

Information Propagation Model of Public Health Events

Qingwen Zhai¹, Zhaodu Zhang¹, Riyue Zhang¹, Mengjie Yu¹, Xinxin Ma¹, Shuli Yan²

¹School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang Henan

²School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing Jiangsu

Email: 1367438205@qq.com

Received: Jun. 1st, 2020; accepted: Jun. 15th, 2020; published: Jun. 23rd, 2020

Abstract

Through the method of combination prediction, this paper studies the phenomenon of information propagation of public health events under the background of new media. By using the grey GM(1,1) model and the weighted Markov chain prediction theory, the grey-weighted Markov chain prediction model is constructed to predict the future trend of public health emergencies, which effectively improves the prediction accuracy. Finally, taking CSFV in Africa as an example, the feasibility and validity of the model are verified.

Keywords

Grey Prediction Model, Weighted Markov Chain, Public Health Events

公共卫生事件信息传播规律模型

翟晴雯¹, 张照都¹, 张日月¹, 余梦洁¹, 马新新¹, 闫书丽²

¹河南科技大学数学与统计学院, 河南 洛阳

²南京信息工程大学管理工程学院, 江苏 南京

Email: 1367438205@qq.com

收稿日期: 2020年6月1日; 录用日期: 2020年6月15日; 发布日期: 2020年6月23日

摘要

针对新媒体背景下突发公共卫生事件信息传播现象, 通过组合预测的方法, 运用灰色GM(1,1)模型和加权马尔可夫链预测理论, 构建了灰色-加权马尔可夫链预测模型, 对突发公共卫生事件的未来趋势进行预测, 有效提高了预测精度。并以非洲猪瘟事件为例进行了比较分析, 验证了模型的可行性和有效性。

文章引用: 翟晴雯, 张照都, 张日月, 余梦洁, 马新新, 闫书丽. 公共卫生事件信息传播规律模型[J]. 统计学与应用, 2020, 9(3): 482-490. DOI: 10.12677/sa.2020.93051

关键词

灰色预测模型, 加权马尔可夫链, 公共卫生事件

Copyright © 2020 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

数据时代新媒体的产生和发展, 加快了信息传播的速度, 扩大了信息传播的范围, 增强了信息传播的影响。但是目前的众多新媒体平台仍然缺乏有效的管理制度, 一些媒体通过报道虚假信息博取关注、制造舆论, 引发公众情绪恐慌, 影响社会的和谐安定[1] [2]。

突发公共卫生事件对人体健康、环境、经济都有不可忽视的影响, 一直是我国公民关注的热点话题, 其信息传播已经成为足够影响社会安定, 国家和谐的重要因素之一, 所以, 研究、探索如何引导、控制公共卫生事件信息的传播, 对降低其带来的负面影响、对研究方法和应用领域的拓展与创新、为决策部门提供预警, 以及提高我国政府解决突发事件的能力都具有重要的价值和意义[3]。

目前, 已有一些学者对公共卫生事件信息传播规律进行了研究, Galam 运用舆情演化模型, 对少数人的意见在公共争论过程中的传播路径进行了研究[4]; 靳松等利用数据生成 H7N9 舆情信息传播网络拓扑结构对舆情传播特点进行了分析[5]; 洪巍等结合 Logistic 多项式模型对食品安全网络舆情中网民信任网络意见领袖的影响因素进行了研究[6]; 滕文杰建立了时间序列 ARIMA(p,d,q)模型对 2011 年以来发生的典型突发公共卫生事件百度搜索指数进行了建模分析[7]; 刘小洋等结合了传播动力学提出了一种动态扩散网络舆情演化模型[8]。

本文基于灰色预测理论和马尔可夫链的理论知识, 通过组合预测的方法, 构建了灰色 - 加权马尔可夫模型, 以非洲猪瘟事件为例, 对突发公共卫生事件的未来趋势进行预测, 帮助了解突发公共卫生事件发展和影响范围, 及时辅助决策者做出科学合理的决策行为。

2. 灰色 - 加权马尔可夫链预测模型

2.1. 灰色预测模型 GM(1,1)

