

doi: 10.16104/j.issn.1673-1891.2022.04.016

基于 LDA 模型和卡方检验的网络暴力话题挖掘方法

谢 静^a, 刘玉文^{b*}

(蚌埠医学院 a. 公共基础学院; b. 卫生管理学院, 安徽 蚌埠 233030)

摘要: [目的] 社交网络中存在着许多暴力话题, 暴力话题识别对网络舆情的精准干预和管控具有十分重要的意义。当前网络暴力研究主要集中在用户负面情感计算、暴力用户识别等领域, 缺乏对网络暴力组织构成研究, 无法在复杂网络环境中精准识别网络暴力的附着载体。[方法] 通过分析网络暴力在话题内的聚焦特性, 提出了一种基于 LDA 模型和卡方检验的网络暴力话题识别方法, 该方法首先运用 LDA 模型识别网络语料库中的话题, 并用相似度计算方法对话题文本进行分类; 然后运用卡方检验筛选话题文本中的暴力特征; 最后依据情感词典计算各话题内的暴力值, 按照暴力密度判断话题的暴力属性。[结果/结论] 在真实的网络语料库上实验验证了本文方法, 实验结果表明: 本文方法的暴力话题识别性能 (F 值) 均值为 80.64%, 优于对比方法, 达到了良好的网络暴力话题识别效果。

关键词: LDA 模型; 卡方检验; 暴力话题; 文本分类; 情感计算

中图分类号: TP311.13; G206 **文献标志码:** A **文章编号:** 1673-1891(2022)04-0097-07

An LDA Model and Chi-Square Test Based Approach to Cyber Violence Topic Mining

XIE Jing^a, LIU Yuwen^{b*}

(a. School of Public Basic; b. School of Health Management, Bengbu Medical College, Bengbu, Anhui 233030, China)

Abstract: [Objective] Violence topic identification is important for accurate intervention and control of network public opinion. Current research on cyber violence focuses mainly on the calculation of net users' negative emotions, violence sponsors' identification etc., and there's lack of analysis and research on the composition of cyber violence organizations, resulting in failures to accurately identify the attached carriers of cyber violence in the complex network environment. [Method] By analyzing focuses of cyber violence topics, an approach to identifying violence topics based on LDA model and chi-square test is proposed. Firstly, the topics in cyber language corpus are identified by LDA model, and the topic texts are classified by similarity calculation method. Then, the chi-square test is used to determine the violence characteristics of the topic texts. Finally, the violence value of each topic is calculated according to the dictionary of emotions, and the violence attribute of the topic is judged according to the violence value. [Result/Conclusion] Experiments on a real cyber corpus verify our proposed method. The experimental results show that the violence topic identification performance (F value) of this method is averaged at 80.64%, which, superior to the contrast, has achieved a better result in the identification of cyber violence topics.

Keywords: LDA model; chi-square test; violence topic; text classification; emotion computing

0 引言

随着交互式多媒体网络技术的迅速发展, 网络上出现了众多自媒体交互平台(如抖音、快手、小红

书等)。自媒体平台无限上传、转发、评论、点赞等互动功能, 使信息交互更加快捷多样, 人们只要拥有一部智能手机就能轻松成为一个自媒体人。由于网络的跨时空特性^[1], 社会热点事件极易在网

收稿日期: 2022-06-07

基金项目: 安徽省哲学社会科学规划项目(AHSKQ2019D070)。

作者简介: 谢静(1985—), 女, 安徽滁州人, 讲师, 硕士, 研究方向: 网络语义。*通信作者: 刘玉文(1982—), 男, 安徽滁州人, 教授, 硕士, 研究方向: 数据挖掘, 网络舆情。

上形成热点话题,加速了网络舆情的形成。

目前,由于网络缺乏道德约束和有效的监管^[2],网络上滋生着大量网络暴力,网络话题是网络暴力的主要载体。由于网络具有无边界性^[3],暴力话题一旦形成,围观网民数量会迅速增长,对当事人会造成非常大的心理伤害。网络暴力的存在严重破坏了网络生态环境,阻碍了社会的和谐发展,所以,开展网络暴力话题识别研究有助于对网络暴力进行组织分类,对舆情管控和引导具有十分重要的意义。

