

文章编号:1007-2985(2011)04-0035-04

2009 年加拿大 H1N1 疫情的 ARIMA 模型预测与分析*

曹 灿^{1,2}, 赵联文², 昌春艳², 刘 娟²

(1. 吉首大学数学与计算机科学学院, 湖南 吉首 416000; 2. 西南交通大学数学学院, 四川 成都 610031)

摘 要:2009 年初, 世界各地先后发生了甲型 H1N1 流感. 针对加拿大 2009 年疫情, 建立了恰当的 ARIMA 模型, 以实现每日 H1N1 疫情的预测. 经过实证分析, 预测的绝对误差在 11% 以内, 总的平均误差是 8.39%, 该模型成功地对加拿大 2009 年疫情进行了预测.

关键词:ARIMA 模型; H1N1 疫情; 预测; 加拿大; 时间序列分析

中图分类号:O212.1

文献标志码:B

甲型 H1N1 流感为急性呼吸道传染病, 其病原体是一种新型的甲型 H1N1 流感, 在人群中传播. 早期被称为猪流感. 2009 年 3 月, 墨西哥和美国先后发生甲型 H1N1 流感, 其病毒为 A 型流感病毒^[1]. 据国家卫生部通报, 截至 2010 年 1 月 10 日, 中国内地已有 124 764 例甲型 H1N1 流感确诊病例, 其中 744 例死亡. 为有效预防流感的爆发, 应用时间序列分析中的差分自回归移动平均模型(Autoregressive Integrated Moving Average, ARIMA)对全球 H1N1 流感最为严重的几个国家的疫情进行分析, 预测出加拿大的 H1N1 疫情变化情况, 以便对中国的疫情防治提供统计学方面的依据.

1 资料与方法

1.1 资料来源

资料来源于新华网由世界卫生组织发布的有关国家和地区甲型 H1N1 流疫情数据, 获取了 2009 年 5 月 5 日到 6 月 20 日加拿大 H1N1 流感确诊病例的人数^[2].

1.2 方法原理

对 5 月 5 日到 6 月 15 日的数据进行预测, 并用 6 月 16 日到 6 月 20 日加拿大 H1N1 流感确诊人数进行模型的检验.

20 世纪 60 年代, 美国学者 Box 和英国统计学者 Jenkins 提出了一整套关于时间序列分析、预测和控制的方法, 称为 Box-Jenkins 建模方法^[3] 或自回归移动平均(Auto Regression Moving Average, ARMA)模型. ARMA 模型中一种比较成熟的模型, 只适合平稳时间序列的短期预测. 对于非平稳序列, 需先对其进行差分, 将非平稳序列转化为平稳序列, 因此文中应用 ARIMA(p, d, q)模型, 其中 d 为差分阶数, p 为自回归阶数, q 为移动平均阶数^[4]. 模型的具体表达式为:

* 收稿日期:2011-03-26

基金项目:中央高校基本科研业务费专项资金资助项目(SWJTU09CX075)

作者简介:曹 灿(1979-), 男, 湖南衡阳人, 硕士生, 主要从事随机过程、统计学习理论研究

通讯作者:赵联文(1964-), 男, 四川巴中人, 西南交通大学数学学院教授, 硕士生导师, 主要从事随机过程、贝叶斯分析研究.

$$\begin{cases} \phi(B)(1-B)^d X_t = \theta(B)\epsilon_t, \\ E(\epsilon_t) = 0, \\ \text{Var}(\epsilon_t) = \sigma_\epsilon^2, \\ E(\epsilon_t \epsilon_s) = 0 \quad s \neq t, \\ E(X_s \epsilon_t) = 0 \quad \forall s < t, \end{cases} \quad (1)$$

其中 $\phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$ 和 $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ 之间无公共因子. 该研究在 SPSS16.0 中进行数据处理.

2 时间序列分析

2.1 数据预分析

ARIMA 建模的基本思路就是先通过 d 阶差分后变成平稳序列, 然后利用 ARMA 模型对差分后的平稳序列建模. 一个平稳的时间序列应符合以下 3 个条件: 均数不随时间变化; 方差不随时间变化; 自相关系数只与时间间隔有关^[5]. 实际上, 大多数用初始数据建立的时间序列都是不平稳的, 所以在建模之前应该先验证时间序列的平稳性, 并将不平稳序列转化为平稳序列. 常见的序列平滑的方法有差分法、季节性差分法、中心移动平均法、向前移动平均法等. ARIMA 模型采用差分法将时间序列平滑化.

2.2 模型建立和检验

通过 SPSS16.0 可得到时间序列的线性图(见图 1). 从线性图可以看出, 序列 $X(t)$ 具有一定的趋势性.

应用 SPSS16.0 可以得到时间序列 $X(t)$ 的样本自相关系数图(见图 2, 3).

首先对序列进行一阶差分, 将非平稳序列转化为平稳序列. 初步确定时间序列的模型为 ARIMA $(p, 1, q)$. 应用 SPSS16.0, ARIMA $(3, 1, 2)$ 模型的参数估计结果如表 1 所示. 从 t 统计量的显著性可以看出, 此模型的自回归模型系数不显著, 所以有必要对模型结构进行改进, 降低自回阶的阶数.

