

工程与应用

基于元胞自动机的传染病传播模型研究

余 雷¹, 薛惠锋¹, 高晓燕², 李 刚¹YU Lei¹, XUE Hui-feng¹, GAO Xiao-yan², LI Gang¹

1.西北工业大学 资源与环境信息化工程研究所, 西安 710072

2.第四军医大学 护理系, 西安 710032

1.Resource and Entertainment Institute, Northwestern Polytechnical University, Xi 'an 710072, China

2.Nurse Department, Fourth Military Medical University, Xi 'an 710032, China

E-mail: yul525@163.com

YU Lei, XUE Hui-feng, GAO Xiao-yan, et al. Epidemic spread model based on cellular automata. Computer Engineering and Applications, 2007, 43(2): 196-198.

Abstract: Based on the theory of complexity adaptive system, we build a cellular automata model to simulate the complex process of epidemic spread. Based on this model, we simulate the spread process of SARS successively. Then several factors, such as move of people, whether being treated in time, which have great effects on the propagation of epidemic are analyzed, and whose effects on epidemic are studied. Some suggestion to deal with this kind of problem is given finally.

Key words: epidemic model; cellular automata; complexity adaptive system; computer simulation

摘 要: 从复杂适应系统的观点, 通过建立元胞自动机模型的方法模拟疾病传播这个复杂的过程. 并对 SARS 地传播过程成功地进行了模拟. 同时以此为基础针对可能对传染病产生影响的几种因素作了具体地考察, 如人员的移动、及时就医等, 考察这些因素对控制传染病达到稳定的具体影响, 并给出一些控制这类问题的建议.

关键词: 传染病模型; 元胞自动机; 复杂适应系统; 计算机模拟

文章编号: 1002-8331(2007)02-0196-03 文献标识码: A 中图分类号: N941; TP273

1 引言

人类历史上曾多次受到危害极其严重的传染病的威胁。其实对传染病的描述和预测是人们由来已久的研究课题。从 20 世纪的四五十年代开始, 以微分方程为主的决定论模型逐渐受到重视, 一直到现在仍然具有非常重要的学术地位, 其中最有影响的是 SIR 和 SIS 模型^[1]。这些模型在应用中往往能给出与实际统计结果符合得相当不错的结果, 但是大多基于微分动力系统, 一般计算繁杂, 并且方程的解对初始条件极为敏感, 不能很好处理实际过程的突发和随机事件。

传染疾病的传播过程中充满了偶然因素的影响, 它的传播显然是一种复杂现象。这是因为传染病传播的载体——人和人之间的交流、接触、联系所形成的系统是复杂的。易感染、有传染性的个体, 全都是有主动性的个体, 它们的行为方式对于病情的发展具有影响, 同时也会随外界情况的变化产生适应性的改变, 这也造成了要研究的问题的复杂性^[2-5]。

对传染病爆发的干预和控制措施对于传染病发展的影响过程更为复杂。可能根据传染病发展的情况采取一些控制措施, 影响传染病发展的具有主动性的个体对于控制措施这一外界因素会产生适应性的变化。针对不同的传染病, 不同的措施效果可能不同, 如何来确定对于特定类型的传染病, 哪种控制措施的效果最好, 找到答案并不容易, 但无疑是非常重要的。解

决了这些问题, 有助于合理地制定决策: 如何及时采取措施能够有效地控制传染病的爆发, 或者具体在传染病发展的不同阶段采取哪些措施会比较有效。

元胞自动机是描述自然界复杂现象的常用工具, 最初被用于模拟生命系统特有的自复制现象。它是时间和空间都离散的动力模型。20 世纪 40 年代初, 现代计算机的创始人之一冯·诺依曼开始思考描述生物自我繁衍的逻辑形式, 提出了元胞自动机的概念。但受制于当时计算机能力的限制, 直到 20 世纪 80 年代, S.Wolfram^[6]才对元胞自动机进行了全面地研究。将其动力学行为分为 4 大类: 平稳型、周期型、混沌型、复杂型, 其中复杂型为出现复杂的局部结构, 由于其演化可以表现出极其复杂的形态, 所以经常用于复杂系统地建模与模拟。

