

基于情感-主题-讽刺混合模型的讽刺检测研究

付月, 史伟

引用本文

付月, 史伟. 基于情感-主题-讽刺混合模型的讽刺检测研究[J]. 计算机科学, 2023, 50(6A): 220300018-6.

FU Yue, SHI We. [Study on Satire Detection Based on Sentiment-Topic-Satire Hybrid Mode](#)[J].

Computer Science, 2023, 50(6A): 220300018-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于prompt和知识增强的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on Prompt and Knowledge Enhancement

计算机科学, 2023, 50(6A): 220300279-7. <https://doi.org/10.11896/jsjcx.220300279>

[方面级情感分析综述](#)

Summarization of Aspect-level Sentiment Analysis

计算机科学, 2023, 50(6A): 220400077-7. <https://doi.org/10.11896/jsjcx.220400077>

[基于多级多尺度特征提取的CNN-BiLSTM模型的中文情感分析](#)

Chinese Sentiment Analysis Based on CNN-BiLSTM Model of Multi-level and Multi-scale Feature Extraction

计算机科学, 2023, 50(5): 248-254. <https://doi.org/10.11896/jsjcx.220400069>

[基于多事件语义增强的情感分析](#)

Sentiment Analysis Based on Multi-event Semantic Enhancement

计算机科学, 2023, 50(5): 238-247. <https://doi.org/10.11896/jsjcx.220400256>

[基于情感知识的双通道图卷积网络的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on Dual-channel Graph Convolutional Network with Sentiment Knowledge

计算机科学, 2023, 50(5): 230-237. <https://doi.org/10.11896/jsjcx.220300008>

基于情感-主题-讽刺混合模型的讽刺检测研究

付月¹ 史伟²

1 湖州学院经济管理学院 浙江 湖州 313000

2 浙江海洋大学经济与管理学院 浙江 舟山 316022

(286824081@qq.com)

摘要 讽刺检测是观点挖掘的一个子任务,主要目的是识别用户在书面文本中表达的观点或情感。文本中讽刺句往往具有混合的情感极性,正确识别讽刺句和非讽刺句在情感分析中起着至关重要的作用。讽刺检测方法一般都采用机器学习分类器,其中分类器的训练主要基于简单的词汇或基于词典的特征。本研究的目的是建立一个无监督的概率关系模型,根据微博中词语的情感分布来识别讽刺主题。模型基于主题级分布估计相关情感,评估出现在短文本中的情感相关词,给出情感相关标签。实验结果表明,该模型在讽刺检测方面优于其他最新的基线模型,非常适合于短文本的讽刺预测。

关键词: 讽刺;情感分析;观点挖掘;主题模型

中图法分类号 TP391.1

Study on Satire Detection Based on Sentiment-Topic-Satire Hybrid Model

FU Yue¹ and SHI We²

1 School of Economics and Management, Huzhou College, Huzhou, Zhejiang 313000, China

2 School of Economics and Management, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, China

Abstract Satire detection is a subtask of opinion mining. Its main purpose is to identify the opinion or emotions expressed by users in written texts. Satirical sentences in texts often have mixed sentiment polarity. Correctly identifying satirical sentences and non satirical sentences plays an important role in sentiment analysis. Various satire detection methods are based on machine learning classifiers, in which the training of classifiers is mainly based on simple words or dictionary features. The purpose of this study is to establish an unsupervised probabilistic relationship model to identify satirical themes according to the sentiment distribution of words in microblog. The model estimates the related sentiment based on the topic level distribution, evaluates the sentiment related words appearing in the short text, and gives the sentiment related labels. Experimental results show that the model is superior to other latest baseline models in satire detection, and is very suitable for satire prediction of short text.