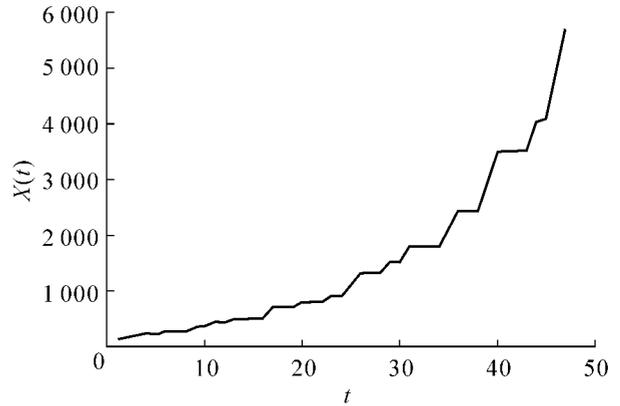


图 1 时间序列 $X(t)$ 的线性图

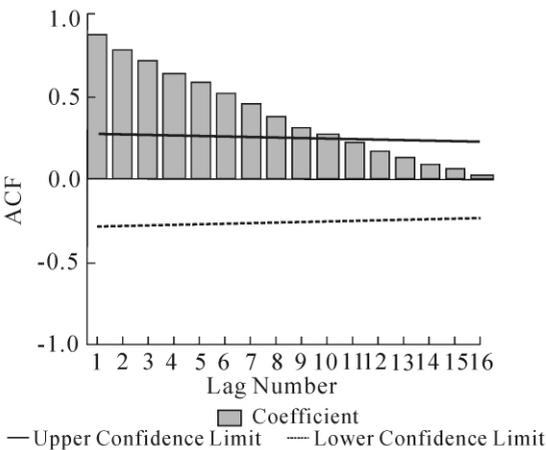


图 2 时间序列 $X(t)$ 的样本自相关系数图

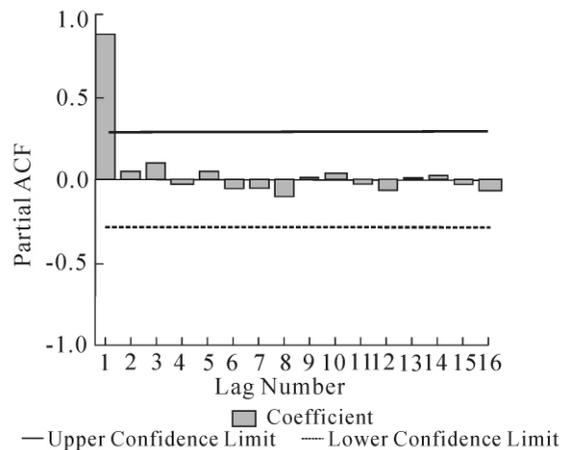


图 3 时间序列 $X(t)$ 的相关偏相关系数图

表 1 ARIMA $(3, 1, 2)$ 参数估计

		估计	标准差	t	显著性
AR	滞后 1	1.12	0.181	6.17	0.000
	滞后 2	-0.29	0.006	-45.50	0.000
	滞后 3	0.17	0.181	0.363	

续表

		估计	标准差	t	显著性
差分	1	1			
MA	滞后 1	1.21	0.28	4.31	0.000
	滞后 2	-0.23	0.25	-0.93	0.357

经改进后,指定模型的结构为 ARIMA(1,1,1).改进后的模型参数如表 2 所示.从表 2 可以看出,改进后此模型的自回归系数是显著的.

表 2 ARIMA(1,1,1)参数估计

		估计	标准差	t	显著性
AR	滞后 1	1.00	0.001	1 336	0.000
差分	1	1			
MA	滞后 1	0.99	0.04	22.39	0.000

改进后的模型的残差序列的自相关图形和偏相关图形如图 4 所示.图 4 中的 2 个图形都没有显著的趋势特征(拖尾或截尾),故可以初步判断这个改进模型是比较恰当的.图 5 描绘了实际观测序列和模型拟合序列的变化趋势,从 2 条曲线高度接近可以推断改进模型较为合理.

