

基于时间序列模型和BP神经网络的山东省GDP预测分析

孙欢

云南师范大学数学学院, 云南 昆明

收稿日期: 2022年11月13日; 录用日期: 2022年12月3日; 发布日期: 2022年12月16日

摘要

GDP是衡量国家经济水平的重要指标, 准确预测显得尤为重要。本文主要选择1978年至2021年的山东省GDP数据, 其中将1978年至2018年的数据作为训练集用于构建模型, 2019年至2021年GDP数据作为测试集用于检验模型的准确度, 文章运用ARIMA对山东省GDP建立时间序列模型, 同时采取BP神经网络对GDP数据进行预测。根据模型预测的结果与实际数据对比, 确定最优时间序列模型ARIMA (1, 3, 2), 并对山东省2022年的GDP数值进行预测, 为山东省经济发展提供借鉴和参考。

关键词

时间序列, 山东省GDP, ARIMA模型, BP神经网络

GDP Forecast Analysis of Shandong Province Based on Time Series Model and BP Neural Network

Huan Sun

School of Mathematics, Yunnan Normal University, Kunming Yunnan

Received: Nov. 13th, 2022; accepted: Dec. 3rd, 2022; published: Dec. 16th, 2022

Abstract

GDP is an important indicator to measure the level of national economy, and it is particularly important to forecast it accurately. This paper mainly selects the GDP data of Shandong Province from 1978 to 2021, in which the data from 1978 to 2018 are used as the training set to build the

model, and the GDP data from 2019 to 2021 are used as the test set to test the accuracy of the model. In this paper, ARIMA is used to establish a time series model of GDP in Shandong Province, and BP neural network is used to predict the GDP data. According to the comparison between the predicted results of the model and the actual data, the optimal time series model ARIMA (1, 3, 2) is determined, and the GDP value of Shandong Province in 2022 is predicted to provide reference for the economic development of Shandong Province.

Keywords

Time Series, GDP in Shandong Province, The Model of ARIMA, BP Neural Network

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

国内生产总值(GDP)是一个国家或地区所有常住单位在一定时期内生产活动的最终结果。它反映一国家或地区的经济生产规模与综合实力,是国民经济核算的核心指标,也是衡量一个国家或地区经济状况和发展水平的重要指标。历年来,山东都是我国的经济大省,在国内占据重要地位。因为近几年受疫情等多因素影响,所以山东省经济发展受到冲击,因此面对错综复杂的国内国外经济贸易形势,山东省多年来的增长优势能否继续保持是当地人民最主要关心的事件之一,下面就针对这一主要问题进行研究分析。

本文首先选用 ARIMA 模型对山东省 1978 年~2018 年的 GDP 数据作时间序列,根据数据的变化规律,建立时间序列模型,预测 2019 年至 2020 年的 GDP 数据,与 BP 神经网络得出的山东省 GDP 作对比,检验结果表明 ARIMA 模型具有较好的预测效果,为山东省制定经济发展目标提供决策参考。

2. ARIMA 模型构建

2.1. ARIMA 模型的结构

ARIMA 模型又称为求和自回归移动平均模型,由 Box-Jenkins 首次提出。ARIMA 模型简记为 ARIMA (p, d, q) 模型[1]:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_t \varepsilon_t) = 0, \forall s < t \end{cases}$$

式中, B 为延迟算子, $\nabla^d = (1-B)^d$, $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, 为平稳可逆 ARMA (p, q) 模型的自回归系数多项式; $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, 为平稳可逆 ARMA (p, q) 模型的移动平滑系数多项式。

2.2. 建模步骤

- 1) 判断序列是否为平稳时间序列,如果序列是非平稳,可以通过差分等方法,将它化为平稳时间序列。
- 2) 确定 ARIMA (p, d, q) 模型,选取适当的 p, d, q 值。
- 3) 模型预测和优化[2]。

3. BP 神经网络

3.1. BP 神经网络理论

人工神经网络是一种运算模型,由大量的节点(神经元)和节点之间相互连接构成,每个节点代表一种特定的输出函数,称为激励函数。每两个节点间的连接都代表一个对于通过该连接信号的加权值,称为权重。它具有通过学习逼近任意非线性映射的能力,将神经网络应用在非线性系统的建模与辨别中,可以不受非线性模型的限制。当前应用最广的是多层前向神经网络(BP 神经网络),由三部分组成:输入层、隐含层和输出层。

$$\mu_k = \sum_{i=1}^n \omega_{ik} x_i, y_k = f(v) f(\mu_k + b_k)$$

式中, x_i 表示输入信号, y_i 表示输出信号, $\omega_{ik} (i=1, \dots, m)$ 表示神经元 k 的突触权值, μ_k 表示输入信号线性组合器的输出, $f(v)$ 表示激活函数, 其中 $v = \mu_k + b_k$ 。