定义 1 记观测到的 n 个观测值为时间序列 $\hat{X}(0)$ [9], 令

$$X^{(0)} = (X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n))$$

通过一次累加, 得到一次累加序列

$$X^{(1)} = (X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n))$$

定义 2 称

$$X^{(0)}(i) + aX^{(1)}(i) = b \quad (1)$$

为 GM(1,1)模型的原始形式, 其中, 参数 a 为发展系数, b 为灰色作用量。

建立 GM(1,1)模型对应的微分方程为

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = b \tag{2}$$

定义 3 若 $\hat{\alpha} = [a, b]^T$ 为参数列, 且

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -\frac{1}{2}(X^{(1)}(1) + X^{(1)}(2)) & 1 \\ -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(X^{(1)}(i-1) + X^{(1)}(i)) & 1 \end{bmatrix}, \quad i = 1, 2, \dots, n$$

则 GM(1,1)模型 $X^{(0)}(k) + aX^{(1)}(k) = b$ 的最小二乘估计参数列满足

$$\hat{\alpha} = (B^T B)^{-1} B^T Y \tag{3}$$

微分方程的解(即时间响应函数)为

$$x^{(1)}(t) = \left(x^{(1)}(1) - \frac{b}{a} \right) e^{-at} + \frac{b}{a} \tag{4}$$

方程(1)的时间响应序列为

$$\hat{x}^{(1)}(i+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ai} + \frac{b}{a}; \quad i = 1, 2, \dots, n \tag{5}$$

以方程(5)作为 GM(1,1)模型的预测模型, 其还原值为

$$\hat{x}^{(0)}(i+1) = \hat{x}^{(1)}(i+1) - \hat{x}^{(1)}(i) = (1 - e^{-a}) \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ai}; \quad i = 1, 2, \dots, n \tag{6}$$

2.1.1. 相对关联度检验

令 $\eta_{(i)}$ 为关联系数, 根据关联系数定义可得

$$\eta_{(i)} = \frac{\min \min |X^{(0)}(i) - \hat{X}^0(i)| + \rho \max \max |X^{(0)}(i) - \hat{X}^0(i)|}{|\hat{X}^{(0)}(i) - X^{(0)}(i)| + \rho \max \max |\hat{X}^{(0)}(i) - X^{(0)}(i)|} \tag{7}$$

ρ 为分辨率, $0 < \rho < 1$, 一般取 $\rho = 0.5$, $\min \min |X^{(0)}(i) - \hat{X}^0(i)|$ 为两级最小差,

$\max \max |X^{(0)}(i) - \hat{X}^0(i)|$ 为两级最大差, 令原始序列与预测序列的关联度

$$r = \frac{1}{n} \sum_{i=1}^n \eta(i) \tag{8}$$

一般当 $\rho = 0.5$ 时, 若 $r > 0.6$ 表示通过检验。

2.2. 马尔可夫链模型

定义 4 实际值与 GM(1,1)模型预测值的相对误差为 $\varepsilon(k)$, 可表示为

$$\varepsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \tag{9}$$

定义 5 S 为状态空间, 对任意的 $i, j \in S$, 称 $P_{ij}^{(k)}(n) = P\{X_{n+k} = j | X_n = i\}$ 为 X_n 在 n 时刻的 k 步转移

概率, $P^{(k)} = (P_{ij}^{(k)})$ 为随机序列 $\{X_n, n \geq 0\}$ 的 k 步转移概率矩阵[10]。记作

$$P^{(k)} = \begin{bmatrix} P_{11}^{(k)} & P_{12}^{(k)} & \cdots & P_{1n}^{(k)} \\ P_{21}^{(k)} & P_{22}^{(k)} & \cdots & P_{2n}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1}^{(k)} & P_{n2}^{(k)} & \cdots & P_{nn}^{(k)} \end{bmatrix}$$

