

基于百度指数的突发事件网络舆情预测分析

徐泽华, 纪欣媛, 王智宇, 赵怡恒, 宋宇榜

上海工程技术大学数理与统计学院, 上海

收稿日期: 2024年6月1日; 录用日期: 2024年6月21日; 发布日期: 2024年6月30日

摘要

如何在公共卫生事件发生时, 正确引导和管理网络舆情, 形成积极健康的舆情氛围, 一直是相关部门需要关注的问题。本文以上海新冠疫情为对象, 爬取网络舆情相关的百度搜索指数样本数据(2022.11.10~2023.01.31)。首先, 依据舆情生命周期理论, 将该阶段上海新冠疫情网络舆情发展划分为萌芽期-成长期-成熟期-衰退期, 对不同时期网络舆情的传播特点进行了定性分析, 并给出各阶段舆情管理建议。然后, 考虑到样本数据的非线性和非平稳的特点, 分别利用最小二乘支持向量机(LSSVM)和自回归移动平均模型(ARIMA)进行预测。最后, 将LSSVM和ARIMA两个模型结果赋以适当权重进行组合预测, 相较于之前单一模型预测, 预测精度明显提升。本文对突发公共事件网络舆情进行预测分析, 了解舆情发展趋势, 以期为相关部门引导和管控网络舆情提供一定的参考依据。

关键词

新冠疫情, 网络舆情, 支持向量机, ARIMA, 预测模型

Prediction and Analysis of Network Public Opinion in Emergencies Based on Baidu Index

Zehua Xu, Xinyuan Ji, Zhiyu Wang, Yiheng Zhao, Yubang Song

School of Mathematics, Physics and Statistics, Shanghai University of Engineering and Technology, Shanghai

Received: Jun. 1st, 2024; accepted: Jun. 21st, 2024; published: Jun. 30th, 2024

Abstract

How to correctly guide and manage online public opinion in public health emergencies to foster a positive and healthy atmosphere has always been a concern for relevant departments. This article

文章引用: 徐泽华, 纪欣媛, 王智宇, 赵怡恒, 宋宇榜. 基于百度指数的突发事件网络舆情预测分析[J]. 统计学与应用, 2024, 13(3): 923-933. DOI: 10.12677/sa.2024.133094

focuses on the COVID-19 epidemic in Shanghai and crawls sample data from the Baidu search index related to online public opinion, spanning from November 10, 2022 to January 31, 2023. Firstly, based on the life cycle theory of public opinion, the article classifies the development of online public opinion into four stages: Germination, growth, maturity, and decline. A qualitative analysis of the propagation characteristics of online public opinion in these different periods is conducted, followed by suggestions for public opinion management in each stage. Then, considering the nonlinear and non-stationary characteristics of the sample data, the least squares support vector machine (LSSVM) and autoregressive moving average model (ARIMA) are utilized for prediction. Finally, the results of the LSSVM and ARIMA models are combined with appropriate weights to enhance prediction accuracy, significantly outperforming previous single model predictions. By predicting and analyzing the development trend of public opinion during sudden public events, this article aims to provide a reference for relevant departments in guiding and controlling online public opinion.

Keywords

COVID-19, Online Public Opinion, Support Vector Machine, ARIMA, Prediction Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

新冠肺炎疫情突发公共卫生事件是由新型冠状病毒肺炎引发的重大国际公共卫生事件，是近 100 年来人类遭遇的最广泛的全球重大流行，其特点是传播速度快，传播范围广、防控难、危害强，严重危害世界人民的生命健康权。这一事件的爆发和蔓延是近期人类历史上从未发生和预测到的危害事件，大大影响力人们的生活和社会的平稳局面，其破坏性和危害性程度极为广泛。

随着突发公共卫生事件的曝光率越来越高，往往会受到大众媒体的广泛关注，由于其传播过程的复杂性和广泛性，由此产生的负面情绪往往容易造成负面舆论。因此，如何在发公共卫生事件发生时，提出适当的应对策略，正确引导和管理网络舆情，稳定大众情绪，最小化控制损失，并且形成积极健康的舆情氛围，成为当前需要关注的问题。本文将利用百度指数网站上的突发公共卫生事件中的新冠疫情搜索数值进行数据处理与统计预测，了解舆情发展趋势，以期为相关部门引导和管控网络舆情提供一定的决策支持，从而有助于稳定社会发展。

