网络意见挖掘、摘要与检索研究综述

侯 锋[1] 王传廷[2] 李国辉[1]
(国防科技大学信息系统与管理学院 长沙 410073)[1] (武汉理工大学理学院 武汉 430070)[2]

摘 要 网络上带有人的主观感情色彩的评论性文本反映了人们的意见、态度和立场,因而具有很大的利用价值。信息挖掘技术针对这些主观文本进行处理,获得有用的意见、结论和知识。首先介绍了意见挖掘出现的背景和应用意义,然后从词汇情感极性识别、粗粒度的情感分类、细粒度的意见挖掘与摘要、意见检索和相关语言资源与系统5个方面综述了研究历程和现状,最后总结了研究难点与研究趋势。

关键词 情感极性,情感分类,意见挖掘,意见检索
中图法分类号 TP391 **文献标识码** A

Survey on the Opinion Mining, Summarization and Retrieval

HOU Feng[1] WANG Chuan-ting[2] LI Guo-hui[1]
(School of Information System and Management, National University of Defense Technology, Changsha 410073, China)[1]
(School of Science, Wuhan University of Technology, Wuhan 430070, China)[2]

Abstract The review texts with subjective sentiments on the Web are valuable for many applications, since they expressed the opinions, attitudes and standpoints of users. The opinion mining technique process these subjective texts specially to generate useful opinion summarization and knowledges automatically. The background and application of opinion mining was introduced firstly; and then the sate of the art of opinion mining was presented in five subtasks: sentiment polarity identification of words, coarse sentiment classification, fine grade opinion mining and summarization, opinion retrieval and language resources and application systems; The difficulty and trend of opinion mining was concluded finally.

Keywords Sentiment polarity, Sentiment classification, Opinion mining, Opinion retrieval

人类自然语言承载的信息可以分为两类:一类是客观事实信息,一类是带有人的主观感情色彩的评论性信息。随着互联网上论坛、社区、博客、评测网站等平台的发展,这类评论性信息也越来越多。评论性信息反映了人们对于特定产品、政策法规、人物和事件的态度、立场和意见等,因而比客观事实信息具有更大的价值。

另一方面,自20世纪70年代开始,计算语言学界就开始研究计算机对自然语言的自动语义分析,但取得的进展非常有限。2002年以后,使用统计学习的方法进行语义角色标注(semantic role labeling)[1]成为浅层语义分析的主要实现方式,但对自动语义理解的贡献还很有限。自动语义分析目前仍然是很难的问题,而识别主观性信息的语义表达及褒贬倾向则相对简单,加上这类信息潜在的商业价值,引起了广泛研究,并逐渐形成了一个新的研究方向——意见挖掘(Opinion Mining)或情感分析(Sentiment Analysis)。

2004年,Kim[2]提出将意见表述分为4个语义成分:主题(topic)、表达者(holder)、陈述(claim)和情感(sentiment)。另外,在产品评论中还应包括产品的属性(Feature)。属性指产品的某个部件或功能,例如数码相机的镜头、屏幕,或电影的剧情、配乐、摄影等。

例1 "我买的那本书虽然包装很精美,但是内容却实在让人不敢恭维"。

在这个意见陈述句子中,"我"是意见的表达者,"那本书"是意见表达的主题,而"精美"和"实在让人不敢恭维"是对意见的陈述,其情感态度分别是褒义和贬义。"包装"和"内容"分别是产品的两个属性。

意见挖掘的过程就是要在网络文本信息中自动识别这5种语义成分并关联这些语义成分之间的关系,从而提取有意义的语义信息。意见挖掘涉及的关键技术有:

(1)主题抽取(Topic Extraction):识别意见所针对的对象;

(2)意见表达者识别(Holder Identification):识别表达意见的人或组织;

(3)产品属性提取(Feature Extraction):确定意见所针对的产品属性;

(4)情感词的提取与褒贬倾向分析(Sentiment Analysis):

到稿日期:2008-09-18 返修日期:2008-11-30 本文受国家自然科学基金项目(60273066),国防科技大学校预研项目(JC06-05-01)资助。
侯 锋(1980—),男,博士研究生,主要研究方向为自然语言处理、文本挖掘、信息检索,E-mail:pundit80@yahoo.com.cn;王传廷(1985—),男,硕士研究生,主要研究方向为自然语言处理、中文分词;李国辉(1963—),男,教授,博士生导师,主要研究方向为多媒体信息系统、多媒体信息检索、数据挖掘。

提取陈述，并判断其情感极性。

意见挖掘技术可以应用于生活中的很多方面，如商业情报获取、电子商务、民意调查、报刊编辑等。在电子购物网站、论坛、博客上，有人们对产品、电影、书籍、时事的评论信息，使用意见挖掘技术对这些信息进行处理后，厂商可以据此了解顾客的反馈意见，潜在的购买者也可以事先了解产品，政府部门可以将其作为决策参考，报社也可以从中选择新闻素材。

