基于句法信息的微博情绪识别方法研究

黄 磊 李寿山 周国栋
(苏州大学计算机科学与技术学院 苏州 215006)

摘 要 情绪识别旨在自动识别文本是否含有情绪。情绪识别是情感分析研究中的一项基本任务。针对该任务,提出了一种基于句法信息的微博文本情绪识别方法。该方法的特色在于充分考虑了微博文本的句法信息。具体实现中,首先利用词性标注(POS)序列和结构句法树来表示句法信息,以分别提取POS序列模式、重写规则和二元句法标签作为特征进行文本表示;然后利用最大熵分类算法对微博文本进行情绪识别。实验结果表明,所提方法能够获得较好的识别效果。

关键词 自然语言处理,微博,情绪识别,POS序列模式,句法树
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.02.040

Emotion Recognition of Chinese Microblogs with Syntactic Information

HUANG Lei LI Shou-shan ZHOU Guo-dong
(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract Emotion recognition aims to predict the involving emotion towards a piece of text. Automatic emotion recognition is a basic task for sentiment analysis. In this paper, an emotion recognition for Chinese microblogs approach based on syntactic information was proposed. One distinguishing feature of the proposed method is that the microblog's syntactic information is employed. Specifically, we took advantage of POS (part of speech) sequence and syntactic tree to represent syntactic information in order to extract POS sequence pattern, rewrite rules and bigrams of syntactic labels as features for text representation. Then, we utilized the maximum entropy algorithm to perform the classification. Experimental studies demonstrate that our approach is very effective for emotion recognition.

Keywords Natural language processing, Microblog, Emotion recognition, POS sequence pattern, Syntax tree

1 引言

随着互联网的日益普及,尤其是Web2.0蓬勃发展,网络用户的参与度大大提高,网络信息的规模迅速扩大。微博(Microblog)作为典型的Web2.0互联网应用,自问世以来便受到了人们的喜爱,迅速发展为人们分享和获取信息的核心社交平台。越来越多的用户在以微博为代表的社交网络中表达自己的观点/情感和点评当前热点事件。在微博平台庞大的文本信息中,有很大一部分是带有情感的文本信息。处理和分析这些海量的带情感的数据信息具有巨大的价值。例如,情感分析可以应用于微博监管、突发事件预警、舆情发现、舆论引导和商业竞争情报分析等实际问题中。因此,对微博情绪分析的研究具有较高的理论价值和应用价值。

一般而言,情绪分析有两个基本任务[1]:情绪识别和情绪分类。情绪识别(Emotion Recognition)是指通过对文本进行分析,判断该文本是否含有情绪。下面例子中的文本来自于腾讯微博。

A:你一点也不关心、不在意我。(有情绪)

B:有什么事情请到微信留言。(无情绪)

*A*文本包含强烈的情感表达,含有情绪。*B*文本只是在客观地陈述,不含情绪。

情绪分类(Emotion Classification)是指在情绪识别的基础上对含有情绪的文本进行分析,进行文本情绪具体类别的判定,例如高兴、愤怒、恐惧等。其中,情绪识别是情绪分类的前提和基础。本文重点研究微博文本的情绪识别方法。

目前,基于机器学习的情绪识别方法通常使用词袋(bag-of-words)模型,选用词作为文本特征进行文本表示,这忽略了文本的语法、句法等要素。由于微博文本内容短小(一般不超过140个字符),其携带的信息相对较少,因此如何充分利用微博文本所提供的信息变得非常重要。一种可能扩充分类信息的策略是利用文本的句法信息。词性序列和句法树通常能够很好地表示文本的句法信息。相对于词的n-grams特征,词性序列与文本内容结合得更加紧密,可以表示更多的文本信息,而句法树含有丰富的句法信息。已有的研究表明,词性序列特征[2-3]、短语结构可以有效地用于文本的情绪识别[4]和情绪分类[5]。

因此,本文尝试利用句法信息的方法来帮助提高微博文本的情绪识别效果。句法信息通过Part of Speech (POS)序

到稿日期:2015-12-01 返修日期:2016-04-26 本文受国家自然科学基金(61331011,61375073,61273320)资助。

黄 磊(1989—),男,硕士生,主要研究方向为自然语言处理、情感分析,E-mail:20134227029@stu.suda.edu.cn;李寿山(1980—),男,教授,主要研究方向为自然语言处理、情感分析;周国栋(1967—),男,教授,主要研究方向为自然语言处理、信息挖掘、统计机器翻译等。

