

doi: 10.12052/gdutxb.210022

基于观点挖掘的突发事件微博意见领袖识别方法

刘高勇, 谭依雯, 艾丹祥, 黄靖钊

(广东工业大学 管理学院, 广东 广州 510520)

摘要: 基于意见领袖概念的本质, 运用观点挖掘技术研究突发事件中微博意见领袖的识别, 为网络舆情治理提供参考。提出三步识别方法框架: 首先采用文献分析法构建指标模型, 评价微博博主的信息影响力; 其次构建文本主客观分类模型, 计算高影响力博主事件相关博文观点输出性, 识别观点博文; 然后针对观点博文的评论文本构建情感极性分类模型, 计算博文观点获得的支持度以及博主观点的支持率, 最终将输出了观点并获得较多支持的高影响力博主作为该事件的“意见领袖”。同时, 运用上述方法对典型实际案例的微博舆情数据进行分析, 识别该事件的微博意见领袖, 对其特征和舆情参与行为进行观察。并将该结果与社会网络分析法、专家人工分析法识别的结果进行了对比, 验证了本方法的科学性和有效性。结果表明, 本方法能够识别在突发事件舆情生命周期中真正拥有舆论引导力, 具备治理价值的微博意见领袖。

关键词: 观点挖掘; 文本主客观分类; 文本情感极性分析; 突发事件; 意见领袖

中图分类号: TP3-05; G202

文献标志码: A

文章编号: 1007-7162(2021)04-0041-11

A Microblog Opinion Leader Identification Method Based on Opinion Mining in Emergencies

Liu Gao-yong, Tan Yi-wen, Ai Dan-xiang, Huang Jing-zhao

(School of Management, Guangdong University of Technology, GuangZhou 510520, China)

Abstract: Based on the nature of opinion leaders, the opinion mining technology is used to study the identification of microblog opinion leaders in emergencies in order to provide a reference for the governance of online public opinions. A three-step opinion leaders identification framework is proposed: first, using a literature analysis, an indicator model is built to evaluate the influence of microblog bloggers; second, a text subjective and objective classification model is built to determine whether high-influence bloggers publish event-related opinions and information, and to identify opinion blog posts; then an emotional polarity classification model is built based on the comment text of the opinion blog post to calculate the support degree of the blog post opinions and the support rate of the blogger's opinions so that those high-influence bloggers who have first output their opinions and then received more support can be rated as the event "Opinion Leaders". These methods are used in the following text to analyze the microblog public opinion data of the typical practical cases, identifying the microblog opinion leaders of the incident, with their characteristics and public opinion participation behaviors observed and analyzed, and the results compared with those of social network analysis methods and the expert manual analysis method to verify the scientificity and effectiveness of this method. The results show that this method can identify microblog opinion leaders who truly have the guiding power of public opinions and can have governance value in the evolution of public opinions in emergencies.

Key words: opinion mining; text subjective and objective classification; text emotional polarity analysis; emergency events; opinion leaders

收稿日期: 2021-02-01

基金项目: 国家自然科学基金资助项目(71740024); 广东省哲学社会科学“十三五”规划2020年度课题(GD20CTS01); 广州市哲学社会科学“十三五”规划2018年度课题(2018GZYSB67)

作者简介: 刘高勇(1975-), 男, 教授, 博士, 主要研究方向为大数据分析、智能信息处理

通信作者: 艾丹祥(1978-), 女, 副教授, 博士, 主要研究方向为知识工程、文本挖掘, E-mail: aidx@gdut.edu.cn

突发事件是指突然发生,造成或可能造成较大社会影响的公共事件,经过网络的发酵后形成网络舆情。在突发事件网络舆情的传播过程中存在着一些“意见领袖”,他们通常是各个领域的媒体、精英、名人或者“网红”,时刻关注着网络上的热点,积极地扮演着“信息引导者”和“信息评论者”的角色,高频率地表达自己对事件的看法和态度,有意识地追踪和推动舆情的发展。社交媒体平台往往是各类意见领袖输出观点、引导舆论的主要渠道,他们利用社交网络响应速度快、受众面广等特点,获得更高的舆情影响力。在我国最大的社交媒体平台——新浪微博中,聚集了各个领域大量的意见领袖。在社会性突发事件发生时,普通网民往往处于不知情状态或持有负面态度,而微博意见领袖是他们进一步了解事件态势的重要信息来源,大众也容易受到其观点的影响。如果出现少数意见领袖为获取网络流量而有意散布错误信息和负面观点,则可能引发社会矛盾,导致社会治理危机。因此准确地识别微博意见领袖及其舆论导向,是事件舆情治理的关键。

从前人对意见领袖概念的定义^[1]中可以看出,作为意见领袖需要满足2个条件:(1)能够传播信息,并具有一定的影响性;(2)需要有自己的观点和态度,并能够获得群体的关注和支持。然而现有的研究在识别事件意见领袖时往往更注重对前者的评估,而忽略了第二项本质条件。事实上,某些媒体或“大V”用户发布的信息可能因为其在其他领域获得的影响力而得到广泛传播,但其信息只是起到陈述事实的作用,并未输出观点并形成意见引导,因此不应将其判定为意见领袖。针对此问题,本文将传统的意见领袖指标模型评价和基于机器学习的观点挖掘技术结合起来,综合考虑“信息影响力、观点输出性、观点支持度”3个方面的衡量标准,形成一种新的突发事件微博意见领袖识别方法。

1 相关研究

“意见领袖”是传播学者拉扎斯菲尔德在20世纪40年代《人民的选择》一书中提出的。在书中,他提出了“两级传播”理论,认为大众传播并不是直接“流”向一般大众,而是通过意见领袖在两者之间进行信息传达,其模式是:大众传播—意见领袖—一般受众^[1],因此意见领袖在信息传播过程中扮演了重要角色。此后,意见领袖的概念在多个不同领域得到了广泛的研究,国内主要是网络舆情、新闻传播等领域。进入21世纪后,web2.0的发展促进了大量网络社

区的产生,而网络社区中同样存在着意见领袖,这些网络空间的意见领袖影响和改变着用户的思想倾向。

在国内,微博已经成为突发事件网络舆情形成的重要平台,从2010年开始逐渐变成网络舆情研究的重点。微博意见领袖识别主要是采用指标分析法和社会网络分析来识别意见领袖,王佳敏等^[2]从影响力和活跃度两个维度出发,结合微博的传播特点,构建微博意见领袖指标体系,应用一种改进的层次分析法确定指标权重。彭丽徽等^[3]从影响力、活跃度和认同度3个维度出发,构建微博意见领袖指标体系,应用一种改进的模糊—层次分析法确定指标要素权重,构建一种意见领袖识别模型。吴江等^[4]融合用户个人属性、网络特征、行为特征和文本特征,构建意见领袖识别的综合指标体系。