3.2. BP 神经网络的步骤

1) 对样本数据进行预处理。BP 神经网络首先对原始数据进行归一化,把数据转化为区间[-1, 1]的数字,这样可以尽可能地平滑数据,消除预测结果的噪声,初始化公式为:

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

其中, x 表示样本序列输入值, x_{\min} 表示输入序列的最小值, x_{\max} 表示输入序列的最大值, X_i 表示归一化后的数据。

2) 划分训练集和测试集。依次将连续 8 年的数据作为网络的一个输入数据,第 9 年的数据作为网络输出,按此方式进行滚动式的排列,形成神经网络的训练样本。

3) 构建 BP 神经网络。

4) 网络参数配置。

5) BP 神经网络训练[3]。

4. 实证分析

4.1. 数据的来源

本文选取的数据是 1978~2021 年期间山东省的 GDP 数据。将 1978~2018 年的数据作为训练集,2019~2021 作为测试集,并预测 2022 年的 GDP。所有数据均来源于《山东统计年鉴》。

4.2. 基于 ARIMA 模型的 GDP 预测

4.2.1. 数据的预处理

将 1978 年至 2018 年的 GDP 数据绘制时间时序图,如图 1 所示,GDP 数据呈现出明显的递增趋势,认定数据是不平稳的;同时对数据做单位根(即 ADF)检验,验证得出数据为非平稳序列。为了使其平稳化,选择用差分方法对其进行平稳化处理,我们对原序列进行差分运算,再者对差分后序列进行 ADF 检验,该序列所有 ADF 检验统计量的 P 值均小于显著性水平($\alpha = 0.05$),所以确定三阶差分后序列平稳(图 2)。再对三阶差分后序列进行纯随机性检验。检验结果显示,各阶延迟下 LB 统计量的 P 值均小于显著性水平,说明差分后序列不是白噪声序列。

山东省GDP

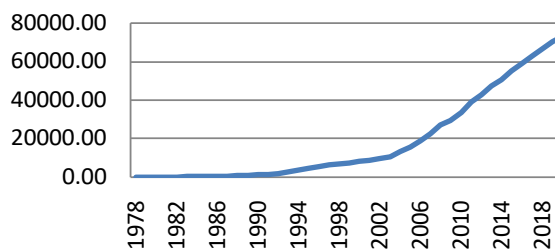


Figure 1. Time series of Shandong GDP

图 1. 山东省 GDP 时序图

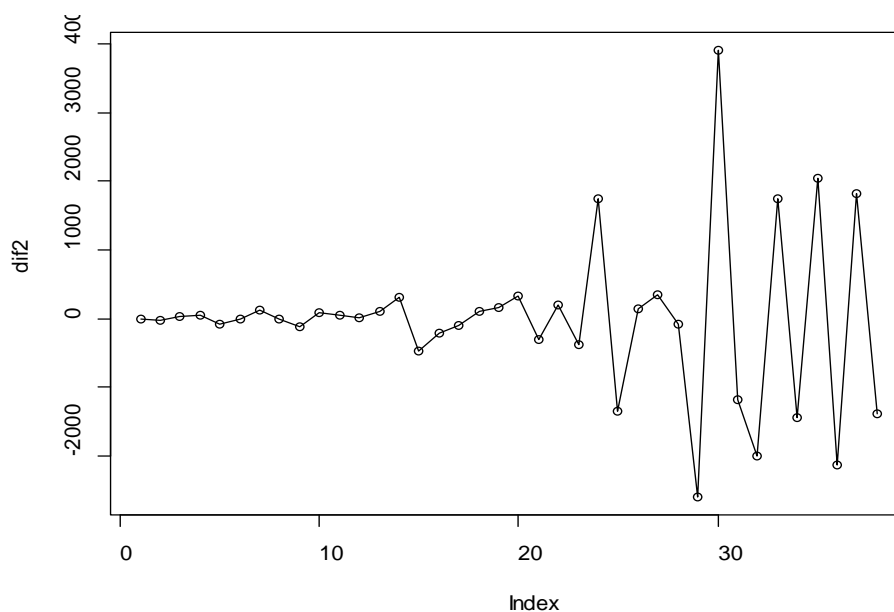


Figure 2. Third order difference time series of Shandong GDP

图 2. 山东省 GDP 三阶差分时序图

4.2.2. ARIMA (p, d, q)模型的识别

ARIMA (p, d, q)模型中未知参数的确定主要依靠分析自相关图和偏自相关图,从图 3 可以观察到序列的自相关图在第一阶差分之后急速衰减,自相关图呈现一阶截尾状态,确定 $q = 1$ 。从图 4 观察偏自相关图在三阶差分之后衰减,为确保 p 、 q 值的选择更加合适,根据 AIC 最小准则,比较不同参数(q)时的 AIC 值,最后选择 ARIMA 模型为 ARIMA (1, 3, 2),此模型下的 AIC = 607.34。

