

文章编号:1002-980X(2006)12-0082-03

基于最优组合预测模型的港口集装箱吞吐量预测

童明荣, 薛恒新, 林琳

(南京理工大学 经济管理学院, 南京 210094)

摘要:根据港口集装箱吞吐量非线性增长等特点,建立了三次指数平滑预测模型、灰色系统预测模型及BP神经网络预测模型等单项预测模型。鉴于单项预测模型的局限性,提出了以测试数据的预测误差绝对值加权和最小为最优化准则的最优组合预测模型,采用线性规划的方法确定最优组合的权系数。最后,给出一个实例进行应用和分析。

关键词:组合预测;指数平滑预测;灰色预测;BP神经网络;集装箱吞吐量

中图分类号:U652.1+4 **文献标志码:**A

港口集装箱吞吐量预测对于港口未来的定位和发展战略、集装箱码头的规划和建设等具有重要意义。集装箱吞吐量受到多种因素的影响和制约,任何单一预测模型都难以对这种非线性系统作出准确的预测。各种单项预测法在数据处理及推理准则方面均有其独到之处,其预测结果都具有一定的参考价值。如果简单地将误差较大的一些方法舍弃掉,将会丢失一些有用的信息,造成预测的片面化。目前流行的做法是,将不同的预测方法以某种适当的方式进行组合,综合利用各种预测方法所提供的信息,于是就形成了组合预测方法。组合预测方法是1969年Bates^[1]等提出来的,理论研究和实际应用表明,组合预测模型能提高预测的精度、增强预测的稳定性。目前最成熟的、最常用的是最优组合预测方法。

一、最优组合预测模型

设某一预测问题在某一时段的实际值为 y_t ($t=1, 2, \dots, n$), 现有 m 种单项预测模型得到的预测值分别为 f_{it} ($i=1, 2, \dots, m; t=1, 2, \dots, n$), 加权向量为 $W=(w_1, w_2, \dots, w_m)^T$, 则可得组合预测公式: $\hat{y}_t = \sum_{j=1}^m w_j f_{jt}$ 。组合预测的关键是如何恰当地确定各个单项预测模型的权重, 本文采用“误差绝对值之加权和最小”作为组合预测的最优准则。由于近期的数据对预测结果所产生的影响要大于远期的数据的影响, 所以在对近期和远期预测误差给予不

同的折扣系数, 以提高预测精度。

基于误差绝对值加权和最小的组合预测模型为:

$$\min S = \sum a_t |w_1 e_{1t} + w_2 e_{2t} + \dots + w_m e_{mt}|$$
$$s. \begin{cases} w_1 + w_2 + \dots + w_m = 1 \\ w_i \geq 0 (i=1, 2, \dots, m) \end{cases}, a_t \text{ 为折扣系数, 本文取作 } a_t = a^{(n-k)}$$

由于目标函数 S 的表达式中带有绝对值, 直接求解不方便, 可令:

$$u_t = \frac{1}{2} |w_1 e_{1t} + w_2 e_{2t} + \dots + w_m e_{mt}| + \frac{1}{2} (w_1 e_{1t} + w_2 e_{2t} + \dots + w_m e_{mt}) \geq 0$$
$$v_t = \frac{1}{2} |w_1 e_{1t} + w_2 e_{2t} + \dots + w_m e_{mt}| - \frac{1}{2} (w_1 e_{1t} + w_2 e_{2t} + \dots + w_m e_{mt}) \geq 0, (t=1, 2, \dots, n)$$

则可将模型化为一个线性规划问题:

$$\min S = \sum a_t (u_t + v_t)$$
$$s. \begin{cases} w_1 e_{1t} + w_2 e_{2t} + \dots + w_m e_{mt} = u_t - v_t \\ w_1 + w_2 + \dots + w_m = 1 \\ u_t \geq 0, v_t \geq 0, w_i \geq 0 \end{cases} (i=1, 2, \dots, m; t=1, 2, \dots, n)$$

