融合多种情感资源的微博情感分类研究

顾益军[1] 刘小明[2]
(中国人民公安大学网络安全保卫学院 北京 100038)[1] (中原工学院计算机学院 郑州 450007)[2]

摘 要 为了通过融合多种情感资源库中的词汇情感特征来提高微博情感分类精度,提出了一种词汇情感确定性度量的计算方法,并以此为基础将在多种情感词汇上获取的情感特征融合为词汇的综合情感特征,然后采用机器学习的分类方法实现微博观点句识别和观点句情感倾向性判定。实验表明,本方法利用词汇的情感确定性度量,统一了词汇情感倾向性的强度度量,在观点句识别和观点句情感倾向性判定两个情感分类任务中都取得了较好的性能。

关键词 情感确定性度量,情感分类,观点句识别,情感倾向性判定,微博
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.4.042

Research of Microblog Sentiment Classification Based on Emotional Polarity Certainty for Vocabulary

GU Yi-jun[1] LIU Xiao-ming[2]
(College of Network Security, People's Public Security University of China, Beijing 100038, China)[1]
(School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, China)[2]

Abstract To improve the sentiment classification precision, the paper proposed a method to determine the certainty of sentiment. The method takes account of multiple sentiment features of vocabularies in the sentiment database and identifies the clause giving opinions. It classifies the clauses into different emotinoal polarities through machine learning. The experiment shows that the proposed method employing the certainty of the sentiment unifies the emotional polarities measurement. It has a better performance on identifying the clause giving options and emotional polarity classification.

Keywords Emotional polarity certainty, Sentiment classification, Sentiment sentence recognition, Emotional polarity classification, Microblog

1 引言

微博作为一种新兴的网络信息传播媒体,因快捷、方便、操作简单等特点,已成为当前用户数量最多的互联网信息传播方式和平台。微博中蕴含了广大网民最为丰富的主观情感信息。例如,微博中对某些热点事件的批评、赞扬或反对、赞同等情感倾向性信息,最能反映当前民众对各种热点事件的舆论倾向。对相关微博按不同情感类别进行分类,可迅速获取不同阶层网民群众的情感、情绪,可为分析人员提供可靠的数据依据,有助于迅速、快捷、全面、准确了解广大网民的诉求,为制定相关决策提供可靠支持。同时,微博情感分类在市场预测、信息安全、网络舆情、商业决策、电子商务等众多领域都有广阔应用前景。

本文提出了一种融合了多种情感资源特征的微博文本情感分类方法,该方法具有较高准确性和运行效率,在实际应用中也取得了较好效果。实验结果表明,通过使用特征的抽取模式以及多特征融合的分类方法,缩减了特征集合中特征的数量,增强了对中文微博情感分类的效果。

2 相关研究

微博情感计算任务分为两个部分:1)微博中情感观点句的识别,即识别微博句子是否(Yes/No)表述了作者的某种观点或态度,或者蕴含有作者的某些情感、情绪信息,这是一个二元分类问题;2)情感观点句的情感倾向性判定,即判定情感观点句中蕴含的情感倾向性是正面(Positive)、负面(Negative)、中立或不置可否(Other),因此是一个多元分类任务[7]。

文本情感分析方面的研究在国外开展较早,已取得了非常多的研究成果。PangBo 等人[1]最早将机器学习方法应用于文本情感分类的研究,发现以 unigram 为特征并结合 SVM 算法时的结果最好;Whitelaw 等人[2]利用带有形容词的词组为特征,实现了对评论文档正负面分类;Tourney 等人对手机、银行、电影等相关的一些评论做了情感倾向分析研究,他们选出了两个基本情感词:正向词 excellent 和负向词 poor,然后使用 PMI 计算短语与基本正向情感词的关联度和负向情感词的关联度,根据两种关联度的差值来判定评论情感倾向性。

国内对文本情感分类的研究相对较少且起步较晚,但近些年也已成为研究热点,特别是对中文情感分类在网络舆情分析与监控研究中发展迅速。近些年来,自然语言处理与中文计算年会(NLP&CC)设立了中文微博情感分析的评测[3],并且全国信息检索学术会议(CCIR)设立了中文倾向性分析评测(COAE)[4],掀起了国内外学者对微博情感分析研究的