从交互式网络技术出现以来,网络暴力就随即出现。随着自媒体技术的普及应用,网络暴力更加盛行。网络暴力的形成涉及网络、社会、心理等多种因素,一直以来都是网络舆情领域研究的重点^[4]。为了应对网络暴力给社会生态文明带来的冲击,降低网络暴力危害,众多学者从多个层面开展了相关研究工作。刘绩宏等^[5]从道德层面分析了网络暴力演化过程中的影响因素,并以 821 位网民的上网数据为基础,构建了网民心理和行为的结构方程模型,发现道德焦虑能够使网民对相关主体产生消极的道德判断,从而促使网民实施网络暴力行为;石经海等^[6]从法律层面分析了网络暴力刑法规制的困境,建议将网络暴力现象纳入相关法律治理体系中予以立体化治理,增设网络暴力罪,以解决网络暴力刑事责任追究难的问题;田圣斌等^[7]从社会管理层面分析法律在应对网络暴力时存在的缺陷,强调网络暴力治理首先要精准识别网络暴力,然后实施阶段性治理举措,加强行政干预,规范网络治理行为,实现网络社会的综合治理。上述研究主要从社会和心理行为角度分析了网络暴力形成原因及治理方式,为网络暴力的管理提出了新思路,但存在的问题是不能从网络大数据上了解网络暴力特性,无法运用技术手段对网络暴力进行引导。为了能及早探测网络暴力,缩短舆情干预响应时间,一些学者从网络本身入手研究网络暴力形成及发展规律,提出了观点演化^[8]、主题识别与演化^[9]、情感计算^[10]、社区发现^[11]等相关技术。谷学汇^[12]针对网络暴力单一模态检测方法精度低等问题,提出了基于文本、视频以及音频多模态融合算法,将文本、视音频分类器当作预分类器完成视频的初始分类,获得候选暴力影视集,随后再运用视音频分类器对候选集进行分类,最终完成对网络暴力的识别;范涛等^[13]为了提高负面情感内容识别精度,提出了基于图卷积神经网络和依存句法分析的网民负面情感分析模型,该模型运用双向长短期记忆网络和自注意力机制抽取文本特征作为依存句法图中的节点

特征,再应用图卷积神经网络对生成的节点特征邻接矩阵进行训练学习,输出负面情感类别。以上研究成果实现了暴力内容的精确查找,但存在的问题是网络暴力组织属性识别能力较差,无法实现网络事件与网络暴力的关系映射。

针对上述问题,本文提出了一种基于 LDA 模型和卡方检验的网络暴力话题识别方法。该方法首先运用 LDA 模型对网络数据进行主题分类,然后再用卡方检验对主题内暴力特征进行识别,最后通过情感计算技术定量计算话题内的暴力密度,依据密度阈值对网络暴力主题进行识别。

1 相关技术介绍

1.1 卡方检验

社交网络中,文本形式的语言暴力(以下简称文本暴力)是网络暴力的主要表现形式,文本暴力语言带有强烈的负面情感特征。本质上说,文本暴力识别是自然语言处理技术的一种具体应用。识别过程包括:文本分词、特征词提取、情感计算、语义关联等,其中特征词提取是关键。目前,文本特征词识别方法^[14]主要包括信息增益法、文档频率法及卡方检验法等,其中卡方检验法是常用的文本特征选择方法。

卡方检验^[15]是用于度量样本理论推断值与实际观测值相关程度的统计方法,卡方值越大,表明二者相关性越大,卡方统计量用 χ^2 表示,计算公式如(1)所示。

$$\chi^2(F_i, C_j) = \frac{m \times (a \times d - b \times c)^2}{(a + b) \times (c + d) \times (a + c) \times (b + d)} \quad (1)$$

式中:变量 F_i 表示特征; C_j 表示特征类别; m 表示训练数据集内文本个数; a 表示包含 F_i 的 C_j 类文本数量; b 表示包含 F_i 的非 C_j 类文本数量; c 表示不包含 F_i 的 C_j 类文本数量; d 表示不包含 F_i 的不属于 C_j 类文本数量。

1.2 LDA 模型

1.2.1 模型描述

隐狄利克雷分配(Latent Dirichlet Allocation, LDA)是一种基于无监督学习的文档生成模型^[16]。该模型认为一篇文档包含多个主题,每个主题又对应着不同词汇;一篇文章的构造过程依赖文档、主题、词汇之间的概率假设。首先文档以一定的概率选择某个主题,然后主题又以一定的概率选择某个词,不断重复执行这个过程,直到完成整个文档的生成为止。LDA 模型根据文档生成的逆过程,对文

