



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

安璐^{1,2}, 胡俊阳², 李纲¹

1 武汉大学 信息资源研究中心, 武汉 430072

2 武汉大学 信息管理学院, 武汉 430072

摘要:社交网络上兴起的意见领袖深刻影响着大众,而评论区的高影响力用户越来越引起人们的重视,其评论可能会引导突发事件的舆论走向,对事件的后期演变和结果产生积极或消极影响,识别评论区的意见领袖对于掌握突发事件的网络舆情、引导舆论走向等方面具有重要的实践意义。

提出基于主题一致性和情感支持的网络评论意见领袖识别方法,以山东辱母杀人事件为例,使用基于Relevance公式改进的潜在狄利克雷分配模型提取微博评论主题,根据可视化输出中评论与其回复所属主题节点之间的距离远近进行主题校正;通过基于情感词典的情感分析,计算评论和回复的情感值,确定回复者对评论者的情感支持态度;构建意见领袖特征指标组,包括获点赞数和获回复数2个1级指标以及获回复数下的直接获回复数、间接获回复数、直接回复数3个2级指标;赋予1级指标权重,计算意见领袖值,根据评论者的意见领袖值的高低确定具有正面和负面高影响力的微博评论意见领袖;从用户主页基本信息和用户评论内容属性两个角度,探讨评论意见领袖的行为特征。

研究表明,评论者的获点赞数、获回复数等指标可以作为用户成为评论意见领袖的基础,根据直接获回复数、间接获回复数和直接回复数之间的相互作用可勾勒出用户在交互中的行为特征,评论内容的效应与用户概况之间没有显著的相关性。提出的网络评论意见领袖识别方法并不局限于只考虑用户本身影响力,而是着重关注评论内容以及用户之间的交互,能有效识别网络评论意见领袖。

研究发现可为管理部门提供网民对突发事件的关注焦点和情感态度,有助于其及时掌握网络舆情,迅速准确地定位网络评论意见领袖,为突发事件应急决策和管理提供有效支持。

关键词:意见领袖;网络评论;主题一致性;情感支持;微博

中图分类号:G203

文献标识码:A

doi:10.3969/j.issn.1672-0334.2019.01.001

文章编号:1672-0334(2019)01-0003-11

引言

近年来,突发事件的爆发往往在社交媒体上引发大量关于该事件的播报和讨论。在这些社交媒体信息中,意见领袖在引导事件的网络舆情走向方面

发挥着重要作用。现有对网络意见领袖的研究绝大多数是分析针对某一事件发表微博并在网络中产生一定影响力的博主,很少有研究涉及对博主的博文进行评论的高影响力评论用户。现在有一种普遍现

收稿日期:2018-07-04 **修返日期:**2018-12-03

基金项目:教育部哲学社会科学研究重大课题攻关项目(17JZD034);国家自然科学基金(71790612,71603189)

作者简介:安璐,管理学博士,武汉大学信息资源研究中心和信息管理学院教授,研究方向为可视化知识发现和网络数据分析等,代表性学术成果为“Visual topical analysis of Chinese and American library and information science research institutions”,发表在2014年第1期《Journal of Informetrics》,E-mail:anlu2009@whu.edu.cn

胡俊阳,武汉大学信息管理学院硕士研究生,研究方向为网络舆情分析等,E-mail:hujy2015@126.com

李纲,管理学博士,武汉大学信息资源研究中心教授、教育部长江学者特聘教授,研究方向为信息资源管理和网络舆情分析等,代表性学术成果为“科研合作超网络下的知识扩散演化模型研究”,发表在2017年第3期《情报学报》,E-mail:imiswhu@aliyun.com

象,即博主博文的评论区有许多网友根据自己的立场和看法进行讨论,在讨论过程中产生观点的分化并形成代表人物。根据两级传播理论^[1],意见领袖通常是作为信息与受众者的中间传递人,因此可以将扮演中间传递人角色的评论区代表用户视为评论意见领袖。与普通网络意见领袖直接针对事件发表微博不同,评论意见领袖活跃于微博的评论区,结合事件和博主博文的内容发表评论,并与其他网民产生交互,评论观点可能会影响他人判断、改变他人认知。发现评论区的意见领袖是发现网络意见领袖的一种新思路,本研究通过识别与评论主题一致且情感上支持的回复,构建评论意见领袖识别体系,并探讨评论意见领袖的行为特征,为管理部门寻求在公共事件中如何利用网络意见领袖有效引导网络舆情的策略提供参考。

1 相关研究评述

1.1 意见领袖识别

意见领袖的概念最早是由 LAZARSFELD et al.^[1]于20世纪50年代提出,他们认为意见领袖不仅是信息的提供者,还在一定程度上能够对他人产生影响;HOLLANDER^[2]于1961年提出一种新的定义意见领袖的方法,即意见领袖最好是以其影响追随者态度和行为的能力来定义;ROGERS^[3]总结了意见领袖的特点,并认为意见领袖必须为他人做出示范。相对于普通群众,意见领袖的特点在于拥有更高的社会地位、更好的教育背景、更强的应对挑战的能力^[4]。意见领袖不仅将普通群众的注意力集中于特定问题,而且能暗示群众应该以怎样的回应和行动面对该问题^[5]。发展至今,对于意见领袖有许多不同的定义,但是所有定义的核心都集中在意见领袖的“中介”作用上。传统上,意见领袖社会地位更高,并与社会事务大量接触,他们比非领袖更频繁地注意新闻媒体的内容^[6]。意见领袖可以是社区、群组或社会中有影响力的成员,他人可向其寻求建议、观点和看法^[7]。意见领袖能够塑造他人想法,并且对普通群众具有影响力^[8]。意见领袖的影响过程类似于一个金字塔,一些极有影响力的人占据了顶层,其影响力作用于下层的群众^[9]。

目前,即时通信工具功能不断改进,与天涯论坛等传统网络社区功能相融合,随着Twitter、新浪微博等带有即时性特点的社交应用的兴起,出现了越来越多的网络平台使网络意见领袖充分发挥其自身影响力,豆瓣、新浪微博、网易新闻等网络应用也逐渐成为网络意见领袖聚集的空间^[10]。网络意见领袖是指将互联网作为主要的活动场所,通过互联网技术支持的各项服务为普通网民提供意见和信息,能够回答网民在某些领域提出的问题,在网民之中具有一定影响力,并在特定时间、空间条件下能够制造和引领社会舆论的一类人^[10]。国外对网络意见领袖的研究早于中国,不仅提出基于在线社区网站的网络意见领袖影响力扩散模型^[11],还以商品的推广情况

为例分析发现网络意见领袖所具有的明显不同于传统意见领袖的特征^[12]。中国对网络意见领袖的代表性研究类型主要有采用数据挖掘等方法筛选网络论坛舆论领袖并从多个维度刻画论坛舆论领袖的特点^[13]、构建指标体系以建立舆论领袖影响力传播模型^[14]、识别社交网络意见领袖^[15]。

考量每一次公共事件的网络舆情演化过程,几乎都有网络意见领袖参与其中并发挥重要作用^[16]。目前国内外对网络意见领袖的研究主要针对其在电子商务活动中的作用,认为意见领袖能够影响消费者的消费倾向^[17]。虽然有学者从微博评论出发识别微博转发网络中的意见领袖^[18],但是对微博评论的研究仍非常少见。相对于现有对意见领袖的研究,本研究以微博评论为研究对象,在已有研究思路的基础上加入回复与原评论的主题一致性,再结合评论回复的情感支持,识别微博评论中有影响力的意见领袖。