到稿日期:2014-04-21 返修日期:2014-07-14
顾益军(1968—),男,博士,副教授,主要研究方向为网络情报技术,E-mail:yj_gu@163.com;刘小明(1979—),男,博士,讲师,主要研究方向为网络舆情、情感计算。

高潮，极大地推动了中文本情感分析研究的深度和广度。

朱嫣岚等人[5]以知网（HowNet）情感词集[6]为基础资源，利用语义相似度和语义相关场方法实现对汉语常用词中文本的情感分类；徐军等基于具有情感倾向词汇特征和否定词特征，采用二值作为特征项权重，利用最大熵和朴素贝叶斯的方法实现了对新闻及评论极性情感的分类；Jiang 等人提出了引入主题相关特征的情感分析方法，其最高准确率达到了66.0%。

通过对以上研究的分析发现，中文微博文本情感计算研究面临的一个重要问题是相关资源匮乏，特别是缺乏相对完备、质量较高的情感词库。为此，本文利用词汇的情感确定性度量作为词汇情感强度特征的补充，实现了将多种情感词典中的词汇情感特征的融合，从而得到词汇的综合情感特征，再利用支持向量机分类器实现了对微博文本情感分类性能的提升。

3 词汇的情感特征

在现代汉语中词汇是表达语义信息的最小单位，情感信息作为语义信息的重要组成部分也是由词汇表述的。同样，微博作为一种网络自媒体工具，虽然文本描述风格、方式多种多样，但其最基本的语言单位还是词汇。微博文本中词汇的情感信息是构成微博情感信息的基础。对微博文本的情感分类的关键在于充分获取其词汇的情感特征。

在目前文本情感分类研究中，词汇情感特征的获取方式主要依赖于在词典中词汇情感特征的约定与描述。目前，关于词汇是如何表述作者的情感信息在语言学界还没有统一的观点和认知，这就造成了各家所构建的情感词库差别较大，没有一致的标准。词典中词汇所描述词汇情感信息的差异主要存在于以下两个方面：首先，不同词典中包含的情感词汇有很大不同，这是由人类自然语言的本质特征所决定的。自然语言中用来表述情感概念的本身是模糊的，因此某些词汇可能被一些学者认为是情感词汇，而另一些学者则可能认为不是。其次，即使不同情感词典都认为是带有情感信息的词汇，但情感特征可能存在很大不同，如相同的情感词汇在一个词典中认为是正面情感，在另一些词典中则可能为负面情感。

例如，大连理工大学的《情感词汇本体》和清华大学的《情感极性此表》都标注有词汇的情感倾向性和词汇的情感强度。这些词库资源中的情感信息比较充分，可以直接用来确定词汇在微博文本中的情感特征。但《知网情感词库》、《台湾大学中文通用情感词典》和《互联网非正式文本词库》仅标注了词汇的情感倾向性信息，而没有给出词汇的情感强度信息。

目前，在对中文文本情感分类中使用的词汇情感特征主要包括以下 4 种。

3.1 TFIDF 特征

TDIDF 值是在文本分类方法中使用最为广泛的特征，也是当前绝大多数搜索引擎、信息检索和问答系统等应用和研究领域使用的基础方法。现有的各种情感词库的覆盖率都不高，仅利用情感词库将很难保证获取充分的情感特征。比如有些句子中虽然不含任何情感词汇，但句子中的关键词汇所表达的意义却带有作者的情感信息。因此，我们将词汇的 TFIDF 特征作为微博句子的基础情感特征。微博文本中的词汇 TFIDF 值的计算公式如下：

$$tfidf(w_i)=tf\cdot idf=tf\cdot\log\frac{N}{df}$$

其中，tf 表示词汇 w_i 在微博某个句子中出现的次数，N 为全部微博句子的总数目，df 为包含词汇 w_i 的微博句子数目。

3.2 知网情感特征

知网（HowNet）情感词集包含了中英文双语的正负面情感评价词和正负面情感词。将其中共 3000 中文词作为特征词。基于 HowNet 情感词集可简单将词的正面或负面情感极性作为情感特征。例如，微博句子 S 中的某个词汇 w_i 的情感特征值如下：

$$hownet(w_i)=\begin{cases}1, & w_i\in\{PlusFeeling\}\cup\{PlusSentiment\}\\ -1, & w_i\in\{MinusFeeling\}\cup\{MinusSentiment\}\end{cases}$$

