

基于时间序列分析的网络谣言传播数据研究

刘新柔, 张靖, 周煜

新疆农业大学数理学院, 新疆 乌鲁木齐

收稿日期: 2025年3月22日; 录用日期: 2025年4月13日; 发布日期: 2025年4月24日

摘要

网络谣言是没有事实依据, 且通过网络介质传播带有攻击性、目的性的不实信息。突发性强、流传速度极快, 且内容往往耸人听闻, 容易引发公众恐慌和不安。研究网络谣言的传播机制, 有助于及时揭露谣言, 为公众提供准确信息, 避免其受到错误信息的误导, 从而保护公众的知情权和判断力。本文介绍了谣言传播机制, 使用SIR和LSTM时间序列模型研究网络谣言在不同时间的传播状态及趋势, 分析网络谣言的传播规律, 识别关键传播节点。本文基于真实案例“房屋养老金”谣言, 使用两个模型研究谣言传播机制, 这对于制定有效的谣言控制策略具有重要意义。

关键词

网络谣言传播, SIR模型, LSTM模型

Research on Network Rumor Propagation Data Based on Time Series Analysis

Xinrou Liu, Jing Zhang, Yu Zhou

School of Mathematics and Physics, Xinjiang Agricultural University, Urumqi Xinjiang

Received: Mar. 22nd, 2025; accepted: Apr. 13th, 2025; published: Apr. 24th, 2025

Abstract

Internet rumors are unfounded pieces of information that are disseminated through online media, characterized by their aggressive and purposeful nature. They often emerge abruptly, spread rapidly, and contain sensational content that can easily incite public panic and unease. Investigating the propagation mechanisms of internet rumors is crucial for promptly debunking these falsehoods, providing the public with accurate information, preventing the public from being misled by misinformation, and thereby safeguarding the public's right to know and their ability to make informed judgments. This article introduces the mechanisms of rumor propagation, employing the SIR and LSTM time series models to study the state and trends of internet rumor dissemination over time,

analyze the patterns of rumor spread, and identify key nodes in the propagation network. This paper is based on the real case of the “House Pension” rumor, utilizing two models to investigate the mechanisms of rumor propagation, which holds significant importance for formulating effective strategies to control rumors.

Keywords

Internet Rumor Propagation, SIR Model, LSTM Model

Copyright © 2025 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

中国互联网络信息中心发布第 54 次《中国互联网络发展状况统计报告》中显示,截至 2024 年 6 月,我国网民规模近 11 亿人(10.9967 亿人),较 2023 年 12 月增长 742 万人,互联网普及率达 78.0%。社交媒体的普及和便捷性为网络谣言的传播提供了便利条件。人们可以通过社交媒体快速地分享和传播信息,同时也可以通过社交媒体进行互动和讨论,从而加速网络谣言的传播。

例如 2024 年 1 月 17 日,“清华大学辟谣 80%毕业生出国”的消息,迅速传播。清华大学公布的毕业生就业去向显示,本科生选择继续深造的比例为 80.8%,硕士生选择继续深造的比例为 12.8%。2024 年 4 月 19 日,网传湘潭大学有学生在外卖中投毒致偷外卖的学生被毒死。湘潭大学回应:网传偷外卖被投毒不属实。不法分子编造虚假信息,将谣言炒得沸沸扬扬,换取不正当利益。虚假信息容易导致公众不安,危害社会稳定,尤其是在涉及公共安全或健康的事件中。

传统谣言传播研究方法存在一定的局限性,如:难以量化分析、缺乏预测能力等。将时间序列分析方法应用于网络谣言传播研究,能够捕捉传播趋势、识别关键节点、预测未来走向等。网络谣言传播的研究不仅具有现实意义,还具有重要的学术价值。通过对网络谣言传播规律深入研究,可以丰富传播学、社会学、心理学等相关学科的理论体系。

2. 网络谣言传播特性

网络谣言的定义建立在网络传播的基础之上,互联网是新时代下谣言传播的全新工具及媒介。随着微博、抖音、小红书等新兴网络软件的兴起,网络谣言的传播具有极高的效率。针对于近年来网络谣言传播的频繁发生,此前已有学者针对谣言传播进行了系列研究[1]-[4]。

网络谣言传播的整个过程所需时间极大的缩短,从谣言产生、传播到爆发极大的舆论浪潮,只需要几天甚至几小时的时间。通过分析近几年主要平台谣言传播的规律,可将网络谣言的传播过程分为图 1 中的五个阶段。

