



面向网络舆情事件的多层次情感分歧度分析方法^{*}

华 玮 吴思洋 俞 超 吴婕洵 徐 健

(中山大学信息管理学院 广州 510006)

摘要:【目的】面对网络舆情事件,从情感分歧角度出发,为舆情分析提供新的分析角度。【方法】引入情感分歧度概念,创建多层次情感分歧度算法,构建网络舆情事件多层次情感分歧度分析模型,对网络舆情事件层、评论对象层、用户层进行情感值及情感分歧度的计算,并将三个层次进行关联分析。【结果】实验结果表明,引入情感分歧度可以弥补原有情感分析研究中对网民意见分歧角度的缺失,本模型可以实现舆情事件关键节点及争议较大的评论对象的识别、判断舆论引导效果,并对舆情争议产生的原因实现精准定位。【局限】仅选取微博作为数据源,未从豆瓣、知乎等其他平台获取数据。【结论】本文模型可应用于监控舆情事件关键节点、根据争议原因选择不同的舆论引导方式以及判断舆论引导效果。

关键词: 情感分歧度 网络舆情分析 多层次 关联分析

分类号: G252

DOI: 10.11925/infotech.2096-3467.2022.0370

引用本文: 华玮, 吴思洋, 俞超等. 面向网络舆情事件的多层次情感分歧度分析方法[J]. 数据分析与知识发现, 2023, 7(4): 16-31.(Hua Wei, Wu Siyang, Yu Chao, et al. Analyzing Divergence of Multi-layer Sentiment for Online Public Opinion Events[J]. Data Analysis and Knowledge Discovery, 2023, 7(4): 16-31.)

1 引言

“情感分歧”指在社交网络中不同评论主体、不同评论源之间,以及相同评论主体在不同时间段对某一特定事物或话题所表达出的情感差异性^[1]。现有研究在利用情感分析技术进行网络舆情分析时,往往忽略了情感分歧的存在,而消解各方的猜疑与分歧,是处置网络舆情、引导网络舆论的重要目标。本文参考已有的情感分歧度量化算法,将情感分歧度这一指标应用于对网络舆情事件的情感分析,提出网络舆情事件多层次情感分歧度分析模型。该模型引入情感分歧度,从网络舆情事件层、评论对象层、用户层三个层面对网络舆情事件进行分析。提

出针对三个层次特点的情感分歧度量化方式并对其进行计算,得出每一层次情感分歧度随时间演化趋势,并进行层次间的关联分析,弥补了原有情感分析指标忽略网民意见存在分歧的缺陷,扩展了网络舆情分析的角度。而且该模型对网络舆情进行的单层次分析以及关联分析,使得网络舆情的分析更加透彻与深入。

2 相关研究综述

2.1 情感分析在网络舆情分析中的应用

从分析指标来看,当前研究主要使用各极性评论数量占比^[2-5]、情感平均值^[6-8]、情感分类与情感强度相结合^[9-10]等指标判断网民的情感倾向,此类方法

通讯作者(Corresponding author): 徐健(Xu Jian), ORCID: 0000-0003-4886-4708, E-mail: issxj@mail.sysu.edu.cn.

*本文系广东省自然科学基金项目(项目编号: 2018A030313981)的研究成果之一。

The work is supported by the Natural Science Foundation of Guangdong Province (Grant No. 2018A030313981).

会忽略网民在表达意见时存在的情感分歧。而找出各方的分歧所在并对其进行合理的引导,是网络舆情处理的关键环节。现有研究对情感分歧度指标在网络舆情分析中的探究还很少。

从分析方法来看,当前研究在利用情感分析方法进行舆情分析时,常与其他分析方法或研究方法相结合,如时间序列分析、主题分析等。例如,赵晓航^[11]将情感分析与主题分析、回归分析相结合,探究在突发事件下,政府如何做好信息公开;李勇等^[5]、崔彦琛等^[12]将情感分析与时间序列分析相结合,分析网络用户在舆情事件中的情感演化趋势,有助于舆情监测。

从分析层面来看,舆情分析主要从事件层、评论对象层、用户层三个层面展开研究。

(1)对事件层的研究,陈兴蜀等^[6]探究了不同时空下,网民对“新冠肺炎疫情”的情感特征,分析预测疫情后续的时空发展趋势;邵琦等^[13]、梁晓贺等^[14]对主题发现方法进一步优化,有助于舆情主题的演化分析。

(2)对评论对象层的研究,梁晓敏等^[15]分析了网民对舆论对象随时间演化的情感表达和关系认知;辜丽琼等^[16]通过对比企业进行危机舆情应对前后网民各评论对象的情感变化评估企业应对效果。

(3)对用户层的研究,王志英等^[17]验证了情感领袖在集群情感的形成中产生群际情感效应和群际认同效应;邓君等^[18]利用舆情事件中相关的 5 类舆情主体对象的情感演化分析舆情的情感走势;孔婧媛等^[19]探究了舆情当事人的回应对网民情感的影响,并得出了有效的证据和诚恳的态度有助于平抑舆情中的负面消极情绪的结论,为舆论引导提供了参考;何跃等^[20]分析了不同用户群体在雾霾舆情中的情感倾向,以及关键用户对公众情绪的引导作用。

现有研究通常从特定层次进行分析,很少将事件、评论对象、用户三个层次进行关联分析。三个层次的关联关系可实现对网民情感变化原因的精准定位,更有助于为舆论的引导提供决策支持。

2.2 情感分歧度相关研究

现有研究对情感分歧度量化方法进行了探索。例如,Cao 等^[21]提出计算两个群体间情感分歧度的方法,将所有情感表达者划分成两个社区群体,并且采用平方误差和的形式计算两个社区群体之间的

情感分歧度。Zhang 等^[22]提出基于 Kullback-Leibler 差异计算的方法、Subašić 等^[23]提出基于 Jensen-Shannon 散度的情感分歧度计算方法,这两种情感分歧度量化算法均用于两组评论间的情感分歧度计算。在计算单一用户情感分歧度时,Pfitzner 等^[24]研究用户评论分歧大小对信息传播与信息分布的影响时,提出计算单一用户情感分歧度的方法,为该用户评论赋予正向情感值(介于 1~5 之间)与负向情感值(介于-5~-1 之间),利用正负情感差值除以 10 作为该用户的情感分歧度。在计算群体内部间情感分歧度时,徐健等^[1]提出了基于情感值差、标准差、变异系数、信息熵、情感分布概率、情感的正负值特点的 6 种情感分歧度量化算法,并利用豆瓣影评数据对 6 种量化算法进行了验证。在应用方面,情感分歧度指标主要应用于经济学领域^[25-28]、产品满意度评价^[29],目前直接从情感分歧度角度进行情感分析的相关研究尚处在初步探索阶段。

因此,本文提出面向网络舆情事件的多层次情感分歧度分析模型,将情感分歧度这一指标用于舆情分析中,弥补了现有情感分析指标忽略网民情感分歧的问题,为网络舆情分析提供更加细致与深入的思路,有助于舆情监测、舆论导向任务的深入展开。同时也扩展了情感分歧度指标的应用领域。

3 网络舆情事件多层次情感分歧度分析研究设计

本文提出网络舆情事件多层次情感分歧度分析模型,该模型主要分为语料获取、情感词库构建、情感计算、多层次关联分析四大模块,如图 1 所示。

3.1 语料获取模块

语料获取模块主要包括数据源选择、数据抓取、数据预处理与存储三个步骤。其中,数据源应选择舆情发酵的主要网络平台;为尽量保证数据召回率,选择与事件相关的话题或名词作为检索词,检索时间段为舆情事件的发酵时间。对抓取的数据进行预处理,主要步骤包括:数据去重、删除未表达意见内容、仅保留检索时间段内的内容。将预处理后的数据存储至数据库,构成研究所需语料库。

3.2 情感词库构建模块

虽然基于机器学习、深度学习的情感分析方法

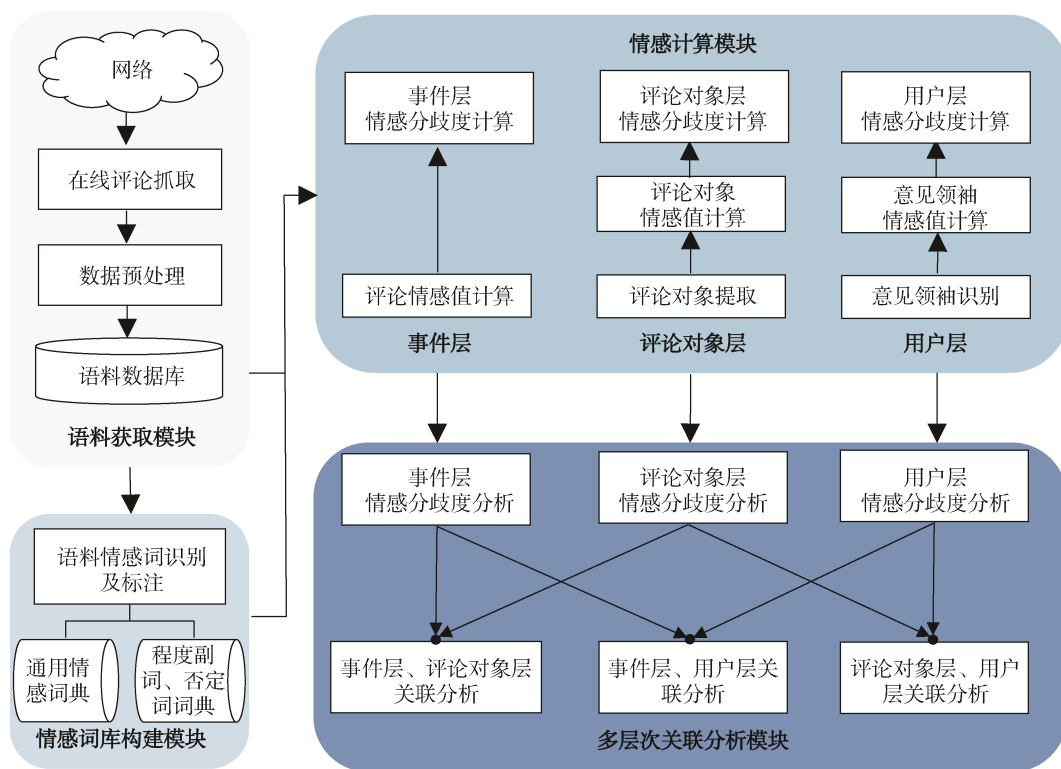


图1 网络舆情事件多层次情感分歧度分析模型

Fig.1 The Multi-layer Sentiment Divergence Degree Analysis Model of Network Public Opinion Events