关于突发公共卫生事件的相关论文研究，多集中于公共卫生事件频发的年份。近年，突发公共卫生事件大多选取新冠疫情为研究对象，雷园[1]选取微博中的传播信息作为研究对象，重点分析政务微博以及意见领袖的微博传播效果，运用框架分析的研究方法，田红梅[2]运用传染病模型对微博信息进行仿真模拟。针对突发公共卫生事件的网络传播舆情分析，郭圳凝等[3]从多元主体的共同价值基础、我国舆情治理的现实环境方向分析，通过建立公民需求为首的服务精神、整合多元主体共同参与的治理架构、建立一站式信息发布渠道、营造舆情演化的良性生态、促进舆情治理法制化建设等举措。朱家安[4]提到舆论爆发前网民们的意见多以舆情形式蛰伏一段时间，同时网络舆论在突发公共卫生事件引起的社会治理问题中是否应用得当也是能否有效解决突发公共卫生事件的关键环节。王旭等[5]描述了突发事件网络舆情与社会网络分析，借此来研究魏则西事件的扩散过程。国外有学者对于“突发公共卫生事件的网络传播”议题，大致基于危机传播理论，从传播主体和传播效果进行研究分析，Singh S 等[6]提出混合方法包

括对新冠肺炎造成的死亡数据,应用离散小波分解,将输入数据分成若干组成序列,然后对每个组成序列应用适当的计量经济学模型,以预测未来的死亡病例。对于网络传播的研究,国内学者从信息传播者的角度出发,多从突发事件内容、传播媒介和网络舆情层面展开了研究[7] [8] [9],国外学者则更倾向于使用理论对突发事件主体和网络舆情效果进行研究。

本文选取新冠疫情这一突发公共卫生事件,利用百度指数数据,基于信息生命周期理论划分突发公共卫生事件网络舆情传播的不同阶段。然后,对突发公共卫生事件网络舆情的统计数据整理与描述性统计,并给出舆情不同阶段的管理建议。接着,分别利用最小二乘支持向量机(LSSVM)和自回归移动平均(ARIMA)模型两种单一模型,及其组合模型对新冠疫情网络舆情进行预测分析。

2. 基于百度指数的新冠疫情网络舆情的定性分析

2.1. 新冠疫情网络舆情阶段划分与数据来源

突发事件网络舆情是网络舆情的一种特殊形式,由于突发事件具有事发突然,情况复杂,影响广泛,危害严重等特点。突发事件在此基础之上其主要包括:突发性、迅速性、自由性、内容复杂性、丰富性及内容非理性特点。根据不同研究角度,学者们对其传播阶段的划分有所不同,大多将突发事件网络舆情传播划分为三、四或者五个阶段。在借鉴学者研究基础上,本文以信息生命周期理论为基础并根据生长曲线的特点,设定突发事件网络舆情传播过程包括萌芽期,成长期,成熟期,衰退期四个时期更为合适[5],如图1描绘出舆情传播过程的不同时期。

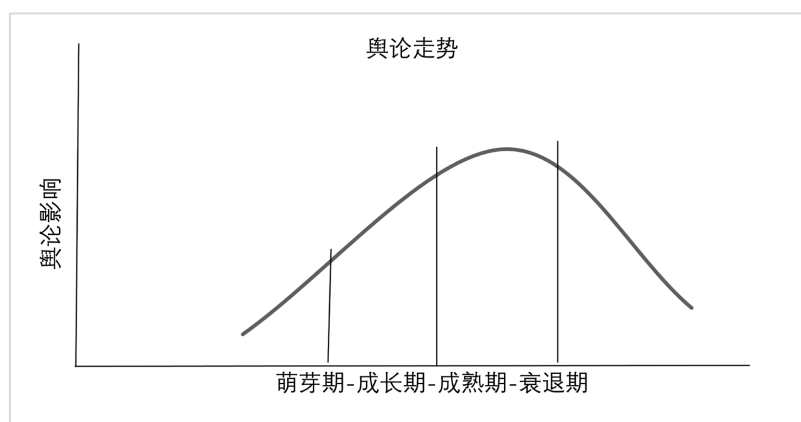


Figure 1. Emergency network public opinion at different stages
图 1. 突发事件网络舆情不同阶段

本论文选取百度指数网站上,地区为上海市关于新冠疫情的相关搜索指数,时间跨度从2022.11.10~2023.01.31。通过百度指数搜索“新冠疫情”在所给定期限内,可以从中可以发现在新十条出台之前,搜索数据虽有增长但是峰值不高。新十条出台期间有关于放开管控居家隔离不再集中隔离的相关文件出台后,疫情局面全面放开,隔离管控区域全面解除,异常情况实行自我监控,居家隔离。因为这是完全不同以前的政策决策,再加上这是切实关系每个人的生命健康以及全体利益的决策,所以关键词数据搜索量出现大幅度增长。进入12月中下旬以后疫情蔓延,但是因为之前已经有了一轮爆发期,相关专家和感染者为大众提供了自己的经验体会和建议,人们的恐慌情以及焦虑情绪得到明显缓解。相关搜索数据趋于稳定。1月上旬到一月底,正逢春节,虽然春运局面规模仍然很大,但是一轮感染基本上已经结束,疫情局面基本稳定,搜索数据明显下降,基本可以算作信息衰退期。