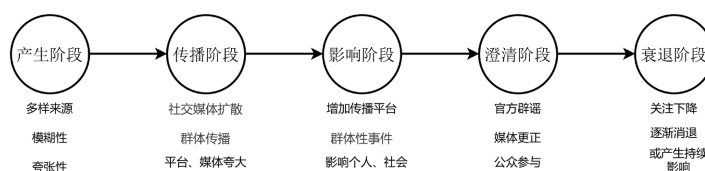


Figure 1. The process of spreading online rumors

图 1. 网络谣言传播过程图

通过研究网络谣言的传播规律，可以识别谣言传播的关键节点、路径和高峰期，从而为及时辟谣、阻断谣言扩散提供科学依据。下面基于 2024 年网络真实案例(房屋养老金)，研究网络谣言传播规律。

3. 网络谣言传播数据研究

在 2024 年 8 月 25 日，一则关于“房屋养老金的收费标准”的传言在社交平台引起热议。传言称，“上海浦东试点房屋养老金的收费标准提前透露，主要依据房屋的面积和楼层而定”，还像模像样地标注了不同房屋的收费细则。在极短时间内，“房屋养老金”相关谣言，通过社交平台迅速以图片、视频、真人解读等形式在个人账号、营销号急速传播。社交平台的信息传播速度快、范围广，谣言往往借助情绪化内容吸引关注，并在短时间内形成“信息爆炸”。

本文通过网络舆情监测平台“知微事见”收集得到“房屋养老金”谣言的传播数据(2024 年 8 月 25 日~2024 年 9 月 6 日)，数据包括微博、微信、网媒、抖音等主要网络平台的谣言传播量。图 2 分别给出了每天和每小时“房屋养老金”谣言传播量数据图。