Keywords Satire, Sentiment analysis, Opinion mining, Topic model

1 引言

社交媒体已经成为沟通的重要手段,人们通过这种媒介来表达自己的观点、意见、情绪、想法、态度等。这些想法通常发布在基于物联网的智能设备上。观点挖掘和情感分析是用来分析大量在线文本数据的有效手段,其中自动检测讽刺表达将是一个重要的挑战,因为人们在互联网上用讽刺方式表达的观点,人和机器都很难理解。

百度百科对“讽刺”的定义为“用比喻、夸张等手法对人或事进行揭露、批评”。讽刺在更深的意义上是高度相关的语言和常识,讽刺是一种情感,人们总是倾向于在文本中使用正面或强化的正面词语来表达负面感受或不喜歡。在文本数据中发现讽刺表达是一件非常有意思的事情,并且很难被普通人识别,这引起了研究者在社交媒体文本中特别是在微博中发现反讽表达的浓厚兴趣。

讽刺检测是观点挖掘的一个子任务,它的主要目的是识别用户在文本表达中的观点或情感。正确识别讽刺句和非讽刺句在情感分析中起着至关重要的作用,讽刺句往往具有混合的情感极性,比如句子“我喜欢被忽视!”,其中“喜欢”表示正面情感,而“忽视”则是一个负面情感词。很少有夸张讽刺句存在(只有肯定词没有否定词),比如“他的样子真是太棒了!”。因此需要一种新的能够检测讽刺表达水平和讽刺流行主题的方法。理解讽刺表达对人类和机器都是一个相当困难和具有挑战的任务。

本研究的主要目的是基于短文本中的情感分布确定讽刺的流行主题,一定程度上有助于讽刺表达的检测。模型背后的重要思想是:1)短文本或微博文本中的一些主题比其他主题更倾向于讽刺;2)与纯粹的正面或负面微博文本相比,讽刺表达微博文本中正面和负面词的分布完全不同。提出的情感-主题-讽刺模型的架构如图1所示,其中预处理将微博

基金项目:国家社会科学基金一般项目(20BXW013)

This work was supported by General Program of National Social Science Foundation of China(20BXW013).

通信作者:史伟(shiwei@zjhu.edu.cn)

文本或评论引入到情感讽刺模型中,基于情感本体,该模型纯粹通过得分来学习词语的分布。这个模型可以发现讽刺的流行主题,其次是正面和负面主题,也可以清楚地估计主题的概率分布。

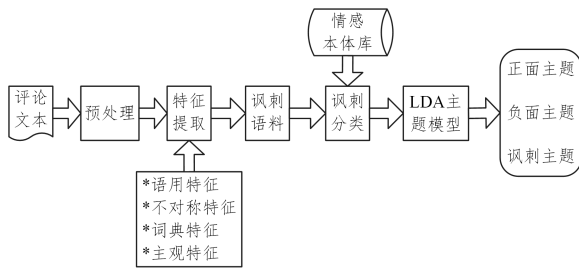


图1 情感-主题-讽刺模型的结构

Fig. 1 Structure of sentiment-topic-satire model

情感-主题-讽刺模型认为评论文本具有3种情感标签:正面的,负面的,讽刺的。模型使用隐藏变量,如主题变量、情感变量以及一个转换变量来识别讽刺流行主题。提出的情感-主题-讽刺混合模型能够识别数据集中特定主题下的词语。模型评价包括定性和定量两个方面,定性评价以情感相关词为基础,定量评价包括精度、准确率、召回率和F值等指标。

本文第2节是相关文献综述;第3节阐明运用情感主题模型进行讽刺检测的动机;第4节描述了设计原理、符号和模型生产过程;第5节分说明了实验设置和用于测试的数据集模型;第6节从定性和定量两方面分析了情感-主题-讽刺混合模型进行讽刺检测的结果;最后总结全文并展望未来。

2 相关工作

在过去几年里,自然语言处理领域研究者比较关注社交媒体的情感分析。最近的一些文章主要采用机器学习或深度学习技术的方法进行情感分类,然而分类技术和特征提取在很大程度上取决于结果。也有一些学者开始应用社交化短文本中的一系列特征因素发现社交媒体中的讽刺表达,并将短文本分为两类:讽刺和非讽刺。讽刺是一种比喻性语言,字面意思表达的信息并不是它真实的观点。对于新闻标题数据集,检测系统检测文本或主题是讽刺或不是,所选特征的重要性在文献[1]中做了评价。Chun-Che等在短文本中增强了一种用于讽刺检测的机器学习算法,他们通过应用一系列特征比如 Unigrams、bigrams、主题建模等证明了系统的精确性[2]。Ashraf等提出了一种在Twitter上检测自嘲讽刺(SDS)的深度学习方法,并在7个Twitter数据集上进行了实验,结果发现提出的方法明显优于许多基于神经网络的基线和最先进的方法[3]。Rajadesingan等通过探索用户的行为特征进行讽刺的检测,这些行为特征通常通过用户过去的对话获取,构建行为模型框架并评估模型的效率[4]。Yogesh等介绍了情感分析、机器分类和深度学习在识别讽刺推文中的应用,并使用各种评估指标比较了各种方法的使用结果,从结果得到讽刺检测可以极大地提高情感分析模型的性能[5]。

