

The Grey Neural Network Combination Model Based on Complex Simpson Formula is Used to Predict the Gross National Income of China

Lang Yu, Lizhi Yang, Guoxin Tang

School of Science, Southwest University of Science and Technology, Mianyang Sichuan
Email: swust2018yl@163.com

Received: Nov. 20th, 2018; accepted: Dec. 5th, 2018; published: Dec. 12th, 2018

Abstract

Aiming at the non-linearity and randomness of China's gross national income data, this paper establishes a grey neural network combined forecasting model which is improved by using the complex Simpson formula to predict China's gross national income. Firstly, the GM(1,1) model and the GM(1,1) model based on the complex Simpson integral formula to improve the background value are established. Secondly, a BP neural network prediction model is established. Furthermore, the GM(1,1) model and the BP neural network prediction model improved by the complex Simpson formula are optimally linearly weighted, and the objective function is solved by using the genetic algorithm to obtain the objective function with the smallest squared error sum. Weighting factors are to improve prediction accuracy. Finally, the GM(1,1) model, the GM(1,1) model of the modified Simpson formula, the BP neural network model and the improved gray neural network model of the modified Simpson formula are used to compare the results of the gross national income forecast. From the prediction results, the gray neural network model improved by the complex Simpson formula is significantly better than the GM(1,1) model, the GM(1,1) model of the modified Simpson formula and the BP neural network model.

Keywords

Complex Simpson Formula, Grey Neural Network, Combined Forecast Model, Least Square Method, Genetic Algorithm

基于复化Simpson公式改进的灰色神经网络组合模型对国民总收入的预测

于 浪, 杨析智, 唐国鑫

文章引用: 于浪, 杨析智, 唐国鑫. 基于复化 Simpson 公式改进的灰色神经网络组合模型对国民总收入的预测[J]. 统计学与应用, 2018, 7(6): 593-604. DOI: 10.12677/sa.2018.76068

西南科技大学理学院, 四川 绵阳
Email: swust2018yl@163.com

收稿日期: 2018年11月20日; 录用日期: 2018年12月5日; 发布日期: 2018年12月12日

摘要

针对我国国民总收入数据具有非线性、随机性等特点, 本文建立复化Simpson公式改进的灰色神经网络组合预测模型来对我国国民总收入进行预测研究。首先, 建立了GM(1,1)模型和基于复化Simpson积分公式改进背景值的GM(1,1)模型。其次, 建立了BP神经网络预测模型。进一步, 将复化Simpson公式改进的GM(1,1)模型和BP神经网络预测模型进行最优线性加权组合, 运用最小二乘法思想, 误差平方和最小为目标函数, 结合遗传算法求解目标函数, 得到各模型的加权系数, 以提高预测精度。最后, 对比分析GM(1,1)模型、复化Simpson公式改进的GM(1,1)模型、BP神经网络模型和复化Simpson公式改进的灰色神经网络模型对国民总收入预测的结果。从预测结果看出, 复化Simpson公式改进的灰色神经网络模型对国民总收入的预测明显优于其他预测模型。

关键词

复化Simpson公式, 灰色神经网络, 组合预测模型, 最小二乘法, 遗传算法

Copyright © 2018 by authors 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

GNI, 即国民总收入(Gross National Income), 是指“一国经济总体当期从国内生产和国外生产中获得的生产性收入总量”, 是该经济体所有常住单位在一定时期内收入初次分配的最终结果。GNI 等于 GDP 加上来自国外的净要素收入, 一般来说来自国外的要素净收入远小于本国 GDP 的规模, 因此 GNI 与 GDP 往往规模相当, 二者都被用于反映一国的经济规模, 见(蒋萍, 2011 [1]; 侯瑜, 2012 [2])。因此, 科学、准确地对人均 GNI 做出预测为政府部门有效的调控经济政策和制定发展规划有着重要的作用。

鉴于 GDP 在经济分析与政策制定中的重要性, GDP 数据质量历来备受关注, 国内外很多文献从不同角度、使用不同方法对 GDP 数据预测做了大量研究工作。侯青霞[3]利用线性自回归模型和函数系数自回归模型对我国人均 GDP 进行建模预测, 张雅清[4]运用 GM(1,1)模型对山西省人均 GDP 进行模拟预测, 并用残差修正的 GM(1,1)模型对预测进行改进, 王永杰[5]采用神经网络算法对 GDP 和 CPI 进行预测, 并提出了 PCA-PSO-BP 模型。其中 GM(1,1)模型对“小样本、贫信息”系统具有良好的预测效果。然而, 经典 GM(1,1)模型对原始数据只进行一阶累加, 导致模型的自由度较小。为此, 沈燕与张丽玲[6]将复化 Simpson 引入到灰色模型中, 详细讨论了模型的性质并给出了详尽的数值计算实例。

