

Forecast of Overseas Students Based on Grey Prediction and Improvement Model

Tianhao Jin, Shaoling Ding

College of Science, Guilin University of Technology, Guilin Guangxi
Email: 839760790@qq.com

Received: Aug. 6th, 2019; accepted: Aug. 16th, 2019; published: Aug. 26th, 2019

Abstract

By comparing and analyzing the grey forecasting model with the improved grey forecasting model based on moving average, this paper establishes a fitting and forecasting model for the number of overseas students and returnees in China. The results show that the grey prediction model smoothed by moving average method is better than the grey prediction method alone in this empirical analysis. The new method improves the fitting accuracy, facilitates relevant state institutions to obtain the trend of studying abroad and the number of returnees, and gives priority to formulating relevant policies.

Keywords

Grey Prediction, Moving Average Method, The Number of Students Studying Abroad, The Number of Returned Students Studying Abroad

基于灰色预测及改进模型的留学人员预测

金天昊, 丁少玲

桂林理工大学理学院, 广西 桂林
Email: 839760790@qq.com

收稿日期: 2019年8月6日; 录用日期: 2019年8月16日; 发布日期: 2019年8月26日

摘要

本文通过对灰色预测模型与使用移动平均改进后的灰色预测模型进行比较和分析,建立了我国出国留学人数以及学成归国人数的拟合及预测模型。结果表明,在本实证分析中,使用移动平均法修匀的灰色预测模型预测效果优于单独使用灰色预测法。新方法提高拟合精度,便于国家有关机构获取出国留学以及

归国人数的趋势, 优先制定相关政策。

关键词

灰色预测, 移动平均法, 出国留学人数, 学成归国留学人员

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着近些年留学热潮的持续升温, 留学现象以及留学归国情况也成为引起学术界关注的经济社会现象之一。而对于未来我国出国留学人数以及留学归国人员走向的准确预测, 毫无疑问, 对把握人才流动趋势、经济科技趋势, 推动教育、经济改革, 政府因地制宜的制定相关教育政策等起到重要作用。

当前 CNKI 收录文献涉及留学研究的文章对于留学生人员的预测方面研究较少, 现有论著包括使用时间序列[1]或利用灰色预测法及其改进模型[2]预测我国出国留学人数。针对灰色预测模型及其改进形式, 陈有为使用指数平滑模型改进离散 GM (1, 1)模型, 但并未利用到留学数据中[3], 柯普使用 GM (1, 1)对出国人数进行预测[4]。2010 年卫彦琼[5]使用灰色神经网络模型对 1999 年到 2006 年我国出国留学学生总人数以及留学生学成归来人数进行了实证分析。但笔者在对比作者使用此模型进行预测 2007 年至 2011 年的数据与国家统计局已经发布的真实数据来看, 偏差较大。因此, 使用灰色预测模型针对留学数据进行预测已有先例但预测效果还有改进空间。

2. 方法介绍

2.1. 灰色预测及 GM (1, 1)模型基本概念与具体方法

本文将灰色预测理论、指数平滑和移动平均法以及改进后的灰色预测模型应用于我国留学生规模以及学成归来学生人数预测, 为我国留学生规模预测提供有效的模型, 从而为政府相关决策提供数据支持。

灰色理论是由我国学者邓聚龙于 20 世纪 80 年代前期提出, 用于控制和预测的新理论和新技术。灰色预测法是一种对含有不确定因素的系统进行预测的方法。灰色系统内的一部分信息是已知的, 另一部分信息是未知的。其中, 灰色预测 GM (1, 1)模型因其所需信息少、运算方便、建模精度较高而被广泛应用于各种预测领域。其理论已初步形成了较为完善的模型、方法和技术体系, 基于灰色建模理论的灰色数列预测方法经过 20 多年的发展, 已经在人口规模、交通事故、农业、社会经济、旅游、生态足迹等方面得到了广泛的应用[6], 且均获得较为理想的预测结果。