正是鉴于元胞自动机对复杂问题的建模能力, 该文尝试利用元胞自动机建立流行病传播模型来模拟流行病传播这一复杂现象, 做一些在现实世界中通过其它方法所不能做的实验, 将结果进行对比分析。建立能够模拟现实的传染病发展过程的计算机模型, 通过改变参数情况, 观察哪些因素对于特定传染病爆发流行最重要, 采取哪类措施能够更有效地控制, 何时采取效果最好。该模型可以省去一般微分方程模型的复杂计算过程, 并且比传统微分方程模型更易于直观分析, 利用其模拟结果可以更好地为流行病防治决策提供支持^[7-10]。

2 元胞自动机

按照元胞自动机的定义, 用数学符号来表示, 元胞自动机可以视为由一个元胞空间和定义于该空间的变换函数所组成^[11]。元胞自动机用形式语言的方式来描述, 可以用一个四元组表示:

$$A=(L_d, S, N, f)$$

在这里 A 代表一个元胞自动机系统, L_d 代表一个规则划分的网格空间, 每个网格单元 (cell) 就是一个元胞; L 通常为二维或三维空间, 但理论上可以是一个任意正整数维的规则空间。 S 代表一个离散的有限集合, 用来表示各个元胞的状态; N 代表元胞的邻居集合, 对于任何元胞的邻居集合 $N \subset L$, 设邻居集合内元素数目为 n , 那么, N 可以表示为一个所有邻域内元胞的组合 (包括中心元胞), 即包含 n 个不同元胞状态的一个空间向量, 记为:

$$N=(s_1, s_2, s_3, \dots, s_n)$$

n 是元胞的邻居个数, $S \subset Z$ (整数集合), $i=(1, 2, \dots, n)$

f 表示一个将 S_n 映射到 S 上的一个局部转换函数, 即根据 t 时刻某个元胞的所有邻居的状态组合来确定 $t+1$ 时刻该元胞的状态值。 f 通常又被称为状态转换函数, 或局部规则。

因此, 由上面公式可知, 元胞自动机最基本的组成是元胞 (cell)、元胞空间 (lattice)、邻居 (neighbors) 及局部规则 (local rules) 4 部分。简单来讲, 元胞自动机可以视为由一个元胞空间和定义于该空间的变换函数所组成^[12-13]。

3 建模

本研究模型构架如下:

(1) 元胞: 某特定空间的所有人群;

(2) 元胞空间: $N=n \times n$ 的二维空间;

(3) 邻居形式: 邻居半径为 1 的 Moore 型;

(4) 元胞状态空间: 设元胞的状态变量为 $S_{i,j}(t)$, 表示第 i 行 j 列元胞在时刻的状态, 根据流行病学的研究方法, 元胞的状态空间 (S) 有 3 种不同取值, 即 $S=\{0, 1, 2\}$, 分别对应流行病传播过程中个体的 3 种状态:

$S_{i,j}^t=0$: 易感状态, 即个体未被传染, 并且没有免疫;

$S_{i,j}^t=1$: 发病期状态, 即个体已被传染, 此时个体具有传染性;

$S_{i,j}^t=2$: 免疫期状态, 个体退出患病状态, 保留较高免疫。

另外, 对每个元胞引入时间参数 $t_n(S_{i,j}^t)$ 和 $t_m(S_{i,j}^t)$, $t_n(S_{i,j}^t)$ 代表个体的发病时间, $t_m(S_{i,j}^t)$ 代表个体的免疫持续时间, t_n 为发病时间的最大值, t_m 为个体免疫时间的最大值。

(5) 元胞演化规则: 将所要进行模拟的二维空间进行均匀的网格划分, 设定流行病特性参数。置所有元胞初始状态 $S=0$, 将病原体状态设为 1。从 0 时刻开始, 在每个时间步对空间内所有元胞进行扫描, 并按以下规则进行元胞状态更新:

当 $S_{i,j}^t=0$ 时, 计算 (i, j) 处元胞的邻接矩阵 $M=$

$$\begin{bmatrix} S_{i-1,j-1}^t & S_{i-1,j}^t & S_{i-1,j+1}^t \\ S_{i,j-1}^t & S_{i,j}^t & S_{i,j+1}^t \\ S_{i+1,j-1}^t & S_{i+1,j}^t & S_{i+1,j+1}^t \end{bmatrix}, \text{ 而矩阵 } N=\begin{bmatrix} C_{i-1,j-1}^t & C_{i-1,j}^t & C_{i-1,j+1}^t \\ C_{i,j-1}^t & C_{i,j}^t & C_{i,j+1}^t \\ C_{i+1,j-1}^t & C_{i+1,j}^t & C_{i+1,j+1}^t \end{bmatrix} \text{ 为}$$