近年来,随着大数据和人工智能技术的发展,自然语言处理技术也被越来越多地运用于网络文本信息的分析与挖掘中。其中观点挖掘和情感分析技术受到研究者的广泛关注,该技术能分析文本中隐含的情绪状态,进而对文本发表者的态度、意图、评价等进行推测。基本的观点挖掘技术包括文本主客观分析、情感分类等。其中主客观分析用于判断文本是否表达了情感或含有观点。情感分类则是用于判断主观性文本中情感的类型。简单的情感分类将文本中情感分为正向和负向,也称为情感极性分析或情感倾向性分析,通常正向代表褒义的情感,负向代表贬义的情感。而更复杂的情感分类技术则是将文本中蕴含的情感进一步划分为“喜”“怒”“哀”“惧”等多种类型。一些研究者已将这些技术应用于网络意见领袖的识别,并提出了一些新的思路与方法。如陈芬等^[5]采用多级文本倾向性分析识别网民评论的5种情感倾向,表征网民对博主的支持度,并将其融合进识别网络意见领袖的指标体系。安璐等^[6]采用主题分析和情感分析方法,基于主题一致性和情感支持识别评论区的意见领袖。他们的研究表明,观点挖掘技术有助于深入判断意见领袖的舆论影响力和引导力,能识别出更为有效的结果。

尽管目前识别网络意见领袖的方法众多,但判定意见领袖的标准仍然缺乏统一性,尤其是缺乏对意见领袖“观点引导性”这一本质特征的考量和评判。而文本则针对这一问题,在现有研究成果的基础上,进一步将观点挖掘和情感分析技术应用于对突发事件中微博意见领袖本质特征的识别与量化评价,从而构建一套新的识别方法。

2 突发事件微博意见领袖识别方法

2.1 方法框架

本文提出的方法框架如图1所示。该方法的初始数据集为 B_0 ,包含与突发事件相关的全部微博博文及其博主,然后分3个步骤从博主中筛选出事件的微博意见领袖。

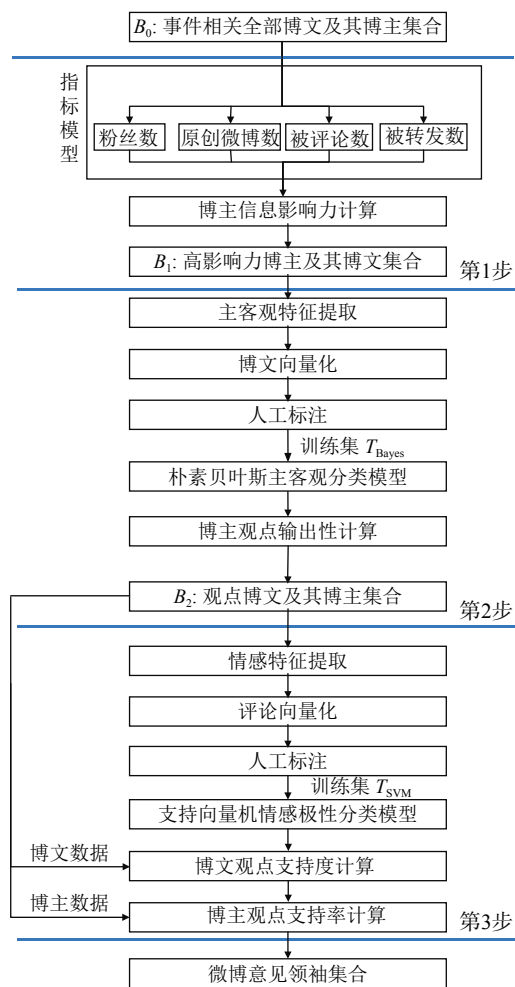


图1 突发事件微博意见领袖识别方法框架

Fig.1 Identification method framework of microblog opinion leaders in emergencies

第1步:基于指标模型的博主信息影响力计算。

信息影响力大小决定了博主的事件相关表述能否被足够多的人看到,也是博主成为事件意见领袖的首要条件。因此本方法首先根据博主的用户属性及其在社交网络中的重要性构建其信息影响力指标模型,通过模型计算,从 B_0 中找出高影响力博主及其博文,形成集合 B_1 。由于指标模型的计算相对简单,因此将其作为本方法框架中的第1步,可以快速筛选掉大批不具意见领袖潜质的普通博主,提高方法的整体计算效率。

第2步:基于博文主客观分类的博主观点输出性计算。

本步骤判断 B_1 中的博主是否在其事件相关表述中形成了观点输出。如果博主要表达自己的观点,则往往会在博文中对事件的某些方面进行具有主观色彩的加工,而不是单纯的客观报道,因此本步骤首先提取 B_1 中所有的博文,然后基于观点挖掘技术中的文本主客观分类,计算每篇博文表述的观点输出性,判断其是否为观点博文,并将 B_1 中的所有观点博文及其博主选出,形成集合 B_2 。

第3步:基于评论情感极性分类的博主观点支持率计算。

本步骤判断 B_2 中的博主观点是否获得了大众支持。博文所获得的评论能直接反映出大众对博文内容的看法和态度,如果评论者认同博文中的观点,则其评价往往是正面和积极的,反之则是负面和消极的。因此本步骤首先获得 B_2 中所有博文的评论,然后基于观点挖掘技术中的文本情感极性分类,计算每篇博文获得的观点支持度,判断其是否获得较多支持,并将 B_2 中的博文划分为获支持的博文和未获支持的博文,最终计算 B_2 中所有博主获支持观点的比例,将大部分观点均获得支持的博主认定为意见领袖。

2.2 博主信息影响力计算

对于网络用户信息影响力的计算,不同的研究者从各个维度提出了评价指标,但大部分可以分为2类,一类考虑用户自身的属性,如用户的活跃度、拥有的粉丝数、是否认证等;一类考虑用户在网络传播中的重要性,如被转发数、被评论数、被点赞数、被@数等。为了构建一个精简的指标模型计算博主的信息影响力,作为筛选意见领袖的粗略标准,选取7篇不同时期识别意见领袖的代表性文献成果。论文1:孙乃利等^[7]发表于2012年;论文2:魏志惠等^[8]发表于2014年;论文3:王佳敏等^[2]发表于2016年;论文4:彭丽徽等^[3]发表于2017年;论文5:陈芬等^[5]发表于2018年;论文6:胡若涵等^[9]发表于2018年;论文7:陈芬等^[10]发表于2019年,并对比总结了其中使用的影响力评价指标,提取了4个被所有文献共同使用的核心指标:粉丝数 w_1 、原创微博数 w_2 、被评论数 w_3 和被转发数 w_4 。其中 w_1 和 w_2 作为博主用户属性上的代表性指标,而 w_3 和 w_4 作为突发事件中博主网络传播重要性上的代表性指标,仅限为与当前事件相关的博文的被评论和被转发数量。

