

文章编号: 2095-2163(2022)10-0036-07

中图分类号: G206

文献标志码: A

# 突发环境事件中微博影响力的预测研究

姚婷<sup>1</sup>, 赵锦栋<sup>2</sup>, 杨莉<sup>1</sup>

(1 南京邮电大学 地理与生物信息学院, 南京 210023; 2 南京邮电大学 经济学院, 南京 210023)

**摘要:** 构建突发环境事件中微博影响力预测模型, 有助于相关部门及早发现问题, 并采取行动, 保障网络舆情平稳发展。因此, 本文以“xx 突发环境事件”为研究对象, 根据微博转发、点赞和评论次数衡量其影响力, 选取微博用户、内容、时间特征, 建立基于 AdaBoost 算法的微博影响力预测模型, 并对微博特征重要性进行分析, 结果表明该模型的预测精度可达到 93.2%, 能够较好地完成预测任务。此外, 微博的内容、时间特征、用户基本信息、用户活跃度均对微博影响力产生作用, 长文本的、通知类的微博影响力更高, 白天发布的微博影响力要比晚上发布的高, 政府媒体用户和大 V 用户发布的微博影响力更高, 原创度高的用户发布的微博影响力也会更高。

**关键词:** 突发环境事件; 微博影响力; AdaBoost 算法; 预测; 不良言论

## Prediction research on the microblog influence in emergent environmental incidents

YAO Ting<sup>1</sup>, ZHAO Jindong<sup>2</sup>, YANG Li<sup>1</sup>

(1 School of Geographic and Bioinformatics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 2 School of Economics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

**【Abstract】** The construction of microblog influence prediction model in environmental emergencies will help relevant departments find problems as soon as possible and take actions to ensure the stable development of online public opinion. Therefore, this paper takes "xx Emergent Environmental Incident" as the research object, measures its influence according to the numbers of microblog forwarding, praise and comments, selects the microblog user, content and time characteristics, establishes the microblog influence prediction model based on AdaBoost algorithm, and analyzes the importance of microblog characteristics. The results show that the prediction accuracy of the model can reach 93.2%, and better complete the prediction task. In addition, the content, time characteristics, basic user information and user activity of microblog all play a role in the influence of microblog. The influence of microblog with long text and notice is higher. The influence of microblog released during the day is higher than that released at night. The influence of microblog released by government media users and big V users is higher, and the influence of microblog released by users with high originality is higher.

**【Key words】** emergent environmental events; microblog influence; AdaBoost algorithm; prediction; bad speech

## 0 引言

突发环境事件, 是指由于污染物排放或自然灾害、生产安全事故等因素, 导致污染物或放射性物质等有毒、有害物质进入大气、水体、土壤等环境介质, 突然造成或可能造成环境质量下降, 危及公众身体健康和财产安全, 在导致生态环境破坏的同时, 也可能对社会城市化发展产生不利影响, 亟需采取紧急措施予以应对的事件<sup>[1]</sup>。近年来, 国内突发环境事件不时出现, 不仅发生了伤亡事故, 还会形成直接经济损失, 同时也给社会稳定和日常生活带来了威胁。

随着互联网的迅速发展, 微博从最初的社交属性延展到如今的自媒体属性, 改变着人们的信息传

播模式以及传媒习惯, 并逐渐成为了在社会媒体中即时性最强、用户范围最广的信息传播平台。但因微博的使用门槛较低、而且具有无需实名认证、用户量大、连通性高等特点, 则也较易滋生出不良舆论。在突发环境事件发生时, 不良舆论就可能通过微博这一途径出现大范围传播, 甚至发生不可控的事件与结果。

因此, 当突发环境事件发生时, 如何监测微博中的不良舆论的传播态势, 正确引导舆论走向即已成为目前学界面临的热点问题。本文通过提取突发环境事件中的微博特征, 利用机器学习技术建立突发环境事件中微博影响力的预测模型, 研究微博影响力的影响因素, 并分析微博用户各类言论的影响力。

**作者简介:** 姚婷(1999-), 女, 硕士研究生, 主要研究方向: 大数据挖掘; 赵锦栋(1996-), 男, 硕士研究生, 主要研究方向: 大数据统计; 杨莉(1977-), 女, 博士, 教授, 主要研究方向: 大数据分析及建模。