灰色预测通过鉴别系统因素之间发展趋势的相异程度, 即进行关联度分析, 并对原始数据进行累加和累减的处理来寻找系统变动的规律。首先生成有较强规律性的数据序列, 然后建立相应的微分方程模型, 从而预测事物未来的发展趋势状况。

首先得到生成列, 为弱化原始时间序列的随机性, 需要进行数据处理即产生生成列, 本文叙述累加法生成(累加即将原始序列通过依次叠加得到生成列)。记原始时间序列为:

$$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), X^{(0)}(3), X^{(0)}(4), \dots, X^{(0)}(n)\} \quad (1)$$

上标为 0 表示原始时间序列, 通过累加构造生成列为:

$$X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), X^{(1)}(3), X^{(1)}(4), \dots, X^{(1)}(n)\} \quad (2)$$

对于非负数据。累加次数越多, 则随机性弱化越多, 当累加次数足够大后, 可认为时间序列已由随机序列变为非随机序列了。一般随机序列的多次累加序列, 大多可用指数曲线逼近。

之后需要计算关联度。关联度分析是分析系统中各因素关联程度的方法, 在计算关联度之前需要先计算关联系数。关联系数公式为:

$$\eta(k) = \frac{\min \min |\hat{X}^{(0)}(k) - X^{(0)}(k)| + \rho \max \max |\hat{X}^{(0)}(k) - X^{(0)}(k)|}{|\hat{X}^{(0)}(k) - X^{(0)}(k)| + \rho \max \max |\hat{X}^{(0)}(k) - X^{(0)}(k)|} \quad (3)$$

式中具体含义此处不予赘述。

在算出 $\hat{X}^{(0)}(k)$ 序列与 $X^{(0)}(k)$ 序列的关联系数后, 计算各类关联系数的平均值:

$$r = \frac{1}{n} \sum_{k=1}^n \eta(k) \quad (4)$$

这个平均值 r 被称为关联度。为建立 GM (1, 1)模型, 设时间序列有 n 个观察值, 通过累加得到序列, 则相应的 GM (1, 1)模型的微分方程为

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu \quad (5)$$

式中, a 称为发展灰数 μ 称为内生控制灰数, 设 \hat{a} 为待估计参数向量, 其中 $\hat{a} = \begin{pmatrix} a \\ \mu \end{pmatrix}$, 利用最小二乘法求解得:

$$\hat{a} = (B^T B)^{-1} B^T Y_n \quad (6)$$

解出微分方程, 得到预测模型

$$\hat{X}^{(1)}(k+1) = \left[\hat{X}^{(0)}(1) - \frac{\mu}{a} \right] e^{-ak} + \frac{\mu}{a} \quad (k = 0, 1, 2, \dots, n) \quad (7)$$

2.2. 移动平均模型

移动平均是技术分析中一种分析时间序列数据的工具, 其基本思想是: 根据时间序列资料、逐项推移, 依次计算包含一定项数的序时平均值, 以反映长期趋势。因此移动平均可抚平短期波动, 反映长期趋势或周期。移动平均法可分为简单移动平均、加权移动平均、指数移动平均等。本文以简单移动平均为例展开。简单移动平均是某变数 k 个数值的未作加权算术平均。这种方法是根据过去 k 期的观察值来计算平均数的, 以之作为测算数据进行决策[7]。

若使用移动平均法, 照其定义 $t+1$ 期的预测值为

$$F_{t+1} = \bar{Y}_t = \frac{Y_{t-k+1} + Y_{t-k+2} + \dots + Y_{t-1} + Y_t}{k} \quad (8)$$

式中 Y 为观测值, k 为观测期数。这种方法的优点是计算简单, 缺点是适用范围具有局限性, 一般只适用于观测值发展比较平稳、起伏不大的数据。对于下文实证分析部分数据来说, 留学生数量与留学生归国数量均是递增的序列。使用移动平均法得出后一项的预测值是前 k 项观测值的算术平均值, 导致预测值偏小, 受到该方法适用范围的局限, 因此不能对本文实证分析的原始数据直接使用移动平均法。

