

# The Forecast of Yanji GDP Based on BP\_Neural Network Model

Wenyuan Sun, Zhe Yin, Jinghu Shen\*

Mathematics Department, Yanbian University, Yanji Jilin  
Email: [\\*yinzhe@ybu.edu.cn](mailto:yinzhe@ybu.edu.cn)

Received: Jun. 24<sup>th</sup>, 2015; accepted: Jul. 6<sup>th</sup>, 2015; published: Jul. 13<sup>th</sup>, 2015

Copyright © 2015 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

---

## Abstract

Economic prediction problem is a complex system prediction problem of typical multi-index small sample. In this paper, by using SAS software with stepwise regression analysis method to do the preprocessing of the raw data, the data are easier to analyze. Using the methods of simulation fitting, neural network prediction and so on based on the statistical data of Yanji's GDP from 2001 to 2009 after pretreatment, relevant factors which significantly affect economic development situation of Yanji city are worked out. Among them,  $X_1$  is the total cost of the first industry;  $X_4$  is the number of tourists in Yanji city;  $X_5$  is the capital introduced and  $X_6$  is the rate of population growth. Next we put the independence as the target of principal component analysis and predict Yanji city's economic form in the future, having a certain reference value to better grasp Yanji city's economic development.

## Keywords

GDP, Stepwise Regression Analysis, Principal Component Analysis, Neural Network Model

---

# 基于BP\_神经网络模型的延吉市GDP预测

孙文渊, 尹 哲, 沈京虎\*

延边大学数学系, 吉林 延吉  
Email: [\\*yinzhe@ybu.edu.cn](mailto:yinzhe@ybu.edu.cn)

收稿日期: 2015年6月24日; 录用日期: 2015年7月6日; 发布日期: 2015年7月13日

\*通讯作者。

## 摘要

经济预测问题是典型的多指标小样本复杂系统的预测问题。本文利用SAS软件做逐步回归分析法对原始数据进行预处理,使数据便于分析研究,预处理后利用仿真拟合、神经网络预测等方法根据2001~2009年延吉国民生产总值近九年的统计数据,研究出有关影响延吉市经济发展情况的显著影响因素 $X_1$ 第一产业总值、 $X_4$ 延吉市内游客数、 $X_5$ 引入资金及 $X_6$ 人口增长率。再进行独立性为目标的主成分分析并预测延吉市未来经济形式,对更好地掌握延吉市经济发展有一定参考价值。

## 关键词

GDP, 逐步回归分析, 主成分分析, 神经网络模型

## 1. 引言

国内生产总值(GDP)是现代国民经济核算体系的核心指标,是衡量一个国家综合国力的重要指标。这个指标把国民经济全部活动的产出成果概括在一个极为简明的统计数字之中,为评价和衡量国家经济状况、经济增长趋势及社会财富的经济表现提供了一个最为综合的尺度。可以说,它是影响经济生活乃至社会生活的最重要的经济指标。因此对 GDP 进行预测显得十分重要, BP\_神经网络是预测 GDP 的一个很好的方法。

经过几十年的发展, BP\_神经网络也有了多种。1943年,心理学家 W. McCulloch 和数理逻辑学家 W. Pitts 在分析、总结神经元基本特性的基础上首先提出神经元的数学模型。此模型沿用至今,并且直接影响着这一领域研究的进展。因而,他们两人可称为人工神经网络研究的先驱。1945年冯·诺依曼领导的设计小组试制成功存储程序式电子计算机,标志着电子计算机时代的开始。BP神经网络是1986年由Rumelhart和McClelland为首的科学家小组提出,是一种按误差逆传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一[1]。

基于国内生产总值对经济发展的重要性,本文根据延吉统计局的数据,对影响延吉的经济发展因素做了很好的分析,选择其中最主要的四个因素分别为  $X_1$  第一产业总值,  $X_4$  延吉市内游客数,  $X_5$  引入资金,  $X_6$  人口增长率,并根据这些因素建立模型,其中只运用了2001年~2009年的GDP数据,用2010年和2011年数据进行预测检验模型的准确性,从而得出利用神经网络对延吉未来的经济发展情况可以做很好的预测。