根据样本数据可知转态 i 出现的次数为 m_i , 由状态 i 经过 k 步转移到状态 j 的次数为 $m_{ij}^{(k)}$, 状态 i 经过 k 步转移到状态 j 概率值的近似值为

$$P_{ij}^{(k)} = \frac{m_{ij}^{(k)}}{m_i} \quad (10)$$

2.3. 灰色 - 加权马尔可夫链模型

定义 6 预测第 n 天的相对误差的状态, 选取离第 n 天最近的 m 天的相对误差所处的状态作为初始状态, 按离预测第 n 天的远近, 转移步数分别为 $1, 2, \dots, m$; 在转移步数对应的转移矩阵 $P^{(k)}$ 中, 取起始状态所对应的行向量 $P_i^k = (p_{i1}^{(k)}, p_{i2}^{(k)}, \dots, p_{im}^{(k)})$, $i \in S$, 构造新的概率矩阵[11]

$$P = \begin{bmatrix} P_{11}^{(k)} & P_{12}^{(k)} & \cdots & P_{1n}^{(k)} \\ P_{21}^{(k)} & P_{22}^{(k)} & \cdots & P_{2n}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{m1}^{(k)} & P_{m2}^{(k)} & \cdots & P_{mn}^{(k)} \end{bmatrix}$$

定义 7 将状态空间其中一个状态的状态概率进行加权求和, 得到该状态的预测概率, 即

$$P_i = \sum_{k=1}^m w_k \times p_i^{(k)}, i \in S \quad (11)$$

其中 w_k 为马尔可夫链的权重, 由下式算得

$$w_k = \frac{|r_k|}{\sum_{i=1}^m |r_k|} \quad (12)$$

式中 m 为计算所需的最大阶数, r_k 为样本数据序列的第 k 阶自相关系数。

定义 8 样本数据序列的第 k 阶自相关系数[12]

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sqrt{\sum_{i=1}^{n-k} (x_i - \bar{x})^2 \sum_{i=1}^{n-k} (x_{i+k} - \bar{x})^2}}, k = 1, \dots, 5 \quad (13)$$

其中 x_i 为第 i 天的样本数据; \bar{x} 为样本均值。

定义 9 设状态空间有 n 个状态[13], 若 $\varepsilon(k) \in (a_i, b_i), i = 1, 2, \dots, n$; 则称相对误差 $\varepsilon(k)$ 处于状态 S_i , 其中 a_i, b_i 分别为状态 S_i 的下界和上界。

最后将 $\max\{P_i, i \in S\}$ 所对应的状态作为第 n 天的相对误差的预测状态。则第 n 天的预测值为

$$\hat{Y}(k) = \hat{x}^{(0)}(k) \left[1 + \frac{1}{2}(b_i + a_i) \right] \quad (14)$$

3. 案例分析

3.1. GM(1,1)模型对非洲猪瘟事件百度搜索指数的预测和状态划分

以国内 2019 年 6 月 29 日至 7 月 18 日的非洲猪瘟事件的百度搜索指数为原始数据, 运用灰色 - 加权

马尔可夫链模型对 7 月 13 日至 7 月 18 日搜索指数进行预测。

以 6 月 29 日至 7 月 2 日的百度搜索指数为基础，建立灰色预测模型 GM(1,1)对 7 月 3 日的数据进行预测，且模型通过相对关联度检验；同理，以 6 月 30 日至 7 月 3 日的百度搜索指数为基础，可以得到 7 月 4 日的数据，以此类推，直至得到 7 月 18 日的百度搜索指数预测值。根据实际百度搜索指数和预测数据，计算相对误差。结果见表 1。

根据相对误差法，将马尔可夫链状态空间划分为 5 种状态，见表 2，按照表 2 的划分标准得出表 1 中相对误差值对应的状态。

Table 1. The actual Baidu search index, predicted Baidu Index and relative error status of African swine fever from April 4 to 18

表 1. 4 至 18 日非洲猪瘟实际百度搜索指数和预测百度指数及相对误差状态

日期	真实值	预测值	相对误差	状态
7.3	12,598.00	13,280.08	-0.0541	S ₄
7.4	13,782.00	13,341.06	0.0320	S ₂
7.5	14,256.00	14,546.83	-0.0204	S ₃
7.6	12,489.00	15,268.78	-0.2226	S ₅
7.7	13,536.00	12,297.54	0.0915	S ₁
7.8	13,509.00	12,697.29	0.0601	S ₂
7.9	14,002.00	14,215.04	-0.0152	S ₃
7.10	14,905.00	14,157.66	0.0501	S ₂
7.11	14,708.00	15,596.71	-0.0604	S ₄
7.12	17,070.00	15,248.86	0.1067	S ₁
7.13	16,350.00	17,911.72	-0.0955	S ₄
7.14	16,490.00	17,697.11	-0.0731	S ₄
7.15	17,807.00	16,059.73	0.0981	S ₁
7.16	17,793.00	18,407.33	-0.0345	S ₃
7.17	17,690.00	18,687.97	-0.0564	S ₄
7.18	20,417.00	17,646.74	0.1357	S ₁