表1显示了从7篇文献成果中提取出的4个核心

指标的原始权重,由于各篇论文中评价指标权重的范围不统一,为了便于各指标的计算,使用式(1)对每篇论文中4个指标的权重进行归一化处理。

$$w_i^* = \frac{w_i}{\sum w_i}, i = 1, 2, 3, 4 \quad (1)$$

表1 7篇文献的指标权重
Table 1 Index weight table of 7 articles

指标	论文1	论文2	论文3	论文4	论文5	论文6	论文7
w_1	0.26	0.30	0.14	0.19	0.05	0.01	0.12
w_2	0.07	0.07	0.06	0.06	0.07	0.19	0.07
w_3	0.13	0.28	0.33	0.21	0.13	0.24	0.20
w_4	0.13	0.26	0.33	0.22	0.08	0.24	0.25

表2显示了归一化处理后的新的权重值。

表2 归一化后的指标权重
Table 2 Index weights after normalization

指标	论文1	论文2	论文3	论文4	论文5	论文6	论文7
w_1^*	0.43	0.32	0.17	0.28	0.16	0.01	0.19
w_2^*	0.13	0.08	0.07	0.09	0.21	0.29	0.11
w_3^*	0.22	0.31	0.38	0.31	0.38	0.35	0.31
w_4^*	0.22	0.29	0.38	0.32	0.25	0.35	0.39

然后取各指标的7篇论文权重均值,形成如表3所示的指标模型。

表3 博主信息影响力计算指标模型
Table 3 Blogger information influence calculation index model

评价指标	指标权重
粉丝数 w_1	0.23
原创微博数 w_2	0.13
被评论数 w_3	0.33
被转发数 w_4	0.31

运用上述指标模型,可对初始数据集 B_0 中的每位博主进行信息影响力的评价计算,并将影响力大于一定阈值的博主选出,作为高影响力博主。

2.3 博主观点输出性计算

对于 B_1 中的每位高影响力博主,如果是真正的意见领袖,则其发表事件相关博文时,不仅会陈述事件情况,还会表达个人的感受、意见或观点,直观体现在其博文文本含有较强烈的主观色彩。因此,为了计算和识别这些博主的观点输出性,可以对其博文进行主客观分类分析,认为主观性概率高的文本观点输出性高,而客观性概率高的文本观点输出性低。多项前人研究结果表明,在判断文本的主观性

问题上,朴素贝叶斯分类器表现效果较好^[11-13],因此本方法采用基于主客观特征的朴素贝叶斯分类模型对博文的主客观概率进行计算,具体过程如下。

第1步:对于 B_1 中的每一篇博文,提取其中的主客观特征。在区分主客观的特征选择问题上,前人研究成果已提供了较多的经验:杨武^[12]通过分析主客观句的差异,认为可以选取语义层面和语法层面两方面的特征。在语义层面,选取情感词、指示性动词、指示性副词、语气词、标点符号作为主观线索特征,选取时间、地点、描述性无感情色彩名词、冒号等作为客观线索特征;在语法层面,选择2-POS模型作为类别特征。丁晟春^[13]则使用句内特征、句式特征和隐性特征对微博博文进行主客观分析,句内特征包括词性、程度副词、主观指示词、客观词等;句式特征包括是否为否定句、疑问句、感叹句等;而隐性特征则同样使用N-POS模型来表征。刘培玉^[14]提取了6类主观性特征:指示性动词、指示性副词、形容词、情感词、人称代词和指示性标点符号,以及少量的客观词作为客观性特征。综合上述研究成果,本文选取了5类最具代表性的主客观特征。

(1) 指示动词。博主在输出观点时,常常使用“认为”“觉得”等主观动词来表明观点的所属,通常与第一人称代词连用。本方法采用《知网》情感分析用词语集中的中文主张词语作为指示动词特征词表。

(2) 程度副词。博主如果对事件有着强烈情感,则往往会采用“很”“非常”等程度副词来加强情感表达。本方法采用《知网》情感分析用词语集中的中文程度级别词语作为程度副词特征词表。

(3) 情感词。博主的博文中如果含有较多的含有情感倾向的词语,则其输出主观性观点的可能性较大,本方法采用BosonNLP情感词典作为特征词表,主要因为BosonNLP情感词典的构建来源中包含了较多微博博文数据,因此词典囊括了很多网络用语和非规范文本,比较适合进行微博的情感分析和观点挖掘。

(4) 语气助词。语气助词常在句尾或句中停顿处出现。微博文本的口语化决定了人们会将口语习惯带入文本中,常用“哟”“啊”“呵”等语气助词表明个人的某种态度。本方法根据百度百科对现代汉语语气助词的解释,手工列举构建特征词表。

(5) 标点符号。选取问号和感叹号作为特征。其中,问号表示对事件的疑问,具有不确定性;而感叹号表明对事件的惊讶。两者皆在一定程度上传达了博主的态度倾向。

第2步:通过Word2Vec方法对第1步选取的主客观特征进行向量化,并将每篇博文的主客观特征词向量相加,构成该博文的文本向量 $\mathbf{d} = (f^{(1)}, f^{(2)}, \dots, f^{(n)})$,其中 $f^{(1)}, f^{(2)}, \dots, f^{(l)}, \dots, f^{(n)}$ 为 \mathbf{d} 具备的 n 个独立的主客观特征。

第3步:构建朴素贝叶斯分类模型。首先,从现有的博文向量中选择一部分样本,人工标注其主客观性,构建朴素贝叶斯分类模型训练数据集 $T_{\text{Bayes}} = \{(d_1, s_1), (d_2, s_2), \dots, (d_k, s_k), \dots, (d_N, s_N)\}$,其中 d_k 和 s_k ($k = 1, 2, \dots, N$)分别代表第 k 个博文样本及其主客观类别; $s_k \in \{0, 1\}$,“1”代表主观类,“0”代表客观类。基于 T_{Bayes} 采用极大似然法估计博文主客观类别的先验概率 $P(s_k = 1)$,计算方法为将属于主观类的博文数除以博文样本总数。

$$P(s_k = 1) = \frac{\text{num}(s_k = 1)}{N}, \quad k = 1, 2, \dots, N \quad (2)$$

其次,基于 T_{Bayes} 采用极大似然法估计条件概率 $P(f^{(l)}|s_k = 1)$,即主客观特征 $f^{(l)} (l \in \{1, 2, \dots, n\})$ 出现在主观类博文中的概率,计算方法为将特征 $f^{(l)}$ 在主观类博文中的出现次数除以主观类博文中所有特征的出现次数之和。

$$P(f^{(l)}|s_k = 1) = \frac{\text{frequency}(f^{(l)}|s_k = 1)}{\sum_{l=1}^n \text{frequency}(f^{(l)}|s_k = 1)}, \quad k = 1, 2, \dots, N \quad (3)$$

同理,可以估计博文属于客观类的先验概率 $P(s_k = 0)$ 、主客观特征 $f^{(l)} (l \in \{1, 2, \dots, n\})$ 出现在客观类博文中的条件概率 $P(f^{(l)}|s_k = 0)$ 。

