

文章编号: 1001—9081(2008)11—2725—04

情感分析研究综述

周立柱, 贺宇凯, 王建勇
(清华大学 计算机科学与技术系, 北京 100084)
(deszlj@tsinghua.edu.cn)

摘 要: 由于 Web 文本迅速增多, 对这些文本, 特别是用户主动发布的评论数据进行挖掘和分析, 识别出其情感趋向及演化规律, 可以更好地理解用户的消费习惯, 分析热点舆情, 给企业、政府等机构提供重要的决策依据。首先对情感分析的研究对象和目标进行了定义和说明, 并给出基本的研究思路。然后, 在主观性句子识别任务上, 详细回顾和分析了主要的处理方法; 在观点分类的特征抽取上, 重点介绍和讨论了两类主流的处理思路——基于情感词和基于频繁模式挖掘。接着简要介绍了其他一些相关的情感分析问题。最后总结了情感分析的现有成就和不足, 以及面临的挑战, 并对其发展前景进行了展望。

关键词: 情感分析; 综述; 观点分类; 主观性识别; 特征抽取
中图分类号: TP311.13 **文献标志码:** A

Survey on research of sentiment analysis

ZHOU Li-zhu HE Yu-kai WANG Jian-yong
(Department of Computer Science and Technology Tsinghua University Beijing 100084, China)

Abstract: With the rapid growth of the Web text data, mining and analyzing these text data, especially the online review data posted by the users, can greatly help better understand the users' consuming habits and public opinions, and plays an important role in decision making for the enterprises and the government. This survey first introduces the motivation, research problems and goals of sentiment analysis, and presents some basic technologies used in sentiment analysis. It then describes one of the major tasks in sentiment analysis, subjective sentence detection, by reviewing and analyzing some recent work in this area. Next, it focuses on another important task in sentiment analysis, opinion classification, and discusses two leading feature extraction techniques for opinion classification, sentimental word based and frequent pattern based methods. Furthermore, it also introduces several other relevant sentiment analysis problems. Finally, the paper summarizes the current status, remaining challenges, and future directions in the field of sentiment analysis.

Key words: sentiment analysis; survey; opinion classification; subjectivity detection; feature extraction

0 引言

Web 已经越来越成为现代社会各种信息的载体。随着 Web 2.0 的兴起与普及, 由普通用户主动发布的文本越来越多, 如新闻、博客文章、产品评论、论坛帖子等。情感分析, 就是对这些信息进行有效的分析和挖掘, 识别出其情感趋向——高兴、伤悲, 或得出其观点是“赞同”还是“反对”, 甚至情感随时间的演化规律。这样就可以更好地理解用户的消费习惯, 分析热点事件的舆情, 为企业、政府等机构提供重要的决策依据。例如, 文献 [1] 就开发了一种把对产品各个部分的用户意见可视化显示出来的系统, 使产品之间比较时, 各部分优劣一目了然, 极大方便了用户的选购。

然而, 目前通常的信息检索技术, 尤其是广为应用的搜索引擎技术, 是以关键词为基础的, 无法对这种需求提供支持, 表现在: 1) 情感或观点很难简单地用关键词来表示和索引, 2) 信息检索领域的排序策略也并不适合观点。

一般来说, 广义的情感分析包括了分析文本中包含的说话人的心理态度, 而狭义的情感分析则主要指的是针对说话人关于某物或某事的观点进行“赞同”或是“反对”的分类。

本世纪初以来, 这方面的研究逐渐多了起来, 不过目前还并没有很完整的综述性文章。文献 [2] 主要从观点分析方面给出了一些例子, 介绍了几个观点分析的系统和技术, 并特别探讨了汉语观点分析的一些问题。本文将偏重技术路线, 从总体上对情感分析的研究进行介绍、总结和展望。

1 情感分析的基本思路及关键技术概述

1.1 情感分析的对象和目标

情感分析最核心和基本的研究任务如表 1 所示。

表 1 情感分析的研究任务

研究对象	研究目标
Web 上的文本, 尤其是用户主动发布的与主题 (或话题) 相关的文本, 大体上可以分为针对文档和针对句子两个不同粒度。	识别出文本中所包含的主观性句子, 并对其情感趋势进行判断。