档主题特征和主题词汇特征进行学习,把文档建模成主题分布,又把文档中的主题建模成词汇分布。从生成过程来看,LDA模型是一个3层贝叶斯网络,根据概率依赖关系,定义了4个变量:文本-主题分布矩阵 θ ,主题-词汇分布矩阵 φ ,主题向量 Z 以及词汇 W 。

1.2.2 模型参数介绍

除了4个变量之外,LDA模型还包含2个超参数 α 和 β 。 Z 是隐变量, W 是唯一可实际观测到的文档词汇。各变量在模型中的含义如表1所示,模型结构如图1所示。

表1 LDA模型变量含义

变量	含义说明
α	主题在文档内服从Dirichlet分布的超参数
β	词汇在主题内服从Dirichlet分布的超参数
θ	文档-主题分布矩阵
φ	主题-词汇分布矩阵
Z	文档中的主题向量
W	文档中可实际观测到的词汇
N	文档内词汇总数
M	语料库中文档总数

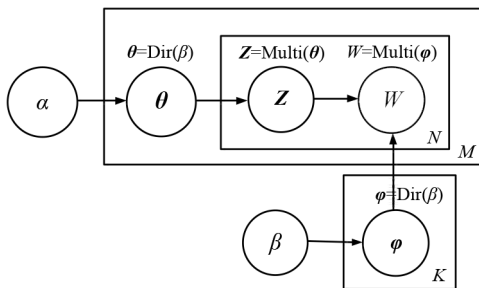


图1 LDA模型的贝叶斯网络结构

1.2.3 文档生成公式

LDA模型先运用超参数 α 生成一个文档主题概率分布,然后运用 β 生成 N 个主题概率分布,最后生成整篇文档的 N 个词的联合概率,如式(2)所示。

$$P(\theta, Z, W|\alpha, \beta) = P(\theta|\alpha) \prod_{n=1}^N P(z_n|\theta) P(w_n|z_n, \beta) \quad (2)$$

由于 θ 和 Z 是隐变量,需要对其进行边缘概率求解,从而消除 θ 和 Z ,最后计算得到文本中每个单词的生成概率,计算公式如(3)所示。

$$P(w|\alpha, \beta) = \int P(\theta|\alpha) \left(\prod_{n=1}^N \sum_{z_n} P(z_n|\theta) P(w_n|z_n, \beta) \right) d\theta \quad (3)$$

通过文档中可观测的词汇,运用吉布斯采样过

程对公式(3)进行反复迭代,并利用最大期望算法(expectation Maximization,简称EM算法)对参数 θ 和 φ 值进行估计,最终训练出文档-主题分布矩阵 θ 和主题-词汇分布矩阵 φ 。虽然LDA模型存在着运算效率高、建模方便等诸多优点,但也存在着主题数 K 需要依赖人为经验事先设定等缺陷,大大增加了建模结果的不确定性。

1.2.4 动态话题数计算

为了降低主题数人为设定给建模结果带来的影响,研究者提出了一系列动态主题数计算方法。其中,基于贝叶斯主题数计算方法是最流行的方法,计算公式如(4)和(5)所示。

$$P(w|z) = \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \right)^K \sum_{z=1}^K \frac{\prod_w \Gamma(n_z^w + \beta)}{\Gamma(n_z + W\beta)} \quad (4)$$

$$\frac{G}{P(w|K)} = \sum_{g=1}^G \frac{1}{P(w|z^g)} \quad (5)$$

式(4)~(5)中: $\Gamma(x)$ 是伽马函数; n_z^w 表示主题 z 内词汇 w 出现的次数; n_z 表示主题 z 内词汇总数; K 表示主题数量; G 表示吉布斯采样次数; z^g 表示第 g 次采样时的主题。

结合公式(4)和(5)对 $P(w|K)$ 进行求解,当 $P(w|K)$ 达到极值时,选择 K 值为最优主题数。

2 网络暴力话题识别方法

网络暴力话题识别包括2个步骤:第1步,运用主题模型从语料文本中对主题进行识别,构造出文档-主题和主题-词汇2个分布矩阵;再用主题词与文档进行相似度计算,对文档进行主题文本分类,得到主题文本集。第2步,对每个主题文本集进行暴力特征筛选,运用情感计算技术计算话题的暴力情感值,根据暴力密度阈值对话题的暴力属性进行判断。具体识别过程如图2所示。

