Computer Science and Application 计算机科学与应用, 2021, 11(6), 1738-1746
Published Online June 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.116179

Hans汉斯

遗传算法优化灰色神经网络旅游人数预测研究

甄洁玲，张　悦

东北大学，辽宁 沈阳
Email: jielingzhen1220@163.com, zhangyue@mail.neu.edu.cn

收稿日期：2021年5月20日；录用日期：2021年6月17日；发布日期：2021年6月24日

摘　要

该文以旅游人数为研究对象，初步选取影响旅游人数的13个因素，首先建立基于遗传算法优化灰色神经网络模型对旅游人数进行预测。为了进一步提高模型预测精度，采用灰色关联度法和平均值影响法两种方法对影响因素进行变量筛选，选取影响程度大的因素代入模型进行预测。分析比较模型的结果，可得经过平均值影响法变量筛选后的模型预测精度最高，误差最小。

关键词

旅游人数，遗传算法，灰色神经网络，灰色关联度法，平均值影响法

Research on the Number of Tourists Prediction Based on Genetic Algorithm Optimization Grey Neural Network

Jieling Zhen, Yue Zhang

Northeastern University, Shenyang Liaoning
Email: jielingzhen1220@163.com, zhangyue@mail.neu.edu.cn

Received: May 20th, 2021; accepted: Jun. 17th, 2021; published: Jun. 24th, 2021

Abstract

Taking the number of tourists as the research object, this paper preliminarily selects 13 factors that affect the number of tourists, and firstly establishes a grey netural network model based on genetic algorithm to predict the number of tourists. In order to improve the prediction accuracy of the model, grey correlation degree method and average influence method were used to screen the variables of factors, and the factors with high influence degree were selected and substituted into the model for prediction. By analyzing and comparing the results of the model, it can be concluded that the model with the influence of the average value has the highest prediction accuracy and minimum error after selecting the normal variables.

文章引用: 甄洁玲, 张悦. 遗传算法优化灰色神经网络旅游人数预测研究[J]. 计算机科学与应用, 2021, 11(6): 1738-1746.
DOI: 10.12677/csa.2021.116179

Keywords

The Number of Tourists, Genetic Algorithm, Grey Neural Network, Grey Relational Degree Method, Mean Influence Method

1. 引言

自从改革开放政策实施以来，我国社会经济呈现出高速增长的态势，国民收入水平显著上升，旅游日益成为人们现代化进行休闲、娱乐活动的主要途径之一。目前，我国的国内旅游业是发展潜力最大的旅游市场，在经济的发展中具有重要的地位[1]。所以，对旅游人数进行较为准确的预测分析，为当地旅游业相关部门制定发展战略决策和其他相关政策措施提供一定的理论基础和技术支撑，从而有效地推动旅游行业的健康发展。主要的预测方法有两大类：定性分析与定量分析。其中，定性分析主要是根据人的主观判断和已有经验进行大致趋势和范围的分析；定量分析通过建立统计数学模型得到具体的一个数值。如：Seetanah 等人通过 VAR 模型研究相对平均费用、旅游基础设施、旅游行业发展水平等相关因素对旅游市场具有长期影响[2]。Pai 等人提出了一种新型预测旅游人数模型，建立模糊聚类和支持向量回归的组合模型，该方法得到的预测精度优于传统的方法[3]。E. Hadavandi 等人首次把遗传模糊系统方法应用到旅游需求预测的问题中[4]。Fong-Lin Chu 建立基于分段线性的方法对澳门旅游需求预测[5]。陈萍萍根据游客量、客运量和旅游航班数的数据，建立了三种模型，分别是指数平滑模型、季节 ARMA 模型和 Elman 模型，得到 Elman 模型预测效果最好，并且三种模型的组合预测效果最优[6]。廖治学等人通过建立以 GMDH 非线性叠加对季节性 ARMA 模型和神经网络模型的组合模型来进行预测，组合模型的预测精度最优[7]。

本文首先将灰色理论和神经网络模型进行结合，建立灰色神经网络模型，为得到最优的初始参数，采用遗传算法对模型进行优化。由于模型预测误差较大，为提高模型预测精度，对影响因素进行变量筛选，将影响程度大的因素留下，删除掉影响程度小的因素。为了增强模型的对比性，用两种方法对变量进行筛选，即分别建立基于灰色关联度和平均值影响法两种方法的变量筛选模型，对旅游人数进行预测。