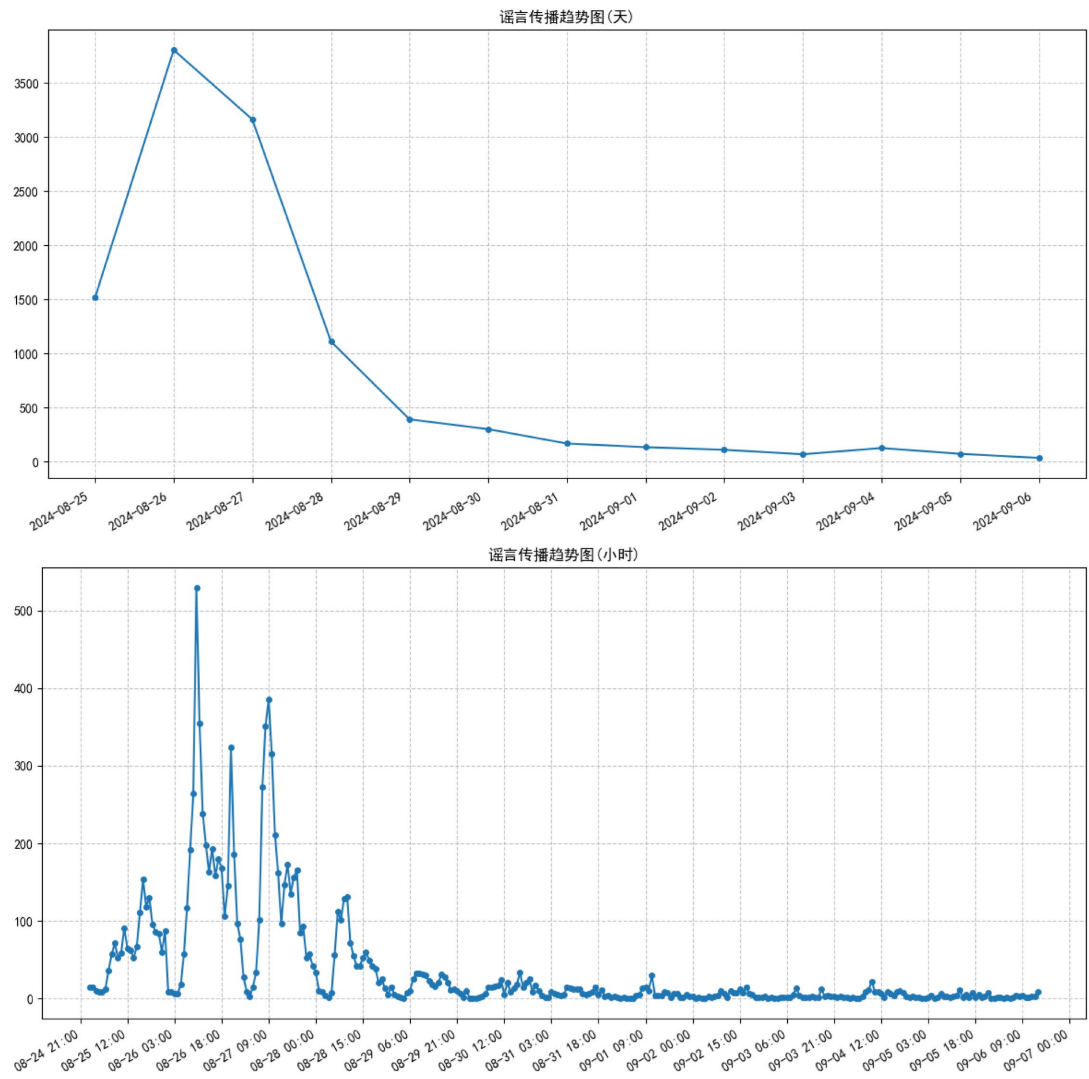


Figure 2. Trends in the spread of rumours
图 2. 谣言传播趋势图

图 2 中第一图数据表示每日新增谣言传播量,共 13 天的观测值,传播周期约为两周。从最低 34 到最高 3804,跨度较大,表明谣言传播过程经历了显著的变化。第一天到第二天传播速度极快,谣言在初期迅速扩散,高峰期持续较短,衰减期较长,网络谣言传播逐渐减弱。

图 2 中第二个图数据表示每小时新增谣言传播量,共 240 小时的观测值。从最低 0 到最高 529,跨度较大,传播过程经历了显著的变化。传播过程中存在多个峰值和波动,传播动态复杂。

3.1. SIR 模型

网络谣言的传播和传染病传播在扩散方式、人群分类上表现出极高的相似性,故在传染病模型视域下探索网络谣言传播问题。SIR (susceptible infected recovered)模型被运用于研究谣言传播扩散[5] [6],也有学者们在此基础上提出了带辟谣机制[7]、怀疑机制[8]谣言传播模型。

SIR 模型中 S 是尚未接触谣言但可能传播谣言的人群、I 是已经接触并正在传播谣言的人群、R 是已经知道谣言但不再传播的人群(包括对谣言失去兴趣或认为谣言不真实的人)。易感人群按照一定的比例 A 转化为感染人群,感染人群按一定的比例 B 转化为康复人群,如图 3 所示。参数 A 的估计通过数据拟合方法确定具体采用最小化模型预测值与实际传播数据的均方误差作为目标函数,并运用优化算法进行迭代调整。参数 B 主要利用传播量下降阶段的数据进行衰减期拟合。通过最大似然估计方法确定 B 值,同样辅以初始值设定和敏感性分析来确保估计的准确性。



Figure 3. SIR rumor propagation model

图 3. SIR 谣言传播模型图

SIR 谣言传播模型中,假定总人口 N 为一个固定值,通过微分方程描述传播动态:

$$\begin{cases} S(t) + I(t) + R(t) = N, \\ \frac{dS}{dt} = -AS(t)I(t), \\ \frac{dI}{dt} = AS(t)I(t) - BI(t), \\ \frac{dR}{dt} = BI(t). \end{cases}$$

根据图 2 提供的每天、每日“房屋养老金”谣言传播时间序列数据,分别拟合 SIR 模型,绘制拟合曲线。

在拟合 SIR 模型后,可计算 SIR 模型的预测值与实际值之间的残差,即

$$\varepsilon_t = I(t) - \hat{I}(t)$$

并绘制残差图。图 4 中左侧上下两幅图为 SIR 模型每天、每时拟合曲线图,右侧上下两幅图为每天、每时残差图。

通过图 4,我们可以直观发现 SIR 模型在拟合谣言传播数据时,能够较好地描述传播的基本动态,但通过残差图,我们可以看到预测值和实际数据存在差异,预测精度稍微有点低。

