

基于文本语义的热点事件网络暴力分析方法

刘玉文¹, 翟菊叶¹, 朱文婕^{1,2}, 谢静¹

(1. 蚌埠医学院, 安徽 蚌埠 233030;
2. 中国科学技术大学 计算机科学与技术学院, 安徽 合肥 230027)

摘要:网络暴力识别与多维度分析对网络舆情态势感知和管控具有十分重要的意义,当前的网络暴力研究主要集中在用户负面情感分析、舆情危机治理及外部网络生态优化等方面,缺乏对网络暴力的定量分析研究,无法在复杂的网络环境中及时感知网络暴力发展态势和组成结构。通过分析网络暴力在文本中的存在形式和结构特征,提出了一种基于文本语义的网络暴力分析方法(text semantic based approach for cyber violence analysis, TSCA)。该方法首先运用互信息理论创建暴力领域情感词典,根据暴力领域情感词典和语义环境从评论语料库中生成负面情感词组集;然后,通过卡方检验对负面情感词组集进行暴力特征筛选,生成暴力词组集;最后,从文本和用户角度对网络暴力进行定量计算和多维度分析。在真实的网络热点事件评论文本数据集上与其他方法进行了对比,实验结果表明:该方法达到了良好的网络暴力特征识别效果。

关键词:文本语义;网络暴力;互信息;情感词典;暴力计算

中图分类号:TP391.1

文献标识码:A

文章编号:1673-629X(2022)07-0208-08

doi:10.3969/j.issn.1673-629X.2022.07.036

A Text Semantics Based Approach for Cyber Violence Analysis on Hot Event

LIU Yu-wen¹, ZHAI Ju-ye¹, ZHU Wen-jie^{1,2}, XIE Jing¹

(1. Bengbu Medical College, Bengbu 233030, China;
2. School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: Cyber violence identification and multidimensional analysis are of great significance to the situation awareness and control of network public opinion. The current research on cyber violence mainly focuses on the analysis of users' negative emotions, the governance of public opinion crisis and the optimization of external network ecology, which is unable to sense the development trend and composition structure of cyber violence in a complex network environment. By analyzing the form and structure of cyber violence in text, a text semantic based approach for cyber violence analysis (TSCA) is proposed. Firstly, a violence domain emotion dictionary is created based on mutual information theory, and a negative emotion phrase set is generated from the comment corpus according to the violence domain emotion dictionary and semantic environment. Then, through χ^2 test, the violence features set is screened based on the negative emotion phrase set. Finally, quantitative calculation and multi-dimensional analysis are carried out from the perspective of text and users. The TSCA approach is compared with other methods on the real text data set of network hot event reviews. The experimental results show that the proposed approach achieves a better performance on cyber violence feature recognition.

Key words: text semantics; cyber violence; mutual information; emotion dictionary; violence computing

0 引言

随着计算机与通讯技术的快速发展,网络朝着终端移动化、平台多元化、带宽高速化的方向发展^[1]。中国互联网络信息中心(CNNIC)发布的第47次《中国互联网络发展状况统计报告》显示:截止2020年12月,中国网民数量为9.89亿,已占全球网民总数的五

分之一。网络打破了时间限制,在给人们带来便利的同时,也给网络暴力提供了滋生环境。当前,社会热点事件是诱发网络暴力的主要因素之一,是因为社会热点事件具有突发性和聚焦性,能在短时间内吸引网民聚集,其所产生的网络暴力威力大,不仅会对当事人的心理造成极大的影响,也会极大地破坏网络生态环境,

收稿日期:2021-07-19

修回日期:2021-11-22

基金项目:安徽省哲学社会科学规划项目(AHSKQ2019D070)

作者简介:刘玉文(1982-),男,副教授,研究方向为数据挖掘、网络舆情。

给社会和谐发展带来严峻挑战。所以,热点事件中网络暴力分析对掌握网络暴力内部规律,寻找合适的网络暴力治理策略具有十分重要的意义。

自社交网络诞生之日起,网络暴力就如影随形。为了应对网络暴力带来的威胁与挑战,众多学者从社交网络入手开展了大量的网络舆情治理相关研究,提出了多种网络舆情处理框架,确立了网络话题识别^[2]、网络社区检测^[3]、意见领袖发现^[4]、用户肖像刻画^[5]、情感计算^[6]等相关核心技术,取得了一系列研究成果。如 Chang 等^[7]提出了一种基于智能语义框架的网络话题识别方法,该方法通过知识框架生成机制从文档中提取规则模式,再通过句法结构和语义关联来有效检测文档主题。Xing 等^[8]对新媒体环境下负面网络舆情检测指标体系进行研究,提出了基于信息熵的网络舆情检测模型,实现了舆情等级的自动分类。Yang 等^[9]对情绪因素影响下的突发事件网络舆情演化机理进行研究,把包含情绪函数的 RDEU 理论引入突发事件网络舆情演化博弈中,构建以网民和政府为代表的动态博弈模型,最后通过假设收益数值实现情绪影响下的演化仿真。

近些年,又有学者依托知识图谱提出了舆情危机治理新思路,实现了舆情感知智能化。如 Shen 等^[10]提出了基于网络新闻语料的公共危机事件知识图谱构建方法,该方法从新闻语料中抽取实体,构建公共危机事件知识图谱模型。实现了实体关系抽取、知识融合、知识加工和知识推理等多项技术的整合。Bao 等^[11]提出了非结构化语料知识图谱构建方法,该方法运用词法分析和语义技术构建评论文本中实体与实体、实体与属性关系,把非结构化数据映射成知识图谱,实现了文本内容的逻辑表示,解决了网络舆情回溯和推理问题。梁野等^[12]提出了一种面向舆情分析与预警领域的跨语言知识图谱架构,完成了涵盖多来源的面向舆情分析与预警领域的跨语言知识图谱构建平台 CLOpin,高效地实现了多源数据整合,解决了夸语言舆情分析与预警问题。