列和句法树表示，具体地，特征选取 POS 序列模式、重写规则和二元句法标签。实验结果表明，结合句法信息能够明显提高微博文本的情绪识别性能。

本文第 2 节介绍情绪分析的相关工作；第 3 节给出实验语料的描述；第 4 节介绍情绪识别方法；第 5 节介绍实验设置并对实验结果进行分析；最后给出结论，并对下一步工作进行展望。

2　相关工作

近年来，随着互联网的发展，文本情绪分析逐渐成为自然语言处理中的一项热点研究任务，受到了国内外研究人员的广泛关注。情绪分析的研究涉及多个方面：情绪资源建设[6-8]、作者情绪和读者情绪[9]、文档级的情绪分类[10]和句子级或短文本级情绪分类[11-13]。

Pang 等[14]首次将机器学习的方法应用到情感分析中，对电影评论进行情感极性的判别。Davidov 等[15]在 Twitter 上使用了 50 种 Twitter 标签和 15 种笑脸符作为情感标签，利用标点符号特征和词的 N-grams 特征对 Twitter 进行情感分类。Liu 等[16]通过新闻文本和评论文本之间的联系，研究了作者情绪和读者情绪的分类。Li 等[17]将微博文本中的词分为 3 种不同类型（情绪词、常用词、无情感或生僻词）并以此构建情感词典，用产生的情感向量来表示文本，进而对新浪微博进行文档级的情绪分类研究。卢伟胜等[3]利用词性标注序列作为文本特征，对腾讯微博文本进行情感分类研究。本文的研究集中在情绪识别方面。

刘欢欢等[4]在中文情绪语料库（Ren-CECps）上研究了句子级的情绪识别方法，并比较了不同特征（词、词＋词性、词＋词）对情绪识别的影响。Aman 等[1]使用基于情感词典的方法进行句子级的有无情绪的分类，准确率（accuracy）达到了 73.89%。Quan 等[18]使用中文情绪语料库（Ren-CECps）研究了基于情绪词的句子级情绪分析，获得了较好的识别效果。姚源林等[19]在 NLP&CC2013 中文微博情绪分析评测任务中介绍了关于微博情绪识别的相关结果。Maeda 等[20]使用 3 种写作风格方式，在 Twitter 语料库上进行情绪识别研究。张晶等[21]以情绪因子中常用的情绪词和情绪短语为基础构建情绪词典，通过情绪词典和情绪规则的匹配和计算实现对中文微博情绪的识别和分类。

以往研究大多采用基于情感词典的方式或基于词特征和词性特征的监督学习方式来进行情绪识别。然而，由于微博表达较为口语、网络用语较多，现有的情感词典往往很难适用；词特征、词性特征忽略了文本语言的多变性、情感的隐蔽性等特点。因此，在本文的微博文本的情绪识别任务中，结合词性标注序列和结构句法树来表示文本的句法信息，进一步提升了情绪识别性能。

3　语料收集与分析

到目前为止，与中文微博情绪分析相关的研究相对较少，相关的公共语料库比较缺乏。本文的语料来源于腾讯微博（http://t.qq.com），通过腾讯微博 API 接口收集。该语料采用基本情绪与复合情绪相结合的情绪分类体系进行标注。基本情绪包括快乐、愤怒、悲伤、恐惧；复合情绪定义为除基本情绪之外的情绪，分为正面复合情绪、中性复合情绪和反面复合情绪。此外，还设置了无情绪类别。在标注体系中，关注的是微博发布者的情绪状态。因此，根据腾讯微博 API 中 Type 字段值（不同的值代表不同的微博类型，如原创发表、转载、评论等）选择用户原创发表的微博进行语料标注。该语料库已完成 15540 条微博文本的情绪标注，其中 6548 条为有情绪微博文本，8992 条为无情绪微博文本。