## 2. 原始数据的预处理

我们首先分析了影响吉林省GDP的几个重要因素,通过搜集资料和相关经济理论,首先选择第一产业,第二产业,第三产业,因为他们是GDP的构成部分,它们的发展直接体现着GDP的发展;其次市内游客数通过增加消费和交通费来增加经济的增长;引进资金、人口增长率、失业率、人均绿地面积和固定资产投资,它们和GDP密切相关,不可分割均会对GDP造成不可忽视的影响。所以选择这几个因素对GDP进行分析,这样会导致预测结果的精度更高。分析出了影响吉林省GDP的主要因素,我们对这几个因素和GDP之间的关系进行建立模型,然后用建立的方程来预测2010,2011的GDP。

影响延吉市经济发展的原始数据(来源于吉林省统计年鉴)如表1。

在做数据分析处理时,首先我利用了SPSS软件[2]做了对数据的简单处理。结果如表2。

**Table 1. Yanji city in 2001-2009 index data**  
**表 1. 延吉市 2001 年~2009 年各项指标数据**

Y (GDP)	X <sub>1</sub> 第一产业	X <sub>2</sub> 第二产业	X <sub>3</sub> 第三产业	X <sub>4</sub> 市内游客数	X <sub>5</sub> 引进资金	X <sub>6</sub> 人口增长率	X <sub>7</sub> 失业率	X <sub>8</sub> 人均绿地面积	X <sub>9</sub> 固定资产投资
12303.15	1.5	11.3	9.4	20.09	2.66	0.38	4	4	20.89
15079.87	3.4	9.3	11.6	26.96	5.2	3.33	4	4	23.04
17690.29	4.1	13.2	11.5	39.17	2.7	1.94	2	3.94	34.58
21246.8	3.6	23	17.7	133.8	1.65	2.01	3	5.64	56.33
26372.59	3.8	18.8	20.3	214.28	15.3	3.72	2	5	80.03
29392.04	4.6	16.1	20.7	253.5	21.88	3.7	7	8.41	126.07
34180.23	3.1	19.3	19.4	321.1	34.86	1.71	6	8.57	164.1
40902.36	3.6	14.7	17.8	371.93	56.8	2.66	5	8.66	213.58
48048.08	4.5	20.5	11.1	431.76	77.36	2.52	3	8.85	155.58

**Table 2. Statistics description**  
**表 2. 统计量描述**

	N	极小值	极大值	均值	标准差	方差	偏度		峰度		
	统计量	统计量	统计量	统计量	标准误	统计量	统计量	标准误	统计量	标准误	
Y	9	12,303	48,048	27,246	4033.3	12,100	1E+08	0.514	0.717	-0.751	1.4
X <sub>1</sub>	9	1.5	4.6	3.578	0.3072	0.9217	0.849	-1.453	0.717	3.069	1.4
X <sub>2</sub>	9	9.3	23	16.244	1.5091	4.5272	20.495	-0.099	0.717	-1.029	1.4
X <sub>3</sub>	9	9.4	20.7	15.5	1.5055	4.5166	20.4	-0.178	0.717	-2.131	1.4
X <sub>4</sub>	9	20.09	431.76	201.4	51.847	155.54	24193	0.136	0.717	-1.554	1.4
X <sub>5</sub>	9	1.65	77.36	24.268	9.039	27.117	735.33	1.176	0.717	0.339	1.4
X <sub>6</sub>	9	0.38	3.72	2.4411	0.3591	1.0773	1.161	-0.585	0.717	0.25	1.4
X <sub>7</sub>	9	2	7	4	0.577	1.732	3	0.557	0.717	-0.643	1.4
X <sub>8</sub>	9	4	9	6.34	0.745	2.234	4.992	0.085	0.717	-2.332	1.4
X <sub>9</sub>	9	20.89	213.58	97.133	23.414	70.241	4933.9	0.434	0.717	-1.287	1.4
有效的 N	9										

上面表格当中数据可以看出各项指标的极大值、极小值、均值、方差、峰度以及偏度等等。

在实际问题中，影响因变量 Y 的因素很多，希望从中挑选出影响显著地自变量来建立模型，也就是对 Y 影响显著的变量。选择“最优”子集回归的方法，所以利用 SAS 软件用逐步回归分析法对原始数据进行预处理，剔除影响小的变量，表 3 为第四步剔除结果[3]：