1.2 主流的研究思路和基本技术

基于文本的情感分析是一个交叉方向的研究, 它涉及自然语言处理、数据库、信息检索、数据挖掘、人工智能等多个领

收稿日期: 2008—07—10。

作者简介: 周立柱 (1947—), 男, 江苏连云港人, 教授, 博士生导师, CCF 高级会员, 主要研究方向: 数据库、Web 数据处理、海量信息系统、数字化图书馆; 贺宇凯 (1983—), 男, 陕西西安人, 博士研究生, 主要研究方向: 数据挖掘、文本情感分析; 王建勇 (1969—), 男, 山东青州人, 副研究员, CCF 高级会员, 主要研究方向: 数据挖掘、Web 数据管理。

域。如图 1 所示,一个基本和典型的情感分析问题的解决,包括下面几个步骤:

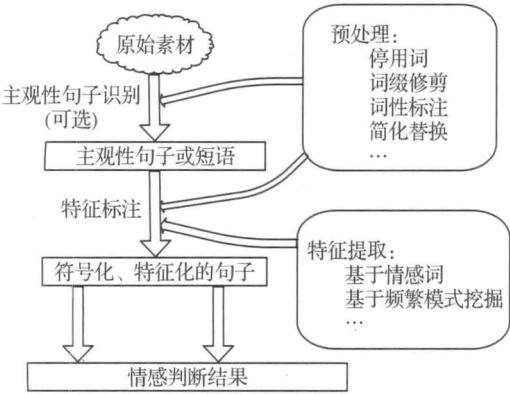


图 1 一个基本和典型的情感分析问题的解决流程

1)收集素材。

目前,大部分的研究所进行情感分析的素材主要来自博客、专业的评论站点、新闻站点和一些电子商务站点。其中各种产品的用户评论,以及作为 Web 2.0 突出代表的博客文章,是众多研究者的首选。

2)预处理。

文献 [3-5]都进行了发现和辨别普通文本中主观性句子的工作,然后还进行了相应的情感分析。而大部分工作则直接利用现成的评论等句子进行分类。但无论怎样,为了减少干扰、提高分类的准确率,预处理都是必不可少的。

根据不同的素材特性和算法需要,预处理也有不少方式,比如停用词、词缀修剪、N 元词、词性标注、必要的简化替换(如把“书”替换成“NOUN”,或把“照相机”替换成“产品名”)等。

文献 [6]详细地列举了种种可能的方法,但结果并不乐观:在大多数情况下,这些有意“美化”的处理都降低了情感分析的准确率。

3)特征抽取。

特征抽取是分类的前提。根据分类的需要,最直接的选择就是各种代表情感的特征词,如“Good”、“不错”、“好”等;特征词可以自动抽取,但有时也需要人工参与构造情感词表,尤其是针对特定产品或特定话题的时候,如文献 [7]。而另一些文献则使用频繁模式挖掘技术来自动挖掘相关的主题和特征,如文献 [8-9]使用关联规则挖掘来寻找特征,文献 [10-12]则使用频繁序列挖掘。

4)分类。

一般都是采用基于机器学习的标准分类器,最常用的是支持向量机 (Support Vector Machine SVM) 和朴素贝叶斯 (Naïve Bayes NB)。当然,也有利用表意较为明显的特征直接进行判断的,这可以看成是基于规则的分类器。

2 主观性句子的识别技术

进行了针对性的预处理以后,识别主观性句子,将随着素材特性、目标类型及算法思路的不同而呈现出不同的技术特色。

2.1 主要工作概述

文献 [13]专门研究了“很好”、“高兴”等各种形容词对于识别主观性句子的重要性。这可以看作识别主观性句子最基

本的思路,但往往还不够。

文献 [4]提到的系统,可以在类似于新闻这样主要讲“事实”的普通文本中,找到并发现主观性句子。它先用一个叫 SimFinder 的工具度量了整个文档中各个句子的相似度,然后依此来构造训练集。随后,用一些情感词,结合了单词、二元组、三元组、词性信息等特征,构建了 NB 分类器。而为了解决训练集构造的不确定性,它还提出了利用多个 NB 分类器来对训练集进行筛选优化,以保证训练集的质量。

而文献 [3]则另辟蹊径,利用以往的一个工作,假定主客观属性相同的句子,位置分布上也会比较接近,从而把候选的句子构成了一幅图,然后利用求图最小割的方法来分类主、客观句子。

此外,文献 [5]提到的 OpinionFinder 系统利用了一些现有的自然语言处理工具来对句子进行记号化、词性标注、词缀修剪等操作,然后结合必要的上下文信息后,直接送到 NB 分类器里面进行分类识别。