新冠三年以来，国家和政府以及相关部门为此做出了努力和牺牲，新十条的出台使得疫情局面逐渐放开，而且网络舆情的发展趋势往往会受到相关文件或者草案的出台。新十条的出台让疫情防控不再层层加码，使得局面逐渐放开。所以本文将在划分新十条政策出台前的半个月为网络舆情潜伏期，新十条出台后的前后十天为舆情爆发期，后面依次为网络舆情蔓延期和消亡期。根据信息生命周期理论：信息的生命周期包括萌芽期，成长期，成熟期，衰退期四个阶段。本论文的研究对象为 2022.11.10~2023.01.20 期间的百度指数网站中关于新冠疫情的网络舆情发展趋势，划分研究对象的四个时期的时间节点为 2022.11.10~2022.11.31 - 萌芽期，2022.12.01~2022.12.20 - 成长期，2022.12.21~2023.01.10 - 成熟期，2023.01.20~2023.01.31 - 衰退期。借助此搜索指数可以反映出网络舆情的发展状况。

2.2. 新冠疫情网络舆情传播阶段特点分析

为了研究新冠疫情网络舆情传播过程并保持该事件网络舆情扩散的完整性，在“新冠疫情”选取时期进行了数据实时跟踪观察收集。通过百度指数这一可视化平台，输入关键词：“新冠疫情”，定位地区选定为“上海”。检索到“新冠疫情”网络搜索指数变化图 and 关注趋势变化图，如图 2 所示，可以看出网络搜索指数的变化。

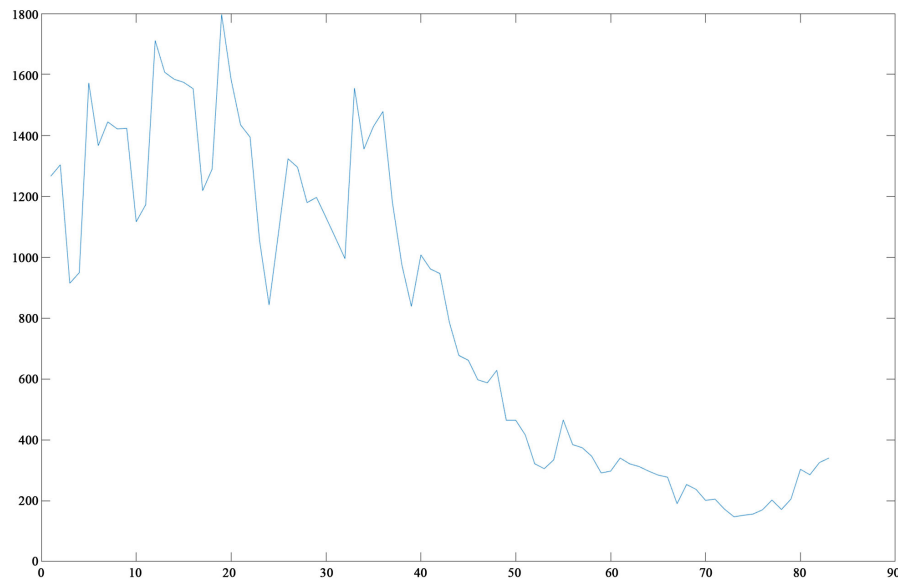


Figure 2. Trend chart of search index of “COVID-19” in Baidu index
图 2. 百度指数中“新冠疫情”搜索指数变化趋势图

从图 2 中可以看出，有关新冠疫情网络舆情传播随着时间的推移数据增长减少直至逐渐消失[10]。同时，根据上文阐述的突发事件网络舆情的四个阶段的生命周期，本文大致将新冠疫情网络舆情传播分为四个阶段，即萌芽期、成长期、成熟期、衰退期，具体传播特点及内容见表 1，总结了不同阶段的传播特点和内容。

Table 1. Four stages of network public opinion dissemination of “COVID-19”
表 1. “新冠疫情”网络舆情传播四阶段

传播阶段	传播特点及其内容
萌芽期	通过百度指数搜索“新冠疫情”在给定时期内，可以从中可以发现在新十条出台之前，搜索数据虽有增长但是峰值不高

续表

成长期	新十条出台期间有关于放开管控居家隔离不再集中隔离的相关文件出台后，疫情局面全面放开，隔离管控全面解除，异常情况实行自我监控，居家隔离。因为这是完全不同以前的政策决策，再加上这是切实关系每个人的生命健康以及全体利益的决策，所以关键词数据搜索量出现大幅度增长
成熟期	进入 12 月中下旬以后疫情蔓延，但是因为之前已经有了一轮爆发期，相关专家和感染者为大众提供了自己的经验体会和建议，人们的恐慌情以及焦虑情绪得到明显缓解。相关搜索数据趋于稳定
衰退期	1 月上旬到一月底，正逢春节，虽然春运局面规模仍然很大，但是一轮感染基本上已经结束，疫情局面基本稳定，搜索数据明显下降，基本可以算作信息衰退期