利用 SAS 软件进行编程，运行结果如表 3，从表中我们可以看出对 Y (GDP)影响因素最大变量为 X<sub>1</sub> 第一产业总值，X<sub>4</sub>延吉市内游客数，X<sub>5</sub>引入资金，X<sub>6</sub>人口增长率，这四项目标对于研究问题结果有很大影响。经过逐步回归法对原始数据进行预处理，处理后的数据如表 4。

但是这四个因素之间还具有一定线性关系，因此得用主成分分析对其变量进行归一化处理，用综合指标来预测未来的延吉经济发展情况。

Table 3. Analysis of variance

表 3. 方差分析表

Source	Df	Sum of Squares	Mean Squares	F Value	Pr > F
Model	4	1,169,351,574	292,337,893	606.36	<0.0001
Error	4	1,928,485	482,121		
Corrected total	8	1,171,280,059			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	8859.87258	1025.64965	35,976,018	74.62	0.0010
X <sub>1</sub>	1850.54780	431.24438	8,877,894	18.41	0.0127
X <sub>4</sub>	47.56768	4.34999	57,650,654	119.58	0.0004
X <sub>5</sub>	157.44478	24.22661	20,362,326	42.23	0.0029
X <sub>6</sub>	-669.97247	349.77198	1,768,887	3.67	0.1279

Table 4. The indicators in 2001-2009 years after data preprocessing

表 4. 数据预处理后 2001 年~2009 年的各项指标

Years	Y	X <sub>1</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
2001	11182.12	1.2	27.85	5.01	0.56
2002	12283.78	1.24	33.88	5.01	1.01
2003	12303.15	1.5	20.09	2.66	0.38
2004	15079.87	3.4	26.96	5.2	3.33
2005	17690.29	4.1	39.17	2.7	1.94
2006	21246.8	3.6	133.8	1.65	2.01
2007	26372.59	3.8	214.28	15.3	3.72
2008	29392.04	4.6	253.5	21.88	3.7
2009	34180.23	3.1	321.1	34.86	1.71

主成分分析也称主分量分析，旨在利用降维的思想，把多指标转化为少数几个综合指标。在实证问题研究中，为了全面、系统地分析问题，我们必须考虑众多影响因素。这些涉及的因素一般称为指标，在多元统计分析中也称为变量。因为每个变量都在不同程度上反映了所研究问题的某些信息，并且指标之间彼此有一定的相关性，因而所得的统计数据反映的信息在一定程度上有重叠。在用统计方法研究多变量问题时，变量太多会增加计算量和增加分析问题的复杂性，人们希望在进行定量分析的过程中，涉及的变量较少，得到的信息量较多。对数据用 SAS 软件进行主成分分析[3]。

分析结果如表 5~8。

从表中的相关阵的特征值可以看出，前两个主成分的累计贡献率已达 94.4%，因此只需用前两个主分量就能很好的概括这组数据。由最大的两个特征值对应的特征向量可以写出第一、第二主分量：

$$Z_1 = 0.486535X_1^* - 0.540823X_4^* - 0.473899X_5^* + 0.486192X_6^*$$

$$Z_2 = 0.489525X_1^* - 0.395792X_4^* - 0.585389X_5^* + 0.510915X_6^*$$

主成分之间是相互独立的，不存在包含不包含的关系，因此标准化后的因变量直接用主成分做自变量。将四个因素  $X_i = (x_{i1}, x_{i2}, \dots, x_{i9})'$  的值代入  $Z_t$  的表达式，计算得到的值为第  $i$  个因素在第  $t$  个主分量的得分，记为  $z_{it} (t = 1, 2)$ 。

**Table 5. Simple statistics**  
**表 5. 简单统计量**

	X <sub>1</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
Means	2.948888889	118.9588889	10.47444444	2.040000000
Std	1.299234818	116.4114893	11.37401876	1.292710331

**Table 6. Correction matrix**  
**表 6. 相关矩阵**

		X <sub>1</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
X <sub>1</sub>	X <sub>1</sub>	1.0000	0.5439	0.3364	0.8426
X <sub>4</sub>	X <sub>4</sub>	0.5439	1.0000	0.9095	0.4919
X <sub>5</sub>	X <sub>5</sub>	0.3364	0.9095	1.0000	0.3394
X <sub>6</sub>	X <sub>6</sub>	0.8426	0.4919	0.3394	1.0000