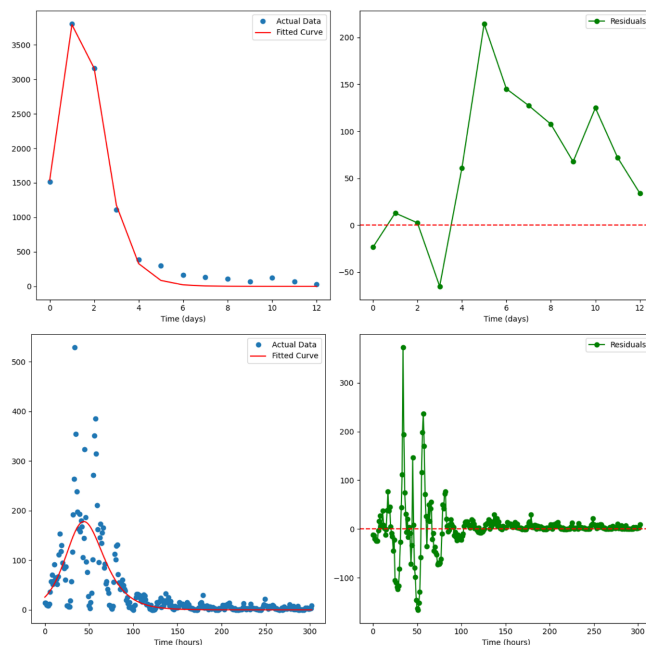


Figure 4. The SIR model fits the curve

图 4. SIR 模型拟合曲线图

此外, 还可使用 SIR 模型分别可视化每天、每小时谣言传播变化趋势、谣言传播速度。通过 SIR 模型, 还可以提取谣言传播过程中的关键特征(峰值点、爆发期、衰减期、传播关键节点等)。

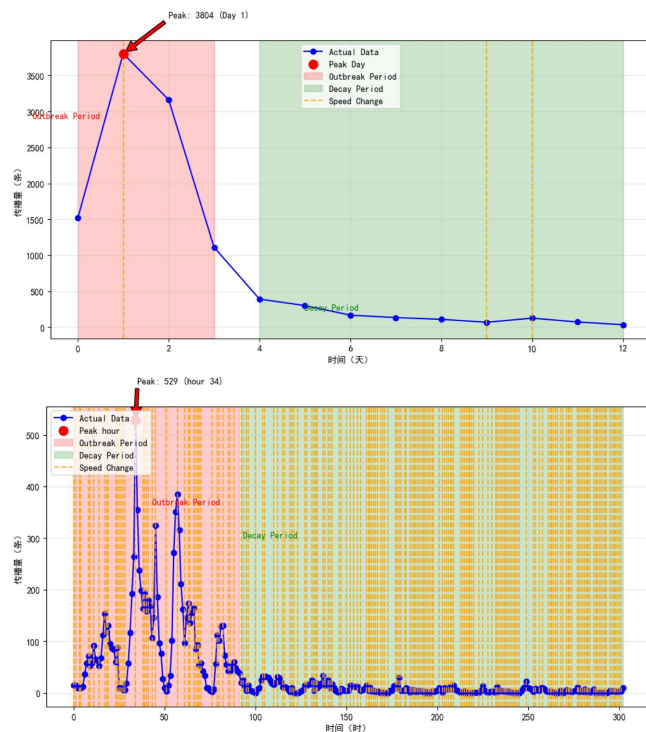


Figure 5. Analysis of the trend of rumor dissemination

图 5. 谣言传播趋势分析图

通过图 5，我们可以清晰地识别谣言每天、每时传播速度、爆发期、衰退期、传播关键节点等关键特征。图 5 的分析结果可以帮助我们更好地理解谣言传播的动态过程，并根据关键节点的确定、谣言传播阶段的划分，合理制定防控策略。

3.2. LSTM 模型

长短期记忆网络(LSTM)对时间序列数据进行建模是一种常见的方法，是一种特殊的循环神经网络(RNN)，擅长处理时间序列、文本等具有序列依赖性的数据[9]。此外，也有学者将 LSTM 模型应用于传染病[10]、股票分析[11] [12]等研究。