Zhang等提出了一种在使用少量标注训练数据的情况下,应用对抗学习框架来提升深度学习模型在讽刺识别任务中性能的方法,该方法具有一定的优越性,但存在模型训练

不稳定等问题[6]。Han等提出了一种结合用户嵌入、论坛主题嵌入与内容编码的上下文语境讽刺检测模型,该模型运用 ParagraphVector 模型与 Bi-GRU 模型分别编码上下文语境信息和内容信息,有效提高了讽刺检测分类准确率,但由于用户文档与主题文档的训练样本较少,实验过程中会出现过拟合、分类错误等问题[7]。Valdivia等提出了一种新的自匹配网络,通过分析词与词之间的交互来捕获句子的不一致信息,该工作吸收了句子的成分信息,以便更好地检测讽刺[8]。Joshi等进行了一项关于对话中讽刺发现的新研究,实验表明在每个连续场景中都会发现讽刺表达,两种序列标记算法的性能优于分类算法[9]。Silvio等利用了与词汇信号一致的用户嵌入来识别讽刺,他们的模型利用了一套普通但精致的特征来识别讽刺表达[10]。Wang等利用上下文信息自动检测 twitter 中的讽刺表达,利用马尔可夫支持向量机将标签分配给 tweets 的整个序列的类别中,实验结果表明序列分类法能有效地利用上下文信息检测讽刺[11]。Hernandez等在构建的模型中考虑了结构特征及情感特征,如 tweet 的整体情感和极性得分等,用于区分讽刺和非讽刺的 tweet[12]。

Nimala等讨论了在 twitter 数据集上进行主题建模,基于聚合策略的哈希标签的重要性和性能[13]。Rajadesingan等讨论了利用用户的行为特征来检测推特中讽刺的可能性,并提出了一个计算行为模型,该模型包含了用户个人信息特征[14]。Weitzel等提出了一个独立于反讽检测领域的无监督框架,包括单词嵌入以获得领域感知词语的反讽倾向,实验结果表明将主题反讽模型与词语嵌入相结合,在真实场景中取得了很好的效果[15]。Tao等提出了一种新的自匹配网络,通过分析词与词之间的相互作用来捕获句子中的不一致信息,这项工作吸收了句子的成分信息,以便更好地检测讽刺[16]。

对讽刺表达识别的研究刚刚展开,虽取得了一定的进展,但多数是在英语语境中,对语义表达更为复杂的中文语境中的讽刺检测还缺乏比较有效的识别手段,并且比较少从情感主题的角度去分析讽刺表达的存在规律和形式。情感主题模型主要是在大规模语料库中发现有情感倾向的主题结构,本文使用情感主题模型检测讽刺表达的背后驱动力是识别是否存在讽刺流行主题,并捕捉讽刺和非讽刺文本的情感分布,主要观点是很少存在这样的主题——它们比其他主题更能自动引发讽刺。

3 模型提出

图2描绘了构建的情感主题讽刺模型的符号图,表1列出了相应的符号和缩写。

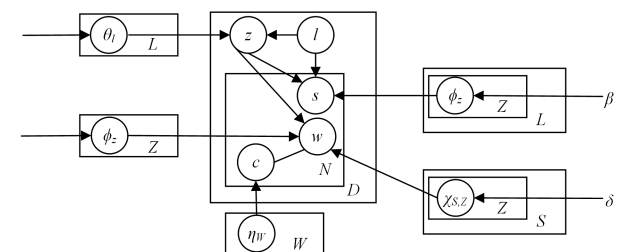


图2 情感主题讽刺模型的符号图

Fig. 2 Symbol diagram of sentiment-topic-satire model

表1 模型使用的符号

Table 1 Symbols used in the model

符号	含义
W	单条微博上的的词语数
L	微博标签,正面、负面和讽刺
C	转换变量(在情感词或主题词间转换的值)
Z	微博主题
S	微博中一个词语的情感(正面、负面)
η_w	转换变量的分布
θ_l	给定标签的主题分布
ϕ_z	给定主题 z 的词语分布
$\lambda_{s,z}$	给定情感 s 和主题 z 的分布
$\Psi_{z,l}$	给定主题 z 和标签 l 的情感分布

