

## 基于灰色加权马尔可夫链的移动通信市场预测

王增民, 王开珏

(北京邮电大学 经济管理学院, 北京 100876)

**摘 要:** 在传统的用灰色预测模型预测的方法基础上, 建立了灰色加权马尔可夫链模型. 以中国移动通信市场预测作为实例, 介绍了使用这种模型的方法与步骤. 灰色加权马尔可夫链模型既考虑了从时间序列中挖掘数据的演变规律, 又通过规范化各阶自相关系数为权重, 用加权的马尔可夫转移概率矩阵的变换, 考虑数据的随机波动, 具有严密的科学性, 能较好地应用于中国移动通信市场的预测.

**关键词:** 灰色预测模型; 加权马尔可夫链; 移动通信市场; 预测

### 1 引言

中国作为通信市场大国, 特别随着移动通信技术的迅猛发展, 到 2008 年底中国已经成为世界移动通信第一大网, 移动通信用户也成为世界第一. 据统计, 截止到 2008 年底中国移动通信用户已达到 6.4 亿多户, 移动通信市场已经发展成为中国通信产业的拉动力. 移动通信市场的高速发展不仅为中国 GDP 贡献巨大的产值, 而且对通信服务的创新和发展起到积极的推动作用, 因此对移动通信市场的预测尤为重要. 为了能够更清晰地认识和研究它, 把移动通信市场抽象为没有物理原型、因素空间难以穷尽、行为轨迹无法控制、信息不完全的灰色系统, 据此可以进行灰色模型预测的研究<sup>[1]</sup>.

但是因为灰色模型预测在其使用条件上存在着一定的限制并且受到原始数据的起伏性波动影响比较大, 同时利用 GM(1,1) 进行预测时, 由于累加生成数列的缘故, 也会对其发展的规律产生影响, 削弱了被预测系统发展的阶段性规律, 这就导致了灰色预测模型在预测的过程中产生较大的误差. 针对以上的问题, 赵佳妮, 吴兆麟 (2005)<sup>[2]</sup>; 田俊改, 许红军 (2009)<sup>[3]</sup>进行了研究并且在预测精度方面取得了良好的效果. 作为进一步研究, 笔者拟对移动通信市场进行灰色预测研究, 并且通过对灰色预测模型的结果进行加权马尔可夫改进来提高其预测的准确性, 为移动通信市场的精确预测提供方法依据.

### 2 灰色加权马尔可夫链模型建立

#### 2.1 建立灰色预测模型

灰色系统理论<sup>[4]</sup>是邓聚龙教授与 1982 年提出并加以发展的, 它对时间序列短, 统计数据少, 信息不完全系统的分析与建模具有独特的功效, 灰色预测模型 (Gray Forecast Model)

作为灰色系统理论在预测方面得到了广泛的应用<sup>[5-8]</sup>. 灰色预测模型是通过少量的, 不完全的信息, 建立数学模型并做出预测的一种预测方法.

GM(1,1) 的建模步骤如下<sup>[9]</sup>:

**定义 1** 设原始数据序列  $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(N)\}$ , 令

$$x^{(1)}(k) = x^{(0)}(1) + x^{(0)}(2) + \dots + x^{(0)}(k), k = 1, 2, \dots, N$$

$$\Delta x^{(1)}(k) = x^{(1)}(k) - x^{(1)}(k-1), k = 2, \dots, N$$

称  $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(N)\}$  为原始数据序列  $x^{(0)}$  的一次累加序列.

**定义 2** 设  $x^{(0)}$  为原始数据序列,  $x^{(1)}$  为原序列的一次累加序列, 对  $x^{(1)}$  建立如下的灰色模型, 简记为 GM(1,1) 模型:

$$\frac{dx^{(1)}}{dk} + ax^{(1)} = u \quad (1)$$

其中  $a$  为发展系数,  $u$  为控制系数.