精确率和召回率相对更高^[30],但需要大规模的训练数据。由于本文的数据集规模较小,基于情感词典的情感分析方法的精确率和召回率已经可以满足本文的分析需求,方法相对简单且易于理解^[31],因此本文使用基于情感词典的分析方法。

情感词库的构建主要包括构建舆情分析专用情感词典、否定词及程度副词词典两部分。目前所使用的情感词典多为通用词典,但由于情感词在不同领域之间存在差异,仅用通用词典进行情感分析会存在一定偏差,故需构建舆情专用情感词典。首先,选取大连理工大学的情感词汇本体库^[32]作为情感分析的基础词典,选取 BosonNLP 情感词典^[33]对网络用语和非正式用语进行补充;其次,提取数语料库中有明显情感倾向的词语补充基础词典,并标注情感极性及其情感强度;最后,收集常用程度副词^[34-36]、否定词^[37]并设置相应权重。

3.3 情感计算模块

(1) 情感值算法

本文采用基于情感词典的情感值算法^[38],如公

式(1)所示。情感值为正则该文本的情感倾向为积极,为负则为消极。

$$Sentiment = \sum_{i=1}^N \left[\sum_{j=1}^n (W_i S_i + k \times 2) - \sum_{j=1}^m (W_j S_j + k \times 2) \right] \quad (1)$$

其中, $Sentiment$ 表示文本的情感值; N 表示分句的条数; n 和 m 分别表示分句所包含的积极情感词数量和消极情感词数量; S 表示情感词强度; W 表示情感词前的否定词或程度副词的权值,若无则为1; k 表示该情感词后是否包含感叹号,有感叹号则 k 值为1,否则为0,由于感叹号对情感词有加强作用^[2],因此当情感词后存在感叹号时,情感强度加2。

(2) 多层次情感分歧度算法

针对不同层次的特征,本文提出三种不同的情感分歧度量化算法。

① 事件层情感分歧度算法

事件层的内容包括该网络舆情事件发生后,网民所发表的所有原创内容、转发、评论等三类数据。对事件层进行分析,可从整体把握该事件的情感走

向。在实际场景中,常发原创内容的用户与常转发的用户表达情感方式不同,前者更加直接强烈,后者相对委婉内敛,故针对不同的人群,应分别计算不同群体的情感平均值、情感分歧度。网络舆情事件的事件层情感平均值以及情感分歧度的计算如公式(2)和公式(3)所示。

$$\overline{Sentiment} = 0.55 \times \overline{Sentiment}_{original} + 0.3 \times \overline{Sentiment}_{transmit} + 0.15 \times \overline{Sentiment}_{comment} \quad (2)$$

$$SentiDiv = 0.55 \times SentiDiv_{original} + 0.3 \times SentiDiv_{transmit} + 0.15 \times SentiDiv_{comment} \quad (3)$$

其中, $\overline{Sentiment}$ 表示情感平均值; $SentiDiv$ 表示情感分歧度;下标 *original*、*transmit*、*comment* 分别代表原创内容、转发、评论,三者的系数为影响力权重,影响力权重的和为1。原创内容、转发、评论的影响力权重分别取 0.55、0.30、0.15^[39-40]。

原创内容、转发、评论的情感平均值的计算如公式(4)所示。原创内容情感平均值绝对值通常大于转发与评论情感平均值的绝对值,为消除情感平均值相差较大的影响,采用基于变异系数的情感分歧度量化算法^[1],如公式(5)所示。

$$\overline{Sentiment}_{event} = \frac{\sum_{j=1}^N Sentiment_j \cdot n_j}{\sum_{j=1}^N n_j} \quad (4)$$

$$SentiDiv_{event} = \frac{\sqrt{\frac{\sum_{j=1}^N (Sentiment_j - \overline{Sentiment})^2 n_j}{\sum_{j=1}^N n_j}}}{|\overline{Sentiment}|} \quad (5)$$

其中, $\overline{Sentiment}_{event}$ 、 $SentiDiv_{event}$ 分别表示情感平均值、情感分歧度; $Sentiment_j$ 表示第 j 条文本的情感值;表示每条评论的情感认同数; N 表示文本总数; $|\overline{Sentiment}|$ 表示情感平均值的绝对值。

公式(4)和公式(5)中包含了情感认同数指标,根据文本的类型不同,情感认同数计算方式不同。对于转发与评论内容,情感认同数与其点赞数直接相关。对于原创内容,情感认同数与用户点赞数以及直接转发数相关。博文转发数权重的大小一般为点赞数的 2 倍左右^[41-42],因此本文的情感认同数 n_j 的计算如公式(6)和公式(7)所示。

$$n_{j_{original}} = 1 + count_{j_{transmit}} \times 0.2 + count_{j_{like}} \times 0.1 \quad (6)$$

$$n_{j_{transmit/comment}} = 1 + count_{j_{like}} \times 0.1 \quad (7)$$

其中, $count_{j_{transmit}}$ 表示第 j 条文本直接转发数; $count_{j_{like}}$ 表示第 j 条文本点赞数。

② 评论对象层情感分歧度算法

评论对象是指在舆情事件中受到网民讨论与评价的人物、行为、事物、环境等,因此评论对象层所分析的内容仅包含提及该评论对象的文本。对评论对象层进行分析,可对该舆情事件实现更加细致的讨论,满足不同主体的关注需求。

评论对象层以分句为语义单元,仍采用公式(2)和公式(3)计算评论对象的情感平均值和情感分歧度。事件层的情感认同数是针对完整的文本计算得来的,对完整的文本的赞同并不等同于对其中每个分句的赞同,因此在评论对象层的计算中情感认同数全部取 1。由于以分句为单位计算出的评论对象情感值会相对较小,因此在计算评论对象情感分歧度时,使用基于标准差的情感分歧度量化算法^[1]。原创内容、转发、评论的评论对象层的情感平均值和情感分歧度的计算如公式(8)和公式(9)所示。

$$\overline{Sentiment}_{object} = \frac{\sum_{j=1}^N Sentiment_j}{N} \quad (8)$$

$$SentiDiv_{object} = \sqrt{\frac{\sum_{j=1}^N (Sentiment_j - \overline{Sentiment})^2}{N}} \quad (9)$$

其中, $\overline{Sentiment}_{object}$ 和 $SentiDiv_{object}$ 分别表示评论对象层的原创内容、转发、评论的情感平均值和情感分歧度。

③ 用户层情感分歧度算法

由于多数用户在单个舆情事件中的意见表达并不具有持续性,因此本文分析的重点放在舆情事件中的意见领袖。在对用户层情感分歧度进行计算之前,首先要进行意见领袖的确定。意见领袖是指在社交平台上能够因传播信息和表达意见而影响广大受众的人^[43]。本文基于灰色关联分析的意见领袖识别方法确认事件的意见领袖^[44]。

由于发文用户是否为意见领袖与被转发数、被评论数、点赞数密切相关,因此抓取选取事件中“被转发数+被评论数+点赞数”的总值排名前 50 的用户的相关数据,利用公式(10)将所抓取的数据进行归一化,利用公式(11)计算关联系数,利用公式(12)计算灰色加权关联度。

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (10)$$

$$\zeta_j(k) = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \rho \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(k) - x_s(k)| + \rho \max_s \max_t |x_0(t) - x_s(t)|} \quad (11)$$

$$r_j = \sum_{j=1}^n w_j \zeta_j(k) \quad (12)$$

其中, $\zeta_j(k)$ 为比较数列 x_i 对参考数列 x_0 在第 k 个指标上的关联系数; $\min_s \min_t |x_0(t) - x_s(t)|$ 和 $\max_s \max_t |x_0(t) - x_s(t)|$ 为两级最小差和两级最大差; ρ 为分辨系数, 本文取 0.5; r_j 为第 j 个评价对象对理想对象的灰色加权关联度, 即其意见领袖指数; n 为指标个数。

与事件层和评论对象层不同, 用户层所分析的内容为单一用户所发表的原创、转发、评论内容, 关注特定用户与大众之间的情感分歧, 因此采用基于情感均值差的方法计算用户的情感分歧度, 即将用户的情感均值与当天的大众情感均值作差, 得到当天该用户与大众之间的情感分歧度, 如公式(13)所示。

$$SentiDiv_{ji} = \overline{Sentiment}_{ji} - \overline{Sentiment}_i \quad (13)$$

其中, $\overline{Sentiment}_{ji}$ 表示用户 j 在第 i 天的情感均值; $\overline{Sentiment}_i$ 表示第 i 天大众的情感均值, 若该用户发表的内容为原创内容, 则使用该天原创内容的大众情感均值, 若该用户发表的内容为评论或转发, 则使用该天评论或转发的大众情感均值; $SentiDiv_{ji}$ 表示用户 j 在第 i 天的情感分歧度, 有正负之分, 数值为正意味着该用户的评论对大众情感存在正向影响, 数值为负意味着该用户的评论对大众情感存在负向影响。 $SentiDiv_{ji}$ 的绝对值代表用户情感与大众情感之间的距离, 即用户 j 在第 i 天的情感分歧强度, 数值越大, 说明用户情感与大众情感的分歧越大。