本文在上述研究的基础上, 通过复化 Simpson 改进了灰色模型的背景值, 利用神经网络能够逼近任意函数的优点, 将影响 GNI 发展的 12 种因素(第一产业、第二产业、第三产业、农林牧渔、工业、建筑、批发零售、交通运输、住宿和餐饮、金融、房地产、人均国内生产总值)作为神经网络的输入, 将复化 Simpson 改进的 GM(1,1), CSGM(1,1)与 BP 神经网络进行误差最小线性加权组合, 利用遗传算法能够智能寻优的

特点,求得最优权系数,并对我国国民总收入进行了预测研究。从中国统计年鉴获得的1987年~2007年的我国国民总收入数据作为建模数据,2008~2016年的数据作为模型的检验数据,对比分析GM(1,1)模型、CSGM(1,1)模型、BP神经网络模型、复化Simpson公式改进的灰色神经网络模型的预测精度。

2. 研究方法

2.1. GM(1,1)预测模型

设原始序列 $X^{(0)}$ 为非负原始序列: $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$,对于满足光滑条件序列,可建立灰色微分方程。经过累加生成 $X^{(1)}$,称 $X^{(1)}$ 为 $X^{(0)}$ 的1-AGO序列: $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$,其中:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad (k=1, 2, \dots, n) \quad (1)$$

$Z^{(1)}$ 为 $X^{(1)}$ 的紧邻均值生成序列: $Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n))$,其中:

$$z^{(1)}(k) = 0.5(x^{(1)}(k) + x^{(1)}(k-1)), \quad k=2, 3, \dots, n \quad (2)$$

为此,灰微分方程为:

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad k=2, 3, \dots, n \quad (3)$$

其中, a 发展系数, b 为灰色作用量。若 $u = [a, b]^T$ 为参数列,且有:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \quad (4)$$

则式(3)的最小二乘估计参数满足:

$$\hat{a} = (B^T B)^{-1} B^T Y \quad (5)$$

若 $B, Y, \hat{u}$ 满足式(4)和式(5),相应的白化微分方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (6)$$

则,GM(1,1)模型的时间序列为:

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

还原后, $x^{(0)}$ 的预测值为:

$$\begin{aligned} \hat{x}^{(0)}(k+1) &= \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \\ &= (1 - e^a) \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak}, \quad k=1, 2, \dots, n \end{aligned} \quad (8)$$

2.2. 背景值误差产生分析

由原始GM(1,1)的还原时间响应序列知,GM(1,1)预测模型的精度取决于参数 a, b ,然而决定参数 a, b 的是 $z^{(1)}(k)$,因此GM(1,1)模型误差主要来源于 $z^{(1)}(k)$ 的计算。对 $\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b$ 在 $[k-1, k]$ 进

行积分, 有:

$$I = \int_{k-1}^k dx^{(1)}(t) + a \int_{k-1}^k x^{(1)}(t) dt = b \int_{k-1}^k dt, \quad k = 2, 3, \dots, n-1 \quad (9)$$

式(9)联合 $x^{(0)}(k) + az^{(1)}(k) = b$, 有:

$$z^{(1)}(k) = \int_{k-1}^{k+1} x^{(1)}(t) dt \quad (10)$$

GM(1,1)模型的背景值的几何意义如图所示, 可以明显看出, 曲线 $x^{(1)}(t)$ 在区间 $[k-1, k]$ 上与横轴 t 所围成的曲边梯形的面积即是 GM(1,1)模型的背景值。传统计算 $z^{(1)}(k)$ 的方式是计算梯形的面积, 如图 1 所示。而用复化 Simpson 公式计算的 GM(1,1)模型, 背景值是区间 $[k-1, k]$ 上的曲边梯形的面积, 如图 2 所示。可以看出, 其计算产生的误差明显小于传统的计算方式。

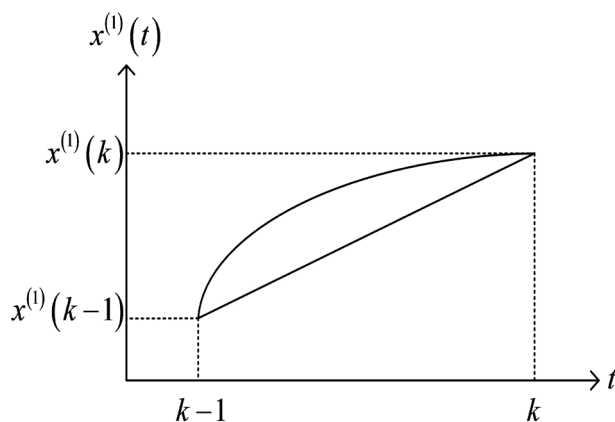


Figure 1. Comparison between the real value and the trapezoid formula background value