2. 算法介绍

2.1. 灰色神经网络模型

灰色理论可以解决样本少、信息少和不确定性的问题，其通过在已有的信息的基础上充分提取有用的信息后对未知的信息进行合理推测。灰色神经网络将样本数据经过一次累加处理后转化成微分方程的形式，利用已知的相关信息，对微分方程的系数进行求解，从而得到微分方程的表达式。首先将原始数据 $x(t)(t=0,1,\cdots,n-1)$ 进行一次累加处理后得到 $y(t)$，然后建立微分方程如下：

$$\frac{d_{y_1}}{d_t}+ay_1=b_1y_2+b_2y_3+\cdots+b_{n-1}y_n$$

其中，$y_2,\cdots,y_n$ 为输入数据，y_1 为输出数据，$a,b_1,\cdots,b_{n-1}$ 为方程系数。

求解微分方程，可得预测结果为 $z(t)$：

$$z(t)=\left(y_1(0)-\frac{b_1}{a}y_2(t)-\frac{b_2}{a}y_3(t)-\cdots-\frac{b_{n-1}}{a}y_n(t)\right)e^{-at}+\frac{b_1}{a}y_2(t)+\frac{b_2}{a}y_3(t)+\cdots+\frac{b_{n-1}}{a}y_n(t)$$

令

$$m=\frac{b_1}{a}y_2(t)+\frac{b_2}{a}y_3(t)+\cdots+\frac{b_{n-1}}{a}y_n(t)$$

将 m 代入上述 $z(t)$ 中，最终可以化简得到：

$$z(t)=\left(\left(y_1(0)-m\right)-y_1(0)\cdot\frac{1}{1+\mathrm{e}^{-at}}+2m\cdot\frac{1}{1+\mathrm{e}^{-at}}\right)\cdot\left(1+\mathrm{e}^{-at}\right)$$

将 $z(t)$ 映射到 BP 神经网络模型中就可以得到了一个网络结构为 n 个输入数据和 1 个输出数据的灰色神经网络模型[8]。

2.2. 遗传算法

遗传算法(Genetic Algorithm)是通过模拟自然界中自然选择和遗传机制中的选择、交叉、变异的现象，得到具有适应能力的个体，使得种群不断进化，最终得到适应能力最优的个体，同时也就得到了对应问题的最佳解。遗传算法以个体适应度函数为标准，通过模拟自然选择和遗传原则来寻求最佳参数[8]。其中，基本的操作步骤如下：

选择：根据个体适应度值的大小，按照一定的概率在当前种群中选择基因优良的个体到下一代中。

交叉：从种群中任意选择两个染色体，以一定概率选择一处或多处的位置交换两者的染色体，从而得到新的优良个体。

变异：从群体中任意选择一个染色体，以一定的概率随机改变该染色体上某一处的值，从而产生新个体。

2.3. 变量筛选

灰色关联度法通过计算影响因素的关联度来评价因素的相关程度。关联度绝对值越大表明因素之间的关联性越强[9]。具体步骤为：首先将原始数据进行预处理，消除单位量纲化的影响；选择参考序列 $X_0(t)$，t 表示时刻，n 个比较序列 $X_i(t)$，$i=1,2,\cdots,n$，比较序列对参考序列的关联系数为：

$$\xi_i(t)=\frac{\min\limits_s\min\limits_t\left|X_0(k)-X_s(k)\right|+\rho\max\limits_s\max\limits_t\left|X_0(k)-X_s(k)\right|}{\left|X_0(t)-X_i(t)\right|+\rho\max\limits_s\max\limits_t\left|X_0(k)-X_s(k)\right|}$$

