基于 BP 网络的海洋生产总值预测*

殷克东，张雪娜

（中国海洋大学 青岛 266100）

摘　　要：Lapedest 和 Farber 首次应用人工神经网络进行预测以来，国内外专家学者不断将其与经济学结合，应用于宏观经济预测、企业经济战略预测、证券市场预测等领域。针对当前我国对海洋经济系统研究有限及海洋经济数据缺乏等特点，文章通过建立 BP 网络，对我国 2010 年的海洋生产产值做出预测。

关 键 词：海洋经济；预测；BP 网络

海洋经济是开发利用海洋的各类产业及相关经济活动的总和。进入 21 世纪以来，党中央、国务院又将海洋事业的发展提升到国家战略的高度。2003 年 5 月，《全国海洋经济发展规划纲要》明确提出了“逐步把我国建设成为海洋强国”的目标，表明国家层面的高度重视。对海洋经济发展做出可靠的预测，对我国海洋经济发展前景和目标、制定科学的海洋强国战略和措施，具有极其重要的理论意义和现实意义。

1 BP 神经网络

人工神经网络通过模拟人的大脑神经处理信息的方式，能够对信息进行并行处理和非线性转换。人工神经网络模型，是一种近乎“黑箱”的操作，只需要利用历史数据训练网络，而具体变量之间的函数表达关系不必知晓。尤其是笔者对海洋经济预测研究，由于相关理论基础薄弱，海洋总产值与其影响因素之间很难建立具体的数学函数表达式。鉴于人工神经网络的上述特点，笔者弃用了传统的统计计量模型建模方法，通过建立 BP（Back Propagation）神经网络模型对海洋经济进行预测。

1.1 BP 神经网络的结构

BP 神经网络[1]是一种具有 3 层或 3 层以上神经元的神经网络，上下层神经元之间实现权连接，即下层的每个神经元与上层的每个神经元都实现权连接，而层内各神经元间无连接。典型的 BP 网络是 3 层前向网络，即输入层、隐含层和输出层。BP 神经网络结构见图 1。

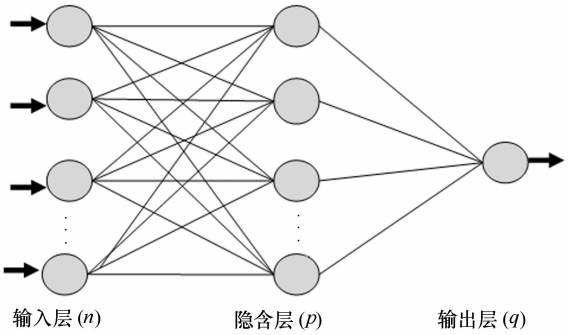

图 1 BP 神经网络结构

1.2 BP 神经网络的学习规则

以下是文章中各符号的形式及意义。

网络输入向量 $\boldsymbol{P}_k=(a_1, a_2, \cdots, a_n)$

网络目标向量 $\boldsymbol{T}_k=(y_1, y_2, \cdots, y_q)$

中间层单元输入向量 $\boldsymbol{S}_k=(s_1, s_2, \cdots, s_p)$，输出向量 $\boldsymbol{B}_k=(b_1, b_2, \cdots, b_p)$；

输出层单元输入向量 $\boldsymbol{L}_k=(l_1, l_2, \cdots, l_q)$，输出向量 $\boldsymbol{C}_k=(c_1, c_2, \cdots, c_q)$；

输出层至中间层的连接权重 w_{ij}，$i=1, 2, \cdots, n$，$j=1, 2, \cdots, p$；

中间层至输出层的连接权重 v_{jt}，$j=1, 2, \cdots, p$，$t=1, 2, \cdots, q$；

中间层各单元的输出阈值 r_t，$t=1, 2, \cdots, q$；

* **基金项目**：国家社科基金课题“沿海地区海洋强国（省）综合实力测评研究”（08BJY006）和国家海洋公益专项“海洋经济监测预警技术与典型区域示范研究（200805081）”阶段性研究成果.

