

基于微博文本的个性化兴趣关注点 及情绪变迁趋势研究

王九硕¹,高 凯¹,赵 捷²,高国江¹

(1. 河北科技大学信息科学与工程学院, 河北石家庄 050018; 2. 国家保密科技测评中心(河北省)分中心, 河北石家庄 050000)

摘 要:随着社交网络的快速发展,人们可利用微博平台发表、分享自己的观点以及抒发某种情绪,进而产生了大量针对不同话题的博文和情绪信息,但传统的文本挖掘算法在处理这些短小且富含个性化情感信息的微博文本方面有所欠缺。在此提出一种基于微博文本的特征权重计算方法,可据此得到博主在不同时间段的关注点,通过情绪分类,分析用户在不同时间段内的情绪变迁情况。实验结果证明此方法具有一定的可行性。

关键词:自然语言处理;微博;个性化建模;个性化兴趣关注点;情绪变迁

中图分类号:TP391 **文献标志码:**A

Research on micro-blog based personal interests and emotion transition tendency

WANG Jiushuo¹, GAO Kai¹, ZHAO Jie², GAO Guojiang¹

(1. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China; 2. National Secrecy Science and Technology Evaluation Center (Hebei Branch), Shijiazhuang, Hebei 050000, China)

Abstract: With the rapid development of social network, people can present or share their opinions and emotions on the micro-blog, and as a result, large amount of blog texts and emotional information of different topics are produced. As the traditional text mining algorithms cannot deal with these blog texts with fractional contents and different personal emotions effectively, this paper presents a novel weighting computation approach based on the key characters of micro-blog posts to get the dynamic interests of bloggers on different times. Furthermore, based on the emotion classification, the emotion transition tendency analysis can also be implemented. Experimental results show the feasibility of the presented approach.

Keywords: natural language processing; micro-blog; personal modeling; personal interests; emotion transition

随着互联网的迅速发展,新兴的社交网络正快速走进人们的生活。由于微博具有即时性、互动性等特点,较传统媒体等,它更迎合了广大用户,特别是年轻一代网民的需求,博文也传达了社会各方面的舆情和用

收稿日期:2014-10-08;修回日期:2014-12-23;责任编辑:李 穆

基金项目:国家自然科学基金(61272362);河北省自然科学基金(F2013208105);河北省高等学校科学技术研究重点项目(ZD2014029);河北科技大学五大平台开放基金课题资助项目

作者简介:王九硕(1989—),女,河北石家庄人,硕士研究生,主要从事自然语言处理、微博计算等方面的研究。

通讯作者:高 凯副教授。E-mail:gaokai@hebust.edu.cn

王九硕,高 凯,赵 捷,等.基于微博文本的个性化兴趣关注点及情绪变迁趋势研究[J].河北科技大学学报,2015,36(2):188-194.

WANG Jiushuo,GAO Kai,ZHAO Jie,et al. Research on micro-blog based personal interests and emotion transition tendency[J]. Journal of Hebei University of Science and Technology,2015,36(2):188-194.

户的情感信息。对博主个性化兴趣关注点及情绪变迁分析的研究是自然语言处理、文本挖掘与心理学交叉领域的重要研究课题,它能分析微博用户的兴趣及情感,分析其情绪变迁趋势,这对探究自然语言信息背后隐藏着的舆情及情感趋势分析具有重要的应用价值。例如,在2013-04-20到2013-04-23的几天时间内,广大的新浪微博博主的关注点就从“4.20雅安地震”事件转为“4.23新疆暴力恐怖案”;而在一年之后的这个时间段内,用户的关注点又从“马航MH370失联客机”事件转移到“兰州自来水苯污染”事件上。及时有效地分析社会网络用户的兴趣关注点及其变迁,对舆情分析具有重要价值,这就需要有效地处理机制来对此进行分析。然而,由于公众关注的信息(如社会热点事件)是经常变化的,每个人的关注点亦有所不同,人们在不同的时期对不同的热点事件也有不同的关注度,且在缺乏言论主体背景知识的前提下,很多情感分析面临的歧义问题得不到有效解决,这就使得传统的文本挖掘算法在分析和处理微博话题的过程中,不能有效地与话题变迁过程结合起来。本文提出一种特征权重计算方法,并将特征词权重与话题变迁时间段结合起来进行分析。具体地,首先利用点互信息方法提取出情绪特征词,为使得到的特征词权重更有针对性,本文通过基于距离的语义相似度选择方法,选择具有相似语义的不同特征词,然后将其应用到特征权重计算中,以此来修正特征词的权重;最后将时间因素加入到特征词权重计算方法中,以便能反映用户在不同时间段关注的相关信息。另一方面,使用情绪分类方法,将博主的情绪分成高兴、悲伤、厌恶、愤怒和恐惧,同时也将时间因素添加进来,从而实现了对博主在不同时期情绪变迁过程的分析。

