

文章编号: 2095-2163(2020)06-0167-04

中图分类号: TP393

文献标志码: A

# 基于关键词提取的网络舆情监管预测算法研究

阮梦黎

(山东管理学院, 济南 250357)

**摘要:**为解决传统网络舆情监管预测算法对舆情态势预测准确性较低的不足,提出了基于关键词提取的网络舆情监管预测算法,根据关键词提取的网络舆情监管预测算法运算确定关键流程,依托网络舆情关键词的提取以及权值的计算,实现了网络舆情时间序列的构建,基于网络舆情二元语义拟合以及最终舆情预测结论的显示,完成了提出的算法研究。试验数据表明,关键词提取的网络舆情监管预测算法,较传统网络舆情监管预测算法具有更高的有效性。

**关键词:**关键词提取;网络舆情;监管预测;预测算法

## Research on network Public opinion supervision prediction Algorithm based on keyword extraction

RUAN Mengli

(Shandong University of Management, Jinan 250357, China)

**[Abstract]** In order to solve the traditional algorithms based network public opinion supervision of public opinion trend of low prediction accuracy, put forward the network public opinion supervision prediction algorithm based on keyword extraction, based on keyword extraction algorithm of network public opinion supervision prediction operation to determine the critical process, relying on the network public opinion the extraction of keywords and calculation of the weights, implements the network public opinion in order to construct a time series, based on binary semantic matching network public opinion and public opinion finally forecast conclusion shows that completed the algorithm research. The experimental data show that the network public opinion supervision and prediction algorithm based on keyword extraction is more effective than the traditional network public opinion supervision and prediction algorithm.

**[Key words]** Keywords extraction; Online public opinion; Regulatory forecasts; Prediction algorithm

### 0 引言

网络舆情监管预测作为网络监管的重要组成部分,对保障信息安全和社会稳定都有着十分重要的意义。传统网络舆情监管预测采用人工监管预测,如网警或采用大数据分析等衍生出灰度阈值监测法等,从监测预测效果上看,由于大数据运算体量较大,人工完全取决于个人能力等主观因素,现有网络舆情监管预测算法存在网络舆情反应较慢,舆情态势预测准确性较低等不足<sup>[1]</sup>,为此,本文提出了基于关键词提取的网络舆情监管预测算法研究。

关键词提取是指利用信息技术和智能算法从待查询的文本或网页中自动获取标志性核心词汇,从而实现认知原文档大致内容要义的过程,它是通过抽取代表信息或核心词语来进行文本挖掘进而实现舆情监控、预测或决策分析等后续工作的关键环节。本文设计的根据关键词提取的网络舆情监管预测算法,通过确定关键流程以及监测到的网络舆情集合,

计算网络舆情在路由链路层的域间关联特征,实现频繁词汇的筛选,利用傅里叶变换分解舆情语义特征,确定筛选度,使其转化成为关键词,利用信息检索与数据挖掘计算,完成网络舆论权值的计算,确定网络舆情时间与计算次序的对应关系,基于此进行网络舆情二元语义的拟合,确定网络舆情综合评价映射,并进行舆情预测结论的最终显示。为了保证研究的正确性,进行仿真试验,试验数据表明本算法具有较高的预测准确性,适合网络舆情监管预测。

### 1 算法运算关键流程的确定

关键词提取网络舆情预测算法是根据网络舆情二元语义的拟合,分析网络舆情综合评价映射,控制筛选度实现的,其网络舆情监管预测算法基本流程如图1所示。

### 2 网络舆情时间序列的构建

#### 2.1 网络舆情关键词的提取

设监测到的网络舆情集合为  $H\{h_1, h_2, h_3, \dots, h_n\}$ ,

**基金项目:**山东省社会科学规划研究项目(17CQXJ11);山东省重点研发计划(公益类科技攻关)(2019GGX105013)。

**作者简介:**阮梦黎(1981-),女,硕士,副教授,主要研究方向:数据挖掘。

**收稿日期:**2019-05-10

一般监测到的舆情集合均较大,不适合整体的监管以及预测,为此进行关键词的提取,针对关键词进行网络舆情监管和预测。网络舆情关键词的提取过程示意图如图2所示<sup>[2]</sup>。

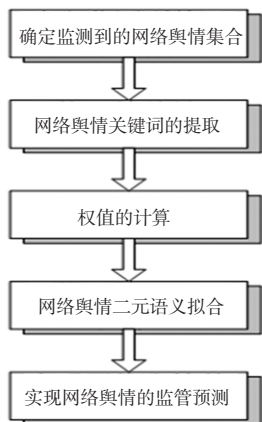


图1 网络舆情监管预测算法基本流程

Fig. 1 Basic flow of the algorithm network public opinion supervision and prediction