在标注过程中，每条微博由两名标注者标注，标注结束后如果结果一致，一致的结果作为最终结果；如果标注不一致，第三名标注者参与标注，将获得一致的标注结果作为最终结果。本文中的有情绪类别包含基本情绪和复合情绪两大类。在微博文本情绪有无标注中得到的一致性（Kappa 值）约为 0.72，取得了较好的一致性效果。关于该语料更详细的信息，可以参考文献[22]。

4　情绪识别方法

传统的基于机器学习的情绪识别方法通常使用词作为文本特征进行文本表示，也就是 BOW（bag-of-words）模型。BOW 模型假定对于一个文档，忽略它的单词顺序和语法、句法等要素，将其仅仅看作是若干个词汇的集合。

本文提出的基于句法信息的微博文本情绪识别方法充分利用微博文本的句法信息，通过词性标注（POS）序列和结构句法树（Phrase Structure Parsing）来表示微博文本的句法信息。具体而言，首先在 POS n-grams 中挑选 POS 序列模式集作为特征，同时在句法树中选取重写规则（Rewrite-rule）和二元句法标签[23]（Syntactic-label Bigram）特征，然后根据这些特征将文本转化为文本特征向量，构建分类器，并对待分类样本进行分类。

4.1　基于 POS 序列的情绪识别

POS n-grams 能够捕捉文本的句法信息[23]。本文中的 POS 序列模式不同于传统的 POS n-grams。每个 POS 序列模式是一连串满足相应限制条件的词性标注序列（POS tags）。不同于使用所有的 n-grams 序列，我们要寻找能够真正表示有无文本情绪的序列模式，序列模式的长度是灵活的（不固定长度为 n）。POS 序列模式能够满足上述要求。限制条件包括用户指定的最小支持（Ninimum Support，Minsup）和最小依附（Minimum AdherenceMminadherence）约束。这些限制条件确保挖掘到的模式能代表真正表示有无文本情绪的规律。

算法的主要思路是执行分层搜索来寻找满足最小支持和最小依附的 POS 序列模式。模式的支持是指包含模式的文档所占的比例。如果一个模式出现的次数较少，那么它很可能是虚假的。一个序列若满足最小支持，它被称为频繁序列。模式的依附通过对称条件概率（Symmetrical Conditional Probability，SCP）权衡。对称条件概率最早由 Silver 提出，用于衡量任意字符串的成词概率，SCP 计算公式如下：

$$fairSCP(w_1 w_2 \cdots w_n) = \frac{f(w_1 w_2 \cdots w_n)^2}{\frac{1}{n-1}\sum_{i=1}^{n-1} f(w_1 \cdots w_i) f(w_{i+1} \cdots w_n)} \quad (1)$$

其中，$f(w_1\cdots w_i)$表示序列 $w_1\cdots w_i$ 出现的次数，n 为序列的长度。序列的 $fairSCP$ 值越高，表明该序列越重要。

为了方便介绍 POS 序列模式挖掘算法，在此先约定相关符号：设 $D=\{d_1,d_2,\cdots,d_n\}$表示微博文本的词性标注文档集合，n 表示语料中的微博条数；$T=\{t_1,t_2,\cdots,t_l\}$表示词性标注(POS tag)集合，l 表示词性标注个数。具体的算法过程如算法 1 所述。

算法 1 POS 序列模式挖掘算法

输入：数据集 D：由词性标注序列的文档组成，词性标注集合 T，用户指定的最小支持 minsup 和最小依附 minadherence

输出：所有满足最小支持和最小依附的 POS 序列模式(存储在 SP 中)

过程：

统计数据集 D 中每个标注 t∈T 出现的次数，获得 C_1

对于 f∈C_1，当满足$\frac{\text{f. count}}{n}\geqslant$minsup 时，获得 F_1

循环迭代 Max-length−1 次

S1)C_k=candidate-gen(F_{k-1})

S2)对于数据集 D 中每个文档 d，d∈D

S3)对于 C_k 中任意候选序列 c，c∈C_k

若 c 包含在文档 d 中

c. count++

S4)对于 c∈C_k，当满足$\frac{\text{f. count}}{n}\geqslant$minsup 时，获得 F_k

S5)对于 f∈F_k，当满足 fairSCP(f)≥minsup 时，获得 SP_k

最终得到 $SP=\bigcup_k SP_k$

函数 candidate-gen(F_{k-1})