自2006年Twitter出现以来,微博等社交网络快速发展,越来越多的研究人员开始研究微博信息的传播模式,或者通过分析网络结构来识别具有影响力的博主^[1]。文献[2]对早期研究工作中通过文本内容来分析社会网络的方法进行探讨。文献[3]指出了发现社交网络中热点话题的问题,并提出一个融合话题、社会关系和微博的概率框架来实现有效的社区发现。模型方法方面,文献[4]展示了一些常见的统计模型方法(如逐步回归、基于偏最小二乘回归的径向基函数、偏鲁棒M-回归和主成分回归等),将其应用到多重共线性域中。文献[5]提出一种基于传统的多信息特征选择的改进方法,通过对不同类别中词的不同表示来构建域特征词。文献[6]提出一个基于TFIDF的权重计算框架,通过文档词频率归一化来决定对应词的重要性。文献[7]实现了对微博网络结构的分析。

在情绪分析方面,文献[8]提出一种利用情绪诱因提取技术进行微博文本情绪分类的算法;文献[9]通过一个多任务多标记的分类模型,来实现情绪与话题的同步分类;文献[10]通过抽取特征向量和使用SVM分类方法实现情绪的分类;文献[11]基于社会网络理论抽取博文的情绪关系;文献[12]使用SVM算法完成了对微博文本的情绪识别;文献[13]通过分析微博文本的特性,包括表情符号、标点符号以及语法框架中核心情绪词间的距离,并采用改进的依存句法分析来识别文本中包含的情绪;文献[14]通过使用微博中的表情符号、由频率统计和标签传递算法构造的情绪词典、以及微博中的语言特征来实现情绪分类;文献[15]提出以深信度网为基础框架,并利用伪标记数据进行句子的表达学习,以实现微博文本中的情绪分类。文献[16]提出一种基于类序列规则的微博情绪分类方法。文献[17]提出了一个高阶的隐马尔可夫模型来进行文本情绪探测。文献[18]分别使用SVM以及人工神经网络(ANN)这两种方法进行情绪分类,并全面论证了这两种方法在情绪分类的性能。文献[19]通过利用上下文情绪词以及句子的语法结构来提取出特征集,并将其应用到情绪分类中,最后使用不同的分类方法来评估特征集的性能。

和上述工作不同,本文是从博主的个性化兴趣关注点进行分析,通过一种基于微博文本的特征权重计算方法,将时间因素加入到此方法中,来计算不同时间段内特征词的权重,从而得到博主的关注信息,并根据此关注信息来分析博主在此时间段内情绪的变化情况,进而完成了社会网络个性化兴趣关注点及情绪的变迁分析工作。

1 基于微博文本的权重计算方法

定义1 博主和它对应的博文可用四元组 Q 形式化表示,如公式(1)所示:

$$Q = (\text{User-id}, \text{keywords}, \text{weight}, \text{timestamp}). \quad (1)$$

式中:User-id代表博主;keywords表示在某一时间段从博文中提取出的特征词;weight代表对应特征词权重;timestamp表示相应的博文发布时间。

首先,对于四元组 Q 中特征词keywords参数的确定,使用基于篇章分析、中文分词、同义词和未登录词

处理的算法,完成对博文主题词的抽取,此方法可弥补单独使用统计方法的缺陷以及避免语义分析中的低频词;其次,对于四元组 Q 中特征词权重 $weight$ 参数的确定,是以传统的 TFIDF 为基础的,但传统的 TFIDF 方法只反映静态文档集中特征词的权重,不能有效地表示特征词在不同时间段内的分布信息。因此,本文首先利用基于点互信息(PMI)的方法提取出特征词,然后提出与时间参数 $timestamp$ 相结合的特征词权重计算方法(参见公式(4))。在此过程中,本文从已有的 4 种情感词库(清华大学词典、台湾大学词典、大连理工大学词典、WordNet)中构造关于(高兴、愤怒、厌恶、恐惧、悲伤)基本情绪的标准词库,然后从微博语料库中选取特定词汇,并根据它与标准情绪词汇在语料库中的互信息大小来确定特征词^[20-21],以下为互信息计算公式,见式(2):

$$PMI(WD_k, ST_{kj}) = \log \frac{P(WD_k ST_{kj})}{P(WD_k)P(ST_{kj})}, \quad (2)$$