LSTM 模型的网络结构包含输入层、隐藏层和输出层。输入层的时间步长设为 24 (小时数据)或 7 (天数据)，与传播周期相匹配。隐藏层采用 2 层 LSTM 结构，每层 128 个神经元，并加入 Dropout 层(比率 0.3)防止过拟合。输出层使用全连接层输出单值预测结果。损失函数采用均方误差；优化器选择 Adam。调整过程中，采用 7:3 的数据划分进行交叉验证，并实施早停法监控验证损失。敏感性分析表明，128 个神经元和 0.3 的 Dropout 比率为最优选择。该结构设计确保了模型预测的准确性。

下面将根据网络谣言真实案例的传播量数据，利用 LSTM 模型分析网络谣言传播的整体趋势、周期性特征，建模及分析过程如图 6 所示。

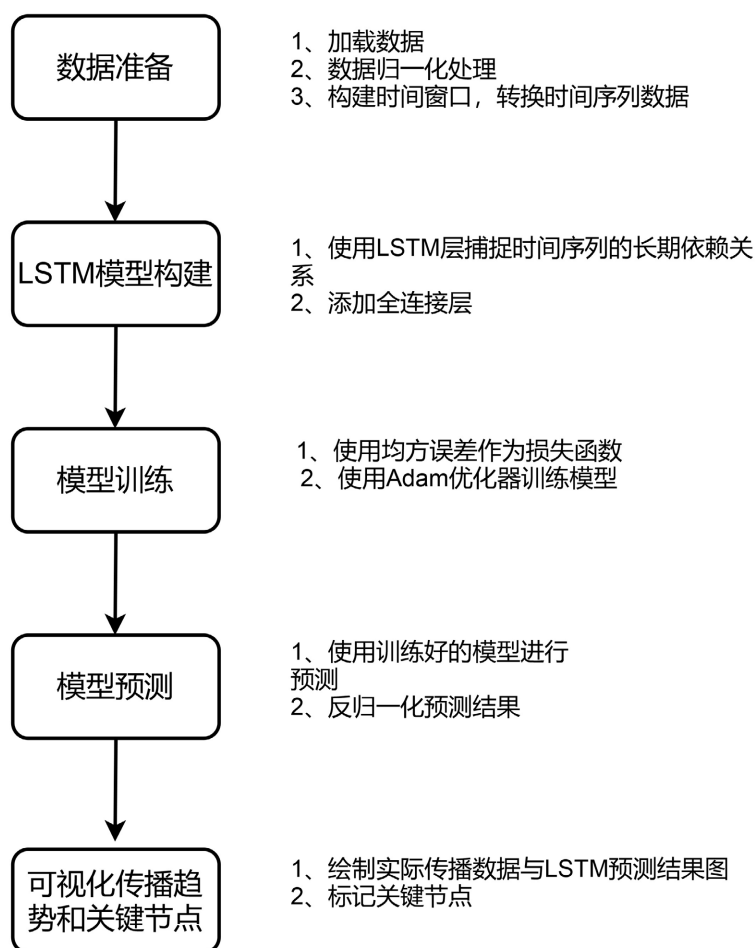


Figure 6. LSTM model steps
图 6. LSTM 模型步骤图

LSTM 模型通过引入记忆单元和门控机制来解决传统 RNN 的梯度消失问题。LSTM 单元的核心结构包括以下部分：

1. 遗忘门：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

2. 输入门：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

3. 记忆单元更新：

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

4. 输出门：

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = O_t \cdot \tanh(C_t)$$

在时间序列预测中，常用的损失函数时均方误差

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

常用的优化算法是 Adam，结合了动量法和自适应学习率。Adam 算法的关键在于同时计算梯度的一阶矩(均值)和二阶矩(未中心的方差)的指数移动平均，并对它们进行偏差校正，以确保在训练初期时梯度估计不会偏向于 0。算法的更新规则如下：

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot \nabla L$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot (\nabla L)^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$\theta_t = \theta_{t-1} - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