假设语料库由讽刺微博构成,并且这些微博按照位置进行收集。对于这个模型,用 l 来表示评论的标签,包括正面的、负面的和讽刺的; c 是转换变量,分别表示用户情感或主题词。模型使用 z 表示主题, s 为词语的情感, η_w 为转换变量的分布, $\lambda_{s,z}$ 表示给定情感和主题的分布, $\Psi_{z,l}$ 为给定主题 z 和标签 l 的情感分布。

4 生成过程

给定文档 D 和具有超参数 α 和 β 的主题数量,以及情感标签 l ,算法输出 D 文档中基于讽刺流行主题的情感。

算法1 STSM

1. /在情感标签 l 下画出主题分布/
2. For every label l
3. $\theta_l \sim \text{Dir}(\alpha)$
4. End For
5. /对于情感标签 l 中的每个主题画出词语分布/
6. For each topic z in l
7. $\Psi_{z,l} \sim \text{Dir}(\beta)$
8. End For
9. /画出主题 z 和情感 s 的分布/
10. For each topic z in s
11. $\lambda_{s,z} \sim \text{Dir}(\delta)$
12. End For
13. /画出主题 z 下的词语分布/
14. For each topic z
15. $\phi_z \sim \text{Dir}(\gamma)$
16. End For
17. /画出指定句子的分布/
18. For each sentence, do
19. $Z_k \sim \theta_{l,k}$
20. Switch between values for all words $C_{k,j} \sim n_j$
21. Sentiment for all sentiment words, $S_{k,j} \sim \Psi_{z,k}, l_k$
22. All topic words $w_{k,j} \sim \phi_{z,k}$
23. All sentiment words $w_{k,j} \sim \Psi_{z,l}, Z_k$
24. End For

5 实验

微博是一个非常流行的在线社交平台,微博用户在上面分享从主流话题(比如饮食、音乐、电影、商品、政治等)到私密爱好的各种信息。截至2021年3月,中国的微博月活跃用户达到5.3亿,移动端占比94%,日活跃用户达到2.3亿,微博已成为展现公众情感和观点的重要平台。微博数据的提取主要有3种方式:1)采用网络爬虫抓取;2)通过API获取;

3)通过微博平台高级搜索获取。本文主要采用第三种方式获取微博信息,这些信息都是公开可免费获取的。微博的主要特点之一就是“即时分享”,用户通过微博平台将自己的观点和情感分享给听众,所以在研究过程中提取他人公开的微博信息是完全合乎国家相关法律许可和基本道德规范的,不存在侵犯他人隐私的问题。

以#讽刺#、#嘲讽#、#嘲笑#、#挖苦#、#讥讽#等为主题收集微博近一周的16000条讽刺微博文本和32000条非讽刺微博(未含有#讽刺#等标签),并将其存储在数据库中。非讽刺类微博进一步被分为正面和负面。基于已构建的情感本体^[17],将含有“期待”“高兴”“喜爱”和“惊讶”等情感类词语的微博设为正面标签,将含有“焦虑”“悲伤”“生气”“讨厌”等情感类词语的微博设为负面标签。

按照以下步骤对微博进行预处理:

- (1)对在空白边界上的个别词进行分离。
- (2)从微博文本中去除所有非文字的数字字符,例如逗号、破折号等。
- (3)去除1208个标准停用词,包括常见的一些动词形式。
- (4)删除一些不相关的微博信息,从微博中过滤掉额外的链接,如含有“http:”或者“www.”的表达,以及各种符号标签和用户的名字(用符号@标志的)。