本文第1节介绍词汇的情感极性识别；第2节分别从文本和句子的情感分类介绍粗粒度的意见挖掘；第3节分别从3个方面介绍细粒度的意见挖掘；第4节介绍意见检索的相关研究；第5节介绍意见挖掘相关的语言资源与应用系统；最后是总结。

1 词汇的情感极性识别

词汇的情感极性识别或褒贬倾向识别，就是识别词汇是褒义、贬义还是中性词。早期是以经典集合论的观点分配某个词的情感极性，即认为词汇只具有一种情感极性。由于有些词在不同的语境下具有不同的褒贬倾向，目前比较合理的做法是以模糊集合的观点分别给出某个词汇属于3种情感极性的概率[11]。词汇的情感极性识别研究是意见挖掘方法和技术研究的基础，可以分为基于语料库的方法[3-6]和基于词典的方法[7-12]。

基于语料库的方法主要利用词汇之间的连词和共现模式。Hatzivassiloglou[3]等人利用词汇之间的连词（and，but，either-or和neither-nor等）生成词汇间的同义或反义关系的连接图，比如'and'连接的两个词应该有相同的褒贬倾向，而'but'连接的两个词有相反的褒贬倾向；然后根据连接图将词汇分为褒义和贬义两类。对于由常用连词连接的词，其识别准确率高于90%。Wiebe等人[4]的方法首先人工标记一部分种子形容词的褒贬倾向，然后从语料库中统计其他形容词与种子词的互信息，从而判断新词的情感极性。Turney[5]的方法类似，除了点互信息（PMI）外，还采用了潜在语义分析（Latent Semantic Analysis），研究对象包括形容词、副词、名词和动词。香港城市大学Yuen等人[6]利用Turney的PMI，以小规模的语料库识别汉语词汇的情感极性；一种算法是像Turney那样以汉语已知极性的词汇为种子，另一种算法是以汉语已知极性的单个汉字（如"幸"、"贪"等）为种子，计算词汇与这些汉字的PMI。实验结果表明，前一种算法的准确率随种子词的增加而提高，但查全率很低，最高才59.57%；以单个汉字为种子时，10个种子汉字即可达到79.89%的准确率，查全率也提升到73.26%。

基于词典的方法类似于通过词典计算词汇的语义相似度，弊端是无法发现那些与上下文相关的情感词。利用WordNet识别英语词汇的情感极性，可以像基于WordNet的语义相似度计算那样，计算词汇与种子词（如Good，Bad）的语义相似度[7]；也可以利用WordNet所具有的层次结构和词汇之间的同义词关系，扩充已知情感极性的种子词集合[2,8]，但是某些词与它们的同义词不一定具有相同的情感极性[9]。Esuli等人[10]也是利用词典同义词扩大种子词集合，然后从词典中提取所有词条的解释信息作为文本分类对象，利用扩大后的种子词集合和解释信息训练文本分类器，再由该分类器根据解释信息预测未知词的极性。这种方法提高了准确率，且可以计算所有词汇的情感极性。Andreevskaia[11]则利用WordNet词条的解释信息确定某词汇与所属情感词集内其他词语义关联的强弱，从而确定其隶属该情感词集的概率大小，也就是将3个情感词集看作模糊集。朱嫣岚等人[12]提出了基于中文HowNet的两种词汇情感极性识别方法。实验表明，基于HowNet语义相似度的方法比基于相关场的方法准确率更高，词频加权后的判别准确率可达80%以上。

2 粗粒度的意见分类研究

2.1 篇章级的情感分类

篇章级的情感分类，可以看作文本分类的特殊形式。情感分类将整个文本分为褒义、贬义和中性等几种类型，对意见句子并不做深入分析。与基于主题的传统文本分类不同之处在于，基于主题的文本分类将主题关键词作为特征；而文本的情感分类则把情感词（比如"漂亮"、"丑陋"等）作为关键特征。Turney[13]的非监督法分3个步骤：1）词性标注后，提取符合某些词性序列模式（比如，形容词＋名词）的两个连续词，作为特征短语；2）分别计算特征短语与excellent和poor的PMI，前者减后者作为短语的语义倾向（Semantic Orientation，SO）值；3）计算所有特征短语的平均SO，作为分类依据。该算法对汽车类文本分类准确率最高达84%，电影类最低为65.83%。Pang[14]则首先人工标记若干电影评论中常有的特征情感词，然后用特征词在文本中出现的频率作为分类特征，采用朴素贝叶斯、最大熵和支持向量机3种分类器进行实验。如果不考虑频率，而只考虑特征词是否出现作为分类特征，则SVM最高达到82.9%的准确率。Dave等人[15]则从CNET网站上下载带有情感极性标记的产品评论，作为有监督学习的语料库，从中选出分类特征词，经过参数平滑后，其最高准确率可达88%。Yu[16]也是采用了有监督学习法。