$C_k=\emptyset$

对于 F_{k-1}集合中任意序列 c，c∈F_{k-1}

1)词性标注集合 T 中任意标注 t，t∈T

2)向序列 c 添加标注 t 作为后缀，获得 c′，并将 c′加入 C_k

4.2 基于句法结构分析的情绪识别

句法分析(Syntactic Parsing)是自然语言处理领域研究的关键问题之一，其基本任务是确定句子的句法结构(Syntactic Structure)或句子中词汇之间的依存关系。本文通过构造句法结构树，提取重写规则和二元句法标签作为特征，用于微博文本的情绪识别。

利用 Stanford Parser 句法分析工具对每个微博文本分析后可生成词性标注序列、上下文无关的短语结构树等。比如对句子“我们总是像智者一样劝慰别人，像傻子一样折磨自己。”进行句法解析后，生成的短语结构树如图 1 所示。

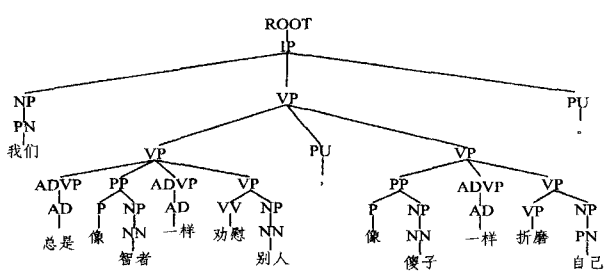

图 1 短语结构树

为方便理解重写规则和二元句法标签，下面将对其进行简要的介绍。重写(短语结构)规则在文中指的是非终结符节点 P(除了叶子节点和其直接父节点以外)产生其孩子节点 C，记作 P−>C。图 2 中含有的重写规则有：IP−>NP_VP_PU，VP−>ADVP_PP_ADVP_VP，VP−>VP_PU_VP 等。标签流[23]可以看作是句子句法结构的一种表示，图 2 是图 1 中例子的标签流。图 2 中含有的二元句法标签包括 IP-NP，NP-PN，PN-VP 等。

IP NP PN VP VP ADVP AD PP P NP ADVP AD VP VV NP
NN PU VP PP P NP NN ADVP AD VP VV NP PN PU

图 2 图 1 中短语结构树的标签流

根据上面的介绍，需要从短语结构树中提取重写规则和二元句法标签。具体的提取算法流程如下所示。

输入：数据集 D：由句法树文档组成，包含 n+个正类样本，n−个负类样本

输出：每个句法树文档的特征集合 T：由重写规则(rewrite-rules)和二元句法标签(syntactic-label bigrams)组成

过程：

对于数据集 D 中每个句法树文档 d，d∈D

a)构造文档 d 的句法树

b)先序遍历句法树

若节点不属于叶子节点

将节点标签加入到句法树标签流序列 labelSet

若节点是非终结符节点

将节点标签加入到非终结符节点集合 PSet

c)遍历标签流序列 labelSet，将所有二元句法标签加入特征集合 T_d

d)t=rewrite-rules(PSet)

e)将 t 加入到特征集合 T_d

f)得到文档 d 的特征集合 T_d 并将其输出

得到每个文档的短语结构树特征

函数 rewrite-rules(PSet)

对于 PSet 集合中的任意节点 c，c∈PSet

该节点的所有孩子节点形成重写规则 r

将 r 加入到重写规则集合 t

最终得到重写规则集合 t

5 实验结果与分析

5.1 实验设置

随机选择有、无情绪(正、负)样本各 5000 个作为实验语料。其中，训练样本为 8000 样例(正、负样本各 4000 个)，测试样本为 2000 样例(正、负样本各 1000 个)。为了获取词特征，使用复旦大学自然语言处理实验室开发的分词软件 FudanNLP 对文本进行分词操作。

分类算法采用最大熵分类方法(ME)，其中最大熵使用 MALLET 机器学习工具包[1]。在使用过程中，这些工具的所有参数都设置为它们的默认值。本文使用用户发表的微博状态文本，特征的选择来源于词、词性和句法 3 方面，如表 1 所