本文将网络谣言传播过程中的数据按照时间顺序排列，形成时间序列数据。使用 LSTM 模型学习这些时间序列数据中的模式，理解谣言传播的趋势和规律。

通过 LSTM 模型特征提取，提取谣言传播过程中的关键特征，如传播速度、爆发期、衰减期。

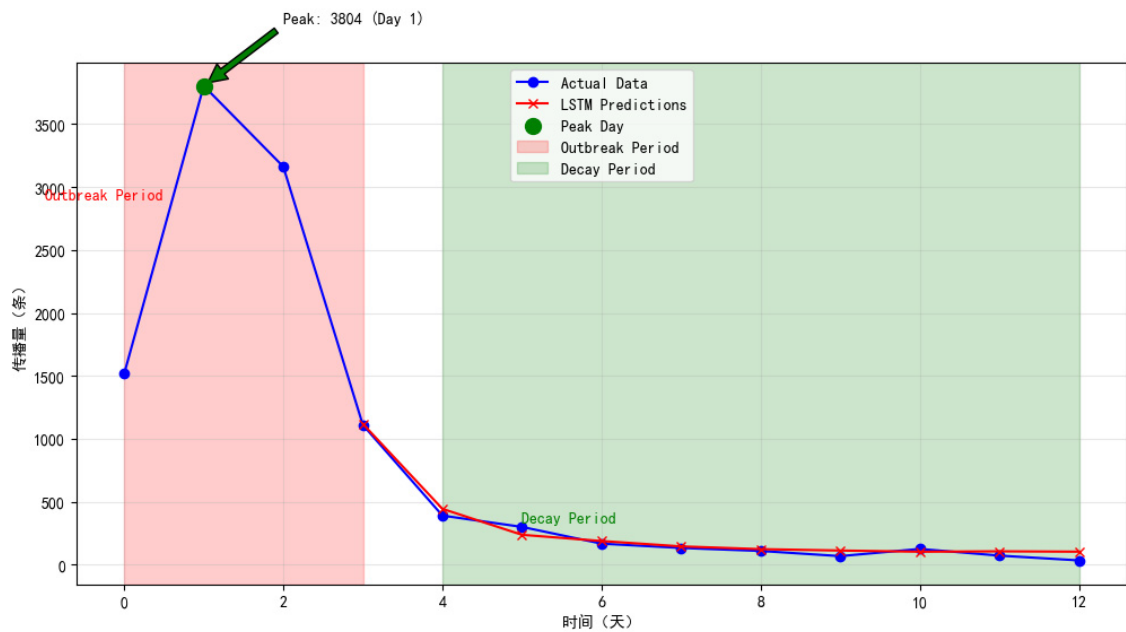


Figure 7. The LSTM model fits the amount of rumor spread (days)

图 7. LSTM 模型拟合谣言传播量(天)图

通过图 7，发现谣言传播初期迅速达到峰值(第 1 天，3804 例)，随后快速下降。衰减期持续时间较长，但新增感染者数量逐渐减少。还可通过图 5 看到谣言传播的关键节点：峰值点：第 1 天为传播高峰，是防控干预的关键时间点。爆发期：前 4 天为传播加速阶段。衰减期：从第 4 天起传播速度下降，进入消退阶段。