④ 多层次情感分歧度关联分析模块

为了研究三个层次情感分歧度之间的影响和关系, 对其进行关联分析, 本文提出评论对象加权情感分歧度、评论对象间情感分歧度、用户加权情感分歧度三个指标。其中, 评论对象加权情感分歧度指标可以衡量该评论对象的情感分歧度对事件层整体情感分歧度的影响程度, 加权情感分歧度越大, 对事件

层整体情感分歧度影响越大; 评论对象间情感分歧度使用基于情感值差的情感分歧度量化算法^[1]计算, 探究评论对象间的情感差异对事件层情感分歧度的影响, 因评论对象情感平均值数值个数较少, 选取该算法可以降低极端值的影响; 用户加权情感分歧度指标用以衡量用户层情感分歧度对事件层整体情感分歧度的影响程度。

1) 评论对象加权情感分歧度算法

评论对象加权情感分歧度指标的计算如公式(14)所示。

$$WeightedSentiDiv_j = SentiDiv_j \cdot \frac{m_j}{M} \quad (14)$$

其中, $WeightedSentiDiv_j$ 表示评论对象 j 加权情感分歧度; $SentiDiv_j$ 表示评论对象 j 的情感分歧度; M 表示事件整体的文本总数; m_j 表示提及评论对象 j 的文本数。

2) 评论对象间情感分歧度算法

评论对象间情感分歧度指标的计算如公式(15)所示。

$$SentiDiv_{between} = \frac{2}{N(N-1)} \sum_{j=1}^N \sum_{k=j+1}^N |\overline{Sentiment}_j - \overline{Sentiment}_k| \quad (15)$$

其中, $SentiDiv_{between}$ 表示评论对象间情感分歧度; $\overline{Sentiment}_j$ 、 $\overline{Sentiment}_k$ 分别代表评论对象 j 、 k 的情感平均值; N 代表评论对象个数。

3) 用户加权情感分歧度算法

用户加权情感分歧度指标的计算如公式(16)所示。

$$WeightedSentiDiv_i = \ln(|SentiDiv_i| + 1) \frac{\sum_{j=1}^n \frac{m_{ij}}{M_j}}{n} \quad (16)$$

其中, $WeightedSentiDiv_i$ 表示用户 i 的加权情感分歧度; $SentiDiv_i$ 表示用户 i 情感分歧度, 取自然底数对数是为了减少用户层情感分歧强度受用户发内容长短的影响, 并且为了减少极端值的影响, 对情感分歧强度做了加 1 的处理; n 表示用户 i 的影响力持续天数; j 表示用户 i 发文后的第 j 天; M_j 表示第 j 天事件整体的文本总数; m_{ij} 表示用户 i 下第 j 天引起的评论、转发文本总数。

4 实验与分析

4.1 实验数据集

选取微博作为数据源。为验证网络舆情事件多层次情感分歧度分析模型的有效性,本文分别从正面舆情、负面舆情、中性舆情三类^[45]舆情事件中选取代表性事件作为实验数据源,对其事件层、评论对象层以及用户层的情感分歧度进行分析。因为在网络舆情事件中,负面舆情事件占比高于另外两类舆情事件,且对于负面舆情事件的舆情分析以及舆论控制更加重要,因此本文选取“网传滴滴司机直播性侵”“虎牙员工自曝被HR抬出公司”两个负面舆情事件、“非全日制研究生系列事件”中性舆情事件以

及“新冠疫苗全民免费”正面舆情事件作为研究数据源,这4个事件的详情经过见支撑数据。

选择与事件相关词汇作为事件的检索词,舆论发酵时间作为检索时间段,4个事件的数据抓取策略与结果见支撑数据。

对所抓取的数据进行预处理,包括:只保留一次转发、评论内容;删除与转发重复的评论内容;删除未发表意见内容;仅保留检索时间段内的内容。4个事件预处理后的数据情况见支撑数据。

4.2 事件层情感分歧度计算及分析

利用3.3节提出的事件层情感平均值以及情感分歧度的计算方式,分别计算4个事件每天的情感平均值及情感分歧度,计算结果如图2所示。

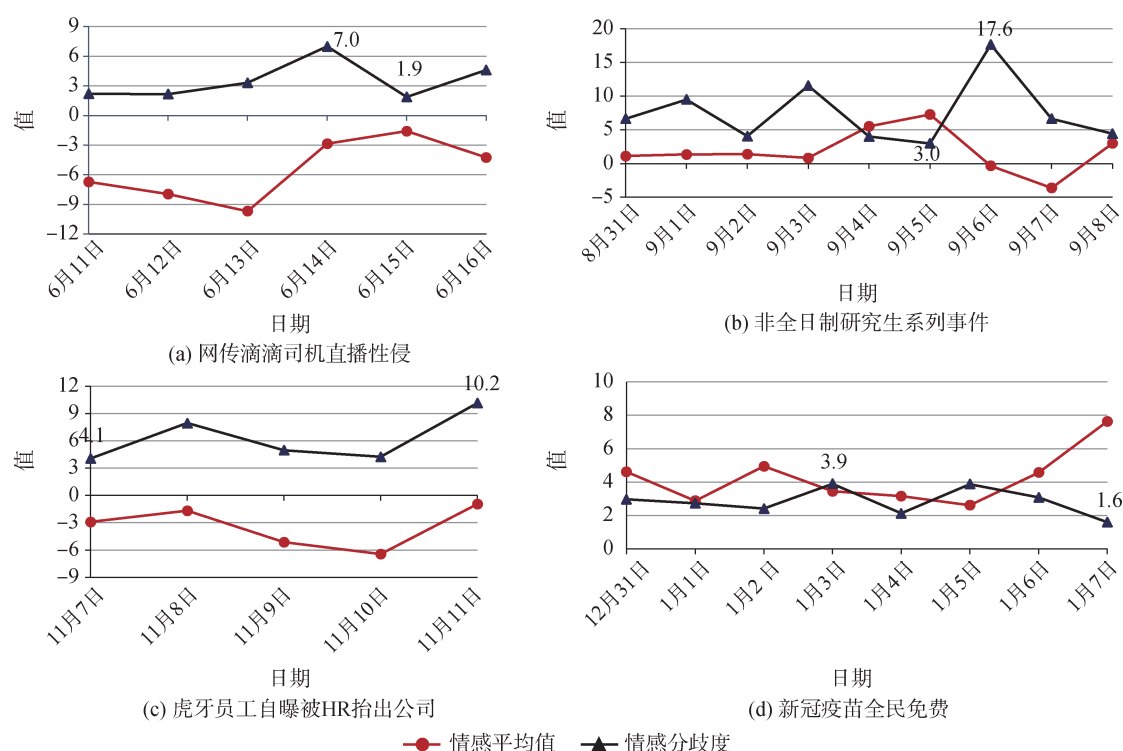


图2 事件层情感平均值与情感分歧度计算结果

Fig.2 Event Layer Emotional Mean and Sentiment Divergence Calculation Results

如图2所示,在对事件层情感分歧度的变化进行分析时,对微博舆情发展阶段可划分为舆情酝酿期、信息阐释期、舆情交互期以及舆情消解期^[11],情感分歧度表现最显著的是以下三个阶段。

(1)在舆情酝酿期,即舆情中关键事件刚刚发生

时期,网民的情绪聚集且所表达的情绪比较接近,因此负面关键事件会导致整体的情感平均值较低,正面关键事件会导致整体的情感平均值较高,且情感分歧度较低,如图2(a)事件6月11日至13日以及图2(d)事件12月31日至1月2日。

(2)在信息阐释期,随着信息的公开,网民所拥有的信息增多,情绪开始变得理性,讨论变得激烈,想法呈现多样性,因此该时期情感分歧度有所上升,且情感平均值更接近于0,如图2(a)事件6月13日至14日。

(3)舆情事件过渡到舆情交互期,此时发表看法的网民更加理性,此类网民的想法会相对一致,情感分歧度相对信息阐释期有所下降,如图2(b)事件9月6日至8日、图2(d)事件1月5日至7日。

值得注意的是,情感分歧度的突然上升或突然下降通常与关键事件有关,如“虎牙员工自曝被HR抬出公司”事件中,11月10日当事人回应简历造假之后引发了情感分歧度的突然上升,其他事件也存在类似的情况。因此,通过情感分歧度的变化可以及时识别关键事件,并对舆情应对措施的效果进行评估。如果舆情应对措施有力,则情感分歧度会明显下降,而情感均值这一指标无法很好地反映这一信息。

4.3 评论对象层情感分歧度计算及分析

对4个事件的所有文本进行分词、去停用词处理、统计词频并按照词频降序排列,人工筛选出与评论对象相关的153个名词并进行分类,选取4个事件中讨论热度较高的评论对象作为本文的研究对象,选取结果见支撑数据。

利用3.3节提出的评论对象层情感平均值以及情感分歧度的计算方式,分别计算4个事件评论对象每天的情感平均值及情感分歧度,计算结果如图3所示。可以看出,无论是情感平均值还是情感分歧度,同一舆情事件中的不同被评论对象的表现都有所不同。

通过评论对象层的情感分歧度分析,可以识别出事件中争议较大的评论对象。例如,“网传滴滴司机直播性侵”事件中,6月12日至14日,“表演者”和“直播平台”这两个被评论对象的情感分歧度明显增大,而其他被评论对象的情感分歧度几乎没有变化,说明这两个被评论对象是造成整个事件意见分歧的主要因素。