其中，$\{PlusFeeling\}$和$\{PlusSentiment\}$分别为正面情感词集和正面评价词集；$\{MinusFeeling\}$和$\{MinusSentiment\}$分别为负面情感词集与负面评价词集。

3.3 情感本体库特征

情感本体库（Sentiment Ontology）是由大连理工大学信息检索研究室独立整理标注完成的情感知识库[8]。该情感本体词库包含 27466 个常用中文情感词。对其中的每一个词，在标注了词性、词义数、词义序号等基本信息的基础上，进一步标注了词的情感分类、强度、极性、辅助情感分类、辅助强度和辅助极性，共 8 项情感描述信息。在我们的情感分类系统中利用词的情感分类、强度和极性值生成词的情感特征值。其中，对于任意的 w_i 在 Sentiment Ontology 上获取的情感特征值 $sentonto(w_i)$ 如下所示：

$$sentonto(w_i)=type(w_i)\times intensity(w_i)\times polarity(w_i)$$

$$type(w_i)=\begin{cases}1, & \text{情感分类标识以“N”开头；}\\ -1, & \text{情感分类标识以“Y”开头；}\end{cases}$$

其中，$type(w_i)$是对情感分类的数值化结果，$intensity(w_i)$为强度，$polarity(w_i)$为极性。

3.4 情感极性词库特征

情感极性词库（SentimentPolarity）是清华大学中文系原博博士构建的包含有 23419 个汉语情感的词集。对于该词集中的每一个词，给出了词条目的格式为“词语\t 极性值”，其中极性值的“+”、“-”符号标识词汇的情感倾向性为正面或负面的强度，实数表示的词汇极性值标识词汇的情感强度。因此，对于任意词 w_i，在情感极性词库上获得的情感特征值为 $sentimentpolaity(w_i)=polarity\_value(w_i)$。

3.5 互联网变异词特征

微博是一种网络信息传播工具，用户在发表言论时使用了大量的“互联网非正式文本”。例如，用户经常会使用“☺”、“☹”等表情符号表示“高兴”或“悲伤”情感；会用“74”代替“去死”，用“7456”代替“气死我了”等。这些非正式文本强烈地表述了用户的情感信息，对微博文本情感分类结果影响重大，有时可直接标识用户的情感倾向。

为了实现对这些互联网非正式文本所蕴含情感信息的处理，我们手工构建了一个《互联网非正式文本词库》，为其中的每个词条名标明了情感倾向性。该词库的部分示例内容已经公开发布（http://www.datatang.com/data/45752），可供免费下载试用。

4 词的情感确定性度量特征

词的情感强度可以看作词在不同情感倾向性上的确定性度量(Emotional Polarity Certainty, EPC)。词汇的情感确定性度量定义如下：

定义 1(词汇情感不确定性度量) 利用词在训练语料中的统计值，可计算出词对不同情感倾向性的确定性度量。

词汇的正面(Positive)情感倾向性确定性度量可用如下公式计算：

$$C_{pos}(w_i|S)=\begin{cases} ROUND\left(5*\dfrac{P(w_i|D_{pos})-P(w_i)}{1-P(w_i)}\right), & P(w_i)<P(w_i|D_{pos}) \\ 0, & P(w_i|D_{pos})\leqslant P(w_i) \end{cases}$$

其中，$P(w_i|D_{pos})$为词 w_i 在情感倾向为正面的文档集中出现的文档概率；$P(w_i)$为在均衡文本集中出现的文档概率。

可以计算得到词汇正面情感倾向性的确定性度量值 $C_{pos}(w_i|S)$是一个值域为 0 到 5 的整数。首先，由 $P(w_i|D_{pos})\geqslant P(w_i)$可计算得到$\dfrac{P(w_i|D_{pos})-P(w_i)}{1-P(w_i)}>0$。由此可得如下推理过程：

$$\frac{P(w_i/D_{pos})-P(w_i)}{1-P(w_i)}\leqslant 1$$

$$\Leftrightarrow P(w_i/D_{pos})-P(w_i)\geqslant 1-P(w_i)$$

$$\Leftrightarrow P(w_i/D_{pos})\leqslant 1$$

因此，$0<\dfrac{P(w_i|D_{pos})-P(w_i)}{1-P(w_i)}\leqslant 1$。对 $ROUND\left(5*\dfrac{P(w_i|D_{pos})-P(w_i)}{1-P(w_i)}\right)$取整，得到的值为区间在 0 到 5 的整数。