式中: WD_k 表示在语料库中属于 $k(1 \leq k \leq 6)$ 类情绪下的词; ST_{kj} 表示标准词库中第 k 类情绪下的第 j 个标准词。最后本文从语料库中选择与标准词汇的互信息最大的词作为特征词。

鉴于微博中不同特征词可能表达相同或者相似的语义,且在不同时间段内博主关注的内容在不断变化,故使用公式(3)来计算特征词的相似度,并为特征权重计算做准备,其中 δ 代表可调参数, $Dis(w_{i,t}, w_{k,t})$ 代表语义距离, i 和 k 分别代表不同的关键词序号,若 2 个特征词的语义相似度 $Sim(w_{i,t}, w_{k,t}) > 0$,意味着在给定时间段 t 内的 2 个特征词($w_{i,t}, w_{k,t}$)属于博主所关注的相关信息,故可将两者作为同一个关注点对待。

$$Sim(w_{i,t}, w_{k,t}) = \frac{\delta}{\delta + Dis(w_{i,t}, w_{k,t})}; \quad (3)$$

$$\begin{cases} W_{i1} = S_{i1}, & t=1; \\ W_{it} = \alpha(S_{it} + \beta WSim(w_{i,t-1}, w_{k,t-1}) + \gamma_{t-1} W_{i(t-1)}), & t > 1. \end{cases} \quad (4)$$

在式(4)中, t 表示时间段, α 和 β 分别代表经验因子。本文设置博主关注点变化的初始时间段 $t=1$,此时用每个特征词的频率 S_{i1} 来计算权重 W_{i1} ; $WSim(w_{i,t-1}, w_{k,t-1})$ 代表在 $Sim(w_{i,t-1}, w_{k,t-1}) > 0$ 情况下特征词 $w_{k,t-1}$ 的权重; γ_{t-1} 代表在特定时间段($t-1$)下特征词的影响度,即在时间段($t-1$)中的特征词在时间段 t 下受到外界因素的影响而发生变化的程度,“total number of keyword”表示特征词的总数,“the number of ranking”表示当前特征词按照权重由大到小的顺序进行排序后的名次,其定义如公式(5)所示。如果 W_{it} 值低于一定的经验阈值,说明用户此时已对该关注内容不感兴趣。

$$\gamma_{t-1} = \frac{\text{the total number of keyword at time } (t-1) - \text{the number of ranking at time } (t) + 1}{\text{the total number of keyword at time } (t-1)}. \quad (5)$$

从上面给出的定义和公式可见,特征词权重在不同时期会有所改变。因此,随着特征词权重排名的不断更改,一些新的特征词将代替旧的特征词出现在特征词集合中,这与实际情况中博主关注点的变化是一致的,也反映了特征词与关注点之间的关系。

2 基于 SVR 算法的博主情绪分析

虽然目前在情绪认知分析领域还没有形成一个统一的基础情绪类别标准,但有些基本的情绪类别通常是被认可的。本文使用基本情绪(高兴、愤怒、厌恶、恐惧、悲伤)作为基础情绪来分析博主的情绪及其变迁过程。另外,除利用自然语言处理工具从文本信息中挖掘用户的基础情绪外,对表情信息的分析也至关重要。因此,建立一个微博表情库,形成一个情绪类型与表情图片的对应关系库来对表情符号进行分析。借鉴文献[8]的情绪分析方法,采用基于 SVR(support vector regression)的情绪分类方法,分析过程如图 1 所示。


首先,进行数据预处理(包括过滤链接博文、繁体与简体的转换以及博文去重等);其次,提取出微博文本特征,比如在微博文本中往往包含大量的表情信息,而这些表情最能直接反映用户当时的情绪,拿  这一表情来说,它包含了强烈的高兴情绪。本文将抽取出的表情符号与基础情绪联系起来,形成一个表情库,如表 1 所示。对于修饰词、否定词、关联词、标点符号等,这些特征在影响情绪产生过程中也起到一定的作用。