[1] http://mallet.cs.umass.edu

[2] http://nlp.stanford.edu/software/lex-parser.shtml

[3] http://nlp.stanford.edu/software/tagger.shtml

列。其中，句法树利用 Stanford Parser 句法分析工具[2]获取；而词性标注序列通过利用 Stanford Postagger 工具[3]获取；使用布尔权重作为文本向量中的特征值。具体地，特征使用 0 或 1 来表示，若特征在特征集合中出现，用 1 来表示；否则，用 0 表示。实验中，采用准确率(Precision)、召回率(Recall)和 F1 值(F1)作为分类结果的评价指标。

表 1　特征描述

特征	特征说明
Word_U	词特征(词的一元特征)
Word_U +POS_P	词特征+句法特征组合 1 (POS 序列模式特征)
Word_U+Syntactic_T	词特征+句法特征组合 2 (重写规则特征+二元句法标签特征)
Word_U +POS_P+ Syntactic_T	词特征+句法特征组合 3 (POS 序列模式特征+重写规则特征+二元句法标签特征)

5.2　实验结果及分析

为了方便进行实验比较，实现了下面几种监督学习方法。

• ME-WU：特征选用词特征，分类算法使用最大熵算法，并作为基准系统；

• ME-WU+PP：特征选用词特征+句法特征组合 1，分类算法使用最大熵算法；

• ME-WU+ST：特征选用词特征+句法特征组合 2，分类算法使用最大熵算法；

• ME-WU+PP+ST：特征选用词特征+句法特征组合 3，分类算法使用最大熵算法。

其中，特征的具体描述见表 1。

(1)不同特征的实验结果

图 3 和图 4 给出了在训练样本为数据集的 80%而剩余的 20%为测试样本的情况下采用不同特征的最大熵方法的情绪识别的分类结果。

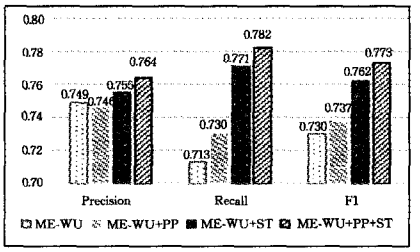

图 3　有情绪类别的分类结果

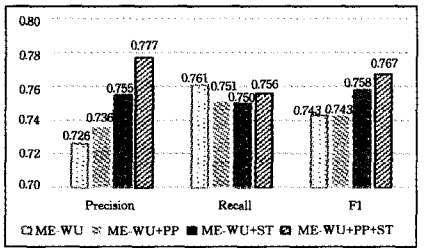

图 4　无情绪类别的分类结果

从图 3 可以看出，随着词性特征、句法树特征的加入，有情绪类别的分类 F1 值逐渐提高，且词特征、词性特征和句法树特征相结合时的分类性能最好，相对于仅使用词特征(基准系统)的 F1 值提高了大约 4%。从准确率(Precision)看，4 类特征识别有情绪类别时的分类性能相差在 1%～2%左右；从召回率(Recall)看，随着词性特征和句法树特征的加入，有情绪类别的分类性能得到了大幅度的提升，约为 7%。

从图 4 可以看出，在识别无情绪类别时，4 类特征的召回率大致相等(相 1%左右)。从准确率看，随着词性特征和句法树特征的加入，无情绪类别的分类性能逐渐提升；词特征、词性特征和句法树特征相结合时识别无情绪类别的 F1 值取得的效果最好，相对于仅使用词特征(基准系统)提高了约 2%。

对比图 3 和图 4 可以看出，随着词性特征、句法树特征的加入，在有情绪的识别结果中召回率(Recall)得到了大幅度的提升，在无情绪的识别结果中准确率(Precision)的提升幅度最为明显。这说明句法信息在识别微博文本是否含有情绪的任务上，能提高有情绪文本识别的召回率，同时帮助提升无情绪微博文本识别的准确率。此外，选取不同特征的组合方式在识别有情绪类别和无情绪类别上表现出了不同的性能(F1 值)：1)选取词特征的方式和选取词特征、词性特征相结合的方式识别无情绪类别的效果优于识别有情绪类别效果；2)选取词特征、句法树特征相结合的方式与选取词特征、词性特征和句法树特征相结合的方式在识别有情绪类别上表现更佳。

(2)不同训练样本大小的识别结果

分别选择实验数据集的 20%，40%，60%和 80%作为训练样本，剩下的 20%作为测试样本，比较不同特征在不同训练样本集下的情绪识别效果。其中，选取有情绪类别和无情绪类别 F1 值的平均值作为最终结果。