评论性文本来自多个领域，如数码产品评论和电影评论，这些来自不同领域的评论之间存在着差异，造成情感分类器的领域适应性很差。Blitzer[17]利用SCL（Structural Correspondence Learning）算法，根据领域共有高频词作为纽带特征（Pivot Feature），寻找目标领域评论中的新特征词，从而使源领域的分类器能适应目标领域的分类。Li[18]则通过多个独立领域分类器之间的互相学习和分类特征共享，实现多领域训练语料的融合，从而提高多领域文本情感分类的准确性。

2.2 句子级的情感分析

由于大部分文本都是主观句与客观句混合在一起，因此篇章级的情感分类太粗糙。句子级的情感分析，有的将句子分为客观或主观[19,20]，有的对主观句子再进一步分为褒义、贬义和中性[2,16,21]。Rilloff等人[19]首先利用已知情感词构造高精度的分类器，自动识别主观和客观句，然后根据句法模板从识别出的两类句子中提取一系列句法模式，最后使用这些模式作为分类特征，识别更多的主观和客观句子，其准确率高达91%。但是由于提取到的模式有限，故recall很低。后来他们又采用朴素贝叶斯分类器进行迭代地自我训练[20]，由于NB能使用更多的特征，从而提高了Recall，但准确率又有下降。Yu等人[16]则采用句子相似度、朴素贝叶斯分类器和朴素贝叶斯多分类器3种方法识别主观句子。对句子的情感极性分类，是根据句子中的情感词极性，其方法与Turney[5]类似，但使用了更多的种子词。Kim[2]在预测主观句的情感极

性时，首先用命名实体识别出可能的意见表达者，然后在附近寻找情感词，把多个情感词的情感极性相加，以此预测整个句子的情感极性。McDonald[21]定义了基于无向图结构的篇章-句子关系模型，将篇章级和句子级的情感极性标记看作序列标记问题，并利用条件随机场(CRF)求解最优标记序列。这种算法可以同时预测篇章级和句子级的情感极性，篇章级准确率可达87%，但句子级的准确率只有60%左右。

3 细粒度的意见挖掘与摘要研究

3.1 产品评论的意见挖掘与摘要

篇章级的情感分类和句子级的情感分析无法识别意见的语义成分。而且，对一个对象的某个属性的情感不能代表意见表达者对该对象的整体意见。所以需要进行细粒度的意见挖掘，其目标是识别意见表达的多个语义成分。产品评论的意见挖掘目标是发现评论者或意见表达者喜欢哪些属性、不喜欢哪些属性。对同一产品的评论可能会有很多，所以应该产生一个意见摘要。

网络上的产品评论可以分为3类[8]：1)首先分别指出优点和缺点，然后给出详细的评论；2)分别指出较详细的优点和缺点；3)优点和缺点混合在一起的自由评论。网站上对同一产品的评论一般集中在同一网页，产品名称即评论意见的主题，意见表达者也可以默认为用户群体。因此这类研究主要是识别被评论产品的属性和情感词，而不涉及主题的识别以及意见表达人的识别，这一类的意见摘要一般是图形化显示(产品每个属性给出正面评论的数量和负面评论的数量[8])，也可以是结构化的句子列表[22]。产品评论的意见挖掘和摘要的技术流程可以概括为图1。

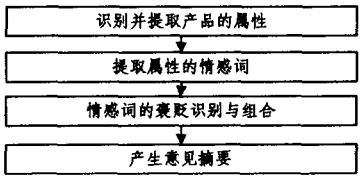

图1 产品评论挖掘与摘要技术流程

产品属性的识别可以分为无监督法[8,23-25]和有监督法[26-27,44]。Hu[8,23]假定产品属性都是名词，首先对产品评论进行词性标注，然后通过联合规则挖掘(Association Rule Minning)找出高频出现的名词作为产品属性。将高频属性和情感词抽取之后，再在剩下的带有情感词的句子中提取名词和名词短语作为低频属性。Popescu[24]等人在此基础上，通过去除那些可能不是特征的高频名词短语提高了准确性，但是Recall有所降低。Li[25]通过构造电影评论句中的属性-情感词对，利用WordNet和演员表提取属性、情感词列表。无监督法的准确率和召回率都不高，因此Liu等人[26]采用有监督法提取产品属性，首先由人工在分词后的语料中标注属性词，然后提取包含属性词的3-gram，通过联合规则挖掘学习(Association Rule Minning)其中的规则，根据规则提取产品属性；其针对的是上述2类评论。对于规则的挖掘，也可以使用标记序列规则(Label Sequential Rules，LSR)[27]，其方法是将评论语句序列转换为带词性标记的序列，然后将其中的名词提取为属性。Skomorowski[44]的方法则是统计形容词性的情感词与产品属性的位置关系，根据情感词提取产品属性。有监督方法准确率较高，但需要人工标记语料库，而且其领域适用性也有限制。识别出的属性有可能是同义词，可以使用WordNet对其合并[26]。后来Carenini等人[28]使用了更复杂的方法，但是该方法需要提供一个特征分类树。