此外,可以通过结合情感分歧度及情感平均值的变化判断评论对象的舆论引导效果。当评论对象的情感平均值变大、情感分歧度下降或不变时,说明

绝大多数人被成功地正向引导,舆论引导效果较好;当情感平均值变大、情感分歧度变大时,说明有部分人被成功地正向引导,舆论引导有一定效果但效果并不好。例如,“网传滴滴司机直播性侵”事件中的评价对象“滴滴公司”,在6月13日至14日其情感分歧度不变,但是情感平均值上升,且从负值转向正值,说明滴滴公司针对该事件进行的及时有效回应使网友的情感向正向转变,舆论引导的效果较好。

4.4 用户层情感分歧度计算及分析

运用3.3节意见领袖识别方法确定事件的意见领袖,得到4个事件的意见领袖排名,其中4个事件排名前5的意见领袖见支撑数据。

运用3.3节提出的用户层情感分歧度以及情感分歧强度的计算方式,分别计算每个事件的机构团体账户以及个人账户情感分歧度平均值以及情感分歧强度平均值,比较机构团体账户以及个人账户情感表达差异,结果如表1所示。

与机构团体账户相比,个人账户的情感分歧度强度更大,即个人账户情感与大众情感的分歧更大。机构团体账户通常事实性报道较多,主观情感内容较少;个人账户则相反,因此个人账户对网友的情感引导作用比机构团体账户大。本文主要关注意见领袖中的个人账户。

选取4个事件排名靠前的8位意见领袖,利用公式(13)分别计算他们所发表的原创微博的情感分歧度与情感分歧强度、原创微博的转发与评论用户的情感分歧度与情感分歧强度,以及转发与评论用户的情感分歧度平均值与情感分歧强度平均值,并进行对比,结果如图4和图5所示。

如图4(a)所示,意见领袖原创微博下转发与评论用户的情感均值极性与意见领袖的情感极性基本保持一致,仅有用户“陈*n”不符合这一规律,原因是作为滴滴出行公司的总裁,其在“网传滴滴司机直播性侵”事件中表达了对滴滴司机被冤枉的愤慨之情,情感为负向,而评论与转发用户表达了对滴滴公司以及滴滴司机的支持,情感为正。

如图4(b)所示,当意见领袖的原创微博的情感分歧度为正时,该意见领袖的原创微博对大众情感存在正向影响,其转发与评论用户的情感分歧度也基本为正,反之亦然。仅有用户“耿*顺”不符合这一

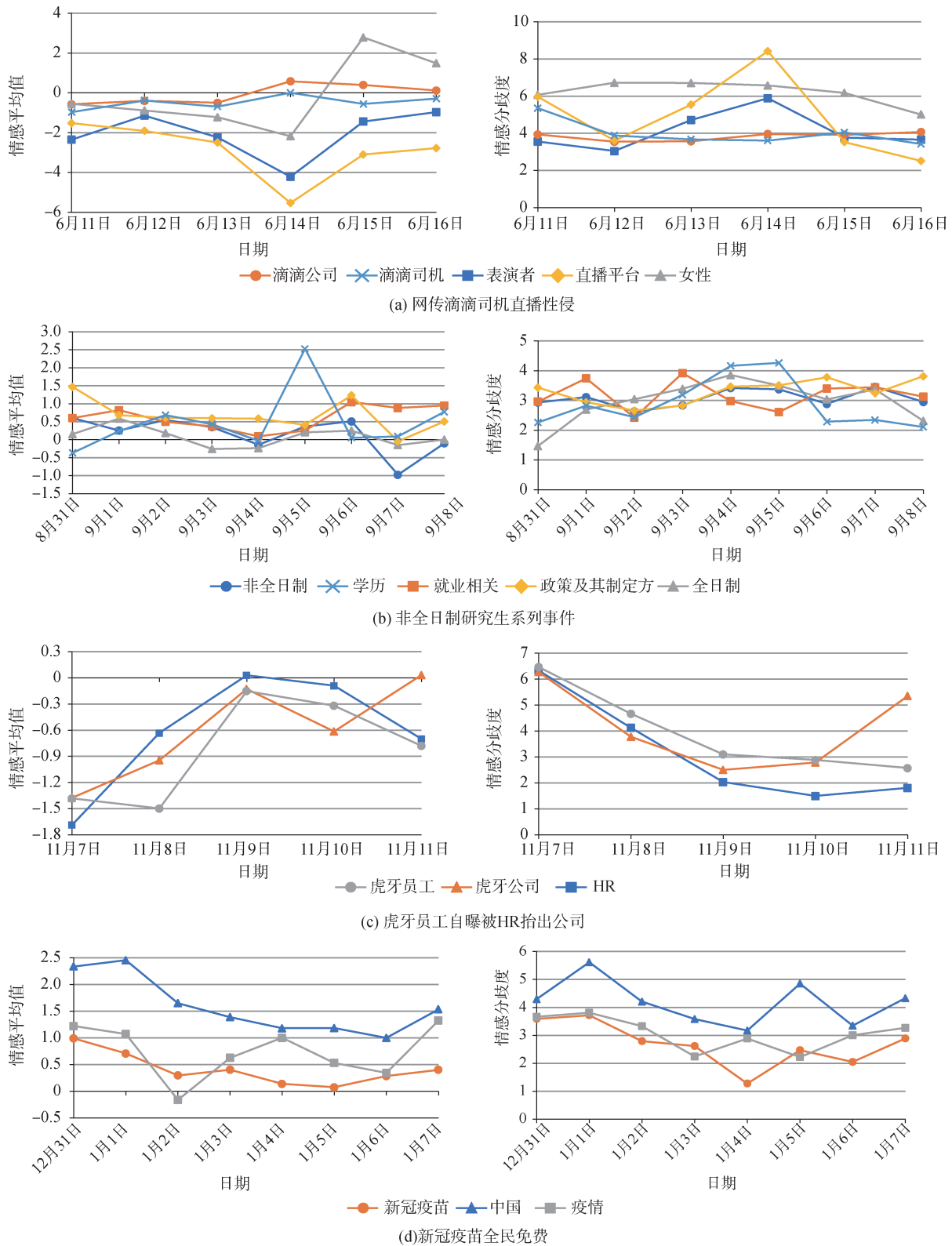


图 3 评论对象层情感平均值与情感分歧度计算结果

Fig.3 Comment Object Layer Emotional Mean and Sentiment Divergence Calculation Results

表1 机构团体账户及个人账户情感分歧度计算结果

Table 1 Account Sentiment Divergence Calculation Results

事件	账户类型	情感分歧度 平均值	情感分歧强度 平均值
网传滴滴司机直播性侵	机构团体账户	6.541	10.818
非全日制研究生系列事件	个人账户	1.815	23.204
虎牙员工自曝被HR抬出公司	机构团体账户	-0.071	4.553
新冠疫苗全民免费	个人账户	4.365	7.877
	机构团体账户	-3.746	10.106
	个人账户	-2.550	12.531
	机构团体账户	0.842	6.251
	个人账户	5.193	10.911

规律,其在“非全日制研究生系列事件”中表达了对非全日制研究生的支持,引起了转发评论用户的

积极情感,因此意见领袖“耿*顺”的情感分歧度为负,但其转发与评论用户的情感分歧度为正。

如图5所示,当意见领袖的情感分歧度强度较大时,一般情况下会导致其评论与转发用户的情感分歧强度较大。用户“耿*顺”与“陈*n”两个用户例外,不符合这一规律的原因与图4的原因一致。

综上所述,具有明显情感倾向的意见领袖在情感值、情感分歧度、情感分歧强度三方面都有一定的引领性作用,因此可通过意见领袖对舆论的情感走向进行引导。

4.5 多层次情感分歧度关联分析

(1) 事件层与评论对象层关联分析

为探究评论对象层与事件层情感分歧度的关

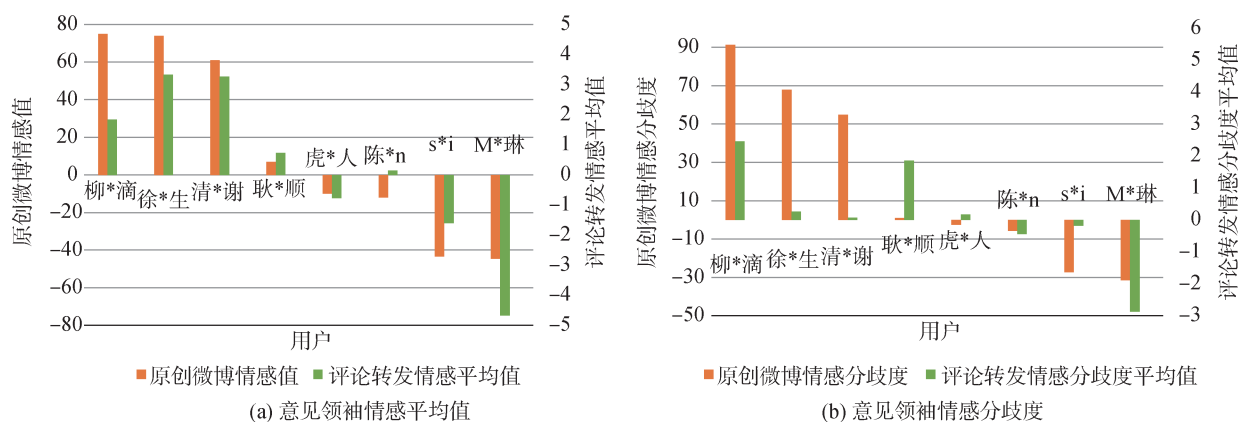


图4 意见领袖原创微博与评论转发对比

Fig.4 Comparison Between Opinion Leader's Original Microblog and Comment Forwarding