与词汇的正面情感确定性度量值计算方法类似，情感极性为负面(Negative)的词情感确定性度量可用如下公式计算：

$$C_{neg}(w_i|S)=\begin{cases} ROUND\left(5*\dfrac{P(w_i|D_{neg})-P(w_i)}{1-P(w_i)}\right), & P(w_i)<P(w_i|D_{neg}) \\ 0, & P(w_i|D_{neg})\leqslant P(w_i) \end{cases}$$

其中，$P(w_i|D_{neg})$为 w_i 在负文本集中出现的文本概率，$P(w_i)$为在均衡文本集中出现的文本概率。$C_{neg}(w_i|S)$的值域同样为 0 到 5 的整数。

《知网情感词库》、《台湾大学中文通用情感词典》和我们构建的《互联网非正式文本词库》中，都仅给出了这些情感词汇的情感倾向性，而没有给出词汇的情感倾向性度量值，因此可以利用 $C_{pos}(w_i|S)$和 $C_{neg}(w_i|S)$值作为词汇的情感强度值，从而将各种情感词汇的倾向性强度值统一。

5 基于综合特征的情感分类

词汇的情感确定性度量弥补了现有情感词典中对词汇情感标注不充分的问题，将词汇的情感确定性特征与现有情感词典中的情感标注内容综合，可以获取微博文本中更多关键词的情感特征。本文将多种情感词库中词汇的情感极性和极性强度值相融合，从而构建了一个综合情感词库。利用综合情感词库中的词汇情感特征，实现对微博句子的情感分类改进。基于上述方法得到的词汇综合情感特征，实现对“微博观点句识别”和“微博观点句情感倾向性判定”这两个情感分类任务。

基于综合情感特征的情感分类的流程如图 1 所示。

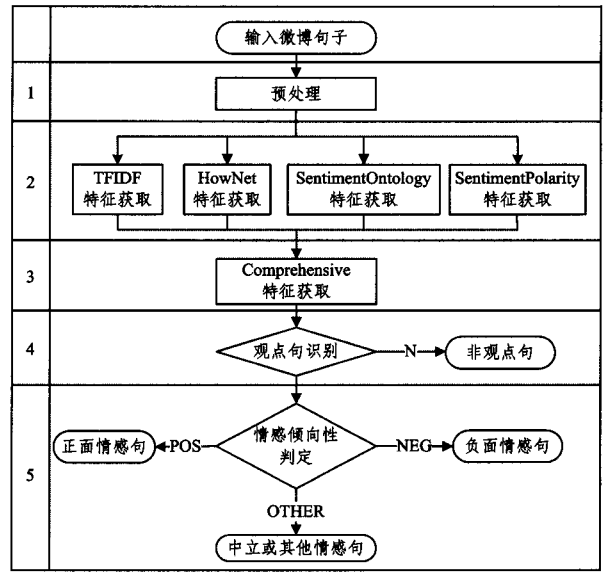

图 1 基于综合情感特征的情感分类的处理流程

1. 预处理

预处理主要是分词与词性标注(Part of Speech, POS)。本文使用了张华平博士研发的 NLPIR 汉语分词系统[9]。上述多种情感词库中的很多词和我们构建的《互联网变异词》中的词条目对 NLPIR 系统来说都是未登录词，因此，需要首先将这些词的词文本和词性加入到用户词典(userdic. txt)中。

2. 获取微博句子在各种情感资源上的情感特征

按第 4 节所述，对句子中的每一个关键词，分别在每一个情感词库上获得该词条目所对应的情感特征。

3. 生成句子的综合情感特征

基于词汇的综合情感特征，将句子每个词在各种情感资源上获取的情感特征融合，从而得到句子的综合情感特征向量。

4. 观点句识别

首先，获取微博句子 S 的综合特征向量 $FV(S)$，以及 S 的情感类别表示 Label。本文情感观点句识别的类别标识分别为 1(情感句)和 0(非情感句)。将情感类别的标识与特征向量按照 libsvm 软件要求的数据格式生成所需训练数据。然后，将训练数据送入 libsvm，生成情感观点句识别的分类器 opinion_classifier。对于任意给出的微博句子 S，按照相同的方式得到情感特征向量 $FV(S)$，将其输入到训练好的分类器 opinion_classifier中，即可得到对该句子观点句的识别结果和相应的概率值。