**通讯作者:** 杨莉 Email: yangli@njupt.edu.cn

收稿日期: 2022-05-06

在突发环境事件发生时,有助于相关机构快速做出反应,更加迅速、准确定位高影响力微博与用户,从而抑制微博上不良舆论的传播,营造良好舆论氛围。

## 1 相关研究

### 1.1 突发事件的舆情研究

目前国内外学者对于突发事件中的舆情研究主要分为 3 方面:

(1) 从舆情本体入手,研究舆情生成及演变路径,主要通过采用搜索指数、发帖量等指标对舆情演变阶段进行划分,分析各阶段的情感强度和主题特征,把握舆情走势,从而提出各阶段的应对措施。

(2) 通过剖析突发事件应急管理中网络舆情的传播规律,构建传播规律模型,对突发事件的舆情发展进行预测,并采取相应的措施。

(3) 对舆情发展过程中的主体展开了深入研究,涉及政府部门、意见领袖、网民等,兼顾不同舆情主体的利益,研究突发事件网络舆情主体博弈的过程<sup>[2]</sup>,从而提出舆情演化过程中各主体需要采取的措施,以确保事件发生时网络舆情的良性运转。

### 1.2 微博影响力的相关研究

国内外有关微博影响力的评价和研究主要包括 2 方面:

(1) 对微博用户影响力的研究和预测。学者们提出了多种针对评价和预测微博用户影响力的方法和模型,包括 h 指数<sup>[3]</sup>、BCI 指数<sup>[4]</sup>、TwitterRank 模型<sup>[5]</sup>等,使用的主要衡量因素包括关注数、粉丝数等。

(2) 单条微博的影响力的评价。学者们普遍用转发数来衡量微博影响力,每一条微博影响力的扩大须依靠转发行为,基于这一理论,提出了不同的微博影响力的预测模型,分析方法主要有多层线性模型<sup>[6]</sup>、PageRank 算法<sup>[7]</sup>等,但由于数据采集困难、微博影响力难以量化等原因,该方面的研究较少。

### 1.3 国内外研究综述

综上所述,国内外关于微博影响力和突发事件中的舆情研究都已经取得了长足的进步,建立了较为完整的研究体系,但是目前的研究仍然有着一定的局限性,依然存在有待改进的方面:

(1) 突发事件中的舆情方面的研究视角较为单一。现有的研究多半为针对舆情演化走势的研究和预测,并没有对舆情演化过程中的各类人群特征做深入剖析,重复劳动较多,且对策、建议研究尚不成熟。

(2) 微博影响力的研究大多以微博用户为主

体,对单条微博影响力的研究仅仅考虑粉丝、注册时间等用户层面的简单影响因素,并且由于微博影响力量化的不科学,这样的分析方法和模型不具备很好的泛化性。

本文对当前的研究不足进行改进,并提供了方法和研究视角上的创新。在突发环境事件中,创新地从微博影响力角度入手,从多个层面综合选取微博影响力的影响因素,利用机器学习技术建立微博影响力预测模型,并对各影响因素进行分析。本研究有助于识别突发环境事件舆情演化过程中的高风险人群和高风险时期,提出舆论引导的建议和实践上的参考,给予管理者决策支持。

## 2 微博影响力预测模型设计

在微博影响力预测模型的选取中,本文经过对多个模型的性能比较,选取基于 AdaBoost 算法的预测模型作为最优模型。

AdaBoost 是一种迭代学习算法<sup>[8]</sup>,其核心思想是将弱的基础分类器迭代成一个强的最终分类器,本文使用决策树作为基学习器。虽然在最初的迭代中,基础分类器的分类效果并不理想,即误判率较高,但经过几轮迭代,不断增加误分类样本的权值,并在更新了权值样本的基础上不断构建新的分类器,从而使得分类精度得到有效提升。

AdaBoost 算法没有关于数据的假设分布,实用性很强,而且对过拟合不那么敏感。

以本文研究的二分类问题为例,AdaBoost 算法步骤如下:

假设  $n$  个初始测试数据为  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ,  $y_i \in \{-1, +1\}$ ,  $-1$  为负样本(低影响力样本),  $+1$  为正样本(高影响力样本);每个样本  $x_i$  包括  $k$  个特征,累计迭代  $M$  次。