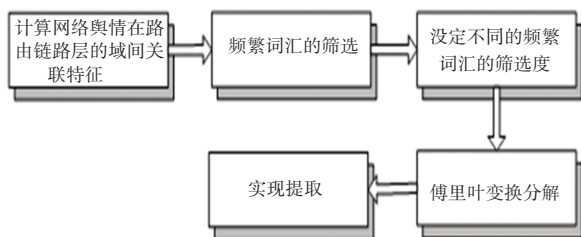


图2 关键词提取过程示意图

Fig. 2 Process of keyword extraction

首先,确定网络舆情在路由链路层的域间关联特征,可用公式(1)表示<sup>[3]</sup>:

$$\begin{cases} I_{ka0} = \frac{1}{3}(I_a + I_b + I_c), \\ I_{ka1} = \frac{1}{3}(I_a + aI_b + a^2I_c) \frac{1}{\sqrt{3}}I_b, \\ I_{ka2} = \frac{1}{3}(I_a + a^2I_b + aI_c) - \frac{1}{\sqrt{3}}I_b. \end{cases} \quad (1)$$

式中,  $I_{ka0}$ 、 $I_{ka1}$ 、 $I_{ka2}$  表示舆情序列波特率;  $I_a$ 、 $I_b$ 、 $I_c$  表示监测频率响应;  $a$  表示舆情监测范围加权数。根据网络舆情在路由链路层的域间关联特征,进行频繁词汇的筛选,可根据设定的不同监测级别,设定不同的频繁词汇的筛选度,实现对关键词的筛选,其逆函数表达式如公式(2)所示<sup>[4]</sup>:

$$f(I) = \frac{(I_{ka2} + I_{ka1} + I_{ka0})}{ax} dI. \quad (2)$$

式中,  $x$  代表频繁词汇的筛选度,根据筛选级别设定,筛选级别越高,筛出关键词越多,反之筛选级别

越低,筛出关键词越少,一般筛选度取值范围为  $[0.4 \sim 0.8]$ <sup>[5]</sup>。

基于频繁词汇筛选出的词语,为关键词的原型词汇,对原型词汇进行网络舆情的综合相对贴度验算,将验算结果进行傅里叶变换分解,分解出舆情语义特征,使其转化成为关键词。其傅里叶变换分解过程可用公式(3)表示<sup>[6]</sup>:

$$S_{ka} = \xi U_{ka} \hat{I}_{ka} f(I). \quad (3)$$

式中,  $U_{ka}$  表示语义环境;  $I_{ka}$  代表分解系数;  $\xi$  语境筛选级别。

基于监测到的网络舆情集合的确定,计算网络舆情在路由链路层的域间关联特征,以及频繁词汇的筛选度,利用傅里叶变换分解计算,实现了网络舆情关键词的提取。

## 2.2 权值的计算

基于网络舆情关键词的提取,得到需要监管预测的关键词,但此时计算出的关键词为平行关键词,计算次序是按照采集次序实现的,为此进行网络舆情时间序列的构建。

时间序列是监管预测的基础序列,利用时间序列定义计算次序,得出非平行关键词组即热度排名,进行信息检索与数据挖掘计算,其信息检索与数据挖掘计算公式如(4)所示<sup>[7]</sup>:

$$Q = \frac{IDF}{d + \frac{cI_{ka}}{S_{ka} + I_{ka}}}. \quad (4)$$

式中,  $IDF$  表示逆文本频率指数 (Inverse Document Frequency),  $IDF$  越大,说明词的区分度越大,切合主题的特点<sup>[8]</sup>;  $d$  表示词频,即  $d$  在集合  $H$  中出现的频率,  $c$  表示语料库中的文件总数,即集合文件和,利用信息检索与数据挖掘计算实现权值的计算。

基于网络舆情关键词的提取,依托不同候选关键词的权重计算,构建一个完成的网络舆情时间序列。

## 3 网络舆情二元语义拟合

基于网络舆情时间序列的构建,完成网络舆情关键词的提取,实现了对网络关键词的加权计算,确定了网络舆情时间与计算次序的对应关系,进行网络舆情二元语义的拟合,以实现基于获取的关键词和网络环境,进行网络舆情预测。

二元语义拟合是由文字语言到计算机语言的拟合,由于文字语言无法直接进行计算机计算,为此二元语义拟合十分必要,通过决策计算,实现对网络舆

情相关的综合评价<sup>[9]</sup>。将获取的关键词转换成计算机语言,首先进行网络舆情序列语义本体模型表达,是计算机获取文字语言转换成计算机语言的常用方法,本文采用借调的方式,利用网络舆情序列语义本体模型表达,将获取的文字语言转换成计算机语言<sup>[10]</sup>。