输出层各单元的输出阈值 γ_j，$j=1$，2，…，p；

参数 $k=1$，2，…，m。

（1）初始化。给每个连接权值 w_{ij}、v_{jt}、阈值 θ_j 与 γ_j 赋予区间（−1，1）内的随机值。

（2）随机选取一组输入和目标样本 $\boldsymbol{P}_k=(a_1^k, a_2^k, \cdots, a_n^k)$、$\boldsymbol{T}_k=(s_1^k, s_2^k, \cdots, s_p^k)$ 提供给网络。

（3）用输入样本 $\boldsymbol{P}_k=(a_1^k, a_2^k, \cdots, a_n^k)$、连接权重 w_{ij} 和阈值 θ_j 计算中间层各单元的输入 s_j，然后用 s_j 通过传递函数计算中间层各单元的输出 b_j。

$$s_j=\sum_{i=1}^{n} w_{ij}a_i-\theta_j, \qquad j=1,2,\cdots,p$$

$$b_j=f(s_j), \qquad j=1,2,\cdots,p$$

（4）利用中间层的输出 b_j、连接权重 v_{jt} 和阈值 γ_t 计算输出层个单元的输出 L_t，然后通过传递函数计算输出层个单元的响应 C_t。

$$L_t=\sum_{j=1}^{p} v_{jt}b_j-\gamma_t, \qquad t=1,2,\cdots,q$$

$$C_t=f(L_t), \qquad t=1,2,\cdots,q$$

（5）利用网络目标向量 $\boldsymbol{T}_k=(y_1^k, y_2^k, \cdots, y_q^k)$，网络的实际输出 C_t，计算输出层的各单元一般误差 d_t^k。

$$d_t^k=(y_t^k-C_t)\cdot C_t(1-C_t), \qquad t=1,2,\cdots,q$$

（6）利用连接权 v_{jt}、输出层的一般误差 d_t 和中间层的输出 b_j 计算中间层个单元的一般化误差 e_j^k。

$$e_j^k=\Big[\sum_{t=1}^{q} d_t v_{jt}\Big]b_j(1-b_j), \qquad j=1,2,\cdots,p$$

（7）利用输出层个单元的一般化误差 d_t^k 与中间层个单元的输出 b_j 来修正连接权重 v_{jt} 和阈值 γ_t。

$$v_{jt}(N+1)=v_{jt}(N)+\alpha.\,d_t^k.\,b_j$$

$$\gamma_t(N+1)=\gamma_t(N)+\alpha.\,d_t^k$$

$$t=1,2,\cdots,q, \qquad j=1,2,\cdots,p, \qquad 0<\alpha<1$$

（8）利用中间层个单元的一般化误差 e_j^k，输出层个单元的输入 $\boldsymbol{P}_k=(a_1, a_2, \cdots, a_n)$ 来修正连接权重 w_{ij} 和阈值 r_t。

$$w_{ij}(N+1)=w_{ij}(N)+\beta e_j^k a_i^k$$

$$r_t(N+1)=r_t(N)+\beta e_j^k$$

$$t=1,2,\cdots,q, \qquad j=1,2,\cdots,p, \qquad 0<\beta<1$$

（9）随机选取下一个学习样本向量提供给网络，返回到步骤（3），直到 m 个训练样本训练完毕。

（10）重新从一个学习样本中随机选取一组输入和目标样本，返回步骤（3），直到网络全局误差 E 小于预先设定的一个极小值，即网络收敛。如果学习次数大于预先设定的值，网络就无法收敛。