从当前研究成果中可以发现,网络舆情治理主要集中在负面舆论分析、舆情知识表示、舆论等级划分、舆情预警等方面,缺少对网络暴力内部数据特征的研究,无法深入了解网络暴力的内容组织关系。针对以上问题,在文本语义和情感词典技术的基础上,该文提出了一种基于文本语义的网络暴力分析方法(TSCA)。该方法从网络暴力数据特征入手,首先根据语境创建网络暴力领域情感词典,实现文本实体情感词汇的识别;然后根据文本分词的位置关系,运用语义规则生成负面情感词组集,并用卡方检验对暴力特征进行筛选;最后通过语义相似度对暴力特征进行计

算,从暴力词和用户两个维度实现对网络暴力的多尺度分析。

1 相关技术

1.1 卡方检验

从文本内容的组成结构角度,语义分析包括词语级语义分析、句子级语义分析和文档级语义分析。但不管从哪个层级进行语义分析,首要任务都是解决文本特征的识别问题。文本特征识别是自然语言处理的核心技术,特征识别方法包括文档频率法(document frequency, DF)、信息增益法(information gain, IG)及卡方检验法等。其中,卡方检验法是最常用的文本特征选择方法。

卡方检验^[13]主要功能是统计样本的实际观测值与理论推断值之间的相关程度,卡方值越大,则两者相关的可能性就越大, χ^2 统计量的计算公式如式(1)所示。

$$\chi^2(\text{Feature}_i, C_j) = \frac{M \times (A \times D - C \times B)^2}{(A + C) \times (B + D) \times (A + C) \times (B + D)} \quad (1)$$

其中, Feature_i 表示特征, C_j 表示特征所属类别; M 表示训练集中文本总数, A 表示属于 C_j 类且包含 Feature_i 的文本频数, B 表示不属于 C_j 类但包含 Feature_i 的文本频数, C 表示属于 C_j 类但是不包含 Feature_i 的文本频数, D 表示既不属于 C_j 类也不包含 Feature_i 的文本频数。

1.2 文本情感分析

1997年,MIT媒体实验室Picard教授首次提出情感计算(affective computing, AC)概念^[6],目的是通过外部表露出的信息来研究人的内在心理活动。经过20多年的发展,情感分析延伸出了面部情感分析、姿态情感分析、自然语言情感分析、多模态情感计算等多个研究分支,该文研究的文本情感分析属于自然语言情感分析范畴。

当前,网络成为了民众信息交互和观点表达的主要平台,网络上存储了大量的文本数据,隐含着很多有价值信息,对文本进行情感分析能够感知网络舆情态势,了解民众需求,对促进社会管理水平的提高具有十分重要的意义。文本作为自然语言的一种存储形式,最小的语义单位是词,词语级情感分析是最基础的文本语义分析技术,同时也为句子级和文档级情感分析提供了知识支撑。词语级情感分析技术运用情感词典来衡量测试词汇的情感值,并依据语义规则进行情感计算,具体分析流程如图1所示。

从图1中可以看出,计算规则是情感计算的核心,它的主要任务是依据情感词典对新词情感进行量化,

量化方式包括互信息计算、相似度对比。

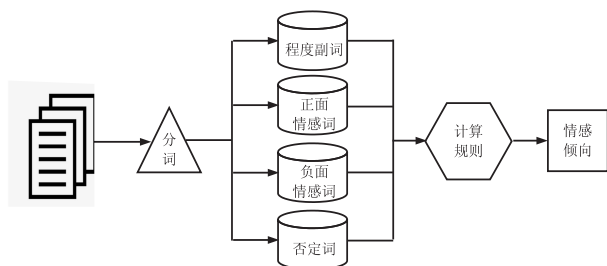


图1 文本情感分析框架

2 基于文本语义的网络暴力分析方法 (TSCA)

从本质上说,文本暴力分析是情感分析的一个典型应用。由于热点事件的突发性和聚焦性,网民评论紧紧围绕着热点事件展开,语言相似性大,领域特性强,用经典的情感词典进行语义对比,情感计算结果准确度不高。另外,网络暴力是带有侮辱性的攻击语言,负面情感色彩强烈,但负面情感不一定是网络暴力,如何从负面情感中筛选出暴力特征是网络暴力分析的关键。所以,在网络暴力分析流程中要解决两个核心问题:(1)暴力领域情感词典建立;(2)基于负面情感特征的网络暴力筛选。

2.1 文本暴力语义规则

情感词典是一种按照文本内词汇之间的逻辑关系搭建起来的知识库,它的建立需要遵循文本语义规则。文本由句子组成,句子又由词组成。词是文本的最小语义单位,但是词在表达语义的时候,由于缺少前缀修饰词,意义表达往往不明确。如果单纯的从词的角度去挖掘暴力信息,挖掘出的信息可能会不准确。比如:“人渣”是个暴力词,但如果语境是“他不是人渣”,在“人渣”的前面有个否定前缀,那整个语义就不具备暴力,所以,在文本暴力挖掘时,往往词组才是具备有效语义的基本单位。