1.2 微博文本主题分析

对微博文本的主题发现方法可以大致概括为文本聚类法和主题模型法两大类。相比较而言,由于文本聚类方法是基于统计层面的,很难解决文本中一词多义以及歧义的问题,而主题模型方法对隐含在文本中的主题建模,能够克服文本聚类方法中文档相似度计算方法的缺点,因此主题模型方法在文本主题发现中应用较多^[19]。基于潜在狄利克雷分配(latent Dirichlet allocation, LDA)模型的主题发现在微博文本主题发现的研究中较为瞩目。LDA是一种3层贝叶斯概率模型,通过无监督的学习方法发现文本中隐含的主题信息^[20]。许多学者对LDA的应用进行拓展。LDA-SVM短文本分类流程是将LDA主题模型和SVM分类器结合起来,使用LDA主题模型对短文本的特征项进行扩展,然后将扩展后的特征向量代入到SVM分类器中进行分类^[21];以标签对潜在主题的贡献为出发点,通过Gibbs算法将标签映射到具有特定意义的主题上也是一种行之有效的LDA潜在主题发现方法^[22]。由于传统LDA主题模型是根据词语在某一个主题下的概率高低来判断词语与主题的相关性,而实际上LDA模型识别出的主题及其包含的词语可读性欠佳。鉴于基于Relevance公式改进的LDA主题模型^[23]中的 θ 矩阵可确定文档-主题的分布、 ϕ 矩阵可确定主题-词语分布矩阵,且引入的权重参数 λ 可用于调节词语与主题之间的相关性,因此本研究采用可以改善主题可读性的基于Relevance公式改进的LDA主题模型,并通过LDAVis^[23]进行可视化分析,判断评论与其回复的主题一致性。

1.3 微博文本情感分析

情感分析,又称意见挖掘,由PANG et al.^[24]在2002年提出,通过对文本进行语义分析判断文本的情感极性,主要包括正面、负面、中立3种情感态度。基于情感词典的分析方法和机器学习方法是最为常见的两种情感分析方法^[25]。基于情感词典的分析方法的核心在于“词典”,很大程度上依赖于词典的质

量,目前中国学者在研究微博平台数据时常用的情感词典有台湾大学 NTUSD 中文情感极性词典、大连理工大学情感词典^[26]、HowNet 情感词典等。在基于机器学习的情感分析中经常使用朴素贝叶斯^[27]、支持向量机^[28]、最大熵模型^[29]等经典分类模型,其中多数分类模型的性能依赖于标注数据集的质量,而获取高质量的标注数据需要耗费大量的人工成本^[30]。综合上述讨论,本研究采用基于情感词典的分析方法。此外,网民经常在微博评论及回复中使用表情符号以辅助自身情感的表达,然而随着网络文化的不断变化,微博表情符号的现实含义早已与微博平台对表情符号的定义大相径庭,如“微笑”表情符号本意是用以表达友好亲切,现在多用于嘲讽、无言的情绪表示。因此,本研究构建与当下网络文化情绪传递相吻合的表情符号词典,作为情感词典的补充。

1.4 意见领袖特征指标

许多学者从不同角度构建意见领袖特征体系对潜在意见领袖进行评估。传统上,主贴数、回帖数、总跟帖数是常见的意见领袖特征指标,在此基础上,根据研究对象和研究角度的不同,特征指标的选取也出现一些差异。响应值测量用户对意见领袖的响应类型和响应强度^[31];活跃度、认同度、关注度分别用于揭示意见领袖的发帖频率、受其他用户认同情况和观点的影响扩散程度^[14];支持力则是其他用户对意见领袖支持程度的体现^[32];平均回复长度、平均被回复长度等量化指标也从侧面反映意见领袖与其他用户的交互^[33];还有学者采用威望度测量意见领袖自身影响力的效应^[34]。已有研究对意见领袖特征指标的选择通常会考虑用户本身影响力等因素,并且尽可能涵盖意见领袖所有方面的特征,以期识别结果准确度更高。本研究并非从用户自身影响力出发,而是从用户评论内容出发,探讨评论内容造成的实际影响力。本研究选取的意见领袖特征指标包括获点赞数和获回复数2个1级指标,以及获回复数下的直接获回复数、间接获回复数、直接回复数3个2级指标,以此评估评论意见领袖的影响力。

2 研究框架和方法

2.1 研究框架

通常,用户发布的微博可能收到若干评论,而每

条评论又可能得到若干回复。微博的评论和回复结构见图1。主题分析有助于了解用户的观点,情感分析有助于识别用户的情感极性。本研究基于主题一致性和情感支持构建研究框架,包括数据收集和预处理、评论主题获取、主题校正、情感计算、意见领袖值计算5个步骤,见图2。

数据收集和预处理阶段的主要工作是利用网络爬虫获取新浪微博评论及其评论回复数据,经过预处理后的语料集留作下一步使用;评论主题获取阶段主要使用基于 Relevance 公式改进的 LDA 主题模型获取微博评论的主题;主题校正阶段是将每条评论与其回复结合,再次使用基于 Relevance 公式改进的 LDA 主题模型进行主题分类,并通过 LDAvis 进行可视化分析,利用多维尺度分析法剔除与评论主题不相关的回复;情感计算阶段基于情感词典和扩展词典,计算每条评论和回复的情感值,以此确定回复用户对评论用户的情感支持态度;意见领袖值计算阶段构建意见领袖特征值指标体系,采用标准离差法确定各指标的权重,根据评论用户意见领袖值的高低确定意见领袖。

2.2 基于 Relevance 公式改进的 LDA 主题分析

LDA 是 BLEI et al.^[35]在2003年提出的一个主题模型。为了改善主题的可读性,使用基于 Relevance 公式改进的 LDA 主题模型对微博评论进行潜在主题提取和主题校正。给定一个权重参数, Relevance 计算公式为

$$r(w, k | \lambda) = \lambda \cdot \log(\theta_{k,w}) + (1 - \lambda) \cdot \log\left(\frac{\theta_{k,w}}{p_w}\right) \quad (1)$$

其中, w 为词语, $w \in \{1, \dots, V\}$; k 为主题, $k \in \{1, \dots, K\}$; λ 为权重参数, $0 \leq \lambda \leq 1$; $r(w, k | \lambda)$ 为在指定的 λ 下, 词语 w 与主题 k 的相关程度; $\theta_{k,w}$ 为主题 k 下词语 w 出现的概率分布矩阵; p_w 为词语 w 在主题-词语矩阵 θ 中的边际概率。通常使用变分贝叶斯方法或吉布斯抽样以及 p_w 计算 θ 。可以用 λ 调节词语与主题之间的相关性, 如果 λ 接近于 1, 主题词语的选择是基于在主题下出现越频繁的词语与主题更相关这一思想, 即为传统的 LDA 主题关键词选择方法^[36]; 如果 λ 越接近于 0, 在该主题下更特殊、更独有的词语与主题更相关, 即在该主题下出现的次数较多但在其他主题中很少出现的词语跟主题更相关。