最后,对于给定的待分类博文文本向量 \mathbf{d} ,可基于其含有的 n 个主客观特征 $f^{(l)} (l \in \{1, 2, \dots, n\})$,计算其属于主观类的概率 $P(s_k = 1|\mathbf{d})$ 和其属于客观类的概率 $P(s_k = 0|\mathbf{d})$ 。

$$P(s_k = 1|\mathbf{d}) = (P(s_k = 1) \prod_{l=1}^n P(f^{(l)}|s_k = 1)) / (P(s_k = 1) \prod_{l=1}^n P(f^{(l)}|s_k = 1) + P(s_k = 0) \prod_{l=1}^n P(f^{(l)}|s_k = 0)), \quad k = 1, 2, \dots, N \quad (4)$$

$$P(s_k = 0|\mathbf{d}) = (P(s_k = 0) \prod_{l=1}^n P(f^{(l)}|s_k = 0)) / (P(s_k = 1) \prod_{l=1}^n P(f^{(l)}|s_k = 1) + P(s_k = 0) \prod_{l=1}^n P(f^{(l)}|s_k = 0)), \quad k = 1, 2, \dots, N \quad (5)$$

比较 $P(s_k = 1|\mathbf{d})$ 和 $P(s_k = 0|\mathbf{d})$ 的值,当

$$P(s_k = 1|\mathbf{d}) > P(s_k = 0|\mathbf{d})$$

则博文 \mathbf{d} 为主观意愿较大的观点博文,反之则为非观点博文。

2.4 博主观点支持率计算

真正的意见领袖不仅输出观点,而且其大部分观点会在群体中获得广泛的认可和支持,从而形成舆论引导力。对于已经筛选出的观点博文,其观点是否获得大众支持往往体现在博文对应的评论中,正向评论表明评论者持有与博主一致的立场,而负向评论和中立评论表明评论者持有与博主相反的立场或者无关的立场。因此,为了计算博文的观点支持度,可以对其评论文本的情感极性进行分析,当正向情感评论数大于非正向情感评论数时,可认为博文的观点获得了较大的支持。

前人研究表明,在针对微博评论短文本的机器学习分类算法中,支持向量机(Support Vector Machine, SVM)的情感分类效果较好^[15],因此本方法采用基于情感特征的SVM分类模型对观点博文的用户评论情感极性进行计算,具体过程如下。

第1步:对于 B_2 中的每一篇观点博文,获取其对应的所有评论,并从评论文本中提取情感特征。特征词表同样采用BosonNLP情感词典,包含约7万个含有情感色彩的词语和网络用语。

第2步:通过Word2Vec方法对第1步选取的情感特征进行向量化,并将每篇评论的情感特征词向量相加,构成该评论的文本向量 $\mathbf{r} = (t^{(1)}, t^{(2)}, \dots, t^{(m)})$,其中 $t^{(1)}, t^{(2)}, \dots, t^{(m)}$ 为 \mathbf{r} 具备的 m 个情感特征。

第3步:构建SVM分类模型。首先,从现有的评论向量中选择一部分样本,人工标注其情感极性,构建SVM分类模型训练数据集 $T_{\text{SVM}} = \{(r_1, c_1), (r_2, c_2), \dots, (r_i, c_i), \dots, (r_M, c_M)\}$,其中 r_i 和 c_i ($i = 1, 2, \dots, M$)分别代表第 i 个评论样本及其情感极性; $c_i \in \{-1, 1\}$,“1”代表支持博文的正向评论,“-1”代表不支持博文的非正向评论。采用径向基函数(Radial Basis Function, RBF)作为SVM的核函数 K ,即

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) \quad (6)$$

构造最优化问题:

$$\min_{\alpha} \left(\frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j c_i c_j K(r_i, r_j) - \sum_{i=1}^M \alpha_i \right) \quad (7)$$

$$\begin{cases} \text{s.t.} & \sum_{i=1}^M \alpha_i r_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, M \end{cases}$$

基于 T_{SVM} , α 为解, i, j 为第 i, j 个评论样本,选择合

适的参数 C 值,求得最优解 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_M^*)^T$ 。再选择 α^* 的一个正分量 $0 \leq \alpha_j^* \leq C$,计算

$$b^* = c_j - \sum_{i=1}^M \alpha_i^* c_i K(r_i, r_j) \quad (8)$$

其次,对于给定的评论 r ,可采用以下的分类决策函数判断其情感极性。

$$f(r) = \text{sign}\left(\sum_{i=1}^M \alpha_i^* c_i K(r, r_i) + b^*\right) \quad (9)$$

当 $f(r) = 1$ 时, r 为正向评论;反之当 $f(r) = -1$ 时, r 为负向评论。

最后,对于 B_2 中的每一篇观点博文,计算其观点支持度Support,即该博文的正向评论数占总评论数的比率。

$$\text{Support} = \frac{\text{num}(r|f(r) = 1)}{\text{num}(r)} \quad (10)$$

当Support > 0.5时,认为该博文观点获得了大众的支持,反之则未获得支持。

第4步:对于 B_2 中的每一位博主,在其发表的所有观点博文中进行统计,如果其获支持的观点博文数未达到一定比例,即观点支持率较小,则表明其大部分观点不被大众接受和认可,思想引导性较小,不具备意见领袖的价值,因此将其剔除。最终保留的博主即为突发事件舆情中真正的微博意见领袖。

3 实验

3.1 实验数据和预处理

本文选取2020年我国发生的重大舆情突发事件——“杭州女子失踪案”作为微博意见领袖识别方法的验证案例。2020年7月5日,在浙江省杭州市江干区三堡北苑小区发生一起女子离奇失踪案件。同月23日,杭州公安发布通报:“杭州女子离奇失踪案”侦破取得重大突破,失踪女子已遇害,嫌疑人是其丈夫许某。该事件被网络媒体报道后,迅速获得广泛关注,并引发网络用户“全民推理破案”。通过查询百度指数发现,网民对事件舆情的搜索和关注从2020年7月16日起始,7月24日案件真相曝光后达到最大热度,7月28日之后逐渐降低。

在该事件舆情的全生命周期中,微博成为传播最新案情和网民参与讨论的主要渠道之一,相关话题频繁登上微博热搜榜,大量“微博大V”进行了事件的报告和评价,积极充当意见领袖的角色。为了更好地识别其中真正的意见领袖,本文使用后羿采集器工具,以“杭州失踪”为关键词,搜索并抓取2020年7月16日~8月4日间的热门微博数据,共获得1 673位