(5)移除“回复”“转发微博”等词和转发的内容(只是转发没有增加任何评论的帖子)。

(6)去除重复的微博,将微博文本分成一个个单句,而后进行情感词标记、基本词性标注和讽刺表达识别。

最后,数据集中8632条微博为正面情感,1978条微博为负面情感,7027条微博有讽刺表达。数据集的2.5%用于测试,其余用于模型训练。本文基于带#标签的微博展开研究,分类如下: $L=3$,正面的、负面的和讽刺的;情感 $S=2$,正面的和负面的;主题 Z 设置为10。这里使用折叠吉布斯抽样来估计分布,并根据它们的联合概率分布来寻找隐藏参数或潜在变量的值。特征提取是从数据集中提取各种特征,使机器学习算法能够正常工作。模型中使用的主要特征有:语义特征、基于不一致性的特征、词汇特征和主观特征等。

(1)语义特征。语义特征是那些基于语句的实际应用,而不是理论知识。有多种类型的语义特征正在生成,以便对模型进行训练。

1)表情符号:表情符号是以脸部表情的形式在文本中表示各种情感。微博中有不同种类的表情符号用于表示人类的不同情感,如表2所列。表情符号可以体现人们在发表评论时所要传达的语气,因此可以用于讽刺表达的检测。Python中的编辑码器模型可以用来阅读表情符号。

表2 人工建立的微博表情符号字典

Table 2 Manually created microblog sentiment dictionary

情感倾向	表情符号
肯定	
否定	

2)标点符号:标点符号对读者来说传递着各种不同的情感

信号。比如感叹号(!)会增加正面或负面情感的强度,同样其他标点符号如“。”“?”“……”等也都表达着不同的语义含义。

(2)基于不一致性的特征。这个特征主要基于如下理论:每一句讽刺表达基本都是在负面情景下正面情感的崩溃。比如对于句子“我非常开心能在星期天工作”,在这个句子中“我非常开心”是正面情感与负面情景“在星期天工作”形成了对比,因此这就形成了一种讽刺的表达。这个特征在模型中的使用如下所示。

1)情感不一致:正面情感词后接负面情感词的次数,反之亦然。

2)最大子序列:表示文本块中正面或负面情感的最大子序列的次数。

3)情感极性计数:表示极性情感词出现的次数,基于已经构建的情感本体^[17],将“期待”“高兴”“喜爱”和“惊讶”等情绪类词语和正面(G)类评价词设为正面情感词,将“焦虑”“悲伤”“生气”“讨厌”等情绪类词语和负面(B)类评价词设为负面情感词。

(3)词汇特征。Unigrams 被用于提取微博文本中基于词汇特征的信息。这种方法的一个扩展是使用 N-grams 模型,它可以表示讽刺。比如,“是的,没错”是一个表示讽刺存在的语句。

(4)主观特征。这是在对话语境或文本中表达个人主张的特征。个人主张包括观点、情感、评价和推测等。本文以知网提供的 38 个中文主张词语(见表 3)作为判断主观特征的标准^[18],比如主观性句子“我觉得这部电影很有意思”,这个句子就包含主张词语“觉得”。

表 3 中文主张词语
Table 3 Chinese assertion words

感知	察觉 触目 耳闻 发觉 发现 风闻 感觉 感觉 感觉到 感受到 见到 见得 觉得 看得出来 窥见 领教 听说 痛感 预感 自觉
认为	抱定 当道 感到 感觉 觉得 看看 待论 认定 认为 认准 相信 以为 主张

6 评价结果

模型的评价分为定性和定量两部分。通常用定性的方法表示从情感主题讽刺模型中提取的主题;而定量方法则讨论定量的测量,比如发现主题情感标签的概率分布,模型的查全率、查准率和 F 值,还有模型与其他讽刺检测方法的比较等。

6.1 定性评价

定性评价的目的是展现由情感主题讽刺模型提取的主题。这项工作分为两个步骤进行探讨,第一步,对该模型发现的主体进行评估,然后对整个语料库进行测量。因为讽刺微博的数据集被输入到模型中,所以生成的主题是讽刺流行主题。在后一步中,联合情感-主题分布模型捕捉讽刺的存在。该模型既能评估主题,又能评估情感词。表 4 列出了仅针对讽刺性微博的综合主题和情感相关主题。标题是为主题手动指定的,下划线词语是携带主题信息的词语,这些信息在表 5 和表 6 中分别被列出,其中包含每个讽刺流行主题的情感主题。仔细观察可以发现,生成的词语具有相反或混合的情感极性。比如,关于天气的讽刺微博“我的头发不湿的时候看起来非常漂亮”“是的,但是现在天气很潮湿”。表 4—表 6 讨论了何时将讽刺微博输入到模型中。