情感词的提取方法主要利用属性与情感词在位置上的邻接关系。Hu[8]假定属性与情感词在评论句子中会一起出现，在得到评论中的属性后，选取属性前后一定长度的字符串，取出其中的形容词作为该属性的情感词。但这种方法没有处理属性和情感词的对应问题，而且只提取形容词的情感，所以性能相对较低。Popescu[22]将句法依存、词性与规则相结合，这样可以提取所有词性的情感词，但是由于规则需要人工整理，如果情感词出现的形式不在规则中就无法提取，而且一套规则不能适用于其他语言。一个意见中的同一个属性可能由多个情感词修饰，必须合并这多个情感词的褒贬值，可以简单相加[8]。如果考虑属性和情感词之间的距离，则可以提高准确性[29]。

3.2 比较评论句的识别与挖掘

人们就某一对象发表意见时，通常带有与其他对象的比较。比如：

例2 我觉得新版《末代皇帝》的配乐比老版的差很远。

比较评论句的语义成分可以分为意见表达者、实体1、实体2、比较谓词、属性、情感词[32]。例2中，“我”是意见表达者，“新版《末代皇帝》”是实体1，“老版”是实体2，“比”是比较谓词，“配乐”是属性，“差很远”是情感词。比较评论句可以分为排名比较和非排名比较[30]。排名比较指的是对评论的两个对象排出名次，这种情况又分3种：相等排名(如“A和B的配乐都很精彩”)、最高级排名(如“A是市场上同类产品中最好的”)、比较级排名(如“A的性能比B要好”)。非排名比较指的是不对评论对象做排序，比如“A价格很贵，B性能很好”。

目前对比较评论句的意见挖掘研究还很少。Jindal等人[30]采用类序列规则(Class Sequential Rule，CSR)和机器学习相结合的方法识别英语比较句，并对其分类。CSR是一种带有类标签的序列模式，而CSR中的序列模式又作为机器学习的特征。由于一个句子可能会同时满足多个规则，而规则之间有可能是相互冲突的。通过朴素贝叶斯分类模型可以组合多个规则，从而减轻或避免规则的冲突。实验也表明，采用CSRs的朴素贝叶斯模型取得最高的准确率为80%。Jindal[31]用比较关系表示比较评论句的核心语义，而比较关系表示为五元组〈关系词，产品属性，实体1，实体2，比较类型〉。比较关系的5个元素识别是通过标签序列规则(Label Sequential Rules，LSR)匹配实现的；从词性组合序列到带有类标签的CSR的转换称为一个LSR规则。

对于汉语的比较评论句挖掘，Hou[32]将比较评论句的语义角色分为比较谓词、评论者、评论属性、实体1、实体2、情感词，并采用基于CRF的语义角色标注方法识别这6类角色，根据情感词的情感极性判定比较关系。

3.3 博客与新闻的意见挖掘与摘要

这一类的意见挖掘与摘要主要研究对象是博客文章和新闻报道，产生的意见摘要是文本形式的。这类研究主要涉及意见主题和意见表达人的识别。对一个或多个文档进行意见挖掘之后，可以形成意见摘要。传统的自动摘要算法主要是