4.2.3. ARIMA (1, 3, 2)模型预测

利用最优模型,对山东省 GDP 进行预测并且与实际值比较如表 1。

4.3. BP 神经网络山东省 GDP 模型预测

本小节以 2001 至 2018 年的数据为初始数据,作为训练集,用于建立模型,2019 至 2021 年数据为模型测试集,并预测 2022 年的 GDP 数值,得到预测结果如表 2。

隐含层神经元的个数通过经验公式确定: $\sqrt{n+m+a}$, n 为输入层层数, m 为输出层层数, $a \in [1,10]$, 隐含层神经元的传输函数选择 $tansig$, 输出层的传输函数为 $purelin$, 反向传播的训练函数为 $trainlm$ [4]。

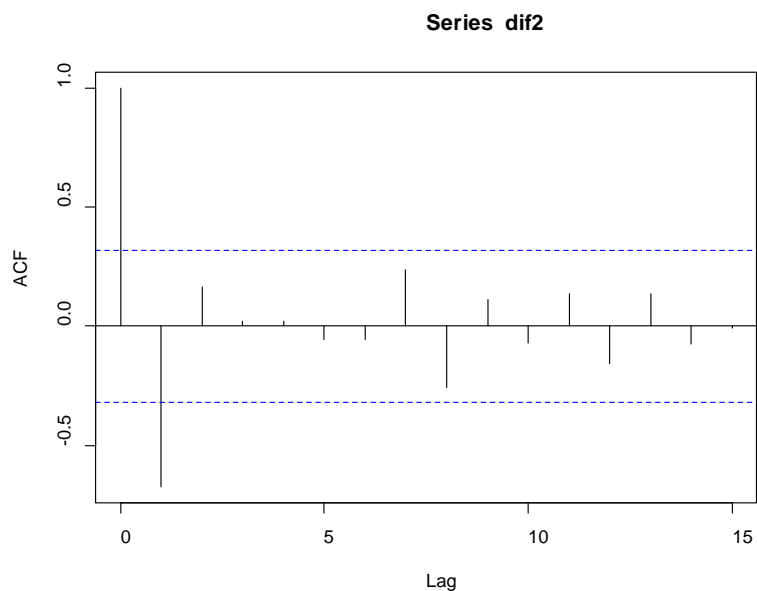


Figure 3. ACF of the third differenced series of GDP
图 3. 三阶差分后 GDP 的 ACF 图

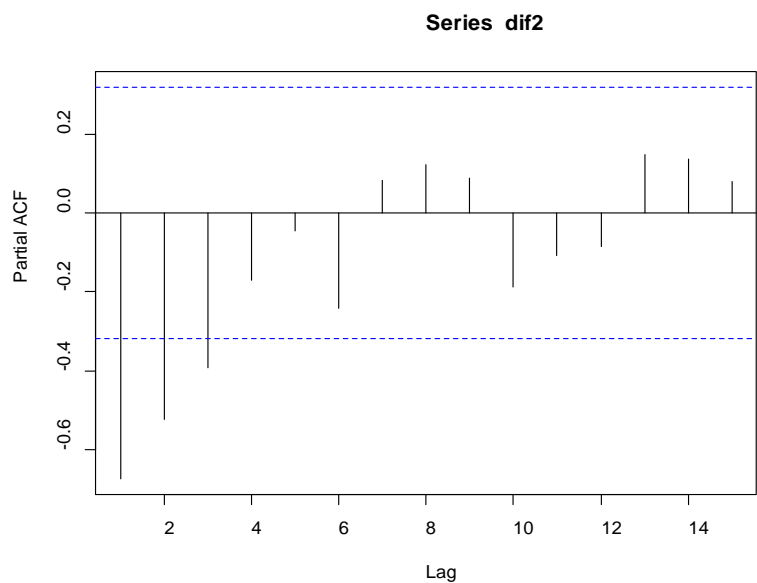


Figure 4. PACF of the third differenced series of GDP
图 4. 三阶差分后 GDP 的 PACF 图

Table 1. Comparison of the forecasts and actual observations in ARIMA (1, 3, 2)
表 1. ARIMA (1, 3, 2) 预测值与实际值的比较

时间	实际值	预测值	相对误差
2019	70540.50	70693.28	0.217%
2020	73129.00	74744.72	2.21%
2021	83095.90	78923.74	-5.02%
2022	---	83194.08	---