表 1 表情集
Tab.1 Emotion sets

表情类型	表情图片示例
高兴	[呵呵] [嘻嘻] [哈哈]
愤怒	[怒] [哼] [怒骂] [鄙视]
厌恶	[吐]
恐惧	[惊恐]
悲伤	[伤心] [悲伤] [失望] [泪]

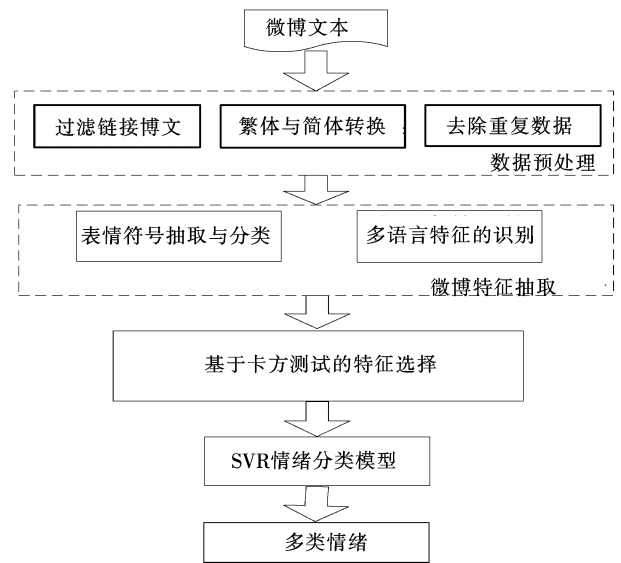


图 1 情绪分析过程

Fig.1 Process of emotion analysis

最后,利用式(6)所示的卡方测试方法(式中 N_i 表示观察频数, n 是总频数, p_i 是期望频率),将文本中的高频词以及类别相关度大的情绪特征词挑选出来,作为辅助的情绪特征,并选择最优的特征数量形成 VSM 向量维度,以此为基础来调节 SVR 参数和分类使用的阈值,实现对微博文本的最佳情绪分类效果。

$$\chi^2 = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i} \tag{6}$$

3 实验结果与分析

3.1 实验数据集

为了更好地分析微博文本,文本采用基于模拟登录技术的方法,从新浪微博平台 weibo.com 获取了大量的微博数据。模拟登录的主要方法是通过使用网页浏览器获得相应的数据,图 2 显示了从新浪微博中采用上述方法随机获取的 154 678 条微博的统计图(水平轴表示博文的长度,纵轴则表示对应的比例),从图中可以看出,博文均较短,即使对博文进行忽略词处理,也难掩其内容碎片化、口语化、不规范等特点,这使得常规的基于关键词或主题词的分类、聚类等传统算法难以发挥应有的作用。

为验证本文算法的有效性,在数据采集阶段,有针对性地抓取了一些特殊人群的微博数据,完成对核心话题的数据采集。随机选取李开复(<http://weibo.com/kaifulee>)的微博数据进行分析。首先,抓取其在 2013-04-20 到 2013-05-20 期间发表的博文作为数据源,并分成 10 个时间段作为观测区间(如 Time=1, Time=2, ..., Time=10),具体数据集如表 2 所示。

表 2 数据集表
Tab.2 Data set table

Time 序号	提取时间段	原始微博	预处理之后的
		总数	微博总数
1	2013-04-20-2013-04-23	62	50
2	2013-04-23-2013-04-26	52	45
3	2013-04-26-2013-04-29	57	52
4	2013-04-29-2013-05-02	68	65
5	2013-05-02-2013-05-05	49	45
6	2013-05-05-2013-05-08	62	54
7	2013-05-08-2013-05-11	58	50
8	2013-05-11-2013-05-14	68	63
9	2013-05-14-2013-05-17	63	56
10	2013-05-17-2013-05-20	54	47

