

一种新的灰色马尔科夫预测模型及其应用

杜雯

(鄂东职业技术学院,湖北 黄冈 438000)

摘要 以奥运会奖牌数预测的实际问题为背景,结合灰色系统理论与马尔科夫预测的相关理论与方法,提出了一种新的奥运会奖牌数预测的灰色马尔科夫预测模型,从而为未来奥运会奖牌数预测提供了一种有效的方法。

关键词 奥运会奖牌数预测;马尔科夫链;灰色马尔科夫预测模型

中图分类号 O211.62 **文献标识码** A **文章编号** 1003-8078(2013)03-0055-04

收稿日期 2013-04-14 **doi** 10.3969/j.issn.1003-8078.2013.03.16

作者简介 杜雯,女,湖北英山人,助教,硕士研究生。

A novel grey-Markov prediction model and its application

DU Wen

(Edong Institute of Vocation and Technology, Huangzhou 438000, Hubei, China)

Abstract Based on the practical issue of predicting the Olympic medal number, a new grey Markov prediction model is presented by using grey system theory and Markov chain. The prediction method can be used to predict the medal number of the future Olympic Games.

Key words prediction of Olympic medal number; markov chain; grey-Markov prediction model

现代奥林匹克运动会是展示世界各国最高竞技运动水平的舞台,其所获奖牌数及国家排名不仅是一个国家体育运动竞技水平的标志,也是国家经济政治和综合实力的体现^[1-2]。奥运会奖牌预测问题也是经济学领域研究的热点问题之一,其理论发展一直受到关注,产生了众多预测方法^[3-6],如趋势直线外推法、统计回归方法、色预测模型、时间序列预测法、Markov方法等。

本文将灰色预测与马尔科夫预测模型相结合,提出了一种新的灰色马尔科夫预测模型,并通过对奥运会奖牌数进行分析预测,结果表明该模型能够降低运算量,提高预测水平,并克服了传统奥运会奖牌数预测模型的不足。

1 灰色马尔科夫预测模型

灰色马尔科夫预测模型的基本思路是:先建立灰色GM(1,1)模型,求出其预测曲线;再以平滑的预测曲线为基准划分若干动态的状态区间,计算出马尔科夫预测未来状态,从而得出预测值区间,取区间中点,最终得到精度较高的预测值。

(1) GM(1,1)模型

① 原始数据的累加处理

设原始数据序列为 $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)$,记为:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$$

对序列 $x^{(0)}$ 作一次累加处理,得到新的生成序列为:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)), k = 1, 2, \dots, n$$

其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$ 。

累加生成序列克服了原始数据序列的随机性和波动性,将原始数据序列转化成规律性较强的递增数据序列,其目的是为建立微分方程形式的预测模型作好准备。

② GM(1,1)模型的建立

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

微分方程(1)就是灰色 GM(1,1)预测模型^[4-6],其中 a, u 为常数,可通过最小二乘法拟合得到:

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

其中

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{bmatrix}, Y_n = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)),$$

可求得微分方程(1)的解(称为时间响应函数)为:

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a} \quad (2)$$

(2)式就是累加生成序列 $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ 的预测公式,它是对一次累加生成数据序列的预测值,可通过下列(3)式求出原始数据序列的还原预测值

$$\hat{y}(k) = \hat{x}^{(0)}(k+1) = (x^{(1)}(k+1) - x^{(1)}(k)) \quad (3)$$

(2) 马尔科夫预测

对于一个符合 n 阶马尔科夫非平稳随机序列 $y(k)$, 其任一状态 Q_k 可表示为 $Q_k \in [Q_{1k}, Q_{2k}]$, 其中 $Q_{1k} = \hat{y}(k) + a_k \bar{y}$, $Q_{2k} = \hat{y}(k) + b_k \bar{y}$, $k = 1, 2, \dots, m$, \bar{y} 为原始数据的均值。根据所研究对象的实际意义和样本数据的多少来确定 Q_k 的含义, 以及状态划分的数目和常数 a_k, b_k 的值。由状态 Q_i 经过 k 步转移到 Q_j 的次数记为 $n_{ij}(k)$, 状态 Q_i 出现的次数为 n_i , 则由状态 Q_i 经过 k 步转移到 Q_j 的转移概率为 $p_{ij}(k) = \frac{n_{ij}(k)}{n_i}$, 则 k 步转移矩阵记为 $P = (p_{ij}(k))_{m \times m}$, 于是系统未来时刻最可能的预测值为

$$Y^*(t) = \frac{Q_{1k} + Q_{2k}}{2} = \bar{y}(t) + \frac{\bar{y}}{2}(a_i + b_i)。$$