Table 2. State division table

表 2. 状态划分表

状态	状态空间%
S ₁	(-23,-10)
S ₂	[-10,-5)
S ₃	[-5,0)
S ₄	[0,7)
S ₅	[7,14)

3.2. 加权马尔可夫链预测

根据表 1 中每个数据对应的状态，可以根据式(10)求出 1 到 5 步长的概率转移矩阵。

$$P^{(1)} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{3}{5} & \frac{1}{5} & 0 & \frac{1}{5} & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$P^{(2)} = \begin{bmatrix} 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$P^{(3)} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$P^{(4)} = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$P^{(5)} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

根据原始数据，运用式(12)和式(13)计算各阶的自相关系数和各阶权重，得到表 3。

Table 3. Autocorrelation coefficient and weight of each order
表 3. 各阶自相关系数及各阶权重

k	r_k	w_k
1	0.7838	0.291679
2	0.6862	0.255359
3	0.5682	0.211447
4	0.3757	0.139811
5	0.2733	0.101704

现在以 7 月 13 日非洲猪瘟事件的百度搜索指数预测为例进行说明。由 7 月 8 日至 7 月 12 日的概率转移矩阵以及表 3 中的数据对 7 月 13 日的百度搜索指数进行预测。可得 7 月 13 日相对误差状态预测表，见表 4。

Table 4. Prediction of relative error state on July 13
表 4. 7 月 13 日相对误差状态预测

日期	状态	权重	S_1	S_2	S_3	S_4	S_5
7.12	S_1	0.29	0	1/3	1/3	1/3	0
7.11	S_4	0.14	1/4	0	1/2	1/4	0
7.10	S_2	0.26	1/3	0	0	2/3	0
7.9	S_3	0.21	0	0	1/2	1/2	0
7.8	S_2	0.26	1/3	0	1/3	1/3	0
	P_i		0.2052	0.0972	0.3580	0.4933	0.0000

从表 4 可以观察到，7 月 13 日的相对误差处于 S_3 ，通过计算得该日期百度搜索指数预测值为 16,458.34。以此类推计算 14 日至 18 日的预测值，得表 5。

Table 5. Forecast results
表 5. 预测结果

日期	真实值	预测值	相对误差
7.13	16,350	16,458.341	-0.0066
7.14	16,490	16,369.827	0.0073
7.15	17,807	17,746.002	0.0034
7.16	17,793	17,947.147	-0.0087
7.17	17,690	17,286.372	0.0228
7.18	20,417	19,499.648	0.0449