提取文档中的重要信息，而将不重要的冗余信息去除。意见文本中重复出现的相同倾向的意见不能丢掉，因为这种重复加强了情感程度。

意见主题的识别，可以是句子级[35,36]的意见主题，也可以是(多)篇章级[33,34]的意见主题。如果意见摘要是针对篇章级的主题，则首先要识别篇章级的意见主题，然后提取相关意见句组成摘要。Ku[33]首先识别多篇汉语文本的主题，然后提取各文档中与主题相关的意见句子，最后根据意见的极性形成正反面意见摘要。其提取多文档意见主题是根据词在文档中出现的频率。$Disp_{P_i}$，$Disp_{S_i}$分别表示一个词在所有段和所有文档中出现的频率，$Dev_{P_{j,t}}$，$Dev_{S_{i,t}}$分别表示一个词在某段和某文档中出现的频率。满足 $Disp_{P_i} \leqslant Disp_{S_i}$，$\exists S_i$，$\forall P_j \in S_i$，$Dev_{S_{i,t}} \leqslant Dev_{P_{j,t}}$或者 $Disp_{P_i} > Disp_{S_i}$，$\exists S_i$，$\forall P_j \in S_i$，$Dev_{S_{i,t}} > Dev_{P_{j,t}}$的词可以认为是主题词。Hu[34]认为读者回复一篇博客文章通常针对文章的中心意见主题，因此可以利用博客文章的回复内容进行篇章级意见主题识别，再从博客文章中提取摘要句子。Kim[35]采用语义角色标注方法识别句子级的意见主题。Qiu[36]首先提取意见句子，然后采用基于规则的方法提取意见句子的主题，而规则是采用汉语依存语法根据一个词的句法角色构造。

意见表达者的识别，一般采用语义角色标注方法。Bethard[37]针对的是以情感动词(believe，realize)表达意见的句子，从 FrameNet 和 PropBank 两个语义角色标注的学习语料选择一部分句子重新标注，作为训练语料。Kim [35]首先识别情感词，然后对含有情感词的句子进行语义角色标注，最后从语义角色中选择意见表达者和意见主题，因此识别出的是单个句子中的主题。Kim[38]首先通过命名实体识别找出所有可能的意见表达者，然后用最大熵模型根据句法特征选择最可能的表达者。然而由于这几种方法依赖于语料库，因而具有较低的领域适应性。Choi[39]则将意见表达者识别看作一种信息提取，采用基于条件随机场(CRF)的序列标记和模式匹配相结合的方法，但性能提高不大。

以上的意见表达者识别并没有考虑非命名实体类(如人称代词 he，they 等)的意见表达者，因而意见摘要中可能会存在让人感到不知所云的人称代词。而且在包含意见的篇章中，现实世界的同一个意见表达人可能以不同的名称出现。Stoyanov[40]提出采用共指消解的方法解决这一问题。共指消解[41]就是将篇章内的所有表述划分为现实世界中不同实体等价描述的过程，主要包含人称代词消解和名词短语消解。如果通过共指消解将人称代词替换为对应的实体名称，可以生成更容易理解的意见摘要。

TAC 2008 设有意见摘要专题研究，目标是从博客文档集中产生关于特定话题的意见摘要，这些意见摘要能够回答关于指定话题的一些问题，其测试数据和评价指标与意见问题回答专题研究相同。

4 意见检索的研究

意见检索是传统信息检索的更高形式。检索出的文档必须满足两个条件：与检索词紧密相关；含有与检索词相关的意见，不管意见是正面还是负面的。根据 Mishne[42]对检索日志的统计分析，博客搜索可以分为两类：上下文检索和概念检索。上下文检索查找博客中评论命名实体的上下文信息；概念检索则是要查找那些讨论某个概念和话题的博客。博客搜索比一般的网络搜索更关注技术、娱乐和政治，尤其是对当前热点时事；用户只关注排名最靠前的几个检索结果。

当前的意见检索采用了两种不同的检索架构(如图 2 所示)，即一阶段法[43]和二阶段法[44]。二阶段法首先采用传统的 IR 模块检索出相关文档，然后利用粗粒度的意见分类技术选择其中带有意见的文档，并抽取其中的意见句子。一阶段方法是首先利用意见分类技术抽取所有文档中的意见句子，然后单独为意见句子构造索引，并保存情感词的情感极性。在进行意见检索时，就可以像传统的检索方法那样直接找到相关的意见句子。一阶段方法丢失了一些上下文信息，使得无法判断某些意见句与检索词的相关程度；二阶段方法则增加了检索时间。当前的意见检索主要针对的是上下文检索，命名实体可以是意见主题(产品、旅游景点、新闻事件等)，也可以是意见表达者。

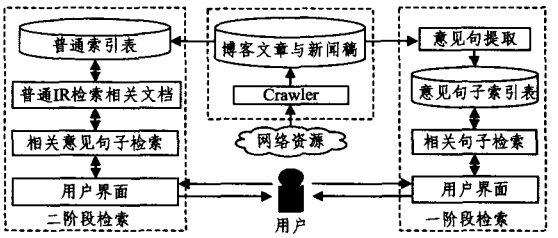

图 2 两种不同的意见检索架构

检索出的文档中含有的意见，其主题不一定是针对检索词的，因此还要判断意见句的主题相关性，并以此作为文档排序的依据。Furuse[43]认为意见句与检索词相关的条件是：在一段连续的多个意见句，如果出现检索词，则认为所有的意见句都相关，出现检索词的单个意见句也是相关的。这种判断方法无法给出文档排序。Skomorowski[44]假设情感词的修饰对象就是要查询的命名实体，一阶段检索出相关文档后，在检索词前后 10 个词中提取情感形容词，根据先前统计学习得到的概率，计算检索词被情感形容词修饰的概率，将概率值的和作为排序的依据。Zhang[45]则采用 3 种指标对相关文档排序：普通 IR 模块计算出的主题相关性；文档中的句子为主观句子的概率和；文档中含有意见句的数量。