2.3. 新冠疫情网络舆情不同阶段的管理建议

萌芽期：舆论预测和管理。在初期阶段，突发事件的网络舆情表现出了分散、无序、浮动等特点，要对其展开汇集、分析、预警和研究判断、争取提前做好准备，这对抢占引导主动权是有利的。首先，加强政府对网络媒体的认识，组建舆论应对团队，关注各平台的舆论发展趋势，形成科学的舆论导向。其次，对潜在的舆论危机信息进行预警，采取信息公开、公众参与、舆论导向等方式，在第一时间疏导人们的负面情绪。最后，形成一套针对突发事件的网上舆情组，为政府在应对类似的舆情事件时，提供有效的借鉴。

成长期：舆论的整理和现场的解决。在成长阶段，网上舆论呈现出聚集化、有序化的趋势，应对的关键是对舆论进行梳理，并采取相应的对策。在这个阶段，要判断突发事件的网络舆情的风险率会变得更加难以，要想从大量的舆情中找出有可能引起突发事件的信息来源，就需要成立一个专门的舆情应对团队，从庞大、复杂、动态的突发事件的网络舆情中找出关键的信息，然后对症下药，充分利用媒体的“议事日程”，给这些信息赋予不同的显著性，及时地发布权威的信息，还要充分地发挥舆论领袖的作用，并对其进行正确的引导。

成熟期：舆情抑制与防止蔓延。已经进入成熟阶段的网络事件，影响巨大，涉及广泛，在这个阶段要对其保持高度的重视，具体可从以下两个方面进行管理。第一，关注百度指数实时搜索数值，通过搜索形成的网络词云，切实了解目前民众所关注的事件以及可能会产生恐慌情绪的公共事件。第二，监控各平台，关注文章标题，发表时间，点击率，回复率，回复率，回复中所反映的主要态度。有目的地收集并分析事件的网络舆情，并对其发展趋势进行预测，为更好地处理网络舆情提供有价值的参考。

衰退期：舆论监督和信心重建在这段时间里，舆论会慢慢地走向消退和沉寂，有关的议题也会趋向于低潮和沉寂，但是，政府和有关机构仍然需要对此保持密切的关注，以防止民众再次陷入恐慌。加强对政府行为的宣传、处罚措施的宣传和行业自律等方面的工作。与此同时，筛选出有利于社会健康发展、生活保障坚定有力、社会管理平稳有序、社会风气不断提高的内容，为广大群众提供积极的社会氛围。

3. 基于 LSSVM 和 ARIMA 组合模型的新冠疫情网络舆情预测

组合预测是提高预测精度的最佳方法之一。在过去的 20 年中，国内外学者对此进行了大量的研究。在组合预测中，一个重要的方法就是加权系数的估计，其中最优加权模型中各个方法的权重，就是按照一定的标准构建的目标函数，并且在有限制的情况下，使目标函数的权重达到最小。其中，目标函数多数依据误差确定，绝对误差、相对误差、对数误差等，目标函数极小化的准则也有多种，如最小二乘法，极小极大法等。本文将赋以单一模型权重，运用基于最小二乘支持向量机(LSSVM)和自回归移动平均

(ARIMA)模型的组合模型对新冠疫情网络舆情进行预测分析。

3.1. 基于 LSSVM 模型的预测分析

根据最小二乘支持向量机(LSSVM) [11]的理论基础,对数据进行处理,将处理后的数据划分为4组,每组基本都有20条数据纪录。在此数据支撑基础之上,建立模型并验证模型可靠性,预测未来一个月的搜索指数数据变化趋势。

3.1.1. 模型理论基础

最小二乘支持向量机(LSSVM) [11]是以统计学习理论为基础,发展起来的一种新的机器学习技术。学习能力和泛化能力,正逐渐成为继神经网络之后机器学习领域新的研究热点[12]。支持向量机可以很好地解决高维数、小样本和非线性等困难,并将其成功地应用到了分类、近似和预测等问题中。因为网络舆情数据具有非线性、复杂性等特点,所以可以利用支持向量机来构建新冠疫情的预测模型。

传统的 LSSVM 模型中的参数设定具有较大的主观性,为解决这一问题,采用交叉验证法来寻找最优模型的参数,以建立更为科学合理的新新冠疫情预测模型[13]。LSSVM 将平方项作为优化指标,并引入等式约束,将其损失函数从传统的二次规划函数替换为最小二乘法函数,这样可以方便地解决分类和函数估计问题,从而让最优化问题变成了求解线性方程,在降低算法复杂程度的同时,还可以提高运算效率[14] [15] [16]。

最小二乘支持向量机的核心思想是把输入数据用非线性映射到一个高维的特征空间,并在该特征空间上做线性回归[17]。

即假设训练集为 $D = \{(x_1, y_1)(x_2, y_2) \cdots (x_m, y_m)\}$, 其中 x_i 为第 i 个样本的特征向量, y_i 为第 i 个样本的真实值。确定 LSSVM 的目标函数。

