

基于灰色 Verhulst 的互联网上网人数动态预测模型*

朱苗苗,牛国锋,乐德广

(常熟理工学院 计算机科学与工程学院,江苏 常熟 215500)

摘要: 在分析最近几年我国互联网上网人数实测数据的基础上,根据数据本身所具有的特征,引入灰色 Verhulst 预测理论,通过分析建立了我国互联网上网人数的灰色 Verhulst 动态模型,并通过所建立的模型对 2000~2005 年我国互联网上网人数进行实例验证和误差检验。检验结果表明,所建立的灰色 Verhulst 动态模型对实测数据具有很高预测精度,具有一定的研究和应用价值。

关键词: 互联网; Verhulst 模型; 残差; 相对误差

中图分类号: TP302

文献标识码: A

文章编号: 1674-7720(2011)23-0091-03

Prediction of the number of people using the Internet based on the gray verhulst dynamic model

Zhu Miaomiao, Niu Guofeng, Le Deguang

(School of Computer Science and Engineering, Changshu Institute of Technology, Changshu 215500, China)

Abstract: Based on the data of the people using the internet in recent years, this paper introduce the prediction theory of gray Verhulst according to the feature of the data, analyze and build the dynamic model of gray Verhulst. By the model, the paper conducts the specimen verification and deviation inspection on the number of the people using the internet. The result expresses that the gray Verhulst dynamic model is helpful to increase the precision of the testing data, and it can plays a significant role in research and application.

Key words: Internet; Verhulst model; residue; relative error

如今,互联网已经成为人们生活中不可或缺的一部分,它的出现不仅改变了人们传统的交流方式,开阔了人们的视野,而且改变了传统企业的营运模式,催生出大量新兴企业,对人们的生活和社会发展带来了巨大的影响。分析预测互联网上网人数,可以全局掌握网络的应用情况,为网络建设、应用、管理及与网络相关的各种技术和业务提供决策依据^[1]。

Verhulst 预测理论是 1837 年德国生物学家 Verhulst 在研究生物繁殖规律时提出的。其基本思想是生物个体数量是呈指数增长的,受周围环境的限制,增长速度逐渐放慢,最终稳定在一个固定值。Verhulst 模型主要用来描述具有饱和状态的过程,即“S”型过程,常用于人口预测、生物生长、繁殖预测及产品经济寿命预测等^[2]。

本文在分析最近几年我国互联网上网人数实测数据的基础上,根据数据本身所具有的特征,引入灰色 Verhulst

预测理论,通过分析建立了我国互联网上网人数的灰色 Verhulst 动态模型,并通过所建立的模型对 2000~2005 年我国互联网上网人数进行实例验证和误差检验。检验结果表明,所建立的灰色 Verhulst 预测模型对实测数据的预测优于 GM(1,1)模型、线性回归、指数曲线等其他方法。

1 Verhulst 预测模型的建立

灰色系统是指既含有已知信息、又含有未知信息的系统,因为它具有对各种现象进行分析判断的能力、对宏观系统进行规划与决策的功能,其立足点是对系统输出序列进行研究,而不过多地涉及系统输入序列的研究,对原始数据序列长度要求也不高。它的特点是所需信息量少,不仅能够将无序离散的原始序列转化为有序序列,而且预测精度高,能够保持原系统的特点特征,较好地反映系统的实际情况^[2]。

灰色 Verhulst 模型主要是用来描述非单调的摆动发展序列或具有饱和状态的 S 型序列。表 1 和图 1 分别显

《微型机与应用》2011 年第 30 卷第 23 期

* 基金项目:江苏省高校自然科学基金项目(10KJB520001)

示了从 2000~2005 年我国互联网上网人数的统计。从图 1 可以很明显的看出,自从 2000 年以来,我国互联网上网人数一直保持着比较强劲的增长势头,并在一定时间段上是呈部分“S”型变化的,所以在一定程度上互联网上网人数的变化情况可以通过建立灰色 Verhulst 模型进行预测。