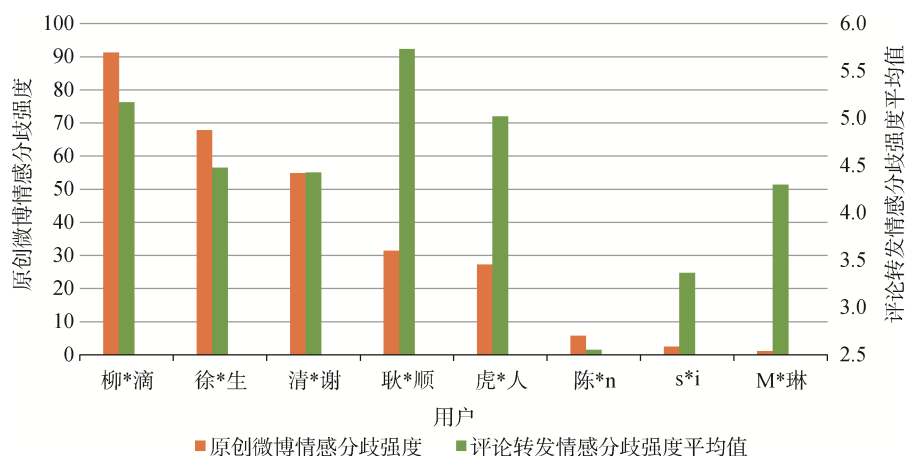


图5 意见领袖原创微博情感分歧强度与评论转发情感分歧强度均值对比

Fig.5 The Sentiment Divergence of Opinion Leader's Microblog and of Comment Forwarding

联关系,首先利用公式(14)计算 4 个事件各评论对象的加权情感分歧度,进而得到评论对象层加

权情感分歧度与事件层情感分歧度的变化情况,如图 6 所示。

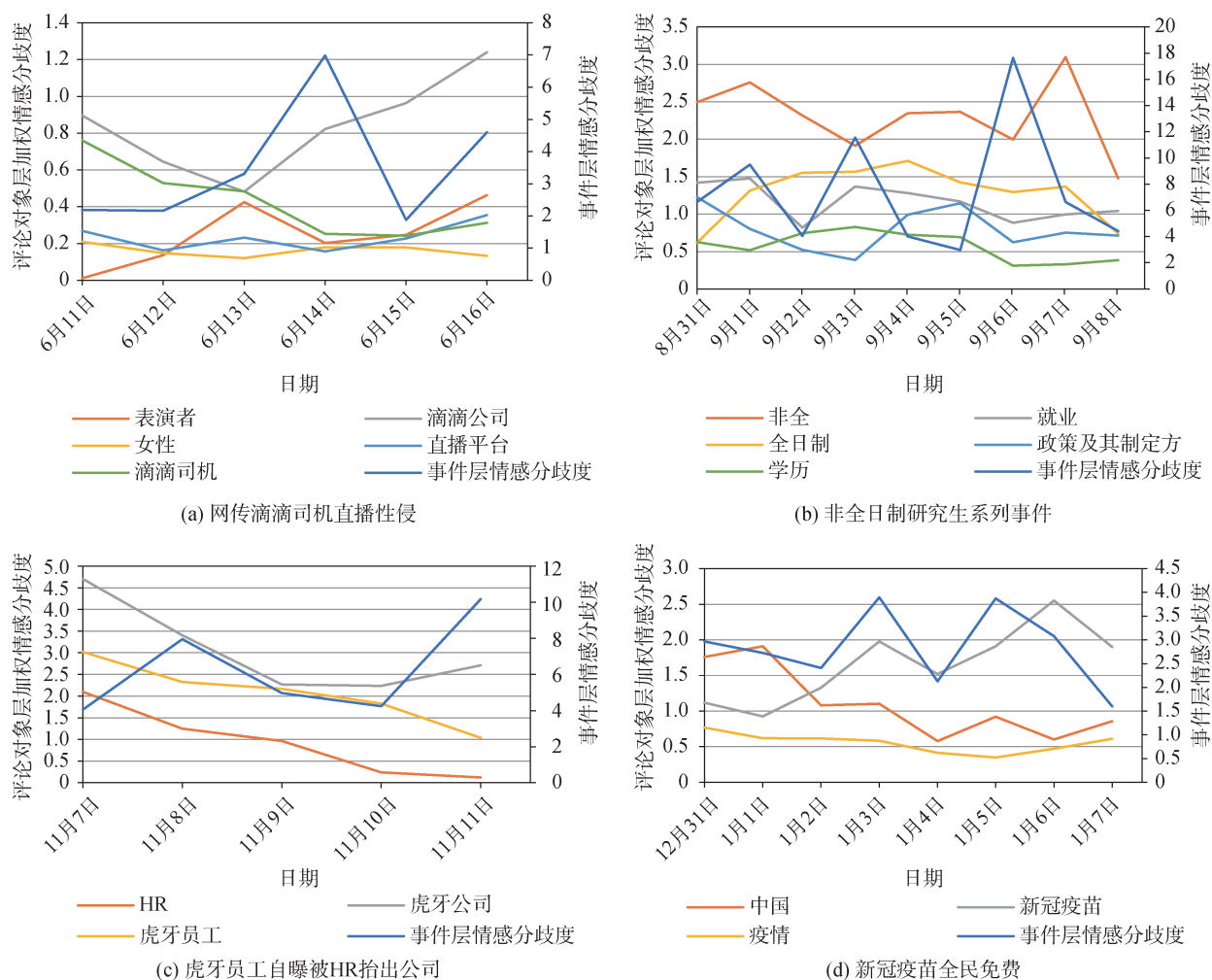


图 6 评论对象层加权情感分歧度与事件层情感分歧度变化

Fig.6 Changes of Sentiment Divergence at Comment Object Level and Event Level

在图 6 中,将 4 个舆情事件中的事件层与评论对象层分歧度变化趋势进行比较,由于被评论对象的计算将其情感分歧度和讨论热度进行了结合,可以看到,核心评论对象的情感分歧度变大时,通常会引起事件层整体情感分歧度变大。

也存在部分事件层情感分歧度走势与评论对象层加权情感分歧度走势不同的情况。由于每个事件包含多个评论对象,评论对象间的情感差异也会对事件层情感分歧度有影响。为了更清晰地观察这一影响,利用公式(15)计算评论对象间的情感分歧度,

将其与事件层整体情感分歧度比较,结果如图 7 所示。事件中不同评论对象间情感分歧度的变化与事件层整体的情感分歧度变化存在较强的一致性。评论对象间的情感分歧度变大,也会导致事件层整体情感分歧度变大。

由评论对象层对事件层情感分歧度的影响,可以更准确地定位事件层情感分歧度变化原因,从而为舆论引导提供方向。

(2) 用户层与事件层、评论对象层关联分析

通过对比去掉意见领袖及其引发的转发、评论

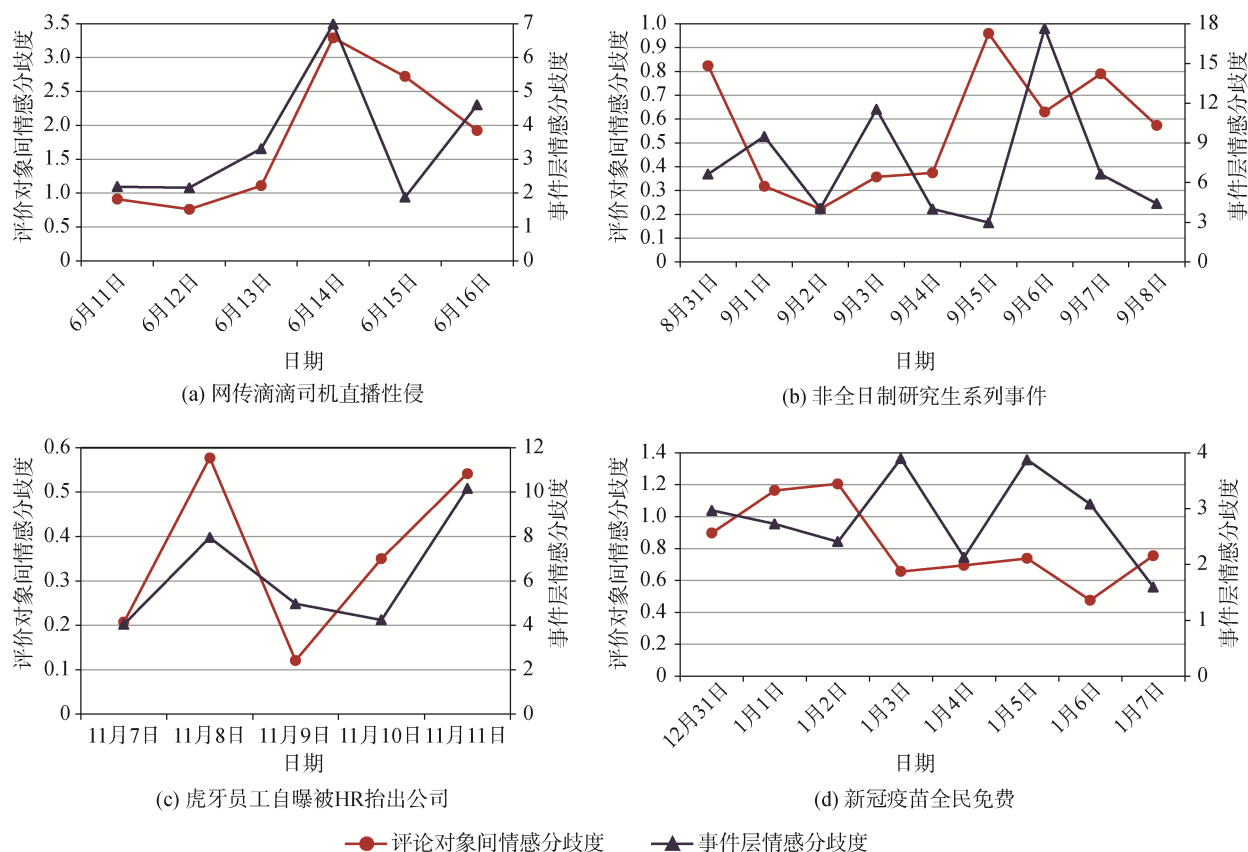


图7 评论对象间情感分歧度与事件层情感分歧度变化

Fig.7 Changes of Sentiment Divergence among Comment Objects and Changes at Event Layer