目标函数的第一项表示最小化权重向量的 L2 范数,以防止过拟合。第二项是一个平方损失函数,用来表示最小化预测值与真实值之间的误差平方和。 C 是正则化参数,它用于平衡最小化误差和最小化权重向量 L2 范数之间的权衡。 C 越大,模型越倾向于最小化误差,而 C 越小,模型越倾向于最小化权重向量的 L2 范数。

约束条件与 LSSVM 类似,LSSVM 的优化问题也存在约束条件。训练样本的真实值 y_i 与预测值 $f(x_i)$ 之间的差异需要小于等于 ϵ , 即 $|y_i - f(x_i)| \leq \epsilon$, ϵ 为一个给定的容忍度。这个约束条件可以表示为:

$$-\epsilon \leq y_i - f(x_i) \leq \epsilon \quad (1)$$

将约束条件代入目标函数,可以得到 LSSVM 的对偶问题。

3.1.2. 模型分析步骤

- (1) 通过训练集数据来建立最小二乘支持向量机(LSSVM)回归模型。
- (2) 将建立的支持最小二乘支持向量机(LSSVM)回归模型应用到训练、测试数据,得到模型评估结果。
- (3) 由于最小二乘支持向量机(LSSVM)回归具有随机性,每次运算的结果不一样。
- (4) 支持最小二乘支持向量机回归无法像传统模型一样得到确定的方程,通常通过测试数据预测精度来对模型进行评价。

3.1.3. 模型拟合结果

从图 3 模型绘制的图像,对比分析模型得到的预测值和数据真实值之间的误差。可以看出最小二乘支持向量机的拟合效果较好,可以将其纳入组合模型的选择中。

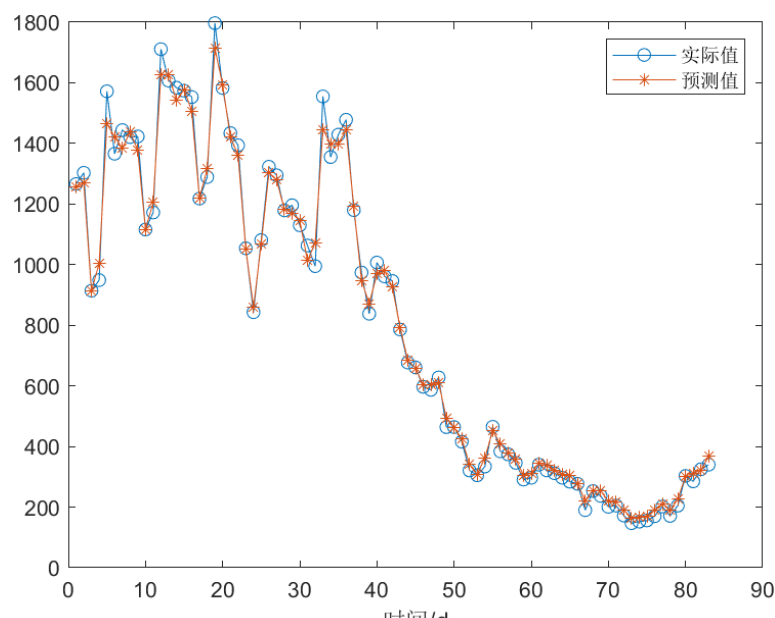


Figure 3. LSSVM model fitting effect diagram

图 3. LSSVM 模型拟合效果图

3.2. 基于 ARIMA 模型的预测分析

对时间序列数据进行分析和预测比较完善和精确的算法是博克斯 - 詹金斯(Box-Jenkins)方法,其常用模型包括:自回归模型(AR 模型)、滑动平均模型(MA 模型)、自回归 - 滑动平均混合模型(ARMA 模型)、自回归移动平均模型(ARIMA 模型),其中 $ARIMA(p,d,q)$ 模型是 $ARMA(p,q)$ 模型的扩展。

3.2.1. 模型理论基础

自回归移动平均(ARIMA)模型将预测对象随时间推移而形成的数据序列视为一个随机序列,用一定的数学模型来近似描述这个序列。本论文中的相关数据为非平稳性数据,需对其进行处理,直至得到的数据具有自相关函数值及偏相关函数数值无异于 0。通过对数据的自相关、偏相关函数的计算,可以确定该序列是适用于 MA 模型还是属于 ARMA 模型。并对模型的参数进行了估算,和对其进行了统计学上的显著性测试。通过对残差数列作了有统计学意义的假设检验,以判断残差数列是不是白噪声序列。最后运用经过测试的模型对未来一个月的网络舆情进行预测分析[6]。

3.2.2. 模型建立步骤

(1) 数据的获取

时间序列的获取方法有两种,一种是通过实验分析得到,另一种就是由有关部门提供的统计数据得到。在收集到的数据中,我们要先检查一下有没有异常,然后再分析是不是有人的失误,或者是其他的原因。确保获取的数据是准确的,是构建恰当的模型、做出恰当的分析的首要保证。本文所用的数据就是以天为单位的时间序列数据。