据统计，网络上约有 20%是垃圾(Spam)博客。由于这些垃圾博客的存在，造成博客意见检索的准确率很低。Mishne[46,47]考虑了博客文章的时效性，提高时间上最集中的那些博客文章的相关性，同时将回复数量作为衡量博客质量高低的指标，并加入垃圾博客过滤技术。这样做可以提高具有商业倾向的检索词的准确率。而没有商业倾向的检索词则会降低准确率。Weerkamp 等人[48]在检索过程中利用博客文章可信性指标过滤那些质量低的博客，从而提高检索的准确率；他们利用的 11 个指标包括：是否正确大小写、使用表情符号(emoticons)的多少、单词拼写错误的多少、文章长度、文章时效性等，博客空间级指标包括是否是垃圾博客、博客空间的回复数量、博客主人写作是否有规律。

5 语言资源与应用系统

5.1 用于意见挖掘的语言资源

MPQA 意见语料库最初是为了 ARDA 资助下的一个关于多角度自动问答(Multi-Perspective Question Answering)

的专题研究，主要建立者是 Wiebe。该语料库包含从各种新闻语料资源中选出的 535 篇新闻稿，11114 个句子；每个句子中的意见，人工标注了意见的发表人、话题和意见的态度（意见的强弱、重要性、极性）。电影评论语料资源（Movie Review Data）包括情感极性数据集（两种极性的评论各 1000，两种极性的句子各 5331）和主观数据集（主客观句子各 5000）；评论文本标注整体的情感极性和主观程度，句子标注主客观类别和情感极性，主要创建者是 Pang 等人。SentiWordnet 是一个基于 WordNet 2.0 的情感词典，每一组同义词都标有 3 种情感极性的概率，作者是意大利 Padova 大学的 Andrea Esuli 等人。TREC Blog 2006 提供了博客文章语料，NTCIR-6 也提供了用于意见挖掘的多国语言语料。

中文《HowNet》也发布了“情感分析用词语集（beta 版）”，中、英文词语集各 6 个文件，包括正、负面情感词语，正、负面评价词语，程度级别词语和主张词语，包含词语约 17887 条。台湾大学的陈信希、古伦维制作了用于情感挖掘的繁体汉语情感词典 NTU SD。目前还没有用于汉语意见挖掘的语料库。

5.2 意见挖掘系统

意见挖掘系统可分为两类：一类是利用意见挖掘技术开发的商业情报系统，一类是博客搜索和挖掘系统。NEC 美国实验室的 Dave 等人[15]所开发的 ReviewSeer，通过对评论性文章的语义倾向分析，为商品的受欢迎程度打分，获取商品的用户信息。Liu 的商品信息反馈系统 Opinion Observer[26]，利用网络上丰富的顾客评论资源，进行商品的市场反馈分析，为生产商和消费者提供直观的针对商品各个属性的评价报告。与之类似的系统还有 WebFountain[49]。Blogdigger 是第一个博客和 RSS 搜索引擎。马里兰大学开发的 BlogVox[50] 在检索到相关的博客文章后，能计算文章含有主观意见的概率，并据此排序。

在中文方面，姚天昉等人开发了汉语汽车论坛的意见挖掘系统[51]。该系统可在各大论坛上挖掘顾客对各种汽车品牌的不同性能指标的评论和意见，并且在最后给出意见摘要的可视化结果。

结束语 本文总结了国内外对网络意见挖掘、摘要和检索的研究现状。虽然这方面已经取得了很多进展，并出现了实际应用系统，但对意见的语义理解还处在相对较低的水平上。

意见文本由不同背景的网民写成，其书写质量和风格差别很大，给研究造成了一定的困难。用于意见挖掘的语料资源，特别是汉语的语料资源还相对缺乏，标注也不够规范和统一。目前的意见挖掘相对还是比较粗糙的，主要是由于对各语义成分的识别精度还很低。

进一步深入的研究需要更丰富的语料资源和规范的标注。以后还要提高对意见语义成分的识别精度，语义成分间的关系也需要考虑更复杂的情形。对情感词，尤其是严重依赖上下文的情感词要进行精细的情感分析。意见主题，尤其是（多）篇章主题的识别，意见发表人的共指消解也要继续深入。自然语言处理的其他技术也会应用到意见挖掘。

参 考 文 献

[1] Gildea D, Jurafsky D. Automatic labeling of semantic roles[J]. Computational Linguist, 2002, 28(3): 245-288