矩阵 M 中值为 1 的元素。

计算 (i, j) 处元胞可被传染概率

$$p=\frac{k^* (C_{i-1,j}^t+C_{i,j-1}^t+C_{i,j+1}^t+C_{i+1,j}^t)}{4}+\frac{l^* (C_{i-1,j-1}^t+C_{i-1,j+1}^t+C_{i+1,j-1}^t+C_{i+1,j+1}^t)}{4}$$

其中, k 为该处元胞的上、下、左、右 4 处元胞的影响值, l 为左上、左下、右上、右下 4 处元胞的影响值, 很显然 $k>l$ 。此时该处元胞依概率 p 由状态 $S_{i,j}^t=0$ 转为 $S_{i,j}^t=1$, 同时个体的发病时间 $t_n(S_{i,j}^t)=t_n(S_{i,j}^t)+1$;

当 $S_{i,j}^t=1$ 时, 判断 $t_n(S_{i,j}^t)$ 的值。当 $t_n(S_{i,j}^t)>t_n$ 时, $S_{i,j}^t=2$, 同时 $t_n(S_{i,j}^t)=0$, $t_m(S_{i,j}^t)=t_m(S_{i,j}^t)+1$, 否则 $S_{i,j}^t=1$, $t_n(S_{i,j}^t)=t_n(S_{i,j}^t)+1$;

当 $S_{i,j}^t=2$ 时, 判断 $t_m(S_{i,j}^t)$ 的值。当 $t_m(S_{i,j}^t)>t_m$ 时, $S_{i,j}^t=0$, 同时 $t_m(S_{i,j}^t)=0$, 否则 $S_{i,j}^t=2$, $t_m(S_{i,j}^t)=t_m(S_{i,j}^t)+1$;

(6) 元胞随机行走运行办法: 以二维元胞自动机模型为例, 设置元胞每次随机行走的最大步长, 对所有元胞产生一个随机扫描顺序, 按此随机顺序扫描元胞。对扫描到的处的元胞生成两个独立的随机数 dx 和 dy ($|dx| \leq L$, $|dy| \leq L$), 然后将 (i, j) 处的元胞与 $(i+dx, j+dy)$ 处元胞进行交换^[14, 15]。

4 模型的模拟及其分析

通过以上所描述的方法建立模型以后, 进行了几组实验并对实验结果进行分析, 希望能在成功模拟北京 SARS 疫情的基础上, 观察不同因素对传染病疫情的影响, 从而发现对于具有特定属性的传染病的及时有效的控制途径。

4.1 SARS 疫情的模拟

根据以上所描述的元胞自动机模型, 对北京 2003 年 4 月 21 日至 2003 年 7 月 1 日 SARS 疫情进行模拟。具体设定值为: $n=100$, $k=0.5$, $l=0.1$, 发病持续时间 t_n 为 5 d, 由于至今无 SARS 的愈后复发报导, 所以可取免疫持续时间 t_m 为 300 d。在图 1 中横坐标为天数, 纵坐标为每天累计发病人数, 可以看出, 模拟结果与实际数据吻合得很好。