5. 情感倾向性判定

若微博句子 S 观点句识别的结果为“1”，即 S 是观点句，则可再次利用该综合情感特征判定 S 的情感倾向性。对观点句情感倾向性判定的类别标识分别为 1(负面)、-1(正面)、0(中性或不置可否)。与情感观点句识别类似，生成综合情感特征向量，利用相同的情感特征向量有助于降低系统复杂度。

同样，将训练数据送入 libsvm，生成情感观点句的倾向性

分类器 opinion_classifier。对于任意给出的微博句子 S，按照相同的方式得到情感特征向量 $FV(S)$，将其输入到训练好的分类器 opinion_classifier 中，即可得到对该句子观点句的识别结果和相应的概率值。

6 实验

实验分为微博观点句的识别和微博观点句的情感倾向性判定两部分。为对比验证本文方法的效果，分别实现了基于 TFIDF 特征、基于 HowNet 情感词库特征、基于情感本体词库(SentiOnto)特征、基于情感极性词库(SentiPola)特征的情感分类方法。

实验数据以 NLP&CC2012 中文微博情感分析评测公布的样例数据为基础，再以从互联网上抓取的 2 万微博句子作为补充。抓取的数据参照 NLPCC 格式进行人工标注。实验中，随机抽取 80%的句子作为训练数据，其余 20%作为测试数据。

实验结果参考 NLP&CC 中文微博情感分析评测任务的评测标准，同样利用正确率 *Precision*、召回率 *Recall* 和 F 值 *F-measure* 3 个评测指标作为我们实验结果的评测标准。

6.1 实验结果

为充分验证本文所提出词汇综合情感特征的效果，分别在微博文本情感分类的观点句识别和观点句情感倾向性判定任务中实验，结果如下。

(1)观点句识别结果

对情感观点句识别的结果如表 1 所列。

表 1 情感观点句识别结果

情感特征	Precision	Recall	F-Measure
TFIDF	0.6550	0.7505	0.6995
HowNet	0.6652	0.7760	0.7164
SentiOnto	0.6272	0.6519	0.6393
SentiPola	0.7298	0.7854	0.7566
AllDict	0.8070	0.8532	0.8295

从表中可以看出，在观点句识别二元分类任务中，随着情感词典覆盖率的提高和情感信息的丰富，情感特征向量的质量得以提升，因此情感观点句识别的准确率逐步提高。

对比实验中各种方法结果的趋势如图 2 所示。

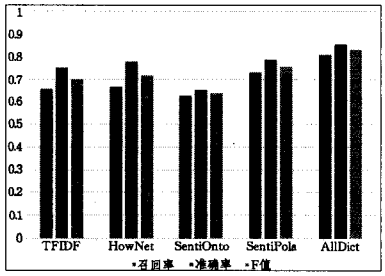

图 2 情感观点句识别实验结果

从图 2 可以看出，基于综合词典情感特征的情感观点句识别结果的召回率、准确率以及综合 F 值都取得了最好的效果，这也说明了词库情感特征对情感观点句识别的效果影响显著。

(2)情感倾向性判定结果

在观点句的情感倾向性判定实验中，同样对比了以上 4 种情感特征获取方法，实验结果如表 2 所列。

表 2 情感倾向性判定实验结果

情感特征	Precision	Recall	F-Measure
TFIDF	0.6652	0.7760	0.7163
HowNet	0.5263	0.5793	0.5515
SentiOnto	0.6384	0.7023	0.6688
SentiPola	0.7026	0.7500	0.7255
AllDict	0.7384	0.7717	0.7547

观点句的情感倾向性判断结果的柱状图如图 3 所示。

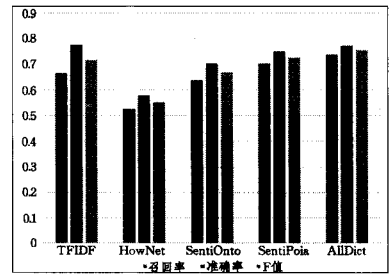

图 3 观点句情感倾向性判定实验结果

从观点句的情感倾向性判定实验结果可以看出，随着情感词库对情感词汇覆盖率的提高，以及对情感特征描述完整性的提升，情感倾向性判断实验结果的性能也得到逐步的提高；而且，利用本文所提的词汇情感确定性度量得到的词汇综合了情感词特征，结合 SVM 分类器取得了最好的情感倾向性分类效果。