图 1. 标准试验系统结果曲线

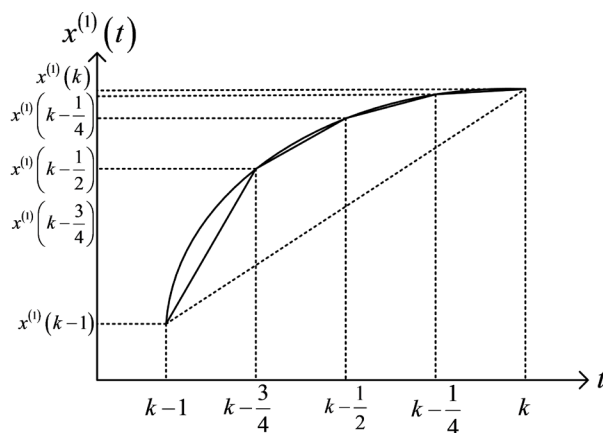


Figure 2. Comparison of the true value and the background value of complex Simpson formula

图 2. 真实值与复化 Simpson 公式的背景值的比较

2.3. 基于复化 Simpson 公式改进的 GM(1,1)模型

由于在使用复化 Simpson 积分公式时需要知道在每一个小区间内的插入分点所对应的值, 所以在此引入 1-AGO 动态序列预测模型[7], 从其他学者的研究成果可知, 1-AGO 序列成非齐次指数形式, 这条曲线只

是一条近似曲线。故可以使用这条曲线构造小区间端点值的一个动态序列模型。由此, 记 $[k, k+1]=[a, b]$, 则可以求出每个区间 $[k, k+1]$ 上 n 等分点处的函数值(即 $f(a+k)$ 的值), 然后利用复化 Simpson 求积公式计算出优化的背景值。设非负原始数据序列为 $X^{(0)}=\{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\}, x^{(0)}(k)>0$

令其 1-AGO 累加序列为: $x^{(1)}(k)=Ae^{\alpha(k-1)}+B$, 则有:

$$\alpha = \ln \frac{x^{(0)}(k)}{x^{(0)}(k-1)} \quad (11)$$

$$A = \frac{x^{(0)}(k)}{\left(\frac{x^{(0)}(k)}{x^{(0)}(k-1)}\right)^{k-1} - \left(1 - \frac{x^{(0)}(k-1)}{x^{(0)}(k)}\right)} \quad (12)$$

$$B = x^{(0)}(k_1) - A = x^{(0)}(k_1) - \frac{x^{(0)}(k)}{\left(\frac{x^{(0)}(k)}{x^{(0)}(k-1)}\right)^{k-1} - \left(1 - \frac{x^{(0)}(k-1)}{x^{(0)}(k)}\right)} \quad (13)$$

称 $x^{(1)}(t)=Ae^{\alpha(t-1)}+B, t \geq 1$ 为其 1-AGO 动态序列预测模型。

记积分区间为 $[k-1, k]$, 将区间等分成 $2m=p$ 份, 各节点为

$$x_{i_{k-1}}^{(1)} = x^{(1)}(k-1+i_{k-1}h), (i=0, 1, 2, \dots, 2m), h=\frac{1}{p}, m=\frac{p}{2}, m=1, 2, \dots, \frac{p}{2} \quad (14)$$

则考虑(15)式在 $[k-1, k]$ 上的积分, 有

$$I = \int_{k-1}^k dx^{(1)}(t) + a \int_{k-1}^k x^{(1)}(t) dt = b \int_{k-1}^k dt, k=2, 3, \dots, n \quad (15)$$

式(15)可进一步表示为:

$$I = x^{(1)}(k) - x^{(1)}(k-1) + a \int_{k-1}^k x^{(1)}(t) dt = b, k=2, 3, \dots, n \quad (16)$$

结合 1-AGO 动态序列预测模型, 运用数值积分方法中的复化 Simpson 积分公式, 计算 $\int_{k-1}^k x^{(1)}(t) dt$ 在 $[k-1, k]$ 上的积分, 有:

$$\begin{aligned} I_1 &= \int_{k-1}^k x^{(1)}(t) dt \\ &= \frac{1}{3} \left[x^{(1)}(k) - x^{(1)}(k-1) + 2 \sum_{i_{k-1}}^{m-1} x_{2i_{k-1}}^{(1)} + 4 \sum_{i_{k-1}}^m x_{2i_{k-1}-1}^{(1)} \right], k=2, 3, \dots, n \end{aligned} \quad (17)$$

则(15)式可表示为:

$$\begin{aligned} I &= x^{(1)}(k) - x^{(1)}(k-1) + a \int_{k-1}^k x^{(1)}(t) dt \\ &= x^{(0)}(k) + \frac{a}{3} \left[x^{(1)}(k) - x^{(1)}(k-1) + 2 \sum_{i_{k-1}}^{m-1} x_{2i_{k-1}}^{(1)} + 4 \sum_{i_{k-1}}^m x_{2i_{k-1}-1}^{(1)} \right] \\ &= b, k=2, 3, \dots, n \end{aligned} \quad (18)$$