表 1 2000~2005 年我国互联网上网人数统计

年份	2000	2001	2002	2003	2004	2005
互联网上网人数/万人	2 250	3 370	5 910	7 950	9 400	11100

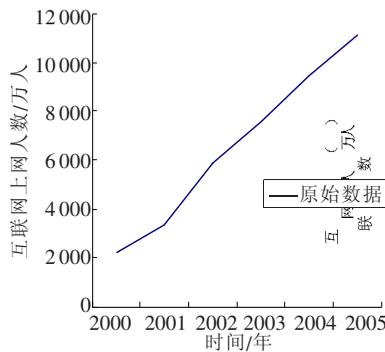


图 1 2000~2005 年我国互联网上网人数数据统计

1.1 Verhulst 预测模型的建立

设在一个模型预测中,原始数据数列为 $X^{(1)}$,为了弱化呈离散状态的原始数据,对原始时间序列进行数据的一次累减处理,经过处理的时间序列称为生成累减序列 $X^{(0)}=(x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$;然后通过对原始数据序列累加生成的递增序列实行紧邻均值生成 $Z^{(1)}$,同时令 $X^{(0)}(1)=X^{(1)}(1)$,这时所建立的灰色模型为^[3]:

$$X^{(0)}+aZ^{(1)}=b(z^{(1)})^2 \quad (1)$$

相应的白化方程为:

$$dx^{(1)}/dt+ax^{(1)}=b(x^{(1)})^2 \quad (2)$$

解得:

$$x^{(1)}(t)=1/e^{at}[1/x^{(1)}(t)-b(1-e^{-at})/a]=ax^{(1)}(0)/e^{at}[a-bx^{(1)}(0)(1-e^{-at})]=ax^{(1)}(0)/[bx^{(1)}(0)+(a-bx^{(1)}(0))e^{at}] \quad (3)$$

所求的灰色 Verhulst 模型时间响应式为:

$$\hat{x}^{(1)}(k+1)=\frac{ax^{(1)}(0)}{bx^{(1)}(0)+(a-bx^{(1)}(0))e^{ak}} \quad (4)$$

1.2 Verhulst 预测模型的求解

式(4)中, k 为时间序数, a 、 b 都为待求系数, a 为发展系数,其大小及符号反映 $X^{(1)}$ 及 $X^{(0)}$ 的发展态势, b 为系统的输入,其内涵是具有灰色信息覆盖的系统作用量^[4],不能直接观测得到,在灰色理论中, a 、 b 的值可以用最小二乘法求取,即:

$$\hat{a}=\begin{bmatrix} a \\ b \end{bmatrix}^T=(B^TB)^{-1}B^TY \quad (5)$$

$$\text{其中, } B=\begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^2 \\ -z^{(1)}(3) & (z^{(1)}(3))^2 \\ -z^{(1)}(4) & (z^{(1)}(4))^2 \\ -z^{(1)}(5) & (z^{(1)}(5))^2 \\ -z^{(1)}(6) & (z^{(1)}(6))^2 \\ -z^{(1)}(7) & (z^{(1)}(7))^2 \end{bmatrix}$$

其中, $Z^{(1)}$ 为由原始数据进行累加生成的紧邻均值生成序列,即:

$$Z^{(1)}=\{z^{(1)}(k)\}_2^7=(\frac{x^{(1)}(1)+x^{(1)}(2)}{2}, \dots, \frac{x^{(1)}(k-1)+x^{(1)}(k)}{2})$$

式(5)中, $Y=[x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(K)]^T$ 。

然后将所求的 a 、 b 值分别代入式(4)中,即求得原始数据序列的预测值。

2 原始数据的分析和模型预测

互联网上网人数预测对于整个互联网发展和相对业务决策有着至关重要的意义。近年来对于互联网上网人数的预测方法比较多,有线性回归、指数曲线等方法。虽然这种方法能够简单、直观地得到预测结果,但由于它一般采用线性或者指数形式,而不完全符合某些事物的发展趋势,所以预测精度存在一定的缺陷。同样对于参考文献 [2] 所介绍的 GM(1,1)模型方法,由于 GM(1,1)模型适用于具有较强指数规律的序列,只能描述单调的变化过程,预测精度同样也存在一定的缺陷。下面根据表 1 中的原始数据,利用本文所建立的 Verhulst 模型对我国近几年来互联网上网人数进行预测。

