中文微博的去抑制化现象研究*

王国梁，昝红英
郑州大学信息工程学院 郑州 450001
E-mail:iamwgliang@gmail.com，iehyzan@zzu.edu.cn

摘要：近年来，随着微博、朋友圈以及 QQ 空间的出现和快速发展，人们越来越多的在网络上发布自己的情绪和心情，随之而来的是海量的大众情感数据。本文通过基于词典、朴素贝叶斯和最大熵等多种情感分析理论的实验来研究微博话题下的情感数据，分析人们在微博中表现的情感特征分布和强度，再通过数据分析的方法来验证社会心理学中的去抑制化现象。实验表明，使用情感分析理论来研究社会心理学现象是可行的，中文微博中存在去抑制化现象。

关键字：中文微博 情感分析 去抑制化

Research on the Disinhibition of Chinese Micro-blog

WANG Guoliang, ZAN Hongying
School of Information Engineering, Zhengzhou University,Zhengzhou Henan 450001,China
E-mail:iamwgliang@gmail.com，iehyzan@zzu.edu.cn

Abstract: In recent years, with the emergence and rapid development of micro-blog, circle of friends and QQ space, more and more people release their emotions and feelings on the Internet, followed by the mass of the public sentiment data. Through theoretical and experimental analysis based on a variety of emotional dictionary, Naive Bayes and Maximum Entropy to study the microblogging topic of emotional data, people in the micro-blog the affective characteristics of the distribution and intensity analysis, based on the data analysis method to verify the social psychology to disinhibition phenomenon. Experiments show that it is feasible to use the theory of affective analysis to study the phenomenon of social psychology and micro-blog in the presence of disinhibition phenomenon.

Keywords: Micro-blog Chinese sentiment analysis Disinhibition

1 引言

随着互联网爆发性的发展，微博用户的规模也越来越大，《第 36 次中国互联网络发展状况统计报告》显示，截至 2015 年 6 月，我国微博客用户规模为 2.04 亿，网民微博使用率为 30.6%[1]。在近年来的社会心理与行为学研究中，

* 本文承国家社会科学基金项目（14BYY096）、国家自然科学基金项目（61402419）、国家高技术研究发展 863 计划（2012AA011101）、国家重点基础研究发展计划 973 课题（2014CB340504）支持资助。

互联网心理学研究也越来越受到关注，通过分析互联网数据可以更加方便的研究社会大众的心理。在互联网中，由于网络中的匿名性、言论自由性以及互不可见性和权威弱化性，人们更加愿意表达出自己的心情变化和行为态度，微博成了人们情感和情绪的交汇点和集中区，通过微博我们可以直接获取到社会大众的心情、情绪、行为以及对某个事件的情感倾向等。所以，研究人们在网络上的情感和行为特征对于研究社会心理具有重大的意义。

在微博的研究中，我们发现一种现象，即人们在微博上表现的情绪和行为往往比现实更加极端，例如出现的争吵、过分追捧、过分褒扬甚至谩骂等在一定程度上都被放大和增强了。也即在网络的虚拟环境中，基于个体的内心准则和社会规范的制约而形成的行为的自我克制大大削弱或不符存在，从而人们的网上行为表现出一种解除抑制的状态，心理学称之为去抑制化现象。本文主要介绍如何通过情感分析的方法来挖掘中文微博中人们的情感信息，验证人们在网络中是否更倾向于表现为极端性的行为和言论，探索使用情感分析的方法代替传统问卷调查来研究社会心理学现象。

2 相关研究

2.1 去抑制化现象

去抑制化(Disinhibition)现象分为两个方面：良性的（Benign Disinhibition)，也即在网上会做出更多的亲善行为，如对英雄的褒扬等等；劣性的（Toxic Disinhibition)，也即一些攻击性言论等等。但有时候这两种行为还会混杂在一起出现。

Starr Roxanne Hiltz[2]等在 1989 年首次提到了去抑制化(Disinhibition)现象，后来 JOHN SULER[3]对这种现象进行归因并作出了理论解释；实证研究也正在一步步完善这一理论，比如 2011 Jeffrey G. Snodgrass[4]等的一篇实证研究就证明了对于负性的去抑制行为，缺乏目光接触（Eye contact）比匿名性（Dissociative anonymity）发挥着更重要的作用；Lapidot-Lefler 等[5]则从匿名性、隐蔽性、缺乏目光接触这三个方面通过对比试验来研究对去抑制化现象的影响，并最终得出缺乏目光接触是网络去抑制现象的主要因素。朱韩兵[6]分析了去抑制化行为的积极和消极意义及其应对策略。