[2] Kim S M, Hovy E. Determining the Sentiment of Opinions[A]//Proceedings of COLING-04[C]. Geneva, Switzerland, 2004: 1367-1373

[3] Hatzivassiloglou V, McKeown K R. Predicting the semantic orientation of adjectives[C]//Proceedings of the ACL. 1997: 174-181

[4] Wiebe J, Learning subjective adjectives from corpora[C]//Proceedings of the 17th National Conference on Artificial Intelligence. 2000: 735-740

[5] Turney P D, Littman M L. Measuring Praise and Criticism: Inference of Semantic Orientation from Association[J]. ACM Transactions on Information System(TOIS), 2003, 21(4): 315-346

[6] Yuen R W M, et al. Morpheme-based Derivation of Bipolar Semantic Orientation of Chinese Words[C]//Proceedings of COLING. Geneva, Switzerland, 2004: 1008-1014

[7] Kamps J, Marx M, Mokken R J, et al. Using WordNet to measure semantic orientation of adjectives[C]//Proceedings of 4th International Conference on Language Resources and Evaluation. Lisbon, Portugal, 2004: 1115-1118

[8] Hu M, Liu B. Mining and summarizing customer reviews[C]//the Proceedings of KDD. Seattle, Washington, USA, 2004

[9] Kim S M, Hovy E. Identifying and Analyzing Judgment Opinions[C]//Proceedings of HLT-NAACL 2006. New York, US, 2006

[10] Esuli A, Sebastiani F. Determining the Semantic Orientation of Terms Through Gloss Classification[C]//Proceedings of CIKM. Bremen, Germany, 2005: 617-624

[11] Andreevskaia A, Bergler S. Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses[C]//Procee-dings of EACL. Trento, Italy, 2006

[12] 朱嫣岚，闵锦，等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报，2006，20(1)：14-20

[13] Turney P D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews[C]//Proceeding of 40th Annual Meeting of ACL. Philadelphia, July 2002: 417-424

[14] Pang B, Lee L. Thumbs up? Sentiment Classification Using Machine Learning Techniques[C]//Proceedings of EMNLP. Philadelphia, July 2002: 79-86

[15] Dave K, Lawrence S, Pennock D. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews[C]//Proceedings of the WWW-03. Budapest, Hungary, 2003

[16] Yu H, Hatzivassiloglou V. Towards answering opinion question: Separating facts from opinions and identifying the polarity of opinion sentences[C]//Proceedings of EMNLP. 2003

[17] Blitzer J, Dredze M, Pereira F. Biographies, Bollywood, Boomboxes and Blenders: Domain adaptation for sentiment classification[C]// Proceedings of ACL. 2007

[18] Li F, Zong C. Multi-domain Sentiment Classification[C]//Proceedings of ACL-08. Ohio, USA, 2008: 257-260

[19] Riloff E, Wiebe J. Learning extraction patterns for subjective expressions[C]//Proceedings of EMNLP. 2003

[20] Wiebe J, Riloff E. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts[C]//Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing. 2005

（下转第 51 页）

ternational Conferrence on System Sciences. 2002

[15] 唐文,陈钟. 基于模糊集合理论的主观信任管理模型研究[J]. 软件学报,2003,14(8):1401-1408

[16] Yahalom R, Klein B, Beth T. Trust relationships in secure systems a distributed authentication perspective[C]// Proc. 1993 IEEE Symp. on Research in Security and Privacy. 1993:150-164

[17] Yahalom R, Klein B, Beth T. Trust- based navigation in distributed systems[J]. Special Issue "Security and Integrity of Open Systems"of the Journal "Computing Systems", 1994

[18] Reiter M K, Stubblebine S G. Resilient authentication using path independence[J]. IEEE Transactions on Computers, 1998, 47 (12)

[19] 白保存,李中学,陈旺. PKI 信任度模型路径算法研究[J]. 计算机工程与应用,2005,41(21):182-185

(上接第 19 页)

[21] McDonald R, Hannan K, Neylon T, et al. Structured Models for Fine-to-Coarse Sentiment Analysis[C]// Proceedings of ACL. 2007

[22] Puspesh K. Multi-document Update and Opinion Summarization [D]. Partial fulfillment of the requirement for the degree of Masters of Technology, Indian Institution of Technology, 2007-2008

[23] Hu M, Liu B. Mining Opinion Features in Customer Reviews[C] //Proceedings of 19th National Conference on Artificial Intelligence. San Jose, USA, July 2004

[24] Popescu A-M, Etzioni O. Extracting Product Features and Opinions from Reviews[C]//Proceedings of EMNLP. 2005