其中 ρ 为分辨系数，$\min\limits_s\min\limits_t\left|X_0(k)-X_s(k)\right|$ 和 $\max\limits_s\max\limits_t\left|X_0(k)-X_s(k)\right|$ 分别为两级最小差和两级最大差。关联度为：

$$\gamma_i=\frac{1}{n}\sum_{k=1}^{n}\xi_i(t)$$

平均影响值法(MIV)可以用来评价输入变量对输出变量的影响程度，其正负号代表相关性的方向，绝对值大小代表影响程度大小。该方法可以用于筛选变量，留取影响程度大的变量，剔除影响程度小的变量。具体计算步骤为：首先使用原始数据训练神经网络，使网络达到稳定；将原始数据的每一个输入变量分别加减 10%，得到新样本 P_1，P_2；分别将 P_1 和 P_2 代入到已训练好的模型中，得到输出结果 Q_1 和 Q_2，计算 Q_1 和 Q_2 的差值即可得到输入变量对输出变量的影响变化值，最后将该值按照样本个数进行平均得到 MIV 值[8]。

3. 实证分析

3.1. 数据来源与预处理

通过阅读文献和查阅相关资料，从《中国统计年鉴》和《中国旅游统计年鉴》中得到原始数据，首先对原始数据进行简单处理，初步得到 13 个影响因素：X_1：总里程；X_2：居民消费价格指数；X_3：国内

旅游收入；X_4：城镇居民可支配收入；X_5：旅行社个数；X_6：城镇居民旅游人均花费；X_7：城镇居民消费水平；X_8：人均国内生产总值；X_9：星级饭店个数；X_{10}：城镇居民人口数；X_{11}：客运量；X_{12}：国内生产总值；X_{13}：第三产业生产总值；被解释变量 Y：国内旅游人数。

数据预处理：首先将数据分为训练集和测试集，具体如：选取 1994~2019 年的相关数据进行模拟，其中，将 1994~2016 年数据作为训练数据，2017~2019 年数据作为测试数据。其次，为了消除单位量纲化的影响，将训练数据和测试数据分别进行归一化处理[10]，消除奇异数据，采用 mapminmax 函数对数据进行归一化处理，并运用 matlab 进行计算。mapminmax 具体计算公式：

$$x_k = \left(x_k - x_{\min}\right) / \left(x_{\max} - x_{\min}\right)$$

其中，$x_{\min}$ 为数列的最小值，$x_{\max}$ 为数列的最大值。

3.2. 模型预测

3.2.1. 遗传算法优化灰色神经网络模型预测(GA-GM 神经网络模型)

遗传算法具有随机性、全局性和鲁棒性强的特点，故用遗传算法优化灰色神经网络模型的初始参数，得到全局最优解[11]。首先将归一化训练数据进行一次累加处理，将其作为模型的输入变量，网络结构为 1-1-14-1。经过训练，根据迭代次数和训练误差，最终确定网络的传递函数为 purelin，训练函数为 traincgf。经 matlab 运算可得，模型均方误差为 9.1526，由图 1 可得，训练集的可决系数为 0.99965，测试集可决系数为 0.99844，整体的可绝系数为 0.99963，说明模型的拟合效果好。预测结果如表 1，可看出相对误差较高，只有一个低于 5%，且绝对误差均大于 2。