2.3. 组合模型的构建

随着灰色预测理论研究的进一步深入, 一些学者发现灰色预测存在一些局限。陈有孝, 林晓言指出灰色模型预测由于其原始数据的起伏性和无序性, 且原始数据的个数有限, 难以将预测带限制在一个较小的范围之内, 导致灰色预测模型在大多数情况之下是粗糙的[8]。

具体来说, 灰色系统模型的一大优点在于不要求通过足够的原始数据建立模型, 然而对于某些具有递增(或递减)趋势的数据, 由于数据量有限, 灰色系统模型偏向于过分提取其递增趋势, 从而高估预测结果。因此笔者尝试利用移动平均的移动平均可削弱原序列的上下波动的特点从而对灰色预测数据进行修匀或平滑, 抑制其较高的递增趋势, 得到准确的预测值。本文尝试使用两种方式进行组合模型的构建。首先使用原移动平均公式的结合灰色预测模型。具体做法为将灰色预测得到的数据使用移动平均法修匀。

结合灰色模型公式与移动平滑公式得到如下:

$$\begin{aligned} F_{t+1} = \bar{Y}_t &= \frac{Y_{t-k+1} + Y_{t-k+2} + \cdots + Y_{t-1} + Y_t}{k} \\ &= \frac{\hat{X}^{(1)}(t-k+1) + \hat{X}^{(1)}(t-k+2) + \cdots + \hat{X}^{(1)}(t-1) + \hat{X}^{(1)}(t)}{k} \end{aligned} \quad (9)$$

值得注意的是, 若使用移动平均的原始公式与灰色预测结合, 在多次测算后发现, 移动平均在改进数据的同时会带有滞后特点, 过分的削弱趋势性。因此为避免移动平均只是在处理水平型历史数据时才有效的特点, 可将移动平均模型进行修改再与灰色预测模型组合, 即将灰色预测得到的数据使用改进后的移动平均法修匀:

$$\begin{aligned} F_{t+1} = \bar{Y}_{t+1} &= \frac{Y_{t-k+2} + Y_{t-k+3} + \cdots + Y_t + Y_{t+1}}{k} \\ &= \frac{\hat{X}^{(1)}(t-k+2) + \hat{X}^{(1)}(t-k+3) + \cdots + \hat{X}^{(1)}(t) + \hat{X}^{(1)}(t+1)}{k} \end{aligned} \quad (10)$$

其中 $\hat{X}^{(1)}$ 为灰色预测模型得出的结果。

对比两种组合模型, 并通过数据进行实验分析, 发现相比于改进后的组合模型, 未改进的移动平均与灰色预测的组合模型对于灰色预测数据的递增趋势的抑制作用更加强烈, 因此针对递增趋势强、弱的不同数据情况, 可分别使用未改进的组合模型和改进后的组合模型。此外还可以通过调整观测系数 k 对不同数据进行适当的修匀。

3. 实证分析

本文通过对灰色预测模型与使用移动平均改进后的灰色预测模型进行比较和分析, 利用国家统计局自 2006 年至 2018 年的我国留学生相关数据, 分别对出国留学总人数以及留学生学成归来人数进行拟合分析, 并对 2018 至 2024 年的留学人数及学成归来人数进行预测。

3.1. 灰色预测模型及其问题

首先分析留学生样本, 数据呈单调递增的变化趋势所选样本数据列适合 GM(1, 1)建模。根据 GM(1, 1)建模原理, 运用 RStudio 软件计算, 得到发展系数 $a = 0.1161707$ 、灰色作用量 $u = 15.27659$, 灰色预测模型的拟合情况如表 1 所示。

由表 1, GM(1, 1)模型的预测误差较小, 软件运算得到从 2005 年到 2018 年预测值的平均误差 11.39092%。根据预测结果, 进行模型检验: 后验差检验结果为 $p = 1$, C 值 = 0.09690246, 预测模型预测精度好, 可以进行后续预测。2019~2024 年的预测留学生数量如表 2 所示。

Table 1. The forecast table of the number of overseas students based on the grey prediction model GM (1, 1)
表 1. 出国留学人数基于灰色预测模型 GM (1, 1) 的预测表