(1) 初始化训练集数据的权重分布,使每个训练样本具有均匀的权值分布,设  $w_{i1}$  为第  $i$  个样本的初始化权重。对应的数学表述可写为如下形式:

$$D_1 = (w_{11}, \dots, w_{i1}, \dots, w_{n1}), w_{i1} = \frac{1}{N}, i = 1, 2, \dots, N \quad (1)$$

(2) 循环训练  $M$  次,  $m = 1, 2, \dots, M$ 。

① 对权值分布更新为  $D_m$  的样本进行迭代,得到基本分类器  $G_m(x)$ 。

② 计算  $G_m(x)$  的分类误差率,具体公式如下:

$$e_m = \sum_{i=1}^N w_{m,i} P(G_m(x_i) \neq y_i) \quad (2)$$

③ 计算  $G_m(x)$  的系数  $\alpha_m$ , 具体公式如下:

$$\alpha_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} \quad (3)$$

由式(3)可知,随着  $e_m$  的减小,  $\alpha_m$  不断增大,同时分类精度越高,基本分类器在最终分类器中所占的权重越大。

④ 更新训练数据集的权重分布,具体数学公式如下:

$$w_{m+1,i} = \frac{w_{m,i}}{Z_m} e^{-\alpha_m \gamma_i G_m(x_i)} \quad (4)$$

其中,  $Z_m$  是规范化因子,推导出的数学定义式可写为:

$$Z_m = \sum_{i=1}^N w_{m,i} e^{(-\alpha_m \gamma_i G_m(x_i))} \quad (5)$$

从公式中可以看出,基本分类器  $G_m(x_i)$  错分类的样本权值在迭代过程中不断扩大,正确分类的样本权值通过迭代不断缩小,从而在此后的机器学习中更加重视错误分类的样本<sup>[9]</sup>。

(3)实现基本分类器的组合,可由如下公式来表示:

$$f_{(x)} = \sum_{m=1}^M \alpha_m G_m(x) \quad (6)$$

得到最终分类器,即:

$$G_x = \text{sign}(f_{(x)}) \quad (7)$$

该模型用于对微博的影响力进行预测,将提取得到的微博特征数据输入模型,预测微博的影响力等级。

### 3 实证研究

#### 3.1 数据获取及处理

本文以“xx 突发环境事件”为例,通过编写爬虫程序,对与事件相关的热门话题下的微博展开爬取,共得到 46 265 条微博数据,每条微博数据都包括:用户 ID、所属地址、微博内容、发布日期、点赞数、评论数、转载量等属性。

获取数据之后,对数据中重复、无效以及无关数据进行删除,经处理后的微博数据剩余 7 540 条。

#### 3.2 指标体系构建及微博特征提取

##### 3.2.1 指标体系构建

本文参考濮小燕<sup>[10]</sup>提出的微博特征提取思路,结合突发环境事件的特点和实际情况,最终选择从微博内容、时间、用户基本信息和用户活跃度四个层面上对微博特征进行提取,选取的微博特征及特征值划分情况见表 1。本文提取了 18 个微博特征变量。

表 1 微博特征提取表

Tab. 1 Microblog features extraction table

微博特征		微博特征值		影响力倾向		
内容特征	文本	长度	长	0.103	正	
			短	-0.109	负	
		情感倾向	积极	0.017	正	
			消极	-0.003	负	
			中性	-0.014	负	
		是否有表情	有	0.014	正	
			无	-0.005	负	
		是否有@	有	0.037	正	
			无	-0.007	负	
		是否有链接	有	0.094	正	
			无	-0.074	负	
		主题	通知类	0.182	正	
			祈福类	0.035	正	
			批判类	0.011	正	
			其他	-0.108	负	
		关键字	地点	有	0.046	正
				无	-0.107	负
			人物	有	0.111	正
			无	-0.045	负	
		事件	有	0.035	正	
			无	-0.120	负	
	原创性	是否原创	是	0.069	正	
			否	-0.103	负	
时间特征	所处舆情演化阶段	爆发阶段	0.063	正		
		暂衰阶段	-0.041	负		
		反复阶段	0.015	正		
		消退阶段	-0.060	负		
		发表时间	0~6 时	-0.105	负	
			6~12 时	0.058	正	
			12~18 时	0.040	正	
	18~24 时	-0.038	负			
用户基本信息	认证类型	政府媒体	0.558	正		
		个人大 V	0.319	正		
		普通用户	-0.114	负		
	所在地区	xx 省	0.040	正		
		其他地区	-0.010	负		
	所在行业	政府	0.315	正		
		媒体	0.478	正		
		化工	0.216	正		
		公益组织	0.148	正		
		其他	-0.176	负		
	粉丝数	多	0.143	正		
		少	-0.141	负		
用户活跃度	发布本条微博前一周	多	-0.106	负		
		少	0.108	正		
	发布本条微博前一周发布原创微博数	多	0.158	正		
		少	-0.132	负		