2 灰色马尔科夫预测模型在伦敦奥运会奖牌数预测中的应用

根据第 24 届至第 29 届奥运会奖牌榜情况, 选取多次出现在奖牌榜前 10 名的 5 个国家^[7], 并重点参考近两届的排名情况, 其基本情况如表 1 所示(其中俄罗斯数据的第 25 届为独联体的数据, 第 24 届为前苏联的数据, 德国第 24 届的数据为民主德国和联邦德国的奖牌数和)。

对于表 1 中的相关数据, 依次对中国、美国、俄罗斯等国家编号 $i = 1, 2, \dots, 5$, 将剩余的国家和地区的综合编号 $i = 10$, 由此确定了所要研究的 10 个国家或地区。

在计算奖牌数比例时必须考虑到“东道主效应”^[8]。竞技体育中的“东道主效应”是指运动员在自己的家乡参加比赛要比在其他地方参加比赛能取得更好的成绩。邓运龙统计了 1992-2000 年这 3 届奥运会中, 东道主举办当届、上一届、再上一届和后一届获得金牌、奖牌排名情况, 论证了夏季奥运会确实存在着东道主效应^[9]。对于“东道主效应”的测算有很多方式, 选用参考文献^[1]中简单平均法综合所有各届奥运会东道主效应, 所得结果是: 金牌数的东道主效应增幅为 11.31%, 整体实力的东道主效

应增幅为 11.71%。而第 30 届奥运会时中国队是客场作战,因此需要在每届奥运会中去掉东道主效应对奖牌数的影响。选取东道主效应指数 $\alpha = 0.1171$,则去除东道主效应后东道主奖牌数 $n_i' = \frac{n_i(t)}{1 + \alpha}$ 。

表 1 第 24 届至第 29 届奥运会奖牌榜

| 国家 | 第 29 届 | 第 28 届 | 第 27 届 | 第 26 届 | 第 25 届 | 第 24 届 |
|-----|--------|--------|--------|--------|--------|--------|
| 中国 | 1 | 2 | 3 | 4 | 4 | 11 |
| 美国 | 2 | 1 | 1 | 1 | 2 | 3 |
| 俄罗斯 | 3 | 3 | 2 | 2 | 1 | 1 |
| 德国 | 5 | 6 | 5 | 3 | 3 | 7 |
| 英国 | 4 | 10 | 10 | 36 | 13 | 12 |

在所研究的数据中有东道主效应的国家有第 26 届的美国,第 27 届的澳大利亚,第 29 届的中国,奖牌数分别为 101,58,100 枚,按 $n_i' = \frac{n_i(t)}{1 + \alpha}$ 折合后奖牌数分别为 91,52,90 枚,将其折合后统计每届中各国奖牌数见表 2 所示。

表 2 去除东道主效应后的奖牌数统计表

| 编号 | 国家 | 第 24 届 | 第 25 届 | 第 26 届 | 第 27 届 | 第 28 届 | 第 29 届 |
|----|-----|--------|--------|--------|--------|--------|--------|
| 1 | 中国 | 28 | 54 | 50 | 59 | 63 | 90 |
| 2 | 美国 | 94 | 108 | 91 | 97 | 103 | 110 |
| 3 | 俄罗斯 | 132 | 112 | 63 | 88 | 92 | 72 |
| 4 | 德国 | 142 | 82 | 65 | 57 | 48 | 41 |
| 5 | 英国 | 24 | 20 | 15 | 28 | 30 | 47 |

根据表 2 中各国家或地区奖牌数 $n_i(t)$ 与总奖牌数 $N(t)$,由式 $w_i(t) = \frac{n_i(t)}{N(t)}$, 计算得到各国家或地区奖牌数比例如表 3。

表 3 各国家或地区奖牌数占总奖牌数的比例

| 编号 | 国家 | 第 24 届 | 第 25 届 | 第 26 届 | 第 27 届 | 第 28 届 | 第 29 届 |
|----|-----|--------|--------|--------|--------|--------|--------|
| 1 | 中国 | 0.038 | 0.066 | 0.060 | 0.064 | 0.068 | 0.094 |
| 2 | 美国 | 0.127 | 0.132 | 0.109 | 0.106 | 0.111 | 0.116 |
| 3 | 俄罗斯 | 0.179 | 0.137 | 0.076 | 0.096 | 0.099 | 0.076 |
| 4 | 德国 | 0.192 | 0.100 | 0.078 | 0.062 | 0.052 | 0.043 |
| 5 | 英国 | 0.032 | 0.025 | 0.018 | 0.031 | 0.032 | 0.050 |

下面运用上文提出的灰色马尔科夫预测模型对第 30 届伦敦奥运会中国获奖牌所占比例进行预测。

Step1:构造奖牌比例序列

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(6)) = [0.038, 0.066, 0.060, 0.064, 0.068, 0.094]$$