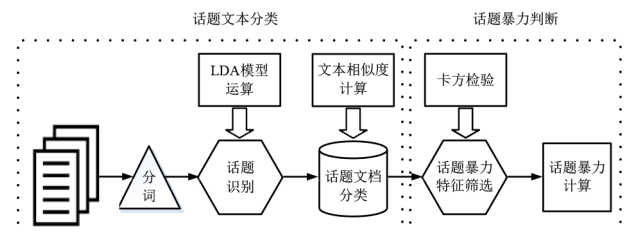


图2 暴力话题识别流程

网络暴力是网络文本内的一种情感形式,带有强烈的负情感,所以,网络暴力可以借助情感实现定量计算。本文运用话题暴力密度作为判定话题暴力属性的依据,话题暴力密度含义如定义1所示。

定义 1:话题暴力密度 设话题 z 由话题词汇集 $A=\{w_1, w_2, \dots, w_x\}$ 表示, 话题内暴力词汇集 $B=\{w_{c1}, w_{c2}, \dots, w_{cy}\}$ 表示, 且 $B \subseteq A$, 则话题 z 的暴力密度 $TD(z)$ 指 B 内词汇情感绝对值之和与 A 内所有词汇情感绝对值之和的比值, 计算公式如(6)所示。

$$TD(z) = \frac{\sum_{w_i \in A} |S(w_i)|}{\sum_{w_j \in B} |S(w_j)|} \quad (6)$$

式中: $S(w)$ 表示词汇 w 的情感计算函数。当 $TD(z)$ 大于设定的阈值时, 则判定话题 z 为暴力话题。

2.1 话题文本分类

话题文本分类的目的是把文本按照话题进行划分, 获取话题文本集。话题文本分类需要运用话题识别和相似度计算 2 项技术。话题识别技术是为了挖掘语料文本中隐含的话题, 相似度计算技术是为了计算话题特征词与文本的相似度, 从而实现依据主题对文档进行归类。本文运用的话题识别技术是 LDA 模型, 该模型通过对网络语料库 D 进行建模, 分别生成大小为 $M \times K$ 的文档-主题分布矩阵 θ 和大小为 $K \times N$ 的主题-词汇分布矩阵 ϕ 。其中, ϕ 的行代表主题, 列代表词汇。对每行词汇生成概率按大小进行排序, 选择 Top- K 词汇作为主题特征词。

从 LDA 建模原理可以看出, LDA 模型本身不能完成话题文本分类, 但主题-词汇分布矩阵 ϕ 中的话题特征词为文本分类提供了聚类特征。在文本语料中, 语义环境是判断词汇相似的重要依据。设主题 $z_i = [w_1, \dots, w_k]$, 文本 $d_i = [w_1, \dots, w_N]$, 对于 z_i 中任意词汇 w_k 和 d_i 中任意词汇 w_n , 如果 w_k 和 w_n 在同一语境中共现频次较大, 满足设定的相似度阈值时, 则 w_k 和 w_n 就具有相关相似性, 即可判定 w_k 和 w_n 属于同一主题。由于语境对词汇意义表达影响较大, 文本运用基于对称差的 KL 距离来进行主题与文本的相似度计算。

KL 距离^[15] (Kullback-Leibler Divergence) 用来度量 2 个概率向量在相同事件空间内分布的差异性。假设 $X = [x_1, x_2, \dots, x_n]$ 和 $Y = [y_1, y_2, \dots, y_n]$ 是 2 个概率分布向量, 则 X 和 Y 差异性计算公式如(7)所示。

$$KL(X|Y) = \sum_{i=1}^n x_i \log \frac{x_i}{y_i} \quad (7)$$

式(7)中, $KL(X|Y)$ 值越小, X 和 Y 的差异性就越小, 说明 X 和 Y 的相似性就越高。但 KL 距离公式存在的缺陷是不具有对称性^[14], 即 $KL(X|Y) \neq KL(Y|X)$, 在实际运用时受到时序因素影响较大。