**Table 7. Eigenvalues correlation matrix**  
**表 7. 相关矩阵的特征根**

	Eigenvalue	Difference	Proportion	Cumulative
1	2.73854290	1.7010791	0.6846	0.6846
2	1.03743499	0.87204677	0.2594	0.9440
3	0.16538823	0.10675435	0.0413	0.9853
4	0.05863388		0.0147	1.0000

**Table 8. Feature vector**  
**表 8. 特征向量**

		Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>	Z <sub>4</sub>
X <sub>1</sub>	X <sub>1</sub>	0.496535	0.489525	-0.657053	0.286531
X <sub>4</sub>	X <sub>4</sub>	0.540823	-0.395792	-0.198097	-0.715274
X <sub>5</sub>	X <sub>5</sub>	0.473899	-0.585389	0.195853	0.627997
X <sub>6</sub>	X <sub>6</sub>	0.486192	0.510915	0.700486	-0.109100

经计算得出：

$$Z = \begin{bmatrix} 18.3043 & 21.8041 & 13.0553 & 20.3521 & 25.4426 & 75.9088 & 126.8337 & 151.5505 & 192.5490 \\ -13.0821 & -15.2192 & -8.5802 & -10.3488 & -14.0855 & -51.1336 & -90.0060 & -108.9994 & -145.1043 \end{bmatrix}$$

上面的矩阵就是主成分分析后的综合指标，这样两个主成分就能代表影响其因素 X<sub>1</sub>、X<sub>4</sub>、X<sub>5</sub>、X<sub>6</sub> 对 GDP 所带来的影响。以下模型就利用其主分量作为因变量进行仿真模拟、预测。

### 3. 模型的建立

为了更好的反应延吉市经济状况的幅度，采用 BP 神经网络模型对延吉市 GDP 进行拟合，得到相应的一个趋势仿真图，并对此加以分析描述得出结论。

首先，我们给出 BP 神经网络模型的一个图形(如图 1)。

BP 算法基本原理利用输出后的误差来估计输出层的直接前导层的误差，再用这个误差估计更前一层的误差，如此一层一层的反传下去，就获得了所有其他各层的误差估计，并逐渐减小误差。该网络模型

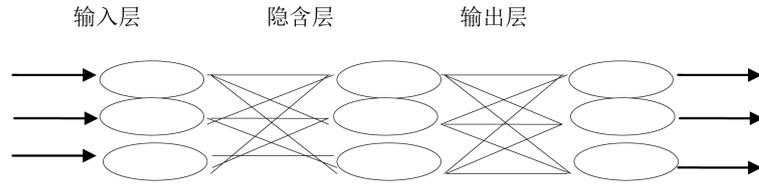


Figure 1. BP neural network model  
图 1. BP 神经网络模型图

分为输入层、隐含层、输出层，共 3 层，其中输入层为量化后的矩阵输入值，隐含层包括影响延吉市 GDP 的数据，分别为  $X_1, X_4, X_5, X_6$ ，输出层是 GDP。通过建立 BP 神经网络模型对历史数据进行提炼，能够在系统内部不知情的情况下，运用软件 Matlab 编程来拟合图像(如图 2)，评价延吉市 GDP，预测并检验延吉市近两年 GDP [4]。

由上面的分析，首先将各项物质进行量化。然后进行网络值初始化，给各连接权值分别赋一个区间  $(0, 1)$  内的随机数，设定误差函数  $e$ ，给定计算精度值  $\varepsilon = 10^{-3}$  和最大学习次数  $M = 5000$ 。随机选取第  $k$  个输入样本及对应期望输出：

$$x(k) = (x_1(k), x_2(k), \dots, x_n(k))$$

$$d_o(k) = (d_1(k), d_2(k), \dots, d_q(k))$$

再计算隐含层各神经元的输入和输出：

$$hi_h(k) = \sum_{i=1}^n w_{ih} x_i(k) - b_h \quad h = 1, 2, \dots, p$$

$$ho_h(k) = f(hi_h(k)) \quad h = 1, 2, \dots, p$$

$$yi_o(k) = \sum_h^p w_{ho} ho_h(k) - b_o \quad o = 1, 2, \dots, q$$

$$yo_o(k) = f(yi_o(k)) \quad o = 1, 2, \dots, q$$

利用网络期望输出和实际输出，计算误差函数对输出层的各神经元的偏导数  $\delta_o(k) \frac{1}{2}$ 。

$$\frac{\partial e}{\partial w_{ho}} = \frac{\partial e}{\partial yi_o} \frac{\partial yi_o}{\partial w_{ho}}$$