将转换好的计算机语言进行决策计算,其目的是为确定网络舆情综合评价映射做数据支持,网络舆情综合评价映射是舆情预测监管的标准,即评价已发生的网络舆情处于何种态势,评判是否触发网络舆情警戒机制,并对网络舆情进行跟踪,直至网络舆情事件结束,其网络舆情综合评价映射可用公式(5)表示为:

$$\Delta S = D_{ka} \left( \frac{Z_{2\Sigma} Z_{0\Sigma}}{Z_{2\Sigma} + Z_{0\Sigma}} \right). \quad (5)$$

式中,  $Z_{2\Sigma}$  表示协方差修正参数;  $Z_{0\Sigma}$  表示语义主题相关度;  $D_{ka}$  表示舆情影响因子。

当  $\Delta S > 1$  时,表示网络舆情态势发展处于超警戒运行状态,监管部门应着力注重,网络舆情监管算法采用跟踪机制,对该舆情进行实时分析研判以提供准确数据;

当  $\Delta S = 1$  时,表示网络舆情态势发展处于临界运行状态,监管部门应着力分析该舆情的发展态势,确定未来监管方向;

当  $\Delta S < 1$  时,表示网络舆情态势发展处于基本可控状态,监管部门根据实时情况进行分析,网络舆情监管算法采用非重点跟踪机制,用于监控舆情的进一步发展。

同时网络舆情的监管预测受突发事件影响,二元语义拟合应受事件关注度的不断增加而加深。

#### 4 网络舆情监管预测的实现

基于网络舆情二元语义拟合,完成协方差修正,确定了舆情影响因子,依托网络舆情的监管预测运行载体,实现网络舆情的监管预测。网络舆情监管预测的运行载体,同时也是关键词提取网络舆情监管预测算法的运行载体。算法运行载体包括算法运算平台和显示平台两部分,其中算法运算平台各参数设置如表1所示,算法运行显示平台用来显示最终监管预测结果,并根据设计警戒参数,实现超临界提醒。

#### 5 实例分析

为保证本文提出的基于关键词提取的网络舆情监管预测算法的有效性,进行实例分析。分析过程中,采用传统人工舆情监测法、灰度阈值监测法作为

实验对比对象,进行算法的态势预测运算验证。

表1 算法运算平台参数

Tab. 1 Platform of algorithm operation

项目	参数	备注
运行平台	Windows	兼容不同版本
编写语言	C++	程序命令
工具	.NET Framework	Windows 系统下的环境模型工具

#### 5.1 实验准备

实验中利用已发生的网络舆情事件作为实验对象进行仿真实验,分别对该网络舆情发生开始后,态势演变至 12.5%、25.0%、37.5%、50.0%、62.5%、75.0% 和 87.5% 这 7 个阶段进行舆情发展态势的预测。

仅对一件网络舆情案件进行发展事态分析,存在实验偶然性和不确定性,为此选择由舆情机构统计的 2018 年全国舆情热点事件进行分析,主要包括“中美贸易争端”、“改革开放 40 周年”、“金庸逝世”、“问题疫苗事件”、“个税改革”、“范冰冰偷逃税事件”、“中概股上市热潮”、“幼儿园虐童事件连续曝光”、“共享单车倒闭潮”、“殴打公交司机等乘客霸凌行为”、“高铁霸座”等网络舆情事件来验证算法对舆情态势预测准确性。

#### 5.2 实验过程

由于本次实验采用的是根据不同网络舆情方法对已完成的网络舆情事件进行分析,用过去参数验证方法的准确性,为此需构建过去实验环境,让关键词提取监管预测算法、传统人工舆情监测法、灰度阈值监测法都将实际事件的网络舆情结果作为未知数据,通过对过去事件的分析得出结论与事实真实结果进行对比,分析其对网络舆情发展态势预测的准确性。

实验过程中,建立的实验环境应基本符合事实发展,采用时间函数控制事态的发展,例如:将“问题疫苗事件”的时间控制在该事件发生开始时,即记为时间点 A,切断所有大于时间点 A 的所有信息,载入需要进行实验对比的 3 种方法,控制时间函数,时间向后移动至 12.5% (事件结束的 12.5%),利用 3 种网络舆情监管预测方法对 A+12.5% 时间内所接收到的所有信息,进行网络舆情事件态势的预测,3 种方法预测完成后,记录预测值,与该事件真实结果对比,并进行记录。再通过控制时间函数,时间向后移动至 25%,利用三种网络舆情监管预测方法进行该事件态势的预测,与真实结果对比并进行记录。



如此往复控制时间函数直到事件结束,将记录的数值形成实验结果图表。

### 5.3 实验结果分析

根据实验过程得出关键词提取监管预测算法、传统人工舆情监测法、灰度阈值监测法,在不同时间段的态势预测情况,根据记录的数据以及真实结果数据形成实验结果数据表,如表2所示。