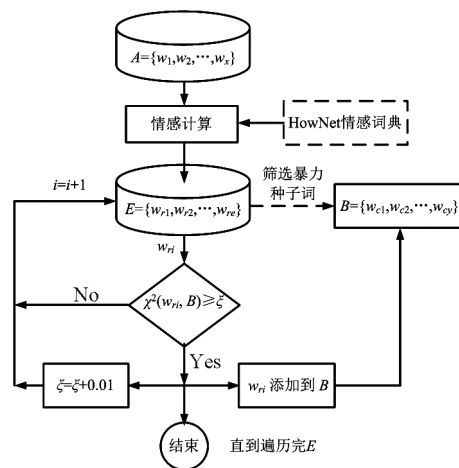
为了弥补这个缺陷, 本文在 KL 距离的基础上引入相对熵概念, 通过变量之间的双向 KL 距离计算消除变量次序对计算结果的影响, 建立了 X 和 Y 对称差异性计算方法, 如式(8)所示。

$$Sim(X, Y) = \frac{1}{2} (KL(X|Y) + KL(Y|X)) = \frac{1}{2} \left(\sum_{i=1}^n x_i \log \frac{x_i}{y_i} + \sum_{i=1}^n y_i \log \frac{y_i}{x_i} \right) \quad (8)$$

对于主题 $z_i = [w_1, \dots, w_k]$ 与文本 $d_i = [w_1, \dots, w_N]$, 通过公式(8)就可得到 z_i 和 d_i 的相似度。运用上述方法, 依次把主题-词汇分布矩阵 ϕ 中的主题 z_i 与语料库中所有文本进行相似度计算, 再把与 z_i 相似度高的文本归类在一起, 得到话题 z_i 的文本集, 最终完成所有主题的文本分类。

2.2 话题内暴力特征分类器建立

话题文本分类完成后, 下一个重要核心任务是在话题文本内进行暴力特征词筛选, 为话题暴力定量计算提供特征词组集。如上文所述, 从情感角度来说, 网络暴力语言带有极强的负面情感, 但负面情感不一定是暴力语言。一般来说, 文本集内暴力特征词集是负面情感词的子集。负面情感特征用情感词典很容易计算得到, 但如何从负面情感词中筛选出暴力词是本文的难点。本文采用 χ^2 检验作为文本特征筛选方法, 以初始的话题暴力词作为种子词对负面词汇集进行反复迭代, 并逐步提升每次迭代的检验阈值, 直到遍历完所有候选词为止, 具体筛选过程如图 3 所示。



注: A 表示话题词汇集; B 表示暴力词汇集; E 表示话题内负面情感词汇集; ξ 表示卡方检验阈值。

图 3 基于 χ^2 检验的网络暴力特征词筛选过程

基于 χ^2 检验的话题暴力特征词筛选过程包括以下几个步骤:

第 1 步: 运用情感词典 HowNet 对话题词汇集 A

中的词汇进行情感计算,得到每个词汇的情感值,并筛选出负情感词汇集 E 。

第2步:在负情感词集 E 中,按情感值大小对情感词排序,用人工方式挑选出负面情感最强烈的10个词作为暴力种子词存放在暴力词汇集 B 中。

第3步:遍历负情感词集 E 中所有词汇,利用卡方检验依次对所有词汇进行特征计算。设定 χ^2 检验阈值 $\xi=0.90$ 。如果 $\chi^2(w_i, B) \geq \xi$,则判定 w_i 为暴力词,并把 w_i 添加到暴力词汇集 B 中,增加检验阈值 ξ ,使得 $\xi=\xi+0.01$,并返回到第2步;否则直接返回到第2步。

第4步:反复执行上述过程,直到选不出暴力特征词为止,最终得到暴力词汇集 B 。

2.3 话题暴力计算与识别

话题暴力计算是运用话题暴力词汇集 B 计算出话题暴力的程度。通常情况下,由于网民存在认知偏差,大多数网络话题内或多或少都存在着暴力成分,但并不是说话题内含有暴力成分就判定该话题一定是暴力话题。只有当暴力成分在话题内占比达到或超过一定比重时才能判定为暴力话题,这个比值用暴力密度来表示。文本暴力与情感具有很强的相关性,所以,暴力计算可以借助情感计算技术来实现。由于修饰词的存在,修饰词对实体情感词的影响较大,在情感计算时需要结合修饰词进行综合计算。