表 4 综合主题和情感相关主题的评估(针对讽刺微博)

Table 4 Evaluation of comprehensive and sentiment related topics (for satirical microblog)

喜爱	工作	天气	聚会	食物
亲爱的	太棒了	超级棒	欢乐	好吃的
感受	成果	气候带	活动	饮料
快乐	行动	气候	狂欢	饮食
喜欢	放弃	雨水	乐队	零食
宠儿	表现	哇	气氛	热爱
宝贝	经典	不好	诱人的	快餐
天使	兼职	真实	厌恶	太棒了
甜蜜	完成	今天	搭档	兴奋
心肝	讨厌	高兴	哈哈	早餐
烦恼	激增	糟糕	夜晚	菜单
欣赏	可怜	下雪	围攻	食材
反感	早晨	气象	无聊	无味

表 5 讽刺微博中由模型评估的主题

Table 5 Topics evaluated by the model

喜爱	工作	天气	聚会	食物
亲爱的	成果	气候	活动	饮料
感受	行动	雨水	乐队	零食
喜欢	表现	今天	气氛	快餐
天使	兼职	下雪	搭档	早餐
心肝	早晨	气象	夜晚	食材

表 6 讽刺微博中由模型学习得到的情感-主题

Table 6 Sentiment — topic obtained from model learning in satirical microblog

喜爱	工作	天气	聚会	食物
快乐	太棒了	超级棒	欢乐	好吃的
宠儿	放弃	糟糕	狂欢	饮食
宝贝	痛苦	哇	诱人的	热爱
情侣	成就	美好的	聚集	太棒了
厌恶	实现	幸存	讨厌	低劣的
痛苦	可怜	忍受	大笑	无味
反感	汗水	高兴的	有趣的	菜单

表 7—表 9 列出了当完整语料库作为模型的输入时,主题、情感相关主题和讽刺流行主题中的词语分布。这些表格中的主题将清楚地区分为讽刺流行主题或基于情感的主题。表格中列出的是语料库中发现的前 5 个主题词,包含微博层面的情感标签:正面,负面和讽刺。与前面的例子一样,表 7 显示了从语料库中评估的综合主题和情感相关主题,并且所有主题的标题都是手动标记的。比如,其中一个主题是“健康”,该主题下最热门的 5 个主题词是“健身”“运动”“早晨”“健康”和“跑步”。

表 7 综合主题和情感相关主题的评估(针对完整语料库)

Table 7 Evaluation of comprehensive and sentiment related topics (for the complete corpus)

健康	工作	音乐	食品	名言
健身	晚上	摇滚	食物	语录
健康	早晨	流行音乐	快餐	黎明
早晨	杰出的	古典的	蛋糕	夜晚
运动	工作	本地的	早餐	渗透
有趣的	讨厌	乡村	健康	激励
享受	厌烦	披头士乐队	喜爱	旅程
跑步	干涸	激情	喜欢	积极的
疲倦	有趣的	热爱	完美	感动
好的	恶心的	快乐	美味的	朗朗上口
开心	喜欢	笑	开心	愉快
有病的	睡眠	糟糕的	蔬菜	坏的
可怜的	更好	可怕的	无味的	极好的

表8 完整语料库中由模型评估的主题

Table 8 Topics evaluated by the model in the complete corpus

健康	工作	音乐	食品	名言
健身	晚上	摇滚	食物	语录
健康	早晨	流行音乐	快餐	黎明
早晨	工作	古典的	蛋糕	夜晚
运动	干活	本地的	早餐	渗透
跑步	睡眠	乡村		激励
		披头士乐队		旅程

