GM (1,1) Evaluation Model of Measles Trends Based on Information Mining

Shiqiang ZHANG[1,3], Salan ZHANG[2], Zheng JIANG[1,3], Chunli WANG[1,3], Ting WANG[1,3]

1. Dept. of mathematics, Chongqing Medical University, Chongqing, China

2. Ping An Asset Management Co. Ltd, Ping An Insurance (Group) Company of China, Ltd, Shanghai, China

3. Lab. of forensic medicine and biomedicine information, Chongqing Medical University, Chongqing, China

1. math808@sohu.com,2.zhangsalan@yahoo.com.cn

Abstract: Measles is an acute respiratory infection and is the most common acute respiratory infectious disease, which is seriously harmful to the child's health. It is also highly infectious and can be very popular in densely populated areas without a general vaccination of the pandemic. Therefore, the establishment of early warning mechanism of measles, measles trend forecasting, prevention and control of measles in the relevant departments to provide a scientific basis for decision making, is an important public health work. Based on the information excavation and the grey system theory, this essay has given a forecast method of measles trend. According to this method and our country's actual situation, we have established a grey system GM(1,1) model of measles trend. This model forecasts the trend of our country's measles disease incidence rate very well.

Keywords: Information Excavation; Grey System; Trend; Measles

基于信息挖掘的麻疹流行趋势 GM(1,1)评价模型

张世强 [1,3], 张飒岚 [2]，蒋 峥 [1,3], 王春丽[1,3], 王婷[1,3]

1.重庆医科大学数学教研室，重庆，中国，400016

2.中国平安保险(集团)股份有限公司平安资产管理有限责任公司，上海，中国，201201

3.重庆医科大学法医学与生物医学信息研究室，重庆，中国，400016

1. math808@sohu.com,2.zhangsalan@yahoo.com.cn

摘　要: 麻疹是一种急性呼吸道传染病，是儿童最常见的急性呼吸道传染病之一，严重危害儿童身体健康，其传染性很强，在人口密集而未普种疫苗的地区可形成社区内的流行。因此，建立麻疹的预警机制，预测麻疹的流行趋势，为有关部门预防和控制麻疹提供科学的决策依据，是一项很有意义的公共卫生工作。基于信息挖掘与灰色系统理论, 给出了一种麻疹流行趋势的预测方法。利用该方法，结合我国的实际情况，建立了一个评价麻疹流行趋势的灰色系统 GM(1,1)模型。该模型较好地评价和预测了我国麻疹发病率的流行趋势。

关键词: 信息挖掘；灰色系统；流行趋势；麻疹

1 引言

麻疹是一种急性呼吸道传染病，接种麻疹疫苗是最有效的办法。麻疹是儿童最常见的急性呼吸道传染病之一，严重危害儿童身体健康，其传染性很强，在人口密集而未普种疫苗的地区易发生流行，大约 2 到 3 年就会发生一次较大的流行。临床上以发热、上呼吸道炎症、眼结膜炎等症状，而以皮肤出现红色斑丘疹和颊粘膜上有麻疹粘膜斑及疹退后遗留色素沉着伴糠麸样脱屑为特征。麻疹患者是惟一的传染源，患儿从接触麻疹后 7 天至出疹后 5 天均有传染性，病毒存在于眼结膜、口、鼻、咽和气管等分泌物中，通过喷嚏、说话和咳嗽等由飞沫传播。本病传染性极强，易感者接触后 90％以上均发病，据统计资料显示，1 到 5 岁小儿发病率最高。使用麻疹减毒活疫苗后，发病率可下降，但因免疫力不持久，故发病年龄后向后推移。

目前发病者在未接受疫苗的学龄前儿童、免疫失败的十几岁儿童和青年人中多见，甚至可形成社区内的流行。因此，建立麻疹的预警机制，预测麻疹的流行趋势，为有关部门预防和控制麻疹提供科学的决策依据，是一

 978-1-935068-47-1 © 2010 SciRes.