2.2 情感分析理论

情感分析是对带有情感色彩的主观性文本、处理、归纳和推理的过程。情感分析的方法主要有两类，基于词典的情感分析和基于机器学习的情感分析。

英文微博的研究学者大多都采用基于 SVM 的距离向量监督学习算法、基于 KNN 的语料强化学习算法、基于语义的关联分析法以及基于情感词的语义标注等方法来对 Twitter 进行情感分类研究[7-9]。由于中文微博的多语义性和复杂性，英文的研究方法并不完全适用于中文的情感分析当中[10]。如谢丽星等[11]的一种基于 SVM 的层次结构多策略中文微博情感分类方法。还有刘志明[12]综合多种算法

构建分类模型对微博正负情感分类进行研究。韩忠明等[13]以 How Net 情感词典为基础,构建了计算短文本情感倾向性的自动机。此外还有研究者结合大量人工标注的情感词典数据加上句法分析有效的提高了情感分析的准确性[14]。

基于词典的情感分析主要通过对语义词典进行情感的规则性研究，找到情感判断的一定规律性。如文献[15]通过定义态度词典、权重词典、否定词典、程度词典以及感叹词词典来计算每条微博的情感指数。基于机器学习的情感分析方法主要是应用机器学习模型，通过对训练集的特征进行学习，构造模型，从而应用于对测试集的分类判断。如文献[16]使用三种机器学习算法、三种特征选取算法以及三种特征项权重计算方法对微博进行了情感分类的实证研究。

3 情感分析模型与去抑制化现象研究

为了增加数据分析结果的准确性，本文使用了三种情感分析模型：第一种是我们实现的基于词典的情感分析模型(Dictionary Model,DM)；第二种是基于朴素贝叶斯的情感分析模型(Naïve Bayes model,NB)；第三种是基于最大熵的情感分析模型(Maximum Entropy model,ME)。

3.1 基于词典的情感分析模型

对于微博数据由于 140 字的编辑限制，内容往往比较精简，同时也要求发布者在有限的文本内容中明确表达出自己的观点或者情感倾向,情感倾向也较为单一。同时，我们会发现上述文本中的否定词、程度词和感叹词在一定程度上可以让我们更加准确的区分积极和消极情感倾向，例如上述文本中的“始终”、“不”、“好”和“!”等。所以我们考虑到了程度词和感叹词对情感的增强，否定词对情感的极性转换，以及补充的新词和网络词提高了词典的完整度。

本模型采用的词典集合:

表 1 基于词典的情感分析模型词典集合

词典类型	示例	规模（条）
情感词典	愉快、忧伤、愤怒	33556
程度词词典	很、比较、相当	125
否定词词典	不、没、未曾	45
感叹词词典	啊、多么、何等	64
网络新词词典	666、坑爹、屌丝	241

情感词典我们结合了台湾大学情感词典（NTUSD）和知网 2007 年发布的情感分析用词语集（beta 版）并在此基础上增加了我们自己发现的新词，其他词典是我们自己维护的词典集合。

3.2 基于朴素贝叶斯的情感分析模型

本文中该模型主要包括训练过程和测试过程。训练过程中，首先载入已经分

类的语料包括积极语料和消极语料经过分句、分词和生成向量最终保存为语料数据模型。在测试过程中，对于任意的语料同样经过分句、分词和生成向量再和训练库中的数据对比，最终计算出特征项和类别的联合概率选出最优分类结果。模型如图 1 所示：

图 1 朴素贝叶斯模型示例图

3.3 基于最大熵的情感分析模型

我们使用了优化器并在学习率上做了优化，学习率决定了每次优化参数时参数变化的增量大小，学习率过小会导致更长的收敛时间，学习率过大可能会导致震荡不收敛甚至是发散到无穷大。模型如图 2 所示：