[25] Zhuang Li, Jing Feng, Zhu Xiaoyan. Movie Review Mining and Summarization[C]// Proceedings of CIKM-06. Virginia, USA 2006

[26] Liu B, Hu M, Cheng J. Opinion Observer: Analyzing and Comparing Opinions on the Web[C]// Proceedings of the 14th International Conference of World Wide Web. Chiba, Japan, 2005

[27] Liu Bing. Web Data Mining: Exploring Hyperlinks, Contents and Usage Data[C]. Springer, December 2006

[28] Carenini G, Ng R T, Zwart E. Extracting knowledge from evalua-tive text[C]// Proceedings of the 3rd international Conference on Knowledge Capture Banff, Alberta, Canada, October 2005

[29] Ding Xiaowen, Liu Bing. The Utility of Linguistic Rules in Opinion Mining[C]// Proceedings of SIGIR-07. Amsterdam, July 2007

[30] Jindal N, Liu Bing. Identifying Comparative Sentences in Text Documents[C]// Proceedings of the 29th Annual International ACM SIGIR Conference. Seattle, 2006

[31] Jindal N, Liu Bing. Mining Comprative Sentences and Relations [C]// Proceedings of 21st National Conference on Artificial Intellgience. Boston, Massachusetts, USA, July 2006

[32] Hou Feng, Li Guo-hui. Mining Chinese Comparative Reviews by Semantic Role Labeling[C]// Proceedings of 7th International Coference on Machine Learning and Cybernetics. Kunming, 2008

[33] Ku Lun-Wei, Liang Yu-Ting, Chen Hsin-Hsi. Opinion Extraction, Summarization and Tracking in News and Blog Corpora[C] //the Proceedings of 21st National Conference on Artificial Intellgience. Boston, Massachusetts, July 2006

[34] Hu Meishan, Sun Aixin, Lim Ee-Peng. Comments-oriented document summarization: understanding documents with readers' feedback[C]// Proceedings of the 31st Annual International ACM SIGIR Conference. Singpore, 2008

[35] Kim S M, Hovy E. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text[C]//Proceedings of the Workshop on Sentiment and Subjectivity in Text of COLING-ACL. Sydney, Australia, 2006

[36] Qiu G, Liu K, Bu J, et al. Extracting Opinion Topics for Chinese Opinions Using Dependence Grammar[C]// Proceedings of SIGKDD-ADKDD. San Jose, California, USA, 2007

[37] Bethard S, Yu H, Thornton A, et al. Automatic Extraction of Opinion Propositions and their Holders[C]// Proceedings of the AAAI Spring Symposium. Stanford, USA, 2004

[38] Kim S M, Hovy E. *Identifying Opinion Holders for Question* Answering in Opinion Texts[C]// Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains. 2005

[39] Choi Y, Cardie C, Riloff E, et al. Identifying sources of opinions with conditional random fields and extraction patterns[C]//Proceedings of HLT/EMNLP-05. Vancouver, B. C., 2005

[40] Stoyanov V, Cardie C. Partially Supervised Coreference Resolution for Opinion Summarization Through Structured Rule Learning[C]// the Proceedings of EMNLP. 2006

[41] Lang Jun, Qin Bing, Liu Ting, et al. Intra-document Coreference Resolution: The state of the art[J]. Journal of Chinese Language and Computing, 17(4): 227-253

[42] Mishne G, de Rijke M. A study of blog search[C]// Proceedings of 28th European Conference on Information Retrieval. London, 2006

[43] Furuse O, Hiroshima N, et al. Opinion Sentence Search Engine on Open-domain Blog[C]// Proceedings of IJCAI. 2007

[44] Skomorowski J. *Topical Opinion Retrieval*[M]. *Dissertation of* Master of Mathematics in Computer Science. Waterloo, Canada, 2006

[45] Zhang Wei, Yu C, Meng Weiyi. Opinion Retrieval from Blogs[C] //Proceedings of ACM 6th CIKM. Lisboa, Portugal, 2007

[46] Mishne G. Using blog properties to improve retrieval[C]// Proceedings of ICWSM. 2007

[47] Mishne G. Applied Text Analytics for Blogs[D]. University of Amsterdam, 2008

[48] Weerkamp W, de Rijke M. Credibility Improves Topical Blog Post Retrieval[C]// Proceedings of ACL08. 2008

[49] Yi J, Niblack W. Sentiment Mining in WebFountain[C]// Proceedings the 21st International Conference on Data Engineering. Tokyo, Japan, 2005: 1073-1083

[50] Java A, Kolari P, Finin T, et al. The BlogVox Opinion Retrieval System[C]// Proceedings of the Fifteenth Text REtrieval Conference. 2006

[51] 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[C]//中国中文信息学会二十五周年学术会议论文集. 北京:清华大学出版社,2006:260-281