA: Association for Computational Linguistics, 2005: 133–140.
- [55] Esuli A, Sebastiani F. SentWordNet: A publicly available lexical resource for opinion mining [C] // Proceedings of the 5th Annual Conference on Language Resources and Evaluation (LREC), Paris, France: European Language Resources Association, 2006: 417–422.
- [56] Esuli A, Sebastiani F. PageRanking WorldNet synsets: An application to opinion mining [C] // Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Morristown, NJ USA: Association for Computational Linguistics, 2007: 424–431.
- [57] Eguchi K, Lavrenko V. Sentiment retrieval using generative models [C] // Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP), Morristown, NJ USA: Association for Computational Linguistics, 2006: 345–354.
- [58] Morinaga S, Yamanishi K, Tateishi K, et al. Mining product reputations on the web [C] // Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), New York, NY, USA: ACM, 2002: 341–349.
- [59] Hu M, Liu B. Mining opinion features in customer reviews [C] // Proceedings of the 19th National Conference on Artificial Intelligence (AAAI), Menlo Park, California, USA: AAAI Press, 2004: 755–760.
- [60] 李实, 叶强, 李一军, Law Rob. 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, Law Rob. Mining features of products from Chinese customer online reviews [J]. Journal of Management Sciences in China, 2009, 12(2): 142–152. (in Chinese)
- [61] Kanayama H, Nasukawa T, Watanabe H. Deeper sentiment analysis using machine translation technology [C] // Proceedings of the 20th International Conference on Computational Linguistics (COLING), Morristown, NJ, USA: Association for Computational Linguistics, 2004: 494–500.
- [62] Godes D, Mayszlin D. Using online conversations to study word of mouth communication [J]. Marketing Science, 2004, 23(4): 545–560.
- [63] Chen P-Y, Su W, Wu S-Y, Yoon J. The impact of online recommendations and consumer feedback on sales [C] // Proceedings of International Conference on Information Systems (ICIS), 2004: 711–724.
- [64] Liu Y. Word-of-mouth for movies: Its dynamics and impact on box office revenue [J]. Journal of Marketing, 2006, 70(3): 74–89.
- [65] Duan W, Gu B, Whinston A B. Do online reviews matter? — An empirical investigation of panel data [J]. Decision Support Systems, 2008, 45(4): 1007–1016.
- [66] Duan W, Gu B, Whinston A B. The dynamics of online word-of-mouth and product sales — An empirical investigation of the movie industry [J]. Journal of Retailing, 2008, 84(2): 233–242.
- [67] Chevalier J A, Mayszlin D. The effect of word of mouth on sales: Online book reviews [J]. Journal of Marketing Research, 2006, 43(8): 345–354.

- [68] Dellarocas C, Zhang X, Awad N F. Exploring the value of online product ratings in revenue forecasting: The case of motion pictures[J]. *Journal of Interactive Marketing*, 2007, 21(4): 23–45.
- [69] Zhang Z, Ye Q, Law R, et al. The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews[J]. *International Journal of Hospitality Management*, 2010, In Press.
- [70] Ye Q, Law R, Gu B. The impact of online user reviews on hotel room sales[J]. *International Journal of Hospitality Management*, 2009, 28(1): 180–182.
- [71] Ye Q, Gu B, Chen W, et al. Measuring the value of managerial responses to online reviews—A natural experiment of two online travel agencies[C] // *Proceedings of the International Conference on Information Systems (ICIS)*, 2008.
- [72] Clemmons E K, Gao G, Hitt L M. When online reviews meet hyperdifferentiation: A study of the craft beer industry[J]. *Journal of Management Information Systems*, 2006, 23(2): 149–171.
- [73] Antweiler W, Frank M Z. Is all that talk just noise? The information content of internet stock message boards[J]. *Journal of Finance*, 2004, 59(3): 1259–1294.
- [74] Mishne G, Galance N. Predicting movie sales from blogger sentiment[C] // *AAAI 2006 Symposium on Computational Approaches to Analyzing Weblogs (AAAI)*, Menlo Park, California, USA: AAAI Press, 2006: 155–158.
- [75] Liu Y, Huang J, An A, et al. ARSA: A sentiment aware model for predicting sales performance using blogs[C] // *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, New York, NY, USA: ACM, 2007: 607–614.
- [76] Ghose A, Ipeirotis P G. Designing novel review ranking systems: Predicting usefulness and impact of reviews[C] // *Proceedings of the 9th International Conference on Electronic Commerce (ICEC)*, New York, NY, USA: ACM, 2007: 303–310.
- [77] Pavlou P A, Dimoka A. The nature and role of feedback text comments in online marketplace: Implications for trust building, price premiums, and seller differentiation[J]. *Information Systems Research*, 2006, 17(4): 392–414.

## L iterature review on sentiment analysis of online product reviews

ZHANG Zi-qiong, YE Qiang, LI Yi-jun

School of Management, Harbin Institute of Technology, Harbin 150001, China

**Abstract** Online review sentiment analysis is a novel field of Web data mining, which has widely attracted attentions of the researchers in computer science, economics and management disciplines. Studies on sentiment analysis mainly concentrate on subjectivity detection, sentiment classification, and the effects of online user reviews on product sale. Methods and technologies of text mining, information retrieval, machine learning, natural language processing and statistics are used in most researches, and some specific methods of review sentiment analysis are proposed. This paper presents a survey on the up-to-date development in sentiment analysis. The methods used in current research are especially emphasized and the existing problems of those studies are discussed. Finally, some possible future directions of research are pointed out.

**Key words** sentiment analysis; online review; subjectivity detection; sentiment classification; online review effects