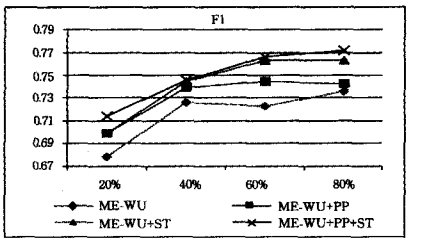

图 5　不同训练样本下不同特征对应的情绪识别结果

从图 5 可以看出，词特征、词性特征和句法树特征相结合的方式的分类效果最佳，且相对于仅使用词特征(基准系统)的分类效果大致提升了 2%～4%。

当训练样本为 20%时，词特征和词性特征相结合的分类效果与词特征和句法树特征相结合的分类结果基本一样。当训练样本为 40%，60%和 80%时，词特征和句法树特征相结合的分类结果都高于词特征和词性相结合的分类结果。随着训练样本的增加，不同特征对应的情绪识别结果都不断地提高。词特征与句法树特征相结合的分类结果明显优于词特征和词性特征相结合的方式的分类结果。图 5 的实验结果表明，本文提出的结合句法信息的情绪识别方法在不同规模的训练样本的情况下都能够取得较好的分类性能。

(3)参数敏感度实验结果

接下来探索 POS 序列模式挖掘算法中参数 *Max-length*，*minsup* 和 *minadherence* 对分类效果的影响。采用控制因素

法，即将多因素问题变成单因素问题，只改变其中的某一个因素，从而研究该因素对实验的影响，分别加以研究，最后综合解决。图 6 为各参数对应的 ME-WU+PP 方法的分类曲线。

从图 6 的结果可以看出，*Max-length* 值为 3～6 时，分类效果稳定。其他参数的实验中都设置 *Max-length* 为 3。对于 *minsup* 和 *minadherence*，当参数值逐渐增大时，初期分类性能呈上升趋势，但达到一定程度后开始下降。当参数 *minsup* 值设置为 0.3 左右且 *minadherence* 值设置为 0.4 左右时，分类效果较稳定。整体而言，参数在一定范围内，算法的分类性能差别不大（变化在 1%左右），这说明所提方法对参数不敏感。

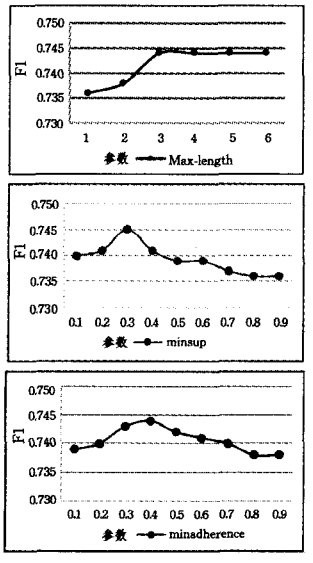

图 6 ME-WU+PP 方法的分类性能随参数的变化情况

(4)公开数据集上的实验结果

为了横向比较所提方法的有效性，在 NLP&CC2013 评测语料上进行同样的实验。随机选择有、无情绪样本各 5000 个作为实验语料，其中 80%的样本作为训练样本，剩余的 20%作为测试样本。其余实验设置与上述实验保持一致，实验结果如图 7、图 8 所示。

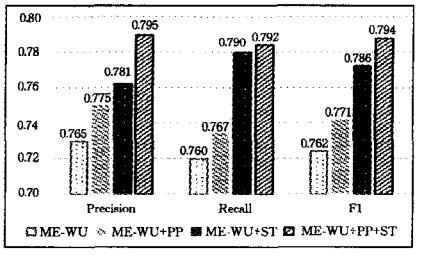

图 7 有情绪类别的分类结果

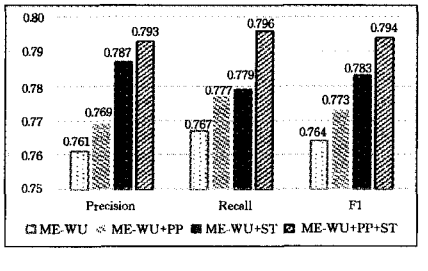

图 8 无情绪类别的分类结果

从图 7 和图 8 的实验结果可以看出，随着词性特征、句法树特征的加入，有情绪类别和无情绪类别的分类性能（F1 值）都在逐渐提高。词特征与句法树特征相结合的方式的分类性能优于词特征与词性特征相结合的方式的分类性能。词特征、词性特征和句法树特征相结合的方式的分类性能最好，相对于仅使用词特征（基准系统）的 F1 值提高了大约 3%。