年份	实际留学生数(万人)	GM (1, 1)模型		
		拟合值(万人)	残差	相对残差
2005	11.8515	11.8515	0	0
2006	13.4	17.65928	-4.259277	0.3178565
2007	14.4	19.83468	-5.43468	0.3774084
2008	17.98	22.27807	-4.298066	0.239047
2009	22.93	25.02245	-2.092446	0.09125365
2010	28.47	28.1049	0.3651002	0.01282403
2011	33.97	31.56707	2.402927	0.07073673
2012	39.96	35.45574	4.504256	0.1127191
2013	41.39	39.82345	1.56655	0.03784853
2014	45.98	44.7292	1.250798	0.02720308
2015	52.37	50.23928	2.130717	0.04068583
2016	54.45	56.42814	-1.978137	0.03632942
2017	60.84	63.37938	-2.53938	0.04173866
2018	66.21	71.18693	-4.97693	0.07516886

Table 2. The forecast table of the number of overseas students and returned students based on the grey prediction model GM (1, 1)

表 2. 出国留学学生以及归国留学生数基于灰色预测模型 GM (1, 1) 的预测表

年份	预测留学生数(万人)	预测学成归国留学生数(万人)
2019	79.95627	76.24983
2020	89.80589	89.13612
2021	100.8689	104.2002
2022	113.2946	121.8101
2023	127.2511	142.3962
2024	142.9269	166.4612

由于数据同样呈单调递增的变化趋势因此使用相同的方法可以分析留学生学成归国人数。根据 GM (1, 1) 建模原理, 计算得出发展系数 $a = 0.1561495$ 、灰色作用量 $u = 8.706992$ (拟合表略)。

使用软件得到从 2005 年到 2018 年预测值的平均误差为 44.93455%、相对精度 = 88.60908%, 效果较差。根据结果, 进行模型检验显示, 后验差检验结果为 $p = 1$, C 值 = 0.09690246, 预测模型预测精度高。继续预测得出 2019~2024 的预计归国留学生数量如上表所示(表 2)。

对于本实证分析来说, 灰色预测存在的问题较为明显。在数据拟合时可以发现, 不管对于留学生数据还是对于留学归国人数的拟合, 2016 年之后的预测值均较实际值偏大, 残差为负值, 且残差绝对值逐渐递增, 说明此灰色模型预测值呈现出过分的递增趋势。而且这种趋势可能将扩大, 在折线图(图 1)中也能明显看出拟合值的指数趋势。

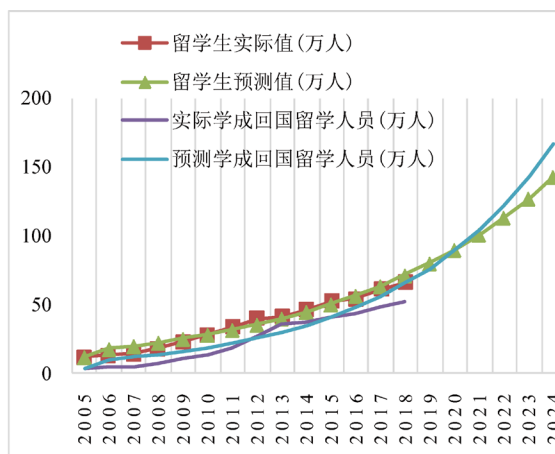


Figure 1. Gray prediction of the number of overseas students and returnees

图 1. 留学生人数及归国人数灰色预测效果图

此外, 学成归国的留学生的灰色预测值在 2021 年之后超过了预测留学生人数(如图 1), 这在现实中很难得到解释。经过分析, 此现象的原因是很可能是预测归国人数时递增趋势经灰色模型过度放大导致超出留学人员的预测值。因此综合本文的预测结果分析, 本数据灰色预测数据指数性过强, 预测效果一般。需要其他方法对此进行改进。

3.2. 移动平均法与灰色预测组合方法

联想到移动平均对序列有修匀或平滑的作用, 使得原序列的上下波动被削弱了, 并结合本例的特点, 在灰色预测拟合结果大大超出了实际值, 预测值呈现较强的指数趋势情况下, 尝试使用移动平均法修匀灰色预测结果, 对两组数据使用改进后的移动平均法(两期), 进行修匀并计算相应 MSE, 与灰色预测模型拟合值进行比较得到如下表格(表 3)。