博主发布的3 217条博文,剔除评论数和转发数都在10次以下的博文及其博主信息,最终保留1 442位博主、2 740条博文作为初始数据集。

3.2 实验流程

3.2.1 信息影响力计算

运用2.2小节中表3的指标模型计算博主信息影响力值。由于各评价指标的具体数值相差巨大,为了便于指标的比较,需要先统一量纲,使用式(11)对各指标值进行非线性归一化。

$$x^* = \lg(x+1)/\lg x_{\max} \quad (11)$$

其中, x^* 为归一化后的指标值, x 为原始指标值, x_{\max} 为该指标中的最大值。

指标值进行归一化处理,按指标模型进行加权求和计算,获得信息影响力值,并从高到低进行排序。

由于模型中的部分指标(转发数和评论数)与具体的博文相关,因此当博主针对事件发表多篇博文时,每篇博文反映的博主信息影响力都会被单独计算。本文设置影响力阈值为0.6,即信息影响力大于0.6的博文及其博主将被保留参与下一步的筛选,最终共有127位博主的369条微博博文被保留。表4展示了部分被保留的博文、博主及其影响力值。

3.2.2 观点输出性计算

从事件相关博文中随机抽取1 200条,人工标注其主客观性。标注时采用4人策略,即先由A和B各自分开对全部博文进行标注,然后对比找出A和B标注不一致博文,再由C和D共同讨论确定其主客观性。最终共标注主观类博文550条,客观类博文640条,删除无法判断的模糊博文10条。

实验采用Pycharm开发环境,Python3.7编程语言。首先调用结巴分词包对博文文本进行自动分词,按照2.3节中确定的主客观特征词表抽取博文特征词,并使用Word2Vec对其进行向量化。然后调用scikit-learn机器学习工具中的naive_bayes包,输入标注好的博文数据,使用高斯朴素贝叶斯算法训练主客观分类模型。采用10折交叉验证的方式测试模型的准确性,即轮流按9:1的比例划分训练集和测试集,取10次测试结果的准确率平均值作为该模型的准确率,最终确定该模型的平均准确率为84.3%。

模型构建完成后,输入上一步保留的高影响力博文数据,以计算其观点输出性,结果表明其中有69位博主的80篇博文含有观点,表5展示了部分博主及其博文观点输出性。

从分类结果可以看出,尽管都具有较高的影响

表4 高信息影响力博文及博主示例

Table 4 Examples of high-influence blog posts and bloggers

博主	博文概括	归一化数值				
		粉丝数	原创微博数	被评论数	被转发数	影响力值
人民日报	警方公布失踪遇害女子丈夫被抓捕视频	0.98	0.93	1.00	1.00	0.98
人民日报	杭州警方公布搜证侦破经过:民警抽取38车粪水筛查发现人体组织	0.98	0.93	0.94	0.98	0.96
中国新闻周刊	杭州失踪女子亲属在化粪池旁痛哭	0.94	0.94	0.97	0.83	0.92
头条新闻	警方在小区化粪池发现杭州失踪女子尸体	0.96	0.96	0.95	0.82	0.91
央视新闻	杭州失踪女子丈夫交代杀人分尸事实	0.98	0.93	0.91	0.83	0.91
头条新闻	失踪女子已在化粪池找到? 警方:系谣言,以通报为准	0.96	0.96	0.91	0.79	0.89
1818黄金眼	女子深夜离奇失踪,最新进展:昨晚警方对小区化粪池进行排查	0.83	0.86	0.91	0.91	0.88
头条新闻	杭州女子蹊跷失联18天,警方辟谣:网传警方通报是假的	0.96	0.96	0.89	0.76	0.88
央视新闻	视频还原杭州女子失踪案侦破过程	0.98	0.94	0.88	0.76	0.87
杭州公安	杭州公安召开杭州江干来某某失踪案通气会	0.72	0.81	0.96	0.89	0.86

表5 博主及其博文的观点输出性示例

Table 5 Example output of opinions of bloggers and their blog posts

博文序号	博主	博文正文(摘录)	主客观类别	观点输出性
1	中国新闻网	#化粪池警告的幽默很恶臭#【化粪池警告? 同款绞肉机? 这种“幽默”很恶臭!】#壹图#鲁迅先生曾说,人类的悲欢并不相通。许多前辈也曾告诉过我们,做人要有“同理心”	主观	有观点
2	搜狐视频	【男子杀妻分尸血迹渗透邻居天花板,部分尸块冲进厕所,被判死缓……】杭州失踪女子被杀案正在进一步侦办中,从嫌疑人作案手段和警方侦破方式中发现,颇似6年前曾经发生在芜湖的一起失踪女子被杀案	客观	无观点
3	封面新闻	【#失踪遇害女子丈夫被抓前最后画面#:经过吸粪车时曾两次看向化粪池】7月22日,浙江杭州。随着警方调查的深入开始对小区化粪池进行排查,傍晚6点46分一个熟悉的身影出现。许某某穿着黑色T恤,黑色裤子白色鞋子步行走了出来	客观	无观点
4	头条新闻	【数据解读:近300起杀妻案如何判罚】杭州来女士失踪案的真相竟是被丈夫杀害。@图数室分析了中国裁判文书网上公开的近300起杀妻案件,发现了这些事实	主观	有观点
5	都市快报	#杭州女子在家睡觉时离奇失踪# 52岁的来女士住在杭州三堡北苑,7月4号晚上,一家人像往常一样入睡,到了凌晨5点半,丈夫许先生发现妻子不见了	客观	无观点

力,但博主发布博文的目的并不完全相同,主观性较大的博文用于输出表达博主观点和态度,而客观性较大的博文用于陈述和报道事实。以表5中序号1~4的博文为例,分析实际数据中博文观点输出性的差异。

序号1的博文:博主首先贴出了“化粪池警告的幽默很恶臭”这个论点,然后引用鲁迅先生以及前辈的话作为论据,表明对用他人生命恶意造梗的风气的批评与反对。

序号2的博文:陈述了杭州失踪女子被杀案疑似6年前芜湖失踪女子被杀案这一事实,没有输出任何观点。

序号3的博文:陈述了失踪遇害女子丈夫经过吸粪车时,曾两次看向化粪池这一事实,也没有输出任何观点。

序号4的博文:博主看似是用数据事实说话,但是“竟”字表明了博主对杭州女子失踪案真相的震惊,同时博主分析了过往的类似案件,认为过往对凶手的处罚偏轻。

从以上分析可以看出,有观点输出的博文更能

反映出博主引导舆论走向的意图,也是其成为意见领袖的关键之一。

3.2.3 观点支持率计算

对上一步保留的80篇观点博文,获取其评论文本,清除其中的空数据、“@某人”但无评论的数据、纯表情和纯符号以及无意义的数字,最终获得43 368条评论数据。随机选取其中的5 000条评论进行人工标注:将正向评论标注为1,非正向评论标注为-1。由于数据较多,标注时分2组进行,每组标注2 500条并同样采用4人策略。最终获得正向评论2 148条,非正向评论2 852条。