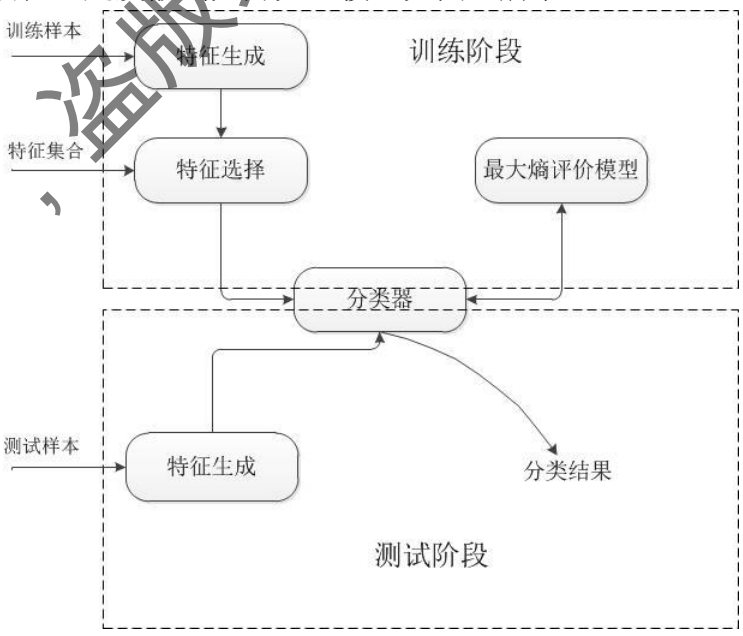

图 2 最大熵模型结构图

3.4 中文微博的去抑制化现象研究

在本实验中，情感分析模型不同于一般的二类情感分析模型，我们在二类情感分析模型基础上采用打分的策略，例如-1、3、-7、10。分值的绝对值越大情感倾向也就越强烈。本质上我们的情感分析模型是一个变形的二类情感分析模型，但在去抑制化现象的研究过程中，我们将这个二类情感分析模型的分值范围划分为三类极性情感，即中性、良性、劣性。对于无任何情感倾向和极性倾向不明显的语料我们都划分为中性，对于有明显情感倾向的正向情感与负向情感，我们划分为良性和劣性。

情感分析分值与去抑制化现象极性对应如表 2 所示：

表 2 情感分析分值与去抑制化现象极性对应表

情感得分 Q	情感极性 J
大于 2	良性
大于等于-2 且小于等于 2	中性
小于-2	劣性

示例语料情感分析值与两种极性类别对应如表 3 所示：

表 3 示例语料情感分析值与两种极性类别对应表

微博语料	情感分析得分	情感模型极性	去抑制化现象极性
笑的太开心啦哈哈，祝小李同志再接再厉！后天一定成功!!狗狗也加油哈哈哈	6	积极	良性
赢了一盘	1	积极	中性
终于输了！再不输，有点反人类啊！	-4	消极	劣性
阿发狗：卧槽！我不能太嚣张，低调低调。否则天网计划容易嗝屁[笑 cry]/(ㄒ o ㄒ)/	-7	消极	劣性
李世石要输啊	-1	消极	中性

4 实验及结果分析

4.1 情感分析

4.1.1 数据集描述

情感分析模型实验数据我们采用第六届中文倾向性分析评测（The sixth Chinese Opinion Analysis Evaluation,COAE2014）任务 4 的提供的微博数据，该数据集共 40000 条微博数据，但官方只公布了 5000 条微博数据的极性，在此我们使用标注了极性的数据,其中包括积极情感语料 2656 条,消极情感语料 2344 条。

表 4 COAE2014 数据样例

积极	消极
三星 N7108，好大台，功能好强劲	我手机也不好用了，闹钟不响，读卡器那也读不出手机卡，三星啊，你让我们情何以堪
浦发银行办事效率真高，赞赞	浦发银行的信用卡一刷 ，卖保险的马上给我打电话。拒绝了，也没有，持续
苹果手机就是比其他手机好啊	手机坏了
太平洋保险的赔付还是很给力的，上午 报险在 20 分内就到现场定损，现在已经将赔付款打到帐上，	车千万别 出 事故保险理赔很麻烦，一会还要拉着旧件去 4S 店