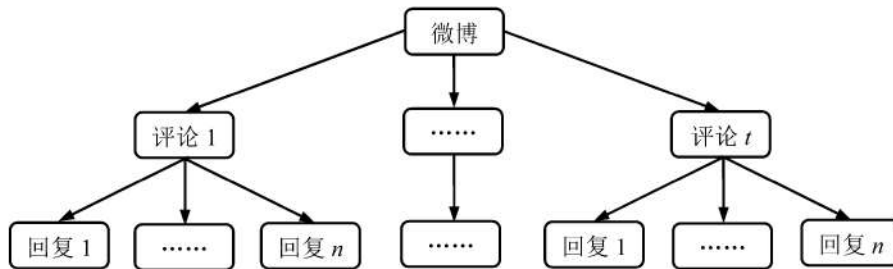


图1 微博评论和回复结构

Figure 1 Structure of Microblog Comments and Responses

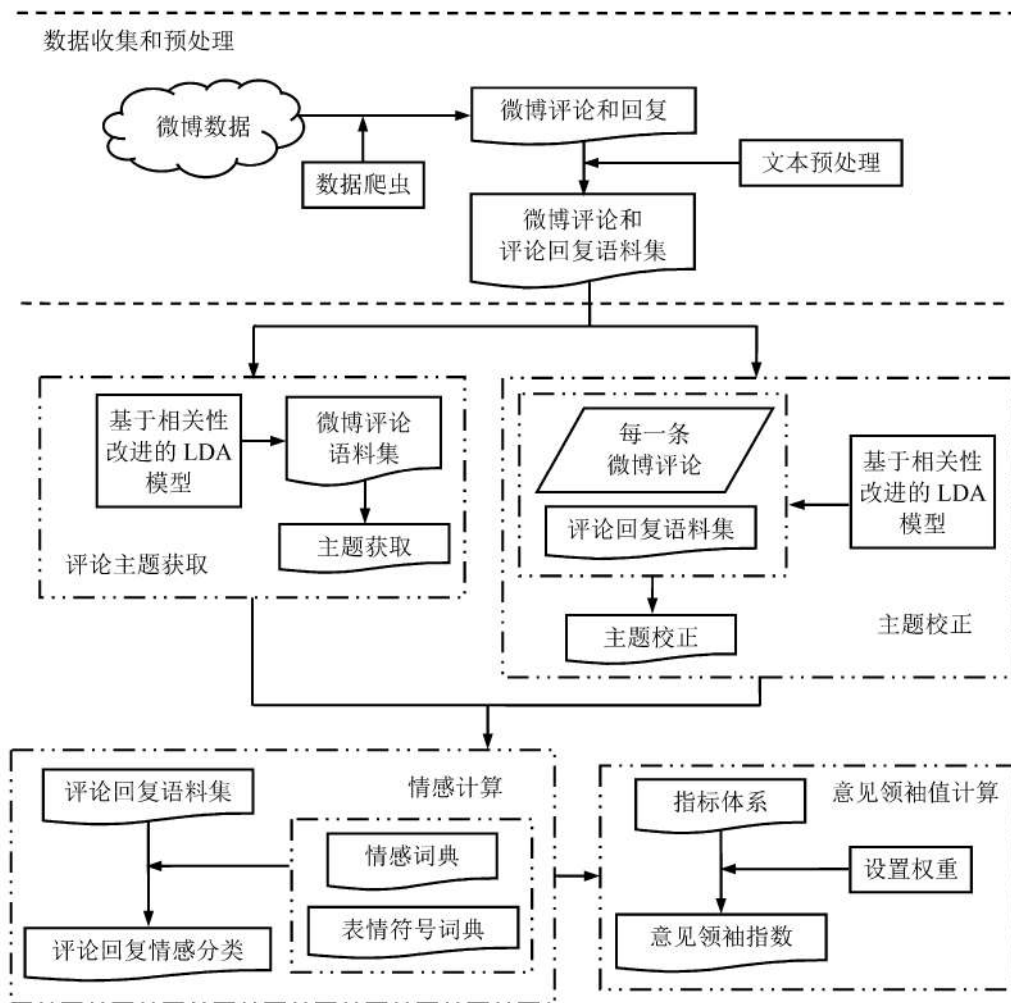


图2 基于主题一致性和情感支持的意见领袖识别总体框架

Figure 2 Opinion Leader Identification Framework Based on Topic Consistency and Emotional Support

2.3 基础情感词典及扩充

本研究通过选取基础情感极性词汇、表情符号、否定词和双重否定词等特征,对微博评论和回复信息进行情感分类。基础情感极性词汇采用台湾大学中文情感极性词典(NTUSD)中被标记为 positive 和 negative 的词表,以及安璐等^[37]在情感分析相关研究中构建的中文微博情感词典;表情符号特征通过构建微博表情符号词典进行匹配;否定词和双重否定词则参考王勇等^[38]构建的否定词典和双重否定词典。

微博平台上的表情符号比语言更直观,它是微博用户经常用来辅助情感表达的一种工具。本研究分析新浪微博自带的表情符号的极性,在已构建好的表情符号词典^[39]的基础上,结合新的网络环境下表情符号的增加和表情符号的情感转变,构建微博表情符号词典。部分新增的表情符号和情感极性发生变化的表情符号见表1。

在汉语表述中经常用到否定词及其多种组合。当使用否定性词语修饰某一词语时,该词的情感极性将发生改变,因此通常使用否定词来表达不同的情感态度;双重否定词语主要表现形式是连用两个

表1 部分新增表情符号和特殊表情符号

Table 1 Some Newly Added Emoji and Special Emoji

情感极性	表情符号
正面情感	[中国赞]、[哆啦A梦花心]、[哆啦A梦微笑]、[伴我同行]、[静音微笑]、[胖虎微笑]、[小夫微笑]、[哆啦A梦笑]、[哆啦A梦美味]、[哆啦A梦开心]、[哆啦A梦亲亲]……
负面情感	[哆啦A梦害怕]、[哆啦A梦吃惊]、[哆啦A梦汗]、[哆啦A梦无奈]、[微笑]、[挖鼻]、[哈欠]、[拜拜]……

否定性词语,与否定性词语不同的是,双重否定有表达肯定、强化语气等功能。构建否定词典时结合实际语料补充21个否定词,新增否定词包括不到、不该、不行、不会、不借、不去、不算、不提、不想、不知、从来不、否、很少、极少、没能、没人、没想到、没用、千万别、勿、只不过。

2.4 微博评论及其回复情感计算

目前许多研究的情感分析涉及到多种情绪分类,如喜、怒、哀、乐等,本研究主要从情感是否支持的角度识别意见领袖,因此对情感倾向性的分析简化为正面情感极性和负面情感极性。本研究认为表情符号表达的情感与情感词表达的情感同样重要,并将表1中的正面情感表情符号与正面情感词汇融合,标注其情感极性值为1;将负面情感表情符号与负面情感词汇融合,标注其情感极性值为-1。使用(2)式修正若干个否定词修饰的情感词的情感极性,即

s'(w_i) = (-1)^φ s(w_i) (2)

其中, w_i 为第 i 个情感词, s'(w_i) 为经否定词修饰后新的情感词极性值, φ 为否定词的个数, s(w_i) 为 w_i 的初始情感极性值。

微博评论及其回复的总体情感值计算公式为

Sen(d) = ∑_{i=1}^m f(w_i) + ∑_{j=1}^n g(x_j) (3)