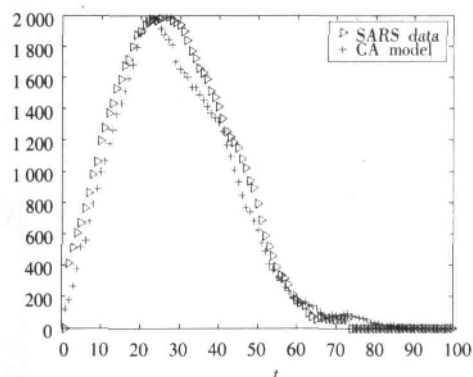


图 1 北京市 SARS 疫情模拟

4.2 人员移动的影响

下面研究人员移动对传染病疫情的影响。具体设定值与上面 SARS 模型相同, 只是在演化过程的每一步首先进行一定比例人员的随机移动。

图 2 不同比例人员移动传染病疫情图, 横坐标为时间, 纵坐标为每日传染病发病人数, 从图中可以看出, 随着人员移动

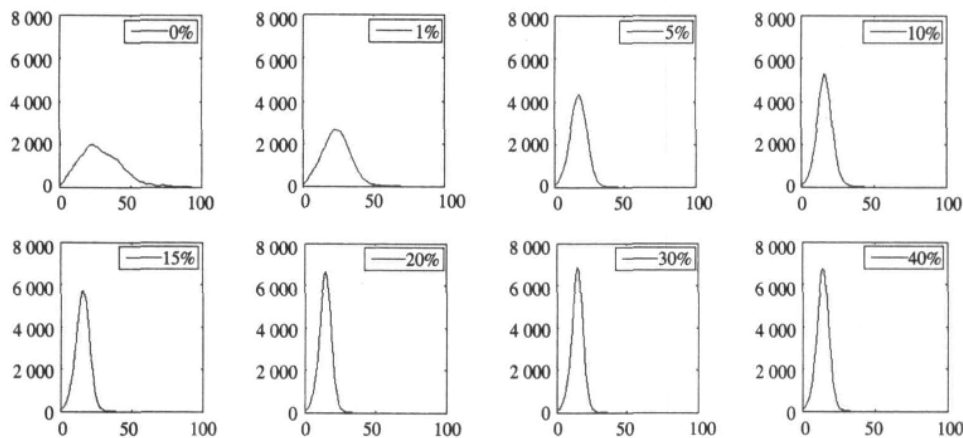


图2 不同比例人员移动传染病疫情图

数目的增加,每日发病人数也显著增加,但当比例增大到一定值时,发病人数达到了一个相对的稳定值。图3即为累计发病人数最大值随人员移动比例变化折线图,从图中可以看出当人员移动的比例逐步增大时,累计发病人数最大值也逐步增大;当比例增大到20%以后,发病人数达到一个相对的稳定值。

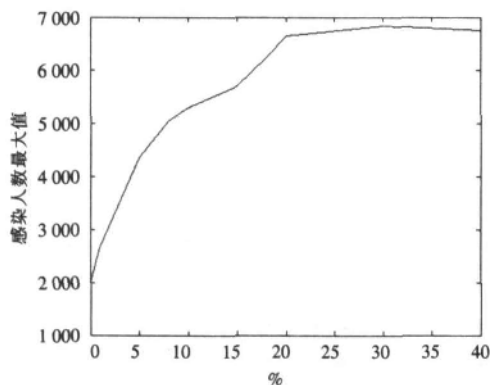


图3 不同移动比例下的感染人数最大值

这种结果为减少人员移动,对病人进行隔离的政策提供了支持,对于像SARS这样的恶性传染病,在发病初期进行人员隔离会收到很好的效果。

4.3 及时就医的影响

现在来看及时就医对于控制病情的重要性。从图4可以看出,在其它值不变,仅仅将 t_1 变为3d以后,发病人数显著减少,由最大值1983人骤减到348人。

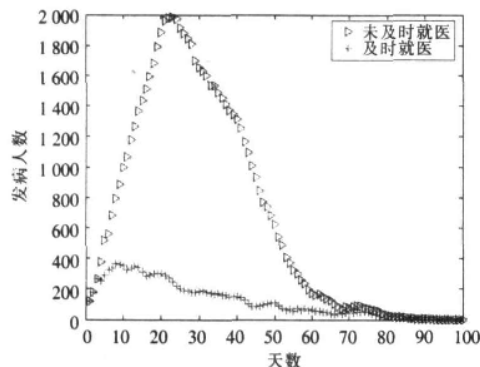


图4 及时就医模拟图

这种对比结果,为及时披露信息,以及促使患者能够尽快就医的政策提供了支持。从现实经验来看,如果人们对发病症状比较了解,对病情的严重性认识比较充分,通常会更快地主动就医。因此如果能够采取有效的措施督促有发病状况的人及

时到医院接受治疗,能比较有效地减少最终感染者的人数。当然为了保证发病者尽快就医,除了宣传和督促,还应结合当地医疗条件,尽量增加就医的方便性^[16-18]。

5 结论

(1) 本文从复杂适应系统的观点,尝试通过建立计算机模型的方法模拟疾病传播这个复杂的过程。在对SARS的流行病传播过程的模拟过程中,得到了与实际数据相一致的结果,说明该模型是可行和有效的。