##### 3.2.2 微博情感特征提取

在微博特征的提取中,大部分微博特征可以直接进行提取,所处舆情演化的阶段这一指标的特征值由

舆情生命周期的划分决定。微博的情感倾向和主题则分别使用了朴素贝叶斯分类器和 LDA 主题模型的方法进行提取。本文选择使用朴素贝叶斯分类器对微博情感倾向特征进行提取。这里需用到的数学公式为:

$$P(C/F_1) = \frac{P(C F_1)}{P(F_1)} = \frac{P(C) * P(F_1/C)}{P(F_1)} \quad (8)$$

式(8)表示对于某个样本, 特征  $F_1$  出现时, 该样本被分为  $C$  类的条件概率。

本文从有关“xx 突发环境事件”的微博数据中随机选取三分之一进行人工标注, 分为“积极”、“消极”和“中性”三类, 这部分数据用于朴素贝叶斯分类器的训练和模型预测精度的计算, 预测精度由十折交叉验证的方法获得。通过计算, 分类器的预测精度可以达到 0.945, 符合本文研究的要求。通过训练得出的模型可以对其余数据的情感倾向进行预测, 提取微博的情感倾向特征。在 7 540 条微博中, 情感倾向为积极的博文数有 2 016 条, 情感倾向为

消极的博文数有 3 311 条, 情感倾向呈中性的博文数有 2 123 条, 这说明在突发环境事件中公众的负向情感强度较正向情感强度更高。

### 3.2.3 微博主题特征提取

在主题特征的提取中, 本文选择有关“xx 突发环境事件”的微博数据, 利用 LDA 主题模型提取主题。

LDA 主题模型认为一篇文档包含多个主题, 文档中的每个单词都属于其中的一个主题。一篇文档的形成, 首先在文档中抽取一个主题, 然后再在这个主题下抽取一个单词, 最后, 重复进行前两步, 直到文档中的全部单词都被遍历。

LDA 的核心公式如下:

$$P\left(\frac{w}{d}\right) = P\left(\frac{w}{t}\right) * P\left(\frac{t}{d}\right) \quad (9)$$

其中,  $d$  为目标语料库中的一篇文档;  $t$  是该篇文章中的一个主题;  $w$  是该主题中的一个单词。

LDA 主题模型提取主题结果见表 2。

表 2 LDA 主题模型提取结果表

Tab. 2 LDA subject model extraction results

主题分类	主题排序	主题特征词及其关联度
通知类	Topic1	xx 省(0.052)、爆炸(0.052)、xx 市(0.043)、事故(0.035)、日(0.022)
	Topic2	爆炸(0.033)、事故(0.026)、xx 市(0.025)、xx 省(0.018)、救援(0.012)
	Topic3	爆炸案(0.029)、化工厂(0.020)、企业(0.018)、大火(0.018)、隐患(0.009)
祈福类	Topic4	哀悼(0.024)、生命(0.021)、真的(0.017)、一点(0.012)、人人有责(0.012)
	Topic5	🙏(0.175)、平安(0.045)、愿(0.037)、安息(0.037)、逝者(0.029)
批判类	Topic6	管理(0.022)、做(0.012)、集体(0.011)、整治(0.010)、邓煌(0.010)
	Topic7	影响(0.021)、原因(0.019)、生产(0.019)、居民(0.017)、企业(0.017)
	Topic8	? (0.025)、监管(0.022)、中国(0.015)、xx 省(0.014)、社会(0.011)
其他	Topic9	遇难(0.045)、可怕(0.016)、太(0.016)、头(0.015)、钱(0.014)
	Topic10	消防员(0.016)、救出(0.016)、妻子(0.015)、废墟(0.014)、21(0.014)