其中, Sen(d) 为评论及其回复的总体情感值,等于该条微博评论或评论回复的情感词语总情感值与表情符号总情感值之和; m 为情感词的个数, i = 1, ..., m; f(w_i) 为 w_i 经否定词修饰后得到的情感修正值; n 为表情符号的个数, j = 1, ..., n; x_j 为第 j 个表情符号; g(x_j) 为 x_j 的情感值。

2.5 意见领袖识别方法

HOLLANDER[2]认为,意见领袖能够影响追随者的态度和行为, ROGERS[3]认为意见领袖须为他人做出示范, WEIMANN[5]认为意见领袖能够将普通群众的注意力集中于特定问题,且暗示群众应如何回应该问题。因此,本研究借鉴以上学者的研究,将主题一致性和情感支持纳入评论意见领袖识别,即要求评论意见领袖既能持续获得同一主题回复又受到他人的情感支持。由于本研究以评论为切入点,识别的是评论区的意见领袖,考虑到用户在评论区与他人的交互,引入直接获回复数、间接获回复数和直接回复数3个指标。

首先,使用基于 Relevance 公式改进的 LDA 主题模型对微博评论进行潜在主题提取,确定各评论所属主题。然后,分别将每条微博评论与其接收到的回复结合,在 LDA 主题建模之后利用 LDAvis 包生成一个可视化的主题模型,揭示主题下词语的显著性和词语-主题相关性,利用多维尺度分析法提取出主成分的维度,将主题分布到这些维度上,主题相互之间的位置远近表达了主题之间的接近性[40]。通过评论所属主题节点与其回复所属主题节点之间的距离,删除主题相关度较低的评论,实现主题校正。之后再基于情感词典和扩展词典,计算每条评论及其回复的情感值,以此确定回复用户对评论用户的情感支持态度。在意见领袖值计算前构建意见领袖特征指标,包括获点赞数和获回复数2个1级指标,以及获回复数下的直接获回复数、间接获回复数和直接回复数3个2级指标。意见领袖判定标准和特征

指标具体结构见表2,获点赞数指1条评论获得的点赞数,获回复数指1条评论获得的总回复数。

表2 意见领袖特征指标设置和判定标准 Table 2 Characteristic Indices Setting and Criterion for Opinion Leaders

Table with 3 columns: 1级指标, 2级指标, 判定标准. Row 1: 获点赞数, 直接获回复数, ①回复的主题与评论的主题一致;②在所有回复中,支持原评论的回复数量大于等于反对原评论的回复数量.

评论与回复之间的局部结构见图3,①为B回复A,即A的直接获回复;②为C回复B,即A的间接获回复;③为A回复B,即A的直接回复;④为B回复C,即A的间接获回复;⑤为C回复A,即A的直接获回复;⑥为A回复C,即A的直接回复。

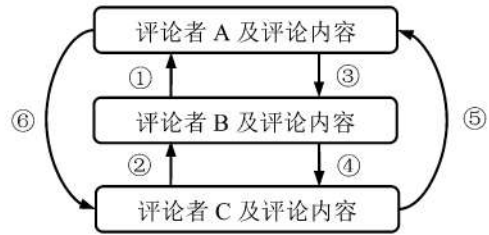


图3 评论与回复的结构示意图

Figure 3 Structure for Comments and Responses

根据图3,本研究给出3个2级指标的含义。直接获回复数指其他用户对该评论回复的数量,直接获回复数 = ① + ⑤ + 类似回复类型;间接获回复数指发生在该评论下但交流双方不包括评论作者的回复的数量,间接获回复数 = ② + ④ + 类似回复类型;直接回复数指该评论作者对他人的回复又进行回复的数量,直接回复数 = ③ + ⑥ + 类似回复类型;获回复数 = 直接获回复数 + 间接获回复数 + 直接回复数。

标准差能反映一个数据集的离散程度[41],通常某个指标的标准差大小反映了其提供的信息量的大小,从而决定该指标在综合评价中所起的作用和权重分配。本研究采用标准离差法确定各指标的权重,首先,以极值法对由指标组构成的数据矩阵的每列(即每个指标)数据进行无量纲化处理;其次,依据标准差计算方法确定获点赞数和获回复数2个指标的标准差;最后,根据1个指标的标准差占有所有指标标准差之和的比重计算该指标的权重。

由于1条评论的获回复数等于该评论的直接获回复数、间接获回复数和直接回复数之和,因此本研究保留获点赞数和获回复数2个1级指标赋予权重计算意见领袖值,根据评论用户意见领袖值的高低确

定意见领袖。意见领袖值的计算公式为

$$Val = W_1 \cdot L_1 + W_2 \cdot L_2 \quad (4)$$

其中, Val 为意见领袖值, W_1 和 W_2 为指标对应的权重, L_1 和 L_2 为经无量纲化处理后的获点赞数或获回复数的值。直接获回复数、间接获回复数和直接回复数用于分析意见领袖值排名靠前的用户的特征。

2.6 意见领袖行为特征分析

本研究从用户主页基本信息和用户评论内容属性两个角度, 探讨评论意见领袖的行为特征。在用户主页基本信息方面, 关注用户的微博账户创建时间、当前等级、是否取得微博认证、关注数、粉丝数、发表微博数等, 以此判断该用户的活跃度和基础影响力; 在用户评论内容属性方面, 除统计评论获点赞数、获回复数(包括直接获回复数、间接获回复数和直接回复数)的数值之外, 还充分讨论直接获回复数、间接获回复数和直接回复数3个2级指标之间的相关性。综合上述两方面的研究, 以实验和数据支撑探讨评论意见领袖的行为特征。

3 实验设计和分析

3.1 实验数据和预处理

本研究以山东辱母杀人事件为例, 将“山东辱母”作为关键词, 利用网络爬虫获取在新浪微博平台上2017年3月1日至2017年12月1日期间所有热门微博的评论及该评论对应的回复信息, 得到2 014条评论的53 564条回复数据。通常意义上, 只有接收到他人的回复的评论用户才可能成为意见领袖, 因此需要剔除评论接收到的回复数为0的用户, 经初步清洗后, 最终有1 547条评论的26 676条回复数据, 平均长度约51个字符。

采用R语言结巴分词^[42]对清洗后的数据进行分词。由于原用户词典的不完善, 没有收录“山东辱母”事件相关词汇、网络流行用语和法律词汇等, 因此本研究首先选用搜狗词库中的搜狗日常用语大词库、搜狗网络流行新词、搜狗法律词汇大全, 再结合图悦生成的辱母事件高频150词, 将上述步骤得到的51 319个词汇一同纳入用户词典并导入分词系统, 分词效果得到极大提升。实验使用哈尔滨工业大学停用词表、四川大学机器智能实验室停用词库和百度停用词等结合的1 893个停用词构成的停用词典删除停用词。

3.2 主题识别与校正

在对所有评论进行分词操作后, 利用基于Relevance公式改进的LDA主题模型确定各主题下的特征词。主题的个数由困惑度确定, 对于不同主题训练出来的模型, 计算其困惑度, 最小困惑度对应的主题个数就是最优的主题数。经试验最优主题数为20, 分别为: 司法判决与道德舆论的平衡(V_1); 微博平台在本案中起到的作用(V_2); 别有用心的人煽动群众情绪(V_3); 致敬于欢的忠孝, 舆论应保护他(V_4); 法律应保护儿子护母的孝顺行为(V_5); 法律丢失了道德底线(V_6); 中小企业融资困难向高利贷借款(V_7);

