

文章编号: 1007-4619(2003)05-0345-05

SARS 预测的 SI 模型和分段 SI 模型

李仲来¹, 崔恒建¹, 杨 华², 李小文^{2,3}

(1. 北京师范大学 数学系, 数据统计与分析中心, 北京 100875; 2. 北京师范大学遥感与 GIS 研究中心, 北京 100875; 3. 中国科学院 遥感应用研究所, 北京 100101)

摘要: 介绍及建立了对 SARS (Severe acute respiratory syndrome) 临床诊断累计病例预测的非线性增长模型: SI (Susceptible and infective) 模型和分段 SI 模型, 并对北京 SARS 累计病例进行了预测。分段 SI 模型转变点的 95% 的置信区间在 4 月 21 日、22 日和 23 日内, 表明我国政府采取了有力措施后, 4 月 24 日以后, SARS 病例的增长率发生显著变化。

关键词: 严重急性呼吸综合症 (SARS); SI 模型; 分段 SI 模型; 转变点

中图分类号: R181.2/O21 **文献标识码:** A

1 引言

2003 年 6 月份以来, 北京的 SARS (Severe acute respiratory syndrome) 疫情已经逐渐得到有效控制。世界卫生组织已于 24 日宣布解除了对北京旅游禁令, 并将北京从疫区名单中删除。至此, 中国已经没有再受到世界卫生组织旅游限制建议的省市, 中国内地已没有被其列入近期有当地传播名单的省市。在中国抵御传染病的历史上, 这场对抗 SARS 的斗争, 无论在人口、组织、经济、舆论规模, 还是科学、技术规模上都是空前的。这场斗争由于有了党和政府的正确领导, 有了全国人民的共同努力, 有了国内外科学家的忘我攻关和医务人员的舍身奉献, 中国的 SARS 防治工作已经取得了阶段性重大胜利。由于 SARS 的传染源、传染途径、传播机制等问题, 短期内无法查清, 这就需要在临床诊断病例资料的基础上, 应用统计分析方法, 研究 SARS 的流行趋势。鉴于 SARS 疫情已得到有效控制, 本文是在参考文献[1]的基础上所做的进一步的回顾性研究。

2 传染病传播的数学模型

早在 1904 年, Ross 首次提倡将数学方法应用于

疟疾等蚊虫传播疾病的控制。自 1911 年发表第一个传染病数学模型以来, 经过许多研究, 疟疾数学模型已从理论探讨进展到现场检验阶段。虽在实际应用方面还存在差距, 但在虫媒传染病的数学模型研究中, 疟疾的模型被认为是较好的。已建立的传染病模型还有艾滋病、麻疹、水痘、流行性感、血吸虫病等传染病的传播过程模型。我们不想从医学角度探讨每一种传染病的传播机理, 只是研究最简单的传染病的传播蔓延过程。

假设传染病人通过空气、食物等接触将病菌传播给健康人, 单位时间内一个病人可传染的人数为 k_0 。记时刻 t 的传染病病例是 $i(t)$, 则

$$i(t+\Delta t) - i(t) = k_0 i(t) \Delta t,$$

即

$$di/dt = k_0 i(t), \quad (1)$$

初始条件为

$$i|_{t=0} = i_0, \quad (2)$$

则方程(1)在(2)下的解是

$$i(t) = i_0 \exp(k_0 t). \quad (3)$$

该结果表明, 传染病病例将按指数规律无限增加, 这与实际情况是不符合的。实际上, 如果不考虑传染病流行期间出生和迁移人数, 一个地区的总人数可认为常数 $n = i(t) + s(t)$, $s(t)$ 为 t 时刻的健康人数。又单位时间内一个病人能传染的人数与当

收稿日期: 2003-07-06; 修订日期: 2003-07-10

基金项目: 国家自然科学基金主任基金项目“SARS 传播时空模型研究”(40341002)和 863 计划课题“SARS 流行病学资料的实时收集、分析和趋势预测”(2003AA208401)资助。