由表 2 可知, 共提取出 10 个话题, 每个话题包含 5 个特征词, 每个特征词都给出了其关联度。根据主题提取结果, 结合事件的具体情况, 大致可将微博数据按话题特征分为 4 类, 分别是通知类、祈福类、批判类和其他。

### 3.2.4 微博影响力量化

本文选取微博的点赞数、评论数和转发数三个指标对微博影响力进行综合衡量, 指标权重由熵权法确定。

若选取  $m$  个评价指标,  $n$  个评价对象, 构成的数据矩阵  $R = (r_{ij})_{m * n}$ , 则:

$$e_i = -k \sum_{j=1}^n p_{ij} \ln(p_{ij})$$

$$i = 1, 2, 3, \dots, m; j = 1, 2, 3, \dots, n \quad (10)$$

其中,  $p_{ij} = r_{ij} / \sum_{j=1}^n r_{ij}; k = 1/\ln(n); p_{ij}$  表示第  $j$  个评价对象在总数为  $i$  个的指标中的占比;  $e_i$  为第  $i$  个指标的熵。

则第  $i$  个指标的熵权  $w_i$  定义为:

$$w_i = \frac{1 - e_i}{m - \sum_{i=1}^m e_i} \quad (11)$$

其中,  $0 \leq w_i \leq 1, w_i$  为第  $i$  个指标的熵权。

计算得到的指标权重及微博影响力计算方法见下式:

$$I = 0.204 * A + 0.387 * C + 0.409 * F \quad (12)$$

其中,  $I$ 、 $A$ 、 $C$ 、 $F$  分别表示该条微博的影响力、点赞数、评论数和转发数。

### 3.3 微博影响力预测模型评价及微博特征分析

#### 3.3.1 微博影响力预测模型评价

本文选取处理后的 7 540 条微博作为实验数据, 每条数据包含 18 个微博特征变量。利用式 (12) 计算每条微博的微博影响力, 根据研究事件的特点, 结合数据分布的真实状态, 对微博影响力进行二分类处理, 设定分类阈值为 3, 即影响力数值大于等于 3 的微博为影响力高的微博, 小于 3 的为影响力低的微博。

基于同样的数据集, 本文分别基于决策树、AdaBoost、Bagging 和随机森林等算法构建预测模型。使用十折交叉验证的方法计算各模型的性能, 选择预测精度、召回率和  $F_1$  度量作为评价指标, 计算结果见表 3, 与其他模型相比, 基于 Adaboost 算法

的预测模型在各项指标上都有所提高, 证明该模型性能最好, 具有较强的泛化能力, 预测效果较好。

表 3 各预测模型性能比较表

Tab. 3 Performance comparison of prediction models

预测模型	Precision	Recall	$F_1$
Decision Tree	0.931	0.976	0.922
AdaBoost	0.932	0.978	0.932
Bagging	0.903	0.969	0.926
Random Forests	0.917	0.961	0.898
SVM	0.919	0.975	0.895

#### 3.3.2 微博特征重要性分析

在本研究中, 使用了 18 个微博特征, 根据 AdaBoost 算法模型基于基尼系数计算变量的重要度的功能, 绘制了变量的相对重要性图, 如图 1 所示, 即条形越长的变量对结果的贡献越大, 条形越短的变量重要程度越低。