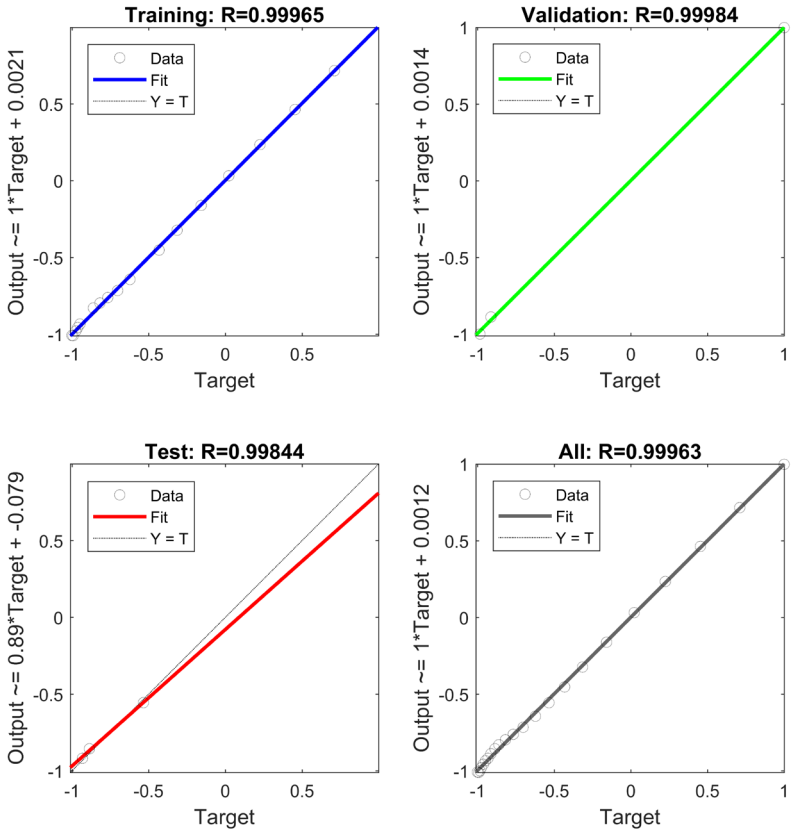


Figure 1. Regression analysis of GA-GM neural network
图 1. GA-GM 神经网络模型结果回归分析

Table 1. Prediction result and error of GA-GM neural network
表 1. GA-GM 神经网络模型预测结果及误差

年份	GA-GM 神经网络模型预测结果	实际旅游人数	相对误差	绝对误差
2017	47.8768	50.01	−4.3%	−2.1332
2018	52.2317	55.39	−5.7%	−3.1583
2019	56.4638	60.06	−6.0%	−3.5962

由于上述的模型输入变量有 13 个，变量较多，为了进一步提高模型预测精度，可对变量进行筛选，将影响效果显著的变量选入模型，效果不显著的变量删除。

3.2.2. 灰色关联度法变量筛选后的 GA-GM 神经网络模型

首先对旅游人数影响因素进行灰色关联度计算。其中，旅游人数作为参考数列，13 个影响因素作为比较序列，计算指标的关联系数，由 matlab 程序进行计算，在求解关联系数中，借鉴前人已有经验将分辨系数设为 0.5，可得到影响旅游人数因素的关联度如表 2。注：表 2 中的灰色关联度为关联系数的绝对值。

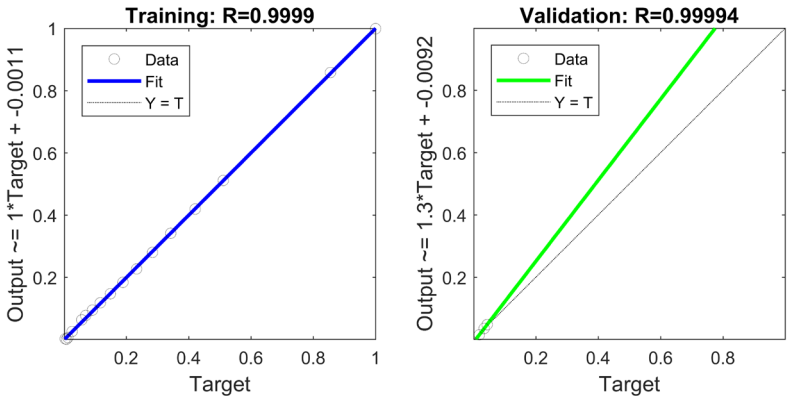
Table 2. Grey correlation degree of influencing factors
表 2. 影响因素的灰色关联度值

影响因素	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
灰色关联度	0.91	0.79	0.65	0.86	0.92	0.87	0.82	0.95	0.76	0.91	0.91	0.72	0.97

通过上述结果可以得到，这 13 个影响因素中，由于 X_3 和 X_{12} 的灰色关联度小于 0.75，故去除变量 X_3 和 X_{12}，选取剩余的 11 个变量作为最终的输入变量，此时网络结构为 1-1-12-1。由 matlab 运算可得，模型的均方误差为 4.9571，由回归图 2 可得，测试集可决系数为 0.99999，模型的拟合效果好，可解释性高。预测结果如表 3，模型的相对误差均在 5%以内，其预测精度显然高于未经过变量筛选的 GA-GM 神经网络模型。

3.2.3. 平均值影响法变量筛选后的 GA-GM 神经网络模型

平均值影响法(MIV)是神经网络中评价变量相关的最好指标之一，正负号代表相关性的方向，绝对值的大小代表指标的重要程度[12]。用 matlab 编程分别对样本数据的 13 个变量求 MIV 值如表 4。注：表 4 中的 MIV 值为其绝对值。