项很有意义的公共卫生工作。

本文基于信息挖掘与灰色系统理论，给出了一种麻疹流行趋势的分析方法。利用该分析方法，结合我国的实际情况，建立了一个基于信息挖掘与灰色系统理论的麻疹流行趋势的GM(1,1)模型。然后利用该模型对我国麻疹发病率进行了麻疹流行趋势分析。

2 原理与方法

灰色系统理论[1]最早基于对社会、经济等的宏观预测和决策而提出，灰色系统理论中的GM(1,1)模型是应用最为广泛的灰色模型。

2.1 传统的GM(1,1)模型的建模方法

设时间序列为 $t=\{t_1, t_2, \ldots, t_n\}$；其对应的原始数据序列为

$$x^{(0)}=\{x^{(0)}(t_1), x^{(0)}(t_2), \ldots, x^{(0)}(t_n)\} \quad (1)$$

令 $\Delta t_k=t_k-t_{k-1}$，当 $\Delta t_k=const$，称序列(1)为等距序列。当 $\Delta t_k\neq const$，序列(1)为非等距序列。原始数据序列(1)的一次累加生成序列为

$$x^{(1)}=\{x^{(1)}(t_1), x^{(1)}(t_2), \ldots, x^{(1)}(t_n)\} \quad (2)$$

其中

$$\begin{cases} x^{(1)}(t_1)=x^{(0)}(t_1) \\ x^{(1)}(t_k)=x^{(1)}(t_1)+\sum_{i=2}^{k}x^{(0)}(t_i)\Delta t_i \\ k=2,3,\cdots,n \end{cases} \quad (3)$$

从一次累加生成序列(2)还原为原始数据序列(1)的计算公式为

$$\begin{cases} x^{(0)}(t_k)=\dfrac{x^{(1)}(t_k)-x^{(1)}(t_{k-1})}{t_k-t_{k-1}} \\ k=2,3,\cdots,n \end{cases} \quad (4)$$

当一次累加生成序列(2)接近于非齐次指数规律变化时，序列(2)的响应函数是微分方程

$$\begin{cases} \dfrac{dx^{(1)}(t)}{dt}+ax^{(1)}(t)=b \\ x^{(1)}(t_1)=x^{(0)}(t_1) \end{cases} \quad (5)$$

的解 $x^{(1)}(t)=(x^{(0)}(t_1)-\dfrac{b}{a})e^{-a(t-t_1)}+\dfrac{b}{a}$，其中未知常数 a 与 b 为待辨参数。序列(2)的离散响应函数为

$$\begin{cases} x^{(1)}(t_1)=x^{(0)}(t_1) \\ x^{(1)}(t_k)=(x^{(0)}(t_1)-\dfrac{b}{a})e^{-a(t_k-t_1)}+\dfrac{b}{a} \end{cases} \quad (6)$$

式(6)中的 $k=2,3,\ldots,n$。为了确定待辨参数 a 与 b，可用下面的差分方程代替微分方程(5):

$$\begin{cases} \dfrac{\Delta x^{(1)}(t_k)}{\Delta t_k}+ax^{(1)}(t_k)=b \\ x^{(1)}(t_1)=x^{(0)}(t_1) \end{cases} \quad (7)$$

其中

$$\frac{\Delta x^{(1)}(t_k)}{\Delta t_k}=\frac{x^{(1)}(t_k)-x^{(1)}(t_{k-1})}{t_k-t_{k-1}}=x^{(0)}(t_k)$$

再用数据代换

$$z^{(1)}(t_k)=\lambda x^{(1)}(t_k)+(1-\lambda)x^{(1)}(t_{k-1}))$$

平滑差分方程(7)中的 $x^{(1)}(t_k)$，可得平滑后差分方程为

$$\begin{cases} \dfrac{\Delta x^{(1)}(t_k)}{\Delta t_k}+az^{(1)}(t_k)=b \\ x^{(1)}(t_1)=x^{(0)}(t_1) \end{cases} \quad (8)$$