对比(16)式得背景值 $z^{(1)}(k)$ 为:

$$z^{(1)}(k) = \frac{a}{3} \left[x^{(1)}(k) - x^{(1)}(k-1) + 2 \sum_{i_{k-1}}^{m-1} x_{2i_{k-1}}^{(1)} + 4 \sum_{i_{k-1}}^m x_{2i_{k-1}-1}^{(1)} \right], k=2, 3, \dots, n \quad (19)$$

进一步化简(18)式有:

$$3x^{(0)}(k) + a \left[x^{(1)}(k) - x^{(1)}(k-1) + 2 \sum_{i_{k-1}}^{m-1} x_{2i_{k-1}} + 4 \sum_{i_{k-1}}^m x_{2i_{k-1}-1} \right] - 3b = 0, \quad k = 2, 3, \dots, n \quad (20)$$

记 $u = [a, b]^T$ 为参数列, 由最小二乘法原理知, 参数列需满足:

$$\hat{u} = (B_2^T B_2)^{-1} B_2^T Y_2 \quad (21)$$

其中:

$$B_2 = \begin{pmatrix} -\frac{1}{3} \left[x^{(1)}(2) + x^{(1)}(1) + 2 \sum_{i_1}^{m-1} x_{2i_1} + 4 \sum_{i_1}^m x_{2i_1-1} \right] & 1 \\ -\frac{1}{3} \left[x^{(1)}(3) + x^{(1)}(2) + 2 \sum_{i_2}^{m-1} x_{2i_2} + 4 \sum_{i_2}^m x_{2i_2-1} \right] & 1 \\ \vdots & \vdots \\ -\frac{1}{3} \left[x^{(1)}(n-1) + x^{(1)}(n-2) + 2 \sum_{i_{n-2}}^{m-1} x_{2i_{n-2}} + 4 \sum_{i_{n-2}}^m x_{2i_{n-2}-1} \right] & 1 \\ -\frac{1}{3} \left[x^{(1)}(n) + x^{(1)}(n-1) + 2 \sum_{i_{n-1}}^{m-1} x_{2i_{n-1}} + 4 \sum_{i_{n-1}}^m x_{2i_{n-1}-1} \right] & 1 \end{pmatrix}, \quad Y_2 = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n-1) \\ x^{(0)}(n) \end{pmatrix} \quad (22)$$

2.4. BP 神经网络

BP (back-propagation)神经网络是由 Rumelhard 等科学家于 20 世纪 80 年代提出的, BP 神经网络是一种按误差反向传播算法训练的多层前馈网络, 也是目前应用最广泛的神经网络模型之一。BP 神经网络具有很强的非线性映射能力, 并且网络的隐含层数、各层的处理单元数及网络的学习系数等参数可根据问题的具体情况而设定, 具有很高的灵活性, 因此被应用到很多领域。图 3 展示了 BP 神经网络的简单结构拓扑图:

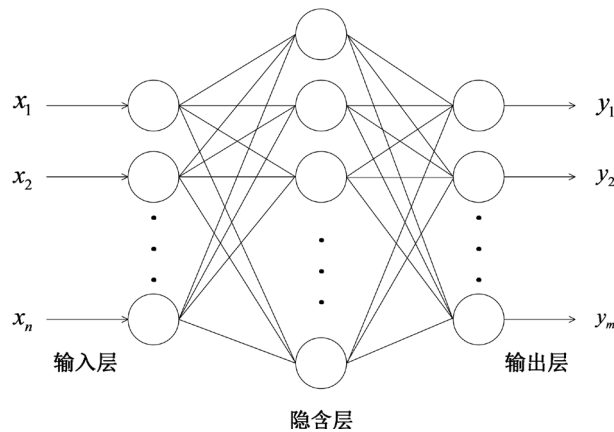


Figure 3. BP network structure topology diagram

图 3. BP 网络结构拓扑图

2.4.1. 神经元模型

一个基本的 BP 神经元模型具有 R 个输入, 每个输入对应一个权值 w , 加权后与下一层连接。网络输出可表示为:

$$Y = f(wp + b) \quad (23)$$

其中为 f 为输入层与输出层的传递函数, w 代表权值, b 为隐含层神经元的偏置值。

2.4.2. 激励函数

BP 神经网络的功能的不同主要取决于激励函数的不同, 激励函数选择的多样性使得神经网络有了不同的信息处理能力。神经元输入层与输出层之间的关系由激励函数 f 决定。神经网络的激励函数有多种, 本文使用的激励函数为 S 型函数, 其表达式为:

$$f(x) = \frac{1}{1 + e^{-cx}} \quad (24)$$

2.4.3. BP 神经网络的学习算法

网络的学习算法分成正向传播和反向传播两部分。在正向传播过程中, 每一层神经元只影响下一层神经元。首先输入样本数据作为训练数据, 将输入数据传送至各个神经元, 经隐含层计算处理后, 输出层输出对应的预测值。如果预测值与实际值的误差较大, 则进入学习算法的第二部分, 输出层进行反向传播该误差, 沿原路返回并逐层调整各层权值, 不断降低误差, 直至误差精度满足我们的要求。学习算法的核心思想是梯度下降法。设一样本 (X, Y) , $X = [x_1, x_2, \dots, x_m]^T$, $Y = [y_1, y_2, \dots, y_m]^T$, 隐含层神经元为 $O = [o_1, o_2, \dots, o_m]^T$ 。输出层与隐含层神经元间的权值矩阵 W_1 和隐含层与输出层神经元之间的网络权值 W_2 分别为:

$$W_1 = (w_{ij}^1)_{l \times m}, \quad i = 1, 2, \dots, l; j = 1, 2, \dots, m \quad (25)$$

$$W_2 = (w_{ij}^2)_{n \times l}, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, l \quad (26)$$

则隐含层神经元的输出为:

$$O_j = f\left(\sum_{i=1}^m w_{ji}^2 x_i - \theta_j^1\right) = f(\text{net}_j), \quad j = 1, 2, \dots, l \quad (27)$$

其中:

$$\text{net}_j = \sum_{i=1}^m w_{ji}^2 x_i - \theta_j^1 \quad (28)$$

$f(\cdot)$ 为隐含层传递函数。

输出神经元的输出为:

$$z_k = g\left(\sum_{j=1}^l w_{kj}^2 O_j - \theta_k^2\right) = g(\text{net}_k), \quad k = 1, 2, \dots, n \quad (29)$$

其中:

$$\text{net}_k = \sum_{j=1}^l w_{kj}^2 O_j - \theta_k^2, \quad k = 1, 2, \dots, n \quad (30)$$

$g(\cdot)$ 为输出层传递函数。

网络输出与期望输出的误差为:

$$E = \frac{1}{2} \sum_{k=1}^n (y_k - z_k)^2 \quad (31)$$

网络的训练误差是各层 w_{ki}^1 和 w_{kj}^2 权值的调整值, 调整权值可改变误差 E , 调整的数学表达式如下:

$$\Delta w_{ji}^1 = -\eta \frac{\partial E}{\partial w_{ji}^1} \quad (32)$$

$$\Delta w_{ji}^2 = -\eta \frac{\partial E}{\partial w_{ji}^2} \quad (33)$$

式(32)、(33)中负号表明梯度下降, $\eta \in (0,1)$ 是常数, 反应了学习过程中的学习梯度。

2.5. 复化 Simpson 神经网络组合预测模型

预测精度是评价一个预测模型好坏的核心指标。在实际预测中, 由于系统复杂, 随机或不确定因素较多, 单一预测模型对系统的考虑较少, 往往不能很好的反映系统的发展过程, 预测存在一定的缺陷。但不同的单一预测模型考虑了系统不同的因素, 把它们设法进行优化组合, 综合考虑单一预测模型有用的信息, 能从各个方面提现系统的发展变化。在将单一预测模型进行组合时, 由不同的方法推导加权系数, 能得到多种不同的组合预测模型。本文采用最优线性组合方式, 其基本思想就是根据某种准则构造相应的目标函数, 在有关约束条件下求得目标函数的最值, 从而得到最优的组合加权系数, 最终得到优化组合预测模型。

本文建立的 GM(1,1)模型只适用于中短期的预测, 而且数据要求呈指数增长规律。BP 神经网络则需要大量的数据样本进行训练, 而且学习时间长。本文综合考虑了单一预测模型的不足, 将 CSGM(1,1)模型和 BP 神经网络模型通过线性最优加权组合的方式, 以误差平方和最小为目标函数, 使用智能算法—遗传算法来求解确定最优组合的权系数, 进行优化组合以达到提高预测精度的目的。组合结构如图 4 所示:

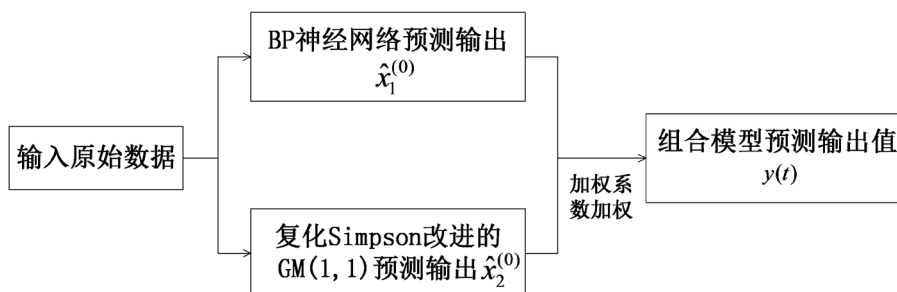


Figure 4. BP network structure topology diagram

图 4. BP 网络结构拓扑图

令 $\hat{x}_1^{(0)}(t)$, $\hat{x}_2^{(0)}(t)$, $\hat{y}(t)$ 分别表示 BP 神经网络模型、复化 Simpson 公式改进的 GM(1,1)模型和组合预测模型在 t 时刻的预测输出值, w_1 , w_2 为 BP 神经网络模型、复化 Simpson 公式改进的 GM(1,1)模型的组合权系数, 则:

$$\hat{y}(t) = w_1 \hat{x}_1^{(0)}(t) + w_2 \hat{x}_2^{(0)}(t) \quad (34)$$

且 w_1 , w_2 满足

$$\begin{cases} w_1 + w_2 = 1 \\ 0 \leq w_1, w_2 \leq 1 \end{cases} \quad (35)$$

令 $e_1(t)$, $e_2(t)$, $e(t)$ 分别为 BP 神经网络模型、CSGM(1,1)模型、组合预测模型在 t 时刻的误差, 则:

$$\begin{cases} \hat{x}_1^{(0)}(t) - x^{(0)}(t) = e_1(t) \\ \hat{x}_2^{(0)}(t) - x^{(0)}(t) = e_2(t) \\ w_1 \hat{x}_1^{(0)}(t) + w_2 \hat{x}_2^{(0)}(t) = e(t) \end{cases} \quad (36)$$

令 $e = \sum_{t=1}^n |e(t)|^2$ ，即误差绝对值平方和，进一步表示为：

$$\min e = \sum_{t=1}^n |e(t)|^2 = \sum_{t=1}^n |w_1 e_1(t) + w_2 e_2(t)|^2 \quad (37)$$

由最小二乘法的思想，对 $e = \sum_{t=1}^n |e(t)|^2$ 中的 w_1 ， w_2 分别求偏导数有：

$$\begin{cases} \frac{\partial e}{\partial w_1} = \sum_{t=1}^n 2(w_1 e_1(t) + w_2 e_2(t)) e_1(t) = 0 \\ \frac{\partial e}{\partial w_2} = \sum_{t=1}^n 2(w_1 e_1(t) + w_2 e_2(t)) e_2(t) = 0 \end{cases} \quad (38)$$

令：

$$E = \begin{bmatrix} e_1(1) & e_1(2) & \cdots & e_1(n) \\ e_2(1) & e_2(2) & \cdots & e_2(n) \end{bmatrix} \quad (39)$$

$$W = \begin{bmatrix} w_1 e_1(1) + w_2 e_2(1) \\ w_1 e_1(2) + w_2 e_2(2) \\ \vdots \\ w_1 e_1(n) + w_2 e_2(n) \end{bmatrix} \quad (40)$$

则上式方程组表示成矩阵形式有：

$$EW = 0 \quad (41)$$

由于为非线性最小二乘问题，若用经典的求解非线性问题的算法，此处求解非常复杂。故本文使用智能算法—遗传算法来求解(38)式，以此来求得最终的组合预测模型的权系数 w_1 ， w_2 。

遗传算法(Genetic Algorithm)是 Holland 在 1975 年提出的一种借鉴生物界的进化规律演化而来的有效解决最优化问题的自适应全局优化搜索算法。遗传算法能有效的解决多峰函数优化、多目标问题以及非线性问题等。遗传算法将问题求解转化成“染色体”，即用编码表示字符串。遗传算法从一群染色体出发，不断的交叉变异，产生新的子代，同时算法将更优的个体保留下来，随着算法的不断进行从而产生更好的个体。与传统的优化算法相比，遗传算法可以有效避免许多求解障碍，直接以目标函数值作为搜索信息。具有自组织、自适应和自学习性等特点。本文以(38)式为目标函数，运用遗传算法进行寻优求解，以得到最优权系数，以此来构建最优组合预测模型。

3. 我国国民总收入预测

为了对上述模型的适用性进行验证，本文从中国统计年鉴选取影响我国 1978~2016 年的国民总收入的 12 个相关因素(第一产业、第二产业、第三产业、农林牧渔、工业、建筑、批发零售、交通运输、住宿和餐饮、金融、房地产、人均国内生产总值)。利用 GM(1,1)和复化 Simpson 改进的 GM(1,1)对我国国民总收入进行建模预测。使用 1978—2007 年的相关数据作为神经网络的输入层，国民总收入作为目标输出，以此对我国国民总收入进行建模预测。最后将复化 Simpson 改进的 GM(1,1)的预测值与 BP 神经网络的预测值进行组合，得到

组合模型的预测值。使用 2008—2016 年的数据作为验证数据检验模型的误差精度。相关数据见表 1:

