

doi:10.16055/j.issn.1672-058X.2022.0003.009

基于 L1 范数的出国留学人数组合预测研究

胡 珍

(湖北工业大学 理学院, 武汉 430068)

摘 要:为更准确地探究我国出国留学人数变化趋势,提出基于 L1 范数的组合预测模型,对出国留学人数进行预测;从多角度选取影响出国留学的因素,利用灰色关联度分析提取影响出国留学人数的典型因子,进而构建 GM(1,3)模型;建立 BP 神经网络模型;提出基于 L1 范数组合预测模型,通过求解线性规划确定单一模型最优权系数;然后,对 2006—2019 年出国留学人数进行预测;选取 GM(1,1)模型为对照模型,通过对照模型以及预测误差评价指标体系比较模型的预测精度,结果表明:基于 L1 范数的组合预测模型效果优于 3 个单一模型,有效地提高了预测精度,能够充分利用单一预测模型提供的信息,从而更加准确地预测出国留学人数;未来几年我国出国留学规模仍有较大的发展空间,预测结果可为全球疫情下我国留学相关工作提供参考。

关键词:GM(1,3)模型;BP 神经网络模型;L1 范数;组合预测;出国留学人数

中图分类号:F224.0

文献标志码:A

文章编号:1672-058X(2022)03-0061-09

0 引 言

留学是我国发展对外开放教育和现代化经济以及实施人才强国战略的重要途径之一,同时也是我国高等教育规模扩张的一个重要过程。2020 年《意见》的印发,强调出国留学仍是为我国提供现代化建设所需的各级各类人才的重要途径之一,疫情只是暂时性地阻碍了出国留学的发展,要把握疫情防控下的教育改革,坚持对外开放教育理念不动摇,同时积极增加同世界各国教育的密切联系^[1]。科学有效地预测出国留学人数,有助于全面把握出国留学工作的总体大局,并对促进我国未来留学工作发展规划的制定和实施具有重要意义。

近年来已有学者对出国留学的相关问题进行了深入研究。如潘昆峰等^[2]发现影响大学生出国留

学的因素包括宏观与微观两个方面;陈玥等^[3]通过研究出国留学政策的历史变迁得出政府决策、市场需求、科技发展这三方面对出国留学影响较大;魏永长等^[4]通过问卷调查和访谈研究,得出个人主观愿望、家庭状况、外部因素对学生出国留学意愿均有影响;陆根书等^[5]通过提出综合选择模型,分析了影响大学生出国留学的因素。陈扬霖等^[6]对高校学生出国的趋势和影响因素等进行了探究,结果发现出国留学人数逐年增加,且家庭背景、人力特征等因素对出国留学均有影响;柯普等^[7]应用 GM(1,1)模型对出国留学人数进行了预测,结果表明 2011—2015 年间我国留学人数呈快速增长的态势;杨彩华^[8]采用了时间序列预测法中的趋势外推法预测了我国留学人数;冯志平等^[9]利用 ARIMA 模型预测了留学人数以及 granger 因果检验分析了留学人数和居民消费之间的关系,结果发现,随着经济快速

收稿日期:2021-05-09;修回日期:2021-06-23.

作者简介:胡珍(1994—),女,安徽铜陵人,硕士研究生,从事预测与决策分析研究.

发展和居民收入的增加,自费出国留学人数激增。

综合已有研究发现,出国留学相关研究仍存在两个方面的问题:其一,大部分研究都是基于定性的方法探究出国留学影响因素;其二,利用统计模型对出国留学人数进行预测的研究相对有限,且模型和预测的方法比较单一,没有进行模型对比和形成公认的最优预测方法或统计模型。其中柯普利用的 GM(1,1) 模型对预测留学人数有一定的参考价值,但通过分析留学现状以及成因可以发现,随着近年来出国留学人数日趋增加,影响留学的因素也在不断变化,呈现多元化趋势,故对模型的预测精度提出了更高的要求,则前人所用的以单一因素建模的 GM(1,1) 模型可能难以全面反映现代数据的复杂状况,拟合精度也会随之降低,并且选用单一模型预测,可能就要承担一定的决策失误风险。

针对现有研究存在的问题,有必要寻求新的预测方法或模型,对预测留学人数加以研究。因此,提出一种基于 L1 范数的组合预测模型。国外学者 Bates 和 Granger,在 1969 年首次研究了组合预测模型^[10],国内一些学者对其也进行了深入探究^[11-13],且取得了良好的研究成果。组合模型一般能利用各种预测方法所包含的信息进而更加有效地提高预测精度,而组合预测方法的关键在于如何有效地选择加权系数,在实际生活中,通常单一模型在不同时刻的预测精度并不相同,故传统的不变权组合模型可能会降低预测精度,而基于 L1 范数的变权组合模型可以提高传统的组合模型的预测精度。因此,基于 L1 范数对单项预测模型赋权,并建立以最小对数误差为准则的组合预测模型,再通过线性规划求解模型;然后,通过对比不同模型的预测效果,表明基于 L1 范数组合预测模型的有效性^[14];最后运用组合预测模型对我国未来几年的出国留学人数进行预测,为分析疫情背景下出国留学人数的总体趋势和相关部门制定决策提供参考依据。