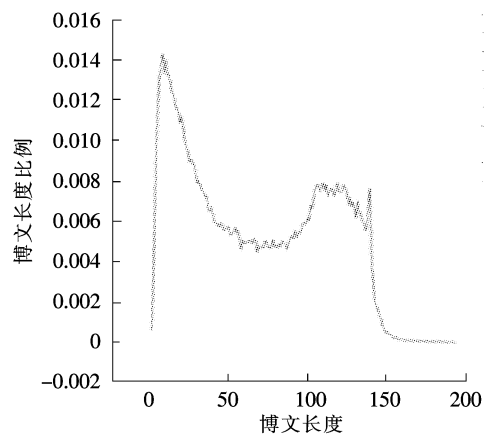


图 2 微博内容统计图

Fig.2 Statistical figure of micro-blog content

表2中的有效微博数指的是经过预处理(指对博文去重、简繁体转换、过滤无关成分、全半角转换等)之后剩下的微博数目。

为了更好地对本算法进行评估,本文将语料库交给3名情绪挖掘领域的人员进行人工标注,标注的内容包括情绪的类别(如:“高兴”、“愤怒”、“厌恶”、“恐惧”、“悲伤”)。具体标注过程如下所示:1)2个标注人员对博文中包含明显情绪类型的博文进行标注;2)如果博文不包含任何情绪,则不对其进行标记;3)如果2个标注人员都无法确定此博文属于哪类情绪,则此博文将被标注为中性;4)如果2个标注人员在标记过程中出现冲突,则最终的结果交由第3个标注人员决定。

3.2 个人关注点与情绪变迁分析

为了更好地分析微博用户在不同时间段对关注话题的变化与情绪变迁情况,本文选取具有较高权重的前 N 个特征词(表3中,经验参数 N 取值5)作为分析博主的兴趣关注点信息。为了分析本方法的实验效果,使用传统的TFIDF方法作为对比。从实验结果中可看到本文方法得到的特征词权重有所提高(表3中的划线部分为特征词权重提高部分),说明时间因素和特征词相似度在特征词权重计算方面也起到了一定的作用,而基于传统的TFIDF得到特征词权重不能较好地反映该词在微博中随时间变化而产生的影响。统计表明,基于时间因素和特征词相似度的算法得到的反映用户特征的关键词集合的权重值,比常规TFIDF算法提高了10.81%(限于篇幅,表3仅给出针对特定博主的部分特征词权重计算结果),从而说明了本文方法的可行性。选取具有最高权重的特征词作为该用户的关注点话题,并利用标签云技术将分析结果显示出来,如图3所示,可更加直观而有效地展示特定博主在某个时期的关注情况。

表3 Top-5特征词权重分析

Tab.3 Analysis of Top-5 feature terms' weights

Time 序号	特征词	TFIDF	改进方法	Time 序号	特征词	TFIDF	改进方法
1	朋友	0.014 4	0.014 4	6	工作	0.008 6	0.008 6
	救援队	0.014 4	0.014 4		调查	0.011 0	0.011 0
	物资	0.014 4	0.014 4		美食	0.012 9	0.012 9
	爱心	0.014 4	0.014 4		中国	0.015 0	<u>0.019 0</u>
	灾区	0.018 1	0.018 1		简历	0.019 3	0.019 3
2	中国	0.021 7	0.021 7	7	媒体	0.008 3	<u>0.008 8</u>
	美国	0.012 6	0.012 6		世界	0.011 0	0.011 0
	捐款	0.009 0	0.009 0		互联网	0.011 0	0.011 0
	地震	0.007 2	<u>0.008 1</u>		产品	0.013 8	0.013 8
	媒体	0.007 2	<u>0.010 8</u>		白宫	0.016 6	<u>0.016 9</u>
3	美国	0.009 5	<u>0.020 5</u>	8	白宫	0.010 5	<u>0.027 4</u>
	红会	0.009 5	0.009 5		台湾省	0.012 3	<u>0.013 7</u>
	物资	0.009 5	0.009 5		失恋	0.012 3	0.012 3
	媒体	0.007 6	<u>0.013 0</u>		中国	0.014 0	<u>0.014 9</u>
	遇难者	0.007 6	0.007 6		美国	0.017 5	<u>0.018 0</u>
4	阿里	0.006 3	0.006 3	9	创业	0.008 4	0.008 4
	世界	0.009 4	<u>0.009 7</u>		爱情	0.010 5	0.010 5
	大都会	0.014 0	0.014 0		智慧	0.010 5	0.010 5
	博物馆	0.018 8	0.018 8		媒体	0.010 5	<u>0.010 9</u>
	纽约	0.020 3	0.020 3		销号	0.010 5	<u>0.010 7</u>
5	社交	0.007 5	<u>0.009 3</u>	10	大学	0.006 4	<u>0.006 7</u>
	台湾省	0.007 5	<u>0.008 3</u>		游戏	0.008 5	<u>0.008 8</u>
	创业者	0.010 0	0.010 0		中国	0.012 8	<u>0.013 8</u>
	媒体	0.010 0	<u>0.010 8</u>		公司	0.015 0	<u>0.015 3</u>
	巴菲特	0.012 5	0.012 5		创业	0.036 3	<u>0.037 7</u>