(2) 时间序列的预处理

本文对时间序列进行了预处理,并对其进行了分析。在此基础上,本文提出了一种基于 ARMA 模型的数据处理方法。在时间序列分析中,最主要的一步就是检验数据的平稳度,常用的方法有序图、相关性图等。

时序图具有直观简单的特征,但误差较大,而自相关图,也就是自相关和副自相关函数图,则比较

复杂,但得到的结果更加精确。文章首先使用序列图来直观地判断,然后使用相关性图来进一步地验证。如果在非平稳的时间序列中出现了增长或者下降的趋势,那么就需要对其进行差分处理,再进行平稳性检验,直到其稳定。

本论文数据明显是非平稳数据,所以对数据进行差分。其中,差分的次数就是模型 $ARIMA(p,d,q)$ 的阶数。在理论上,差分的次数越多,就能更好地从时序信息中提取出不稳定的确定性信息,但在实际操作中,差分的次数不是越多就越好,每一次差分计算,都会导致信息的丢失,因此,应该避免过度的差分。一般在应用中,差分的阶数不超过 2。差分结果如图 4,最终数据变为平稳序列。

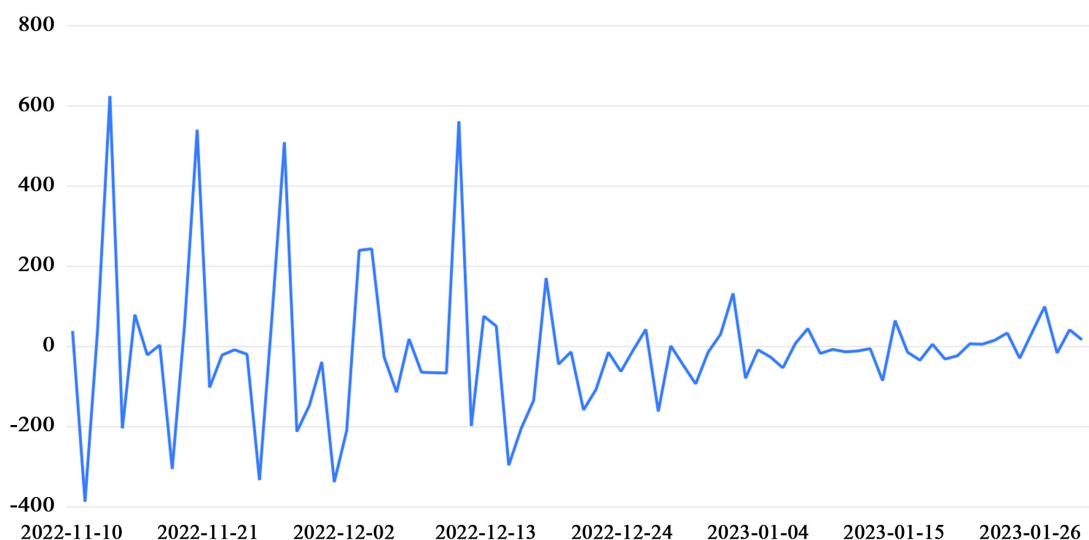


Figure 4. Time series diagram of the original data after first-order difference

图 4. 原始数据 1 阶差分后的时序图

(3) 模型识别

模型识别即从已知的模型中根据所给的时间序列,选取一个合适的模型。目前已有很多模型辨识方法,如 Box-Jenkins 等。

(4) 模型定阶

在确定了模型的类型之后,还需要知道模型的阶数,可使用 AIC 准则法进行定阶。

(5) 参数估计

对模型的参数进行估计的方法通常有相关的矩估计法、最小二乘估计以及极大似然估计等。

(6) 模型的验证

模型的验证主要是对模型的拟合效果进行验证,如果模型完全或者基本解释了系统数据的相关性,那么模型的噪声序列为白噪声序列,那么模型的验证就是噪声序列的独立性检验。贝体的测度方法是运用该定理来构建测度统计量 Q 。如果得到的模型不能通过检查,则需要对其进行重新调整,直到该模型能够通过检查。