对 $x^{(0)}$ 进行一次累加生成,得生成序列

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(6)) = [0.038, 0.014, 0.164, 0.228, 0.296, 0.39]$$

Step2:构造数据矩阵 B 及数据向量 Y

$$B = \begin{pmatrix} -0.071 & 1 \\ -0.134 & 1 \\ -0.196 & 1 \\ -0.262 & 1 \\ -0.342 & 1 \end{pmatrix} \qquad Y = \begin{pmatrix} 0.066 \\ 0.060 \\ 0.064 \\ 0.068 \\ 0.094 \end{pmatrix}$$

Step3:计算 \hat{a}

$$\hat{a}(B^TB)^{-1}B^TY = \begin{pmatrix} -0.099 & 56 \\ 0.0503 & 88 \end{pmatrix} = (a, u)^T$$

于是得到

$$a = -0.099\,56, u = 0.0503\,88$$

Step4: 建立模型, 得预测方程

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a} = 0.544\,107e^{0.09956k} - 0.506\,107$$

$$\hat{y}(k) = \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = 0.056\,96e^{0.09956k}.$$

Step5: 以 $\hat{y}(k)$ 曲线为基准, 划分成与 $\hat{y}(k)$ 曲线平行的三个区域, 每一个区域构成了一个状态如下图 1 所示。

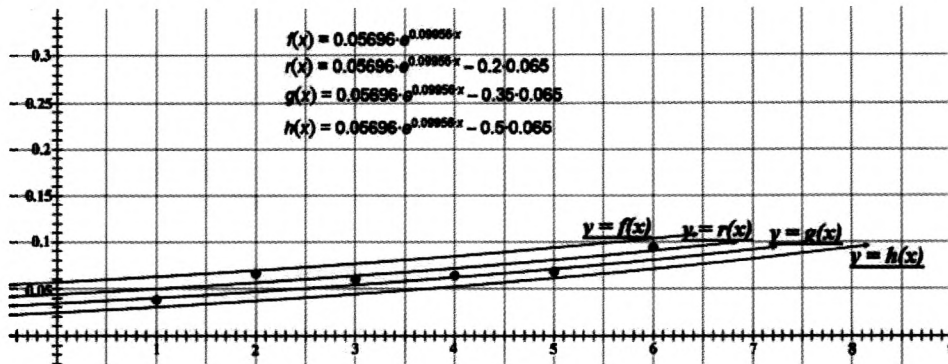


图 1 预测方程状态图

$$Q_1 : Q_{11} = \hat{y}(k), Q_{12} = \hat{y}(k) - 0.2\bar{y}$$

$$Q_2 : Q_{21} = \hat{y}(k) - 0.2\bar{y}, Q_{22} = \hat{y}(k) - 0.35\bar{y}$$

$$Q_3 : Q_{31} = \hat{y}(k) - 0.35\bar{y}, Q_{32} = \hat{y}(k) - 0.5\bar{y}$$

其中 $\hat{y}(k)$ 为第 k 年的预测值, \bar{y} 为历年奖牌比率的平均值 0.065。由图 1 可知, 落入 Q_1, Q_2, Q_3 三个状态的原始数据的样本点数分别为 $n_1 = n_2 = n_3 = 2$ 。由状态 Q_1 一步转移到状态 Q_1, Q_2 和 Q_3 的原始数据样本点数分别为 $n_{11} = 1, n_{12} = 1, n_{13} = 0$; 由状态 Q_2 一步转移到状态 Q_1, Q_2 和 Q_3 的原始数据样本点数分别为 $n_{21} = 0, n_{22} = 1, n_{23} = 1$; 由状态 Q_3 一步转移到状态 Q_1, Q_2 和 Q_3 的原始样本点数分别为 $n_{31} = 0, n_{32} = 1, n_{33} = 1$ 。故可得一步转移概率矩阵为

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

根据该矩阵可以判定 2012 年中国所获奖牌率最有可能处于状态 Q_3 , 即可能在灰区间 $[Q_{31}, Q_{32}]$ 内, 因此 2012 年的中国所获奖牌率预测值

$$Y^*(7) = \frac{Q_{1k} + Q_{2k}}{2} = \hat{y}(6) + \frac{\bar{y}}{2}(0.35 + 0.5) = 0.098\,37$$

同样用此方法可依次计算出各个国家或地区奖牌比率。再根据各届的奖牌总数假设下届的奖牌总数为 960 枚, 最后求得第 30 届奥运会各国家或地区奖牌比例和数量情况如表 4 所示。