内容前后,事件层情感分歧度的变化,分析用户层的情感分歧度对事件层情感分歧度的影响。分别计算该意见领袖发文当天事件层情感平均值、情感分歧度,以及去掉该意见领袖的原创微博、原创微博下的评论及转发后,当天情感平均值、情感分歧度,将去掉该意见领袖相关文本前的情感分歧度减去去掉该意见领袖相关文本后的当天情感分歧度作为情感分歧度的变化量,绘制用户层情感分歧度、包含该意见领袖时的当天情感平均值以及情感分歧度变化量的对比图。使用相同计算方法分析用户层的情感分歧度对评论对象层情感分歧度的影响,绘制用户层情感分歧度、去掉该意见领袖前的该评论对象的当天情感平均值以及情感分歧度变化量的对比图。结果如图8所示。

如图8(a)所示,当用户情感分歧度与当天情感平均值同向时,即当天情感平均值为负,情感倾向为消极,该意见领袖的情感分歧度也为负,将用户向负

向情感引导,此时该意见领袖对事件层情感分歧度的影响无法判断。如图8(b)所示,用户层情感分歧度对评论对象层情感分歧度的影响规律与用户层对事件层影响规律相同。因此,本文主要研究用户情感分歧度与当天情感平均值反向的用户。当用户情感分歧度与当天情感平均值反向时,该意见领袖会使事件层情感分歧度变大。例如,“柳*滴”的情感分歧度与当天情感平均值反向,当天情感平均值为负,整体情感为消极,但其发表了积极的言论,将网友情感向正向引导,当去除“柳*滴”相关文本后,情感分歧度变化量为正,即情感分歧度变小,因此意见领袖“柳*滴”的加入使事件层整体情感分歧度变大。图8中意见领袖“柳*滴”“耿*顺”“陈*n”“离*原”“国*官”“今*吗”“t*员”“紫*希”均符合这一规律。

为进一步探究这一类用户影响力的持续时间,分别计算该用户发文当日、第2日、第3日引起的事件层、评论对象层情感分歧度变化量,如图9所示。

可以看出,这类意见领袖都是在发文的首日及第2日对事件层整体情感分歧度产生影响,到第3日对

整体情感分歧度的影响几乎为0,影响力持续时间非常有限。

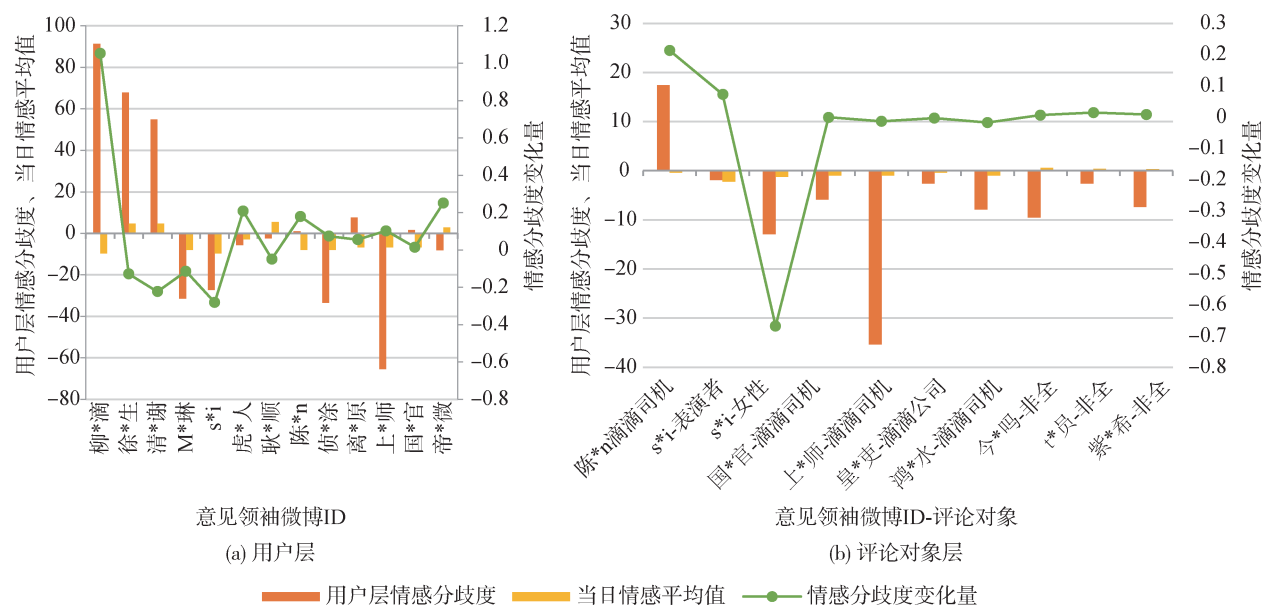


Fig.8 Influence of Opinion Leader's Views on User Layer and Comment Object Layer

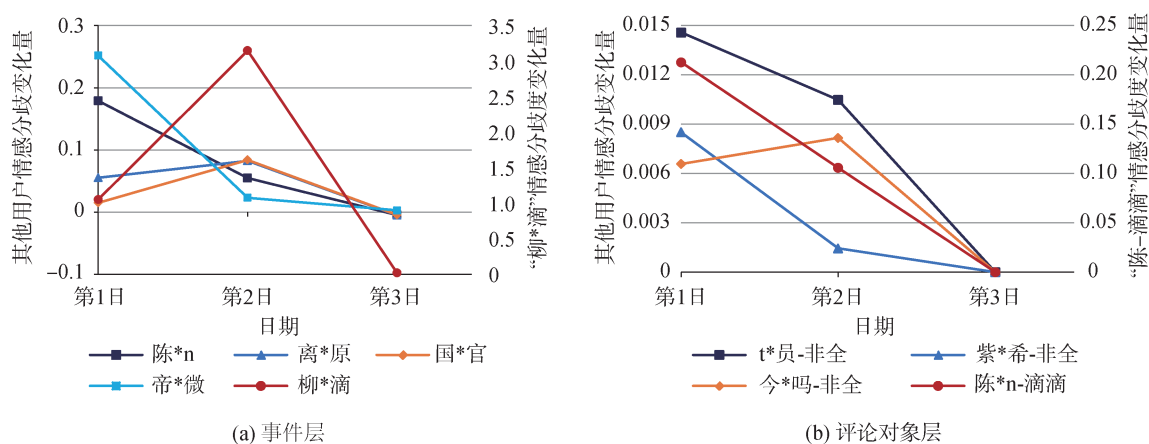


Fig.9 User-induced Changes of Sentiment Divergence of Event Layer and Comment Object Layer over Time

利用公式(16)计算“柳*滴”“耿*顺”“离*原”“国*官”“陈*n”“今*吗”“t*员”“紫*希”这8个用户的加权情感分歧度,并将前5位用户的加权情感分歧度与情感分歧度变化量平均值进行比较,探究用户加权情感分歧度与事件层情感分歧度变化量平均值的关系;对于后4位意见领袖,对比其加权情

感分歧度与情感分歧度变化量平均值的区别。如图10所示,用户加权情感分歧度指标与用户对事件情感分歧度、评论对象情感分歧度的影响程度呈正相关关系,因此可以将该指标作为意见领袖的影响力值,用以确定事件和评论对象的情感分歧度意见领袖。

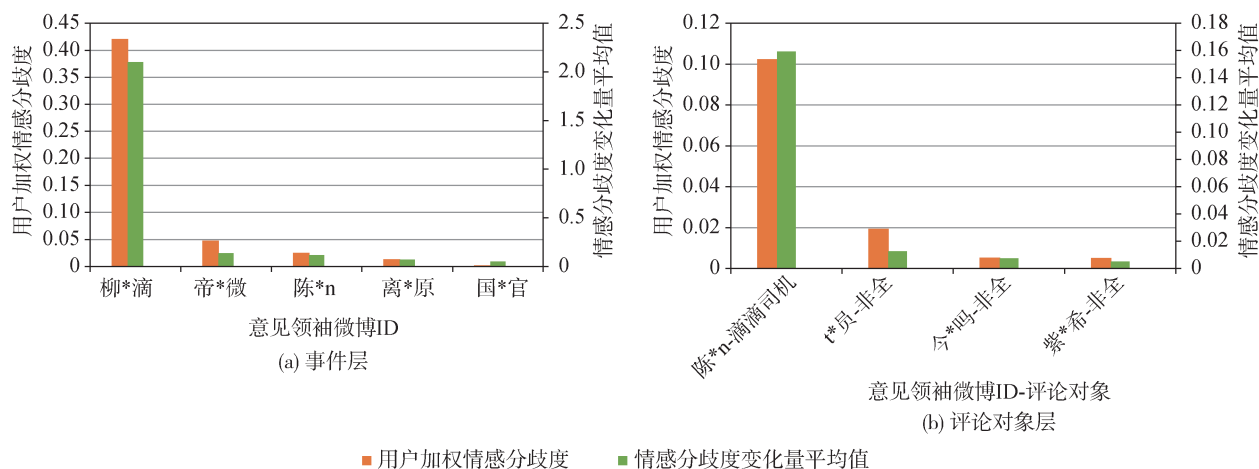


图 10 事件层、评论对象层情感分歧度变化量与用户层加权情感分歧度比较

Fig.10 Comparison Between the Change of Sentiment Divergence of Event Layer, Comment Object Layer and User Layer