2.3.1 词语的修饰处理

修饰词包括2个属性:一是词性,二是位置。从文本词性语义分析来看,形容词、副词及否定词对实体词的情感表达影响较大。另外,否定词的位置也是影响实体词情感表达的重要因素,比如“很不好”和“不很好”二者表达的情感相差很大。所以,在进行情感计算时一定要考虑实体情感词的上下文语义环境。为了方便计算,文本运用语义词来刻画情感词的上下文关系,具体形式如定义2所示。

定义2:语义词 设一个五元组 $w=\langle w_e, \text{Neg}, \text{Adv}, \text{Pos}, \text{Q} \rangle$ 是个语义词, w_e 表示主题内情感实体词, Neg 表示 w_e 的否定前缀, Adv 表示 w_e 的修饰前缀, Pos 表示否定词位置, Q 表示 w_e 的极性。

语义词 w 的情感计算过程如下:按照文本顺序读取一个词 w_e ,首先判断 w_e 是否为情感实体词,如果是,则根据 w_e 在文本中的位置读取出前缀词,并对前缀词进行词性判断,按公式(9)对 w_e 进行情感计算;否则读取下个词语,并重复上述过程。

$S(w)=\text{Neg}(w_e)\text{Adv}(w_e)\text{S}(w_e)\text{Pos}(w_e)\text{Q}(w_e)$ (9)
式中: $\text{Neg}(w_e)$ 表示 w_e 的否定权重; $\text{Pos}(w_e)$ 表示否

定词的位置权重; $\text{Q}(w_e)$ 表示情感翻转系数; $\text{Adv}(w_e)$ 表示 w_e 修饰权重。其中, $\text{Neg}(w_e)$ 和 $\text{Adv}(w_e)$ 可进一步分解,分解公式分别如式(10)和(11)所示。

$$\text{Neg}(w_e) = (-1)^t, t \geq 0 \quad (10)$$

$$\text{Adv}(w_e) = \sum_{i=1}^V \text{Adv}(w_e^i) \quad (11)$$

式(10)~(11)中: t 表示否定词出现的次数; V 表示修饰词出现的数量。根据文本语义规则,否定词位置影响着修饰词对情感实体词的修饰程度,本文对否定词的位置权重做如下规定:若否定词在修饰词前,则 $\text{Pos}(w_e)$ 取值为0.5;若在修饰词后,则取值为1。

2.3.2 话题暴力计算

通过卡方检验从话题负情感集 E 中筛选出话题暴力特征词,按照式(9)计算每个暴力词的情感值,然后再对所有暴力词进行情感求和,得到话题暴力值。由于网络暴力值是负值,为了直观地对网络话题暴力进行描述,文本对话题暴力值进行翻转,把暴力值映射到正数区间内,具体的话题暴力计算公式如式(12)所示。

$$\text{Cyber}(z_i) =$$

$$-\sum_{n=1}^N \text{Neg}(w_n)\text{Adv}(w_n)\text{S}(w_n)\text{Pos}(w_n)\text{Q}(w_n) \quad (12)$$

最后,运用式(9)对话题 z_i 内其他非暴力词的情感进行计算,再运用式(6)计算话题 z_i 的暴力密度,如果暴力密度超过设定的阈值即可判断 z_i 为暴力话题。

3 实验分析

3.1 数据来源及预处理

本文使用“八爪鱼”网络数据采集器在腾讯新闻中下载了4个话题文本数据,数据包括正文内容和评论,具体数据详情如表2所示。为了模拟真实网络环境以便验证本文方法,把所有下载的话题文章混合放在一起,并随机加入20个无主题文本,组成一个混合主题文本集,构造一个小规模网络环境。然后再把主题文本与评论文本建立对应关系。随后使用Python软件中提供的Jieba分词工具对每条新闻文本及评论进行分词,去除停用词、介词、语气词、转折词等无用词后,分别建立新闻文本语料矩阵 D 和评论文本语料库 C 。运用动态主题数计算方法获取 D 中的主题数 K ,再对LDA模型进行参数设置: α 设置为 $0.5/K$, β 设置为0.1,话题特征词数量 $T=15$,抽样次数为1000。

表 2 实验数据集

序号	话题	文章数	评论数
1	环卫工人解救被卡女童事件	12	286
2	合肥母子三人跳楼事件	18	307
3	玛莎拉蒂女司机醉驾事件	15	326
4	“徽州宴”事件	16	265