6.2 结果分析

分析上述情感词库上获取的情感特征发现，虽然微博文本中还有大量丰富的情感信息，但由于《知网情感词库》、《情感词汇本体库》和《情感极性词表》中词汇都基本是书面用语词汇，因此仅在这些词库上获取的微博情感特征很不充分，甚至有很多微博句子上得到的特性向量为空。

利用本文所提方法综合上述词库，并与网络变异词融合后，获取的微博句子情感特征得到补充丰富。本文主要基于以下 3 个方面尽可能保证了微博句子特征向量的有效性：1)利用 TFIDF 特征保证了微博句子的全部关键词的情感特征都可以获取到；2)综合情感词库上的特征保证了情感词汇的特征有效性；3)互联网变异词库保证了对互联网上常用变异词汇情感特征的有效性。通过以上分析可以得出结论：情感词库的构建质量对情感分类性能的影响巨大，情感词库的覆盖率越高，词汇的情感倾向性和强度信息特征越充分，对情感分类的区别效用越强。

结束语 本文利用词汇情感倾向的确定性度量值，统一了词汇情感倾向性强度的度量方法，以此可以融合多种当前质量较好的情感词汇资源库。词汇情感倾向的确定性度量值的计算相对比较简单，在仅标注了情感倾向性的情感语料上简单统计计算即可容易得到。在面向微博本文的情感计算任务中，“观点句识别”与“观点句情感倾向性判定”的实验结果都充分验证了综合情感特征在情感计算中的有效性。

因此，完善和规范化情感词汇资源库将是一项长期性工作。而且，微博文本中所表述的情感信息不仅蕴含在词汇中，也蕴含在更深层次句法、语义中。例如，词汇间的依存搭配特征、词汇的概念以及概念相关的领域知识特征等，这些特征蕴含有更丰富的情感信息。如何利用更深层次情感语义信息来提高对微博文本的情感分类性能，将是我们下一步的研究内容。

(下转第 239 页)

质。为验证对不同规模数据的性能，分别向模式 S 的每个关系中注入 100、500、2000、10000 个元组（分别称为 D1/D2/D3/D4）分别进行对比测试。对比算法使用 COMA++算法，首先使用原始 COMA++算法获取匹配关系 *M1*，然后利用 COMA++与依赖冲突策略相结合获取匹配关系 *M2*；再分别使用 M1 和 *M2* 对 D1/D2/D3/D4 进行数据转换（采用文献[17]中的 Clio 算法）；最后使用文献[18]中方法分别对数据转化结果进行优化以得到核心数据转换方案。对不同转换结果进行优化的时间如表 3 所列。

表 3　优化时间对比

单位：秒	D1	D2	D3	D4
M1	56.7	217	864	3524
M2	12.4	38	146	563

从表 3 可知，在匹配结果选取时考虑其中的依赖冲突能够显著提高映射数据的质量，并缩短优化算法的执行时间。

结束语　本节提出了模式匹配结果中依赖冲突的新概念，给出了依赖冲突的定义以及冲突检测分类算法；最后提出一种对候选匹配结果中的依赖冲突进行分析以选取最终匹配结果的新策略，该策略能够在一定程度上提高匹配结果的准确率。实验结果表明：通过在已有匹配方法中结合新的匹配结果选取策略能够有效提高匹配结果准确率，同时数据映射结果的优化算法所耗费的时间会大幅减少。目前已有的模式匹配方法仅使用元素相互匹配的正面信息，在匹配准确率达到一定程度后再进一步提高会变得很困难，挖掘元素相互匹配的负面信息（证明元素不能匹配）不仅对进一步提高准确率有一定帮助，同时还可有效缩短匹配算法的执行时间（确定某些元素间不能匹配后可不计算其相似度），这将是我们进一步的研究方向。

参 考 文 献

[1] Rahm E, Bernstein P A. A Survey of approaches to automatic schema matching[J]. VLDB Journal, 2001, 10(4): 334-350

[2] Bernstein P A, Madhavan J, Rahm E. Generic schema matching, ten years later[J]. Proceedings of the VLDB Endowment, 2011, 4(11): 695-701