(1)根据表 1 中的数据,其原始数列:

$$X^{(1)}=(2\ 250, 3\ 370, 5\ 910, 7\ 950, 9\ 400, 11\ 100)$$

(2) $X^{(1)}$ 的一次累减(1-IAGO)序列:

$$X^{(0)}=[x^{(0)}(k)]_2^7=(2\ 250, 1\ 120, 2\ 540, 2\ 040, 1\ 450, 1\ 700)$$

(3)紧邻均值生成序列:

$$Z^{(1)}=\{z^{(1)}(k)\}_2^7=(2\ 810, 4\ 640, 6\ 930, 8\ 675, 10\ 250)$$

(4)求得 B 及 Y :

$$B=\begin{bmatrix} -2810 & 7896100 \\ -4040 & 21529600 \\ -6930 & 48024900 \\ -8675 & 75255625 \\ -10250 & 105062500 \end{bmatrix}$$

$$Y=[x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(6)]^T=[1\ 120, 2\ 540, 2\ 040, 1\ 450, 2\ 700]^T$$

(5)将 B 和 Y 分别代入式(5)中,通过 MATLAB 软件求得 a 和 b ,即:

$$\hat{a}=\begin{bmatrix} a \\ b \end{bmatrix}^T=(B^TB)^{-1}B^TY=\begin{bmatrix} -0.6570726998 \\ -0.0000504886 \end{bmatrix}$$

(6)取 $X^{(1)}(0)=X^{(1)}(1)=4\ 456$

根据式(4),求得 Verhulst 互联网上网人数预测时间响应式为:

$$\hat{x}^{(1)}(k+1)=\frac{ax^{(1)}(1)}{bx^{(1)}(1)+(a-bx^{(1)}(1))e^{ak}} \\ =\frac{-1\ 478.413\ 574\ 56}{-0.11359943+(-0.54347327)e^{-0.6570726998k}}$$

(7)由此可以对 2000~2005 年我国互联网上网人数进行模拟、预测:

$$\hat{x}^{(1)}(1)=2\ 250 \quad \hat{x}^{(1)}(2)=3\ 740 \quad \hat{x}^{(1)}(3)=5\ 694$$

$$\hat{x}^{(1)}(4)=7\ 810 \quad \hat{x}^{(1)}(5)=9\ 673 \quad \hat{x}^{(1)}(6)=11\ 038$$

(8)原始数据和预测数据比对,比对结果如图 2 所示。

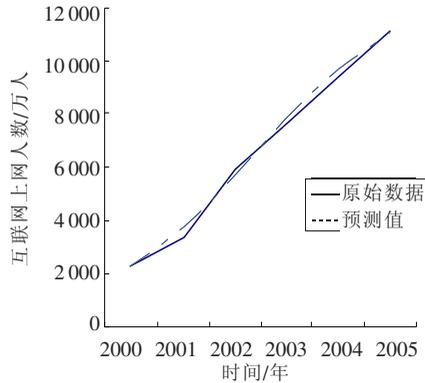


图2 基于 Verhulst 模型预测上网人数统计结果

3 Verhulst 模型检验

一种模型要经过多种检验才能判断其是否合理、合格,也只有通过检验的模型才能用来做预测^[5]。本节中,为了验证所建灰色模型是否能更准确对互联网上网人数进行预测,针对表2中的数据分别采用残差检验法和后残差检验法进行检验。

表2 2000~2005 我国互联网上网人数预测结果

时间/年	2000	2001	2002	2003	2004	2005
原始数据/万人	2250	3370	5910	7950	9400	11100
预测数据/万人	2250	3740	5694	7810	9673	11038

3.1 残差检验法

根据计算出来的预测值和原始数据求出残差和相对误差:

$$\text{残差: } q^{(1)}(k) = x^{(1)}(k) - \hat{x}^{(1)}(k)$$

$$\text{相对误差: } e(k) = |q^{(1)}(k)/x^{(1)}(k)| \times 100\%$$

通过上述检验方法,分别对本文的 Verhulst 模型和 GM(1,1)模型计算出误差检验结果,表3显示了 Verhulst 模型和 GM(1,1)模型预测结果对比。

表3 Verhulst 模型和 GM(1,1)模型预测结果对比

年份	原始数据	Verhulst 模型预测结果			GM(1,1)模型预测结果		
		预测值	残差	相对误差/%	预测值	残差	相对误差/%
2000	2250	2250	0	0
2001	3370	3740	370	10.98	3370	0	0
2002	5910	5694	-216	3.65	5406	503	8.51
2003	7590	7810	220	2.90	7433	517	6.81
2004	9400	9673	273	2.90	9157	243	2.59
2005	11038	11038	-62	0.56	11817	717	6.46
		平均相对误差 3.50%			平均相对误差 4.87%		