3.2.3. 模型拟合结果

通过对数据的差分得到平稳数据后,进行模型的定阶,最终选择模型为: $ARIMA(0,1,2)$ 。如图 5 得到预测值和实际值之间的对比图。

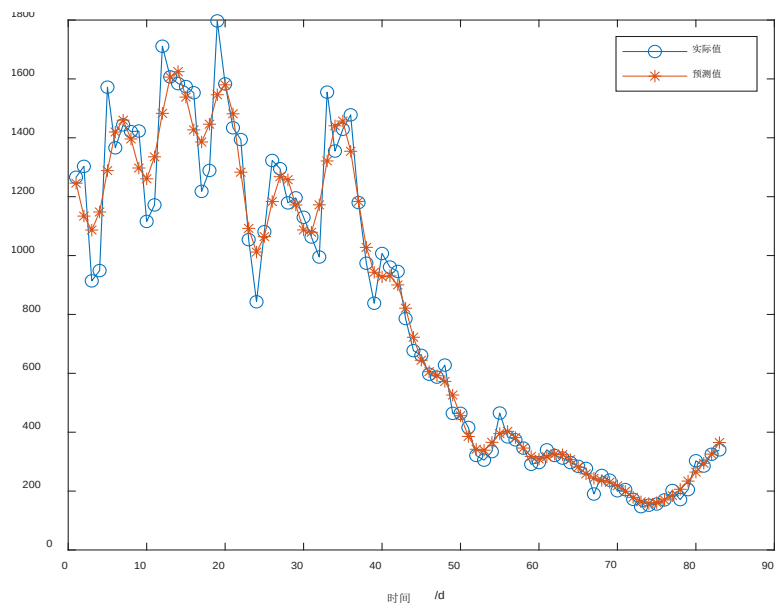


Figure 5. ARIMA model fitting effect diagram
图 5. ARIMA 模型拟合效果图

3.2.4. 模型检验

Table 2. ARIMA model validation
表 2. ARIMA 模型检验

ARIMA 模型(0, 1, 2)检验表		
项	符号	值
样本数量	Df Residuals	78
	N	82
Q 统计量	Q6(P 值)	0.009 (0.926)
	Q12(P 值)	1.244 (0.975)
	Q18(P 值)	34.71 (0.001***)
	Q24(P 值)	56.735 (0.000***)
	Q30(P 值)	69.955 (0.000***)
信息准则	AIC	1059.256
	BIC	1068.833
拟合优度	R ²	0.909

表 2 展示了本次模型检验结果，包括样本数、自由度、Q 统计量和信息准则模型的拟合优度。

ARIMA 模型要求模型的残差不存在自相关性，即模型残差为白噪声，查看模型检验表，根据 Q 统计量的 P 值(P 值大于 0.1 为白噪声)对模型白噪声进行检验。根据信息准则 AIC 和 BIC 值用于多次分析模型对比(越低越好) R²代表时间序列的拟合程度，越接近 1 效果越好。

系统基于 AIC 信息准则自动寻找最优参数，模型结果为 ARIMA(0,1,2) 检验表，基于变量：时序变量转换_Y，从 Q 统计量结果分析可以得到：Q6 在水平上不呈现显著性，不能拒绝模型的残差为白噪声序列的假设，同时模型的拟合优度 R² 为 0.909，模型表现优秀，模型基本满足要求。

Table3. Model significance test
表 3. 模型显著性检验

	模型参数表					
	系数	标准差	t	P > t	0.025	0.975
常数	-12.446	8.044	-1.547	0.122	-28.211	3.319
ma.L1.D.时序变量转换_Y	-0.178	0.107	-1.663	0.096	-0.389	0.032
ma.L2.D.时序变量转换_Y	-0.38	0.109	-3.498	0	-0.593	-0.167

由表 3 可知基于变量时序变量转换得到的 Y，系统基于 AIC 信息准则自动寻找最优参数，模型结果为 ARIMA(0,1,2) 检验表且基于 1 阶差分数数据，最终模型公式如下：

$$y(t) = -12.446 - 0.178 * \varepsilon(t-1) - 0.38 * \varepsilon(t-2) \quad (2)$$

3.3. 组合模型建立及预测

通过比较可以发现其中使用最小二乘支持向量机模型得到的拟合程度最好，而使用 ARIMA 模型得到的结果次之，将这两种模型作为组合模型中的基础模型，分别根据这两个模型与原始数据的不同拟合程度给予不同的权重[10] [18]，从而得到想要的组合模型。

使用得到的组合模型进行未来数据的预测，其中组合模型的预测数据与基础模型的数据对比分析如表 4 所示，可以看出不同模型之间的误差对比。

通过表 4 中的数据对比分析，可以看出组合模型的建立使得预测值的精度提高，使得预测值和真实值之间的误差减小。从中也可以看出预测的百度指数搜索值呈现出逐渐下降趋势，这也与实际情况相吻合。

Table 4. Model comparison analysis table
表 4. 模型对比分析表

日期	真实数据	LSSVM 预测数据	ARIMA 预测数据	组合模型预 测数据	LSSVM 预 测相对误差	ARIMA 预测 相对误差	组合模型预 测相对误差
2022.12.21	946	902	880	923	0.047	0.075	0.025
2023.01.10	321	318	331	323	0.009	0.031	0.006
2023.01.30	325	322	312	324	0.009	0.040	0.003
2023.02.10	269	283	290	275	0.049	0.078	0.021
2023.02.20	391	400	413	389	0.022	0.056	0.005
2023.03.10	218	226	230	223	0.035	0.055	0.022