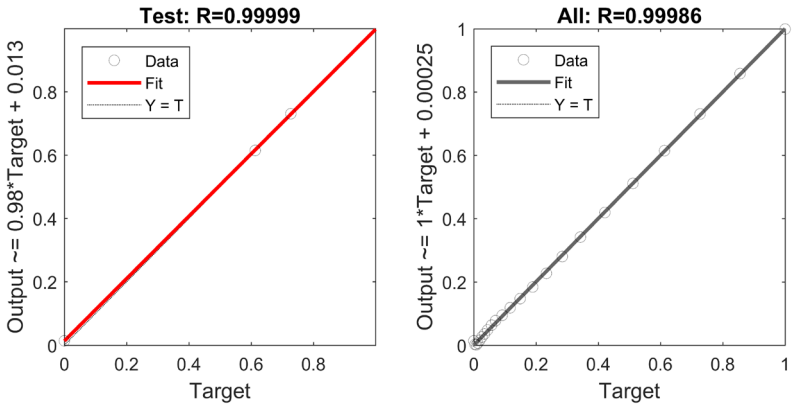


Figure 2. The regression analysis of the results after the grey relational degree screening variables
图 2. 灰色关联度筛选变量后模型结果回归分析

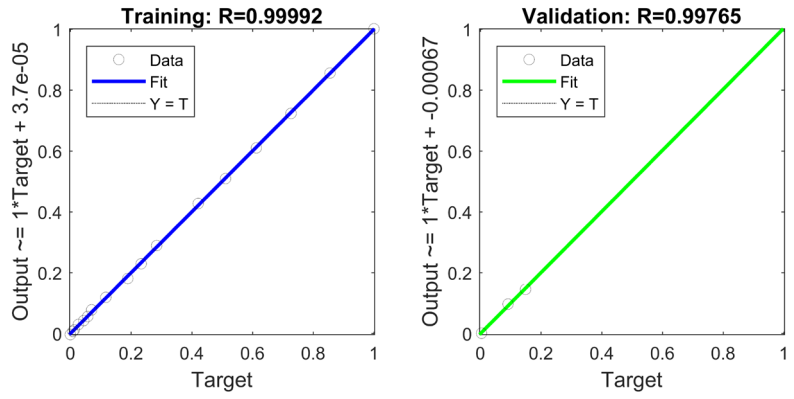
Table 3. The prediction results and errors of the model after the grey relational degree screening variables
表 3. 灰色关联度筛选变量后模型预测结果及误差

年份	灰色关联度筛选后模型结果	实际旅游人数	相对误差	绝对误差
2017	51.5009	50.01	1.0%	1.4909
2018	58.1393	55.39	5.0%	2.7493
2019	62.3161	60.06	3.8%	2.2561

Table 4. The MIV value of the influencing factors
表 4. 影响因素的 MIV 值

影响因素	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
MIV 值	1.42	0.0075	0.58	2.01	0.046	0.31	0.29	0.41	0.78	0.42	0.15	1.06	0.75

根据表 4 中 13 个变量的 MIV 绝对值的大小进行排序，可得 X_2 和 X_5 的影响程度最小，故剔除变量 X_2 和 X_5，选取剩余的 11 个变量作为最终的输入变量。根据回归图 3 可得，训练集可决系数为 0.99992，测试集可决系数为 0.99986，模型的可解释性高。由表 5 可得，模型的相对误差均低于 5%，绝对误差均低于 1.5。模型的均方误差为 0.9470，是当前均方误差最小的模型，其预测精度高于未变量筛选的 GA-GM 神经网络模型和灰色关联度筛选后的模型。



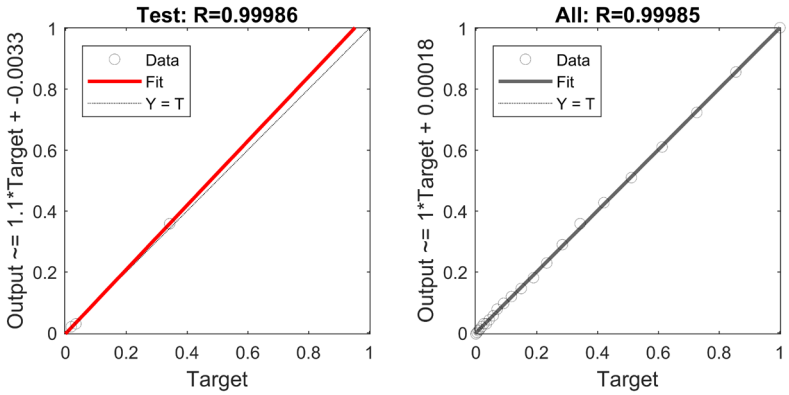


Figure 3. The regression analysis of the results after the average influence method screening variables
图 3. 平均值影响法筛选变量后模型结果回归分析