1 预测模型理论与建模框架

1.1 灰色预测模型

1.1.1 灰色系统

灰色系统理论是由邓聚龙教授^[15]提出的,它着

重研究了生活、经济、科研活动中等难以解决的“小数据”和“贫信息”以及“不确定性”等问题。“灰”即指信息的不完全。考虑影响留学因素过多且复杂以及信息不完全性等问题,故采用灰色系统进行建模。

1.1.2 灰色关联度

设系统行为序列为

$$X_i = (x_i(1), x_i(2), \dots, x_i(n)), i = 1, 2, 3, \dots, m$$

对于 $\xi \in (0, 1)$, 令

$$\gamma(x_0(k), x_i(k)) =$$

$$\frac{\min_i \min_k |x_0(k) - x_i(k)| + \xi \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \xi \max_i \max_k |x_0(k) - x_i(k)|} \quad (1)$$

$$\gamma(X_0, X_i) = \frac{1}{n} \sum_{k=1}^n \gamma(x_0(k), x_i(k)) \quad (2)$$

则 $\gamma(X_0, X_i)$ 称为 X_0 与 X_i 的灰色关联度^[15]。

1.1.3 GM(1, N) 模型

(1) 设有原始序列为

$$X_1^{(0)} = (x_1^{(0)}(1), x_1^{(0)}(2), \dots, x_1^{(0)}(n))$$

关联序列为

$$X_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n)), i = 2, 3, \dots, N$$

(2) 时间响应函数计算公式^[12]为

$$\hat{x}_i(k+1) = [x_1^{(0)}(k+1) - \frac{1}{a} \sum_{i=2}^N b_i x_i^{(1)}(k+1)] e^{-ak} + \frac{1}{a} \sum_{i=2}^N b_i x_i^{(1)}(k+1) \quad (3)$$

再根据式: $\hat{x}_1^{(0)}(k) = \hat{x}_1^{(1)}(k) - \hat{x}_1^{(1)}(k-1)$, 进而求得 $\hat{x}_1^{(0)}(k)$ 的预测值。

1.2 BP 神经网络模型

针对一些无法用线性模型解决的问题,神经网络利用自身较强的映射能力,可以较好地解决这类非线性问题。BP 神经网络是一种误差反向传播和信号正向传播的多层前馈神经网络,一般包括输入、隐含、输出三层或者以上的神经网络,而且能够较好地逼近非线性连续函数^[16]。考虑出国留学影响因素众多以及因素与人数之间复杂的非线性关系,使得难以精准预测留学人数,而神经网络模型恰好使得输入与输出之间,呈现出高度非线性的映射特点,更好地进行非线性预测。

BP 神经网络构建主要分成 3 步,首先,生成 BP 网络,再进行网络训练,最后得到网络仿真,其函数

表达式如下。

若一个神经元模型有 n 个输入变量, w 为连接不同神经元之间的权重, 最终输出为

$$y = f(wx + b)$$

其中, b 为神经元的偏置值, f 为传递函数。

BP 神经网络拓扑图如图 1 所示, 图 1 中输入层、隐含层以及输出层神经元节点个数分别为 n 、 m 、 l , 这种结构称为 $n-m-l$ 三层 BP 神经网络^[16]。其中, $x_i (i=1, 2, \dots, n)$ 为实际的输入值, $y_j (j=1, 2, \dots, m)$ 是隐含层的输出, $O_k (k=1, 2, \dots, l)$ 为实际输出, 隐含层和输出层阈值分别为 a, b , 输入层与隐含层以及隐含层与输出层之间的连接权值分别为 v_{ij}, w_{jk} 。

此处, 权值更新的计算公式^[16]为

$$v_{ij} = v_{ij} + \Delta v_{ij}, w_{jk} = w_{jk} + \Delta w_{jk}$$

$$\Delta v_{ij} = \eta \left(\sum_{k=1}^l \delta_k^o w_{jk} \right) y_j (1 - y_j) x_i$$

$$\Delta w_{jk} = \eta \delta_k^o y_j$$

其中, $\delta_k^o = (d_k - o_k) o_k (1 - o_k)$, η 为学习速率。

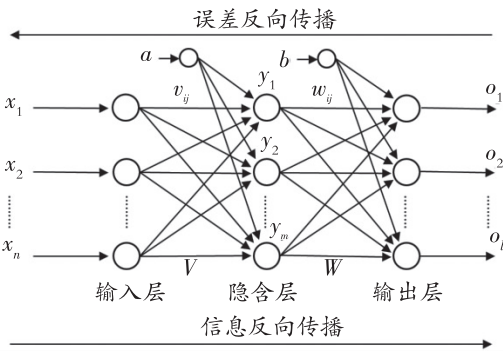


图 1 BP 神经网络拓扑图

Fig. 1 BP neural network topology

1.3 基于 L1 范数的加权几何平均的组合预测模型

采用陈华友等^[17]提出的基于 L1 范数的加权几何平均的组合预测模型, 利用预测误差绝对值的和, 衡量预测精度, 这样可以克服预测误差再平方以后导致误差扩大或者缩小的缺陷, 尤其针对数据中存在异常值的问题, 其模型参数估计的稳健性比预测误差平方和要好。