$$\frac{\partial yi_o(k)}{\partial w_{ho}} = \frac{\partial \left( \sum_h^p w_{ho} ho_h(k) - b_o \right)}{\partial w_{ho}} = ho_h(k)$$

$$\frac{\partial e}{\partial yi_o} = \frac{\partial \left( \frac{1}{2} \sum_{o=1}^q (d_o(k) - yo_o(k))^2 \right)}{\partial yi_o} = -(d_o(k) - yo_o(k)) yo_o'(k)$$

$$= -(d_o(k) - yo_o(k)) f'(yi_o(k)) \triangleq -\delta_o(k)$$

计算全局误差：

$$E = \frac{1}{2m} \sum_{k=1}^m \sum_{o=1}^q (d_o(k) - yo_o(k))^2$$

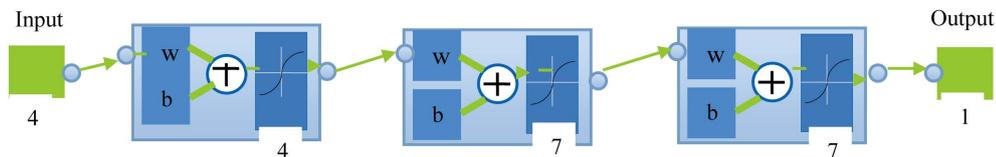


Figure 2. Prediction of structure model  
图 2. 预测模型结构图

最后判断网络误差是否满足要求。当误差达到预设精度或学习次数大于设定的最大次数，则结束算法。

$$\frac{\partial e}{\partial w_{ho}} = \frac{\partial e}{\partial y_{i_o}} \frac{\partial y_{i_o}}{\partial w_{ho}}$$

$$\frac{\partial y_{i_o}(k)}{\partial w_{ho}} = \frac{\partial \left( \sum_h^p w_{ho} h o_h(k) - b_o \right)}{\partial w_{ho}} = h o_h(k)$$

$$\frac{\partial e}{\partial y_{i_o}} = \frac{\partial \left( \frac{1}{2} \sum_{o=1}^q (d_o(k) - y_{o_o}(k))^2 \right)}{\partial y_{i_o}} = -(d_o(k) - y_{o_o}(k)) y_{o_o}'(k)$$

$$= -(d_o(k) - y_{o_o}(k)) f'(y_{i_o}(k)) \triangleq -\delta_o(k)$$

通过表 2 的数据，运用 Matlab 软件[5]进行模拟图形，可得图 3。

从图 3 中可以看出：实际样本与网络输出值之间训练和测试的对比图，显然两者之间非常接近，误差极小，因此能够进行预测。

利用主成分归一化处理后的数据，我们对 2001 年到 2009 年进行预测，其结果如图 4。

从图 4 可以看出，经过量化后的数据图像更加具有说服力，并且 GDP 受这几个指标的显著影响。2002 年至 2003 年一场突如其来的“非典”，使我市的旅游业受到了极大的冲击。因此图形当中模拟值和实际值之间有一定的偏差。但可以从图中看出来模拟数据基本和实际 GDP 基本相符，因此该模型能后很好的对延吉市经济发展做出很好的预测。

#### 4. 模型预测

通过此模型对延吉市经济(GDP) 2010 年和 2011 年的预测结果：

从表 9 可知，预测值与真实数据相比较误差较小，模拟数据基本和实际 GDP 基本相符，因此该模型对于预测未来近几年延吉市 GDP 是有一定的说服力。这样我们就可以通过此模型对延吉市 GDP 进行预测，这也让我们更好的掌握延吉市经济状况，以更好地促进其经济发展。

#### 5. 结论

利用逐步回归分析，得出延吉第一产业总值、延吉市内游客数、延吉市引入资金、延吉市人口增长率等四个显著因素对其影响较大，但保证不了因素之间独立性，通过主成分分析，简化为两个主成分，这两个主成分解释了所有变量 94.4% 的信息，既保证了原始数据的绝大多数信息，又保证了变量间的独立。这样更能简化网络结构，进而通过 BP 网络建立映射关系，得到相应年份的仿真，通过以上的分析，可以清楚反映延吉市过去几年的经济状况变化，实际样本与网络输出值之间训练和测试的对比图，显然两者之间非常接近，误差极小，因此能够进行预测。再对 2001 年~2009 年数据进行模型仿真模拟，由于