采用和上一步相同的Python环境,首先调用结巴分词包对评论文本进行分词,按照2.4节中确定的情感特征词表抽取评论的特征词,并使用Word2Vec对其进行向量化。然后调用scikit-learn机器学习工具中的SVM包,采用RBF(Radial Basis Function)核函数,输入标注好的评论数据训练情感极性分类模型。采用10折交叉验证的方式测试模型的准确性,最终该模型的平均准确率为83.7%。

模型构建完成后,输入剩余未标注的评论数据,获得每篇评论的情感极性,再依据式(11)计算每篇博

文的观点支持度,并判断其是否获得支持。表6显示了部分博文及其博主的观点支持度。

表6 评论对博文的支持情况示例
Table 6 Support rate of comments for blog posts

博文序号	博主	博文正文(摘录)	发帖时间	评论数	支持度	观点支持情况
1	韩东言	这是疑案,不是灵异。这个世界,不存在没有痕迹的失踪,大门口的……	07月18日07:29	287	0.18	未获支持
2	八组兔区爆料	杭州失踪女子凭空失踪12天大家都关注了吗,监控显示没离开过小……	07月18日09:46	418	0.86	获得支持
3	我叫IT科技君	52岁的来女士住在杭州三堡北苑,7月4号晚上,一家人像往常一……	07月18日10:30	1 246	0.61	获得支持
4	今日热点大事件	蹊跷! #杭州一女子睡觉时离奇失踪# 多个监控全无踪迹】来女士……	07月18日11:41	1980	0.56	获得支持
5	新浪新闻	7月4日晚,杭州来女士一家像往常一样入睡。来女士的丈夫称到了……	07月18日21:18	508	0.00	未获支持
6	韩东言	清空所有外界干扰,我们现在获取的所有信息,都是来女士丈夫传递……	07月18日21:47	273	0.08	未获支持
7	韩东言	来女士失踪仅穿一条睡衣,钱包、手机、钥匙、身份证都没有带走……	07月18日22:09	2 136	0.08	未获支持
8	三联生活周刊	随着来女士离奇失踪一事持续发酵,事件本身扑朔迷离,加上网络上……	07月19日14:03	348	0.49	未获支持
9	中国新闻周刊	随着来女士离奇失踪一事持续发酵,事件本身扑朔迷离,加上网络上……	07月19日14:13	347	0.54	获得支持
10	杭州生活头条资讯	着来女士离奇失踪一事持续发酵,事件本身扑朔迷离,加上网络上持……	07月19日15:40	2 325	0.60	获得支持

从表6可以看出,尽管有些博主发表了含有相同或相似内容的博文,但是获得的用户支持率却不一样,表明某些博主不被大众认可,并非由于博文输出的观点,而是出于自身口碑或者其他原因,这也导致此类博主的观点无法起到其预期的传播或者引导效果,因此不能作为真正的意见领袖。

进一步的,根据博文的观点支持情况,对上一步保留的69位博主的观点支持率进行统计,并将支持率低于0.5的博主剔除。最终共有40位博主被认定为此次事件中既有观点输出又获得大众支持的真正意见领袖。

3.3 实验结果分析

为了验证本方法识别的微博舆情博主具有意见领袖的本质特征,文本按照“杭州失踪女子案”事件发展的时间线,对识别出的40位微博意见领袖的属性和舆情参与行为进行了观察和分析。

首先,根据意见领袖的微博认证进行其类型的划分和统计,如图2所示。结果表明,40位微博意见领袖中共有11位为官方微博,大部分为从事社会新闻报道评述的官方媒体号,占意见领袖总人数的27.5%,其余29位为个人自媒体号,其认证的类型包括娱乐、新闻、法律、科技、情感、美食、搞笑幽默、作者、演员和音乐人等。

其次,对官方微博和各类自媒体意见领袖在此次突发事件中发表微博的时间和表述的观点进行分析。

图3展示了各类意见领袖发表观点微博的时间点。图中其他类包含:情感、军事、美食、演员、作家、

音乐人和无微博认证的自媒体博主。从发帖时间上可以看出,7月23日~7月28日期间是意见领袖较为集中的发表观点的时间,与百度指数的高热度期吻合。依据杜洪涛等对突发事件网络舆情演化模式的研究^[6],结合意见领袖的舆情参与时间和人次,可将此次事件的微博平台舆情更为精确地分为3个阶段:7月18日~7月22日为舆情的形成(扩散)阶段,7月23日~7月28日为舆情的高潮阶段,7月29日之后为舆情的消散阶段。

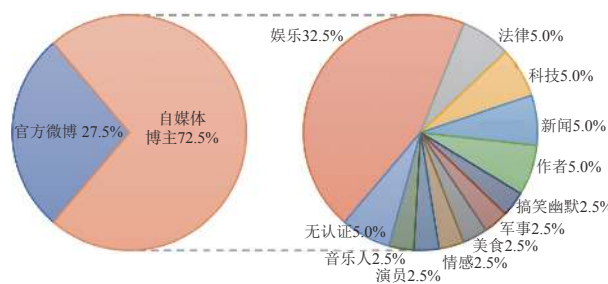


图2 “杭州女子失踪案”微博意见领袖类型占比分布

Fig.2 "A Hangzhou Woman Missing Case" microblog opinion leader type distribution

表7则进一步总结了各微博意见领袖在事件舆情发展中的参与阶段和发表的观点。可以看出事件舆情在高潮阶段出现过2个讨论热点,(1)7月24日某演员在其微博号“郑爽SZ”中发表对事件的评论后,“圈内领袖”“朕硕”“星闻揭秘”“娱乐圈扒姑”“新浪娱乐”“娱舔舔”“八爷圈”“芒果剧集”“天生八卦”“娱乐圈探班”等娱乐类自媒体号纷纷跟进,支持其在热点事件中发声,引发一波舆情热度。(2)在7月25日,杭州市公安局召开记者会,确认许某因家庭生