设某经济现象的指标观测值为 $\{x_t, t=1, 2, \dots, N\}$, 设有 m 个单一预测模型, 其中 x_{it} 为第 i 个模型在

t 时刻的预测值, l_1, l_2, \dots, l_m 为单一模型在组合预测模型中的权系数, 且 $\sum_{i=1}^m l_i = 1, l_i \geq 0, i = 1, \dots, m$

定义 1^[11]: 令

$$e_{it} = \ln x_{it} - \ln x_t$$

$$e_t = \ln x_t - \ln x_t = \sum_{i=1}^m l_i (\ln x_{it} - \ln x_t) = \sum_{i=1}^m l_i e_{it}$$

e_{it} 称为第 i 种单一预测模型第 t 时刻预测值与实际值之间的对数误差, e_t 为第 t 时刻的组合模型预测值和实际值之间对数误差。

定义 2^[17]: 令

$$F = \sum_{t=1}^N |e_t| = \sum_{t=1}^N \left| \sum_{i=1}^m l_i e_{it} \right|$$

$$F_i = \sum_{t=1}^N |e_{it}|, i = 1, 2, \dots, m \quad (4)$$

称 F 为加权几何平均组合模型预测值与实际值之间的基于 L1 范数对数误差, 其中, F_i 为第 i 种单一模型预测值与实际值之间的基于 L1 范数对数误差。

定义 3^[17]: 设 $F(L)$ 为各种预测模型的加权系数向量的函数, 则基于 L1 范数以最小对数误差为准则的加权几何平均组合预测模型为

$$\min F(L) = \sum_{t=1}^N \left| \sum_{i=1}^m l_i e_{it} \right|$$

$$\text{s. t. } \begin{cases} \sum_{i=1}^m l_i = 1 & \dots \\ l_i \geq 0, & i = 1, 2, \dots, m \end{cases} \quad (5)$$

若要求目标函数的最小值, 可做如下变换, 将其转化为线性规划问题, 令

$$\varepsilon_t = \sum_{i=1}^m l_i e_{it}, \varepsilon_t^+ = \begin{cases} \varepsilon_t, & \varepsilon_t \geq 0 \\ 0, & \varepsilon_t \leq 0 \end{cases}, \varepsilon_t^- = \begin{cases} -\varepsilon_t, & \varepsilon_t \leq 0 \\ 0, & \varepsilon_t \geq 0 \end{cases}$$

则 $|\varepsilon_t| = \varepsilon_t^+ + \varepsilon_t^-, \varepsilon_t = \varepsilon_t^+ - \varepsilon_t^-, \varepsilon_t^+ \varepsilon_t^- = 0$

在上述变换下, 式(5)可改写成如式(6)模型:

$$\min F(L) = \sum_{t=1}^N (\varepsilon_t^+ + \varepsilon_t^-)$$

$$\begin{cases} \sum_{i=1}^m l_i e_{it} - \varepsilon_t^+ + \varepsilon_t^- = 0 \\ i = \sum_{i=1}^m l_i = 1 \\ \varepsilon_t^+ \geq 0, \varepsilon_t^- \geq 0, t = 1, 2, \dots, N \\ l_i \geq 0, i = 1, 2, \dots, m \end{cases} \quad (6)$$

此处可以利用 MATLAB 或 LINGO 软件求解,求得的最优解 L 即为组合预测模型的加权系数。

1.4 建模框架

首先选取柯普文章中提到的 GM(1,1) 作为对照模型,但仅以单一因素建模,难以精准预测留学人数,而 GM(1,N) 模型主要用来研究多外部因子影响内部因子的动态变化关系,克服了以单一因素建模的局限性;针对出国留学与各影响因素之间属于非线性函数的问题,若使用传统的线性回归预测方法可能难以反映它们之间的非线性关系,而 BP 神经网络模型可以适应非线性预测,使得输入与输出之间呈现高度的非线性映射特点,因此建立 BP 神经网络模型;然后构建基于 L1 范数改进的组合预测模型;最后,利用预测误差评价指标体系比较 GM(1,1)、GM(1,3)、BP 神经网络模型和基于 L1 范数改进的组合模型预测精度,具体建模框架如图 2 所示。

