

深圳港基于 BP 神经网络的集装箱预测模型

徐 杏¹,史喜军²

(1. 河海大学交通与海洋工程学院,江苏 南京 210098;2. 河海大学国际工商学院,江苏 南京 210098)

摘要 运用 BP 人工神经网络建立了深圳港集装箱吞吐量的预测模型,得出深圳港 2001~2005 年的集装箱吞吐量. 研究表明,该预测方法具有很强的学习与泛化能力,预测结果对深圳港的发展有较强的借鉴作用,可以提高深圳港处理运量的能力.

关键词 BP 神经网络模型;深圳港;集装箱吞吐量;预测

中图分类号 U169;TP183 **文献标识码** A **文章编号** 1000-1980(2002)04-0041-04

近几年来,深圳港集装箱运输迅速发展,从 1994 年的 17.81 万 TEU 增加到 2000 年的 399.3 万 TEU,吞吐量一直以 60% 以上的水平增长,增长率居全球 20 大集装箱港口的首位. 深圳港的集装箱运输主要集中在 3 个港区,即西部的赤湾港区(主要是凯丰集装箱码头)、蛇口港区(主要是 SCT)和东部的盐田港区. 历年来深圳港集装箱吞吐量的数据见表 1.

表 1 深圳港集装箱吞吐量历史数据

Table 1 Historical data of container throughput of Shenzhen Port

万 TEU

年度	1992	1993	1994	1995	1996	1997	1998	1999	2000
集装箱吞吐量	10.88	12.88	17.81	28.36	58.90	114.73	195.18	298.6	399.3

随着深圳港集装箱运输的发展,港口码头设计的年处理集装箱作业的能力已不能满足其发展的需要. 2000 年深圳港的集装箱年设计处理能力为 320 万 TEU,而 2000 年完成的集装箱吞吐量为 399.3 万 TEU,处理能力缺口已达到 70 万 TEU 左右. 由于港口集装箱吞吐量的飞速发展,必然要求深圳港集装箱泊位的规划跟上实际情况的发展,否则就会导致港口设施的增加跟不上港口吞吐量的增长,影响港口的发展.

一般而言,影响港口集装箱吞吐量的因素很多,其中包括港口自然条件、航运市场状况、国家政策走向以及世界经济形势等等^[1]. 由于 BP 神经网络能用于多因素非线性系统的分析,为此本文拟采用人工神经网络中的 BP 网络来预测深圳港的集装箱吞吐量.

1 BP 网络的理论基础

人工神经网络(Artificial Neural Network,简记为 ANN)是一种包含许多简单的非线性计算单元或连接点的非线性动力系统. ANN 在具有大规模并行模拟处理、非线性动力学和网络全局作用等特点的同时,还具有很强的自适应、自学习及容错能力,因此在模式识别、智能控制、图形处理、预测和非线性优化、生物医学工程等领域取得了成功的应用,已成为一种强大的非线性信息处理工具. BP 网络是其中应用得最广泛的一种.

BP 网络属于多层状型的 ANN,由输入层、隐含层和输出层 3 层神经元组成,各层神经元的作用都是不同的. BP 网络的学习过程由正向信号传播和反向误差传播两阶段组成,即输入信息从输入层经隐含层(1 层或多层)传向输出层,如果在输出层得到的实际输出与所期望得到的输出不一致,则转入反向传播,将误差信号(实际输出与期望输出之差)沿原来通路返回,通过学习来修改各层神经元之间的连接权值,从而最后使误差达到最小.

BP 网络可看成是一个从输入到输出的高度非线性映射,根据 Kolmogorov 定理,给定任一连续函数 $f:U^m \rightarrow R^m$, $f(X) = Y$,这里 U 是闭单位区间 $[0,1]$, f 可以精确地用一个三层前向网络实现. 网络模型中输入层有 N 个神经元,输出层有 M 个神经元,隐含层有 L 个神经元,这样的 BP 网络就成为一个高度非线性映射模

型,可以在任意希望的精度上实现任意的连续函数.