上式中的 $z^{(1)}(t_k)$称为背景值，$\lambda\in[0，1]$称为背景参数。背景参数λ目前还没有最佳的取值方法，一般应用时为简单易行，通常按文献[1]取背景参数为1/2。将原始数据序列(1)的一次累加生成序列(2)代入上式，用矩阵方程可确定待辨参数 a 与 b:

$$\begin{bmatrix} a \\ b \end{bmatrix}=(B^TB)^{-1}B^TY$$

其中 $Y=[x^{(0)}(t_2), x^{(0)}(t_3), \ldots, x^{(0)}(t_n)]^T$，

$$B^T=\begin{bmatrix} -z^{(1)}(t_2) & -z^{(1)}(t_3) & \cdots & -z^{(1)}(t_n) \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

将上面求得的参数 a 与 b 代入式(6)中，得到原始数据序列 $x^{(0)}$的灰色系统GM(1,1)模型:

$$\begin{cases} \hat{x}^{(0)}(t_1)=x^{(0)}(t_1) \\ \hat{x}^{(0)}(t_k)=\dfrac{\hat{x}^{(1)}(t_k)-\hat{x}^{(1)}(t_{k-1})}{t_k-t_{k-1}} \\ \hat{x}^{(1)}(t_k)=(x^{(0)}(t_1)-\dfrac{b}{a})e^{-a(t_k-t_1)}+\dfrac{b}{a} \end{cases} \quad (9)$$

2.2 基于信息挖掘的GM(1,1)模型建模方法

建立GM(1,1)模型的传统方法有计算方法简便的优点，美中不足的是拟合精度和预测精度有时较差。文[2]和文[3]将参数的求解与边值的确定综合在一起讨论，提出了一种基于信息挖掘的方法。用该方法建立的GM(1,1)模型既能极大地提高GM(1,1)模型的拟合精度和预测精度，又能保持传统方法建立GM(1,1)模型的计算方法简便的优点。为方便读者，下面简单介绍一下该方法。

首先利用传统(等距及非等距序列)的灰色系统GM(1,1)模型的建模方法得到原始数据序列 $x^{(0)}$的灰色系统GM(1,1)模型(9)，可将模型(9)称为毛坯模型。

然后对毛坯模型(9)进行精加工，即将毛坯模型(9)的第三个公式改写为

$$\hat{x}^{(1)}(t_k)=\alpha e^{-a(t_k-t_1)}+\beta \qquad k=2,3,\ldots,n \qquad (10)$$

其中α与β为新的待辨参数。

利用传统(等距及非等距序列)的灰色系统GM(1,1)模型的建模方法得到的参数 a，再次利用原始数据序列的累加生成序列以及原始数据序列对应的时间序列。将累加生成序列 $x^{(1)}=\{x^{(1)}(t_1), x^{(1)}(t_2), \ldots, x^{(1)}(t_n)\}$和原始数据序列对应的时间序列 $t=\{t_1, t_2, \ldots, t_n\}$代入上式，利用线性代数中的矩阵方程可确定待辨参数$\alpha$与$\beta$[4]:

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=(B^TB)^{-1}B^TY$$

其中 $Y=[x^{(1)}(t_1), x^{(1)}(t_2), \ldots, x^{(1)}(t_n)]^T$，

$$B^T=\begin{bmatrix}1 & e^{-a(t_2-t_1)} & \cdots & e^{-a(t_n-t_1)}\\ 1 & 1 & \cdots & 1\end{bmatrix}$$

将参数α与β代入式(10)中，得到原始数据序列 $x^{(0)}$的新灰色系统 GM(1,1)模型:

$$\hat{x}^{(1)}(t_k)=\alpha e^{-a(t_k-t_1)}+\beta \qquad k=1,2,\ldots,n \qquad (11)$$

将上式进行一次累减生成，可得到原始数据序列 $x^{(0)}$的还原值为:

$$\begin{cases}\hat{x}^{(1)}(t_k)=\alpha e^{-a(t_k-t_1)}+\beta & (k=1,2,\cdots,n)\\ \hat{x}^{(0)}(t_k)=\dfrac{\hat{x}^{(1)}(t_k)-\hat{x}^{(1)}(t_{k-1})}{t_k-t_{k-1}} & (k=2,3,\cdots,n)\end{cases} \qquad (12)$$