**定义 3** 由最小二乘法求出参数  $\hat{U} = \begin{pmatrix} \hat{a} \\ \hat{u} \end{pmatrix}$ , 得

$$\hat{U} = \begin{pmatrix} \hat{a} \\ \hat{u} \end{pmatrix} = (B^T B)^{-1} B^T y$$

其中,

$$B = \begin{pmatrix} -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(1)] & 1 \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(1)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(1)] & 1 \end{pmatrix}, y = \begin{pmatrix} x^{(0)}(2) & x^{(0)}(3) & \dots & x^{(0)}(N) \end{pmatrix}^T$$

则方程 (1) 的时间序列响应方程为:

$$x^{(1)}(k) = [x^{(1)}(1) - \frac{u}{a}]e^{-a(k-1)} + \frac{u}{a}, k = 2, 3, \dots, N \quad (2)$$

模拟值为:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), k = 2, 3, \dots, N \quad (3)$$

**定义 4** 对于灰色模型的精度检验, 采用后验差比值和小误差概率.

残差记为:

$$E(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 2, 3, \dots, N \quad (4)$$

$x^{(0)}$  的均值记:

$$\bar{x} = \frac{1}{N} \sum_{k=1}^N x^{(0)}(k)$$

$x^{(0)}$  的方差记为:

$$S_1 = \sqrt{\frac{1}{N} \sum_{k=1}^N [x^{(0)}(k) - \bar{x}]^2}$$

残差的均值记为:

$$\bar{E} = \frac{1}{N-1} \sum_{k=2}^N E(k)$$

残差的方差记为:

$$S_2 = \sqrt{\frac{1}{N-1} \sum_{k=2}^N [E(k) - \bar{E}]^2}$$

相对误差:

$$e(k) = [x^{(0)}(k) - \hat{x}^{(0)}(k)]/x^{(0)}(k) \quad (5)$$

则后验差比值:

$$C = \frac{S_2}{S_1} \quad (6)$$

小误差概率:

$$P = p\{|E(k) - \bar{E}| < 0.6745S_1\} \quad (7)$$

## 2.2 构造转移概率矩阵

马尔可夫链模型是在满足“后无效性”和“平稳性”的基础上建立的<sup>[10]</sup>. 则状态转移概率

$$P_{ij}(k) = M_{ij}(k)/M_i, \quad i = 1, 2, \dots, N$$

状态转移概率矩阵的构造如下:

$$P(k) = \begin{pmatrix} P_{11}(k) & P_{12}(k) & \cdots & P_{1N}(k) \\ P_{21}(k) & P_{22}(k) & \cdots & P_{2N}(k) \\ \vdots & \vdots & \vdots & \vdots \\ P_{N1}(k) & P_{N2}(k) & \cdots & P_{N3}(k) \end{pmatrix} \quad (8)$$

式中  $M_{ij}(k)$  表示由状态  $i$  经过  $k$  步转移到状态  $j$  的原始数据的个数;  $M_i$  为处于状态  $i$  的原始数据的个数;  $P_{ij}(k)$  表示由状态  $i$  经过  $k$  步转移到状态  $j$  的概率;  $P(k)$  表示  $k$  步状态转移概率矩阵.

## 2.3 计算各阶自相关系数和各种步长马尔可夫链的权重

**定义 5** 设平稳原始数列序列  $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(N)\}$ , 原始数列序列自协方差系数为:

$$\hat{\gamma}_i = \frac{1}{N} \sum_{k=1}^{N-i} (x_k - \bar{x})(x_{k+i} - \bar{x}), 1 \leq i \leq N-1$$

且  $\hat{\gamma}_{-i} = \hat{\gamma}_i$

原始数据序列方差为:

$$\hat{\gamma}_0 = \frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^2$$

则原始数据序列自相关系数为:

$$\hat{\rho}_i = \frac{\hat{\gamma}_i}{\hat{\gamma}_0} \quad (|i| \leq N-1)$$

对各阶自相关系数规范化作为各种步长的马尔可夫链的权重, 即:

$$\omega_i = |\gamma_i| / \sum_{i=1}^{N-1} \gamma_i \quad (9)$$

## 2.4 确定残差预测值

根据 GM(1,1) 模型中的公式 (4), 可得模型的残差为

$$E = \{E(1), E(2), \dots, E(N)\}$$

为了进行马尔可夫预测, 可以将相对误差  $E$  划分若干状态, 并且由于状态的划分与预测正确与否有很大的关系, 因此选择与时间序列数据特点相适应的状态划分方法尤其重要. 目前, 用于状态划分的方法主要有相对法<sup>[11]</sup>, 平行曲线法<sup>[12]</sup>, 人工神经网络法<sup>[13]</sup>. 对于移动通信市场预测, 采用平行曲线法比其他的方法更加合适. 因此运用平行曲线法可以将  $E$  中的相对误差划分为  $s$  个可能状态, 并且记为  $R_1, R_2, \dots, R_s$ .

将同一状态的各个预测概率加权作为移动通信市场处于该状态的预测概率, 即

$$P_i = \sum_{k=1}^s \omega_k P_i^{(k)} (i \in E)$$

其中  $\max P_i$  所对应的即为该时段移动通信市场状态的预测. 当确定预测值的变动区间在范围  $R_i = [R_{i-}, R_{i+}]$  之间时, 便可以用区间中位数作为未来时刻预测值的相对值, 则残差值为

$$\hat{e}_i = 0.5(R_{i-} + R_{i+})$$

## 2.5 预测原始序列

利用加权马尔可夫链预测出残差  $\hat{E}(k)$ , 结合 GM(1,1) 的预测值  $\hat{x}^{(0)}(i)$ , 按照残差的定义求出原始数据序列的预测值:

$$\hat{y}^{(0)}(i) = \hat{x}^{(0)}(i) + \hat{E}(k) \quad (10)$$

# 3 灰色加权马尔可夫链模型在通信市场预测中的应用

## 3.1 基于 GM(1,1) 模型的通信市场预测

用 1999–2008 年共 10 年的全国移动电话年末用户数据来预测 2009–2011 年全国移动电话年末用户数. 按照上述 GM(1,1) 预测步骤并且借助于 MATLAB6.5 软件得出了 1999–2008 年的全国移动电话年末用户数.

如表 1 所示, 样本均能通过序列的规律性检验, 因此将其带入已经编好的 MATLAB 程序中得出:

$$\hat{a} = -0.194269 \quad \hat{u} = 12246.588686$$

则时间响应形式方程组如下:

$$\begin{cases} x^{(1)}(t+1) = 67368.996643e^{0.194269t} - 63039.396643 \\ x^{(0)}(t+1) = x^{(1)}(t+1) - x^{(1)}(t) \end{cases}$$

通过以上的时间响应方程预测结果如表 1 所示.

表 1 中国移动电话年末用户数的实测值和预测值的对照表 (单位: 万户)

年份	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
实测值	4329.6	8453.3	14522.2	20600.5	26995.3	33482.4	39340.6	46105.8	54730.6	64123.0
预测值	4329.6	10445.4	17542.9	21304.5	25872.6	31420.3	38157.5	46339.4	56275.6	68342.3
绝对误差	0	-1992.1	-3020.7	-704	1122.7	2062.1	1183.1	-233.6	-1545	-4219.3

—— 实测值数据来源于国家统计年鉴

从该模型的拟合精度上来看, 计算后验差比值:

$$C = \frac{S_2}{S_1} = 0.1293$$

计算小误差概率

$$P = p\{|E(k) - \bar{E}| < 0.6745S_1\} = 0.8971$$

虽然后验差比值的拟合精度达到了一级, 小误差概率的拟合精度达到了二级, 但是有些预测数据的绝对误差较大, 因此运用加权马尔可夫链模型对残差进行相对的处理从而提高 GM(1,1) 模型的预测结果.