人民的血性被剥夺(V_8); 案件背后存在“警匪一家”等问题(V_9); 案件暴露的问题上升到更高层面(V_{10}); 出警民警及案件审判人员信息(V_{11}); 出警民警的处理结果存在争议(V_{12}); 不满以法官为代表的法院的判决(V_{13}); 警察对民事纠纷敷衍了事、渎职(V_{14}); 法院冤判, 能否相信法律与正义(V_{15}); 于欢案量刑过重应轻判(V_{16}); 鼓励遭受不公的人反抗(V_{17}); 为不公正事件发声, 敢于奋起反击(V_{18}); 聊城市全面调查案件背后公务人员的问题(V_{19}); 严厉打击山东省内涉黑行为等(V_{20})。

经反复试验, $\lambda = 0.6$ 时各主题之间距离最大, 主题词语的选取效果最好。这20个主题可以分为5个大类: V_1 和 V_2 期望以法律形式保护于欢权益; V_3 、 V_4 、 V_8 和 V_{11} 对此次事件涉及到的政府公务人员的行为展开讨论; V_{13} 和 V_{19} 关注舆论如何推动此次事件的发展; V_{16} 鼓励群众为不公正事件发声; V_5 、 V_6 、 V_7 、 V_9 、 V_{10} 、 V_{12} 、 V_{14} 、 V_{15} 、 V_{17} 、 V_{18} 和 V_{20} 探索本案背后更深层次的问题, 如法律完善、贷款制度、司法公正等, 围绕全面调查并严厉打击山东省内公务人员涉黑、渎职、贪腐等问题发表意见。

按照上文描述的主题校正方法, 对每条评论收到的所有回复进行主题一致性筛选。首先, 将经预处理后留下的1 547条评论分别与其对应的回复结合, 形成1 547份评论-回复结合的数据文件; 其次, 利用基于Relevance公式改进的LDA主题模型和LDAVis, 将这些数据文件一一做可视化展示; 最后, 在可视化结果的基础上判断每个数据文件中评论与回复的主题一致性, 识别不相关的回复并将其删除。删除不相关回复的规则如下: ①若获回复数不超过100, 则删除与评论主题距离最远的一个主题的特征词所构成的最相关回复; ②若获回复数多于100但不超过1 000, 则删除与评论主题距离最远的两个主题的特征词所构成的最相关回复; ③若获回复数多于1 000, 则删除与评论主题距离最远的3个主题的特征词所构成的最相关回复。对所有回复内容进行主题一致性筛选后, 与对应评论主题保持一致的回复的数量为25 535条。

同时, 将没有进行主题校正的数据保留, 以在后续处理中进行对比实验。

3.3 情感计算

基于情感词典及其扩展, 本研究采用上文的情感计算方法, 将各主题下每条评论对应的回复信息进行情感分析, 各主题下获得支持和反对态度的评论数量分布见图4。

结合实际数据可以看出, 由于对出警民警的处理结果存在很大争议, 因此有关主题 V_{12} 的评论也引起广泛讨论, 一部分网民认为出警警察玩忽职守, 而另一部分网民认为警察处理并无不妥, 双方争执不下导致在这一主题下很难有一个统一的意见。以警察的行为为切入点, 许多网民分享了自己遇到的报警无果事件, 牵扯出的与主题 V_9 相关的“警匪一家”问题也引起了极大的争论。基于此, 有网民呼吁严

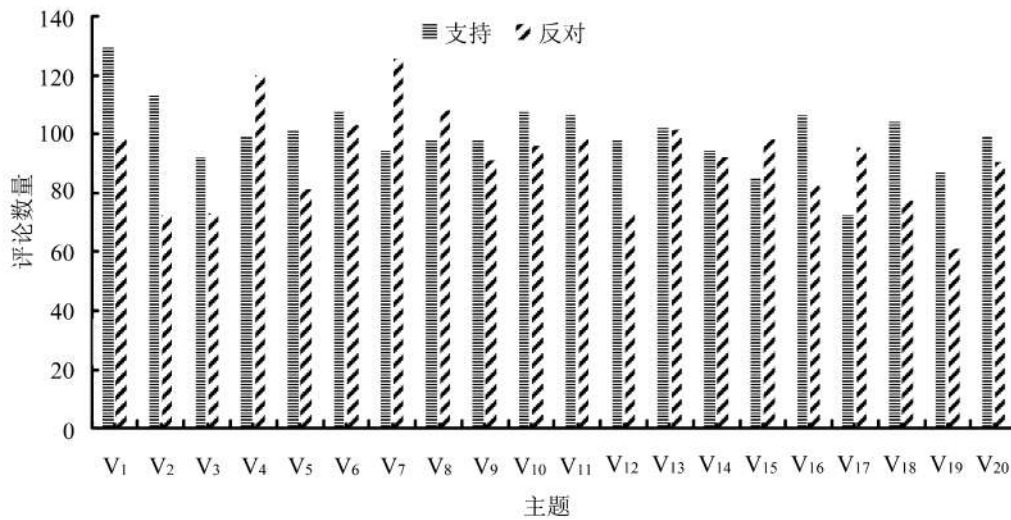


图4 各主题获得支持和反对态度的评论数量分布

Figure 4 Quantity Distribution of Supportive and Objective Comments of Different Topics

表3 排名前10位的评论意见领袖

Table 3 Top 10 Comment Opinion Leaders

评论用户编号	获点赞数	获回复数	意见领袖值	直接获回复数	间接获回复数	直接回复数	评论情感极性
No. 1103	1	0.736	0.860	1	0.421	1	-3(负面)
No. 958	0.571	1	0.799	0.743	1	0	-1(负面)
No. 1104	0.598	0.473	0.532	0.409	0.436	0.086	-2(负面)
No. 959	0.328	0.587	0.465	0.598	0.493	0.022	-2(负面)
No. 1105	0.372	0.338	0.354	0.710	0.077	0.011	0(正面)
No. 1106	0.353	0.258	0.303	0.359	0.156	0.129	-2(负面)
No. 963	0.123	0.404	0.272	0.281	0.394	0.366	-1(负面)
No. 907	0.023	0.483	0.267	0.958	0.142	0	-2(负面)
No. 1112	0.162	0.324	0.248	0.698	0.063	0.022	-3(负面)
No. 1108	0.209	0.244	0.228	0.542	0.039	0	4(正面)

严厉打击山东省内涉黑行为(主题V₂₀),同时聊城市也发出公告,将全面调查于欢杀人护母案件背后公务人员的问题(主题V₁₉),但是多数网民对此并不抱有期待。除此之外,一些网民质疑于欢母亲向高利贷借款的非法行为,但更多的网民认为是银行贷款制度的苛刻和缺乏人性化的设置导致于欢母亲求助于高利贷的做法,而这一现象背后透露出中国中小企业融资困难及难以生存的现状(主题V₇)。民众的信任缺失极易引发社会矛盾,需要意见领袖理智发声,引导网民以正确且积极的心态看待问题,因此识别网络评论意见领袖具有至关重要的意义。通过情感分析,本研究筛选出824个候选意见领袖。