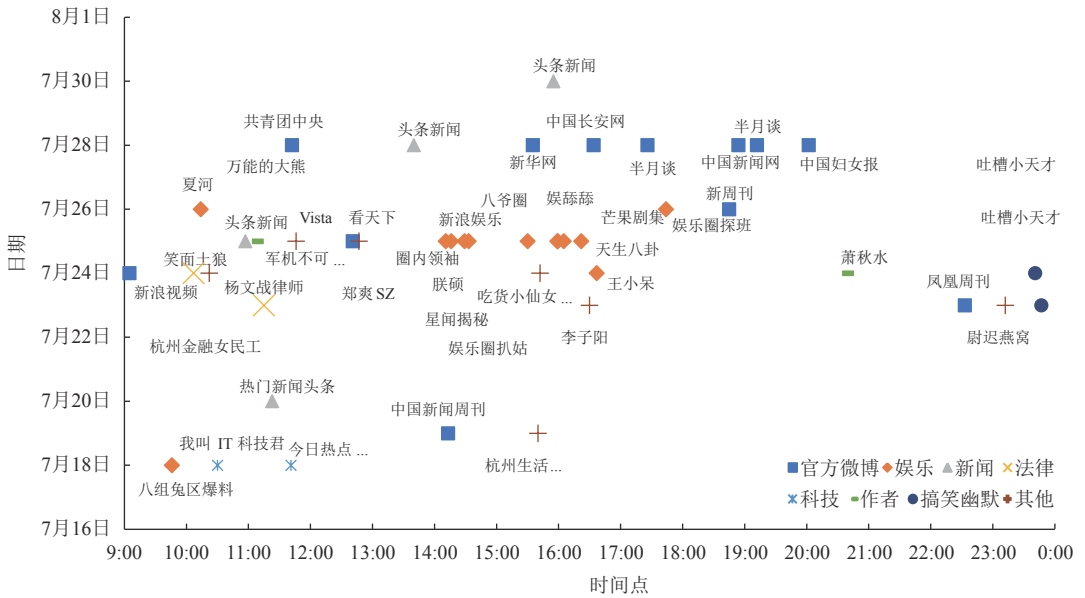


图3 意见领袖发表观点微博的时间序列

Fig.3 Time series of opinions expressed by opinion leaders on microblog

表7 “杭州女子失踪案”微博意见领袖参与阶段及观点

Table 7 "A Hangzhou Woman Missing Case" microblog opinion leaders' participation stage and views

意见领袖	发表观点	舆情演化周期
八组兔区爆料、我叫IT科技君、今日热点大事件	对女子离奇失踪进行猜想	形成(扩散): 7月18日~7月22日
中国新闻周刊、杭州生活头条资讯	批评网红直播的“蹭热点”行为,反对网红对小区的过度关注	
热门新闻头条	赞扬记者的“仁爱之心”	
杨文战律师	希望官媒、大V注意维护自己的公信力	高潮: 7月23日~7月28日
凤凰周刊、吐槽小天才、新浪视频、吃货小仙女de日常、头条新闻、万能的大熊、Vista看天下、郑爽SZ	表达了对案件真相的震惊与难过	
李子阳	旧案重提,希望重新寻找失踪母亲	
王小呆	旧案重提,希望帮助别人寻找失踪母亲	
尉迟燕窝	反对娱乐化“凶案”	
笑面土狼、杭州金融女民工	探讨凶手心理	
萧秋水	提醒女性 in 家庭关系中保护自己,希望减少类似事情的发生	
夏河	感叹婚姻和爱情相互矛盾	
军机不可泄露	反对恶意造谣凶手身份	
圈内领袖、朕硕、星闻揭秘、娱乐圈扒姑、新浪娱乐、娱舔舔、八爷圈、芒果剧集、天生八卦、娱乐圈探班	支持明星为热点事件发声	
新周刊、共青团中央、新华网、中国长安网、半月谈、中国新闻网、中国妇女报	反对恶意造梗,消遣惨剧	消散: 7月29日之后
头条新闻	认为过往杀妻案判罚偏轻	

活矛盾于7月5日凌晨杀害来某,并分尸抛弃。之后几日内社交平台悄悄涌现出“XXX警告”等所谓“网络新梗”,恶意渲染、调侃案件细节,消遣惨剧。这些言语和行为不仅挑战公序良俗的底线,更对社会舆情场造成了极为严重的误导效应。针对这一“恶意玩梗”现象,一些官方微博和自媒体微博大号,如“新周刊”“共青团中央”“头条新闻”“新华网”“中国长安网”“半月谈”“中国新闻网”“中国妇女报”等及时进行了谴责,安抚群众情绪,引导舆情正常化。

从舆情治理的角度上看,识别网络意见领袖的目的在于2个方面:(1)可以通过分析意见领袖的行为判断事件舆情的热点和走向;(2)可以通过意见领袖对事件舆情做正面的引导。从上述针对真实突发事件的实验结果上看,本方法识别出的微博意见领袖无论属于官方微博还是自媒体,都明显注重追踪事件的发展,善于捕捉事件中的新情况和敏感点,并在事件进程的关键节点处发声。加之其博文具有较强的观点导向性,借助其在庞大粉丝群的影响力和

话语权,往往能有效地推动舆情热点的形成和发展。其中,一些官方微博之所以能成为事件中的意见领袖,除了其自身粉丝体量大以外,更多的是由于其输出了强烈而鲜明的是非观点,而其舆论引导作用也非常明显,特别在一些社会道德问题上,其观点和态度能有效地帮助民众明辨是非,形成对社会丑恶现象的舆论批判,消除其不良影响。此次典型事件中,在恶意造梗现象出现后,大批官微在7月28日集中发声批评,促成了网民对此类“网络新梗”的反感和抵制,很快相关话题的热度便降低至消散。此外,自媒体意见领袖在事件中的舆情影响作用也不容小觑,一方面不同类型自媒体博主的参与,扩大了事件舆情的受众面和受关注度,特别是一些娱乐明星的发声,可能会引发大批娱乐营销号的联动,迅速推高舆情热度。另一方面自媒体意见领袖输出的观点自由度较高,增加了事件舆情走向的不确定性,某些观点虽然获得了网民的支持,但客观上也加剧了民众对事件的负面情绪,因此需要舆情治理者密切关注并及时疏导。总之,本方法通过运用观点挖掘技术,能够较为有效地过滤掉“陈述报道事件事实”和“观点不具备影响力”的伪微博意见领袖,挖掘出更具备治理价值的核心意见领袖集合。

3.4 实验结果的验证

为了进一步验证实验结果的科学性,将本文通过观点挖掘法识别的“杭州女子失踪案”事件的微博意见领袖与通过社会网络分析法、专家人工分析法识别的微博意见领袖进行了对比,对比结果见表8。

社会网络分析法是经典的意见领袖识别方法,考虑微博的转发和评论,通过比较微博节点入度、出度以及各类型中心度代表节点的重要性;专家人工分析法是通过专家对事件各维度影响因素进行综合考虑排序,人工确定意见领袖。邀请了20位专家进行综合评分,其中包括3名新闻传播领域专家,7名网络舆情领域的研究学者,以及10位高校研究网络舆情的博士研究生。因观点挖掘法共识别出39位意见领袖,所以选择社会网络分析法、专家人工分析法识别前39位意见领袖,按同样的规则降序排列。通过表8可以发现观点挖掘分析法识别的意见领袖与专家人工分析的结果更为相似;社会网络分析法则更关注信息的传播力,体现在识别的意见领袖均为具有高转发量的博主,而本文方法识别的意见领袖则更具有实际舆情影响力,与在事件舆情进程关键点发挥作用的重点用户高度吻合,真实引导着舆论的发展趋势,是真正意义上的意见领袖。对比发现,本