通过对比图 3、图 4 和图 7、图 8 的实验结果可以看出，随着句法信息的加入，在不同实验语料上，本文提出的基于句法信息的情绪识别方法的分类性能的变化趋势一致，并且都能够提升情绪识别的性能。

结束语 本文提出了一种基于句法信息的情绪识别方法，其利用句法信息提高情绪识别的性能。句法信息分别通过词性标注序列和结构短语句法树来表示。具体实现中，选取 POS 序列模式作为词性特征，选取重写规则和二元句法标签作为句法树特征。实验结果表明，基于 POS 序列和基于结构句法树特征都能够有效提高情绪识别性能。当结合这两种句法信息后，情绪识别性能达到最佳。

在下一步的工作中，将尝试使用最新的深度学习分类方法[24]并结合本文提出的特征来提升情绪识别性能。此外，在情绪识别的基础上，将进行情绪分类研究，具体考察句法特征对情绪分类任务的影响。

参 考 文 献

[1] AMAN S, SZPAKOWICZM S. Identifying Expressions of Emotion in Text [C]// Proceedings of the 10th International Conference (TSD 2007). 2007: 196-205.

[2] LU W S, GUO G D, CHEN L F. Emotion Classification with Feature Extraction Based on Part of Speech Tagging Sequences in Microblog [J]. Journal of Computer Applications, 2014, 34 (10): 2869-2873. (in Chinese)
卢伟胜，郭躬德，陈黎飞. 基于词性标注序列特征提取的微博情感分类 [J]. 计算机应用，2014，34(10)：2869-2873.

[3] MUKHERJEE A, LIU B. Improving Gender Classification of Blog Authors [C]// Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2010). East Stroudsburg, United States, 2010: 207-217.

[4] LIU H H, LI S S, ZHOU D G, et al. Research on Chinese Emotion Recognition [J]. Journal of Jiangxi Normal University (Natural Science Edition), 2013, 37(2): 120-124. (in Chinese)
刘欢欢，李寿山，周国栋，等. 中文情绪识别方法研究 [J]. 江西师范大学学报，2013，37(2)：120-124.

[5] LIU Q, FENG C, HUANG H. Emotional Tendency Identification for Micro-blog Topics Based on Multiple Characteristics [C]// Proceedings of the 26th Pacific Asia Conference on Language, Information and Computation (PACLIC 2012). Bali, Indonesia, 2012: 207-217.

[6] WIEBE J, WILSON T, CARDIE C. Annotating Expressions of Opinions and Emotions in Language [J]. Language Resources and Evaluation, 2005, 39: 65-210.

[7] QUAN C, REN F. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis [C]// Proceedings of

Empirical Methods in Natural Language Processing (EMNLP 2009). East Stroudsburg, United States, 2009: 1446-1454.

[8] XU J, XU R, LU Q, et al. Coarse-to-fine Sentence-level Emotion Classification Based on the Intra-sentence Features and Sentential Context [C]// Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012). United States, 2012: 2455-2458.

[9] LIN K, YANG C, CHEN H. Emotion Classification of Online News Articles from the Reader's Perspective [C]// Proceedings of IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops (WI-IAT 2008). Sydney, NSW, Australia, 2008: 220-226.

[10] ALM C, ROTH D, SPROAT R. Emotions from Text: Machine Learning for Text-based Emotion Prediction [C]// Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2009). East Stroudsburg, United States, 2005: 579-586.

[11] TOKUHISA R, INJI K, MATSUMOTO Y. Emotion Classification Using Massive Examples Extracted from the Web [C]// Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008). 2008: 881-888.

[12] BHOWMICK P, BASU A, MITRA P, et al. Multi-label Text Classification Approach for Sentence Level News Emotion Analysis[C]// Pattern Recognition and Machine Intelligence. Lecture Notes in Computer Science, Berlin, Germany, 2009: 261-266.