3.4 意见领袖值计算

意见领袖特征指标包括获点赞数和获回复数2个1级指标,以及获回复数下的直接获回复数、间接获回复数和直接回复数3个2级指标。指标体系构成的数据矩阵在经过无量纲化处理和标准差计算后,确定了获点赞数和获回复数的权重,分别为0.469和0.531。按照意见领袖值从高到低排序,前10名意见领袖见表3, No. 1103评论用户和No. 958评论用户的意见领袖值远高于其他评论用户。

结合实际数据可知,排名第1的No. 1103评论用户获点赞数为58 000,获回复数为1 680,其中直接获回复数为912,间接获回复数为675,直接回复数为93。

No. 1103 评论用户获点赞数在所有评论者中是最高的,获回复数在所有评论者中排名第2,综合这两个指标分析,该用户无疑是一个具有影响力的意见领袖。No. 1103 评论用户的直接获回复数和直接回复数在所有评论者中是最大的,说明该用户与回复用户之间互动较多;其间接获回复数在所有评论者中排名第4,高于平均水平,即其他用户与回复用户之间也有较多交流。核对 No. 1103 评论用户的评论内容可知,该评论用户的观点在于于欢护母的忠孝,认为无论是社会还是法律都应保护于欢,并鼓励人们敢于发声,许多回复用户用点赞和回复的方式表示支持。No. 1103 评论用户与回复用户之间的讨论基本上达成了共识,这也说明大多数网民对于欢抱以同情的心理,理解于欢的行为。鉴于上述讨论,识别 No. 1103 评论用户为一名意见领袖是无争议的。

No. 958 评论用户获点赞数为33 112,获回复数为2 282,其中直接获回复数为678,间接获回复数为1 604,直接回复数为0。No. 958 评论用户的获回复数在所有评论者中是最高的,获点赞数在所有评论者中排名第3,该用户在这两个指标上的表现均处于较高水平。No. 958 评论用户的直接回复数在所有评论者中是最小的,表明该用户很少与回复用户进行交流。此外,该评论用户的直接获回复数在所有评论者中排名第3,间接获回复数在所有评论者中排名第1,这也说明该评论用户的评论易引起不同观点之间的相互讨论。核对 No. 958 评论用户的评论内容可知,该评论用户重点讨论出警民警及案件审判人员信息,关注聊城市对此案件的调查,对案件处理尤为重视,并且几乎不与回复用户进一步交流。回复用户针对案件的处理情况发表了不同的看法,相互之间因观点不同而产生许多争论。从回复用户的回复内容可以看出,许多网民关注公检法系统在此次案件中的行为,期待正义的声音不会被淹没。因此, No. 958 评论用户是一个能够引导其他用户思考的意见领袖。

结合实际情况考量,识别出的其他评论用户在本次事件中基本上均发挥了意见领袖应有的作用。此外,通过对比实验,对没有校正主题的回复进行情感计算和意见领袖值计算,发现对比实验识别出的、但原实验未识别出的一些意见领袖,其回复者出现偏离评论主题转而讨论其他事情的现象,回复者与评论者讨论的不是一个主题,这种情况下的评论用户不应视为意见领袖。因此,只考虑情感支持不考虑主题一致性的对比实验存在缺陷,不能有效识别真正的意见领袖,而本研究提出的基于情感支持和主题一致性的方法弥补了这一缺陷。

3.5 评论意见领袖行为特征探讨

通过对1级指标与用户基本信息(包括微博账户等级、是否取得认证、关注数、粉丝数、发表微博数等)之间的相关性检验,发现在置信水平为0.010时,获点赞数与微博账户等级、是否取得认证、关注数、粉丝数、发表微博数等均不存在显著相关性,获回复

数与这些用户基本信息也均不存在显著相关性。因此,评论用户的自身影响力不一定对评论的实际影响力产生作用。

本研究还对意见领袖特征指标之间的相关性进行分析,从而揭示了意见领袖的行为特征。在置信水平为0.010时,直接回复数与直接获回复数之间存在显著的弱相关性,直接获回复数与间接获回复数之间存在显著的强相关性,直接回复数与间接获回复数之间存在显著的弱相关性。结合实际情况可初步得出以下推论:评论者的直接回复在一定程度上引起更多的回复者再次回复,从而增加一定数量的直接获回复数;回复者的间接回复是在直接获回复的基础上产生,即直接获回复是间接回复的存在条件,直接回复数与间接获回复数两者之间存在强相关性不难理解;直接回复需经过直接获回复作用于间接获回复,直接回复数与间接获回复数之间的相关程度较低。

综上所述,本研究提出的基于主题一致性和情感支持的网络评论意见领袖识别方法并不局限于只考虑用户自身影响力,而是着重关注评论内容以及用户之间的交互,是一种新的识别网络评论意见领袖的方法。此外,构建的意见领袖特征体系中的3个2级指标能够反映评论意见领袖的行为特征。

4 结论

本研究提出一种基于主题一致性和情感支持的网络评论意见领袖识别方法,以山东辱母杀人事件为例,使用基于Relevance公式改进的LDA模型和情感词典,筛选与评论主题一致且情感上支持评论者的回复,并根据意见领袖指标体系最终确定具有正面和负面高影响力的微博评论意见领袖,同时从用户主页基本信息和用户评论内容属性两个角度,探讨评论意见领袖的行为特征。

研究结果表明,本研究提出的方法能够识别评论区的意见领袖,评论内容的效应与用户概况之间没有显著的相关性,提出的网络评论意见领袖识别方法并不局限于只考虑评论者自身的影响力,而是着重关注评论内容本身以及用户之间的交互。评论者的获点赞数和获回复数等指标可以作为评论者成为评论区意见领袖的基础,而根据直接获回复数、间接获回复数和直接回复数之间的相互作用可勾勒出用户在交互中的行为特征。

本研究结果能够为突发事件管理部门提供网民对事件的关注焦点和情感态度,有助于其及时掌握网络舆情,通过识别、引导正面评论意见领袖积极发声减少消极情绪,通过引导负面评论意见领袖合理回应存在的问题避免情绪的极端化乃至事件恶化,为后续的应对管理提供理论和方法支持。

本研究也有一定的局限性,即识别的评论意见领袖没有考虑时间因素,缺乏事件发生期间高影响力评论用户的观点演变分析,后续将尝试增加时间这一因素,探索不同时间段内意见领袖的观点构成,

进而识别是否存在具有影响力但观点不同于之前意见领袖的“意见扭转者”;以山东辱母案为例,本研究结果的分析是针对该事件的微博数据展开,该结论在其他突发事件中是否成立还有待验证,因此还需结合多类型事件探讨结论的普适性;情感词汇的情感判断方面,可能会出现情感词汇在词汇表中的情感极性与实际表达中不同的问题,未来将在研究中引入反讽计算等方法。

参考文献:

- [1] LAZARSELD P F, BERELSON B, GAUDET H. *The people's choice: how the voter makes up his mind in a presidential campaign*. 3rd ed. New York: Columbia University Press, 1968: 37-39.
- [2] HOLLANDER E P. Some effects of perceived status on responses to innovative behavior. *Journal of Abnormal & Social Psychology*, 1961, 63(2): 247-250.
- [3] ROGERS E M. *Diffusion of innovations*. New York: The Free Press, 1962: 6-7.
- [4] ROBERTSON T S, MYERS J H. Personality correlates of opinion leadership and innovative buying behavior. *Journal of Marketing Research*, 1969, 6(2): 164-168.
- [5] WEIMANN G. *The influentials: people who influence people*. New York: State University of New York Press, 1994: 28.
- [6] PARK C S, KAYE B K. The Tweet goes on: interconnection of Twitter opinion leadership, network size, and civic engagement. *Computers in Human Behavior*, 2017, 69: 174-180.
- [7] LI B Y, WONG K F, ZHOU L J, et al. Pests hidden in your fans: an effective approach for opinion leader discovery // *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*. Berlin: Springer-Verlag, 2013: 227-237.
- [8] ALEAHMAD A, KARISANI P, RAHGOZAR M, et al. OLFinder: finding opinion leaders in online social networks. *Journal of Information Science*, 2016, 42(5): 659-674.
- [9] KELLER E, BERRY J. *The influentials: one American in ten tells the other nine how to vote, where to eat, and what to buy*. New York: The Free Press, 2003: 52.
- [10] 倪邦文. 中国网络青年意见领袖的构成、特征及作用. *中国青年研究*, 2011(9): 5-9, 102.
NI Bangwen. Study of youth opinion leaders. *China Youth Study*, 2011(9): 5-9, 102. (in Chinese)
- [11] MATSUMURA N. Collaborative communication strategies in online community // *Proceedings Fourth International Workshop on Social Intelligence Design (SID2005)*. Enschede, Netherland: Centre for Telematics and Information Technology at University of Twente, 2005: 79-86.
- [12] SOHN Y. *Opinion leaders and seekers in online brand communities: centered on Korean digital camera brand communities*. Florida: The Florida State University, 2005: 4.
- [13] 余红. 网络时政论坛舆论领袖研究: 以强国社区中日论坛为例. 武汉: 华中科技大学, 2007: 53-65.
YU Hong. *Research on the opinion leaders of political bbs: an case study on Sino-Japan bbs of strong nation forum*. Wuhan: Huazhong University of Science and Technology, 2007: 53-65. (in Chinese)
- [14] 薛可, 陈晔. BBS中的“舆论领袖”影响力传播模型研究: 以上海交通大学“饮水思源”BBS为例. *新闻大学*, 2010(4): 87-93.
XUE Ke, CHEN Xi. Research on the influence propagation model of “public opinion leaders” in BBS: a case study on Shanghai Jiao Tong University BBS. *Journalism Quarterly*, 2010(4): 87-93. (in Chinese)
- [15] 丁汉青, 王亚萍. SNS网络空间中“意见领袖”特征之分析: 以豆瓣网为例. *新闻与传播研究*, 2010, 17(3): 82-91, 111.
DING Hanqing, WANG Yaping. Analyzing “opinion leader” attributes in SNS cyberspace: an investigation of Douban.com. *Journalism & Communication*, 2010, 17(3): 82-91, 111. (in Chinese)
- [16] 顾品浩, 蒋冠. 突发性公共事件中的网络意见领袖分析: 以“杨达才事件”为例. *情报杂志*, 2013, 32(5): 20-24.
GU Pinhao, JIANG Guan. Analysis on network opinion leaders in public emergencies: a case study of Yang Dacai event. *Journal of Intelligence*, 2013, 32(5): 20-24. (in Chinese)
- [17] 常亚平, 邱媛媛, 阎俊, 等. 虚拟社区知识共享主体对首购意愿的作用机理研究. *管理科学*, 2011, 24(2): 74-84.
CHANG Yaping, QIU Yuanyuan, YAN Jun, et al. Influencing mechanism of knowledge sharing subject in virtual communities on customers' first-time purchase intention. *Journal of Management Science*, 2011, 24(2): 74-84. (in Chinese)
- [18] 熊涛, 何跃. 微博转发网络中意见领袖的识别与分析. *现代图书情报技术*, 2013(6): 55-62.
XIONG Tao, HE Yue. The identification and analysis of micro-blogging opinion leaders in the network of retweet relationship. *New Technology of Library and Information Service*, 2013(6): 55-62. (in Chinese)
- [19] 唐晓波, 向坤. 基于LDA模型和微博热度的热点挖掘. *图书情报工作*, 2014, 58(5): 58-63.
TANG Xiaobo, XIANG Kun. Hotspot mining based on LDA model and microblog heat. *Library and Information Service*, 2014, 58(5): 58-63. (in Chinese)
- [20] 沈磊. 基于规则与机器学习方法的中文微博情感分析研究. 合肥: 安徽大学, 2015: 22-24.
SHEN Lei. *The research on Chinese microblog sentiment analysis based on rules and machine learning methods*. Hefei: Anhui University, 2015: 22-24. (in Chinese)
- [21] 朱茂然, 林星凯, 陆颀, 等. 基于情感分析的社交网络意见领袖的识别: 以汽车论坛为例. *情报理论与实践*, 2017, 40(6): 76-81.
ZHU Maoran, LIN Xingkai, LU Ting, et al. Identification of opinion leaders in social networks based on sentimental analysis: evidence from an automotive forum. *Information Studies: Theory & Application*, 2017, 40(6): 76-81. (in Chinese)
- [22] 朱义生. 基于SNA面向特定主题的意见领袖发现研究. 合肥: 合肥工业大学, 2012: 18-24.
ZHU Yisheng. *The discovery of opinion leaders for a particular theme based on social network analysis*. Hefei: Hefei