表8 各意见领袖识别方法识别结果比较

Table 8 Comparison of the identification results of various opinion leaders

观点挖掘分析法	社会网络分析法	专家人工分析法
中国长安网	中国新闻周刊	中国长安网
中国新闻周刊	中国新闻网	中国新闻网
中国新闻网	中国妇女报	中国妇女报
中国妇女报	郑爽SZ	郑爽SZ
郑爽SZ	杨文战律师	朕硕
朕硕	央视新闻	娱舔舔
娱舔舔	新周刊	娱乐圈扒姑
娱乐圈探班	新浪视频	杨文战律师
娱乐圈扒姑	笑面土狼	星闻揭秘
杨文战律师	小强热线-浙江教科	新周刊
星闻揭秘	夏河	新闻晨报
新浪娱乐	尉迟燕窝	新浪娱乐
新浪视频	微天下	新浪视频

文提出的突发事件微博意见领袖识别方法更具有科学性与实用性。

4 结论

本文在对网络意见领袖的定义进行解析后,依据其本质提出了一个新的针对突发事件的微博意见领袖识别方法,通过综合运用指标模型方法和观点挖掘技术,对高信息影响力、输出观点、观点获支持等意见领袖的本质特征进行量化评价,形成了一套更完整的评价标准及模式。以“杭州女子失踪案”事件为例对本文提出的方法进行实验验证,结果表明,本方法发现的意见领袖更符合意见领袖的定义,是用户真实支持的有舆论引导力的意见领袖。

本研究的局限性在于:大数据环境下,网络舆情的传播方式和传播途径多元化,而本文未考虑到多平台多渠道的信息流动。同时,由于微博上部分博主存在筛选评论和事后删除微博的行为,可能导致采集到的数据出现偏差,对最终结果的准确性有一定的影响。

参考文献:

- [1] LAZARSFELD P F, BERELSON B, GAUDET H. The people's choice[M]. New York: Columbia University Press, 1948.
- [2] 王佳敏, 吴鹏, 陈芬, 等. 突发事件中意见领袖的识别和影响力实证研究[J]. 情报学报, 2016, 35(2): 169-176.
WANG J M, WU P, CHEN F, et al. Empirical study on recognition and influence of opinion leaders in emergency [J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(2): 169-176.

- [3] 彭丽徽,李贺,张艳丰.基于灰色关联分析的网络舆情意见领袖识别及影响力排序研究——以新浪微博“8·12滨海爆炸事件”为例[J].情报理论与实践,2017,40(9):90-94.
PENG L H, LI H, ZHANG Y F. Research on the identification and influence ranking of network public opinion leaders based on grey relational analysis [J]. *Intelligence Theory and Practice*, 2017, 40(9): 90-94.
- [4] 吴江,赵颖慧,高嘉慧.医疗舆情事件的微博意见领袖识别与分析研究[J].数据分析与知识发现,2019,3(4):53-62.
WU J, ZHAO Y H, GAO J H. Research on Microblog opinion leaders identification and analysis in medical public opinion incidents [J]. *Data Analysis and Knowledge Discovery*, 2019, 3(4): 53-62.
- [5] 陈芬,陈佩帆,吴鹏,等.融合用户特征与多级文本倾向性分析的网络意见领袖识别[J].情报理论与实践,2018,41(7):143-148.
CHEN F, CHEN P F, WU P, *et al.* Online opinion leader identification method integrating user feature and multilevel text sentiment analysis [J]. *Information Theory and Practice*, 2018, 41(7): 143-148.
- [6] 安璐,胡俊阳,李纲.基于主题一致性和情感支持的评论意见领袖识别方法研究[J].管理科学,2019,32(1):3-13.
AN L, HU J Y, LI G. A method of identifying opinion leaders based on topic consistency and emotional support [J]. *Journal of Management Science*, 2019, 32(1): 3-13.
- [7] 孙乃利,王玉龙,沈奇威.微博客意见领袖识别的研究[J].*电信技术*,2012(12):78-80.
SUN N L, WANG Y L, SHEN Q W. Research on the identification of microblog opinion leaders [J]. *Telecommunications Technology*, 2012(12): 78-80.
- [8] 魏志惠,何跃.基于信息熵和未确知测度模型的微博意见领袖识别——以“甘肃庆阳校车突发事件”为例[J].情报科学,2014,32(10):38-43.
WEI Z H, HE Y. Identify microblogging opinion leaders based on information entropy and unascertained measure model—taking “emergencies of Qingyang school bus” as an example [J]. *Information Science*, 2014, 32(10): 38-43.
- [9] 胡若涵,张慧明.中兴通讯案:中美科技贸易摩擦的意见领袖识别研究[J].中国经贸导刊(中),2018(23):11-13.
HU R H, ZHANG H M. ZTE case: research on the identification of opinion leaders in Sino-US technology trade friction [J]. *China Economic and Trade Guide (Part 2)*, 2018(23): 11-13.
- [10] 陈芬,高小欢,彭玥,等.融合文本倾向性分析的微博意见领袖识别[J].数据分析与知识发现,2019,3(11):120-128.
CHEN F, GAO X H, PENG Y, *et al.* Text sentiment analysis and its application in Microblog opinion leader identification [J]. *Data Analysis and Knowledge Discovery*, 2019, 3(11): 120-128.
- [11] WIEBE J, WILSON T, BRUCE R, *et al.* Learning subjective language [J]. *Computational Linguistics*, 2004, 30(3): 277-308.
- [12] 杨武,宋静静,唐继强.中文微博情感分析中主客观句分类方法[J].重庆理工大学学报,2013,27(1):51-56.
YANG W, SONG J J, TANG J Q. A study on the classification approach for Chinese Microblog subjective and objective sentences [J]. *Journal of Chongqing University of Technology*, 2013, 27(1): 51-56.
- [13] 丁晟春,吴靓婵媛,李红梅.基于SVM的中文微博观点倾向性识别[J].*情报学报*,2016,35(12):1235-1243.
DING S C, WU L C Y, LI H M. Chinese Microblogging opinion recognition based on SVM model [J]. *Journal of the China Society for Scientific and Technical Information*, 2016, 35(12): 1235-1243.
- [14] 刘培玉,旬静,费绍栋,等.基于隐马尔可夫模型的主观句识别[J].中文信息学报,2016,30(4):206-212.
LIU P Y, XUN J, FEI S D, *et al.* Subjective sentence recognition based on hidden Markov model [J]. *Journal of Chinese Information Processing*, 2016, 30(4): 206-212.
- [15] 谭有新,滕少华.短文本特征的组合加权方法[J].*广东工业大学学报*,2020,37(5):51-61.
TAN Y X, TENG S H. Combined weighting method for short text features [J]. *Journal of Guangdong University of Technology*, 2020, 37(5): 51-61.
- [16] 杜洪涛,王君泽,李婕.基于多案例的突发事件网络舆情演化模式研究[J].*情报学报*,2017,36(10):1038-1049.
DU H T, WANG J Z, LI J. Research on evolution Model of online public opinion in emergencies events based on multiple cases [J]. *Journal of the China Society for Scientific and Technical Information*, 2017, 36(10): 1038-1049.

(责任编辑:张玮欣)