该模型可以用单纯形法来求解。

二、实例分析

运用上述方法和表1中的历史数据对年宁波港集装箱吞吐量进行预测。根据港口集装箱吞吐量历

收稿日期:2006-09-12

作者简介:童明荣(1980—),男,浙江江山人,南京理工大学博士生,研究方向:物流管理。

史数据不多、非线性快速增长的特点,本文选取三次指数平滑预测模型、BP 神经网络预测模型和灰色系统预测模型作为单项预测模型。为了确定各单项预

测模型中的参数以及各单项预测模型在组合预测模型中的权重,将 1990—2003 年的数据作为试验数据,2004—2005 年的数据作为测试数据。

表 1 宁波港集装箱吞吐量(单位:万 TEU)历史统计数据

年份	1990	1991	1992	1993	1994	1995	1996	1997
吞吐量	2.2	3.6	5.3	7.9	12.5	16.0	20.2	25.7
年份	1998	1999	2000	2001	2002	2003	2004	2005
吞吐量	35.3	60.1	90.2	121.3	185.9	277.2	400.5	520.8

(一)三次指数平滑预测

根据国际上使用三次指数平滑法的经验, α 值一般取之间。本文使 α 以为步长在 $[0.05, 0.5]$ 的区间递增,分别对试验数据作三次指数平滑建模,并计算测试数据的预测值与实际值的均方误差,找到最佳 α 。

经计算当 α 为 0.5 时测试数据的预测误差均方差(MSE)最小,预测模型为: $f(t) = 272.7 + 83.4 \times t + 6.4 \times t^2$,该模型对 2004~2005 年的预测结果及其误差见表 2。在相同参数下,对全部历史数据建模得到预测模型为, $f(t) = 521.0 + 134.4 \times t + 8.7 \times t^2$ 。

表 2 测试数据的预测结果 单位:万 TEU

年份	实际值	指数平滑模型			BP 神经网络预测模型			灰色预测模型		
		预测值	误差	MSE	预测值	误差	MSE	预测值	误差	MSE
2004	400.5	389.7	- 10.8	69.8	411.1	10.6	64.6	373.6	- 26.9	534.8
2005	520.8	525.6	4.8		516.7	- 4.1		539.4	18.6	

表 3 2006~2010 年宁波港集装箱吞吐量的预测结果 单位:万 TEU

年份	指数平滑预测值	神经网络预测值	灰色预测值	最优组合预测值
2006	664	620	798	642
2007	825	735	1146	780
2008	1002	837	1468	919
2009	1198	943	2368	1069
2010	1411	1045	3403	1226

(二)BP 神经网络预测模型

BP 神经网络是一种单向传播的多层前向网络,由输入层、隐层和输出层构成。针对 BP 网络目前存在的几个主要不足,分别作了改进:

1. 网络结构人为性大,网络输入节点数和隐层单元数的选择尚无理论上的指导。Robert Hecht Nielson^[2]等人的研究表明,单隐层的 BP 网络能够以任意精度逼近任一非线性函数,但如果隐层神经元数目过少,网络不能收敛,过多则会导致训练效率低和过拟合,因此本文设计一个隐层神经元数目可变的单隐层 BP 网络,通过对比收敛效果和预测误差,确定最佳的隐层神经元数目。输入维数用以下方法确定:

首先,将给定时间序列用 Box-Cox 变换进行预处理,转变为平稳离散时间序列: $Y(\lambda) = \begin{cases} (Y^\lambda - 1) & \lambda \neq 0 \\ \ln \lambda & \lambda = 0 \end{cases}$,要寻找合适的 λ 值使得变化后的变量具有最小的标准差,即尽量接近正态分布。

然后,将该序列分成 k 组,每组 $m + 1$ 个数,前 m 个作为网络输入节点的输入,后一个作为输出节点的期望值。

即: $X: \{x(i), x(i + 1), \cdots, x(i + m)\} \rightarrow Y: \{x(i + m + 1)\}, (i = 1, 2, \cdots, k)$ 。确定 m 的经验公式为: $m = \min\{i | E_G - E_O\}, (i = 1, 2, \cdots, \max)$, $\max$ 是人为指定的最大输入层节点数上限。其中:

$$E_G = \sum_{r=1}^{n-1} \sum_{s=r+1}^n d_{rs}, E_O = \sum_{r=1}^{n-1} \sum_{s=r+1}^n |Y_r - Y_s|,$$
$$d_{rs} = \frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^m (x_{ri} - r_{si})(x_{ij} - x_{sj}) \delta_{ij} \right]^{1/2},$$
$$\delta_{ij} = \frac{\sum_{p=1}^n (x_{pi} - \bar{x}_{pi})(x_{pj} - \bar{x}_{pj})}{\sqrt{\sum_{p=1}^n (x_{pi} - \bar{x}_{pi})^2} \sqrt{\sum_{p=1}^n (x_{pj} - \bar{x}_{pj})^2}}$$

d_{rs} 表示输入向量 $X(r)$ 和 $X(s)$ 之间的斜交叉距离, $\bar{x}_{pi}$ 和 $\bar{x}_{pj}$ 分别表示所有输入向量 X 的第 i 个和第 j 个分量的均值。