（11）学习结束。

上述学习过程用图 2 表示。

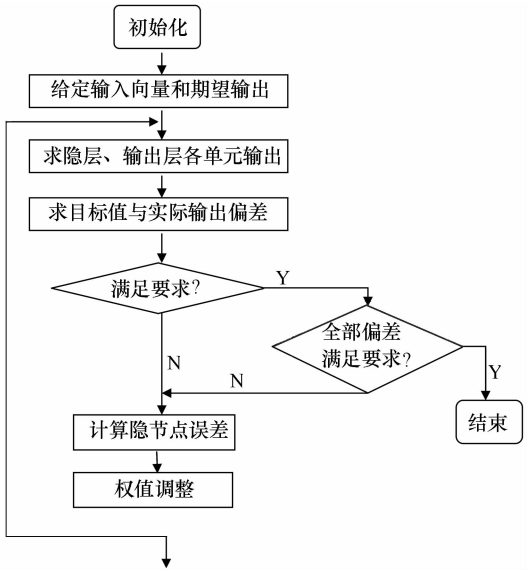

图 2　BP 网络学习流程

总的来说，误差 BP 网络的训练步骤如下。

① 计算输入矢量集中的任一矢量，计算相应的网络输出矢量。② 计算网络实际输出矢量与期望的输出矢量之差值。③ 逐层调整网络的权值，使输出误差减小。④ 重复上述过程，直至误差满足一定的要求。

综上可知，BP 网络具有良好的自适应、自学习、极强的非线性逼近，大规模并行处理和容错能力等特点，具有广泛的应用背景。

2　海洋数据的神经网络模型的建立及预测

在进行 BP 网络预测模型设计时，主要考虑以下因素：网络的层数、隐层中的神经元个数、初始值的选择、学习速率和期望误差。

2.1　网络的层数及神经元的个数

1989 年，Robert Hecht-Nielson 证明了对于任何的连续函数映射关系都可用含有一个隐含层的 BP 网络来逼近，即一个 3 层的 BP 网络

可完成任意的从 n 维到 m 维的映射，并且可通过增加隐含层神经元的个数来提高网络的精度。由此，文章建立三层神经元的 BP 网络。

本文通过检验隐层包含 3～11 个神经元的 BP 网络的性能，最后确定用隐层含 7 个神经元的 BP 网络模型预测海洋生产总值[2]。

2.2 数据预处理及后处理

2.2.1 数据选择及预处理

由于我国海洋经济统计口径在 2001 年重新进行了修订，笔者选择了 2001—2009 年的有连续统计数据且与海洋生产总值密切相关的海洋科研教育管理服务业、海洋相关产业、海洋渔业、海洋交通运输业、滨海旅游业、海洋船舶工业和海洋油气业每年的增加值作为输入层的信息[3]。

考虑到 BP 网络的神经元采用的是 S 型激励函数$\left[f(s_j)=\frac{1}{1+\exp(-s_j)}\right]$，变换后可以防止因净输入的绝对值过大而使神经元输出饱和，使权值调整进入误差曲面的平坦区，同时兼顾网络的训练时间。本文采用了下列预处理方法。

$$P'_{ki}=1-\frac{1}{e\left\{0.162\,5+\frac{(1.897\,1-0.162\,5)\left[\boldsymbol{P}_{ki}-\min(\boldsymbol{P}_k)\right]}{\max(\boldsymbol{P}_k)-\min(\boldsymbol{P}_k)}\right\}},$$

使得 $0.15<\boldsymbol{P}'_{ki}<0.85$。

2.2.2 数据的后处理

数据的后处理即数据的返归一化。根据原公式，求出 $\boldsymbol{P}_k$ 即可。返归一化公式为

$$\boldsymbol{P}_k=\left[\frac{\max(\boldsymbol{P}_k)-\min(\boldsymbol{P}_k)}{1.897\,1-0.162\,5}\right]\left[\ln\left(\frac{1}{1-\boldsymbol{P}'_{ki}}\right)-0.162\,5\right]+\min(\boldsymbol{P}_k)$$

2.3 学习速率

网络的学习通常要有一个合适的学习速率。学习速率决定每一次循环训练中所产生的权值的变化量。大的学习速率可能会导致网络学习的不稳定。但是小的学习速率又会导致训练时间延长，收敛速度较慢，因此，通常倾向于选择较小的学习速率以保证网络的稳定性，一般取值在 0.01～0.6 之间。

2.4 期望误差

期望误差值也是通过网络对不同误差值分别进行训练而确定的最适合值。所谓的最适合值是相对于所需要的隐含层的节点数的，一个较小的误差值的获得需要增加隐含层的节点以及训练时间。

2.5 模型的实现

笔者利用 Matlab7.0 的神经网络工具箱对数据进行处理。其中将 2001—2008 年的数据作为训练样本，用 2009 年的数据对训练好的网络的预测能力进行检验。2001—2009 年海洋生产总值和各海洋产业的增加值见表 1[4]。