[13] LI S, HUANG L, WANG R, et al. Sentence-level Emotion Classification with Label and Context Dependence [C]// Proceedings of 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL 2015). Beijing, China, 2015: 1045-1053.

[14] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment Classification Using Machine Learning Techniques [C]// Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2010). 2002: 79-86.

[15] DAVIDOV D, TSUR O, RAPPOPORT A. Enhanced Sentiment Learning Using Twitter Hastags and Smileys [C]// Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008). 2010: 241-249.

[16] LIU H, LI S, ZHOU G, et al. Joint Modeling of News Reader's and Comment Writer's Emotions [C]// Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). Sofia, Bulgaria, 2013: 511-515.

[17] LI C, WU H, JIN Q. Emotion Classification of Chinese Microblog Text via Fusion of BoW and eVector Feature Representations [C]// Proceedings of the 3rd CCF Conference on Natural Language Processing and Chinese Computing (NLP&CC 2014). 2014: 217-228.

[18] QUAN C, REN F. Sentence Emotion Analysis and Recognition Based on Emotion Words Using Ren-CECps [J]. International Journal of Advanced Intelligence, 2010, 2(1): 105-117.

[19] YAO Y L, WANG S W, XU R F, et al. The Construction of an Emotion Annotated Corpus on Microblog Text [J]. Journal of Chinese Information Processing, 2014, 28(5): 83-91. (in Chinese)
姚源林，王树伟，徐睿峰，等. 面向微博文本的情绪标注语料库构建 [J]. 中文信息学报，2014，28(5)：83-91.

[20] MAEDA H, SHIMADA K, ENDO K. Twitter Sentiment Analysis Based on Writing Style [C]// Proceedings of the 8th International Conference on NLP (NLP 2012). Kanazawa, Japan, 2012: 278-288.

[21] ZHANG J, ZHU B, LIANG L L, et al. Recognition and Classification of Emotions in the Chinese Microblog Based on Emotional Factor [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2014, 50(1): 79-84. (in Chinese)
张晶，朱波，梁琳琳，等. 基于情绪因子的中文微博情绪识别与分类 [J]. 北京大学学报(自然科学版)，2014，50(1)：79-84.

[22] HUANG L, LI S, ZHOU G. Emotion Corpus Construction on Microblog Text [C]// Proceedings of the 16th Workshop on Chinese Lexical Semantics Workshop (CLSW 2015). Beijing, China, 2015: 204-212.

[23] HIRST G, FEIGUINA O. Bigrams of Syntactic Labels for Authorship Discrimination of Short Texts [J]. Literary & Linguistic Computing, 2007, 22(4): 405-417.

[24] CHEN F, CHAO W H, ZHOU Q, et al. Convolution Tree Kernel Based Sentiment Element Recognition Approach for Chinese Microblog [J]. Computer Science, 2014, 41(12): 133-137, 142. (in Chinese)
陈锋，巢文涵，周庆，等. 基于卷积树核的中文微博情感要素识别 [J]. 计算机科学，2014，41(12)：133-137，142.

(上接第 238 页)

[11] GARZA P, QUINTARELLI E, RABOSIO E, et al. Reducing Big Data by Means of Context-Aware Tailoring[M]// New Trends in Databased and Information Systems. Springer International Publishing, 2016: 115-127.

[12] OH K J, KIM Z, OH H, et al. Travel intention-based attraction network for recommending travel destinations[C]// 2016 International Conference on Big Data and Smart Computing (BigComp). IEEE, 2016: 277-280.

[13] ZHANG J W, YANG Z. Collaborative Filtering Recommendation Algorithm Based on Improved User Clustering[J]. Computer Science, 2014, 41(12): 176-178. (in Chinese)
张峻玮，杨洲. 一种基于改进的层次聚类的协同过滤用户推荐算法研究[J]. 计算机科学：2014，41(12)：176-178.

[14] YU X S, SHAN H. Research on Personalized Recommedation Model Based on Network Users' Information Behavior [J]. Journal of Chongqing Institute Technology, 2013, 27(1): 47-50. (in Chinese)
余肖生，孙珊. 基于网络用户信息行为的个性化推荐模型[J]. 重庆理工大学这报(自然科学版)，2013，27(1)：47-50.