从文学角度来说,词组可分成很多种类型,如:主谓词组(大家帮助);动宾词组(发现问题);介宾词组(把他打死);偏正词组(崇高的理想);否定词组(不喜欢)。从这些词组的语义环境来看,能显著对情感进行修饰或改变的只有偏正词组和否定词组。进一步研究发现,否定词和修饰词的位置在文本语义表示中非常重要,比如,“不很美丽”和“很不美丽”两者意义截然相反。另外,正情感词和负情感词前面加否定词时,语义翻转的程度也不同,如:正情感词“好人”加否定后得“不是好人”,语义翻转到了好人的对立面,即“坏人”;而负情感词“坏人”加否定前缀后的“不是坏人”,语义就不能翻转到“好人”,从语境理解,“不是坏人”只能表明“不坏”,但不能表明是“好”。因此,该文对

具有否定的词组结构进行如下规定:

规则1:否定前缀+情感词。

如果正情感词前面有否定前缀,则否定权值为1;若负情感词前面有否定前缀,则否定权值为0.5。

规则2:否定前缀+修饰词。

当情感词前面同时含有否定词和修饰词时,若否定词在修饰词前,则否定权值为0.5;若在修饰词后,则取值为1。

2.2 领域情感词典建立

评论文本是用户发表意见的载体,文本包含用户的情感、态度、行为等特征,文本的情感计算可以转化为对语义词组的情感计算^[14]。由于在不同的语境下词汇表达的含义可能会不一样,所以,首先需要依据语义环境创建情感词典,然后,用测试文本的分词与情感词典进行语义对比,得出测试分词的情感值。情感词典由3个数据表组成:情感词表 W_c 、程度副词表 W_p 和否定词表 W_n 。情感词表包含6个属性:前缀否定词ID、前缀副词ID、词项、性质(褒义或贬义)、情感值、位置;程度副词表包含3个属性:词项、强度值和位置;否定词表也包括3个属性:词项、否定值(否定一般设置为-1)和位置。

(1)情感词表创建。

根据当前语料的语义环境,人工筛选出 N 个核心情感种子词。种子词筛选原则是情感意义非常明确,具有代表性,情感极性最强。首先运用情感种子词建立原始情感词表,然后用循环的方法依次遍历文本分词,当 $i=1$ 时,把文本 d_i 中的新词 w_i 与情感词表中的所有词汇进行语义互信息计算,把互信息最高的词汇情感值作为新词 w_i 的情感值,并填充进情感词表。依次循环,使得每个文本上的分词都会被遍历一次。以下介绍情感词表的扩充方法。

设 w_n 是 D 中的新词, w_i 是情感词表 W_c 内的词项, w_n 的情感值可以通过与 w_i 的点间互信息计算得到,计算公式如式(2)所示:

$$\text{PMI}(w_n, w_i) = \lg \frac{P(w_n, w_i)}{P(w_n) \times P(w_i)} \quad (2)$$

其中, $P(w_n, w_i)$ 表示 w_n 和 w_i 在 D 中共现的概率, $P(w_n)$ 和 $P(w_i)$ 分别表示 w_n 和 w_i 单独出现的概率, $\text{PMI}(w_n, w_i)$ 的取值范围为 $[0, 1]$ 。如果 $\text{PMI}(w_n, w_i) > \delta$ (δ 表示 w_n 和 w_i 的相似性阈值),则判定 w_n 和 w_i 的语义相同。为了综合判断词汇 w_n 的情感倾向,还需要在情感词表 W_c 上对 w_n 进行综合考察,设情感词表 W_c 正面情感词集合为 W_{cp} ,负面情感词集合为 W_{cn} ,词汇 w_n 的综合值计算公式如式(3)所示:

$$S(w_n) = \frac{\sum_{w_p \in W_{cp}} \text{PMI}(w_n, w_p)}{|W_{cp}|} - \frac{\sum_{w_n \in W_{cn}} \text{PMI}(w_n, w_p)}{|W_{cn}|} \quad (3)$$

从公式(3)可以看出, $S(w_n)$ 的取值范围也是 $[-1, 1]$, 绝对值越大, w_n 的极性越强。最后判断 w_n 在 W_c 内是否为新词, 如果是新词, 则把 w_n 直接加入到 W_c 中, 否则对原有词的情感值进行覆盖。按照上述方法计算 D 中所有分词, 完成对情感词表 W_c 的扩充。对任意情感词 w_n , 若 $S(w_n) > 0$, 则 w_n 具有正情感倾向; 若 $S(w_n) = 0$, 则 w_n 具有中性情感倾向; 若 $S(w_n) < 0$, 则 w_n 具有负情感倾向。

(2) 程度副词表创建。

程度副词表扩充与情感词表扩充的过程相似, 不同之处是程度副词的修饰程度取值范围为 $[0, 1]$ 。为了提高程度副词的扩充精度, 根据6级划分理论, 按梯度下降公式对每个级别赋予不同的权值, 具体创建方式如下:

设 w_i 是程度副词, W_v 为程度副词表, 为了度量 w_i 的修饰程度, 需要在 W_v 上对 w_i 进行综合计算, 如公式(4)所示:

$$\text{Adv}(w_n) = \frac{\sum_{w_i \in W_v} k \times \text{PMI}(w_i, w_n)}{|W_v|} \quad (4)$$

其中, k 表示 w_v 的权值。根据公式(4)的计算结果, 首先判断 w_i 在 W_v 内是否为新词, 如果是新词, 则把 w_i 直接加入到 W_v 中; 如果不是新词, 则对原有词的修饰强度进行覆盖。遍历分词预料库中所有副词完成副词表 W_v 的扩充。