表2 实验结果对比表

Tab. 2 Comparison of experimental results %

舆情发展阶段	舆情态势预测准确性		
	关键词提取监管 预测算法	传统人工舆 情监测法	灰度阈值 监测法
12.5	60.5	80.4	50.4
25.0	66.7	60.5	48.6
37.5	69.8	54.1	45.2
50.0	70.0	48.7	48.5
62.5	75.4	62.2	54.5
75.0	76.4	68.2	54.8
87.5	87.5	79.8	68.7

根据实验结果可以得出,人工舆情监测法在舆情发展初期阶段具有较高的灵敏度,但从整体上看,传统人工舆情监测法受网络舆情发展阶段影响较大,随着接收到的信息增多,由于主观因素的存在,易造成判断性失误;灰度阈值监测法具有较高的稳定性,但整体舆情态势预测略低于关键词监测管理算法。

通过对实验数据的统计、计算得出,关键词提取监管预测算法整体态势预测有效性为61.56%,传统人工舆情监测法整体态势预测有效性为55.96%,灰度阈值监测法整体态势预测有效性为45.17%。可见本文提出的关键词提取的网络舆情监管预测算法,较其它两种预测算法具有更高的有效性。

(上接第166页)

## 6 结束语

最近引入的用于光线跟踪加速的专用硬件和图形api中添加的光线跟踪进行创新,并尝试一种混合呈现的新方法,将栅格化和光线跟踪相结合。虽然没有进行在商业级游戏引擎虚幻引擎4中集成光线追踪的工程实践。不过这种创新的重建过滤器,用于渲染随机效果,如光滑的反射、柔和的阴影、环境遮挡和漫射间接照明,每个像素只有一条路径,使这些昂贵的效果在实时中应用成为可能。而且已经成功地使用混合渲染创建了两个电影质量的演示虽

## 6 结束语

本文提出了基于关键词提取的网络舆情监管预测算法,通过关键流程的确定、舆情时间序列的构建、舆情二元语义拟合和最终舆情预测结论显示等过程,并通过仿真实验及其与其它传统算法的比较,验证了算法的有效性。舆情分析和监管是一项涉及信息学、管理学、社会学以及传播学等多学科跨领域的工作,既需要充分利用智能技术协助决策支持和处理研判,又要重视对其潜在的社会和个体影响因素进行合理分析和有效引导,希望后期通过进一步的研究和实验优化,使其在文本信息挖掘、自动摘要生成、舆情精准研判等方面更具广泛的应用价值。

## 参考文献

- [1] 丁晨春,王鹏鹏,龚思兰.基于社区发现和关键词共现的网络舆情潜在主题发现研究——以新浪微博魏则西事件为例[J].情报科学,2018,36;323(7):80-86.
- [2] 王永志,金樑,朱月琴,等.基于大数据技术的地学文档关键词提取算法研发[J].地球物理学进展,2018,33;149(3):372-379.
- [3] 龙震岳,艾解清,邹洪,等.基于改进灰狼优化算法的网络流量预测模型[J].计算机应用研究,2018,35;320(6):251-254.
- [4] 夏一雪,袁野,张文才,等.面向大数据的网络舆情异常数据监测与应用研究[J].现代情报,2018,38;324(06):82-87.
- [5] 马慧芳,刘芳,夏琴,等.基于加权超图随机游走的文献关键词提取算法[J].电子学报,2018,46;424(6):133-137.
- [6] 张素智,赵亚楠,杨茵.基于MPB-Tree索引的空间数据多关键词模糊查询算法研究[J].华中师范大学学报(自然科学版),2017(6):49-55.
- [7] 曾子明,黄城莺.基于BP神经网络的突发传染病舆情热度趋势预测模型研究[J].现代情报,2018,38;323(5):39-46,54.
- [8] 张茂元,孙树园,王奕博,等.基于EKSC算法的网络事件热度预测方法[J].计算机工程与科学,2018,40(2):238-245.
- [9] 闫小勇,李青.基于最佳路径搜索的二进制协议格式关键词边界确定方法[J].计算机应用,2018,38;334(6):206-211.
- [10] 李磊,刘继,张弘魁.基于共现分析的网络舆情话题发现及态势演化研究[J].情报科学,2016,34(1):44-47.

然演示的片段过于短促。

## 参考文献

- [1] 段辉丽,唐丽玉.利用OSG三维图形渲染引擎实现的骨骼动画[J].计算机工程与应用,2015(3).
- [2] 胡起云.UE4游戏引擎的建筑可视化表现探析[J].装饰,2016(8).
- [3] Edward Liu. Cinematic Rendering in UE4 with Real-Time Ray Tracing and Denoising[J]. Ray Tracing Gems 2019.卷期页码
- [4] 盖伟.虚拟现实实时交互方法研究[D].济南:山东大学,2017.
- [5] 杨焕宇.面向虚拟现实的三维点云数据处理关键技术研究[D].上海:东华大学,2016.