

非线性动态系统的递归神经网络预测研究

——对我国钢铁产量的预测分析

刘兰娟, 谢美萍

(上海财经大学 经济信息管理系, 上海 200433)

摘要:钢铁行业是我国国民经济的支柱产业之一,为国民经济的持续发展作出了积极的贡献。因此,对钢铁产量的预测研究已经成为一项非常重要的课题。文章使用基于数据挖掘和知识发现的人工神经网络法、时间序列分析法、递归神经网络技术来预测钢铁产量的方法,并将递归神经网络方法预测的结果与前面的两种方法的预测结果进行比较,比较的结果说明该方法是可行的。

关键词:钢铁产量;人工神经网络;时间序列分析法;递归神经网络技术

中图分类号:F402.4 **文献标识码:**A **文章编号:**1001-9952(2004)11-0026-08

一、引言

钢铁行业是我国国民经济的支柱产业之一,为国民经济的持续发展作出了积极的贡献。钢铁产量更是经济发展的晴雨表,众所周知,钢铁行业是典型的周期性行业,与宏观经济的发展密切相关。1998~2002年我国累计发行国债6 600亿元,带动配套资金2.54万亿元,形成投资总规模3.2万亿元。在此投资拉动背景下,我国的钢铁行业获得了快速发展,尤其是近两年来,我国宏观经济的高速增长,更是为钢铁行业带来了难得的发展机遇。各项指标表明,我国的宏观经济从2003年开始,已开始进入一轮新的经济发展周期,在积极的财政政策和稳健的货币政策支持下,GDP增长呈稳定的上升势头。并且投资主体逐渐发生转变,内生性的投资机制开始出现。不论是政府基础设施类投资,还是汽车、机械、房地产、家电、造船等新的消费性投资,都为钢铁行业的产品销售对象从建筑业向制造业转变创造了条件,从而带来更大的市场空间。同时,钢铁生产仍未摆脱高投入、高消耗、高污染、低产出这种粗放经营的状况。

收稿日期:2004-09-01

作者简介:刘兰娟(1960—),女,上海人,上海财经大学经济信息管理系副教授;

谢美萍(1974—),女,江苏如皋人,上海财经大学经济信息管理系讲师。

钢铁生产需要耗费大量的电、煤、水、矿石,需要铁路、航运运能的配合,现在一些钢厂大量进口铁矿石。由于国际市场价格大幅提高,1吨铁矿石的价格已经上涨到1000美元,相当于生铁的价格,一年涨幅增加一倍,生产成本急剧提高。今后两年我国钢产量还将持续增加,两年内新投产4000万吨钢,到2005年全国钢产量将增至2.6亿吨,由此可见,这是一个非常可怕的局面。首先是需求没有这么大,其次是浪费能源,造成新一轮的重复建设,引起市场的无序竞争。

一份关于我国钢产量需求关系的报告也显示,2005年我国钢需求量只有2.7亿吨,2010年的需求量也仅为3.3亿吨,如果减去每年进口钢的数量,国产钢的产量已经大大超过需求。目前供需紧张是暂时的,如再不加限制,盲目扩产的结果是钢卖不出去,从而造成资源的极大浪费。

钢铁产量的预测方法与手段的缺乏是引起钢铁生产供需不匹配局面的原因之一。因此,对钢铁产量的预测研究就成为一项非常重要的课题。然而,钢铁的年产量与影响它的各种因素之间的关系不能简单地用一次方程的形式来表示,即钢铁的年产量问题是一个非线性动态系统的问题。本文的主要目的就是借助于知识管理和数据挖掘的一些知识以及目前已有的针对非线性动态系统的一些建模预测研究方法,来研究我国钢铁产量的预测问题。

二、非线性动态系统的建模预测研究

(一)非线性动态系统的建模预测

现实世界中的实际过程都是非线性的,而且很多过程都是复杂的非线性动态过程,钢铁产量的预测就是这样一个非线性动态系统。尽管确定非线性动态系统的未知参数非常困难,但非线性动态系统的建模和预测的研究需求十分迫切。

建立描述非线性动态系统的模型是研究非线性问题的基础,由此才能展开有关的分析、设计、预测等研究工作。由于非线性动态系统的性质互不相同,而各个非线性动态系统本身又包含着非常丰富的内容,所以用于描述非线性动态系统的模型也是形态各异,很难像线性系统一样有统一的描述形式,因而非线性动态系统的建模也无法像线性系统的建模那样能进行统一的处理。也就是说,针对不同的非线性模型需要实施不同的建模方法。目前,描述非线性动态系统的模型主要有泛函级数模型、高阶频率响应函数和分块系统等几大类。传统的泛函级数描述方法,如Volterra级数或Wiener级数描述,虽然能充分表示一大类非线性动态系统,但是,即使应用该方法对一个简单的非线性动态系统建模,往往需要几百个参数,估计未知参数的计算量非常大,而且难以解释所得结果和要求使用的特殊输入信号,这些缺点限制了泛函级数描述方法在非线性动态系统建模中的应用,更难于实现工业过程的建模。分块系统在一定程度上解决了泛函级数描述法的不足,但是,由于它先验地假设了模型结构,所以只能表示范围很小的一类非线性动态系统。因此,对非线性动