(3) 否定词表创建。

否定词能让情感词的极性发生翻转, 如果情感词前有否定词, 在情感计算时, 根据否定词出现的次数 e , 用 $(-1)^e$ 乘以情感强度。由于否定词意义明确, 其意义表达不依赖于语义环境, 通常情况下以“不”、“没”、“未”等字组成, 在 HowNet 词典中收录非常完善, 所以, 该文借鉴 HowNet 中的否定词表, 以此填充到创建的领域词典内。

2.3 负面情感计算

负面情感词组检测。

网络暴力本质是极端恶劣的负面情感, 所以文本暴力分析是情感计算技术的一种具体应用, 目标是从负面情感特征中识别出暴力特征, 并对暴力特征进行计算。文本是词汇的集合, 文本的情感隐含在情感词中, 文本情感通过计算文本词汇的综合情感值来实现。为了更清楚地表述词汇语义, 特做如下定义:

定义1: 语义词组。设五元组 $W = \langle w, N, \text{Adv}, M, Q \rangle$ 是个语义词组, w 代表情感词, N 代表 w 的否定前缀, Adv 代表 w 的修饰前缀, M 代表否定词位置, Q 代表 w 的极性。

语义词组 W 能够明确描述情感词的语义环境, 其

情感的计算过程如下: 首先, 依次遍历文本内情感词, 根据文本中第 n 个分词 w_n 所在的位置读取出前缀词, 并与 w_n 连接, 组成语义词组 W , W 的情感值计算如公式(5)所示:

$$S(W) = N(w_n) \text{Adv}(w_n) S(w_n) M(w_n) Q(w_n) \quad (5)$$

其中, $N(w_n)$ 表示 w_n 的否定词向量权重; $\text{Adv}(w_n)$ 表示 w_n 修饰词向量权重, 两个分量可继续分解为公式(6)和公式(7):

$$N(w_n) = (-1)^e, e \geq 0 \quad (6)$$

$$\text{Adv}(w_n) = \sum_{i=1}^V \text{Adv}(w_n^i) \quad (7)$$

在公式(5)~(7)中, e 表示否定词出现的次数, V 表示程度副词出现的数量。 $M(w_n)$ 表示否定词的位置, 初始值为1; $Q(w_n)$ 表示情感翻转系数, 初始值为1。根据该文定义的语义规则: ①若否定词在修饰词前, 则 $M(w_n)$ 取值为0.5; 若在修饰词后, 则取值为1; ②当 $S(w_n) > 0$ 时, 若 w_n 有否定前缀, 则 $Q(w_n)$ 取值为1; 若 $S(w_n) < 0$ 时, 则 $Q(w_n)$ 取值为0.5。由公式可以计算出词组 W 的情感值。如果 $S(w)$ 的值小于0, 则 W 为负情感词组。

2.4 网络暴力计算

(1) 网络暴力词组筛选。

网络暴力作为一种攻击性语言, 具有很强的负情感, 但负情感特征词不一定是暴力语言。暴力是负情感的充分条件, 负情感是暴力的必要条件。所以网络暴力特征需要从负面情感特征中进一步筛选。 χ^2 检验是常用的文本特征筛选方法, 运用 χ^2 检验的网络暴力筛选过程如下:

第①步: 从负情感词组集中, 用人工的方法标注出负面情感最强烈的20个暴力词组作为种子存放在词向量 C 中, 并设定 χ^2 检验阈值 $\xi = 0.90$;

第②步: 利用卡方检验对负情感词组进行特征计算, 按从大到小顺序, 选择 $\chi^2(W^i, C) \geq \xi$ 的特征词组, 添加到 C 中;

第③步: 增加阈值 ξ , 使得 $\xi = \xi + 0.01$, 返回到第②步, 直到选不出暴力特征词组为止。

通过暴力特征词的筛选能把不具备暴力的负面情感词组过滤掉, 大大提高了网络暴力识别精度。从情感计算角度看, 网络暴力是负值, 取值范围是 $[-1, 0)$ 。为了直观地对网络暴力进行描述, 文本对包含网络暴力的负面情感进行翻转计算, 把网络暴力取值范围映射到区间 $(0, 1]$ 内, 转换后的暴力词组计算公式如式(8)所示:

$$\text{Cyber}(W) = -N(w_n) \text{Adv}(w_n) S(w_n) M(w_n) Q(w_n) \quad (8)$$

定义2: 暴力密度。设词组集 $WS = \{W_1, W_2, \dots,$

$W_K\}$,暴力语义词组集 $WC = \{W_{c1}, W_{c2}, \dots, W_{cK}\}$,且 $WC \subseteq WS$,则 WS 的暴力密度计算公式如式(9)所示:

$$\text{Density}(W) = \frac{\sum_{W_i \in WC} \text{Cyber}(W_i)}{|WS|} \quad (9)$$

(2)文本暴力计算。

文本是单词的有序集合,即: $d_i = \{w_1, w_2, \dots, w_N\}$ 。但从词组的角度看,文本又可表示成多个语义词组的顺序排列,即: $d_i = \{W_1, W_2, \dots, W_K\}$,且 $\{W_1 \cup W_2 \cup \dots \cup W_K\} = \{w_1, w_2, \dots, w_N\}$ 。所以,基于词组的文本暴力计算公式如式(10)所示:

$$\text{Cyber}(d_i) = \sum_{k=1}^K \text{Cyber}(W_k), W_k \in C \quad (10)$$

其中, C 表示暴力词组向量。然后对所有文本的暴力值进行归一化处理,计算公式如式(11)所示:

$$\text{Cyber}(d_i^*) = \frac{\text{Cyber}(d_i) - \text{Cyber}_{\min}(d_i)}{\text{Cyber}_{\max}(d_i) - \text{Cyber}_{\min}(d_i)} \quad (11)$$