表 1 2001—2009 年海洋生产总值及各海洋产业增加值 亿元

年份	海洋生产总值	海洋科研教育管理服务业	海洋相关产业	海洋渔业	海洋交通运输业	滨海旅游业	海洋船舶工业	海洋油气业
2001	9 518.40	1 877.00	3 784.80	966.00	1 316.40	1 072.00	73.63	230.45
2002	11 270.50	2 090.50	4 483.20	1 091.20	1 507.40	1 523.70	79.50	261.20
2003	11 952.30	2 383.30	4 814.60	1 145.00	1 752.50	1 105.80	105.02	335.38
2004	14 662.00	2 882.50	5 951.90	1 271.20	2 030.70	1 522.00	141.00	434.00
2005	17 655.60	3 350.90	7 116.60	1 507.60	2 373.30	2 010.60	176.00	467.00
2006	21 260.40	3 805.00	8 638.20	1 708.10	2 842.10	2 619.60	252.00	683.00
2007	25 073.00	4 436.90	10 171.00	1 910.00	3 353.10	3 225.80	448.00	769.00
2008	29 662.00	5 108.00	12 311.00	2 216.00	3 858.00	3 438.00	762.00	874.00
2009	31 964.00	5 753.00	13 222.00	2 509.00	3 748.00	3 725.00	828.00	748.00

数据来源：中国海洋统计年鉴（2002—2009）；2010 年中国海洋统计公报．

根据预处理公式，对原始数据进行预处理，得到取值在 0.15，0.85 的数据（表 2）。

表 2　2001—2009 年海洋生产总值及各海洋产业增加值的预处理后数据

年份	海洋生产总值	海洋科研教育管理服务业	海洋相关产业	海洋渔业	海洋交通运输业	滨海旅游业	海洋船舶工业	海洋油气业
2001	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150
2002	0.258	0.227	0.252	0.262	0.254	0.367	0.161	0.218
2003	0.296	0.322	0.297	0.305	0.369	0.169	0.209	0.359
2004	0.429	0.458	0.429	0.397	0.478	0.367	0.272	0.509
2005	0.547	0.560	0.539	0.538	0.587	0.540	0.328	0.551
2006	0.657	0.641	0.652	0.631	0.700	0.691	0.436	0.749
2007	0.745	0.730	0.737	0.706	0.788	0.792	0.641	0.801
2008	0.821	0.800	0.823	0.791	0.850	0.819	0.825	0.850
2009	0.850	0.850	0.850	0.850	0.838	0.850	0.850	0.789

以 2001—2008 年的 8 组数据作为训练样本，对包含 7 个神经元的 BP 网络进行训练，并进行仿真测试。

图 3 表明网络经过 316 次训练，已经达到了预定的精度。

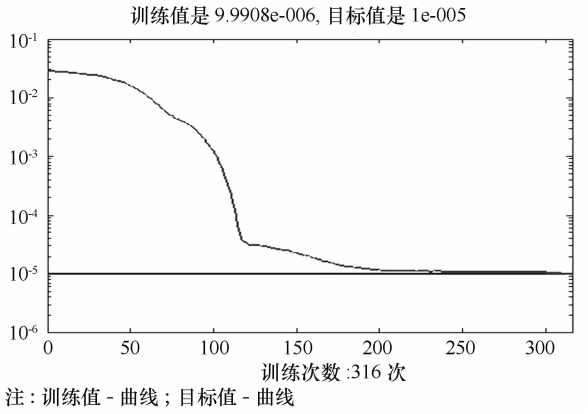

图 3　训练误差变化

由表 3 可见网络拟合相对误差均在 1%以内，效果比较好。然后利用趋势外推法预测 2010 年的海洋科研教育管理服务业、海洋相关产业、海洋渔业、海洋交通运输业、滨海旅游业、海洋船舶工业以及海洋油气业增加值，预测结果见表 4。