3 基于信息挖掘的 GM(1,1)麻疹流行趋势评价模型

基于信息挖掘与灰色系统理论的麻疹流行趋势的分析方法是：利用原始数据，首先利用灰色系统理论建立麻疹流行趋势 GM(1,1)模型；其次利用信息挖掘方法，在已建立的灰色系统的麻疹流行趋势 GM(1,1)模型的基础上，建立基于信息挖掘的 GM(1,1)麻疹流行趋势评价模型。

表 1 是我国 1976 年到 2006 年麻疹发病人数的统计数据。数据来源自中国知网的中国宏观数据挖掘系统以及中国卫生统计年鉴。为节省篇幅，只给出了有代表性的少数数据。

Table 1. Original data
表 1. 原始数据

Time(Year)	The number of incidence (unit: person)	Time(Year)	The number of incidence (unit: person)
1976	2563722	1981	1015331
1986	203940	1991	123442
1996	76738	2001	91253
2002	61143	2003	58539
2004	70583	2005	123172
2006	100163		

利用基于信息挖掘的方法，建立的麻疹疫情预测趋势的灰色系统 GM(1,1)模型为

$$\hat{x}^{(1)}(t_k)=-15720626e^{-0.206161308(t_k-t_1)}+18428783$$

其中，$k=1,2,\cdots,n$。

图 1 是我国 1976 年到 2006 年麻疹发病人数的统计数据及其模拟数据之间的曲线关系图。

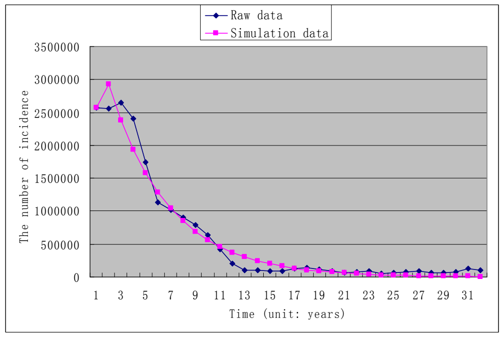

Figure 1. The tendency curve of incidence of measles with time
图 1. 麻疹发病人数随时间变化曲线趋势

从 1976 年到 2006 年麻疹发病人数随时间变化趋势图可以看出，从 1976 年到 2006 年我国麻疹发病人数呈逐年下降趋势。其中 1976 年到 1988 年的年发病人数下降趋势非常明显，1988 年之后的年发病人数趋于平稳。

图 2 是我国 1976 年到 1996 年麻疹发病人数的统计数据及其模拟数据之间的曲线关系图

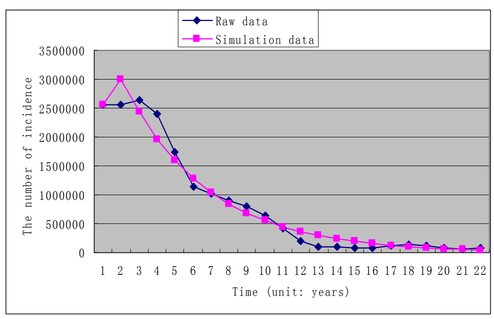

Figure 2. The tendency curve of incidence of measles with time
图 2. 麻疹发病人数随时间变化曲线趋势

从 1976 年到 1996 年麻疹发病人数随时间变化趋

978-1-935068-47-1 © 2010 SciRes.

势图可以看出，从 1976 年到 1996 年我国麻疹发病人数呈逐年快速下降趋势，1988 年到 1996 年发病人数一直稳定在较低的水平。

图 3 是我国 1996 年到 2006 年麻疹发病人数的统计数据及其模拟数据之间的曲线关系图

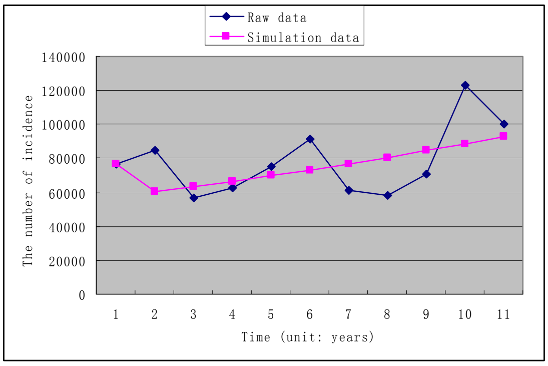

Figure 3. The tendency curve of incidence of measles with time
图 3. 麻疹发病人数随时间变化曲线趋势