3.2 状态的划分

由表 1 可知:

$$E = (0, -1992.1, -3020.7, -704, 1122.7, 2062.1, 1183.1, -233.6, -1545, -4219.3)$$

根据实际情况将  $E$  划分成 3 个状态  $R_i(i = 1, 2, 3)$ , 这三个状态的区间分别为:

高估状态 (预测值相对于实际值较高) $R_1 \in [(\bar{E} - 2S_2) (\bar{E} - 0.5S_2)]$ , 即:  $R_1 \in [-4698.28 -1725.58]$

正常状态 (预测值处于正常范围) $R_2 \in (\bar{E} - 0.5S_2 \quad \bar{E} + -0.5S_2)$ , 即:  $R_2 \in (-1725.58 256.22)$

低估状态 (预测值相对于实际值较低) $R_3 \in [(\bar{E} + 0.5S_2)(\bar{E} + 2S_2)$ , 即:  $R_3 \in [256.22 3228.92]$

因此每年的状态如表 2 所示.

表 2 每年的状态

年份	状态	年份	状态
1999	$R_2$	2004	$R_3$
2000	$R_1$	2005	$R_3$
2001	$R_1$	2006	$R_2$
2002	$R_2$	2007	$R_2$
2003	$R_3$	2008	$R_1$

3.3 构造状态转移矩阵

根据表 2 中的数据可得到原始状态中的样本数 (将最后一个数据状态去掉) 分别为:  $M_1 =$

2,  $M_2 = 4$ ,  $M_3 = 3$ . 通过转移样本数和原始状态样本数的比值可以得到一步状态转移概率:

$$P(1) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \quad P(2) = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix}, \quad P(3) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix}$$

3.4 计算各种步长马尔可夫链的权重

根据式 (9) 介绍的权重计算方法, 先计算出各阶自相关系数, 然后标准化计算出的各阶自相关系数从而得出各步长马尔可夫链的权重, 如表 3 所示.

表 3 各阶自相关系数和各种步长马尔可夫链的权重

步长 $k$	1	2	3
$\gamma_k$	0.449	-0.119	-0.471
$\omega_k$	0.432	0.115	0.453

表 4 2009 年中国移动通信市场预测

初始年	初始状态	滞时	权重	1	2	3	概率来源
2008	1	1	0.432	1/2	1/2	0	$P^{(1)}$
2007	2	2	0.115	2/3	0	1/3	$P^{(2)}$
2006	2	3	0.453	0	1/2	1/2	$P^{(3)}$
$P_i(\text{权重和})$				0.293	0.442	0.265	

表 5 中国移动通信市场预测值比较

年份	实际值	灰色加权马尔可夫模型			灰色GM(1,1)模型		
		预测值	残差	相对误差 (%)	预测值	残差	相对误差 (%)
1999	4329.6	4329.6	0	0	4329.6	0	0
2000	8453.3	7233.47	1219.8	14.43	10445.4	-1992.1	-23.56
2001	14522.2	14331.0	191.2	1.32	17542.9	-3020.7	-20.80
2002	20600.5	20569.8	30.7	0.15	21304.5	-704	-3.42
2003	26995.3	27615.2	-619.9	-2.30	25872.6	1122.7	4.16
2004	33482.4	33162.9	319.5	0.95	31420.3	2062.1	6.16
2005	39340.6	39900.1	-559.5	-1.42	38157.5	1183.1	3.01
2006	46105.8	45604.5	501.3	1.08	46339.4	-233.6	-0.51
2007	54730.6	54533.0	197.6	0.36	56275.6	-1545	-2.82
2008	64123.0	65130.4	-1007.4	-1.57	68342.3	-4219.3	-6.58

3.5 预测状态及残差预测值的确定

依据中国移动电话年末用户数的实测值及相应的状态转移概率矩阵对 2009 年的移动通信市场状况进行预测, 其结果如表 4 所示.

由表 4 可知  $\max P_i = 0.442$ , 从而预测 2009 年的移动通信市场处于  $R_2$  状态. 即 2009 年的残差在  $R_2 \in (-1725.58\ 256.22)$ , 取区间  $R_2 \in (-1725.58\ 256.22)$  的中位数作为 2009 年移动通信市场预测的残差为 -734.68, 同时根据灰色 GM(1,1) 模型得到 2009 年移动通信市场预测值为 82996.5 万户, 因此可以得到 2009 年移动通信市场最后的预测值为 82261.82 万户. 同时表 5 给出了模型预测值与精度的比较.