态系统的建模预测问题成为一个很重要的课题。

通常情况下,假设非线性动态系统的多变量离散模型为

$$y(k) = f(y(k-1), \dots, y(k-n), u(k-1), \dots, u(k-n_u)) \quad (1)$$

其中, $y(k) = (y_1(k), \dots, y_q(k))^T$ 为 q 维输出, $u(k) = (u_1(k), \dots, u_p(k))^T$ 为 p 维输入, n_y 和 n_u 分别为系统输出和输入的最大时滞, $f(k) = (f_1(\cdot), \dots, f_q(\cdot))$ 为光滑的线性函数。

由非线性动态系统(1),令

$$x_j(k) = y_i(k), y_i(k-1), \dots, y_i(k-n_{y_i}), u_j(k-1), \dots, u_j(k-n_{u_j})$$

作为递归神经网络的输入,就可以构造出递归神经网络模型。

通常, k 时刻的输出预测值可以表示成模型参数 θ 、时间 k 和 $k-1$ 时刻以前输入输出数据集 $D^{(k-1)}$ 的函数关系,即

$$\hat{y}(k|\theta) = f(\theta, k, D^{(k-1)})$$

对某种给定的模型结构来说,函数 $f(\cdot)$ 是惟一确定的。

(二) 人工神经网络法

神经网络系统是由大量的、同时也是很简单的处理单元(或称神经元)广泛地互相连接而形成的复杂网络系统。其信息处理功能是由网络单元的输入输出特性(激活特性)、网络的拓扑结构(神经元的连接方式)、连接权的大小(突触联系强度)和神经元的阈值所决定的。利用上述不同特性的不同组合,可获得不同功能的人工神经网络。随着人工神经网络在各个领域的广泛应用和研究,国内已经有许多人把这一方法应用于非线性动态系统建模预测研究中。人工神经网络的数学基础是逼近论,逼近论的基本问题是用简单表示复杂。该方法的优点是只要给网络一定的学习样本,通过训练,网络可以自主地找出映射规律,从而给出期望输出,这样就可以省去数据分析和建模过程,对问题的处理带来很大方便。

尽管人工神经网络逼近非线性函数的能力已在理论和应用方面得到广泛的研究。但是,其逼近还存在着某些缺陷,如在许多情况下,神经网络的 Sigmoid 函数逼近,需有很多步迭代才能达到所要求的下降量,而且当神经元的输出接近于饱和时,该神经元的输出对权值调整将不敏感,误差曲线常常有局部极小值存在,而且这些局部极小点处的误差还比较高,要避免这些局部极小达到全局极小是一个复杂的问题。

在使用神经网络时有几点需要注意:第一,神经网络很难解释,目前还没有能对神经网络做出清晰解释的方法学。第二,神经网络会学习过度,在训练神经网络时一定要恰当地使用一些能严格衡量神经网络的方法,如测试集方法和交叉验证法等。这主要是由于神经网络太灵活、可变参数太多,如果给予足够的时间,神经网络几乎可以“记住”任何事情。第三,除非问题非常简单,训练一个神经网络可能需要相当长的时间才能完成。当然,一旦神经网络建立好了,在它做预测时运行速度还是很快的。第四,建立神经网络需要做的数据准备工作量很大。一个很有

误导性的神话就是不管用什么数据,神经网络都能很好地工作并做出准确的预测。其实,要想得到准确度高的模型必须认真地进行数据的清洗、整理、转换、选择等工作,对任何数据挖掘技术都是这样,神经网络尤其注重这一点。

(三)时间序列分析法

用时间序列分析方法预测非线性动态系统,只需建立时间序列模型来预测系统的未来值,就可以避免人工神经网络法陷入局部极小的麻烦。

所谓时间序列,就是各种各样的社会、经济、自然现象的数量指标依时间次序排列起来的统计数据,每一个数据都是在相同的时间间隙里产生的。人们熟悉的这些序列有产量、收入和GDP数据等。时间序列分析是用来分析这些数据的一种十分有效的方法,其目的在于:预测序列的未来发展情况、分析序列的基本趋势、季节和随机项的组成;分析特定的数据集合、拟合理论模型,尤其是建立数学模型,进而进行模型结构分析和实证研究。