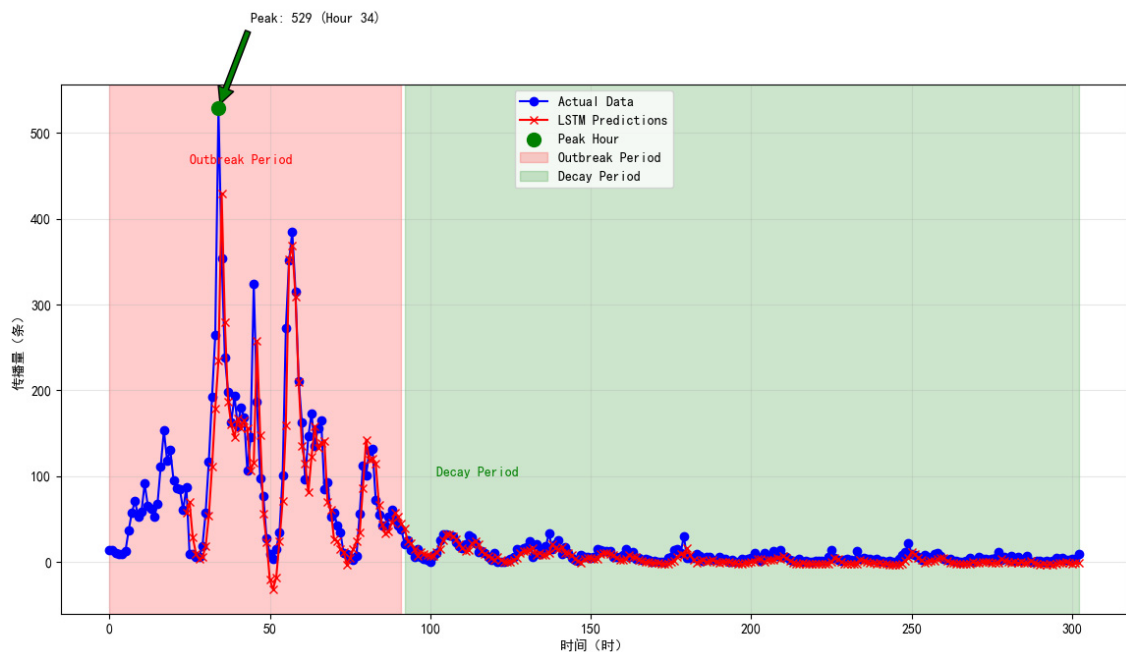


Figure 8. The LSTM model fits the amount of rumor spread (hours)

图 8. LSTM 模型拟合谣言传播量(时)图

通过图 8，发现谣言传播初期迅速达到峰值(第 34 小时，529 例)，随后快速下降。衰减期持续时间较长，但新增感染者数量也是逐渐减少。从图中可看到关键节点，峰值点：第 34 小时为传播高峰，是防控干预的关键时间点。爆发期：前 34 小时为传播加速阶段。衰减期：从第 35 小时起传播速度下降，进入消退阶段。

该条谣言已经通过上海辟谣平台从上海市住建委、市房管局获悉，传言不实。在官方进行辟谣之后，此谣言的传播量几乎为零。辟谣后，谣言传播量迅速下降，但仍有一些用户因情感驱动继续转发。

3.3. 模型对比分析

3.3.1. 定量评价

衡量 SIR 模型和 LSTM 模型预测值与实际值之间的平均平方误差，值越小表示拟合效果越好。均方误差

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

衡量 SIR 模型和 LSTM 模型预测值与实际值之间的平均绝对误差，值越小表示拟合效果越好。平均绝对误差

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

通过 python 软件计算，我们发现 LSTM 模型的 MSE 值和 MAE 值更小，其拟合的效果也更好。

3.3.2. 定性评价

使用 SIR 模型和 LSTM 模型分析网络谣言传播趋势，通过网络谣言传播数据与 SIR 模型、LSTM 模型绘制传播趋势图，可以直观评估模型的拟合效果，比较模型识别的关键节点与实际数据的吻合程度。从上面的几幅拟合图、特征识别图中，我们可以发现 LSTM 模型在网络谣言传播拟合趋势的效果、模型特征识别的效果更好，并由表 1 给出了两种模型的对比分析。

可视化每日、每时谣言传播量的变化趋势，分析传播速度的加速或减速趋势，识别谣言传播的爆发期和衰减期。这些特征有助于更深入地理解谣言的传播机制，能够有效遏制谣言传播，维护社会稳定。

Table 1. Model comparison analysis

表 1. 模型对比分析表

	SIR 模型	LSTM 模型
理论基础	基于传染病模型	基于深度学习
传播趋势分析	适用于简单传播场景	适用于复杂传播场景
关键节点识别	通过求解模型参数识别	通过预测结果识别
计算复杂度	计算简单，速度快	计算复杂，速度较慢
适用场景	理论分析、简单传播场景	数据驱动的复杂传播预测

4. 阻止谣言传播措施

利用时间序列方法研究网络谣言的传播，识别出传播周期、爆发点、衰减特征等，有目的有层次地阻断谣言。通过数据预处理、模型训练等，可以实现对传播数据的准确预测和分析。了解谣言传播机制，在关键节点阻止谣言的扩散和传播。在不同时期阻断谣言传播需要采取不同的策略。

1. 谣言产生期

初期通过监测预警和快速反应。可利用 AI 关键词拦截和权威账号联动等技术手段实时监测网络平台, 及时发现谣言苗头。一旦识别到可能为不实信息, 即可求证, 并截断传播源头。对于出现的谣言及时拦截。在微博、微信等多以文字 + 图片类谣言的传播平台, 可以使用 AI 关键词拦截, 实时监测高频敏感词(如“收费”“试点”), 自动标记可疑内容。