表9 完整语料库中由模型学习得到的情感-主题

Table 9 Sentiment topic obtained from model learning in the complete corpus

健康	工作	音乐	食品	名言
有趣的	杰出的	激情	健康	积极的
享受	讨厌	热爱	喜爱	感动
疲倦的	无聊的	悲痛	讨厌	朗朗上口
好的	压力	笑	完美	掩饰
愉悦	有病	糟糕的	美味的	较少的
不舒服	喜欢	坏的	贫乏的	极度的

6.2 定量评价

定量评价就是通过对主题子集的概率值理解,分析针对一特定主题用户交谈内容的情感标签。表10显示,正面情感最高的是爱情(0.92)、音乐(0.93)、天气(0.86)和聚会(0.83);负面情感概率值较高的是食物(0.89);讽刺流行主题是学校(0.85)、工作(0.85)等。图3表示微博标签的正面情感词分布,这里X轴表示在一条微博中正面情感词所占的百分比,Y轴则是对应的该类微博在数据库中的百分比。图3明确显示负面微博包含的正面情感词较少,而正面微博包含的正面情感词较多,讽刺性微博中正面情感词的比例高于负面情感词。

表10 主题情感标签的概率

Table 10 Probability of topic sentiment label

主题	正面	负面	讽刺
爱情	0.92	0.05	0.03
工作	0.08	0.07	0.85
天气	0.86	0.08	0.06
聚会	0.83	0.11	0.06
食物	0.06	0.89	0.05
学校	0.07	0.08	0.85
音乐	0.93	0.05	0.02
名言	0.86	0.05	0.09

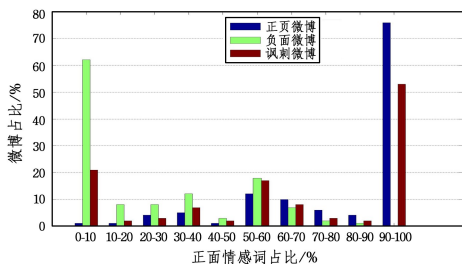


图3 微博标签的正面情感词分布

Fig. 3 Distribution of positive sentiment words in microblog Tags

本文通过SVM分类器进行讽刺检测,能够将微博分为讽刺和非讽刺,通过关键性能指标(KPI)来评估,比如精度、召回率和F值。如表11所列,所提出的情感主题讽刺模型具有更高的精度和召回率,并且具有更好的F值,表明该模型相比其他基线模型具有较好的性能。精度、召回率

和F值的计算公式如下:

$$precision = \frac{a}{a+b} \quad (1)$$

$$recall = \frac{a}{a+c} \quad (2)$$

$$F_{\beta=1} = \frac{(1+\beta^2) \times precision \times recall}{\beta^2 \times precision + recall} = \frac{2a}{2a+b+c} \quad (3)$$

参考混合矩阵, a, b, c 分别对应正确进行分类的具有讽刺表达的微博数、进行分类但不具有讽刺表达的微博数及具有讽刺表达但未进行分类的微博数。

表11 本文模型在讽刺检测中与其他方法的比较

Table 11 Comparison of this model with other methods in satire detection

方法	精度	召回率	F值
Topic irony model	0.75	0.64	0.69
TIM+WE	0.82	0.77	0.79
Hierarchical topic model	0.77	0.75	0.76
本文模型	0.85	0.82	0.83

结束语 情感主题讽刺模型是一种可以有效检测讽刺相关主题的先进模型。本文提出的主题模型使用了包含正面、负面和讽刺的微博数据集,并评估了与讽刺流行主题相关的词语分布。该模型获得了讽刺流行主题,如学校(0.85)和工作(0.85)。同时,模型学习得到的词语分布清楚地区分了讽刺流行主题和相应主题中的词语(包括具有混合极性的词语)。该方法还通过运用一定程度的逻辑推理来识别多种事件的讽刺。构建的模型工作有效,并且非常适合各种讽刺检测的应用。但提出的模型主要依赖于词袋,将来可能会用 bigrams 和 trigrams 进一步扩展,因为大多数情况下,讽刺总是以词组的形式用含蓄的情感表达出来。本文的研究工作涉及到用于讽刺检测的短文本无监督情感和主题分析。随着深度学习技术的发展,利用深度学习的弱监督表示可以有效地检测社交化短文本中的讽刺表达,这将会是我们今后研究的一个方向。