2 集装箱吞吐量预测的 BP 网络建模

2.1 影响因素分析

一般而言,影响港口集装箱吞吐量的因素可以分为可量化因素和不可量化因素.其中港口所在地区的外贸进出口总额对于港口的集装箱吞吐量起决定性作用,因此将其作为可量化因素.而其他的影响因素如政治因素、法律因素、环境因素(指港口所处地理位置及是否是深水港等因素)、科技因素、形势因素(包括世界经济形势和中国经济形势以及航运市场状况等因素)、管理因素、价格因素和服务因素则作为不可量化的因素,这些因素就作为神经网络的输入部分.网络的输出部分则由相应年份的集装箱吞吐量组成.这样输入、输出之间就形成一种对应关系,构成了 9 组训练样本,每组样本中有 9 个值为输入量,即该预测网络输入层有 9 个神经元;1 个值为输出量,即输出层神经元为 1 个(见表 2).

2.2 初始化网络

2.2.1 可量化数据的预处理

为了充分发挥 BP 神经网络的预测功能,提高其预测精度,有必要对输入数据进行预处理,以提高神经网络的泛化能力(即对未学习数据的正确应答能力),一般可将各输入量归一化到 $[0,1]$ 区间,但 Sigmoid 函数在 $[0,0.1]$ 和 $[0.9,1]$ 区间内变化极为缓慢,不利于特征提取,可将各输入量归一化至 $[0.1,0.9]$ 之间,为此采用线性化处理办法,对于值越大越好的指标可采取以下方式进行处理.

令 $\max x = X_{\max}$,则 $X_{\max}$ 为第 i 个样本第 j 个指标中的最好值; $\min x = X_{\min}$,则 $X_{\min}$ 为第 i 个样本第 j 个指标中的最差值.则指标 j 的所有 x_i 转化为无量纲的 x'_i 为

$$x'_i = 0.1 + (x_i - X_{\min}) / (X_{\max} - X_{\min}) \times 0.8 \qquad i = 1,2,\cdots,n$$

在本次应用中, j 取 1; x_i 为第 i 年深圳市外贸进出口总额(i 取 9),即前面表示的 x_{i0} , x 则表示深圳市 9 年的进出口总额序列,对于输出即 9 年的集装箱吞吐量也必须进行归一化处理.在预测中产生的输出数据是归一化后的指标,还必须按相反规则进行变换以求出具体的数据.

2.2.2 不可量化因素的处理

对于不可量化因素的预处理,可采用编码量化的方法.从实际角度来看,可将各影响因素从 1992 年到 2005 年的形势分为 3 级,规定其分别对应于等级编码 0.1,0.5,0.9,增量编码 0.01.一般而言,影响因素在无特殊的情况下不会发生实质性的变化,因此假定若在一段时间内,某因素呈稳步发展态势,则规定其发展的增量为 0.01,若有较大幅度的变化则可规定其变化幅度为 0.02 等.若某一因素向不利的方向发展,则可视其影响程度在原有的基础上减少 0.01 或 0.02 等不同的幅度.若某因素发生了突变,此时就假定其影响程度的级别变了,编码的取值将在原有等级编码的基础上上升一级.对此可以从影响因素发展的历史和预测中找到其具体的变化情况.针对实际,影响深圳港集装箱吞吐量的不可量化因素的发展状况编码见表 3.

表 3 集装箱吞吐量的不可量化影响因素的编码

Table 3 Coding of unquantizable factors influencing container throughput														
影响因素	年 份													
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
政治因素	0.10	0.50	0.51	0.52	0.53	0.54	0.55	0.56	0.57	0.90	0.91	0.92	0.93	0.94
法律因素	0.10	0.11	0.50	0.51	0.52	0.53	0.54	0.55	0.56	0.90	0.91	0.92	0.93	0.94
环境因素	0.10	0.11	0.50	0.51	0.52	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98
科技因素	0.10	0.11	0.12	0.13	0.14	0.50	0.51	0.52	0.53	0.90	0.91	0.92	0.93	0.94
形势因素	0.10	0.11	0.50	0.51	0.52	0.53	0.52	0.55	0.90	0.91	0.92	0.93	0.94	0.95
管理因素	0.10	0.11	0.12	0.13	0.14	0.50	0.51	0.52	0.90	0.91	0.92	0.93	0.94	0.95
价格因素	0.10	0.11	0.12	0.13	0.50	0.51	0.50	0.52	0.53	0.54	0.90	0.91	0.92	0.93
服务因素	0.10	0.11	0.12	0.13	0.50	0.52	0.55	0.90	0.91	0.92	0.93	0.94	0.95	0.96