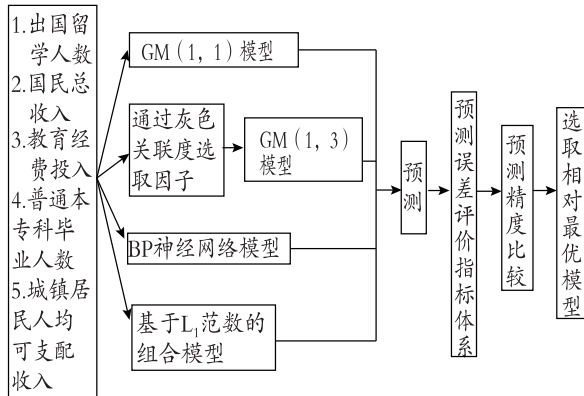


图 2 建模框架图

Fig. 2 The framework of modeling

2 实证分析

2.1 数据说明

数据均来源于《中国统计年鉴》和中华人民共和国教育部官方网站。由于中华人民共和国教育部官方网站中 2020 年我国出国留学人数还未更新,并且考虑各变量数据的时效性和可获得性,仅选取 2006—2019 年出国留学人数以及各变量统计数据

进行分析。

2.2 对照模型预测

对照模型 GM(1,1) 的未知参数求解计算如下:根据式(7)

$$\sigma(k) = \frac{x^{(0)}(k)}{x^{(0)}(k-1)}, k=2,3,\dots,14 \quad (7)$$

可计算级比结果 $\sigma(k)$ 均满足 $\sigma(k) \in (e^{\frac{-2}{n}}, e^{\frac{2}{n}})$, 故可建模。

根据最小二乘估计解得 GM(1,1) 模型的参数估计值: $\hat{a} = -0.10671$, $\hat{b} = 18.49$, 则响应方程为

$$\hat{x}_i^{(1)}(k) = (186.67) e^{0.10671(k-1)} - 173.27, k=1, 2, \dots, n$$

将 k 代入响应方程,可计算 GM(1,1) 模型的预测值及预测精度。

2.3 GM(1,3) 模型预测

2.3.1 数据样本指标的选取

由于数据样本的选取对模型的构建及其预测结果影响较大,而影响留学的因素众多,其中又有一些因素是难以量化的,倘若选取全部的因素构造预测模型是难以实现的。因此,要科学地选取外部影响因素,并尽可能客观真实地反映留学人数的变化规律,这样预测留学人数变化趋势才更准确且具有说服力。刘志民等^[18]认为一国经济发展状况和教育投入是影响留学的重要因素。刘玉君等^[19]通过实证分析发现,教育经费的投入对经济发展有促进作用,同时也是实现经济可持续发展的现实要求。刘倩等^[20]认为经济发展水平和产业结构对中国高等教育有一定的影响。马子健等^[21]选取普通高等学校毕业生数、城镇居民家庭人均可支配收入作为影响大学生出国留学影响因素。根据科学性、可行性等原则,借鉴前人文献指标的研究成果,选取国民总收入、教育经费投入、普通本专科毕业人数、城镇居民人均可支配收入作为影响出国留学的外部因素,具体数值如表 1 所示。

表 1 出国留学人数影响因子

Table 1 Factors influencing the number of students studying abroad

年份	出国留学人数/ 万人	国民总收入/ 亿元	教育经费投入/ 亿元	普通本专科毕业人数/ 万人	城镇居民人均可支配收入/ 元
2006	13.40	219 028.5	981 5	306.8	11 619.7
2007	14.40	270 704.0	12 148	447.8	13 602.5
2008	17.98	321 229.5	14 500	511.9	15 549.4
2009	22.93	347 934.9	16 502	531.1	16 900.5
2010	28.47	410 354.1	19 561	575.4	18 779.1
2011	33.97	483 392.8	23 869	608.2	21 426.9
2012	39.96	537 329.0	28 655	624.7	24 126.7
2013	41.39	588 141.2	30 365	638.7	26 467.0
2014	45.98	644 380.2	32 806	659.4	28 843.9
2015	52.37	686 255.7	36 129	680.9	31 194.8
2016	54.45	743 408.3	38 866	704.2	33 616.2
2017	60.84	831 381.2	42 557	735.8	36 396.2
2018	66.21	914 327.1	46 135	753.3	39 250.8
2019	70.35	988 528.9	50 178	758.5	42 358.8

2.3.2 灰色关联度计算与典型因子的选取

通过式(1)、式(2)计算可确定 N ,由表 2 可知教育经费投入与出国留学人数之间的灰色关联度最大($r=0.8378$),其次是国民总收入($r=0.9959$)、城镇居民人均可支配收入($r=0.6779$)、普通本专科毕业人数($r=0.7317$),从中选取灰色关联度 $r>0.7$ 的 3 个典型因子,即教育经费投入、国民总收入、城镇居民人均可支配收入,则 $N=3$ 。