参考文献

- [1] BOUAZIZI M, OHTSUKI T. Sarcasm Detection in Twitter: "All products are Incredibly amazing !!!"-Are they really? [C]//IEEE Global Communications Conference. Keio University, Japan, 2014.
- [2] CHUN-CHE P, MOHAMMAD L, JAN W P. Detecting sarcasm in text: an obvious solution to a trivial problem[R]. Stanford CS 229 Machine Learning, 2015.
- [3] ASHRAF K, MUHAMMAD A. CATBiGRU: Convolution and Attention with BiDirectional Gated Recurrent Unit for SelfDeprecating Sarcasm Detection[J]. Cognitive Computation, 2021, 1: 78-97.
- [4] RAJADESINGAN R, ZAFARANI H L. Sarcasm detection on Twitter: a behavioural modelling approach[C]// Proceedings of 18th ACM International Conference on Web Search Data Mining. 2015: 9-106.
- [5] YOGESH K, NIKITA G. AI-Based Learning Techniques for Sarcasm Detection of Social Media Tweets: State-of-the-Art Survey[J]. SN Computer Science, 2020, 1(318): 20-34.
- [6] ZHANG Q L, DU J C, XU R F. Sarcasm Detection Based on Adversarial Learning [J]. Acta Scientiarum Naturalium Universita-

tis Pekinensis, 2019, 55(1):29-36.

- [7] HAN H, ZHAO Q T, SUN T Y, et al. Contextual Sarcasm Detection Model for Social Media Comments[J]. Computer Engineering, 2021, 47(1):66-71.
- [8] VALDIVIA A, MARTÍNEZ-CÁMARA E, CHATURVEDI I, et al. What do people think about this monument? Understanding negative reviews via deep learning, clustering and descriptive rules[J]. J Ambient Intell Human Comput, 2020, 11:39-52.
- [9] JOSHI A, VAIBHAV T, PUSHPAK B, et al. Harnessing sequence labeling for sarcasm detection in dialogue from tv series friends[C]// CoNLL. Berlin, Germany; ACL's Special Interest Group on Natural Language Learning, 2016:146-155.
- [10] SILVIO A B C, WALLACE H L, PAULA CARVALHO M J S. Modelling context with user embeddings for sarcasm detection in social media[C]// CoNLL. Berlin, Germany; ACL's Special Interest Group on Natural Language Learning, 2016:167-179.
- [11] WANG Z L, WU Z J, REN Y F, et al. Twitter sarcasm detection exploiting a context based model[C]// Web Information Systems Engineering-WISE. Springer, Miami, FL, USA, 2015:77-91.
- [12] HERNANDEZ-FARIAS, BENED J, ROSSO P. Applying basic features from sentiment analysis for automatic irony detection [C]// Pattern Recognition and Image Analysis. Springer, USA, 2015:337-344.
- [13] NIMALA K, JEBAKUMAR A. A robust user sentiment biterm topic mixture model based on user aggregation strategy to avoid data sparsity for short text [J]. J Med Syst, 2019, 43(93):32-48.
- [14] RAJADESINGAN A, ZAFARANI R, LIU H. Sarcasm detection on twitter: a behavioral modeling approach[C]// Proceedings of the 8th ACM International Conference on Web Search and Data Mining. ACM Press, 2015:97-106.
- [15] WEITZEL L, PRATI R C, AGUIAR R F. The comprehension of figurative language: what is the influence of irony and sarcasm on NLP techniques? [M]// Sentiment Analysis and Ontology Engineering. New York; Springer, 2016:49-74.
- [16] XIONG T, ZHANG P, ZHU H, et al. Sarcasm Detection with Self-matching Networks and Low-rank Bilinear Pooling[C]// Proceeding of WWW '19. Raleigh North Carolina, USA: ACM, 2019:2115-2124.
- [17] SHI W, WANG H W, HE S Y. Study on construction of fuzzy Emotion ontology Based on HowNet[J]. Journal of the China Society for Scientific and Technical Information, 2012(6):595-602.
- [18] HowNet [OL]. <http://www.keenage.com>.



FU Yue, born in 1983, master, lecturer. Her main research interests include network public opinion and text mining.



SHI Wei, born in 1981, Ph.D, professor. His main research interests include business intelligence and affective computing.