Table 5. The prediction results and errors of the model after the average influence method screening variables
表 5. 平均值影响法筛选变量后模型预测结果及误差

年份	平均值影响法筛选后模型结果	实际旅游人数	相对误差	绝对误差
2017	51.0399	50.01	2.1%	1.0299
2018	56.6557	55.39	2.3%	1.2657
2019	60.4821	60.06	0.7%	0.4221

4. 结果比较分析

为了比较上述的三个模型的预测精度和拟合效果，由前面结果可得均方误差、相对误差 MIN、相对误差 MAX、绝对误差 MIN、绝对误差 MAX 和测试集的可决系数。其中，均方误差是衡量估计量与被估计量之间的差异程度，均方误差越大，估计量与被估计量的差异也就越大，具体计算公式为：$MSE\left(\hat{\theta}\right)=E\left(\hat{\theta}-\theta\right)^2$，其中，$\theta$是被估计量的参数真值，$\hat{\theta}$是估计量数值。相对误差是绝对误差与真值之比乘以 100%所得，以百分数表示，是一个无量纲的值。绝对误差是预测值与真值之间的差值，有方向和大小，具有量纲。其中，相对误差和绝对误差越小，说明预测结果越精确。为了便于比较，将结果整理成表格的形式，如表 6：

Table 6. Comparative analysis of model prediction effect
表 6. 模型预测效果比较分析

模型	GA-GM 模型	灰色关联度变量筛选后 GA-GM 模型	平均值影响法变量筛选后 GA-GM 模型
均方误差	9.1526	4.9571	0.9470
相对误差 MIN	−4.3%	3.0%	0.7%
相对误差 MAX	−6.0%	5.0%	2.3%
绝对误差 MIN	−2.1332	1.4909	0.4221
绝对误差 MAX	−3.5962	2.7493	1.2657
可决系数	0.99844	0.99999	0.99986

由表 6 可得，

1) 未经过变量筛选的 GA-GM 模型相比经过变量筛选的模型存在较大的误差，拟合效果也较弱。

2) 比较灰色关联度法和平均值影响法，由表 6 可知，在均方误差、相对误差和绝对误差中，平均值影响法均低于灰色关联度法；但灰色关联度法的可决系数略高于平均值法。

3) 综合考虑上述模型，基于平均值影响法变量筛选的 GA-GM 神经网络模型的各种误差均最小，拟合效果好，所以确定该模型预测国内旅游人数。

为了更加直观的得到各个模型预测结果与实际值之间的关系，通过 matlab 画图，可以得到图 4，可以发现 3 种模型的预测趋势均是逐年上升的，其中，MIV 变量筛选后的预测值最接近实际值，预测的误差最低。

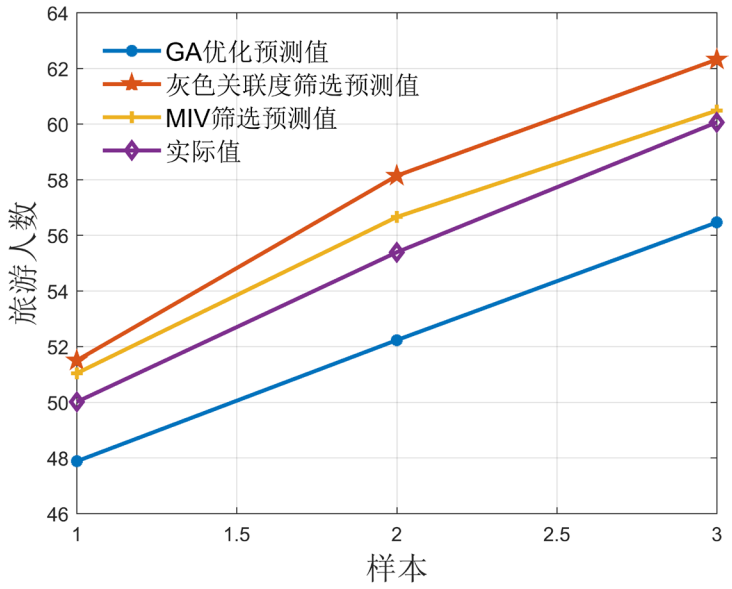


Figure 4. Comparison diagram of predicted value and actual value of each model
图 4. 各个模型预测值与实际值对比图