表 2 灰色关联度系数表

Table 2 Grey correlation coefficient

变量	灰色关联系数
国民总收入	0.9959
教育经费投入	0.8378
普通本专科毕业人数	0.6779
城镇居民人均可支配收入	0.7317

2.3.3 数值计算

利用最小二乘估计可得 GM(1,3)模型参数估计值: $\hat{a}=-0.7964, \hat{b}_2=0.0000432, \hat{b}_3=0.0000378, \hat{b}_4=-0.000169$,再根据式(3)求出 GM(1,3)的预测值及预测精度如表 3 所示。

2.4 BP 神经网络模型预测

选取国民总收入、教育经费投入、城镇居民人均

可支配收入、普通本专科毕业人数作为 BP 神经网络的输入层,出国留学人数为输出层,则输入层神经元个数为 4,而输出层神经元个数是根据具体问题确定的,本文是针对预测出国留学人数的问题,故输出层神经元数目为 1,隐含层结点数为 6。函数选取和参数设定分别为:训练函数用弹性梯度下降法 trainrp,最大训练次数=50 000,精度=0.0045,学习率 $\eta=0.01$,迭代次数=50,具体算法流程如图 3 所示。用 MATLAB 运行后可得网络实际输出值与期望输出值,结果及预测精度如表 3 所示。

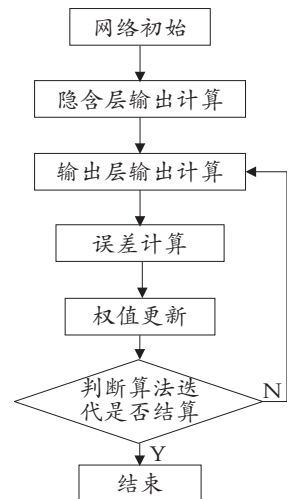


图 3 BP 神经网络算法流程

Fig. 3 BP neural network algorithm flow

表 3 单项预测模型预测结果及其预测精度

Table 3 The prediction value and prediction accuracy of each single prediction model

年份	实际值	GM(1,1)模型		GM(1,3)模型		BP 神经网络模型	
		预测值	预测精度	预测值	预测精度	预测值	预测精度
2006	813.40	13.40	0	13.40	0	12.75	6.92
2007	14.40	21.02	45.98	13.12	8.89	15.91	10.51
2008	17.98	23.39	30.08	17.18	4.45	19.51	8.54
2009	22.93	26.02	13.49	24.61	7.33	22.20	3.19
2010	28.47	28.95	1.70	28.19	0.98	26.80	5.87
2011	33.97	32.21	5.17	32.26	5.03	33.10	2.57
2012	39.96	35.84	10.31	37.70	5.66	39.38	1.46
2013	41.39	39.88	3.66	43.37	4.78	42.93	3.73
2014	45.98	44.37	3.51	46.47	1.07	47.07	2.36
2015	52.37	49.36	5.74	50.29	3.97	51.35	1.95
2016	54.45	54.93	0.77	54.67	0.40	55.29	1.54
2017	60.84	61.11	0.44	58.85	3.27	60.13	1.16
2018	66.21	67.99	2.69	64.38	2.76	64.33	2.84
2019	70.35	75.64	7.53	69.81	0.77	68.18	3.08

2.5 基于 L1 范数的组合预测模型预测

根据表 3 可知,GM(1,3)模型、BP 神经网络模型的拟合精度均高于 GM(1,1)模型,再选用 GM(1,3)模型和 BP 神经网络模型进行组合,将这两个单一模型在各个时期的预测值代入基于 L1 范数改进的组合预测模型式(5),可得到基于 L1 范数改进的最优化组合预测模型:

$$\min F(l_1, l_2) = |0.044 5l_1 - 0.123 3l_2| + |10.039 1l_1 - 0.020 3l_2| + |10.038 7l_1 - 0.092 8l_2| + | -0.014 1l_1 + 0.074 1l_2| + | -0.039 2l_1 + 0.048 3l_2| + | -0.080 6l_1 - 0.014 0l_2| + | -0.071 9l_1 - 0.035 3l_2| + | -0.029 5l_1 - 0.024 4l_2| + | -0.006 7l_1 - 0.029 7l_2| + |10.019 8l_1 - 0.025 6l_2| + | -0.054 7l_1 - 0.115 3l_2|$$

$$\text{s. t. } \begin{cases} l_1 + l_2 = 1 \\ l_1 \geq 0, l_2 \geq 0 \end{cases}$$

利用 LINGO 软件对上述式子求解,得到最优化模型的解为

$$l_1^* = 0.5171, l_2^* = 0.4829$$

再将模型的最优解代入式(4),计算得基于 L1 范数的组合预测模型的预测值和预测精度如表 4 所示。

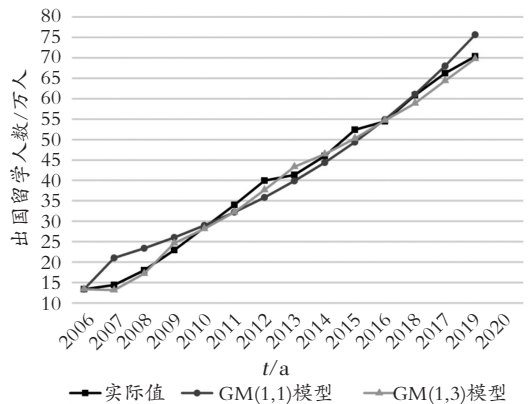
表 4 基于 L1 范数的组合预测模型预测值与预测精度

Table 4 Prediction value and prediction accuracy of combination prediction model based on L1 norm