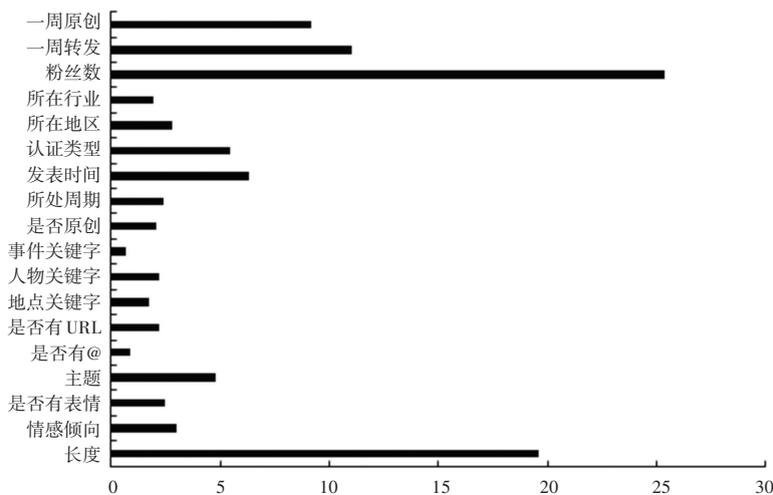


图 1 微博特征值重要性表

Fig. 1 Microblog eigenvalue importance table

结果显示, 模型中的各微博特征重要性相差较大。其中, 重要性较大的 7 个特征从大到小依次是: 用户粉丝数、微博长度、用户一周发微博数、用户一周发原创微博数、发表时间、用户认证类型和微博主题, 而关键字特征、是否有 @ 和链接等特征相对重要性较低。这说明粉丝数、微博长度、用户活跃度等因素对微博影响力大小至关重要。

#### 3.3.3 微博特征值影响力倾向分析

为了进一步探究各微博特征值对微博影响力的影响倾向, 对微博文本长度、粉丝数、发布本条微博前一周转发微博数、发布本条微博前一周发布原创微博数这 4 个指标进行二分类处理, 以中位数作为类别的判定标准。本文参考张继东等人<sup>[11]</sup>的研究思路, 引入支持度的概念, 并对微博特征值的影响力

倾向的计算方法进行了描述, 同时又阐明了各特征值与影响力两者的关联情况。具体计算公式如下:

$$Support(v_i) = \frac{num(v_i = High)}{num(v_i)} \quad (13)$$

$$Support(D) = \frac{num(D = High)}{num(D)} \quad (14)$$

$$Impact(v_i) = Support(v_i) - support(D) \quad (15)$$

式 (13) 中,  $Support(v_i)$  表示特征值  $v_i$  的高影响力支持度, 即在所有特征值为  $v_i$  的微博数据中高影响力微博数据所占比例; 式 (14) 中,  $Support(D)$  表示全部微博数据的高影响力支持度; 式 (15) 中,  $Impact(v_i)$  表示特征值  $v_i$  的影响倾向值, 数值的正负表示该特征值对微博影响力的影响倾向, 数值的绝对值表示该特征值对微博影响力的

影响大小。计算结果见表 1, 基于计算结果, 分别在内容、时间、用户基本信息和用户活跃度层面对微博特征值影响倾向展开研究, 拟做阐释分述如下。

(1) 内容层面。根据计算结果, 分析内容层面的微博特征值的影响倾向, 微博的长度、主题和原创性对微博影响力的影响较大, 这与变量的相对重要性图相符。从微博长度来看, 长文本微博比短文本微博的影响力更大; 从主题特征来看, 一般通知类的微博影响力最大, 祈福类微博和批判类微博的影响力也大于平均水平, 而其他主题的微博影响力较小; 从微博原创性来看, 相比非原创的微博, 原创的微博具有更高的影响力。

(2) 时间层面。时间层面包含微博所处舆情演化阶段和微博发表时间两个特征, 每个特征包含 4 个特征值。从微博所处舆情演化阶段来分析, 处于爆发阶段和反复阶段的微博较暂衰阶段和消退阶段的微博影响力更高, 其中爆发阶段的微博影响力最高。这是因为在舆情爆发阶段, 事件的严重性和突发性导致公众的关注度较高; 而在舆情发展的反复阶段, 因为有伤亡情况和相关责任人刑事处罚等后续信息散布出来, 导致次生舆情出现, 再度引起事件热度提升。从微博发表时间来看, 白天发布的微博比夜晚和凌晨发布的微博影响力更高, 这也符合人们的作息习惯。

(3) 用户基本信息层面。在微博用户基本信息层面, 本文选取了 4 个特征, 这 4 个特征对微博影响力的影响均比较显著。其中, 政府媒体和大 V 用户发表的微博的影响力会更高, 这主要是由于这 2 种用户的粉丝数比普通用户要更多, 根据计算结果, 政府媒体用户和大 V 用户的粉丝数分别是普通用户的 25 392 倍和 1 502 倍; 事件发生地的用户所发表的微博影响力会更高, 说明事件发生地的用户相比其他地区用户容易受到更多关注<sup>[12]</sup>; 相比普通行业, 政府、媒体、公益组织的用户展现出更高的影响力倾向, 同时化工行业的用户也呈现出较高的影响力倾向, 这说明突发环境事件的相关行业工作者发表的言论更容易受到公众的关注, 引发人们的共鸣。