Table 2. Comparison of the forecasts and actual observations in BP neural network
表 2. BP 神经网络预测值与实际值对比

时间	实际值	预测值	相对误差
2019	70540.50	69592	-1.345%
2020	73129.00	72821	-0.421%
2021	83095.90	74889	-9.876%
2022	——	87101	——

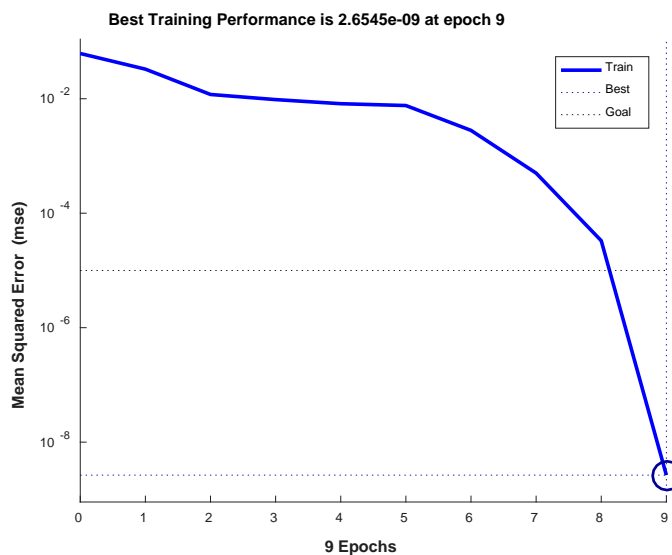


Figure 5. Learning effect curve
图 5. 学习效果曲线图

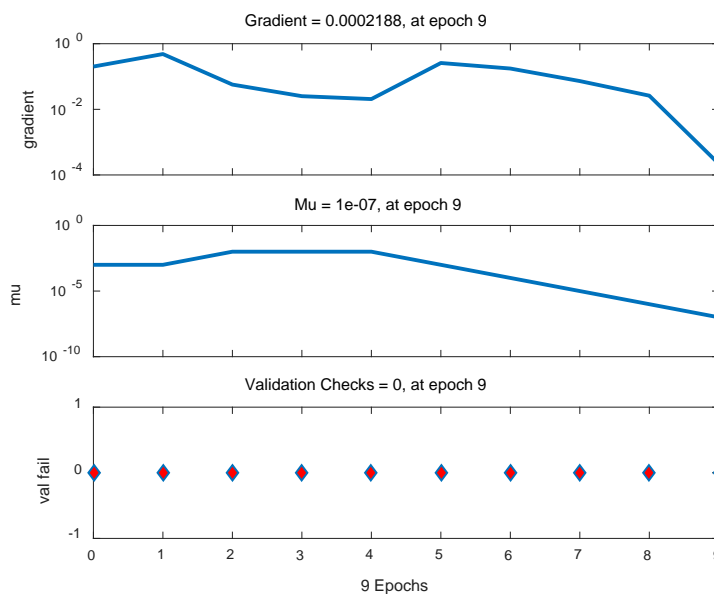


Figure 6. BP neural network training data process chart
图 6. BP 神经网络训练数据进程图

对于给定的数据集，在对 BP 神经网络进行预测性能测试后，最终可以得到神经网络的输出值，在训练次数达到 9 次时，BP 神经网络的均方误差为 2.6545×10^{-9} ，预测得到 2022 年的 GDP 值，得到较好的训练效果，预测精度较高。BP 神经网络的学习效果图如图 5，训练数据图如图 6。

5. 总结

时间序列模型主要是先利用已知数据进行建模，之后对未来数据进行预测的一种方法，BP 神经网络主要通过 *tansig* 可微函数来表达信息输入、输出非线性映射。本文仅使用山东省 GDP 数据这单一指标进行预测，未考虑其他经济因素。文章利用时间序列 ARIMA 模型和 BP 神经网络模型，对山东省 GDP 数值分别进行预测分析，通过预测结果分析，发现在本文中时间序列 ARIMA 模型预测 2019 年到 2021 年的 GDP 数据更接近真实值，相对比 BP 神经网络的预测结果误差小，但从预测 2022 年 GDP 数据来看，BP 神经网络的预测值比时间序列 ARIMA 模型的预测值可靠。

参考文献

- [1] 王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2005.
- [2] 严彦文. 基于 ARIMA 模型的山东省 GDP 的分析与预测[J]. 数学的实践与认识, 2018, 48(4): 285-292.
- [3] 黄雪燕. 神经网络在 GDP 预测中的应用研究[D]: [硕士学位论文]. 吉林: 吉林大学, 2007.
- [4] 尤旭洋. 基于 GA-BP 模型及时间序列分析的 GDP 预测——以四川省为例[J]. 统计理论与实践, 2022(4): 27-34.