从 1996 年到 2006 年的麻疹发病人数随时间变化趋势图可以看出，我国麻疹发病人数呈轻微上升趋势。

下面利用基于信息挖掘的麻疹疫情预测趋势的灰色系统 GM(1,1)模型

$$\hat{x}^{(1)}(t_k) = -15720626e^{-0.206161308(t_k - t_1)} + 18428783$$

预测了 2007 年至 2012 年我国麻疹发病人数，预测结果提示 2007 年至 2012 年我国麻疹发病人数呈缓慢上升趋势。具体数据见表 2。

Table 2. Forecast data
表 2. 预测数据

Time(Year)	The number of incidence (unit: person)	Time(Year)	The number of incidence (unit: person)
2007	97567	2008	102365
2009	107398	2010	112680
2011	118220	2012	124034

这可能与人口的频繁流动有关，提示我们应该加强对流动人群的麻疹疫苗的接种工作。

4 结论

上面运用由基于信息挖掘的方法建立的 GM(1,1)模型对我国 1976 年到 2006 年麻疹发病人数的统计数据进行灰色拟合及动态预测，结果显示该我国麻疹疫情已得到有效控制，不会出现暴发性流行趋势。

GM (1.1)模型是灰色动态模型中应用最广泛的预测模型，主要用于对复杂系统某一主导因素特征值的拟合和预测，以揭示主导因素变化规律和未来发展变化态势。在医学领域，广泛应用于流行病的预测[5]。但是任何一种预测模型都有其局限性，GM(1,1)模型的缺点就在于拟合精度和预测精度有时较差。

本文基于信息挖掘与灰色系统理论，给出一种基于信息挖掘的灰色系统 GM(1,1)模型的建模方式。该方式一方面极大地提高 GM(1,1)模型的拟合精度和预测精度，另一方面又保持了传统方法建立 GM(1,1) 模型计算方法简便的优点。利用该方式，给出了一种基于信息挖掘的 GM(1,1)麻疹流行趋势评价模型。同时结合我国的实际情况建立了麻疹流行预测模型，并利用该模型对我国麻疹发病率进行了麻疹流行趋势的分析与预测。实例分析验证了基于信息挖掘方法的有效性和实用性。

致谢

This project was supported by Chongqing Medical University gran (NSFYY200722)。

References (参考文献)

[1] Julong Deng. Control Problems of Grey Systems[J].Syst & Contr Lett,1982,1(5),P288-294.
[2] Shiqiang Zhang, Discussion about Non-equal-space Sequence Modelling Process [J],*Mathematics in Practice and Theory*, 2007,37(18),P50-56(Ch)
张世强,关于非等距序列建模过程的讨论[J],数学的实践与认识,2007,37(18),P50-56.
[3] Shiqiang Zhang, Modeling Method of Grey System GM(1.1) Model Based Information Reused and It's Application [J],*Mathematics in Practice and Theory*, 2009,39(13),P97-104(Ch)
张世强,基于信息再利用的灰色系统 GM(1,1)模型建模方法及应用[J],数学的实践与认识, 2009,39(13),P97-104.
[4] Shiqiang Zhang, Zheng Jiang, Modeling Method of Logisitic Model Based Grey System[J],*Mathematics in Practice and Theory*, 2010,40(9),P144-148 (Ch)
张世强,蒋峥,基于灰色系统的逻辑斯蒂模型的建模方法[J],数学的实践与认识, 2010,40(9),P144-148.
[5] Jieneng Lv, Shiqiang Zhang, Lei Zhang and Zheng Jiang.Grey System GM(1.1) Model Based Information Reused and It's Application on Infectious Disease Trends Forecast[J],*Moden Preventive Medicine*,2010,37(14),P2601-2603 (Ch)
吕杰能,张世强,张雷,蒋峥,基于信息再利用的灰色系统 GM(1.1)模型在传染病流行趋势预测中的应用[J],现代预防医学,2010,37(14),P2601-2603.