由表 3 可知对于留学生人数的预测来说, 虽然 MSE 略有上升, 但上升的范围很小, 且最近几期的(2016 年~2018 年)的拟合值明显优于单独使用灰色预测的拟合值, 说明过分的指数趋势被平滑和修正, 且效果较好, 对于未来几期的预测效果优于使用单一的灰色预测。对留学归国人数的拟合, 同样的, 表中信息表明 2016 年~2018 年的拟合值更好, 避免了不合理的过分上升趋势。而且 MSE 得到了下降, 说明使用本方法的效果很好, 经过修改的移动平均的对本文两组数据的拟合效果皆有提升。

此外, 大量试验发现, 单独对于留学归国人员数据而言, 使用未修改的移动平均公式结合灰色预测的效果会让 MSE 大幅减小(MSE = 5.4135625)。但若使用此方法拟合留学人数会让对应的 MSE 大大增加(MSE = 46.913)。猜想这样的现象发生的是由于回国留学生的数据变化较为剧烈, 灰色预测的预测指数性过强, 上升趋势极其明显, 需要使用未改进后的移动平均才能更有效的抑制其过分的趋势性, 而留学生数据灰色预测值的指数性相对较弱, 若使用原始的移动平均, 会将趋势性过分抑制, 预测效果下降。

3.3. 预测结果

本文对出国留学学生总人数, 留学生学成归来人数做出预测, 分别使用修改的移动平均和未修改的移动平均修匀灰色预测得到如下预测值, 整合为表 4 及图 2。

列表数据可以看出对上述两组数据来说, 移动平均都有修匀作用, 使强烈的递增趋势得到控制(见图 2), 这符合当前实际情况——留学生多年的快速增加使留学市场接近饱和状态。且改进后的预测留学归国人数小于改进后的预测留学生数量, 这也是合理的, 因此判断预测效果得到改进。

Table 3. Grey moving average method for fitting table of the number of overseas students and returnees
表 3. 灰色移动平均法对留学生及归国人数拟合表

年份	留学生数量(万人)				留学生归国数量(万人)			
	真实值	灰色预测拟合值	灰色移动平均	灰色移动平均残差平方和	真实值	灰色预测拟合值	灰色移动平均	灰色移动平均残差平方和
2005	11.8515	11.8515	#N/A	#N/A	3.4987	3.4987	#N/A	#N/A
2006	13.4	17.65928	14.75539	1.837082	4.2	10.01488	6.75679	6.537175
2007	14.4	19.83468	18.74698	18.89624	4.4	11.7074	10.86114	41.74633
2008	17.98	22.27807	21.05638	9.464083	6.93	13.68597	12.69669	33.25466
2009	22.93	25.02245	23.65026	0.518774	10.83	15.99891	14.84244	16.09967
2010	28.47	28.1049	26.56368	3.634075	13.48	18.70274	17.35083	14.98329
2011	33.97	31.56707	29.83599	17.09008	18.62	21.86352	20.28313	2.766001
2012	39.96	35.45574	33.51141	41.58438	27.29	25.55847	23.711	12.80928
2013	41.39	39.82345	37.6396	14.06554	35.35	29.87788	27.71818	58.24475
2014	45.98	44.7292	42.27633	13.71721	36.48	34.92727	32.40258	16.62539
2015	52.37	50.23928	47.48424	23.87065	40.91	40.83001	37.87864	9.189143
2016	54.45	56.42814	53.33371	1.246103	43.25	47.73031	44.28016	1.06123
2017	60.84	63.37938	59.90376	0.876545	48.09	55.79678	51.76355	13.49493
2018	66.21	71.18693	67.28316	1.151662	51.94	65.22649	60.51164	73.47293
MSE	10.0296 (灰色预测)		10.56803 (平均后)		34.90909 (灰色预测)		23.09883 (平均后)	