Table 1. Actual GNI data from 2008 to 2016
表 1. 2008~2016 年的 GNI 实际数据

年份	实际值	年份	实际值	年份	实际值
1978	3678.7	1991	22,050.3	2004	161,415.4
1979	4100.5	1992	27,208.2	2005	185,998.9
1980	4587.6	1993	35,599.2	2006	219,028.5
1981	4933.7	1994	48,548.2	2007	270,844
1982	5380.5	1995	60,356.6	2008	321,500.5
1983	6043.8	1996	70,779.6	2009	348,498.5
1984	7314.2	1997	78,802.9	2010	411,265.2
1985	9123.6	1998	83,817.6	2011	484,753.2
1986	10,375.4	1999	89,366.5	2012	539,116.5
1987	12,166.6	2000	99,066.1	2013	590,422.4
1988	15,174.4	2001	109,276.2	2014	644,791.1
1989	17,188.4	2002	120,480.4	2015	686,449.6
1990	18,923.3	2003	136,576.3	2016	741,140.4

4. 预测结果

本文利用 MATLAB 编程计算, 分别求得 GM(1,1)模型、复化 Simpson 公式改进的 GM(1,1)模型、BP 神经网络模型建立预测模型的预测值。利用遗传算法, 得到 BP 与复化 Simpson 改进的 GM(1,1)模型的组合权系数分别为 $w_1 = 0.425$, $w_2 = 0.574$, 由此建立出复化 Simpson 公式改进的灰色神经网络组合模型: $Y(t) = 0.425\hat{X}_1^{(0)}(t) + 0.574\hat{X}_2^{(0)}(t)$ 。4 种模型的预测值见表 2 和图 5。从表 2 和图 5 中可看出, 所建立的 GM(1,1)模型的预测结果相对较差, CSGM(1,1)模型和 BP 神经网络对 GNI 预测结果相对 GM(1,1)的预测值有一定的改进, 能反映出 2008 年到 2016 年 GNI 的变化趋势。然而, 复化 Simpson 公式改进的灰色神经网络组合模型的预测值比其他模型更接近真实值。

Table 2. Comparison of the predicted and actual GNI values of China from 2008 to 2016 based on different models
表 2. 基于不同模型的中国 2008~2016 年 GNI 预测值与实际值的比较

年份	实际值	BP	CSGM(1,1)	GM(1,1)	组合预测模型
2008	321,500.5	307,020.8	273,138.6	400,044.6	287,265.4
2009	348,498.5	375,943.9	312,980.9	459,224.2	339,427.2
2010	411,265.2	480,925.3	358,635	527,158.4	410,249.7
2011	484,753.2	591,258.7	410,948.5	605,142.4	487,169.4
2012	539,116.5	648,890.1	470,893	694,662.6	546,070.8
2013	590,422.4	695,708.5	539,581.4	797,425.9	605,395.9
2014	644,791.1	713,382.3	618,289.3	915,391.2	658,085.5
2015	686,449.6	736,190.6	708,478.3	1,050,807	719,547.5
2016	741,140.4	732,242.4	811,823	1,206,256	777,189.4

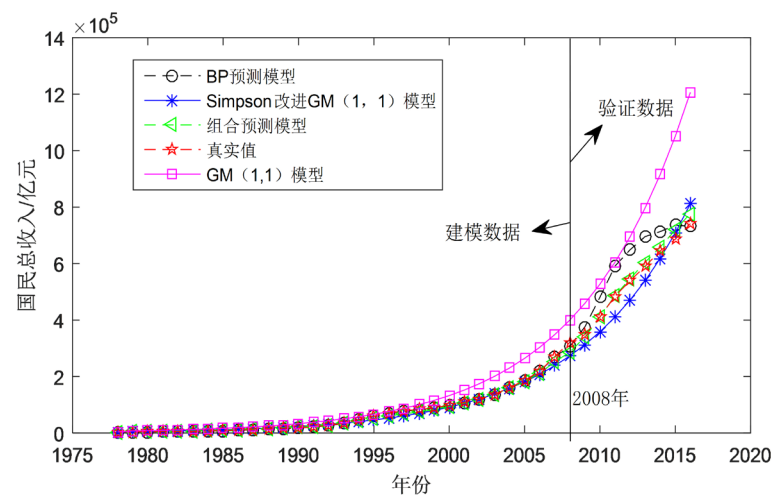


Figure 5. Comparison of predicted and actual GNI values of China from 1978 to 2016 based on different models

图 5. 基于不同模型的中国 1978~2016 年 GNI 预测值与实际值的比较

为了比较 GM(1,1), CSGM(1,1), BP, 复化 Simpson 公式改进的灰色神经网络组合模型的预测精度, 分别计算 4 种模型验证数据的 RMSPEPO 以及相对误差, 其结果见表 3。此外, 为了更直观的比较 4 种模型的预测精度, 图 6 给出了 4 种模型的相对误差。

Table 3. Relative error between predicted and actual GNI of per capita based on different models