(2) 本文通过该模型模拟的一些实验结果进行了简单的对比分析。具体来说是针对可能对SARS产生影响的几种因素作了具体地考察,进而比较了几种可能的控制措施的效果。发现尽量缩短发病到就医的时间间隔、及时地采取隔离措施都会比较有效地遏制类似于SARS这种传染病的发展和最终影响。

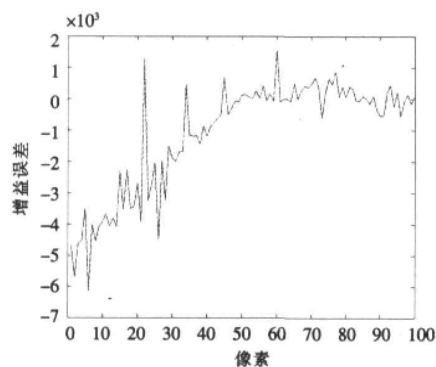
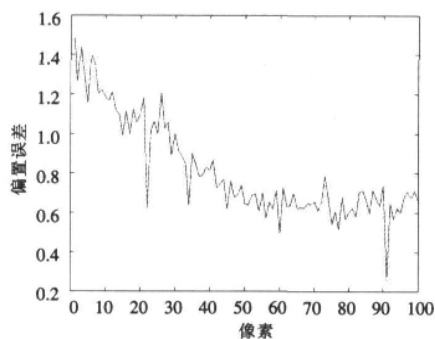
(3) 通过对模型的不断完善,希望模型能够更为有效地帮助了解更为一般的传染病传播蔓延过程,这样会尽快了解一种新出现的传染病。在此基础上,可以通过模型找到抑制病情的更为快速有效的途径,从而减少损失。

总之通过对模拟的初步结果表明,元胞自动机作为研究系统复杂性的一项重要工具,在疾病传播的复杂性研究中将展现出其巨大的潜能。(收稿日期:2006年9月)

参考文献:

- [1] Anderson R M, May R M. Infections diseases of humans[M]. Oxford: Oxford University Press, 1991.
- [2] 约翰·霍兰. 隐秩序[M]. 周晓牧, 韩晖, 译. 上海: 上海科技教育出版社, 2000.
- [3] 钱学森, 于景元, 戴汝为. 一个科学新领域—开放的复杂巨系统及其方法论[J]. 自然杂志, 1990, 13(1).
- [4] 许国志. 系统科学[M]. 上海: 上海科技教育出版社, 2000.
- [5] 迈克尔·沃尔德罗普. 复杂[M]. 陈玲, 译. 北京: 北京三联出版社, 1997.
- [6] Wolfman S. Theory and application of cellular automata[M]. Singapore: World Scientific, 1986.
- [7] Johansen A. A simple model of recurrent epidemics[J]. Journal of Theoretical Biology, 1996, 178(1): 45-51.
- [8] Srakoulis G C, Karafyllidis I, Thanailakis A. A cellular automaton model for the effects of population movement and vaccination on epidemic propagation[J]. Ecological Modelling, 2000, 133: 209-223.

(下转 237 页)

(a) 增益值误差 (均方根误差 2.1×10^{-3})

(b) 偏置值误差 (均方根误差 0.85)

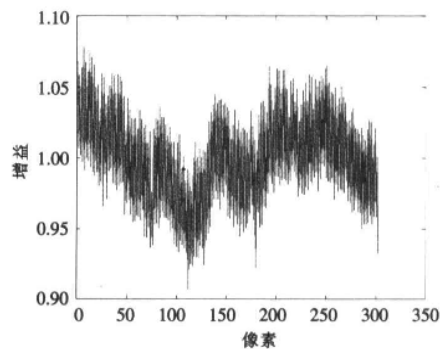
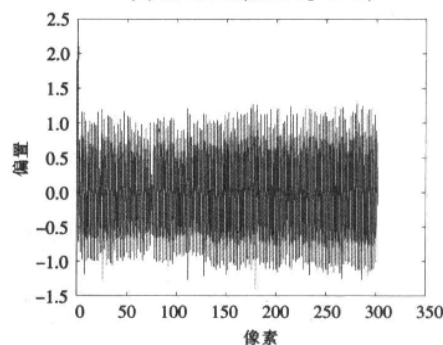
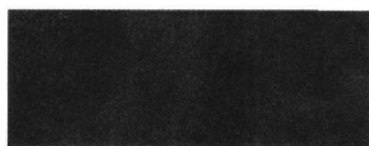
图1 模拟数据真实校正参数和模型结果
误差模型相对误差小于 1%(a) 增益曲线 (权重 $w_g = n/10$)(b) 偏置曲线 (权重 $w_b = n/100$)