表 3　2001—2009 年海洋生产总值仿真结果及误差

年份	输入数据	仿真输出	误差/%
2001	0.150 00	0.149 95	−0.02
2002	0.258 00	0.349 76	0.05
2003	0.296 00	0.408 65	0.15
2004	0.429 00	0.573 30	−0.16
2005	0.547 00	0.685 41	0.47
2006	0.657 00	0.753 09	−0.65
2007	0.745 00	0.810 07	0.58
2008	0.821 00	0.842 20	0.40
2009	0.850 00	0.845 42	−0.54

表 4　2010 年主要海洋产业增加值的预测值　　亿元

海洋科研教育管理服务业	海洋相关产业	海洋渔业	海洋交通运输业	滨海旅游业	海洋船舶工业	海洋油气业
6 619.46	15 479.67	2 828.50	4 279.53	4 436.50	1 137.33	875.94

利用上述归一化公式，在 2009 年之前数据基础上，对 2010 年预测数据进行归一化，即在 2001—2009 年的各个海洋产业增加值中取各产业的最大值和最小值，以保持对训练网络输入数据的一致性。2001—2010 年原始数据所产生的新归一化矩阵见表 5。

表 5 2010 年归一化数据（基于 2001—2009 年归一化后数据）

年份	海洋生产总值	海洋科研教育管理服务业	海洋相关产业	海洋渔业	海洋交通运输业	滨海旅游业	海洋船舶工业	海洋油气业
2001	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150
2002	0.258	0.227	0.252	0.262	0.254	0.367	0.161	0.218
2003	0.296	0.322	0.297	0.305	0.369	0.169	0.209	0.359
2004	0.429	0.458	0.429	0.397	0.478	0.367	0.272	0.509
2005	0.547	0.560	0.539	0.538	0.587	0.540	0.328	0.551
2006	0.657	0.641	0.652	0.631	0.700	0.691	0.436	0.749
2007	0.745	0.730	0.737	0.706	0.788	0.792	0.641	0.801
2008	0.821	0.800	0.823	0.791	0.850	0.819	0.825	0.850
2009	0.850	0.850	0.850	0.850	0.838	0.850	0.850	0.789
2010		0.898	0.901	0.895	0.888	0.906	0.926	0.851

归一化后，代入训练好的网络，进行仿真预测，得到 2010 年海洋生产总值的预测值为 0.864 6。利用还原公式对 2010 年的预测值还原，得到的还原值为 34 781 亿元，即 2010 年海洋生产总值的预测值为 34 781 亿元。

3 模型讨论

3.1 模型优势

（1）神经网络模型以独特的数学方法把变量关系不清晰的海洋经济系统用神经网络模型来处理，弱化了随机因素的干扰，从杂乱无章的现象中揭示了事物的发展规律。

（2）神经网络模型可以在较少的原始数据基础上，得到较高的预测精度。对于统计数据不易得到的海洋经济产业来说，这是一个较好的方法。

（3）神经网络模型具有良好的自适应、自学习、极强的非线性逼近，大规模并行处理和容错能力等特点。对于处理海洋经济系统来说，它更加贴合实际。

3.2 模型的劣势

（1）神经网络模型仍是以历史数据为依据的，但是由于海洋经济数据统计口径的改变（如，2006 年统计口径进行了修正），同时新增了海洋相关产业和海洋科研教育管理服务业，这些改变使得预测数据与实际数据相差较大，影响模型的精度。

（2）神经网络模型是个根据训练样本不断学习的模型，基于较少的训练样本所训练的网络只能用于短期预测。若需长期预测，必须用更多的数据对网络进行训练。文章的模型也只适于短期预测。

（3）BP 算法易陷入局部极小，算法有待进一步改进。

参考文献

[1] 葛哲学，孙志强．神经网络理论与 MATLABR 2007 实现[M]．北京：电子工业出版社，2007：37－39.

[2] 黄雪燕．人工神经网络在 GDP 预测中的应用[D]．长春：吉林大学，2007.

[3] 《中国海洋经济发展趋势与展望》课题组．中国海洋经济预测研究[J]．统计与决策，2005(12)：43－45.

[4] 国家海洋局．中国海洋经济统计公报[DB/OL]．(2010－02－23)[2011－06－03]．http://www.soa.gov.cn/hyjww/hygb/hyjjtjgb/2009/02/1225332549346756.htm.