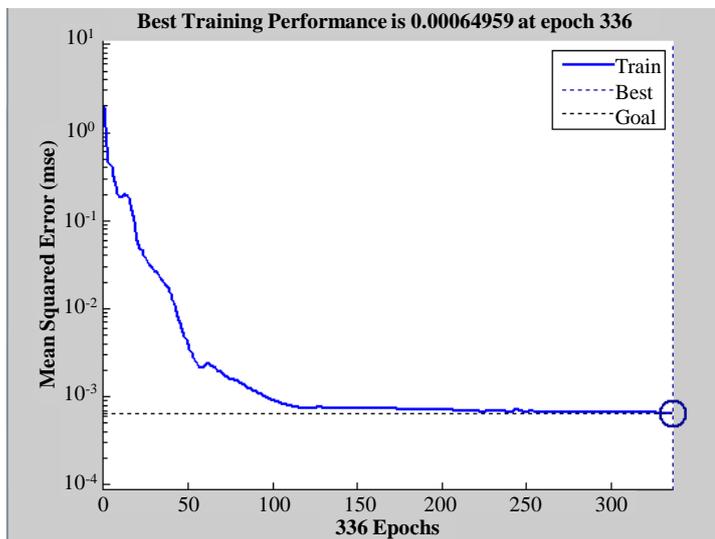


Figure 3. The simulation graphics 1

图 3. 模拟图形 1

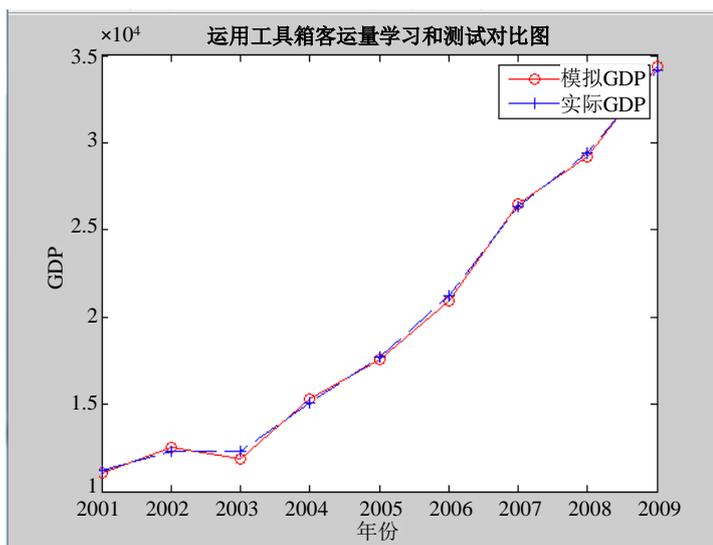


Figure 4. The simulation graphics 2

图 4. 模拟图形 2

Table 9. Forecast map

表 9. 预测图

年份	预测的 GDP	真实 GDP
2010	40991.00	40902.36
2011	47704.00	48048.08

延吉市的旅游业对 GDP 增长有很大影响，所以在 2002 年~2003 年存在“非典”情况，GDP 呈下降状态，以后每年都呈现明显增长趋势，进而总结出延吉市 GDP 的变化情况。然后再对 2010 年和 2011 年的 GDP 进行预测，得出 2010 年和 2011 年的 GDP 值与真实值对比，其误差较小，利用此模型进行预测具有一定参考价值。

### 参考文献 (References)

- [1] 周政 (2008) BP 神经网络的发展现状综述. *山西电子技术*, **2**, 90-92.
- [2] 罗应婷 (2007) SPSS 统计分析从基础到实践. 电子工业出版社, 北京.
- [3] 邓祖新 (2006) 数据分析方法和 SAS 系统. 上海财经大学出版社, 上海.
- [4] 卓金武 (2011) MATLAB 在数学建模中的应用. 北京航空航天大学出版社, 北京.
- [5] 章德斌, 徐家鹏, 许建军, 等 (2010) 基于监测数据和 BP 神经网络的食物安全预警模型. *农业工程学报*, **1**, 221-226.