2.2 主观性句子识别小结

从上面几个典型工作即可看出:识别主观性句子,主要还是以情感词为主,辅助以各种词汇及文法信息,然后或者送入标准分类器(常见的是 NB),或者直接根据标注的特征进行判断。

有时候进行主观性识别和提取能有效减少干扰,帮助进行观点分类,如文献 [3]。单论准确率的话,以文献 [4]的实验结果,文档粒度的主观性识别可以达到 97% 左右,而句子粒度的识别对于客观性句子能达到 80%~90%,但对主观性句子却只能达到 55% 左右。

3 观点分类技术

观点分类是情感分析最常见的任务。其中关键的步骤就是特征抽取。

以判别观点句子情感极性为目标的特征抽取方法,从技术上可分为利用词汇和语义、借助频繁模式挖掘两大类。

3.1 主要利用词汇和语义特征的方法

具有情感趋向的词汇无疑是最有资格成为分类特征的。文献 [7]能自动从一篇文档中抽取若干针对特定主题的情感表达,再通过一个剖析器,来判断其极性是“赞同”还是“反对”。该系统的准确率很高(约 75%~95%),但是召回率则比较低(20%~25%)。不过,即使有人工参与针对特定主题情感词的维护,这种过分依赖情感词的判断方式,也容易对如双重否定这样的复杂句式(如“也不能说这个相机很差”)进行误判。

而文献 [14]作为一个全自动的产品评论分析器,就不存在人工参与构造情感词的过程。它定义了两个词 w_1 和 w_2 之间的逐点互信息量 (Pointwise Mutual Information PMI),用 $PMI = \log \left[\frac{p(w_1, w_2)}{p(w_1)p(w_2)} \right]$ 来计算,这样,一个情感词 w 的情感趋势 $SO(w)$ 可以用 $PMI(w, \text{“excellent”}) - PMI(w, \text{“poor”})$ 来表示。其中 $p(w_1, w_2)$ 的求法,可以借助 AltaVista 搜索引擎中的 $hits(w_1 \text{ NEAR } w_2)$ 来表示,即 w_1 和 w_2 临近出现的页面数,而 $p(w)$ 则直接用 $hits(w)$ 来表示即可。这样,就自动确定了各个情感词的趋向。最后利用评论中 SO 的平均值来区分情感极性。在其实验结果中,汽车评论的准确率要比电影

评论高出一截,究其原因,很可能是因为电影评论中,很多表示负面情感的“坏词”,比如“恐怖”等,并不一定表示评论者的否定情绪。不过,通过 AltaVista 搜索引擎来计算 SO 值,网络开销将会比较大,难以进行大规模的快速分析。

文献 [9]也不需要构造情感词,它找到潜在的情感词以后,利用已有的一些已知趋向的情感词做种子,结合 WordNet 中的同义、近义信息来自动识别新情感词的趋向,然后类似于文献 [14]和 [7],直接根据句子中各个情感词的总体趋向进行观点的极性判断。

还有,文献 [15]则利用一种叫作“Relaxation labeling”的方法,对潜在观点词极性判断的准确率提高较多。而文献 [16]对情感文本采用多种标准和多个信息源进行特征化,并分别建立 SVM 分类器,再把这些分类器集成起来,效果优于只采用单一的信息源。

总的来说,1)除非是针对特定对象,大部分工作都趋于让机器自动去识别、发现并判断情感词;2)这种以情感词为主的特征抽取因为比较精确,指示性好,往往就不再需要如 SVM 这样的分类器了。

3.2 借助频繁模式挖掘的方法

借助频繁模式挖掘的方法出现较晚,不过由于其优良特性,也渐渐受到了更多的关注。

文献 [8-9]都是基于这样的假设:在大量的产品评论中,频繁出现的,必定是和产品特征紧密相关的语句。它们先利用关联规则挖掘技术找到产品的部件或特征名,如相机的尺寸、镜头等,然后在这些词的周围找到相关的情感词,再识别其情感趋向,达到对这些产品评论进行有效分析的目的。这样做的优点很明显,就是对产品部件或特征,以及针对这些特征的描述提取得比较精确,利于进一步分类。

文献 [11]和 [12]则在一定预处理的前提下,利用频繁序列挖掘找出频繁的单词子序列,进行必要的特征选取,然后送入标准的 SVM 或 NB 进行分类的。这种情况下,挖掘出来的频繁子序列作为特征,一个很大的优势就是可以看成规则,可解释性好,而且也很大程度掩盖了文献 [7]中无法对双重否定之类句子结构进行处理的缺点。