4.1.2 评价标准

和 COAE2014 评测[17]一样，我们同样采用正确率、召回率和 F-测度值来评价分类器的性能。

准确率:

$$P = (\frac{RP}{RP+WP} + \frac{WN}{WN+RN})/2 \quad (1)$$

召回率：

$$R = (\frac{RP}{RP+RN} + \frac{WN}{WN+WP})/2 \quad (2)$$

F-测度值：

$$F = \frac{2*P*R}{P+R} \quad (3)$$

其中：针对预测结果正类预测为正类的个数(RP)，正类预测为负类的个数

(RN)；针对预测结果负类预测为负类的个数(WN)，负类预测为正类的个数(WP)；F-测度值，即为准确率和召回率的调和平均值。

微平均：微平均即积极情感和消极情感的平均值。

4.1.3 情感分析模型对比实验

本文采用三种情感分析模型进行对比实验，其中基于词典的情感分析模型实验，在我们的词典集合基础上利用我们设计的权值计算算法，得出最终的情感值；基于 NB 的情感分析模型实验，分为训练集和测试集，首先载入训练集，模型学习特征，语料库采用人民日报 1998 年中文标注语料库，词性标注使用 3-gram 模型，特征值使用 TF-IDF 方法计算，文本相似使用 BM25 计算；基于最大熵的情感分析模型实验，首先从原始语料选取特征对比特征库得出 lib-svm 格式的数据格式文件，最终生成矩阵向量并分为训练集和测试集。

4.1.4 实验结果

根据实验数据我们得出的情感分析模型性能如下：

表 5 三种情感分析模型评测结果统计表

分类器	Pos_P(%)	Neg_P(%)	Micro_P(%)	Micro_R(%)	Micro_F(%)
NB	77.5	80.1	78.7	78.8	66.0
DM	71.3	69.0	70.2	70.3	61.9
ME	97.1	92.3	94.9	95.0	73.5
Best	97.1	92.3	94.9	95.0	73.5
Medians	81.9	80.5	81.3	81.4	67.1

基于词典的模型可以达到 70.22%的正确率；基于 NB 的模型可以达到 78.77%的准确率；基于最大熵的模型可以达到 94.88%的正确率。其中基于最大熵的情感分析模型性能最好。

COAE2014 评测结果最优与平均成绩：

表 6 COAE2014 评测结果最优与平均成绩表

COAE2014	Pos_P(%)	Neg_P(%)	Micro_P(%)	Micro_R(%)	Micro_F(%)
Best	97.7	97.1	96.2	54.7	68.1
Medians	89.1	85.0	85.7	30.5	45.0

从表 5 和表 6 我们可以看出，本文的模型正确率方面和 COAE2014 评测结果相比略有偏低但差别不大，本文模型的优势是在召回率上远远超过 COAE2014 评测结果，在 F 值方面也有一定的优势。

评测实验一是为了检验我们模型的性能，二是为了优化和训练我们的模型，调试模型参数达到最优的性能，为我们的模型在应用的情景中能够最大限度的发挥模型的优势。

4.2 情感分析在去抑制化现象中的应用

4.2.1 数据集描述

我们选取的是那些没有明显情感倾向的微博数据，例如科技新闻、明星八卦、个人发表的对某个事件态度的微博等等没有明显极性的微博数据作为我们的实验数据。这里，我们选取了当前具有代表性的事件，例如“AlphaGo 李世石围棋大战”和“Papi 酱天价广告拍卖”。这些事件客观上并没有明显的极性，对于大众的舆论倾向也是未知的，符合我们的实验对象数据要求。

4.2.2 实验过程

爬虫从网络中抓取了原始微博数据，我们关注的属性类型包括话题、微博、微博附带数据(赞、转发、来源)及微博评论数据等。针对于网页上的微博数据如图 3 所示：

图 3 微博数据截图

我们关注的微博属性如表 7 所示：

表 7 微博数据属性表

数据类别	数据值
微博内容	理论上电脑总是能胜过人脑。但如何实现，这个难度不是一般的大。因此。电脑围棋胜了人类。表明了人工智能算法上取得了重大突破。很多人根本就不了解这些。
赞数	1
回复数	0
发布日期	03 月 13 日
数据来源	华为 Ascend Note7
数据发布者	邵路军（实际抓取的是用户 ID）

情感分析的主题数据是微博内容的分析，包含常规情感语句、情感词、网络词以及特殊符号等等，赞数和回复数也作为补充数据，数据来源和发布者等作为附属属性，可以方便未来去抑制化网络的研究。

本实验中，基于本文上述的情感分析模型，对于每一个话题的全部数据，我们首先预处理爬虫抓取到的微博数据，经过垃圾过滤并生产不同模型需要的向量，利用情感分析模型分析语料数据得出情感值，最终统计出数据分布，并得出去抑制化现象的概率。

4.2.3 实验结果

“AlphaGo 李世石围棋大战”结果：

表 8“AlphaGo 李世石围棋大战”去抑制化现象实验结果

分类器	预测正极性(条)	预测负极性(条)	预测中性(条)	极性语料/总语料(%)
NB	3978	92	2174	65.18
DM	3775	236	2233	64.24
ME	3645	153	2404	61.24