5. 结论

由于旅游人数受到多种因素的影响，样本数据具有高维度、样本量小的特点，神经网络模型具有自学习性和自适应性的优点，故建立灰色神经网络模型，并不断对模型进行改进，最终确定基于平均值影响法变量筛选的 GA-GM 神经网络模型为旅游人数的预测模型。创新点在于一方面为旅游人数预测提供了新的算法，丰富了旅游人数预测方向的理论基础，拓宽了基于平均值影响法变量筛选后的 GA-GM 神经网络算法的应用范围；另一方面该文选取了两种变量筛选的方法，增加了模型的对比性，从而选取最优的模型进行预测。当解释变量已知时，就可以对旅游人数进行预测，有利于旅游业相关部门科学制定政策，开展旅游业相关的活动，对资源进行合理配置，促进旅游行业的健康发展。

本文搜集了 1994~2019 年共 26 年数据，所选取的样本数量太少，属于小样本，在一定程度上会影响预测值的准确性和可信性。若要进行更加准确性的研究分析，要使得样本容量尽可能地大。在模型之外的影响因素被忽略了，显然会对模型的精度造成一定程度的影响，因此所得的模型仍然需要进行优化修正，从而得到的预测效果更加接近于实际。

基金项目

本文由国家自然科学基金(61703083 和 61673100)和国家留学基金委(201706085041)赞助支持。

参考文献

[1] 刘胜. 基于 ARIMA 与 SVM 组合模型的国内旅游市场预测研究[D]: [硕士学位论文]. 南昌: 东华理工大学, 2017.

[2] Seetanah, B., Sannassee, R. and Rojid, S. (2015) The Impact of Relative Prices on Tourism Demand for Mauritius: An Empirical Analysis. *Development Southern Africa*, **32**, 363-376. https://doi.org/10.1080/0376835X.2015.1010717

[3] Pai, P.F., Hung, K.C. and Lin, K.P. (2014) Tourism Demand Forecasting Using Novel Hybrid System. *Expert Systems with Applications*, **41**, 3691-3702. https://doi.org/10.1016/j.eswa.2013.12.007

[4] Hadavandi, E., Ghanbari, A., Shahanaghi, K. and Abbasian-Naghneh, S. (2010) Tourist Arrival Forecasting by Evolutionary Fuzzy Systems. *Tourism Management*, **9**, 1-8.

[5] Chu, F.-L. (2011) A Piecewise Linear Approach to Modeling and Forecasting Demand for Macau Tourism. *Tourism Management*, **32**, 1414-1420. https://doi.org/10.1016/j.tourman.2011.01.018

[6] 陈萍萍. 基于时间序列的旅游需求预测模型[J]. 统计与决策, 2013(18): 11-13.

[7] 廖治学, 戈鹏, 任佩瑜, 等. 基于 AB@G 集成模型的九寨沟景区游客量预测研究[J]. 旅游学刊, 2013, 28(4): 88-93.

[8] 王小川. MATLAB 神经网络 43 个案例分析[M]. 北京: 北京航空航天大学出版社, 2013, 20-22, 207-208, 327-328.

[9] 韩中庚. 数学建模方法及其应用[M]. 北京: 高等教育出版社, 2017: 345-350.

[10] 张德丰. MATLAB 神经网络编程[M]. 北京: 化学工业出版社, 2011: 110-130.

[11] 杨思锐. 基于 GA-MIV-BP 算法的二手车估价模型研究[D]: [硕士学位论文]. 重庆: 重庆理工大学, 2020.

[12] 何芳, 王小川, 肖森予, 李晓丽. 基于 MIV-BP 型网络实验的房地产项目风险识别研究[J]. 运筹与管理, 2013, 22(2): 229-234.