3.2 实验结果及分析

运行 LDA 模型前,运用动态主题数识别方法对

主题数 K 进行运算,计算结果为 4,与 D 中事先混合的主题数相同,说明动态主题数计算方法的精度较高。运用 LDA 模型对语料库 D 中的话题进行识别,再用相似度计算公式对语料库 D 中的文本按话题进行分类,得到话题文本集;然后,对每个话题对应的评论文本集进行情感计算,并用卡方检验从负面情感词中筛选出暴力特征;最后,通过暴力密度判定话题的暴力属性。话题特征词、评论暴力词及话题暴力密度计算结果如表 3 所示。

表 3 网络话题特征词及暴力特征词识别结果

话题	话题特征词及概率	评论暴力词(Top 10)	暴力密度
1	女孩 0.003 6、环卫 0.003 2、工人 0.003 2、缝隙 0.002 9、被卡 0.002 9、解救 0.002 9、脱险 0.002 7、脖子 0.002 7、部位 0.002 5、恐慌 0.002 5、触碰 0.002 3、铁门 0.002 2、质疑 0.002 2、高先生 0.002 0、隐私 0.001 9	隐私、咸猪手、非礼、注意、腿根、故意、肮脏、便宜、大叔、不应该	0.013
2	长丰县 0.005 2、三人 0.004 9、坠 0.004 6、妈妈 0.004 4、孩子 0.004 1、死亡 0.003 9、女孩 0.003 7、男婴 0.003 6、丈夫 0.003 5、婆婆 0.003 4、钱 0.003 2、公安局 0.003 2、家庭 0.003 1、矛盾 0.003 1、绝望 0.003 0	冷血、自私、渣男、残忍、恶心、报应、畜生、地狱、恐怖、死	0.132
3	yuwei 0.004 5、玛莎拉蒂 0.004 4、醉驾 0.004 4、鱼尾 0.004 4、交警 0.004 3、驾驶 0.004 1、酒 0.004 0、江西 0.004 0、直播 0.003 8、女司机 0.003 7、孙某 0.003 5、银行 0.003 5、豪车 0.003 2、检查 0.003 2、违法 0.002 8	脑残、自欺欺人、嚣张、彻查、狂、严惩、瑟瑟发抖、害人、豪横、关系	0.075
4	狗 0.006 5、徽州宴 0.006 5、警察 0.006 3、钱 0.005 9、弄死 0.005 6、女孩 0.005 2、老板娘 0.005 2、主播 0.005 0、嚣张 0.004 9、退 0.004 6、消费者 0.004 6、道歉 0.004 5、外甥 0.004 1、拘留 0.003 8、二姨 0.003 5	嚣张、霸道、缺德、暴发户、德行、恶毒、飞扬跋扈、破产、为富不仁	0.217

表 3 给出了对 4 个话题特征词及评论暴力词的识别结果,根据式(4)计算出每个话题的暴力密度。暴力话题密度阈值设置为 0.1。话题 2 和话题 4 的暴力密度分别为 0.132 和 0.217,根据规则,判断出话题 2 和话题 4 是暴力话题。对实际网络文本分析可以发现,在话题 2 评论中,网友对坠楼女子、女子丈夫及家庭均进行了大量的指责,语言尖锐,负面情感很强;话题 4 是全国“臭名昭著”的“徽州宴”事件,不仅大量网民对徽州宴老板娘实施网络暴力,而且全国网红也围堵了徽州宴实体店,消费者纷纷退订,导致徽州宴经营受阻,线上线下都遭受了暴力对待。

话题 1 和话题 3 的暴力密度分别是 0.013 和 0.075,均没有达到设定阈值,判定不是暴力话题。从网络文本分析看,虽然话题 1 中的环卫工人在前期遭受到较强的网络暴力,部分网民指他触碰女孩隐私部位,但随着真相的逐步披露,紧急关头救助

被卡女童的高尚行为还是被绝大多数网民所赞同,使得先期的暴力语言逐渐被正能量语言所稀释,暴力密度逐渐降低,大众对环卫工人的态度符合大众的理性认知。对于话题 3,从表面上看应该会成为暴力话题,但计算结果却“出乎意料”,它并不是暴力话题。从网络文本分析中可以推断出它没有成为暴力话题的原因可能有 2 个:一是女子酒驾行为没有对其他人造成伤害,没有激怒网民;二是大多数网民在酒驾女子的“滑稽表演”中完全充当吃瓜看客,评论语言既委婉又讽刺,负情感不是非常强烈,也体现出了网民“吃瓜群众”的天然属性。