其中, $\text{Cyber}_{\max}(d_i)$ 表示文本暴力的最大值, $\text{Cyber}_{\min}(d_i)$ 表示文本暴力的最小值。

(3)用户暴力计算。

用户暴力是用户所发表评论文本的暴力之和,所以,对于用户 u_i ,其暴力计算公式如式(12)所示:

$$\text{Cyber}(u_i) = \sum_{n=1}^{|D_i|} \text{Cyber}(d_n) \quad (12)$$

其中, $|D_i|$ 表示用户 u_i 发表的评论文本数量。对用户暴力进行归一化处理,计算公式如式(13)所示:

$$\text{Cyber}(u_i^*) = \frac{\text{Cyber}(u_i) - \text{Cyber}_{\min}(u_i)}{\text{Cyber}_{\max}(u_i) - \text{Cyber}_{\min}(u_i)} \quad (13)$$

其中, $\text{Cyber}_{\max}(u_i)$ 表示用户暴力的最大值,

$\text{Cyber}_{\min}(u_i)$ 表示用户暴力的最小值。

3 实验分析

3.1 数据来源及预处理

以“合肥母子三人跳楼事件”为例,使用八爪鱼数据采集器获取了腾讯新闻中该主题下的评论信息:用户节点 1 052 个,评论文本 3 128 条。首先抽取用户之间的“回复、点赞”关系,建立用户关系库 U ,然后使用 ICTCLAS 分词软件对评论文本进行分词,去除停用词、介词、语气词、转折词等无用词后,建立文本语料矩阵 G 。

U 的大小为 $|U| \times 4$,其中,第 i 行存放第 i 个用户 u_i 的相关信息,存放顺序是:第 1 列存放 u_i 的序号,第 2 列存放 u_i 的 ID,第 3 列存放 u_i 回复的用户 ID 串,第 4 列存放 u_i 点赞的用户 ID 串。 G 的大小为 $|G| \times [2 + N + 3]$,其中,第 i 行存放第 i 条评论文本 d_i 的相关信息,存放顺序是:第 1 列存放 d_i 的序号;第 2 列存放发表 d_i 的用户 ID,后 N 列存放 d_i 的分词结果。初始语料矩阵准备完毕之后,运用该文提出的方法对语料信息进行处理分析。

3.2 实验结果分析

(1)暴力词语义分析。

在语料矩阵 G 的基础上,按照文本暴力分析路线,首先创建领域情感词典 S ,生成负面情感词组集 N ,再通过卡方检验筛选出暴力词组集 C 。组成暴力词组集的高频基础情感词及情感值计算结果如表 1 所示。

表 1 事件评论网络暴力高频词列表(Top 45)

序号	词项	词频	情感值	序号	词项	词频	情感值	序号	词项	词频	情感值
1	男人	112	-0.46	16	恶心	27	-0.75	31	温暖	15	0.46
2	责任	102	0.61	17	死	27	-0.79	32	鸟人	15	-0.79
3	自私	98	-0.73	18	本事	23	0.59	33	赚钱	14	0.61
4	冷血	92	-0.73	19	恐怖	23	-0.62	34	魔窟	14	-0.68
5	残忍	88	-0.91	20	没用	23	-0.71	35	珍惜	14	0.49
6	可怕	87	-0.39	21	孤独	21	-0.21	36	谴责	14	-0.43
7	渣男	76	-0.78	22	人品	21	0.25	37	了断	13	-0.83
8	冷漠	73	-0.83	23	陪葬	21	-0.88	38	活刮	12	-0.92
9	懦弱	51	-0.41	24	悲哀	21	-0.35	39	反省	12	-0.31
10	可恨	43	-0.69	25	嫌弃	21	-0.43	40	葬送	12	-0.43
11	不配	37	-0.63	26	逼迫	20	-0.52	41	地狱	12	-0.83
12	无耻	38	-0.85	27	报应	20	-0.60	42	永不超生	10	-0.83
13	愤怒	37	-0.57	28	东西	19	-0.90	43	老不死	10	-0.92
14	靠谱	31	0.80	29	畜生	18	-0.83	44	丢脸	10	-0.36
15	嫁错	29	-0.38	30	刽子手	17	-0.76	45	枪毙	10	-0.91

(2)暴力多尺度分析。

为了反映暴力特征词之间意义分布形态,对 45 个

暴力特征词进行多维尺度统计分析。根据暴力特征词两两之间共现频数,得到大小为 45×45 的暴力特征词

共现矩阵。由于频次的范围变化大,数据分析不方便,为了消除词频计数对分析的影响,用 Ochiaa 系数将共词矩阵转换为相关矩阵,并对共现频次进行归一化处理,再根据词与词之间的共现关系,建立两词之间的相异矩阵,计算公式如式(14)所示:

$$Ochiaa = \frac{P(A,B)}{\sqrt{P(A)} \times \sqrt{P(B)}} \tag{14}$$

为了便于理解,用“1”与矩阵相减,得到表示两词建相异程度的相异矩阵,转换结果如表 2 所示。

表 2 Ochiaa 相异矩阵

	男人	责任	自私	冷血	残忍	可怕	渣男	冷漠	懦弱	可恨	不配	无耻	愤怒	靠谱	嫁错
男人	0.0000	0.9157	0.9039	0.9129	0.9086	0.9079	0.9435	0.9756	0.9231	0.9847	0.9095	0.9479	0.9621	0.9246	0.9085
责任	0.9086	0.0000	0.9821	0.9953	0.9627	0.9845	0.9132	0.9867	0.9365	0.9084	0.9757	0.9315	0.9043	0.9964	0.9352
自私	0.9764	0.9537	0.0000	0.9647	0.946	0.9031	0.9845	0.9279	0.9754	0.9453	0.9253	0.9154	0.9174	0.9632	0.9258
冷血	0.9532	0.9163	0.9325	0.0000	0.9731	0.9427	0.9325	0.9075	0.9473	0.9137	0.9186	0.9068	0.9285	0.9364	0.9735
残忍	0.9086	0.9073	0.9845	0.9326	0.0000	0.9075	0.9537	0.9148	0.9365	0.9243	0.9256	0.9254	0.9194	0.9243	0.9783
可怕	0.9378	0.9481	0.9052	0.9086	0.9648	0.0000	0.9946	0.9364	0.9787	0.9248	0.9398	0.9375	0.9025	0.9089	0.9635
渣男	0.9768	0.9361	0.9237	0.9452	0.9063	0.9352	0.0000	0.9683	0.9875	0.9463	0.9232	0.9368	0.9143	0.9326	0.9243
冷漠	0.9326	0.9748	0.9637	0.9764	0.9374	0.9086	0.9753	0.0000	0.9654	0.9545	0.9365	0.9647	0.9076	0.9647	0.9673
懦弱	0.9352	0.9638	0.9953	0.9235	0.9953	0.9793	0.9536	0.9574	0.0000	0.9623	0.9245	0.9357	0.9265	0.9251	0.9846
可恨	0.9178	0.9073	0.9057	0.9095	0.9235	0.9241	0.9235	0.9834	0.9642	0.0000	0.9287	0.9356	0.9647	0.9354	0.9274
不配	0.9046	0.9628	0.9836	0.9105	0.9968	0.9085	0.9524	0.9643	0.9874	0.9125	0.0000	0.9474	0.9366	0.9034	0.9735
无耻	0.9834	0.9532	0.9053	0.9576	0.9642	0.9734	0.9342	0.9631	0.9352	0.9056	0.9477	0.0000	0.9483	0.9021	0.9243
愤怒	0.9523	0.9238	0.9426	0.9947	0.9658	0.9213	0.9176	0.9432	0.9543	0.9074	0.9065	0.9286	0.0000	0.9241	0.9362
靠谱	0.9749	0.9068	0.9327	0.9042	0.9084	0.9753	0.9186	0.9332	0.9453	0.9143	0.9264	0.9056	0.9274	0.0000	0.9282
嫁错	0.9241	0.9846	0.9740	0.9542	0.9705	0.9895	0.9397	0.9758	0.9231	0.9562	0.9187	0.9024	0.9086	0.9462	0.0000

把相异矩阵数据输入到 SPSS 软件中,选择二维分析组图输出方式,对其进行 Euclidean 距离分析,分析结果如图 2 所示。图中点与点之间的距离表示词汇所属内容主题的相关程度,距离越小表明主题语义相似度越大,主题内容越集中;反之,距离越远表明表达内容越独立。

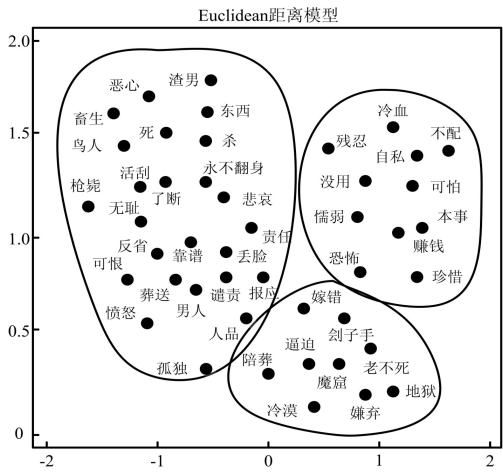


图 2 暴力词汇多尺度分析结果

从图 2 显示的分析结果来看,词汇之间既有渗透交叉,又有群组分布的独立性,说明词汇语义反映出了明显的主题差异性。从词汇表达的内容角度出发,人工对词汇进行主题分组,根据划分边界大致可分成三

个群组,每个群组代表一个被网络暴力攻击的实体对象,分别为:跳楼女子(以下简称“女方”)、女子丈夫(以下简称“男方”)、女子所处的家庭环境(以下简称“家庭”)。按照分组结果,对语料中每个群组词及群组内暴力词进行统计,结果如图 3 所示。

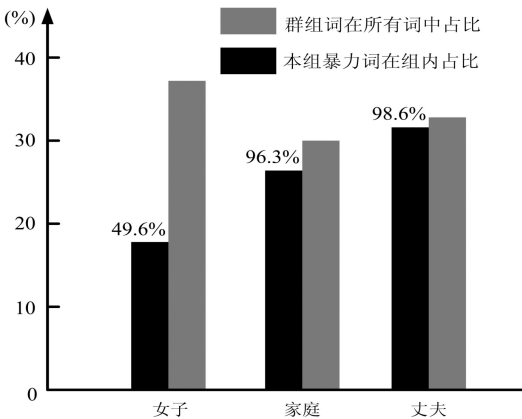


图 3 各群组暴力词占比

从图 3 可以看出,在所有语料词组中,女方群组词数量的占比为 38.2%,家庭群组词占比为 29.7%,丈夫群组词占比为 32.1%。网民针对女方发表的网络评论最多,也从侧面反映出了女方在事件中的主体地位。但从暴力词汇在群组内的占比来看,针对丈夫的暴力词占比最高,达到 98.6%,其次针对家庭的暴力