作者简介: 李仲来(1953—), 男, 教授, 1977 年毕业于北京师范大学数学系。主要从事预防医学、数学生态学方面的科研和教学工作。发表论文 140 多篇。

时健康人数成正比,比例系数为 k (称为传染系数), 则(1)式中的 k_0 应改为 $ks(t)$, 故(1)式可修改为

$$di/dt = ki(t)(n - i(t)), \quad (4)$$

初始条件仍为(2)式。用分离变量法可求得(4)式在(2)式下的解为

$$i(t) = n/[1 + (n/i_0 - 1)\exp(-knt)], \quad (5)$$

(5)式即为 SI (Susceptible and infective) 模型。由(5)式, 当 $t \rightarrow \infty$ 时, $i(t) \rightarrow n$, 这表明所有的人最终都要被传染。但由于被传染的病例或经治愈后免疫(不考虑治愈后再被传染的病, 如性病), 或死亡, 故病例最终应趋于零。模型(4)的进一步修改称为 SIR 模型。

3 北京 SARS 临床诊断累计病例的 SI 模型及预测

北京市的 SARS 临床诊断病例在全国占有相当高的比例。从 2003 年 4 月 20 日到 6 月 24 日, 北京市确诊的临床诊断病例占全国的 47.3%, 因此, 研究北京市临床诊断病例的变化与趋势, 对于研究全国的疫情具有相当重要的意义。我们利用传染病传播的 SI 模型进行研究。

t 从 3 月 1 日算起, 4 月 20 日至 24 日的临床诊断病例, 是截止到当日 20 时; 26 日之后的病例, 是截止到当日 10 时。因此, 4 月 24 日化为 $t = 54.42$, 4 月 26 日化为 $t = 56$, 依此类推, 所用数据来自国家卫生部网站。

将(5)式化为

$$i(t) = n/[1 + \exp(a - \pi)], \quad (6)$$

利用 SAS (Statistical analysis system) 的非线性模型进行拟合和预测。

根据北京的 SARS 临床诊断累计病例, 可以估计模型(6)中的参数 n , a , r 及预测见表 1。表 1 第 1 行为用 (2003-04-20-2003-05-01) 的 SARS 临床诊断累计病例, 估计模型(6)中的参数, 预测 SARS 临床诊断累计病例 n 的 95% 的置信区间, 对次日和 10 日的 SARS 临床诊断病例做出预测; 依此类推。在 6 月份, 每隔 5 日给出预测值, 其余略去。图 1 给出了 (2003-04-20-2003-06-24) 的 SI 模型。

4 北京 SARS 累计病例的分段 SI 模型

如果一个传染病 SI 模型增长特征的累计病例 $i(t)$ 在时刻 t_c 之前具有瞬时增长率 r_0 , 时刻 t_c 后它

转变为 r_1 , 则 SI 的动态模型可用如下的模型来描述:

$$di/dt = r_0 i(1 - i/n), \quad t \leq t_c,$$

$$di/dt = r_1 i(1 - i/n), \quad t_c \leq t,$$

其中 n 是总病例数。对于初始值 $i(t_0) = i_0$, 上述模型可解出

$$i = n/[1 + \{[n - i(t_0)]/i(t_0)\} \exp[-r_0(t - t_0)]], \quad t \leq t_c,$$

$$i = n/[1 + \{[n - i(t_c)]/i(t_c)\} \exp[-r_1(t - t_c)]], \quad t_c \leq t,$$

其中

$$i(t_c) = n/[1 + \{[n - i(t_0)]/i(t_0)\} \exp[-r_0(t_c - t_0)]],$$

它是由累计病例 $i(t)$ 在时刻 t_c 的连续性决定的。上式所确定的函数 $i(t)$ 的导数在时刻 t_c 是不连续的。它表明在时刻 t_c , 增长率发生了转折性变化, 致使 $i(t)$ 的增长速率在 t_c 有一个突然的改变。它也是病例增加过程的一个重要参数, 通常称 t_c 为变点 (Change point, 连续变点或一阶连续变点)^[2]。上述模型称为分段 SI 模型。