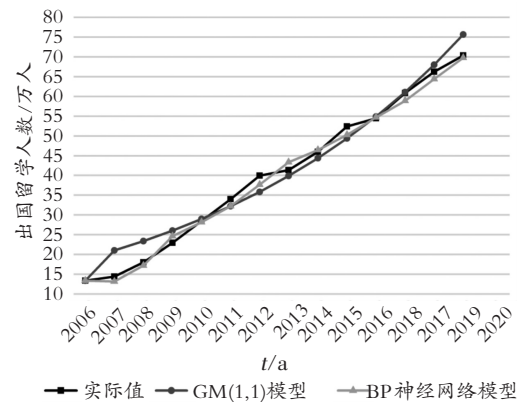
年份	预测值	预测精度
2006	13.09	2.342
2007	14.47	0.467
2008	18.31	1.808
2009	23.45	2.251
2010	27.52	3.341
2011	32.67	3.840
2012	38.51	3.625
2013	43.16	4.270
2014	46.76	1.696
2015	50.80	2.994
2016	54.97	0.954
2017	59.47	2.255
2018	64.36	2.800
2019	69.02	1.886

2.6 各单项预测与组合预测结果对比

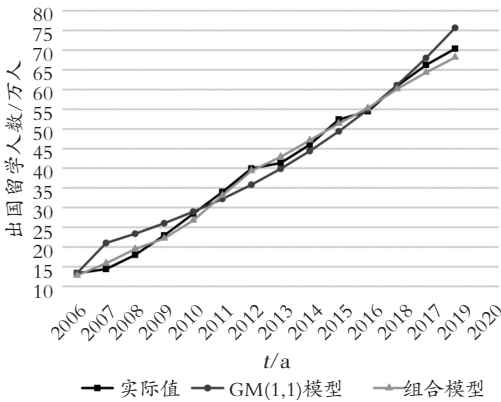
为体现基于 L1 范数的组合预测模型的有效性,将 3 个单一模型和基于 L1 范数的组合模型的预测值与实际值进行对比分析,如图 4 所示。



(a) GM(1,1)与 GM(1,3)模型对比



(b) GM(1,1)与 BP 模型对比



(c) CM(1,1) 与混合模型对比

图 4 不同模型对比图

Fig. 4 Comparison of different models

从图 4 可以直观地看出:GM(1,1)模型与实际值的预测效果不佳,GM(1,3)模型、BP 神经网络模型对实际值的预测效果较好,基于 L1 范数的组合模型预测效果最佳。为了进一步比较基于 L1 范数的组合预测模型与 3 个单一预测模型之间的优劣,利用预测误差平方和、平均绝对误差、平均绝对百分比误差、均方误差、均方百分比误差 5 个评价指标构成的模型预测效果评价指标体系进行比较,结果如表 5 所示。

(1) 预测误差平方和:

$$R_{SSE} = \sum_{i=1}^N (x_i - \hat{x}_i)^2$$

(2) 平均绝对误差:

$$R_{MAE} = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i|$$

(3) 平均绝对百分比误差

$$R_{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right|$$

(4) 均方误差

$$R_{MSE} = \frac{1}{N} \sqrt{\sum_{i=1}^N (x_i - \hat{x}_i)^2}$$

(5) 均方百分比误差

$$R_{MSPE} = \frac{1}{N} \sqrt{\sum_{i=1}^N \left(\frac{x_i - \hat{x}_i}{x_i} \right)^2}$$

表 5 模型预测效果评价指标体系

Table 5 Evaluation index system of model prediction effect

有效性指标		R_{SSE}	R_{MAE}	R_{MAPE}	R_{MSE}	R_{MSPE}
单一预测模型	GM(1,1)	148.331	2.531	2.530	0.831 4	0.042
	GM(1,3)	29.357	1.224	1.224	0.363 8	0.011
	BP 神经网络	23.512	1.199	1.156	0.452 6	0.012
组合预测模型	基于 L1 范数	18.727	1.009	0.988	0.353 1	0.007
	的组合模型					

从表 5 可以看出,3 个单一模型的各项指标误差均高于基于 L1 范数的组合预测模型的误差指标值。即基于 L1 范数改进的组合预测模型的预测效果更优,并且有效地利用了单一模型提供的信息,从而提高了预测精度,可以更加准确地预测我国出国留学人数,同时利用组合模型预测我国未来几年出国留学人数,结果如表 6 所示。