3.3 几种典型分类器效果简评

前面提到,在一些直观、简单的特征已经完善提取之后,复杂的分类器未必很重要,不过,在另一些思路的工作中,基于机器学习的分类器也是不可或缺的工具。

文献 [17]详细对比了 SVM、最大熵分类器 (Maximum Entropy, ME)和 NB 这三种分类器,并把它们和基本的手工分类进行了比较。

结果显示,基于机器学习的分类器要比手工分类效果好得多。其中平均表现最好的是 SVM,其次是 ME,而 NB 最差,但实际上,三者的差异并不非常明显。不过,对于情感分类的任务来说,不管用什么分类器,都很难达到其对于文本类别分类的准确率。

4 其他一些类型的情感分析问题

除了简单地发现并识别主观性句子并判断其情感极性以外,情感分析还有一些别的侧面,简述如下。

1)进一步的挖掘任务:

①主题或意见发表者和情感一起分析,是情感分析中一项重要的扩展。即同时要抽取情感所描述的对象或其发出人,得出“某物→好”或“某人说:某物坏”这样的结论,如文献 [8-9, 18-21]。

②对比型句子的情感识别。文献 [10]提出了识别“X 比 Y 好”这样的句子,并进行了不错的尝试和实践。

③情感的强度分析。这是比“赞同还是反对”这种二元分类更高要求的任务。文献 [22]不仅仅满足于“正”和“负”的判断,而是按情感强度由最差到最好评定其星级(如一星到五星)。

2)在应用上,除了最基本的得到情感分类结果外,还有下面一些工作:

①情感检索。文献 [23]的工作中,用户可以按情感词、主体、和希望的态度进行查询。

②预测。如根据电影宣传时期的用户评论进行票房预测 [24],以及根据用户评价进行图书的销售曲线预测 [25]等。

③情感的演化。文献 [26]提到的 MoodViews 就是一个利用一段时间内的博客文章来分析作者情绪变化的工具。

3)在文献 [27-28]中,首次提到了用户评论的垃圾和反垃圾问题,并初步给了一些解决方案。

5 结语

相比传统的文本分类,情感分类有先天的困难和挑战,主要表现在:1)自然文本中表达方式的多样化:比喻、附和、讽刺、正话反说等。2)句式的复杂性:比较型的句子、各种不同的习惯用语、句式的不同搭配等。3)训练数据的稀疏性和不均衡性。这些都造成了:1)目前还较为狭窄的应用领域。2)不能同时获得较高的准确率和召回率。

即使如此,在巨大商业利益的驱动和广大研究先行者的不懈努力下,情感分析、尤其是观点分类方面,已经取得了不小的成就:1)技术上,在众多领域的技术基础上,已经产生了多种有效的特征抽取和分类策略。2)效果上,在素材来源可信、分析对象确定的情况下,具备较高准确率的众多产品和系统已经问世。

从技术发展和应用需求的角度来看,当今和未来一段时间内的情感分析,很可能将有下列几个热点:1)更好的特征抽取技术和工具,无疑是情感分析中不变的追求;2)和社区挖掘结合,对情感的演化进行建模,以实现需要的分析和监控;3)与其他文本挖掘任务相结合,比如和观点摘要相结合,可以对用户观点形成比“赞同或反对”更有价值的总结报告;4)处理各种来源的信息,而不局限于博客、产品论坛、专业的评论网站。

参考文献:

[1] LIU B, HU M, CHENG J. Opinion observer: Analyzing and comparing opinions on the Web[C]// Proceedings of the 14th International Conference on World Wide Web: WWW 2005. New York: ACM Press, 2005: 342-351.

[2] 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.

[3] PANG B, LEE L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts[C]// Proceedings of the 42nd Annual Meeting of the Association for Computa-