如何利用动态的临床诊断病例给出模型参数的估计是模型组建的重要环节。当变点不存在时, 它是一个普通的非线性模型, 其参数通常可由非线性最小二乘法 (如 Gauss-Newton 法等) 给出估计。当变点存在时, 分段拟合的方法无法保证 SI 模型在变点的连续性, 况且当变点未知时, 选取的人为性很强。我们在利用阶梯函数给出 SI 模型统一表达式的基础上, 给出了得到模型所有参数的最小二乘估计的算法^[3], 并用来描述分段 SI 累计病例动态模型。

记 $H(t - t_c)$ 为在 t_c 点具有单位跳跃函数

$$H(t - t_c) = 0, \quad t < t_c; \quad H(t - t_c) = 1, \quad t \geq t_c,$$

则分段 SI 模型可改写为

$$di/dt = [r_0 + (r_1 - r_0)(t - t_c)H(t - t_c)]i(1 - i/n),$$

对其积分, 模型可解出

$$i = n/[1 + \{[n - i(t_0)]/i(t_0)\} \exp\{-[r_0(t - t_0) + (r_1 - r_0)(t - t_c)H(t - t_c)]\}]$$

$$= n/[1 + \exp[\alpha + \beta + \gamma(t - t_c)H(t - t_c)]],$$

其中 $\alpha = \ln\{[n - i(t_0)]/i(t_0)\} + r_0 t_0$, $\beta = -r_0$, $\gamma = r_1 - r_0$ 。

模型的参数可以由 SARS 累计的病例, 利用非线性回归的 DUD (Does not use derivatives) 法^[4]来估计。对 4 月 18 日至 6 月 24 日按日的 SARS 临床诊断累计病例, 4 月 18 日、19 日和 20 日的临床诊断病例均认为是截止到当日 20 时。先求出非线性 SI 模型

$$i = 2521.4975/[1 + \exp(10.1082 - 0.1730t)],$$

残差平方和 $Q = 235.10$ 。

表1 北京SARS累计病例SI模型的参数估计及预测

Table 1 The parameters estimation and forecasting on SI models of the SARS cumulative cases in Beijing

时间	a	r	n	n 的 95% 的置信区间	次日(实测)	10日(实测)
05-01	9.9751	0.1706	2553.32	[1819, 3287]	102(83)	38(39)
05-02	10.1104	0.1739	2477.55	[2023, 2932]	94(105)	30(43)
05-03	10.0536	0.1726	2505.01	[2172, 2838]	89(62)	27(23)
05-04	10.2294	0.1767	2431.49	[2205, 2658]	78(94)	21(18)
05-05	10.1976	0.1760	2442.83	[2272, 2614]	71(63)	18(17)
05-06	10.2291	0.1767	2432.62	[2304, 2561]	63(89)	15(15)
05-07	10.1182	0.1742	2465.95	[2357, 2575]	58(87)	14(14)
05-08	9.9361	0.1703	2519.72	[2416, 2624]	55(41)	13(3)
05-09	9.8930	0.1693	2532.06	[2445, 2619]	49(50)	11(7)
05-10	9.8668	0.1688	2539.11	[2468, 2610]	44(38)	10(0)
05-11	9.8700	0.1688	2538.30	[2480, 2597]	39(39)	8(12)
05-12	9.8704	0.1688	2538.21	[2489, 2587]	34(43)	7(9)
05-13	9.8460	0.1683	2543.50	[2501, 2586]	29(23)	6(25)
05-14	9.8448	0.1683	2543.73	[2508, 2580]	25(18)	5(9)
05-15	9.8603	0.1686	2540.79	[2509, 2572]	22(17)	4(5)
05-16	9.8807	0.1690	2537.13	[2510, 2565]	19(15)	4(8)
05-17	9.9021	0.1695	2533.54	[2509, 2558]	16(14)	3(2)
05-18	9.9209	0.1698	2530.56	[2509, 2552]	13(3)	3(3)
05-19	9.9518	0.1704	2525.92	[2506, 2546]	11(7)	2(3)
05-20	9.9819	0.1710	2521.64	[2504, 2540]	9(0)	2(1)
05-21	10.0188	0.1718	2516.65	[2500, 2534]	8(12)	1(1)
05-22	10.0426	0.1722	2513.55	[2498, 2529]	7(9)	1(0)
05-23	10.0590	0.1725	2511.52	[2497, 2526]	6(25)	1(0)
05-24	10.0400	0.1722	2512.75	[2499, 2526]	5(9)	1(0)
05-25	10.0372	0.1721	2514.09	[2501, 2527]	5(5)	1(0)
05-26	10.0256	0.1719	2515.42	[2503, 2527]	4(8)	1(0)
05-27	10.0100	0.1716	2517.03	[2505, 2528]	3(2)	1(1)
05-28	9.9993	0.1714	2518.19	[2508, 2529]	3(3)	0(-1)
05-29	9.9888	0.1712	2519.42	[2509, 2530]	2(3)	0(0)
05-30	9.9788	0.1710	2520.46	[2511, 2530]	2(1)	0(0)
05-31	9.9706	0.1708	2521.29	[2512, 2531]	2(1)	0(1)
06-05	9.9503	0.1705	2523.29	[2516, 2531]	1(0)	0(-1)
06-10	9.9465	0.1704	2523.64	[2518, 2530]	0(1)	0(0)
06-15	9.9469	0.1704	2523.61	[2518, 2529]	0(-1)	0(0)
06-20	9.9509	0.1705	2523.26	[2519, 2528]	0(0)	0(0)
06-24	9.9537	0.1705	2523.02	[2519, 2527]	0(0)	0(0)