表 6 2020—2021 年出国留学人数预测

Table 6 Forecast of the number of students studying abroad from 2020 to 2021

年份	2020	2021
出国留学人数预测值/万人	71.42	73.60

3 结束语

出国留学教育作为促进我国经济和高等教育发展的重要方式之一,且留学生规模又作为衡量留学教育的一个重要指标,故有必要选择一个相对最优模型来精准地预测出国留学人数。在借鉴前人研究的基础上,选取 GM(1,1)模型为对照模型,同时构建了 GM(1,3)模型、BP 神经网络模型和基于 L1 范数的组合预测模型,结果显示基于多因素建模的 GM(1,3)模型的预测精度相对于以单一因素建模的 GM(1,1)模型显著提升;同时,BP 神经网络模型的精度相对于 GM(1,1)模型又一次提高,而基于 L1 范数的 GM(1,3)和 BP 神经网络的组合模型预测精度又高于 3 个单一模型预测精度,故所建立基于 L1 范数的组合预测模型相对于单一模型在预测出国留学人数更有优势。通过实证分析可知,未来留学人数将会有增长趋势,但由于一些重大突发事件,如非典、金融危机、新冠病毒等非规律因素,会对预测结果产生影响。故需要结合非规律因素,对当前疫情下留学的发展趋势做进一步讨论。

参考文献 (References):

[1] 人民网. 教育部等八部门关于加快和扩大新时代教育对外开放的意见答问[EB/OL]. (2020-6-18) [2020-7-6]. [http://edu. people. com. cn/GB/n1/2020/0618/c1006-31752128. html](http://edu.people.com.cn/GB/n1/2020/0618/c1006-31752128.html).
 People's Daily. Opinions of the ministry of education and other eight departments on accelerating and expanding the opening up of education in the new era [EB/OL]. (2020-6-18) [2020-7-6]. <http://edu. people. com. cn/GB/n1/>

- 2020/0618/c1006-31752128.html.
- [2] 潘昆峰, 蒋承. 我国大学生留学选择的影响因素分析[J]. 中国高教研究, 2015(3): 15—20.
PAN Kun-feng, JIANG Cheng. Analysis on the influencing factors of Chinese college students' choice of studying abroad [J]. China Higher Education Research, 2015 (3): 15—20.
- [3] 陈玥, 毛立伟. 我国出国留学政策变迁的基本逻辑及理性选择——基于历史制度主义的分析视角[J]. 当代教育论坛, 2020(5): 10—16.
CHEN Yue, MAO Li-wei. The basic logic and rational choice of the policy change of studying abroad in China from the perspective of historical institutionalism [J]. Contemporary Education Forum, 2020 (5): 10—16.
- [4] 魏永长, 陈方宇. 大学生出国留学意愿影响因素实证研究[J]. 高等教育评论, 2017, 5(2): 204—219.
WEI Yong-chang, CHEN Fang-yu. Empirical study on influencing factors of college students' willingness to study abroad [J]. Higher Education Review, 2017, 5 (2): 204—219.
- [5] 陆根书, 田美, 黎万红. 大学生出国留学意愿的影响因素分析[J]. 复旦教育论坛, 2014, 12(5): 36—44.
LU Gen-shu, TIAN Mei, LI Wan-hong. Analysis on influencing factors of college students' willingness to study abroad [J]. Fudan Education Forum, 2014, 12 (5): 36—44.
- [6] 陈扬霖, 岳昌君. 我国高校学生留学趋势及影响因素:2005—2017[J]. 教育学术月刊, 2020(5): 40—45, 111.
CHEN Yang-lin, YUE Chang-jun. Trends and influencing factors of college students studying abroad in China: 2005—2017 [J]. Education Academic Monthly, 2020 (5): 40—45 ,111.
- [7] 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.
KE Pu, WU Guang. Prediction of the number of students studying abroad based on GM (1,1) model [J]. Value Engineering, 2012, 31 (25): 318—319.
- [8] 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.
YANG Cai-hua, DUAN Ling-xiao. Prediction of the number of Chinese students studying abroad: based on time series prediction method [J]. Science and Technology Economic Market, 2016 (1): 119—121.
- [9] 冯志平, 余东. 我国留学生人数的预测及其对居民消费的影响[J]. 南京信息工程大学学报(自然科学版), 2014, 6(4): 369—373.
FENG Zhi-ping, YU Dong. Prediction of the number of overseas students in China and its impact on residents' consumption [J]. Journal of Nanjing University of Information Science Technology (Natural Science Edition), 2014, 6 (4): 369—373.
- [10] BATES J M, GRANGER C W J. Combination of forecasts [J]. Operations Research Quarterly, 1969, 20 (4): 451—468.
- [11] 陈华友. 组合预测方法有效性理论及其应用[M]. 北京: 科学出版社, 2008.
CHEN Hua-you. Validity theory and application of combination forecasting method [M]. Beijing: Science Press, 2008.
- [12] 唐小我, 马永开, 曾勇, 等. 现代组合预测和组合投资决策方法及应用研究[M]. 北京: 科学出版社, 2003.
TANG Xiao-wo, MA Yong-kai, ZENG Yong, et al. Research on Modern portfolio forecasting and portfolio investment decision method and its application [M]. Beijing: Science Press, 2003.
- [13] 刘金培, 汪漂, 黄燕燕, 等. 基于区间时间序列小波多尺度分解的组合预测方法[J]. 统计与决策, 2020, 36(19): 5—9.
LIU Jin-pei, WANG Piao, HUANG Yan-yan, et al. Combined forecasting method based on interval time series wavelet multi-scale decomposition [J]. Statistics and decision, 2020, 36 (19): 5—9.
- [14] 陈华友, 刘春林. 基于L₁范数的加权几何平均组合预测模型的性质[J]. 东南大学学报(自然科学版), 2004(4): 535—540.
CHEN Hua-you, LIU Chun-lin. Properties of weighted geometric average combination forecasting model based on L₁ norm [J]. Journal of Southeast University (Natural Science Edition), 2004 (4): 535—540.
- [15] 邓聚龙. 灰色系统基本方法[M]. 武汉: 华中科技大学出版社, 2005.
DENG Ju-long. Basic method of grey system [M]. Wuhan: Huazhong University of Science and Technology Press, 2005.
- [16] 陈雯柏. 人工神经网络原理与实践[M]. 西安: 西安电子科技大学出版社, 2016.
CHEN Wen-bai. Principle and practice of artificial neural network [M]. XI'an: Xi'dian University Press, 2016.
- [17] 陈华友. 基于L₁范数的加权几何平均组合预测方法[J]. 安徽大学学报(自然科学版), 2004(4): 5—10.
CHEN Hua-you. Weighted geometric average combination forecasting method based on L₁ norm [J]. Journal of Anhui University (Natural Science Edition), 2004 (4): 5—10.