5 结 语

本文参考已有情感分歧度量化算法,提出了网络舆情事件多层次情感分歧度模型,针对事件层、评论对象层、用户层三个层次特点,对每一层次的情感分歧度都进行了计算和分析,并将三个层次进行关联分析,得出以下结论。

(1)事件层情感分歧度可实现从整体观察该事件的争议性,把握该事件情感分歧度演化过程的关键拐点。

(2)评论对象层情感分歧度可满足不同主体的关注需求,实现对舆情事件更加细致的分析,根据评论对象情感分歧度及其情感平均值的大小,可对评论对象的争议性进行判断。

(3)用户层情感分歧度可用于定位情绪化意见领袖,有助于舆论情绪的监控。

(4)三个层次情感分歧度的关联关系可实现对舆情争议产生原因的精准定位。

本文的研究结论可应用于舆情的控制以及舆论的引导,针对本文结论提出以下应用建议。

(1)通过事件层中情感分歧度的突然上升或下降进行关键事件的识别,对舆情事件中关键节点进行监控。

(2)通过情感分歧度以及情感平均值的变化判断评论对象的舆论引导效果,为舆论引导效果提供

评价依据。

(3)评论对象层通过评论对象的情感分歧度、评论对象间情感分歧度共同影响事件层情感分歧度,用户层通过意见领袖影响事件层情感分歧度,根据影响因素的不同选择不同的舆论引导方式。

本文在对网络舆情事件进行多层次情感分歧度量化分析时,选取微博作为数据源,未能从豆瓣、知乎等其他平台获取数据,对网络舆情事件进行更全面的分析。

在未来研究中,可以利用该模型对同一网络舆情事件在不同平台上的数据进行分析,结合多平台数据,以实现舆情事件更全面的分析,为舆论引导提供更有针对性的建议。

参考文献:

- [1] 徐健, 吴思洋. 网络用户评论的情感分歧度量化算法研究[J]. 情报学报, 2020, 39(4): 427-435. (Xu Jian, Wu Siyang. The Quantitative Algorithm of Sentiment Divergence Based on Web User Reviews[J]. Journal of the China Society for Scientific and Technical Information, 2020, 39(4): 427-435.)
- [2] 赵菲菲, 王宇琪, 孔庆超, 等. 面向公共政策的网络媒体内容文本分析应用框架与实证——以新能源汽车政策为例[J]. 情报科学, 2020, 38(4): 140-146. (Zhao Feifei, Wang Yuqi, Kong Qingchao, et al. Application Framework and Empirical Study of Web Media Content Analysis on Public Policy—New Energy Vehicle Policy[J]. Information Science, 2020, 38(4): 140-146.)
- [3] 马续补, 陈颖, 秦春秀. 突发公共卫生事件科研信息报道的网

- 络舆情特征分析及应对策略[J]. 现代情报, 2020, 40(10): 3-10, 61. (Ma Xubu, Chen Ying, Qin Chunxiu. Research on the Characteristics of Online Public Opinion for Scientific Research Information Reporting in Public Health Emergencies and Its Coping Strategies[J]. Journal of Modern Information, 2020, 40(10): 3-10, 61.)
- [4] Fang Y J, Chen X, Song Z, et al. Modelling Propagation of Public Opinions on Microblogging Big Data Using Sentiment Analysis and Compartmental Models[J]. International Journal on Semantic Web & Information Systems, 2017, 13(1): 11-27.
- [5] 李勇, 蔡梦思, 邹凯, 等. 社交网络用户线上线下情感传播差异及影响因素分析——以“成都女司机被打”事件为例[J]. 情报杂志, 2016, 35(6): 80-85. (Li Yong, Cai Mengsi, Zou Kai, et al. Influencing Factors and Differences of the Online and Offline Emotional Spread of Social Network Users: “Female Driver Beaten by a Male Driver in Chengdu” Case as an Example[J]. Journal of Intelligence, 2016, 35(6): 80-85.)
- [6] 陈兴蜀, 常天祐, 王海舟, 等. 基于微博数据的“新冠肺炎疫情”舆情演化时空分析[J]. 四川大学学报(自然科学版), 2020, 57(2): 409-416. (Chen Xingshu, Chang Tianyou, Wang Haizhou, et al. Spatial and Temporal Analysis on Public Opinion Evolution of Epidemic Situation about Novel Coronavirus Pneumonia Based on Micro-Blog Data[J]. Journal of Sichuan University (Natural Science Edition), 2020, 57(2): 409-416.)
- [7] 李吉, 黄微, 郭苏琳, 等. 网络口碑与情感强度测度模型研究——基于 PAD 三维情感模型[J]. 情报学报, 2019, 38(3): 277-285. (Li Ji, Huang Wei, Guo Sulin, et al. Research on the Sentiment Intensity Measurement Model of Internet Word-of-Mouth Public Opinion Based on the PAD Model[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(3): 277-285.)
- [8] 洪小娟, 宗江燕, 黄卫东, 等. 基于情感语义空间的食物安全舆情情感分析[J]. 现代情报, 2020, 40(10): 132-143. (Hong Xiaojuan, Zong Jiangyan, Huang Weidong, et al. Analysis of Food Safety Public Opinion Emotion Based on Emotional Semantic Space[J]. Journal of Modern Information, 2020, 40(10): 132-143.)
- [9] 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[J]. 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[J]. Library and Information Service, 2017, 61(15): 120-129.)
- [10] 张鹏, 崔彦琛, 兰月新, 等. 基于扎根理论与词典构建的微博突发事件情感分析与舆情引导策略[J]. 现代情报, 2019, 39(3): 122-131, 143. (Zhang Peng, Cui Yanchen, Lan Yuexin, et al. Sentiment Analysis of the Micro-Blog Emergency and Related Guiding Strategy Based on the Grounded Theory and Lexicon Construction[J]. Journal of Modern Information, 2019, 39(3): 122-131, 143.)
- [11] 赵晓航. 基于情感分析与主题分析的“后微博”时代突发事件政府信息公开研究——以新浪微博“天津爆炸”话题为例[J]. 图书情报工作, 2016, 60(20): 104-111. (Zhao Xiaohang. The Study on Government News Release in the Era of Post-Microblog Based on Sentiment Analysis and Subject Analysis: A Case Study of the “Tianjin Explosion” on Sina Microblog[J]. Library and Information Service, 2016, 60(20): 104-111.)
- [12] 崔彦琛, 张鹏, 兰月新, 等. 面向时间序列的微博突发事件衍生舆情情感分析研究——以“6.22”杭州保姆纵火案衍生舆情事件为例[J]. 情报科学, 2019, 37(3): 119-126. (Cui Yanchen, Zhang Peng, Lan Yuexin, et al. Time Series-Oriented Study on the Sentiment Analysis of the Derived Public Opinion for Microblog Emergencies[J]. Information Science, 2019, 37(3): 119-126.)
- [13] 邵琦, 牟冬梅, 王萍, 等. 基于语义的突发公共卫生事件网络舆情主题发现研究[J]. 数据分析与知识发现, 2020, 4(9): 68-80. (Shao Qi, Mu Dongmei, Wang Ping, et al. Identifying Subjects of Online Opinion from Public Health Emergencies[J]. Data Analysis and Knowledge Discovery, 2020, 4(9): 68-80.)
- [14] 梁晓贺, 田儒雅, 吴蕾, 等. 基于超网络的微博相似度及其在微博舆情主题发现中的应用[J]. 图书情报工作, 2020, 64(11): 77-86. (Liang Xiaohe, Tian Ruya, Wu Lei, et al. Microblog Similarity Based on Super Network and Its Application in Microblog Public Opinion Topic Detection[J]. Library and Information Service, 2020, 64(11): 77-86.)
- [15] 梁晓敏, 徐健. 舆情事件中评论对象的情感分析及其关系网络研究[J]. 情报科学, 2018, 36(2): 37-42. (Liang Xiaomin, Xu Jian. Sentiment Analysis of Objects in Public Opinion Events and Their Relation Network Research[J]. Information Science, 2018, 36(2): 37-42.)
- [16] 辜丽琼, 夏志杰, 宋祖康, 等. 基于在线网民评论情感追踪分析的企业危机舆情应对研究[J]. 情报理论与实践, 2019, 42(12): 67-73. (Gu Liqiong, Xia Zhijie, Song Zukang, et al. Research on Enterprise Crisis Response Based on Emotional Tracing Analysis of Online Netizens' Comments[J]. Information Studies: Theory & Application, 2019, 42(12): 67-73.)
- [17] 王志英, 杨森, 王念新, 等. 信息安全突发事件情感领袖群际效应研究[J]. 管理科学, 2019, 32(1): 14-27. (Wang Zhiying, Yang Miao, Wang Nianxin, et al. Research on the Intergroup Effect of Emotional Leaders in Information Security Emergencies[J]. Journal of Management Science, 2019, 32(1): 14-27.)
- [18] 邓君, 孙绍丹, 王阮, 等. 基于 Word2Vec 和 SVM 的微博舆情情感演化分析[J]. 情报理论与实践, 2020, 43(8): 112-119. (Deng Jun, Sun Shaodan, Wang Ruan, et al. Evolution Analysis of Weibo Public Opinion Emotion Based on Word2Vec and SVM[J]. Information Studies: Theory & Application, 2020, 43(8): 112-119.)
- [19] 孔婧媛, 滕广青, 王思茗, 等. 舆情当事人回应对网民情感的影响研究[J]. 图书情报工作, 2020, 64(18): 89-96. (Kong Jingyuan,