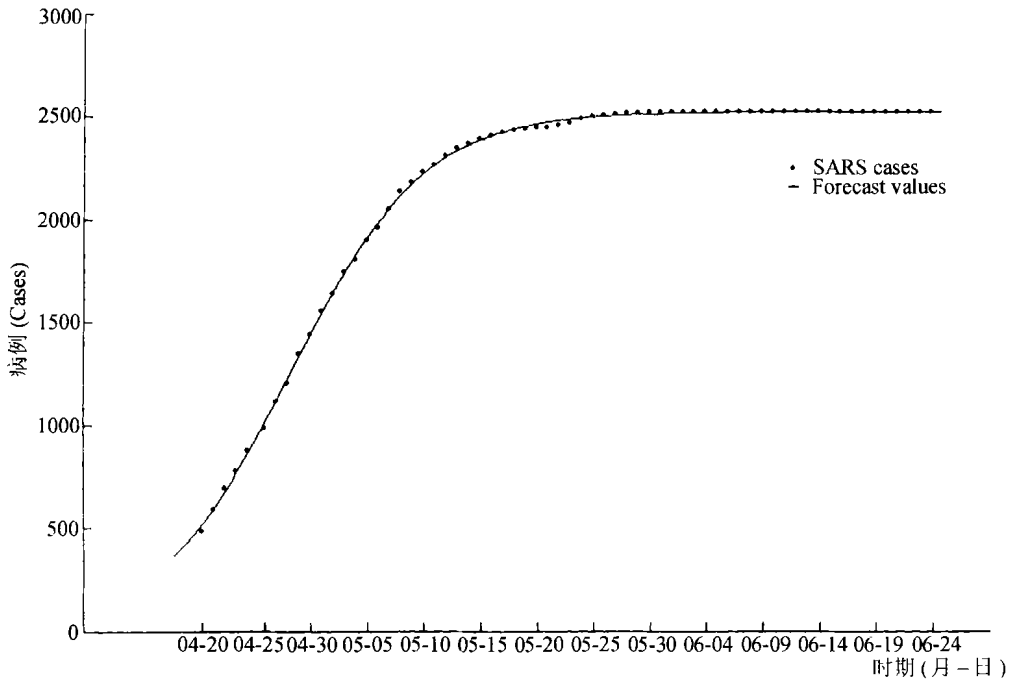


图 1 北京 SARS 累计病例的 SI 模型

Fig.1 The SI model of SARS cumulative cases in Beijing

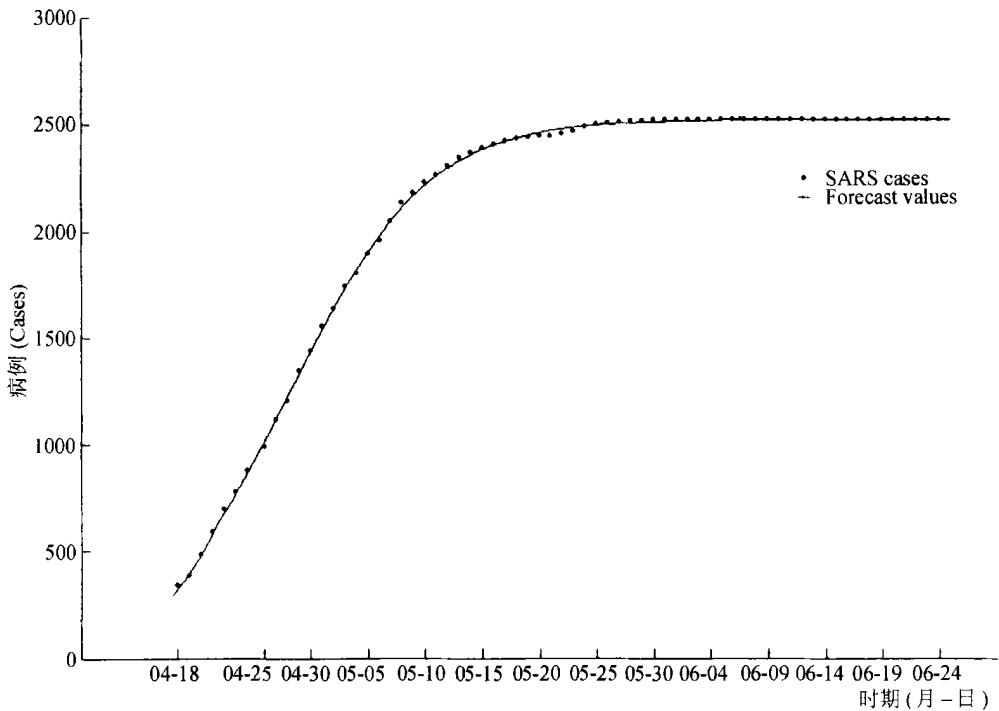


图 2 北京分段 SARS 累计病例的 SI 模型

Fig. 2 The piecewise SI model of SARS cumulative cases in Beijing

分段 SI 模型(图 2)为

$$i = 2523.7678 / \{1 + \exp[13.3327 - 0.2359t + 0.0667(t - 51.8963)H(t - 51.8963)]\},$$

$t = 48.42, 49.42, \dots, 54.42, 56, 57, \dots, 115$ 。

即 $i = 2523.7678 / [1 + \exp(13.3327 - 0.2359t)],$

$$t < 51.8963;$$

$$i = 2523.7678 / [1 + \exp(9.8730 - 0.1690t)],$$

$$t \geq 51.8963,$$

残差平方和 $Q = 129.36$, 转变点 t_c 的 95% 的置信区间是 $[51.01, 52.79]$ 。

5 结论与讨论

在 SARS 流行时,文献[1]所做的预测是前瞻性研究;SARS 流行停止后,所做的预测是回顾性研究。本文与文献[1]的区别主要在于此。由于 SARS 的传染源、传染途径、传播机制等问题,短期内无法查清,数学与 SARS 的主要关系之一就是预测 SARS 临床诊断病例的上限和研究 SARS 的流行趋势。