本例中,对于试验数据,用 Minitab 软件计算得到 $\lambda=0$,将平稳后的数据序列应用于经验公式,从表 4 可知输入维数 m 取 2。

本文利用神经网络工具箱实现 BP 神经网络,隐层神经元均采用双曲正切 s 型传递函数,输出层采用线性传递函数,训练函数用“ trainlm ”,该函数以学习率可变的动量 BP 算法修正神经网络的权值和阈值,初始学习速率取为 0.1,动量因子取 0.95,隐层神经元数目分别取进行试算,比较收敛速度和误差。当隐层神经元数取 6 时,网络收敛效果较好,由于 BP 神经网络每次训练结果都不相同,故进行多次训练,取预测误差最小的一次,对年的预测结果见表 2。然后,保持同样的网络结构和参数对全部历史数据建模,进行多次训练,取最符合增长趋势的一次,训练 1836 步后,网络误差精度能达到 0.001。

表 4 确定神经网络输入维数 m 的经验公式计算值

m	1	2	3	4	5
$E_G - E_O$	2.12	0.42	0.93	2.39	4.6

2. 传统的 BP 算法学习速率是固定的。要选择适当的学习速率比较困难,小的学习速率导致较长的训练时间,而大的学习速率可能导致网络的不稳定。本文在网络训练中采用自动调整学习速率的方法,即自适应学习速率法。

3. 传统的 BP 网络是一个非线性优化问题,不可避免地存在局部极小问题。本文采用附加动量法可以降低网络对于误差曲面局部细节的敏感性,有效地抑制网络陷入局部极小。

(三) 灰色预测

灰色预测^[3]是我国著名学者邓聚龙教授于 1982 年创立的,常用的是一阶预测模型 $GM(1,1)$ 。传统的初值选用原始数列的第一个数据值 $X^{(0)}(1)$,这是没有理论依据的,在运用中会降低建模精度和预测精度,本文设初始值为 $\beta X^{(0)}(1)$,对试验数据建模,用试算的方法寻找最佳 β 值,使得测试数据的预测均方误差最小。构造如下目标函数:

$$J = \sum_{k=1}^n [X^{(1)} - X^{(1)}(k)]^2$$

$$= \sum_{k=1}^n \left[\beta X^{(0)}(1) - \frac{a}{a} e^{-a(k-1)} + \frac{a}{a} - X^{(1)}(k) \right]^2$$

为了求出使目标函数 J 最小的 β 值,可用梯度法,即令 $\frac{dJ}{d\beta} = 0$ 求解得到。

先对试验数据作 2 次方根变换,使之满足灰色建模要求。让 β 以步长在 $[0.0001, 3]$ 区间内递增,

分别对测试数据作灰色预测,当 $\beta = 1.1624$ 时,预测均方误差最小,此时的预测模型为: $X^{(0)}(k+1) = 8.811e^{0.1836k} - 8.811e^{-0.1836+0.1836k}$, ($k=1, 2, \dots$)。

对全部历史数据作 2 次方根变换后,用梯度法计算得到 $\beta = 0.9934$,此时,预测模型为: $X^{(0)}(k+1) = 9.3614e^{0.1813k} - 9.3614e^{-0.1813+0.1813k}$, ($k=1, 2, \dots$)。后验指标方差比 $C = 0.065$,第二后验指标小误差概率 $P = 1$,精度为一级。

(四) 最优组合预测

将表 2 中数据代入组合预测模型,取折扣系数 $a_t = 0.9^{(n-k)}$,即可求得各种预测方法的权重, $\omega_1 = 0.495$, $\omega_2 = 0.505$, $\omega_3 = 0$,因此,组合预测模型为: $\hat{y}_t = 0.495f_{1t} + 0.505f_{2t}$,该模型的预测结果见表 3。

(五) 预测结果分析

从对试验数据年的预测结果看,灰色模型预测的标准差最大,指数平滑法次之,神经网络的标准差最小。这是因为, BP 神经网络的优点是短期预测精度很高,它具有局部逼近网络的优点,可以近似任何非线性函数,该模型用于近期数据的模拟及预测,精度相当高,优于其它模型。灰色模型的优点是对样本数据量要求较少,但该模型的近期预测效果较差。指数平滑法的特点是对样本条件要求较高,它不仅需要大量的统计数据,并且要求这些数据符合一定的统计规律。任何一种模型的预测结果都不可能是完全精确的,这是因为像集装箱吞吐量这样一个复杂的经济系统,随着时间的推移,与历史数据相联系的内部及外部环境条件会产生各种变化,组合预测模型能最大限度地综合各种单项预测模型的优点,克服这些不确定性,因此更加适合于实际预测。