[3] Mecca G, Papotti P, Raunich S. Core schema mappings: Scalable core computations in data exchange[J]. Information Systems, 2012, 37(7): 677-711

[4] Calvanese D, De Giacomo G, Lenzerini M, et al. On simplification of schema mappings[J]. Journal of Computer and System Sciences, 2013, 79(6): 816-834

[5] Sorrentino S, Bergamaschi S, Gawinecki M, et al. Schema label normalization for improving schema matching[J]. Data & Knowledge Engineering, 2009, 69(12): 1254-1273

[6] Bilke A, Naumann F. Schema matching using duplicates[C]// Proceedings of 21st International Conference on Data Engineering. 2005: 69-80

[7] Elmeleegy H, Elmagarmid A, Lee J. Leveraging query logs for schema mapping generation in U-MAP[C]// Proceedings of the 2011 International Conference on Management of Data. Athens Greece, 2011: 121-132

[8] 李国徽，杜小坤，杨兵，等. 基于部分函数依赖的结构匹配方法[J]. 计算机学报，2010，33(2)：240-250

[9] Madhavan M J, Bernstein P A, Rahm E. Generic schema matching with cupid[C]// Proc. of VLDB. 2001: 49-58

[10] 申德荣，余恩运，张旭，等. SKM：一种基于模式结构和已有匹配知识的模式匹配模型[J]. 软件学报，2009，20(2)：327-338

[11] Elmeleegy H, Elmagarmid A, Lee J. Leveraging query logs for schema mapping generation in U-MAP[C]// Proceedings of the 2011 International Conference on Management of Data. Athens Greece, 2011: 121-132

[12] Pinkel C. Interactive Pay as You Go Relational-to-Ontology Mapping[M]// The Semantic Web-ISWC. 2013: 456-464

[13] Aumueller D, Do H H, Massmann S, et al. Schema and ontology matching with COMA++[C]// Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. Chicago, IL, USA, 2005: 906-908

[14] Peukert E, Eberius J, Rahm E. A self-configuring schema matching system[C]// Proceedings of 28st International Conference on Data Engineering. Washington DC, USA, 2012: 306-317

[15] Qian L, Cafarella M J, Jagadish H V. Sample-driven schema mapping[C]// Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. Scottsdale, USA, 2012: 73-84

[16] Zhang C J, Chen L, Jagadish H V, et al. Reducing uncertainty of schema matching via crowdsourcing[J]. Proceedings of the VLDB Endowment, 2013, 6(9): 757-768

[17] Popa L, Velegrakis Y, Hernández M A, et al. Translating web data[C]// Proceedings of the 28th international conference on Very Large Data Bases. VLDB Endowment, 2002: 598-609

[18] Fagin R, Kolaitis P G, Popa L. Data exchange: getting to the core[J]. ACM Transactions on Database Systems (TODS), 2005, 30(1): 174-210

[19] Alexe B, Hernández M, Popa L, et al. MapMerge: Correlating independent schema mappings[J]. Proceedings of the VLDB Endowment, 2010, 3(1/2): 81-92

（上接第 212 页）

参 考 文 献

[1] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts[C]// ACL'04. 2004

[2] Whitelaw C, Garg N, Argamon S. Using appraisal groups for sentiment analysis[C]// CIKM'05. NewYork, NY, USA, 2005: 625-631

[3] 中国计算机学会中文信息技术专业委员会 2013 年会评测[OL]. 2013-03-10[2013-09-30]. http://tcci.ccf.org.cn/conference/2013/pages/page04_eva.html

[4] 第五届中文倾向性分析评测（COAE2013）大纲[OL]. 2013-08-01[2013-09-30]. http://www.cipsc.org.cn/hytx/13.html#23

[5] 朱嫣岚，闵锦，周雅倩，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

[6] 董振东，董强.《知网》情感分析用词语集[OL]. http://www.keenage.com/html.com/html/c_bulletin_2007.htm

[7] 赵妍妍，秦兵，刘挺. 文本情感分析[J]. 软件学报，2010，21(8)：1834-1848

[8] 徐琳宏，林鸿飞，潘宇，等. 情感词汇本体的构造[J]. 情报学报，2008，27(2)：180-185

[9] 张华平. NLPIR 汉语分词系统[OL]. http://ictclas.nlpir.org