(4) 用户活跃度层面。用户使用微博的活跃度可以用博文数来衡量。实验结果显示, 原创微博发布频率与影响力倾向成正比, 转发微博发布频率与影响力倾向成反比。这是因为对于受众来说, 只能接受有限的消息, 某些博主发布微博较为频繁可能导致粉丝的选择分散, 而有的博主发布微博的频率较低, 正是因为这样, 该博主一旦发博, 其粉丝更可

能认真阅读或转发<sup>[13]</sup>。

### 3.4 各类微博言论的影响力分析

本文将“xx 突发环境事件”中获取的微博数据进行人工分类处理, 对各类微博言论的影响力进行分析。首先将数据分为媒体或大 V 用户发布的微博和普通用户发布的微博两类, 对这 2 类微博的影响力进行计算, 发现第一类微博的影响力平均值高达 56.62, 而第二类微博的影响力平均值仅为 0.52; 接下来对第二类微博进行细分, 分为不良言论、鼓舞性言论和其他言论三类。参考崔珊<sup>[14]</sup>提出的不良文本分类体系, 结合数据实际情况, 在不良言论方面主要选取了谣言类、暴恐类、政治有害类等微博言论, 在鼓舞类言论方面主要选取了加油打气、宣传积极事例等具有鼓舞人心作用的积极言论。结合舆情演化路径, 分析舆情演化各阶段中各类言论的影响力水平, 分析结果如下。

舆情演化路径中普通用户各类言论影响力变化图如图 2 所示。根据图 2, 在事件发展的整个过程中, 不良微博言论的影响力水平最高, 平均影响力数值达到了 5.48; 鼓舞类微博言论的影响力水平其次, 平均影响力数值为 1.84; 其他微博言论的影响力水平最低, 仅为 0.29, 这说明在突发环境事件中, 由于不良言论的蛊惑性和欺骗性, 很多不明真相的公众并不能正确地判断信息的可靠性, 容易被外界环境所影响, 而且还可能进行二次转发; 此外, 网络上的鼓舞类言论同样影响力较大, 具有积极的宣传作用。结合舆情发展阶段来看, 在舆情发展的反复阶段, 普通用户发布的各类微博言论的影响力均达到最高峰, 在舆情发展的消退阶段达到最低值, 这说明舆情发展的反复阶段是防范不良言论的紧要时期, 同时也是积极引导的关键时期。

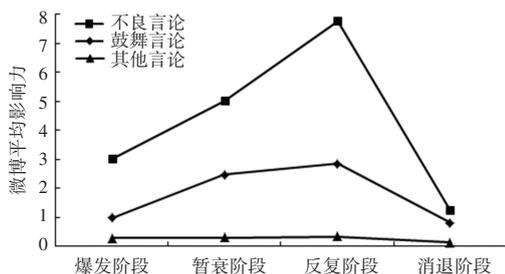


图 2 舆情演化路径中普通用户各类言论影响力变化图

Fig. 2 Changes in the influence of various remarks of ordinary users in the evolution path of public opinion

## 4 结束语

本文构建了微博影响力评价的指标体系,利用AdaBoost算法建立突发环境事件中微博影响力的预测模型,并以“xx突发环境事件”为例进行了实例验证,结合舆情演化的生命周期对微博影响力变化规律进行研究。通过研究,发现影响舆情发展的主要影响因素,帮助相关部门了解舆情发展各阶段的特征,对于突发事件发生后微博网络舆情进行准确预测,及时制止不良言论,并进行正确引导,从而在突发环境事件发生时,确保网络舆情良性运转。

研究表明,模型效果比较理想,具有良好的预测能力,这对于相关部门在环境突发事件初期识别高影响力微博和打造高影响力微博具有重要的现实意义。突发环境事件中的微博的影响力受多层次微博特征的影响,其影响因素重要性排序为:用户基本信息、内容特征、用户活跃度和时间特征。在舆情发展的整个过程中,媒体或大V用户发布的微博影响力始终保持在很高水平;而在普通民众发布的微博中,处于反复阶段微博影响力水平最高,消退阶段最低。从不同言论种类来看,不良言论的影响力最高、鼓舞类言论的影响力其次、其他言论的影响力较低。