三、结束语

本文基于测试数据的预测误差绝对值加权和最小建立了组合预测模型,在对宁波港集装箱吞吐量预测中取得了较好的效果。从理论上讲,该组合预测模型可以预测未来中长期的集装箱吞吐量,但随着时间的推移,未来的诸多不可控因素将不断地进入系统;再者,影响港口集装箱吞吐量的因素很多,这些因素有的是直接影响因素,有的是间接影响因素,对港口集装箱吞吐量的影响也有主有次,可见对港口集装箱吞吐量的预测是较为复杂的过程。因此,对于港口集装箱吞吐量的长期预测模型的建立有待进一步的研究和探索。

(下转第 92 页)

石油、煤炭和森林是东三省最重要的资源产业,围绕这三大资源产业兴起了几百家大中型矿、林企业,以这些企业为基地形成了二十几个资源型城市,这些城市的人口数量、土地面积和经济总量均占很大比重。资源型城市实现经济转型是老工业基地调整改造的重点和难点。其转型策略应定位在:对资源仍较丰富的,要加强资源的综合利用和精深加工,拉长发展链条;对资源近期可能走向衰竭的,要抓紧研究接续产业发展问题;对资源已经枯竭和接近枯竭的,要加快发展接续产业。资源型城市要大力发展石油化工、煤化工、林木产品深加工、林下产业等资源精深加工和新的接续产业,扶持共伴生资源综合利用和废弃物资源化,支持油页岩、煤矸石的

综合利用。鼓励石油企业开发低产、低渗、低效油田和周边低丰度油田。加快矿区环境修复和污染治理,解决好矿山关闭破产、职工安置、沉陷区居民搬迁等紧迫问题。加大对采煤沉陷区治理的支持力度,研究建立资源开发补偿机制和衰退产业援助机制,促进资源型城市经济转型和可持续发展。

参考文献

- ① 陈永杰. 东北老工业基地基本情况调查报告[R]. 简报, 2003—10—24.
- ② 吉林省发展计划委员会. 振兴吉林老工业基地规划纲要[N]. 经济日报, 2004—03—08.
- ③ 黑龙江省人民政府. 黑龙江省老工业基地振兴总体规划[Z]. 2004—02.

Study On the Strategy of Vitalising the Northeast Old Industrial Base

ZHANG Feng-wu

(Heilongjiang University of Science and Technology, Harbin 150027, China)

Abstract : In order to regulate and reconstruct the northeast old industrial base into a new type of industrial base, it's necessary to build pleasant economic development environment for promoting the reform of state-owned enterprises and the rapid economic development of massed ownership, to optimize investment environment vigorously and broaden the opening scopes continuously, to solidify foundational position of modern agriculture to promote the optimization and update of industry composition, to develop tertiary industry energetically, to implement the strategy of developing province of science and education and strengthening province to promote the economic transformation of resource-type cities and the development of various undertaking.

Key words : northeast ; old industrial base ; vitalising strategy

(上接第 84 页)

参考文献

- ① Bates J M, Granger C W J. Combination of Forecast [J]

. Operations Research Quarterly, 1969, 20(4):451 - 468.

- ② Hecht-Nielsen R. Theory of Back Propagation Neural Network [J]. Proc of ICJNN, 1989, (1):593 - 603.

- ③ 邓聚龙. 灰预测与灰决策[M]. 武汉:华中科技大学出版社, 2002.

Forecasting Container Throughput of Port Based on Combined Forecasting Model

TONG Ming-rong, XUE Heng-xin, LIN Lin

(School of Economics & Management, Nanjing University of Science & Technology, Nanjing 210094, China)

Abstract : According to the non-linear increasing characteristics of the container throughput of port, triple exponential smoothing forecasting model, grey system forecasting model and BP artificial neural network forecasting model are established respectively. Due to the limitations of each single forecasting model, this paper put forward the combined forecasting model based on the optimum principle of the minimal sum of the forecasting error absolute values, the weighting coefficients are derived by linear programming. Finally, a practical case is given.

Key words : combined forecasting; exponential smoothing forecasting; grey forecasting; BP artificial neural network; container throughput