3.3 性能评价

文本的实验数据集是事先用 4 个话题文本集与 20 个无主题文本混合而成,为了验证本文方法(LDA- χ^2)的话题暴力识别性能,把实验运算得到的话题内暴力特征词与混合前 4 个原始话题内标注暴力特征词进行对比,再以黄瑞^[14]提出的方法

(NVLD)做对比,2种方法对4个话题内暴力特征词的识别性能如表4所示。从表4可以看出,在4个话题的混合文本集上,文本方法的暴力特征词组识别

性能(F 值)好于NVLD方法,说明文本的话题暴力特征词识别效果达到了良好的水平。

表4 暴力特征识别性能对比

话题数据集	NVLD			LDA- χ^2		
	准确率/%	召回率/%	F 值/%	准确率/%	召回率/%	F 值/%
T_1	79.65	81.12	80.38	81.51	80.94	81.22
T_2	80.16	82.09	81.11	80.86	82.37	81.61
T_3	77.25	79.32	78.27	78.63	80.72	79.66
T_4	79.01	80.27	79.64	80.64	79.51	80.07

4 结语

为了在复杂网络环境中精确挖掘网络暴力话题,本文提出了一种LDA模型和卡方检验网络暴力话题识别方法,该方法首先运用LDA模型识别出网络文本语料库中存在的话题,并运用相似度计算方法对话题文本进行分类;然后运用卡方检验筛选出话题文本中的暴力特征词,并用情感计算技术得出

话题暴力的值;最后运用暴力密度对话题的暴力属性进行判断。经过实验验证,本文方法在网络话题网络暴力特征识别方面达到了较好的性能。

本文方法认为暴力特征词都具有强烈的负面情感,对于其他情感属性的暴力词识别较为困难,如何在多情感分布下提高暴力特征识别精确是下一步需要研究的方向。

参考文献:

- [1] 林泽斐,欧石燕.基于在线百科的大规模人物社会网络抽取与分析[J].中国图书馆学报,2019(6):100-118.
- [2] 周昕,黄微,滕广青,等.网络舆情传播模式解析与重构研究[J].情报理论与实践,2016(12):25-30.
- [3] GUO R, CHEN F J, CHENG X G. Research on dynamic IPO knowledge matching based on evidence reasoning[J]. Journal of the China Society for Scientific and Technical Information, 2017, 36(12):1290-1301.
- [4] 李华君,曾留馨,滕姗姗.网络暴力发展研究:内涵类型、现状特征与治理对策——基于2012—2016年30起典型网络暴力事件分析[J].情报杂志,2017,36(9):139-145.
- [5] 刘绩宏,柯惠新.道德心理的舆论张力:网络谣言向网络暴力的演化模式及其影响因素研究[J].国际新闻界,2018(7):37-61.
- [6] 石经海,黄亚瑞.网络暴力刑法规制的困境分析与出路探究[J].安徽大学学报(哲学社会科学版),2020(4):78-89.
- [7] 田圣斌,刘锦.社会治理视域下网络暴力的识别与规制[J].中南民族大学学报(人文社会科学版),2020(3):168-173.
- [8] LI G Q, LIU S, ZHANG Y N, et al. Research on the opinion evolution in online collective behavior from the perspective of information entropy[J]. Information Science, 2020(1):42-47+86.
- [9] 杨宗亮,张玉茜,李建飞.一种基于地理空间大数据的网络舆情监测软件架构[J].测绘通报,2017(3):96-100.
- [10] 饶元,吴连伟,王一鸣,等.基于语义分析的情感计算技术研究进展[J].软件学报,2018,29(8):2397-2426.
- [11] YU H, ZHAO Y L, CUI K, et al. Community detection algorithm based on cross-entropy method[J]. Chinese Journal of Computers, 2015(8):1574-1581.
- [12] 谷学汇.基于信息融合算法的暴力视频内容识别[J].济南大学学报(自然科学版),2019(3):224-228.
- [13] 范涛,王昊,吴鹏.基于图卷积神经网络和依存句法分析的网民负面情感分析研究[J].数据分析与知识发现,2021(9):97-106.
- [14] 黄瑞.网络暴力语言检测系统的实现[D].武汉:华中科技大学,2016.
- [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.
- [16] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.