Table 4. Grey moving average method for forecasting the number of overseas students and returnees
表 4. 灰色移动平均法对留学生及归国人数预测表

年份	预测留学生人数(万人)		预测学成归国留学生人数(万人)	
	原始灰色预测	修改的移动平均	原始灰色预测	未修改的移动平均
2019	79.95627	#N/A	76.24983	#N/A
2020	89.80589	84.88108	89.13612	#N/A
2021	100.8689	95.3374	104.2002	82.69298
2022	113.2946	107.0818	121.8101	96.66816
2023	127.2511	120.2729	142.3962	113.0052
2024	142.9269	135.089	166.4612	132.1032

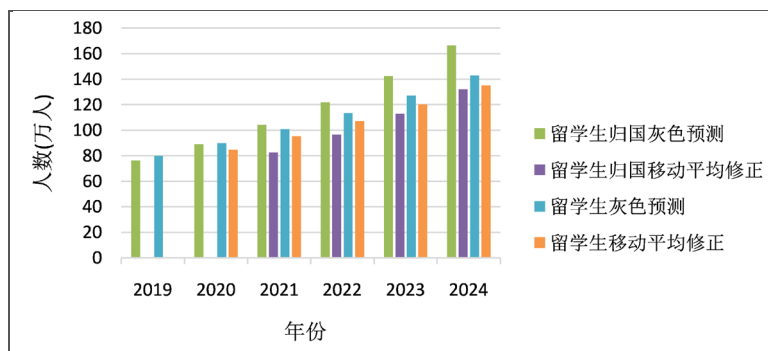


Figure 2. Grey moving average method for forecasting the number of overseas students and returnees

图 2. 灰色移动平均法对留学生及归国人数预测条形图

4. 结论

本文通过对灰色预测模型与使用移动平均改进后的灰色预测模型进行比较和分析, 利用从 2006 年至 2018 年的我国留学生相关数据, 分别对出国留学学生总人数以及留学生学成归来人数进行拟合分析, 并对 2018 至 2024 年的留学人数及学成归来人数进行预测。

首先使用灰色预测法, 发现对本例的数据来看预测拟合效果一般。对于实际情况来说, 预测数据有较强的上升趋势, 而且预测值增速过快, 其中某些结果不符合实际意义。因此, 利用此预测模型进行预测, 要持谨慎态度, 并需要对其方法进行改进。

将移动平均(二期)进行修改并与灰色预测模型结合起来, 发现对本文两组数据都有较好的修匀作用。对比发现, 修改后的移动平均对留学生数据有更好的效果, 未修改的移动平均对留学生归国数据有更好的效果。最终完成预测(表 4), 结果符合实际情况, 且效果优于单独使用灰色预测法。

致 谢

感谢审稿人提出的建议。

基金项目

桂林理工大学 2019 年本科教学建设项目与改革(jxzh201926)。

参考文献

- [1] 杨彩华, 段玲晓. 我国出国留学人数的预测——基于时间序列预测法[J]. 科技经济市场, 2016(1): 119-121.
- [2] 林春艳. 一种新的灰色预测模型及其在留学管理中的应用[C]//中国管理科学学术会议, 2001.
- [3] 陈有为. 基于离散 GM 模型和指数平滑模型组合的统计预测方法[J]. 统计与决策, 2015(10): 74-76.
- [4] 柯普, 吴广. 基于 GM(1,1)模型的出国留学人数预测研究[J]. 价值工程, 2012, 31(25): 318-319.
- [5] 卫彦琼. 基于灰色神经网络模型的留学生和学成归国人员预测[J]. 商业经济, 2010(4): 4-6.
- [6] 赵琳琳. 灰色马氏链预测方法及其在水文序列中的应用研究[D]: [硕士学位论文]. 南京: 河海大学, 2007.
- [7] 李云刚. 基于移动平均法的改进[J]. 统计与决策, 2009(19): 158-159.
- [8] 陈有孝, 林晓言. 灰色-马尔可夫链改进的预测方法[J]. 统计与决策, 2005(8): 36-38.

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询; 或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sa@hanspub.org