表 4 第 30 届伦敦奥运会各国家或地区奖牌比例和数量预测值和实际值比较

| 国家 | 中国 | 美国 | 俄罗斯 | 德国 | 英国 |
|---------|--------|--------|--------|--------|--------|
| 奖牌比例 | 0.098 | 0.110 | 0.081 | 0.043 | 0.064 |
| 奖牌数量预测值 | 94 | 105 | 78 | 42 | 62 |
| 奖牌数量实际值 | 88 | 104 | 82 | 44 | 65 |
| 平均相对误差 | 6.818% | 0.962% | 4.878% | 4.545% | 4.615% |

(下转第 68 页)

远的教育价值。本文从教育游戏的现状出发,提出了教育融入游戏的途径,并以一个教育游戏的开发为案例,展示了 Virtools 技术应用教育游戏开发的部分重要环节。从中可以看出,应用 Virtools 技术开发教育游戏,快捷建立逼真的虚拟场景,实现对文字、声音、图形图像、动画和三维模型等对象操作,完成虚拟世界的碰撞检测和动态交互,能适时融入教育理念和教育内容,使游戏具有趣味性、挑战性和教育性,让学习者有身临其境的感觉,在游戏世界中达成教育目标。

参考文献:

- [1] 钱莉,金科,章苏静. 教育游戏中学习者注意力资源开发策略[J]. 远程教育杂志, 2010(06): 93-97.
- [2] 李彤彤,马秀峰,张明娟. 教育游戏在小学英语教学中的应用模式研究[J]. 中国电化教育, 2010(01): 90-93.
- [3] 爱迪斯通. <http://www.virtools.com.cn/> Virtools User Guide[DB/CD]. 北京:爱迪斯通科技有限公司, 2011.
- [4] 张雪鹏,陈国华,戴莺莺,等. 基于3D的虚拟运动仿真平台设计及 Virtools 功能实现[J]. 北京化工大学学报(自然科学版), 2009(04): 1219-1222.
- [5] 刘明坤. 三维游戏设计师宝典—Virtools 开发工具篇[M]. 成都:四川电子音像出版中心, 2005: 20-58.
- [6] 方利伟. 基于 Virtools 的三维虚拟实验室研究与实现[J]. 实验技术与管理, 2010(05): 83-86.
- [7] 刘志广,武真,张永策,等. 基于 Virtools 的动态 GC 虚拟实验室的设计与实现[J]. 计算机工程与设计, 2009(05): 93-95.
- [8] 杨王黎,王惠影. 基于 Virtools 的注聚站虚拟培训系统的设计与实现[J]. 计算机工程与科学, 2009(01): 134-137.
- [9] 朱琦,李兆君. 基于 Game Maker 的教学游戏的设计与研究[J]. 现代教育技术, 2010(02): 124-126.

(张所滨)

(上接第58页)

由表4可以看出,整体的预测效果比较好,预测的平均相对误差均在7%以内。因此,可运用该预测模型来预测第31届里约热内卢奥运会世界各国的奖牌数,预测值的可信度将比较高。

本文将灰色理论与马尔科夫预测相结合,提出了灰色马尔科夫预测模型。灰色GM(1,1)预测所需信息较少,计算简便,但由于GM(1,1)模型的解为指数型曲线,其预测的几何图形是一条较平滑曲线,因而对波动性较大的数据列拟合较差,预测精度较低。而马尔科夫模型适合于波动性较大的数据列预测问题,但它要求研究对象具有平稳过程等特点。因此,本文将两者取长补短结合起来,形成了灰色马尔科夫预测模型,该模型可有效提高随机波动较大数据列的预测精度。

参考文献:

- [1] 王国凡,唐学峰. 奥运会奖牌预测国内、外研究动态及发展趋势[J]. 中国体育科技, 2009, 45(6): 130-132.
- [2] 王宇鹏,许健. 奥运会奖牌榜影响因素的实证分析[J]. 统计研究, 2008, 25(10): 57-62.
- [3] 张正民. 奥运会男子100m跑成绩的灰色GM(1,1)模型预测[J]. 体育学刊, 2011, 18(4): 112-114.
- [4] 肖新平,宋中民,李峰. 灰技术基础及应用[M]. 北京:科学出版社, 2005.
- [5] 邓聚龙. 灰预测与灰决策[M]. 武汉:华中科技大学出版社, 2002, 60-65.
- [6] Rao C J, Peng J, Li C F, Li W. Group decision making model based on grey relational analysis[J]. The Journal of Grey System, 2009, 21(1): 15-24.
- [7] 张玉超. 第29届奥运会中、美、俄国奖牌分布特点及其启示[J]. 体育学刊, 2009, 16(2): 81-84.
- [8] 吴殿延,吴颖. 2008北京奥运会中国金牌赶超美国的可能性[J]. 统计研究, 2008, 25(3): 61-64.
- [9] 张海波,赵焕成. 北京奥运会中国军团金牌数的预测[J]. 统计与决策, 2008, 15: 76-77.

(张所滨)