对北京 SARS 临床诊断累计病例的 SI 模型,其预测临床诊断病例的结果在 5 月 25 日以前偏向于保守,参数 n , a , r 随时间的变化逐渐趋于稳定。5 月 1 日前后,预测北京 SARS 临床诊断累计病例的上限是 3287 例,从现在来看,当时的这一预测也是合理的,低于文献[1]的 SARS 临床诊断累计病例的上限 3362 例。北京的 SARS 临床诊断累计病例为 2521 例,均落在表 1 总病例 95% 的置信区间内。当然,最终所获得的总临床诊断病例是在强有力的控制措施下得到的,否则,模型的参数可能会有较大的变化。

由于瞒报,缓报和漏报,4 月 18 日以前的病例是不可信的。利用表 1 最后一行(2003-04-20-2003-06-24)的 SI 模型,可以反推出 4 月 20 日之前的理论病例。如 4 月 5 日、10 日、15 日的病例依次为 49 例、113 例和 250 例。

与文献[1]的 SARS 临床诊断累计病例拟合结果比较,在前一段,文献[1]的拟合效果优于 SI 模型,这是因为,在没有后期病例的时候,拟合的优劣就显得非常重要。从 SARS 临床诊断累计预测的病

例看,考虑次日预测与实测病例之差的绝对值之和,本文为 255 例,低于文献[1]近百例;考虑未来 10 日预测与实测病例之差的绝对值之和,本文为 89 例,低于文献[1]200 多例。

由于没有 4 月 18 日以前每日的 SARS 临床诊断累计病例,我们求出了 4 月 18 日后的非线性 SI 模型及分段非线性 SI 模型,分段 SI 模型的残差平方和 Q 低于 SI 模型 Q 的 105.74,当然,我们是以多估计两个参数为代价的。如果有 4 月 18 日以前的病例,分段模拟的效果会更好。从转变点 t_c 的 95% 的置信区间是在 4 月 21 日、22 日和 23 日内,表明 4 月 20 日左右,中国政府采取了有力措施后,4 月 24 日以后,SARS 病例的增长率明显发生变化。

致谢 文中图像由施宏绘制,特此致谢。

参考文献 (References)

- [1] Cui Hengjian, Li Zhonglai, Yang Hua, Li Xiaowen. Nonlinear Regression in SARS Forecasting [J]. *J. of remote sensing*, 2003, 7(4):245-250. [崔恒建,李仲来,杨华,李小文,SARS 疫情预测预报中的分段非线性回归方法[J]. 遥感学报,2003,7(4):245-250.]
- [2] Krishnaiah P. R., Miao B. Q., Review about estimation of change points[M]. In: Krishnaiah P. R., Rao C. R. editors, *Handbook of statistics*. Amsterdam: Elsevier Science Publishers B. V., 1988, 7: 375-402.
- [3] Li Zhonglai, Fitting nonlinear models of ecology with DUD algorithm [J]. *Chin. J. of Ecology*, 1997, 16(2): 73-77. [李仲来. DUD 法拟合生态学中的非线性模型[J]. 生态学杂志,1997,16(2): 73-77.]
- [4] Ralston M. L., Jennrich R. I., DUD, a derivative-free algorithm for nonlinear least squares[J]. *Technometrics*, 1979, 20(1):7-14.

SI Models and Piecewise SI Model on SARS Forecasting

LI Zhong-lai¹, CUI Heng-jian¹, YANG Hua², LI Xiao-wen^{2,3}

(1. Department of Mathematics, Statistical Data Analysis Center, Beijing Normal University, Beijing 100875, China;

2. Remote Sensing and GIS Research Center of Beijing Normal University, Beijing 100875, China;

3. Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101, China)

Abstract: This paper introduces and sets up some kinds of nonlinear growth models, the SI(Susceptible and infective) model and piecewise SI model, for forecasting clinical diagnose cumulative SARS (Severe acute respiratory syndrome) cases in Beijing. The 95% confidence interval of the time change point on piecewise SI model is made well which includes April 21, 22 and 23. It means some control policies in Beijing at the end of this April 24 played an important role for anti-spreading of SARS, after change of increase rate for SARS cases is quite significant.

Key words: Severe Acute Respiratory Syndrome(SARS); susceptible and infective model; piecewise susceptible and infective model; change point