词占比为 96.3%,说明网友的暴力攻击点主要集中在男方和家庭,而针对女子的暴力词占比也有 49.6%,说明了虽然在整个事件中女方是受害者,但女方绑架子女生命的赴死方式也遭到了众多网友的鄙视。

为了度量不同群组遭受的暴力,运用公式(9)对每个群组的暴力密度进行计算,了解不同对象遭受暴力的强度,三个群组的暴力密度计算结果如图 4 所示。

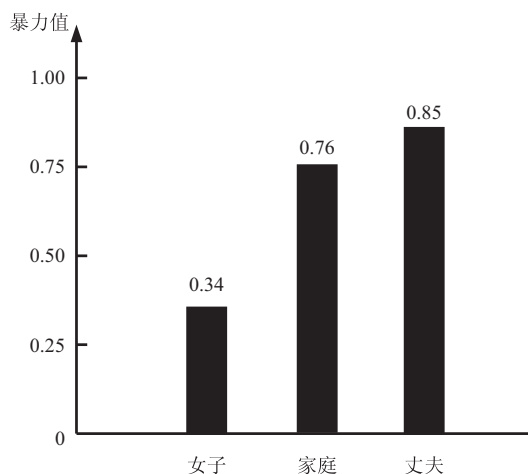


图 4 三个群组网络暴力密度对比

通过群组词汇量整体占比、暴力词在群组内占比以及群组词汇的暴力密度等信息,可以对不同对象所遭受的网络暴力做如下分析。

①对男方的暴力攻击分析:通过图 3 和图 4 可知,攻击男方的暴力特征词占比最大,暴力密度也最高。

说明事件中女子的丈夫是被网民攻击的主要对象。从评论文本中可以看出,网民一致认为他是促使妻子跳楼的最直接元凶,评论用词激烈极端,遭受的网络暴力最强。

②对家庭的暴力攻击分析:从暴力词占比以及暴力密度对比结果来看,网民另一个攻击对象是女子所处的家庭环境。通过对原始语料分析可知,女子独自带孩子,孩子身体不好,丈夫不仅不给生活费,还遭受公婆嫌弃,引起了网民的极大愤恨和道德谴责。

③对女方的暴力攻击:虽然女方是受害方,本应该受到网民同情,但她以极端的方式剥夺了孩子的生命,从性质上来说,她也是杀人犯,展现出了她可怕、冷血的一面;另外也突出了她自私、懦弱的性格特征。网民在同情女子遭受不幸的过程中,同样也对她进行了指责。

(3)暴力用户分析。

用户是网络暴力的发起者,对用户暴力进行分析有助于掌握网络暴力实施的人群分布。具体分析过程是:首先遍历语料数据库 G ,以用户为单位检索出用户发表的评论文本分词,并与暴力词组集 C 对照匹配,找出用户的暴力词组;再根据公式(10)和(11)计算用户文本的暴力值,然后再用公式(12)和(13)计算出用户的暴力值;最后把用户暴力词组分别与三个群组暴力词组进行相似度对比,识别出用户暴力攻击的目标,计算结果如表 3 所示。

表 3 暴力用户信息表(暴力值 Top 40)

序号	ID	暴力语义词组	暴力值	攻击目标	序号	ID	暴力语义词组	暴力值	攻击目标
1	U ₁₂₈	冷血、永不超生、自私、(不)靠谱	0.94	男方/女方	21	U ₆₉	嫁错、冷漠、(非常)可恨、寒心	0.80	男方/家庭
2	U ₅₆₃	(不是)东西、渣男、畜生	0.92	男方	22	U ₂₆₇	(没)资格、自私、剥夺	0.80	女方
3	U ₁₅	报应、去死、(没)责任	0.92	男方	23	U ₇₂₈	(太)自私、(真)狠心、(不能)理解	0.79	女方
4	U ₂₆₈	刽子手、千刀万剐、可恨	0.92	男方	24	U ₃₄₂	无耻、(没)责任心、孤独	0.79	男方
5	U ₃₆₉	万劫不复、畜生、(不是)男人	0.92	男方	25	U ₉₄₄	无能、恐怖、下手	0.79	女方
6	U ₁₅₇	渣男、(很)无情无义、(不)靠谱	0.91	男方	26	U ₈₀₆	死男人、可恨、反省	0.78	男方
7	U ₉₇₁	(太)可恨、天打雷劈、死	0.91	男方	27	U ₇₃₅	无权、自私	0.78	女方
8	U ₈₃₁	老东西、(不是)好鸟	0.91	家庭	28	U ₃₂₇	陪葬、男人、(没有)依靠	0.78	男方/家庭
9	U ₂₀₆	鸟人、恶心、自行了断	0.90	男方	29	U ₄₃₄	无权、残忍、剥夺	0.78	女方
10	U ₇₆₃	渣男、(没)责任、(不是)男人	0.89	男方	30	U ₂₆₁	(太)冷漠、(非常)可怕	0.77	女方
11	U ₃₀₉	魔窟、(没)人情味、(不)疼爱	0.88	男方/家庭	31	U ₇₉₄	(非常)可悲、葬送	0.77	男方
12	U ₅₇₂	(不)珍惜、(不是)男人、人品差	0.86	男方	32	U ₁₈₅	悲哀、(不)靠谱、丢脸	0.77	男方
13	U ₇₂₈	地狱、枪毙、没用	0.83	男方/女方	33	U ₆₇₃	冷漠、无情、(不)和谐	0.77	家庭
14	U ₁₅₅	(不是)东西、男人、去死	0.85	男方	34	U ₃₃₈	(太)残忍、掠夺	0.76	女方
15	U ₄₉₀	嫁错、老不死、(没有)温暖	0.85	家庭	35	U ₅₀₁	(不)可靠、(没)责任	0.76	男方
16	U ₁₀₀₂	无耻、人品(差)、丢脸	0.83	男方	36	U ₆₅	不配、自私、恐怖	0.76	女方
17	U ₈₇₆	(不是)东西、枉为人父、	0.82	男方	37	U ₈	(不)体贴、什么人	0.76	男方
18	U ₆₁₈	老不死、(没有)温暖、冷血	0.82	家庭	38	U ₉₇₂	刻薄、(没)温暖	0.76	家庭
19	U ₃₉₄	可怕、恐怖	0.80	女方	39	U ₄₁₃	(无法)想象、(多么)恐怖	0.75	女方
20	U ₇₅₈	鄙视、啥意义、可悲	0.80	男方	40	U ₆₇₂	(不)和谐、压抑	0.75	家庭