“Papi 酱天价广告拍卖”结果：

表 9“Papi 酱天价广告拍卖” 去抑制化现象实验结果

分类器	预测正极性(条)	预测负极性(条)	预测中性(条)	极性语料/总语料(%)
NB	1227	613	984	67.55
DM	1423	354	1067	65.23
ME	1365	508	1020	69.63

从表 8、表 9 的实验结果我们可以看到，大众对话题“AlphaGo 李世石围棋大战”和“Papi 酱天价广告拍卖”的情感倾向分布客观上应该是中性的，但在我们的结果分布中却有着 61%以上的极性情感。可得出，对于我们随机选出的微博话题，大众的情感确实表现为去抑制化现象。

5 结论

本文使用三种情感分析模型研究去抑制化现象在中文微博中的存在性与其分布特征。由实验结果可见，在微博中去抑制化现象是明显存在的，并且更倾向于表现良性的去抑制化现象，即人们更倾向于表现积极的行为和情绪。此外，本文利用自然语言处理的方法对去抑制化这一社会心理学现象的研究也表明了计算机和互联网技术应用在社会心理学研究领域的可行性和有效性。当然，我们所做的工作还存在一些不足，例如没有针对去抑制化现象的细节进行挖掘，下一步我们将对去抑制化现象的性别、年龄、时空、职业分布等角度细化分析，并构建去抑制化现象关系网络。

参考文献，

[1] 中国互联网络信息中心.《第 36 次中国互联网络发展状况统计报告》[EB/OL].https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201507/t20150722_52624.htm.2015-07-22

[2] Starr Roxanne Hiltz, Murray Turoff, Kenneth Johnson: Experiments in group decision making, 3: disinhibition, deindividuation, and group process in pen name and real name computer conferences. Decision Support Systems 5(2): 217-232 (1989)

[3] Suler, J. (2004). The online disinhibition effect. Cyberpsychology & behavior, 7(3), 321-326.

[4] Jeffrey G. Snodgrass, Michael G. Lacy, H.J. Francois Dengah II, Jesse Fagan

(2011). Enhancing one life rather than living two: Playing MMOs with offline friends, 27(2), 1211 - 1222.
[5] Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. Computers in Human Behavior, 28(2), 434-443.
[6] 朱韩兵等. 网络行为的去抑制化分析综述. 宿州学院学报，2008, 3(6):Vol. 23, No.3
[7] Jansen B J, Zhang Mimi, Sobel K, Chowdury A. Twitter Power:Tweets as ElectronicWord of Mouth. 2009,60(11): 2169-2188
[8] Luciano Barbosa Junlan Feng Robust Sentiment Detection on Twitter from Biased and Noisy Data. Proceedings of the 23rd International Computational Linguistics.Beijing. Tsinghua University Press. 2010:36-44
[9] Dmitry Davidov,Oren Tsur, Ari Rappoport.Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. Proceedings of the 23rd International Computational Linguistics.Beijing. Tsinghua University Press. 2010:241-249
[10]周胜臣,瞿文婷,石英子等.中文微博情感分析研究综述.计算机应用与软件, 2013, 30(3):161- 164, 181
[11] 谢丽星, 周明,孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取.中文信息学报.2012, 26(1): 73-83
[12]刘志明,刘鲁.基于机器学习的中文微博情感分类实证研究.计算机工程与应用, 2012, 48(1): 1-4
[13] 韩忠明,张玉沙,张慧,等.有效的中文微博短文本倾向性分类算法.计算机 应 用 与 软 件 ,2012, 29(10): 89-93
[14] 梁军,柴玉梅,原慧斌,等.基于深度学习的微博情感分析[J].中文信息学报,2014,28(5):155-161
[15] Shen Yang,Li Shuchen,Zheng Ling,et al. Emotion Mining Reasearchon Micro-blog [C] / / Web Society 2009. SWS’09. 1st IEEE Symposium，2009
[16] 刘志明,刘鲁.基于机器学习的中文微博情感分类实证研究[J].计算机工程与应用，2012，48(1) : 1-4.
[17] http://www.liip.cn/ccir2014/pc.html.第二十届全国信息检索学术会议.第六届中文倾向性分析评测（The sixth Chinese Opinion Analysis Evaluation, 简称 COAE2014）[EB/OL].2014-02