利用对基础情绪的分类算法,对相关博主在一个月内的博文进行分析,得到博主在不同时间段内表现出来的情绪如图 4 所示(其中,横坐标表示时间段,纵坐标表示微博数目)。从图 4 中可以看出,博主在时间段内所表现出的基础情绪以波浪式的方式变化,其中,在第 1 个时间段内发生了“4. 20 雅安地震”事件,正好与博主在这个时间段的关注点“灾区”相符,此时用户主要表现为“悲伤”的情绪,说明用户情绪与当时发生的社会环境有关。另一方面,从博主的整体情绪变化来看,用户多数情况下是处于“高兴”或者无情绪的状态,这也说明了用户在日常生活中一直表现积极的态度。从宏观方面来说,有效地提取出用户的情绪可以帮助研究人员研究其心理活动,对于构建健康的网络环境和社会环境具有深远的意义。

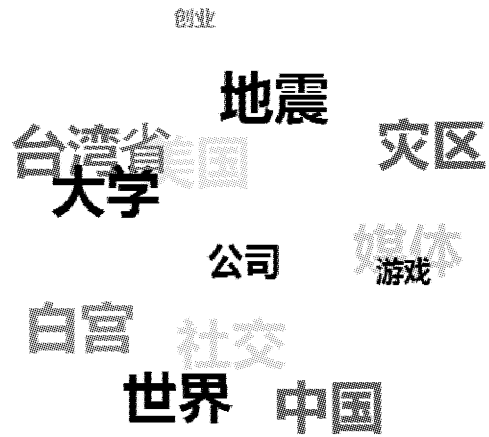


图 3 个性化兴趣关注点标签云
Fig. 3 Tag cloud of personal interests

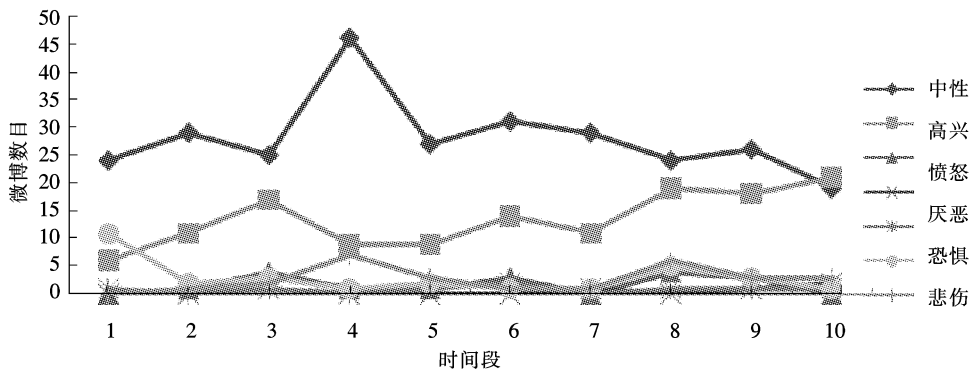


图 4 情绪变迁情况
Fig. 4 Case of emotion transition tendency

4 结 语

本文给出基于微博文本的个人兴趣关注点动态变迁算法以及基于 SVR 的博主情绪变迁方法,该方法虽然达到了预期的目标,但是在特征权重计算方法上仍有改进的空间,在情绪分析方面可能存在部分片面性,如未将博主的性格特征与博主的情绪结合起来进行分析,观察用户的情绪变迁情况。分析结果以标签云的形式展现出来,达到了关注点可视化的效果。未来的工作中,将完善相关方法,并对情绪产生的诱因进行分析,以实现较好的情绪诱因抽取效果,并计算出不同的诱因成分所占的比例。

参考文献/References:

[1] KWAK H, LEE C, PARK H, et al. What is twitter, a social network or a news media [A]. Proceedings of the 19th International Conference on World Wide Web [C]. New York: ACM, 2010:591-600.

[2] DANESCU-NICULESCU-MIZIL C, LEE L, PANG B. Echoes of power: Language effects and power differences in social interaction [A]. Proceedings of the 21st International Conference on World Wide Web [C]. New York: ACM, 2012:699-708.