- [18] 刘志民, 杨洲. 2001—2014年八大留学目的国位序变化的成因分析与启示[J]. 黑龙江高教研究, 2017(8): 17—22.
LIU Zhi-min, YANG Zhou. Analysis of the causes of the change of the order of the eight major countries of study abroad from 2001 to 2014 and its enlightenment [J]. Heilongjiang Higher Education Research, 2017(8): 17—22.
- [19] 刘玉君, 王成武, 应卫平. 教育经费投入对经济发展影响的区域差异研究[J]. 统计与决策, 2020, 36(2): 121—124.
LIU Yu-jun, WANG Cheng-wu, YING Wei-ping. A study on the regional differences of the impact of educational investment on economic development [J]. Statistics and Decision Making, 2020, 36(2): 121—124.
- [20] 刘倩, 王永哲. 中国高等教育规模的经济影响因素分析[J]. 统计与决策, 2019, 35(18): 134—138.
LIU Qian, WANG Yong-zhe. Analysis on the influencing factors of economic development of China's higher education scale [J]. Statistics and Decision Making, 2019, 35(18): 134—138.
- [21] 马子建, 汪蕊. 大学生出国留学影响因素的实证研究[J]. 管理观察, 2017(4): 145—147.
MA Zi-jian, WANG Rui. An empirical study on the influencing factors of college students studying abroad [J]. Management Observation, 2017(4): 145—147.

Research on Combination Forecasting of the Number of Students Studying Abroad Based on L1 Norm

HU Zhen

(School of Science, Hubei University of Technology, Wuhan 430068, China)

Abstract: In order to more accurately explore the trend of the number of people studying abroad in China, a combination forecasting model based on L1 norm is proposed to make a prediction about the number of people studying abroad. The factors that affect the number of students studying abroad are selected from multiple perspectives, and the typical factors that affect the number of students studying abroad are extracted by grey correlation analysis, and the GM(1,3) model is constructed. BP neural network model is established. Combination prediction model based on L1 norm is proposed, and the optimal weight coefficient of single model is determined by solving linear programming. Subsequently, it forecasts the number of students studying abroad from 2006 to 2019. The GM(1,1) model is selected as the control model, and the prediction accuracy of the model is compared through the control model and the prediction error evaluation index system. The results show that the combination forecasting model based on L1 norm outperforms three single models, which effectively improves the prediction accuracy and makes full use of the information provided by the single prediction model, thus more accurately predicting the number of students studying abroad. In the next few years, there is still a large space for the development of the scale of studying abroad in China. The prediction results can provide reference for the relevant work of studying abroad in China under the global epidemic situation.

Key words: GM(1,3) model; BP neural network model; L1 norm; combination forecasting; the number of students studying abroad

责任编辑:田 静

引用本文/Cite this paper:

胡珍. 基于L1范数的出国留学人数组合预测研究[J]. 重庆工商大学学报(自然科学版), 2022, 39(3): 61—69.

HU Zhen. Research on combination forecasting of the number of students studying abroad based on L1 norm [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2022, 39(3): 61—69.