- tional Linguistics. Morristown, NJ, USA: Association for Computational Linguistics, 2004: 271–278.
- [4] YU H, HATZIVASSILOGLU V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences[C]// Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. Morristown, NJ, USA: Association for Computational Linguistics, 2003: 129–136.
 - [5] WILSON T, HOFFMANN P, SOMASUNDARAN S, et al. OpinionFinder: A system for subjectivity analysis[C]// Proceedings of the 2005 Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. Morristown, NJ, USA: Association for Computational Linguistics, 2005: 34–35.
 - [6] DAVE K, LAWRENCE S, DPENNOCK M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews[C]// Proceedings of the 12th International Conference on World Wide Web. New York: ACM Press, 2003: 519–528.
 - [7] NASUKAWA T, YI J. Sentiment analysis: Capturing favorability using natural language processing[C]// Proceedings of the 2nd International Conference on Knowledge Capture. New York: ACM Press, 2003: 70–77.
 - [8] HU M, LIU B. Mining opinion features in customer reviews[C]// Proceedings of the 19th National Conference on Artificial Intelligence. AAAI 2004. Menlo Park, California: AAAI Press, 2004: 755–760.
 - [9] HU M, LIU B. Mining and summarizing customer reviews[C]// Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining. New York: ACM Press, 2004: 168–177.
 - [10] JINDAL N, LIU B. Identifying comparative sentences in text documents[C]// Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2006: 244–251.
 - [11] MATSUMOTO S, TAKAMURA H, OKUMURA M. Sentiment classification using word sub-sequences and dependency sub-trees[C]// Proceedings of the 9th Pacific Asia Conference on Knowledge Discovery and Data Mining. LNCS 3518. Berlin: Springer-Verlag, 2005: 301–311.
 - [12] GAO C, WANG J, HE Y, et al. Efficient mining of frequent sequence generators[C]// Proceeding of the 17th International Conference on World Wide Web. New York: ACM Press, 2008: 1051–1052.
 - [13] HATZIVASSILOGLU V, WIEBE J M. Effects of adjective orientation and gradability on sentence subjectivity[C]// Proceedings of the 18th Conference on Computational Linguistics. Morristown, NJ, USA: Association for Computational Linguistics, 2000: 299–305.
 - [14] 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. Morristown, NJ, USA: Association for Computational Linguistics, 2002: 417–424.
 - [15] POPESCU A M, ETZIONI O. Extracting product features and opinions from reviews[C]// Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing. Morristown, NJ, USA: Association for Computational Linguistics, 2005: 339–346.
 - [16] MULLEN T, COLLIER N. Sentiment analysis using support vector machines with diverse information sources[C]// Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. Morristown, NJ, USA: Association for Computational Linguistics, 2004: 412–418.
 - [17] PANG B, LEE L, VAITHYANATHAN S. Thumbs up?: Sentiment classification using machine learning techniques[C]// Proceedings of the ACL'02 Conference on Empirical Methods in Natural Language Processing. Morristown, NJ, USA: Association for Computational Linguistics, 2002: 79–86.
 - [18] YI J, NASUKAWA T, BUNESCU R, et al. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques[C]// Proceedings of the 3rd IEEE International Conference on Data Mining. Washington, DC: IEEE Computer Society, 2003: 427–434.
 - [19] MEI Q, LING X, WONDRA M, et al. Topic sentiment mixture: Modeling facets and opinions in Weblogs[C]// Proceedings of the 16th International Conference on World Wide Web. New York: ACM Press, 2007: 171–180.
 - [20] KM S M, HOVY E. Determining the sentiment of opinions[C]// Proceedings of the 20th International Conference on Computational Linguistics. Morristown, NJ, USA: Association for Computational Linguistics, 2004: 1367–1373.
 - [21] GAMON M, AUE A, CORSTON S, et al. Pulse: Mining customer opinions from free text[C]// Proceedings of the 2005 International Symposium on Intelligent Data Analysis. LNCS 3646. Berlin: Springer-Verlag, 2005: 121–132.
 - [22] PANG B, LEE L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales[C]// Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics. Morristown, NJ, USA: Association for Computational Linguistics, 2005: 115–124.
 - [23] EGUCHI K, LAVRENKO V. Sentiment retrieval using generative models[C]// Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Morristown, NJ, USA: Association for Computational Linguistics, 2006: 345–354.
 - [24] MISHNE G, GLANCE N. Predicting movie sales from blogger sentiment[C]// Proceedings of the 21st National Conference on Artificial Intelligence. Menlo Park, California: AAAI Press, 2006: 155–158.
 - [25] GRUHL D, GUHA R, KUMAR R, et al. The predictive power of online chatter[C]// Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2005: 78–87.
 - [26] MISHNE G, RIJKE DE M. MoodViews: Tools for blog mood analysis[C]// Proceedings of the 21st National Conference on Artificial Intelligence. Menlo Park, California: AAAI Press, 2006: 153–154.
 - [27] JINDAL N, LIU B. Review spam detection[C]// Proceedings of the 16th International Conference on World Wide Web. New York: ACM Press, 2007: 1189–1190.
 - [28] JINDAL N, LIU B. Opinion spam and analysis[C]// Proceedings of the 2008 International Conference on Web Search and Web Data Mining. New York: ACM Press, 2008: 219–230.