[3] SACHAN M, CONTRACTOR D, FARUQUIE T A, et al. Using content and interactions for discovering communities in social networks[A]. Proceedings of the 21st International Conference on World Wide Web [C]. New York: ACM, 2012:331-340.

[4] GARG A, TAI K. Comparison of statistical and machine learning methods in modelling of data with multicollinearity[J]. International Journal of Modelling, Identification and Control, 2013, 18(4): 295-312.

- [5] LUO Y, OUYANG N. Text similarity calculation based on domain feature word [A]. International Conference on Automatic Control and Artificial Intelligence [C]. New York: IEEE, 2012:2049-2051.
- [6] PAIK J H. A novel tf-idf weighting scheme for effective ranking[A]. Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. New York: ACM, 2013:343-352.
- [7] WENG Jianshu, LIM E P, JIANG Jing. Twiterrank; finding topic sensitive influential twitterers [A]. Proceedings of the Third ACM International Conference on Web Search and Data Mining [C]. New York: ACM, 2010:261-270.
- [8] LI Weiyuan, XU Hua. Text-based emotion classification using emotion cause extraction[J]. Expert Systems With Applications, 2014, 41(4): 1742-1749.
- [9] HUANG Shu, PENG Wei, LI Jingxuan, et al. Sentiment and topic analysis on social media: a multi-task multi-label classification approach[A]. Proceedings of the 5th Annual ACM Web Science Conference[C]. New York: ACM, 2013:172-181.
- [10] CHO S H, KANG H B. Text sentiment classification for sns-based marketing using domain sentiment dictionary[A]. 2012 IEEE International Conference on Consumer Electronics (ICCE) [C]. New York: IEEE, 2012:717-718.
- [11] HU Xia, TANG Lei, TANG Jiliang, et al. Exploiting social relations for sentiment analysis in microblogging [A]. Proceedings of the sixth ACM International Conference on Web Search and Data Mining[C]. New York: ACM, 2013:537-546.
- [12] ZHENG Yuan, MATTHEW P. Predicting emotion labels for chinese microblog texts [A]. CEUR Workshop Proceedings[C]. UK: School of computing University of Portsmouth Buckingham Building, 2012,917(4): 40-47.
- [13] GUO Fuliang, ZHOU Gang. Research on micro-blog sentiment orientation analysis based on improved dependency parsing [A]. 2013 3rd International Conference on Consumer Electronics, Communications and Networks (CECNet) [C]. New York: IEEE, 2013:546-550.
- [14] JIANG Fei, CUI Anqi, LIU Yiqun, et al. Every Term has Sentiment; Learning from Emoticon Evidences for Chinese Microblog Sentiment Analysis [M]. Berlin: Springer Berlin Heidelberg, 2013:224-235.
- [15] TANG Duyu, QIN Bing, LIU Ting, et al. Learning Sentence Representation for Emotion Classification on Microblogs [M]. Berlin: Springer Berlin Heidelberg, 2013:212-223.
- [16] WEN Shiyang, WAN Xiaojun. Emotion classification in microblog texts using class sequential rules [A]. Twenty-Eighth AAAI Conference on Artificial Intelligence[C]. Canada: AAAI Press, 2014:187-193.
- [17] HO D T, CAO T H. A High-order Hidden Markov Model for Emotion Detection from Textual Data [M]. Berlin:Springer Berlin Heidelberg, 2012: 94-105.
- [18] MORAES R, VALIATI J F, NETO W P. Document-level sentiment classification: an empirical comparison between svm and ann[J]. Expert Systems with Applications, 2013, 40(2): 621-633.
- [19] GHAZI D, INKPEN D, SZPAKOWICZ S. Prior and contextual emotion of words in sentential context [J]. Computer Speech & Language, 2014, 28(1): 76-92.
- [20] 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.
XU Linhong, LIN Hongfei, PAN Yu, et al. Constructing the affective lexicon ontology [J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(2): 180-185.
- [21] 于潇,万军,何翔,等.校园微博情感分析系统的设计与实现[J].河北工业大学学报,2013,42(6):24-29.
YU Xiao,WAN Jun,HE Xiang,et al. The design and realization of the micro blog sentiment analysis system for campus network[J]. Journal of Hebei University of Technology,2013,42(6):24-29.