2. 谣言爆发期

高峰期通过技术拦截和法律手段。爆发期对应谣言传播的高峰期, 是防控的关键阶段。例如, “房屋养老金”谣言可在爆发期初期(第 1 天到第 2 天)、(第 25 小时到第 35 小时)采取快速干预措施并对案例中的首发账号封禁并追溯法律责任。又如在视频类谣言传播平台如抖音、快手、B 站等可通过画面分析、语音识别、智能打标系统等技术手段拦截谣言并下架相关视频。如 2023 年“疫苗致死”谣言视频, 通过识别伪造的医院场景, 下架率达 89%。通过权威渠道, 如: 主流媒体、公众号、公告等方式及时澄清谣言。以“房屋养老金”谣言为例, 可在微博热搜页面嵌入住建委的辟谣声明, 并限制相关关键词的搜索推荐。此外, 公众应提高辨别能力, 自觉不轻信、不传播未经证实的信息。

3. 谣言衰退期

衰退期通过持续澄清和教育宣传。衰减期对应谣言传播的消退阶段, 传播速度逐渐减慢。在波动衰减阶段持续监测传播动态, 防止传播反弹。即使谣言热度下降, 仍可持续发布权威信息, 消除残余影响。以“疫苗致死”谣言衰退期为例, 系统持续捕获 32 个死灰复燃的传播节点, 卫健委官微每日发布接种数据对比图, 使残余传播量再降 83%。建立谣言举报渠道, 完善相关法律, 严惩制造和传播谣言的行为。长期则通过媒介素养教育和信息透明。如衰退期在微信朋友圈广告推送“谣言辨识”公益广告, 公众号嵌入“谣言粉碎机”小程序, 提供实时查证功能。

只有多管齐下, 才能有效阻断谣言传播, 维护社会稳定。

5. 结论

谣言不仅造成个体伤害, 也在破坏个体共识。谣言经网络平台快速发酵、传播, 谣言干扰真实信息的传播, 危害社会和谐稳定。本文详细介绍了谣言传播机制, 并使用 SIR 谣言传播模型和 LSTM 模型拟合谣言传播趋势, 研究了网络谣言传播的时间序列特征。通过建模和分析谣言传播的动态过程, 可以为及时制定有效的谣言防控策略提供支持。本文使用的谣言传播动力学模型和深度学习模型分析方法研究网络谣言传播, 下一步可以进行多源数据融合, 应用其他深度学习模型研究谣言的传播。

基金项目

新疆农业大学 2024 年度大学生创新训练计划项目(项目编号: dxscx2024599)。

参考文献

- [1] 翟月荧. 网络谣言的传播与治理[J]. 东岳论丛, 2023, 44(8): 150-156.
- [2] 冯雯璐, 刘乃榕, 田晓丽. 网络舆情事件中的谣言传播与智能化治理[J]. 媒体融合新观察, 2024(6): 63-71.
- [3] 刘立伟, 谢晓娟. 新媒体时代网络舆情的新态势及其治理[J]. 学校党建与思想教育, 2024(23): 71-74.
- [4] 高卫国, 蔡永丽. 网络谣言传播的动力学建模及其平衡点稳定性分析[J]. 扬州大学学报(自然科学版), 2022, 25(5): 7-11+53.
- [5] Daley, D.J. and Kendall, D.G. (1965) Stochastic Rumours. *Journal of Applied Mathematics*, 1, 42-55. <https://doi.org/10.1093/imamat/1.1.42>
- [6] 李佳洋, 宋博伟, 王丹. 基于 SIR 的网络谣言演化模型与控制策略[J]. 沈阳大学学报(自然科学版), 2021, 33(2):

- 140-149.
- [7] 王菽裕, 宋俊芳, 张春玉. 考虑双因素辟谣机制的谣言传播模型及其仿真研究[J]. 网络安全技术与应用, 2023(11): 48-51.
 - [8] 张明菊, 仇丽青. 社交网络中考虑怀疑机制的谣言传播模型[J]. 软件导刊, 2021, 20(4): 123-128.
 - [9] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780.
<https://doi.org/10.1162/neco.1997.9.8.1735>
 - [10] 吴昊, 曹宇, 魏海平, 等. 基于自注意力机制 LSTM 的 COVID-19 感染预测[J]. 计算机应用与软件, 2024, 41(9): 106-113.
 - [11] 何杰, 李素平, 何盈盈, 等. 基于 ARIMA 及 LSTM 模型的股票分析[J]. 现代信息科技, 2024, 8(21): 41-45.
 - [12] 丁文娟. 基于股票预测的 ARIMA 模型、LSTM 模型比较[J]. 工业控制计算机, 2021, 34(7): 109-112, 116.