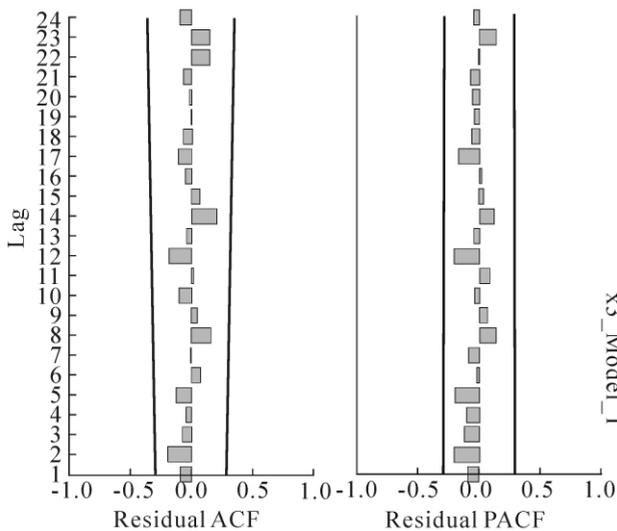


图 4 ARIMA 模型残差的相关函数图

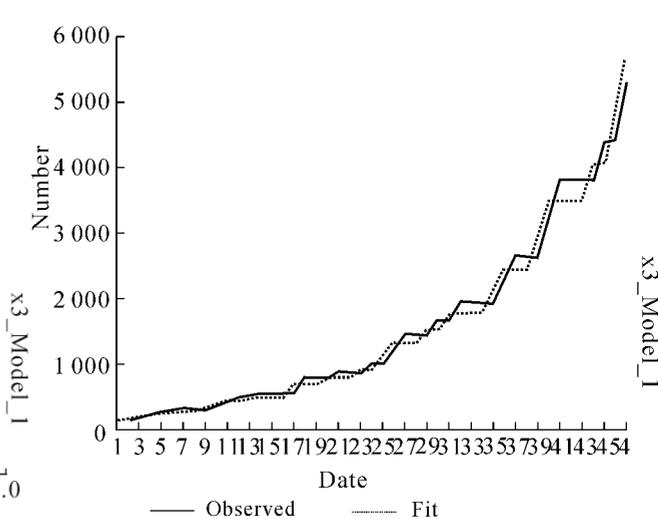


图 5 ARIMA 模型拟合图形

2.3 预测

利用 ARIMA(1,1,1)模型对加拿大 6 月 16 日到 6 月 20 日 H1N1 流感确诊人数进行预测,其结果见表 3.

表 3 加拿大 H1N1 确诊人数预测结果

时间	实际值	预测值	相对误差/%	时间	实际值	预测值	相对误差/%
6 月 16 日	3 520	3 878.4	10.1	6 月 19 日	4 905	4 408.1	10.1
6 月 17 日	4 049	3 894.4	3.8	6 月 20 日	5 710	5 290.5	7.3
6 月 18 日	4 095	4 419.5	7.9				

经计算,ARIMA(1,1,1)平均相对误差为 7.84%.预测的平均相对误差低于 5%的达 1 d(总预测时间为 5 d),预测的平均相对误差在 5%~10%的有 2 d.此模型的预测结果出现较大误差的主要原因是新华网公布的数据没有及时更新,从 2009 年 5 月 5 日到 6 月 20 日 46 d 中有 36 d 的数据是 2 d 或 2 d 以上相同.

3 结语

模型的不足:(1)统计数据的口径不一致.由于世界各国所在位置不同,存在一定的时差,因此数据统计的时间上有一定的差别.(2)部分数据没有及时更新,统计有误.加拿大连续几天的流感确诊人数没有变化,随后 1 d 人数突然变大,这主要是由于数据没有及时更新所致.(3)模型是通过加拿大 H1N1 流感确诊人数间接来预测世界各国的 H1N1 流感的变化情况,同时也可以借鉴该方法预测世界各国的具体情况,但其他国家的具体情况的预测方法还有待进一步研究.

H1N1 流感是近年来对人类影响最为严重的病毒之一,对全人类的生命安全造成了极大的威胁.对 H1N1 流感的发病趋势建立时间序列模型,可以较好地对全球 H1N1 流感的发病进行预测与监测.研究显示,ARIMA 模型能较好地用于 H1N1 流感的短期预测,根据预测数据有针对性地开展防治工作,有助于加强疾病预防控制工作.

参考文献:

- [1] 赵 研,王新春,陆 莹,等.甲型 H1N1 流感患者流行病学分析 [J]. 中国公共卫生,2010,26(8):1 080.
- [2] 新华网.2009 年 5 月 5 日有关国家和地区甲型 H1N1 流感疫情数据 [EB/OL]. http://news.xinhuanet.com/world/2009-05-05/content_11313205.htm,2009-05-05.
- [3] GEORGE E P BOX,GWILYM M JENKINS.时间序列分析预测与控制 [M].第 3 版.顾 岚,译.北京:中国统计出版社,1997.
- [4] 王 沁.时间序列分析及其应用 [M].成都:西南交通大学出版社,2008.
- [5] 杜 强,贾丽艳.SPSS 统计分析从入门到精通 [M].北京:人民邮电出版社,2009.

ARIMA Model Prediction and Analysis of H1N1 Epidemic Situation in Canada in 2009

CAO Can^{1,2}, ZHAO Lian-wen², CHANG Chun-yan², LIU Juan²

(1. School of Mathematics and Computer Science, Jishou University, Jishou 416000, Hunan China;

2. School of Mathematics, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: At the beginning of the year of 2009, H1N1 influenza outbreak took place throughout the world. According to the epidemic situation in 2009 in Canada, this paper establishes an appropriate ARIMA model in order to achieve a daily prediction of H1N1 pandemic. After a positive analysis, the absolute error of the prediction is less than 11%; while the total average error is 8.39%. This model successfully predicts the disease of Canada in 2009.

Key words: ARIMA model; H1N1 epidemic situation; prediction; Canada; time series analysis

(责任编辑 向阳洁)