4. 结论

本文首先以生命周期理论为理论基础，将突发公共卫生事件网络舆情划分为萌芽期、成长期、成熟期和衰退期 4 个发展阶段，针对各阶段给出了相应的管理建议。

然后，以突发事件的百度搜索指数为依据对其网络舆情发展进行预测。以新冠疫情为例，分别利用 ARIMA 模型和 LSSVM 对舆情数据进行拟合预测。基于两种单一模型的拟合预测效果，综合考虑到舆情数据具有时变性特征，以误差平方和最小为建模准则，求解得到最佳权重组合，构建了权重组合预测模型以期提高模型整体的拟合与预测精度。通过实验结果可以表明：赋权组合模型在充分利用已有舆情数据的基础上，能够较好程度地预测出后期舆情的发展态势，验证了该模型的有效性与可行性[19] [20]。

基金项目

上海工程技术大学(SUES)市级大学生创新训练计划(项目编号：cs2121002)。

参考文献

- [1] 雷园. 突发公共卫生事件网络舆情传播效果的影响因素分析[D]: [硕士学位论文]. 北京: 北京化工大学, 2022.
- [2] 田红梅. 突发公共卫生事件下网络舆情传播机理研究[D]: [硕士学位论文]. 石家庄: 河北科技大学, 2022.
- [3] 郭圳凝, 张筱荣. 突发公共卫生事件中网络舆情的整体性治理研究[J]. 安徽理工大学学报(社会科学版), 2022, 24(1): 19-29.
- [4] 朱家安. 突发公共卫生事件下的网络舆论与社会治理[J]. 新闻传播, 2022(1): 40-41.
- [5] 王旭, 孙瑞英. 基于 SNA 的突发事件网络舆情传播研究——以“魏则西事件”为例[J]. 情报科学, 2017, 35(3): 87-92.
- [6] Singh, S., Parmar, K.S., Kumar, J., *et al.* (2020) Development of New Hybrid Model of Discrete Wavelet Decomposition and Autoregressive Integrated Moving Average (ARIMA) Models in Application to One Month Forecast the Casualties Cases of COVID-19. *Chaos, Solitons & Fractals*, **135**, Article 109866. <https://doi.org/10.1016/j.chaos.2020.109866>
- [7] 郝春艳. 基于 SD 的突发事件网络舆情政府应对能力研究[D]: [硕士学位论文]. 大连: 大连理工大学, 2021.
- [8] 牟冬梅, 靳春妍, 邵琦. 基于情感分析的突发公共卫生事件网络舆情热度预测模型仿真[J]. 现代情报, 2021, 41(10): 59-66.
- [9] 靳春妍. 基于情感分析的突发公共卫生事件网络舆情热度预测研究[D]: [硕士学位论文]. 长春: 吉林大学, 2021.
- [10] 程铁军, 王曼, 黄宝凤, 冯兰萍. 基于 CEEMDAN-BP 模型的突发事件网络舆情预测研究[J]. 数据分析与知识发现, 2021, 5(11): 59-67.
- [11] 王启云, 郑中团. CEEMDAN-HURST 算法在新冠疫情预测中的应用[J]. 计算机工程与应用, 2023, 59(7): 261-268.
- [12] 曾杰, 张华. 基于最小二乘支持向量机的风速预测模型[J]. 电网技术, 2009, 33(18): 144-147.
- [13] 王凯, 侯著荣, 王聪丽. 基于交叉验证 SVM 的网络入侵检测[J]. 测试技术学报, 2010, 24(5): 419-423.
- [14] 李琨, 韩莹, 黄海礁. 基于 IBH-LSSVM 的混沌时间序列预测及其在抽油井动液面短期预测中的应用[J]. 信息与控制, 2016, 45(2): 241-247, 256.
- [15] Yang, Y.F., Qin, Y., Jia, L.M. and Dong, H.H. (2016) Traffic Safety Region Estimation Based on SFS-PCA-LSSVM: An Application to Highway Crash Risk Evaluation. *International Journal of Software Engineering and Knowledge Engineering*, **26**, 1555-1570. <https://doi.org/10.1142/S0218194016400179>
- [16] Gao, S.Z., Li, T.C. and Zhang, Y.M. (2020) Rolling Bearing Fault Diagnosis of PSO-LSSVM Based on CEEMD Entropy Fusion. *Transactions of the Canadian Society for Mechanical Engineering*, **44**, 405-418. <https://doi.org/10.1139/tcsme-2019-0114>
- [17] 白鹏, 张喜斌, 张斌, 等. 支持向量机理论及工程应用实例[M]. 西安: 西安电子科技大学出版社, 2008: 14-15.
- [18] 程铁军, 王曼. 基于变权组合的突发事件网络舆情趋势预测[J]. 计算机科学, 2021, 48(S1): 190-195+202.
- [19] 杨茂青, 谢健民, 秦琴, 等. 基于 RF 算法的突发事件网络舆情演化预测分析[J]. 情报科学, 2019, 37(7): 95-100.
- [20] 杜洪涛, 王君泽, 李婕. 基于多案例的突发事件网络舆情演化模式研究[J]. 情报学报, 2017, 36(10): 1038-1049.