表 3. 基于不同模型的人均 GNI 预测值与实际值相对误差

年份	组合预测模型	BP	CSGM(1,1)	GM(1,1)
2008	0.106485	0.045038	0.150426	0.244305
2009	0.026030	0.078753	0.101916	0.317722
2010	0.002469	0.169380	0.127972	0.281797
2011	0.004984	0.219711	0.152252	0.248351
2012	0.012900	0.203618	0.126547	0.288520
2013	0.025361	0.178323	0.086110	0.350602
2014	0.020618	0.106377	0.041101	0.419671
2015	0.048216	0.072461	0.032091	0.530786
2016	0.048640	0.012006	0.095370	0.627568
RMSPEPO	4.4688%	13.9508%	10.9451%	38.8671%

从表 3、图 5 可看出, GM(1,1) 的 RMSPEPO 高达 38.8671%, CSGM(1,1) 模型的 RMSPEPO 为 10.9451%, BP 神经网络模型的 RMSPEPO 为 13.9508%, 复化 Simpson 公式改进的灰色神经网络组合模型的 RMSPEPO 为 4.4687%。因此, 计算结果表明, 复化 Simpson 公式改进的灰色神经网络组合模型对 GNI 的预测结果优于 GM(1,1) 模型, 复化 Simpson 改进的 GM(1,1) 模型, BP 神经网络模型。

5. 结论

本文利用 Simpson 积分公式对 GM(1,1) 模型的背景值进行改进, 再将其与 BP 神经网络模型组合, 建立复化 Simpson 公式改进的灰色神经网络组合模型对中国国民总收入(GNI)进行预测分析, 其结果表明:

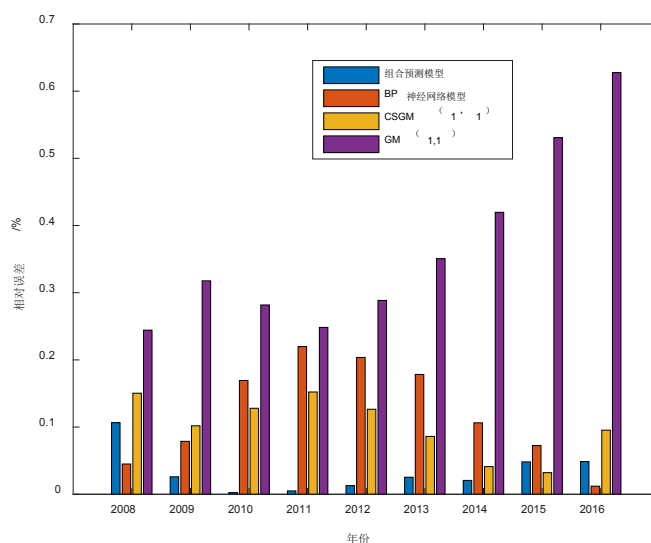


Figure 6. GNI relative error comparison graphs based on different models in verified data

图 6. 基于验证数据中不同模型的 GNI 相对误差对比图

- 1) 复化 Simpson 公式改进的灰色神经网络组合模型用于中国国民总收入(GNI)的方法是切实有效的。
- 2) 复化 Simpson 公式改进的灰色神经网络组合模型预测的精度最高, RMSPEPO 为 4.4687%。其预测精度高于 GM(1,1)模型、BP 神经网络模型和复化 Simpson 改进的 GM(1,1)模型。
- 3) 组合预测的预测精度比单个预测精度更高。同时, 它可以克服单一模型预测的欠缺, 综合结合各种预测模型的长处, 能够更好地降低预测的误差。因此, 组合预测模型可以达到改观单一预测结果的目的。进一步期望本文模型能应用于人口、能源等领域。

基金项目

西南科技大学理学院创新基金项目“基于 scrapy 的房地产数据爬虫系统”(项目编号: LXCX-19), 主持人: 杨羿轩。

参考文献

- [1] 蒋萍. 核算制度缺陷, 统计方法偏颇与经济总量失实[M]. 北京: 中国统计出版社.
- [2] 侯瑜. 基于 DQAF 框架的我国统计数据质量管理及改进[J]. 统计研究, 2012, 29(12): 24-30.
- [3] 侯青霞. 中国人均 GDP 的函数系数自回归模型的预测研究[D]: [硕士学位论文]. 西安: 西安理工大学, 2008.
- [4] 张雅清. 基于残差修正的 GM(1,1)模型在山西省人均 GDP 预测中的应用[J]. 攀枝花学院学报, 2016, 33(2): 68-70.
- [5] 王永杰. 人工神经网络算法在 GDP 和 CPI 中的预测应用[D]: [硕士学位论文]. 太原: 中北大学, 2017.
- [6] 沈艳, 张丽玲. 基于复合辛普森公式的 GM(1,1)模型背景值的优化[J]. 应用科技, 2016, 43(4): 81-84.
- [7] 蒋诗泉, 刘思峰, 周兴才. 基于复化梯形公式的 GM(1,1)模型背景值的优化[J]. 控制与决策, 2014, 29(12): 2221-2225.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 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