- Teng Guangqing, Wang Siming, et al. Impact of Party's Responses on Netizens' Emotions in Public Opinion[J]. Library and Information Service, 2020, 64(18): 89-96.)
- [20] 何跃, 朱婷婷. 基于微博情感分析和社会网络分析的雾霾舆情研究[J]. 情报科学, 2018, 36(7): 91-97. (He Yue, Zhu Tingting. Research on the Haze Public Opinions Based on Sentiment Analysis and Social Network Analysis to Microblogging Data[J]. Information Science, 2018, 36(7): 91-97.)
- [21] Cao N, Lu L, Lin Y R, et al. SocialHelix: Visual Analysis of Sentiment Divergence in Social Media[J]. Journal of Visualization, 2015, 18(2):221-235.
- [22] Zhang Z, Li X, Chen Y B. Deciphering Word-of-Mouth in Social Media: Text-Based Metrics of Consumer Reviews[J]. ACM Transactions on Management Information Systems, 2012, 3(1): 1-23.
- [23] Subašić I, Berendt B. Peddling or Creating? Investigating the Role of Twitter in News Reporting[C]//Proceedings of European Conference on Information Retrieval. 2011: 207-213.
- [24] Pfitzner R, Garas A, Schweitzer F. Emotional Divergence Influences Information Spreading in Twitter[C]//Proceedings of the 6th International AAAI Conference on Weblogs and Social Media. 2012.
- [25] Siganos A, Vagenas-Nanos E, Verwijmeren P. Divergence of Sentiment and Stock Market Trading[J]. Journal of Banking & Finance, 2017, 78: 130-141.
- [26] Farina V. The Impact of Investors' Sentiment Divergence on Stock Market Efficiency[C]//Proceedings of the 2nd International Conference on Management, Economics, and Finance. 2019: 138-161.
- [27] Zheng X L, Tian H, Wan Z, et al. Game Starts at GameStop: Characterizing the Collective Behaviors and Social Dynamics in the Short Squeeze Episode[J]. IEEE Transactions on Computational Social Systems, 2022, 9(1): 45-58.
- [28] Reboredo J C, Ugolini A. The Impact of Twitter Sentiment on Renewable Energy Stocks[J]. Energy Economics, 2018, 76: 153-169.
- [29] Xing T, Wang G, Yuan L, et al. A Systematic Estimation Approach for the Importance of Engineering Characteristics Based on Online Reviews[J]. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 2020, 234(11): 1433-1447.
- [30] 李君轶, 任涛, 陆路正. 游客情感计算的文本大数据挖掘方法比较研究[J]. 浙江大学学报(理学版), 2020, 47(4): 507-520. (Li Junyi, Ren Tao, Lu Luzheng. A Comparative Study of Big Text Data Mining Methods on Tourist Emotion Computing[J]. Journal of Zhejiang University (Science Edition), 2020, 47(4): 507-520.)
- [31] 钟佳娃, 刘巍, 王思丽, 等. 文本情感分析方法及应用综述[J]. 数据分析与知识发现, 2021, 5(6): 1-13. (Zhong Jiawa, Liu Wei, Wang Sili, et al. Review of Methods and Applications of Text Sentiment Analysis[J]. Data Analysis and Knowledge Discovery, 2021, 5(6): 1-13.)
- [32] 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.)
- [33] Min K, Ma C, Zhao T, et al. BosonNLP: An Ensemble Approach for Word Segmentation and POS Tagging[C]//Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing. 2015: 520-526..
- [34] 董振东, 董强. 《知网》中文版[DB/OL]. [2021-03-14]. <https://openhownet.thunlp.org/download>. (Dong Zhendong, Dong Qiang. HowNet[DB/OL]. [2021-03-14]. <https://openhownet.thunlp.org/download>.)
- [35] Qi F, Yang C, Liu Z, et al. OpenHowNet: An Open Sememe-based Lexical Knowledge Base[OL]. arXiv Preprint, arXiv: 1901.09957.
- [36] Dong Z, Dong Q. HowNet-A Hybrid Language and Knowledge Resource[C]//Proceedings of International Conference on Natural Language Processing and Knowledge Engineering. IEEE, 2003: 820-824.
- [37] 卢伟聪, 徐健. 基于二分网络的网络用户评论情感分析[J]. 情报理论与实践, 2018, 41(2): 121-126. (Lu Weicong, Xu Jian. Sentiment Analysis of Network Users' Reviews Based on Bipartite Network[J]. Information Studies: Theory & Application, 2018, 41(2): 121-126.)
- [38] 赵常煜, 吴亚平, 王继民. “一带一路”倡议下的 Twitter 文本主题挖掘和情感分析[J]. 图书情报工作, 2019, 63(19): 119-127. (Zhao Changyu, Wu Yaping, Wang Jimin. Twitter Text Topic Mining and Sentiment Analysis under the Belt and Road Initiative [J]. Library and Information Service, 2019, 63(19): 119-127.)
- [39] 王曰芬, 王一山, 杨洁. 基于社区发现和关键节点识别的网络舆情主题发现与实证分析[J]. 图书与情报, 2020(5): 48-58. (Wang Yuefen, Wang Yishan, Yang Jie. Topic Discovery and Empirical Analysis of Network Public Opinion Based on Community Detection and Key Node Identification[J]. Library & Information, 2020(5): 48-58.)
- [40] 安璐, 易兴悦, 孙冉. 恐怖事件情境下微博影响力的预测及演化[J]. 图书情报知识, 2019(4): 52-61. (An Lu, Yi Xingyue, Sun Ran. Prediction and Evolution of the Influence of Microblog Entries in the Context of Terrorist Events[J]. Documentation, Information & Knowledge, 2019(4): 52-61.)
- [41] 郭顺利, 张向先, 李中梅. 高校图书馆微信公众平台传播影响力评价体系研究[J]. 图书情报工作, 2016, 60(4): 29-36, 43. (Guo Shunli, Zhang Xiangxian, Li Zhongmei. Research on the Evaluation System of the WeChat Official Account Admin Platform of University Libraries[J]. Library and Information Service, 2016, 60(4): 29-36, 43.)

- [42] 雷欢. 图书馆网络影响力研究: 以自媒体微博微信为视角[J]. 图书馆论坛, 2017, 37(1): 109-114. (Lei Huan. Web Impact of Library: From the WeMedia Perspective of Microblog and WeChat[J]. Library Tribune, 2017, 37(1): 109-114.)
- [43] 罗莉, 周婷. 意见领袖对网络舆情的作用[J]. 新闻战线, 2016(8): 26-27. (Luo Li, Zhou Ting. The Role of Opinion Leaders in Internet Public Opinion[J]. The Press, 2016(8): 26-27.)
- [44] 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 Lihui, Li He, Zhang Yanfeng. Research on the Identification and Influence Ranking of Network Public Opinion Leaders Based on Grey Relational Analysis[J]. Information Studies: Theory & Application, 2017, 40(9): 90-94.)
- [45] 徐涵. 大数据、人工智能和网络舆情治理[M]. 武汉: 武汉大学出版社, 2018. (Xu Han. Big Data, Artificial Intelligence and Internet Public Opinion Governance[M]. Wuhan: Wuhan University Press, 2018.)

作者贡献声明:

徐健: 提出研究思路, 设计研究方案;

俞超, 华玮, 吴婕洵: 进行实验;
吴思洋, 吴婕洵: 采集、清洗和分析数据;
华玮, 吴思洋: 论文起草;
徐健: 论文最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: huaw6@mail2.sysu.edu.cn。

- [1] 华玮. Comment_layer.xlsx. 评论对象层计算结果.
[2] 华玮. Event_layer.xlsx. 事件层计算结果.
[3] 华玮. User_layer.xlsx. 用户层计算结果.
[4] 华玮. Correlation_analysis.xlsx. 关联分析计算结果.
[5] 华玮. Original_data.zip. 原始数据.
[6] 华玮. Event_details.zip. 事件详情.

收稿日期: 2022-04-20

收修改稿日期: 2022-05-08

Analyzing Divergence of Multi-layer Sentiment for Online Public Opinion Events

Hua Wei Wu Siyang Yu Chao Wu Jiexun Xu Jian

(School of Information Management, Sun Yat-Sen University, Guangzhou 510006, China)

Abstract: [Objective] This paper provides a new model to analyze public opinion from the perspective of sentiment divergence, aiming to address online public opinion events effectively. [Methods] First, we introduced the concept of sentiment disagreement and proposed a multi-level sentiment disagreement algorithm. Then, we constructed a multi-level sentiment disagreement analysis model for online opinion events. This model could calculate sentiment values and disagreement for the online opinion event, comment object, and user layers to perform correlation analysis among the three layers. [Results] Introducing sentiment disagreement can compensate for the lack of research on netizens' opinion divergence in the original sentiment analysis. This model can identify the critical nodes of public opinion events and the comments generating significant controversy. It also evaluates the effectiveness of public opinion guidance and locates the causes of controversies. [Limitations] We only retrieved the needed data from Sina Weibo (Microblog). More research is needed to collect data from social platforms like Douban and Zhihu. [Conclusions] The proposed model can be applied to monitor the key nodes of public opinion events, select different public opinion guidance methods based on the reasons for controversies, and evaluate the effectiveness of public opinion guidance.

Keywords: Sentiment Divergence Network Public Opinion Analysis Multi-level Relevance Analysis