从表3可以看出,即使参考文献[2]中通过建立残差对 GM(1,1)模型进行修正,其预测结果也没有本文所介绍的 Verhulst 模型预测精度高。

3.2 后残差检验方法

相比较残差检验法,后残差检验法需要在原始数据和所得到的残差的基础上,分别计算出原始数据和残差的均值,从而更进一步计算出其方差。即:原始数据 $x^{(1)}$

$$(k) \text{ 均值 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x^{(1)}(k); \text{ 原始数据方差 } S_1^2 = \frac{1}{n} \sum_{i=1}^n [(x^{(1)}(k) -$$

$$\bar{x})^2]; \text{ 残差 } q^{(m)}(k) \text{ 均值 } \bar{q} = \frac{1}{n} \sum_{i=1}^n q^{(1)}(k); \text{ 残差方差 } S_2^2 = \frac{1}{n}$$

$$\sum_{i=1}^n [(q^{(1)}(k) - \bar{q})^2].$$

而后检验比值 $C = S_2/S_1$, 小误差频率 $p = \{|q^{(1)}(k) - \bar{q}| < 0.6745 S_1\}$ 。一个好的预测,要求 C 值越小越好,一般要求 $C < 0.35$, 最大不超过 0.65。预测好坏的另一个指标是小误差频率要大。所谓小误差是指绝对偏差 $|q^{(1)}(k) - \bar{q}| < 0.6745 S_1$, 一般要求 $P > 0.95$, 不得 < 0.7 。按照 P 和 C 的大小,可将预测精度分为 4 个等级,各等级标准如表 4 所示^[3]。

等级	P	C
好	> 0.95	< 0.35
合格	> 0.80	< 0.45
勉强	> 0.70	< 0.50
不合格	≤ 0.70	< 0.65

通过以上方法对所建立的互联网上网人数模型进行检验拟合,其结果显示:本文所建立的灰色 Verhulst 动态模型后检验比值 $C = 0.0308$, $P = 1$, 根据表 4 判断依据,此模型预测效果为好。

合理科学地对我国互联网上网人数进行预测,可以为网络产业的发展提供科学有效的参考,同时,对于整个社会经济的发展也具有重要的意义。

本文通过分析研究动态 Verhulst 模型的思想、特征和我国互联网上网人数逐年增长,并在一定时间段上原始数据呈部分“S”型变化的趋势前提下,提出将 Verhulst 模型应用于我国互联网上网人数的预测。通过建立模型和结果检验,证明此方法具有方法简单、预测精度高、可进行中长期预测等特点,是一种值得推广的预测方法。

虽然通过此模型的建立,模拟的精度还是比较高的,但是根据灰色理论研究的原理和方法,若想提高预测精度、保证结果的可靠性,可以将 Verhulst 模型与其他预测方法结合使用,同时应该根据不断得到的新数据对预测模型进行不断的修正,因为随着时间的推移和一些外界因素的影响,数据的变化趋势会发生一定的变化,只有这样才能够更加准确地对我国互联网上网人数进行科学准确的预测,从而产生一系列的实用价值。

参考文献

- [1] 陈晶晶,毛谦,刘国辉.我国互联网互用数预测研究[J].光通信研究,2007(1).
- [2] 连飞.基于 GM(1,1)模型的我国互联网上网人数灰色预测[J].统计与咨询,2007(4).
- [3] 邓聚龙.灰色系统理论教程[M].武汉:华中理工大学出版社,1999.
- [4] WU W Y, CHEN S P. A prediction method using the Grey model GMC (1,n) combined with the grey relational analysis; a case study on Internet access population forecast [J]. Applied Mathematics and Computation, 2005,169(1).
- [5] 葛建华.灰色模型在人口预测中的应用[J].西安文理学院学报,2008,11(3).

(收稿日期:2011-05-12)

作者简介:

朱苗苗,女,1981年生,讲师,硕士,主要研究方向:数据挖掘,人工智能和决策研究。

《微型机与应用》2011年第30卷第23期