基于上述数据分析及研究结果,在突发环境事件发生时,用户的基本信息对微博影响力最大,因此相关部门要重视自身微博平台的运营,积极与公众互动,多与政府媒体和大V用户建立合作,提升自身微博影响力;在发布时间方面,白天发布的微博影响力会更大,政府部门应该在此阶段发布重要内容,来提高微博影响力;在发布内容方面,媒体和相关部门要注意用词的准确度、真实性,发布长文本的微博来提高微博影响力,正确引导舆情发展方向;在发布频率方面,要定期发布高质量原创微博,尽量在用户登录微博的高峰期发帖。在舆情发展的整个过程中,根据各阶段演化特征,在环境突发事件发生初

期,相关部门应加强对高风险微博用户和高风险微博的防范,从源头阻止不良言论传播;舆情发展的爆发阶段和反复阶段是容易导致舆情发展失控的高风险时期,相关部门要采取正面教育、科学引导、及时澄清和准确抑制的措施,保证舆情态势的平稳发展。

## 参考文献

- [1] 国务院办公厅. 国家突发环境事件应急预案[Z]. 北京:中华人民共和国国务院,2014.
- [2] 张玉亮,张昊苏. 突发事件网络舆情主体静态博弈分析及相关政策建议[J]. 情报杂志,2014,33(08):150-154,167.
- [3] 陈远,丛振江. 利用h指数评测微博影响力—以新浪校园微博为例[J]. 情报科学,2015,33(05):85-90.
- [4] 陈冬玲. 基于微博传播指数BCI的图书馆新浪微博平台实证研究—以广东省高职院校图书馆为例[J]. 图书馆研究,2018,48(03):101-106.
- [5] WENG Jianshu, LIM E P, JIANG Jing, et al. Twitterrank: Finding topic-sensitive influential twitterers[C]// Proceedings of the Third International Conference on Web Search and Web Data Mining. New York, USA:ACM, 2010: 261-270.
- [6] 赵红,王璨,胡锋,等. 基于多层结构的微博影响力决定因素研究[J]. 管理学报,2014,11(07):1062-1068.
- [7] 易续平. 微博影响力的量化研究[D]. 昆明:云南财经大学,2014.
- [8] 殷存军. 基于深度卷积网络和AdaBoost算法的无人机图像中病害松树识别[D]. 合肥:安徽大学,2020.
- [9] 陈拥权,陈影,陈学三. 基于Adaboost分类器的车辆检测与跟踪算法[J]. 计算机技术与发展,2017,27(10):165-168,176.
- [10] 濮小燕. 基于多层结构的单条微博影响力研究[D]. 成都:电子科技大学,2015.
- [11] 张继东,杨杨. 基于交互行为和情感倾向的移动社交网络用户影响力度量模型研究[J]. 情报理论与实践,2019,42(01):112-117,93.
- [12] BISHT S, TOSHNIWAL D. EventStory: Event detection using Twitter stream based on locality[M]// ORCHADO E, LOZANO J A, QUINTIÁN H, et al. Intelligent data engineering and automated learning - Ideal 2014. Lecture Notes in Computer Science. Cham: Springer, 2014,8669: 394-403.
- [13] 赵蓉英,曾宪琴. 微博信息传播的影响因素研究分析[J]. 情报理论与实践,2014,37(03):58-63.
- [14] 崔珊. 网络内容安全中不良文本过滤研究[D]. 北京:北京邮电大学,2017.

(上接第35页)

- [9] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [10] 周祥全, 张津. 深层网络中的梯度消失现象[J]. 科技展望, 2017, 27(27):284.
- [11] 陈建廷, 向阳. 深度神经网络训练中梯度不稳定现象研究综述[J]. 软件学报, 2018, 29(7):2071-2091.
- [12] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep

- residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, Nevada: IEEE,2016:770-778.
- [13] MILLETARI F, NASSIR N, SEYED-AHMAD A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation[J]. arXiv preprint arXiv:1606.04797, 2016.
- [14] IBTEHAZ N, RAHMAN M S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural Networks, 2020,121:74-87.