图2 增益偏置参数 (截取其中 300 列)



(a) 校正前



(b) 校正后

图3 对图像的校正 (局部, 经过拉伸)

(1) 得到比较准确反映 CCD 性质的相对增益偏置参数, 既校正了图像的条纹噪声, 又没有对图像细节纹理造成破坏, 也没有引进其它噪声。

(2) 得到的参数比较稳定, 可以用于同一 CCD 在一段时间内成像的图像。

(3) 样本选择范围大, 提高了模型的精度和可靠性。

(4) 可扩展性强。对于 CCD 探元其它性质的差异, 可以观察差异的特性, 修改约束条件, 甚至重建基本方程进行描述, 得到一个新的方程组。

不足之处是模型的计算量比较大, 对样本选取的要求高, 以及在 CCD 有对大的区域性差异纠正效果不明显。但这些都可以通过修改模型来改进。实践证明, 这个模型还有很多改进的空间, 十分具有发展前途。(收稿日期: 2006 年 5 月)

参考文献:

- [1] Barker J L, Abrams R B, Ball D L, et al. Radiometric calibration and processing procedure for reflective bands on Landsat-4 proflight thematic mapper[C]//Proc Landsat-4 Early Results Symp, NASA Publication, 1983.
- [2] Desachy J. Investigation of Landsat-4 TM line-to-line and band-to-band registration and relative detector calibration[J]. PE & RS 1985, 51(9): 1291-1298.
- [3] 陈宝林. 最优化理论与算法[M]. 北京: 清华大学出版社, 2002.
- [4] 陈希孺. 线性模型参数的估计理论[M]. 北京: 科学出版社, 1985.
- [5] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003.
- [6] 郑婉勤, 章文毅, 陈甫, 等. 一种新的相对辐射校正统计模型[J]. 遥感学报, 2006, 10(4): 685-692.
- [7] 郑婉勤, 章文毅, 陈甫, 等. 一种新的相对辐射校正统计模型[J]. 遥感学报, 2006, 10(4): 685-692.
- [8] 郑婉勤, 章文毅, 陈甫, 等. 一种新的相对辐射校正统计模型[J]. 遥感学报, 2006, 10(4): 685-692.
- [9] Kleczkowski A, Grenfell B T. Mean-field-type equations for spread of epidemics: the "small world" model[J]. Physica A, 1999, 274(1): 355-360.
- [10] Arcangelis L, Hermann H J. Self-organized criticality on small world networks[J]. Physica A, 2002, 308(1): 545-549.
- [11] Amoroso S, Patt Y. Decision procedures for surjectivity and injectivity of parallel maps for tessellation structures[J]. J Computer System Sci, 1972, 6: 448-464.
- [12] 周成虎, 孙战利, 谢一春. 地理元胞自动机研究[M]. 北京: 科学出版社, 1999.
- [13] Chopard B, Droz M. 物理系统的元胞自动机模型[M]. 祝玉学, 赵学龙, 译. 北京: 清华大学出版社, 2003.
- [14] Martinsa M L, Ceottoa G I. A cellular automata model for citrus variegated chlorosis[J]. Physica A, 2001, 295: 42-48.
- [15] Fuentes1 M A, Kuperman M N. Cellular automata and epidemiological models with spatial dependence[J]. Physica A, 1999, 267: 471.
- [16] 许田, 张培培. 流行病传播模型与 SARS[J]. 自然杂志, 2004, 26(1).
- [17] 于鑫, 段晓东, 刘向东, 等. 基于元胞自动机的流行病传播模型及模拟[J]. 计算机工程与应用, 2005, 41(2): 205-209.
- [18] 刘颖, 陈禹. 复杂适应系统理论对控制 SARS 疫情的模拟分析[J]. 复杂系统与复杂性科学, 2004, 1(2): 74-79.

(上接 198 页)