3 BP 神经网络预测系统的实施

3.1 影响因素的预期

通过预测各影响因素的发展状况,可以对深圳港的集装箱吞吐量发展状况进行预测.

3.1.1 不可量化因素的预期

表 3 给出了不可量化的影响因素的发展状况预测. 2001 ~ 2005 年的编码是根据历史的发展变化规律及变化趋势定出来的,这些数据表示了变化的规律,网络只对编码的变化规律进行识别.

3.1.2 可量化因素的预期

本问题中只有一个可量化因素,即深圳市的外贸进出口总额. 1992 ~ 2000 年的数据见表 4^[2~7].

表 4 深圳市 1992 ~ 2000 年外贸进出口总额

Table 4 Amounts of import and export trade of Shenzhen for 1992 ~ 2000									亿美元
年份	1992	1993	1994	1995	1996	1997	1998	1999	2000
外贸进出口总额	79.77	148.80	349.83	387.70	390.53	450.08	452.7	504.3	639.4

从表 4 中可以看出深圳市外贸进出口总额随着时间的发展以某种方式呈现递增的规律. 在此,采用时间序列线性回归模型来预测其发展状况. 经过计算可得预测模型的表达式为

$$\hat{y}_i = 60.269t_i - 119917.16 \quad (R^2 = 0.9037)$$

其中 t_i 表示时间. 代入时间即可得 2001 ~ 2005 年的外贸进出口总额. 深圳市 2001 ~ 2005 年的外贸进出口总额分别为 681.1,741.38,801.65,861.92,922.19 亿美元. 将 1992 ~ 2005 年的外贸进出口总额进行归一化处理,最后作为 BP 神经网络的输入数据,在此不一一列出.

3.2 深圳港集装箱吞吐量预测

3.2.1 训练及检验阶段

由前面介绍所知,本问题共有 9 组样本,经过对输入、输出数据的预处理,最后得到标准训练样本为

$$\begin{aligned} X^{(1)'} &= [0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1]^T & Y_1' &= 0.1 \\ X^{(2)'} &= [0.1656,0.5,0.11,0.11,0.11,0.11,0.11,0.11,0.11]^T & Y_2' &= 0.1034 \\ \\ X^{(9)'} &= [0.630,0.57,0.56,0.93,0.53,0.9,0.9,0.53,0.91]^T & Y_9' &= 0.9 \end{aligned}$$

将前面 7 组样本作为训练样本,后两组样本作为预测效果检验样本. 应用 BP 网络解决本问题时,输入层有 9 个神经元,输出神经元为 1 个. 隐含层神经元为 10 个,最大容许误差为 0.001,学习步长为 0.8,最大学习次数为 2500 次. 隐含层中采用对数饱和性(Logsit)映射函数,即 $y = \frac{1}{1 + e^{-x}}$; 计算输出层神经元的输出时采用线性(Purelin)变换函数,即 $y = kx$; 误差函数为 $E = \frac{1}{2} (Z^{(i)} - Y^{(i)'})^2$,其中 $Z^{(i)}$ 为输出层神经元的输出, $Y^{(i)'}$ 为 $Y^{(i)}$ 归一化处理后的数据.

经过对前 7 组样本的训练,可对网络的输入层至隐含层、隐含层与输出层之间的连接权值、隐含层神经元及输出层神经元的阈值进行调整,最后得出比较合理的数值,在此不一一列出. 同时笔者利用经过训练的网络以及余下的两组样本来进行网络预测准确与否的检验. 将第 7 组样本中的输入量 $X^{(7)'}$ 输入到网络中,得到归一化后的检验值,将其还原得 $\hat{Y}^{(7)} = 297.6$. 该值与 1999 年深圳港集装箱吞吐量的实际值 $Y^{(7)} = y_7 = 298.6$ 接近. 同理,利用 2000 年的数据样本来检验也可得出比较准确的结论. 由此表明该网络可以用来预测.