从表 5 可以看出, 灰色加权马尔可夫模型的预测精度要远远高于灰色 GM(1,1) 模型, 灰色加权马尔可夫模型预测的相对误差小于灰色 GM(1,1) 模型预测的相对误差, 因此灰色加权马尔可夫模型预测的绝对误差精度也相当的高. 于是, 灰色加权马尔可夫链模型预测出 2009-20011 年中国移动通信市场的预测值, 如表 6 所示.

表 6 2009-2011 年中国移动市场预测值 (单位: 万户)	
年份	预测值
2009	82261.8
2010	101296.8
2011	124147.7

4 结束语

灰色加权马尔可夫链模型, 综合了灰色 GM(1,1) 模型和加权马尔可夫链模型的优点, 能够充分利用历史数据的信息. 该模型针对序列数据首先建立灰色 GM(1,1) 模型进行趋势的预测, 然后利用残差的自相关系数与马尔科夫状态转移概率矩阵相结合寻求系统的微观波动规律, 对其预测值进行二次拟合, 修正预测的范围, 增加了预测的可信度.

参考文献

[1] 刘思峰, 邓聚龙. GM (1,1) 模型的适用范围 [J]. 系统工程理论与实践, 2000, 20(5): 122-124.

[2] 赵佳妮, 吴兆麟. 基于灰色马尔可夫模型的水上交通事故预测 [J]. 大连海事大学学报, 2005, 31(4): 15-18.

[3] 田俊改, 许红军. 基于灰色马尔可夫链的航空货邮预测 [J]. 中国民航大学学报, 2009, 27(1): 35-38.

[4] 邓聚龙. 灰预测与灰决策 [M]. 武汉: 华中科技大学出版社, 2002.

[5] 何文章, 宋国乡, 吴爱弟. 估计 GM(1, 1) 模型中参数的一族算法 [J]. 系统工程理论与实践, 2005, 25(1): 69-75.

[6] 刘斌, 赵亮, 翟振杰, 等. 优化的 GM (1,1) 模型及其适用范围 [J]. 南京航空航天大学学报, 2003, (4): 451- 454.

[7] 刘斌, 刘思峰, 翟振杰, 等. GM (1,1) 模型时间响应函数的最优化 [J]. 中国国管理科学, 2003, (4): 54-57.

[8] 张大海, 江世芳, 史开泉. 灰色预测公式的理论缺陷及改进 [J]. 系统工程理论与实践, 2002, 22( 8): 140-142.

[9] 刘思峰, 党耀国, 方志耕. 灰色系统理论及其应用 [M]. 北京: 科学出版社, 2004.

- [10] Meyn, Tweedie. Markov chains and stochastic stability[M]. Beijing: World Book Publishing Company, 1999.
- [11] 周维新. 交通事故灰色模型的研究 [J]. 西安公路交通大学学报, 2000, 20(2): 32- 39.
- [12] 阳吉宝. 运用灰色马尔柯夫链模型预测地下水位 [J]. 工程勘察, 1994, (5): 44-46.
- [13] 翟雅峤, 翁剑成, 等. 多源数据融合的区间车辆速度预测算法研究 [J]. 交通信息与安全, 2009, 3(27): 74-77.

## Mobile Communications Market Forecasting Based on Gray-weighted Markov Model

WANG Zeng-min, WANG Kai-jue

(School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** Based on the traditional method of using gray prediction model to forecast tendency, the paper gives Gray-Weighted Markov chain Model for forecasting mobile communications Market. Introduce the method and steps of using this model and then take prediction of Chinese mobile communications market as an example. Gray-weighted Markov chain model has not only taken into consideration evolution laws of data from time series, but also considered random fluctuations in data by normalizing autocorrelation coefficient as weight and adopting converting of transition probability matrix, which is rigorous scientific method and could be well applied to the Chinese mobile communications market forecasting.

**Keywords:** gray prediction model; weighted Markov chain; mobile communications market; forecast