根据预测结果，绘制预测结果图，见图 1。

根据表 1 及表 5 中的相对误差值，绘制 GM(1,1)模型和灰色 - 加权马尔可夫链模型的相对误差结果对比表，见表 6。

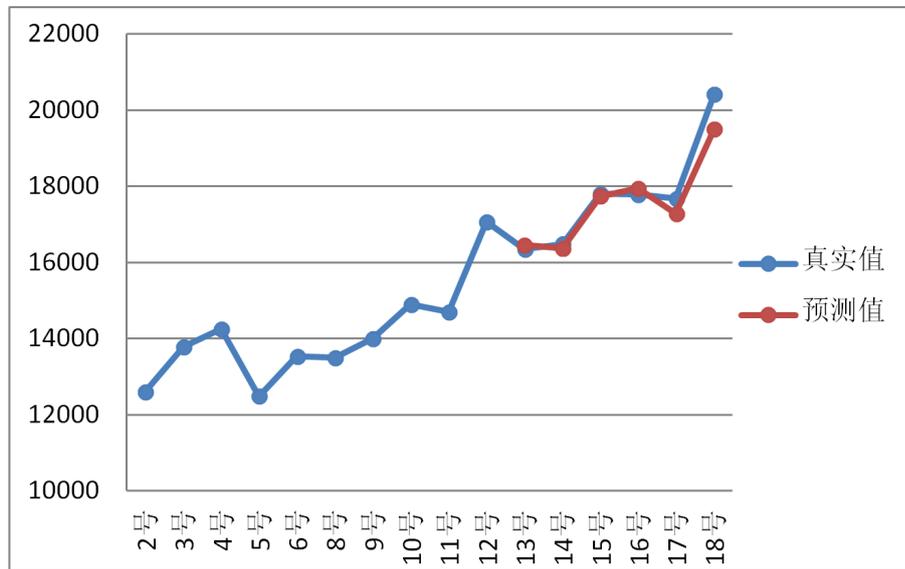


Figure 1. Forecast result chart

图 1. 预测结果图

Table 6. Comparison table of relative error results

表 6. 相对误差结果对比表

日期	GM(1,1)	灰色 - 加权马尔科夫链模型
7.13	-0.0955	-0.0066
7.14	-0.0731	0.0073
7.15	0.0981	0.0034
7.16	-0.0345	-0.0087
7.17	-0.0564	0.0228
7.18	0.1357	0.0449

4. 结论

1) 对于有一定波动性的数据来说, 通过传统的灰色预测模型 GM(1,1) 计算得到的百度搜索指数预测值相对误差较大, 而马尔可夫链对于波动性大的数据预测较为准确, 所以用加权马尔可夫链对模型进行优化, 结果表明, 构建的灰色 - 加权马尔可夫链模型能够有效地减小相对误差值, 提高模型的精度, 使预测结果更具有可信度。

2) 本文对突发公共卫生事件的未来趋势预测提供了一种预测结果较为精确的组合模型, 帮助了解突发公共卫生事件发展和影响范围, 描述突发公共卫生事件传播规律, 对及时辅助决策者做出科学合理的决策行为具有一定指导意义。

基金项目

河南科技大学大学生科研训练计划专项基金(项目编号: 2019203)。

参考文献

[1] 王林, 王可, 吴江. 社交媒体中突发公共卫生事件舆情传播与演变——以 2018 年疫苗事件为例[J]. 数据分析与

知识发现, 2019(4): 42-52.

- [2] 马天娇. 融媒时代突发公共卫生事件信息传播及风险沟通策略探讨[J]. 新闻研究导刊, 2018, 9(21): 85-86.
- [3] 张倩生, 谢柏林, 张新猛. 网络舆情突发事件的应急群决策方法[J]. 统计与决策, 2016(10): 38-41.
- [4] Galam, S. (2002) Minority Opinion Spreading in Random Geometry. *The European Physical Journal B*, **25**, 403-406. <https://doi.org/10.1140/epib/e20020045>
- [5] 靳松, 庄亚明. 基于 H7N9 的突发事件信息传播网络簇结构特性研究[J]. 情报杂志, 2013, 32(12): 12-17.
- [6] 洪巍, 吴林海, 吴祐昕. 食品安全网络舆情中的网络意见领袖[J]. 华南农业大学学报(社会科学版), 2014, 13(4): 101-110.
- [7] 滕文杰. 时间序列分析法在突发公共卫生事件网络舆情分析中的应用研究[J]. 中国卫生统计, 2014(6): 1071-1073.
- [8] 刘小洋, 何道兵. 基于突发公共事件的信息传播动力学模型与舆情演化研究[J], 计算机科学, 2019, 46(5): 320-326.
- [9] 李海晨, 张爱真. 基于灰色理论的科技文献增长规律研究——以微信舆情为例[J]. 情报探索, 2019(2): 7-12.
- [10] 林元烈. 应用随机过程[M]. 北京: 清华大学出版社, 2002.
- [11] 乔松珊, 张建军. 基于灰色马尔可夫修正模型的城市月用电需求预测[J]. 电力需求侧管理, 2013, 15(1): 11-14.
- [12] 蒋峰, 王宗耀, 张鹏. 基于灰色-加权马尔可夫链的光伏发电量预测[J]. 电力系统保护与控制, 2019, 47(15): 55-60.
- [13] 宋巧娜, 唐德善. 基于灰色马尔可夫模型的农业用水量预测[J]. 安徽农业科学, 2007, 36(6): 1788-1789.