- University of Technology, 2012;18-24. (in Chinese)
- [23] SIEVERT C, SHIRLEY K E. LDavis: a method for visualizing and interpreting topics // *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces at the Association for Computational Linguistics*. Baltimore, Maryland, 2014;63-70.
- [24] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques // *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*. Philadelphia, 2002;79-86.
- [25] 唐晓波, 兰玉婷. 基于特征本体的微博产品评论情感分析. *图书情报工作*, 2016, 60(16):121-127, 136.
TANG Xiaobo, LAN Yuting. Sentiment analysis of microblog product reviews based on feature ontology. *Library and Information Service*, 2016, 60(16):121-127, 136. (in Chinese)
- [26] 徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造. *情报学报*, 2008, 27(2):180-185.
XU Linhong, LIN Hongfei, PAN Yu, et al. Constructing the affective lexicon ontology. *Journal of the China Society for Scientific and Technical Information*, 2008, 27(2):180-185. (in Chinese)
- [27] MARON M E, KUHNS J L. On relevance, probabilistic indexing and information retrieval. *Journal of the ACM*, 1960, 7(3):216-244.
- [28] CORTES C, VAPNIK V. Support-vector networks. *Machine Learning*, 1995, 20(3):273-297.
- [29] 司琪. 基于云模型理论的LDA最大熵模型观点挖掘研究. 武汉: 华中师范大学, 2016;22-33.
SI Qi. *Opinion mining research of LDA maximum entropy model based on cloud model theory*. Wuhan: Central China Normal University, 2016;22-33. (in Chinese)
- [30] SINDHWANI V, MELVILLE P. Document-word co-regularization for semi-supervised sentiment analysis // *Proceedings of the 8th IEEE International Conference on Data Mining*. Pisa, 2008;1025-1030.
- [31] 张卓, 高维和. 网络意见领袖的分化与草根精英的崛起: 基于网络游戏论坛的理论与实证检验. *新闻与传播研究*, 2017, 24(3):20-41, 126.
ZHANG Zhuo, GAO Weihe. Network opinion leaders and the raising levels of crowd sourced elites: a theoretical study and empirical test on online game BBS. *Journalism & Communication*, 2017, 24(3):20-41, 126. (in Chinese)
- [32] 蒋翠清, 朱义生, 丁勇. 基于UGC下的意见领袖发现研究. *情报杂志*, 2011, 30(10):82-85.
JIANG Cuiqing, ZHU Yisheng, DING Yong. On discovery of opinion leaders based on UGC. *Journal of Intelligence*, 2011, 30(10):82-85. (in Chinese)
- [33] 王珏, 曾剑平, 周葆华, 等. 基于聚类分析的网络论坛意见领袖发现方法. *计算机工程*, 2011, 37(5):44-46, 49.
WANG Jue, ZENG Jianping, ZHOU Baohua, et al. Online forum opinion leaders discovering method based on clustering analysis. *Computer Engineering*, 2011, 37(5):44-46, 49. (in Chinese)
- [34] 祝帅, 郑小林, 陈德人. 论坛中的意见领袖自动发现算法研究. *系统工程理论与实践*, 2011, 31(S2):7-12.
ZHU Shuai, ZHENG Xiaolin, CHEN Deren. Research of algorithm for automatic opinion leader detection in bbs. *Systems Engineering - Theory & Practice*, 2011, 31(S2):7-12. (in Chinese)
- [35] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003, 3(1):993-1022.
- [36] 陈晓美, 高铨, 关心惠. 网络舆情观点提取的LDA主题模型方法. *图书情报工作*, 2015, 59(21):21-26.
CHEN Xiaomei, GAO Cheng, GUAN Xinhui. Extraction method of network public opinion based on LDA topic model. *Library and Information Service*, 2015, 59(21):21-26. (in Chinese)
- [37] 安璐, 欧孟花. 突发公共卫生事件利益相关者的社会网络情感图谱研究. *图书情报工作*, 2017, 61(20):120-130.
AN Lu, OU Menghua. Social network sentiment map of the stakeholders in public health emergencies. *Library and Information Service*, 2017, 61(20):120-130. (in Chinese)
- [38] 王勇, 吕学强, 姬连春, 等. 基于极性词典的中文微博客情感分类. *计算机应用与软件*, 2014, 31(1):34-37, 126.
WANG Yong, LYU Xueqiang, JI Lianchun, et al. Sentiment classification for Chinese microblogging based on polarity lexicons. *Computer Applications and Software*, 2014, 31(1):34-37, 126. (in Chinese)
- [39] 安璐, 吴林. 融合主题与情感特征的突发事件微博舆情演化分析. *图书情报工作*, 2017, 61(15):120-129.
AN Lu, WU Lin. An integrated analysis of topical and emotional evolution of microblog public opinions on public emergencies. *Library and Information Service*, 2017, 61(15):120-129. (in Chinese)
- [40] 袁慧, 马建霞. 失效专利的主题挖掘与应用分析: 以2007-2016年镜头失效专利为例. *情报理论与实践*, 2017, 40(12):128-133, 74.
YUAN Hui, MA Jianxia. Topic mining and application analysis of invalid patent: a case study of invalid patent on lens from 2007 to 2016. *Information Studies: Theory & Application*, 2017, 40(12):128-133, 74. (in Chinese)
- [41] 肖挺, 刘华. 服务业生产效率异质性对城乡收入差距影响研究. *管理科学*, 2013, 26(4):103-112.
XIAO Ting, LIU Hua. Empirical study on the impacts of productivity heterogeneity in Chinese service industry on the urban-rural income gap. *Journal of Management Science*, 2013, 26(4):103-112. (in Chinese)
- [42] 刘为怀, 才华, 何东杰. 一种基于中文分词和数据聚合的餐饮行为特征挖掘方法. *软件产业与工程*, 2015(4):47-51, 56.
LIU Weihuai, CAI Hua, HE Dongjie. Efficient method for mining restaurant behavior characteristics based on Chinese segmentation and data aggregation. *Software Industry and Engineering*, 2015(4):47-51, 56. (in Chinese)

A Method of Identifying Comment Opinion Leaders Based on Topic Consistency and Emotional Support

AN Lu^{1,2}, HU Junyang², LI Gang¹

1 Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China

2 School of Information Management, Wuhan University, Wuhan 430072, China

Abstract: The emergence of opinion leaders on social networks has a profound impact on the public. Users with high impact in the comment area receive more and more attention. Their comments may guide the public opinions about public emergencies, which will have positive or negative effects on the later evolution and results of events. Identifying opinion leaders in the comment area is of great practical significance in guiding the public opinions about emergencies.

To this end, a new method of identifying online opinion leaders in comments based on topical consistency and emotional support was proposed. We took the Shandong humiliating mother case as an example, using the improved Latent Dirichlet Allocation model based on the relevance formula to extract the topics of microblog comments, and achieved topic correction according to the distances between comment nodes and the response nodes in the visual topic display. By the sentiment analysis based on sentiment lexicons, we calculated the emotion values of comments and responses and determined the emotional support values of responders to commenters. A group of characteristic indices of opinion leaders was constructed, including two first-level indicators (i. e. the number of likes, the number of responders) and three second-level indices (i. e. the number of being responded directly, the number of being responded indirectly and the number of responding to others). Weights were assigned to the first-level indices and the values of opinion leaders were calculated. According to the opinion leader values, the positive and negative high-impact commenters were identified. At the same time, the behavioral characteristics of opinion leaders in comments were explored from two perspectives of user profiles and the contents of these comments.

The result shows that the number of likes and responses can be used as the basis for users that become opinion leaders in comments. According to the interaction between the number of being responded directly, the number of being responded indirectly and the number of responding to others, the behavioral characteristics of users can be revealed. It is found that there is no significant correlation between the effect of comment contents and user profiles. The proposed method of identifying opinion leaders in comments is no longer limited to the effect of the commenter's own influence, but focuses on the comment content and the interaction between commenters and responders, which can effectively identify the opinion leaders in comments.

The research findings can provide management departments with netizens' focus of attention and emotional attitude to the public emergencies, and help them grasp the online public opinion timely. It can also quickly and accurately locate the online opinion leaders in comments, and provide effective support for decision-making and management.

Keywords: opinion leader; Web comment; topical consistency; emotional support; microblog

Received Date: July 4th, 2018 **Accepted Date:** December 3rd, 2018

Funded Project: Supported by the Major Project of Philosophy and Social Sciences Research, Ministry of Education of China(17JZD034) and the National Natural Science Foundation of China(71790612,71603189)

Biography: AN LU, doctor in management, is a professor in the Center for Studies of Information Resources and School of Information Management at Wuhan University. Her research interests include visual knowledge discovery and online data analysis. Her representative paper titled "Visual topical analysis of Chinese and American library and information science research institutions" was published in the *Journal of Informetrics* (Issue 1, 2014). E-mail: anlu2009@whu.edu.cn

HU Junyang is a master degree candidate in the School of Information Management at Wuhan University. Her research interest focuses on online public opinion analysis. E-mail: huju2015@126.com

LI Gang, doctor in management, is a professor in the Center for Studies of Information Resources at Wuhan University and a Yangtze River Scholar distinguished professor. His research interests include information resource management and online public opinion analysis. His representative paper titled "Research on evolutionary dynamics of knowledge diffusion based on collaboration hypernetwork" was published in the *Journal of the China Society for Scientific and Technical Information* (Issue 3, 2017). E-mail: imiswhu@aliyun.com □