3.2.2 预测

将 2001 ~ 2005 年所对应的各影响因素指标值列成如下形式:

$$\begin{aligned} X^{(10)'} &= [0.6711,0.9,0.9,0.94,0.9,0.91,0.91,0.54,0.92]^T \\ X^{(11)'} &= [0.7283,0.91,0.91,0.95,0.91,0.92,0.92,0.9,0.93]^T \\ \\ X^{(14)'} &= [0.9,0.94,0.94,0.98,0.94,0.95,0.95,0.53,0.96]^T \end{aligned}$$

将这几组数据作为网络的输入,用上述介绍的方法求解输出,累积样本,并重新训练更新网络权值,即可实现 2001~2005 年的港口集装箱吞吐量预测. 以上过程通过 MATLAB 神经网络工具箱编程实现^[8],预测结果即深圳港 2001~2005 年的集装箱吞吐量分别为 500,575,650,725 和 800 万 TEU.

4 结 束 语

本文通过 ANN 对深圳港集装箱吞吐量的发展状况进行了预测,由于 ANN 具有很强的非线性映射能力、自学习能力、容错能力等特性,可以较好地解决非线性系统的预测问题^[9],本文最后得出的结论就较好地反映了 ANN 的特性. 由此可以根据该预测结果来发展深圳港的各项业务. 通过预测,可以看出深圳港将来几年的发展将趋向比较稳定,到 2005 年,该港的集装箱吞吐量将突破 800 万 TEU. 据此可以得出结论,即在 2005 年,深圳港的集装箱设计吞吐能力必须要达到 800 万 TEU 以上,各项基础设施也必须要达到能处理 800 万 TEU 的设计能力,这样才不会在将来由于设计吞吐能力的不足导致不能完成日益增长的集装箱业务.

参考文献:

- [1] 郑金海,金菊良. 应用加速基因算法预测港口集装箱吞吐量[J]. 中国港湾建设,2000(3): 33~34.
- [2] 中国经济年鉴社编辑委员会. 中国经济年鉴[M]. 北京: 中国经济年鉴社,1995. 465.
- [3] 中国经济年鉴社编辑委员会. 中国经济年鉴[M]. 北京: 中国经济年鉴社,1996. 510.
- [4] 中国经济年鉴社编辑委员会. 中国经济年鉴[M]. 北京: 中国经济年鉴社,1997. 540.
- [5] 中国经济年鉴社编辑委员会. 中国经济年鉴[M]. 北京: 中国经济年鉴社,1998. 557.
- [6] 中国经济年鉴社编辑委员会. 中国经济年鉴[M]. 北京: 中国经济年鉴社,1999. 565.
- [7] 中国经济年鉴社编辑委员会. 中国经济年鉴[M]. 北京: 中国经济年鉴社,2000. 573.
- [8] 楼顺天,施阳. 基于 MATLAB 的系统分析与设计——神经网络[M]. 第 2 版. 西安: 西安电子科技大学出版社,1999. 114~118.
- [9] 韩玉兵. 基于神经网络的 Smith 补偿 PID 控制设计[J]. 河海大学学报(自然科学版),2000,28(5): 38~91.

BP neural network-based container throughput predicting model for Shenzhen Port

XU Xing¹, SHI Xi-jun²

- (1. College of Traffic and Ocean Engineering, Hohai Univ., Nanjing 210098, China;
2. College of International Industry and Commerce, Hohai Univ., Nanjing 210098, China)

Abstract: A model for predicting container throughput is developed for the Shenzhen Port based on BP neural network, and the throughputs of the Shenzhen Port from 2001 to 2005 are calculated. It is indicated that BP neural network has a strong ability to study and generalize, and that the predicted result can be used as a reference for the development of the Shenzhen Port, and may help improve throughput calculation for the Shenzhen Port.

Key words: BP neural network model; Shenzhen Port; container throughput; prediction