3.3 性能评价

暴力特征词组的筛选是该文的核心,为了验证该方法的暴力特征识别效果,对实验语料中的负面情感词组和暴力特征词组进行人工标注,并把原始语料库分成5个数据集,分别用TSCA和FWCC^[15]方法对数据集的暴力特征进行提取。为了提高人工标注的可信度和准确性,采用三组标注法,以高分表决的方式标注语料中的暴力特征词组,两种方法的评价对比结果如表4所示。

表4 暴力特征识别性能对比 %

数据集	TSCA			FWCC		
	准确率	召回率	F 值	准确率	召回率	F 值
D ₁	86.49	90.76	88.57	85.71	86.85	86.28
D ₂	89.20	88.95	89.07	87.29	84.31	85.77
D ₃	85.61	86.44	86.02	89.37	85.70	87.50
D ₄	90.07	87.62	88.83	84.94	86.78	85.85
D ₅	91.75	90.13	90.93	86.08	87.12	86.60

从表4可以看出,在五组语料数据集上,TSCA方法的暴力特征词组识别性能(F值)高于FWCC方法,说明TSCA方法达到了良好的文本暴力特征识别效果。

4 结束语

运用信息熵理论建立了网络暴力领域情感词典,并以领域情感词典为基础,结合词法分析技术建立了基于文本语义的网络暴力分析方法。该方法能从暴力词和用户两个维度度量热点话题下的网络暴力信息,拓展了网络暴力识别与分析方法。经过实验,该方法在多维网络暴力分析方面达到了良好的效果。由于该方法中暴力阈值的设定是人工实现的,存在一定的局限性。如何实现阈值的自动划分是未来需要进一步研究的方向。

参考文献:

- [1] 端祥宇,袁冠,孟凡荣.动态社区发现方法研究综述[J].计算机科学与探索,2021,15(4):612-630.
- [2] BLEI D M,NG A Y,JORDAN M I,et al. Latent Dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [3] GUENDOUZ M,AMINE A,HAMOU R M. A discrete modified fireworks algorithm for community detection in complex networks[J]. Applied Intelligence,2017,46(2):373-385.
- [4] 张米,张晖,杨春明,等.基于多特征信息传播模型的微博意见领袖挖掘[J].中文信息学报,2018,32(2):129-138.
- [5] RUAN Q,WU Q,WANG Y,et al. Effective learning model of user classification based on ensemble learning algorithms[J]. Computing,2018,101(6):531-545.
- [6] GRATCH J. Editorial[J]. IEEE Transactions on Affective Computing,2010,1(1):1-10.
- [7] CHANG Y C,HSIEH Y L,CHEN C C,et al. A semantic frame-based intelligent agent for topic detection[J]. Soft Computing, 2017,21(2):391-401.
- [8] XING Yunfei,WANG Xiwei,WANG Duo,et al. Research on the negative network public opinion monitoring index system based on information entropy in new media environment[J]. Journal of Modern Information,2018,38(9):41-47.
- [9] YANG Yang,WANG Jie. The evolution of emergency network public opinion influenced by motional factors[J]. Information Science,2020,38(3):35-41.
- [10] SHEN Yunfeng,WANG Yingjie. Knowledge mapping of public crisis events based on internet news corpus[J]. Information Science,2021,39(1):72-80.
- [11] 鲍玉来,耿雪来,飞龙.基于卷积神经网络的旅游信息关系抽取研究[J].现代情报,2019,39(8):132-136.
- [12] 梁野,李小元,许航,等. CLOpin:一种面向舆情分析与预警领域的跨语言知识图谱架构[J].数据分析与知识发现,2020,4(6):1-14.
- [13] QIN H,FEI Q,MA X,et al. A new parameter reduction algorithm for soft sets based on chi-square test[J]. Applied Intelligence,2021,51(11):7960-7972.
- [14] PANG Shanchen,YAO Jiamin,LIU Ting,et al. A text similarity measurement based on semantic fingerprint of characteristic phrases[J]. Chinese Journal of Electronics,2020,29(2):233-241.
- [15] 朱晓旭,钱培德.脏话文本语料库建设[J].计算机工程与应用,2014,50(11):126-129.