一般情况下,非线性动态系统的时间序列模型,即ARMAX(Auto Regressive Moving Average with Exogenous Inputs)模型,可表示为

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t)$$

其中, $y(t)$ 表示数据序列 t 时刻的输出值, $u(t)$ 表示 t 时刻的输入值, $e(t)$ 是均值为零的正态无噪声,即

$$E(e(t)) = 0, \quad E(e(t)e(s)) = \sigma^2 \delta_{ts}$$

其中, δ_{ts} 是Kronecker δ 函数, q^{-1} 是单位滞后算子, $A(\cdot)$ 、 $B(\cdot)$ 和 $C(\cdot)$ 是算子多项式。

需要说明的是,从建立的时间序列模型出发计算的预测值,常常与实际值不符合,这主要有两方面的原因:一是拟合的模型有错误需要修正;二是事物原有的趋势发生了改变,也需要及时修正。但从文献资料分析来看,时间序列分析技术早已在国外应用,成为经济管理数据分析的基本方法。但为了提高预测的有效性,关于这方面进一步的工作还应继续。

(四)递归神经网络的结构和算法

1. 递归神经网络的结构

人工神经网络是模拟生物大脑神经元结构和功能,由彼此之间高度连接的处理单元构成的一种信息处理系统。神经网络通常由三层单元(节点)或人工神经元组成。第一层即输入层,包含一组处理单元,负责接收向网络输入的数据。输入层与隐含层相连接,隐含层既不直接接收输入数据也不直接产生输出数据。不同的网络可能含有一层或多层隐含层,最后一层与输出层相连接。每一个输入单元都与隐含层的隐单元相连接。隐含层最后一层的隐单元都与输出层的单元相连接。输出层含有一个或多个节点,这些节点产生输出数据。

本文所构造的递归神经网络由输入层、隐层和输出层组成。

设在每个离散时刻 k , $X_i(k)$ 为网络输入, $i=1,2,\dots,n$; $O_j(k)$ 为网络输

出; $S_j(k)$ 和 $Z_j(k)$ 分别为隐含层输入和输出, $j=1, 2, \dots, m$; W_j^o 为隐含层至输出层的连接权, W_j^d 为递归神经元连接权, 则有

$$\begin{aligned} S_j(k) &= W_j^d \cdot Z_j(k-1) + \sum_{i=1}^n W_{ji}^h \cdot X_i(k) \\ Z_j(k) &= \varphi(S_j(k)) \\ O &= \sum_{j=1}^m W_j^o \cdot z_j(k) \end{aligned}$$

则非线性动态系统(1)在 $k+1$ 时刻的预测值 $\hat{y}_i(k+1|k)$ 可写为

$$\hat{y}_i(k+1|k) = O(x_j) = f(\theta, k, D^k) \quad (2)$$

其中, θ 是 W_j^o 的集合, D^k 是以前输入输出数据集合。

给定准则函数:

$$\begin{aligned} J(\theta) &= \sum_{i=1}^q W_i E[y_i(k+1|k) - \hat{y}_i(k+1|k)]^2 \\ &= \sum_{i=1}^q W_i E[y_i(k+1|k) - O(x_j)]^2 \end{aligned}$$

2. 递归神经网络算法的稳定性

所有的参数利用最速下降法原理, 从给定的初值 $\hat{\theta}(0)$ 出发, 沿着准则函数 $J(\theta)$ 的负梯度方向修正参数估计值 $\hat{\theta}(k)$, 直至准则函数 $J(\theta)$ 达到最小值, 数学期望表达式可写成

$$\hat{\theta}(k+1) = \hat{\theta}(k) - R(k) \text{grad}[J(\theta)]|_{\hat{\theta}(k)} \quad (3)$$

其中, 步长间隔 1 的选择必须使输入向量 $x(k)$ 的条件协方差是常数阵, 且与 $\hat{\theta}(k)$ 无关; $R(k)$ 是 N 维的对称阵, 为权矩阵; $\text{grad}[J(\theta)]$ 表示准则函数关于 θ 的梯度, 即

$$\text{grad}[J(\theta)]|_{\hat{\theta}(k)} = -[y(k) - x^T(k)\hat{\theta}(k)]x(k) \quad (4)$$

将(4)代入(3)式可以得到 $\hat{\theta}(k+1)$ 的另一种表示方式:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + R(k)x(k)[y(k) - x^T(k)\hat{\theta}(k)] \quad (5)$$

由此, 可以得到关于非线性动态系统(1)参数的收敛性结论: 当输入向量不含噪声时, (5)式可给出渐近无偏参数估计值, 即有 $\lim_{k \rightarrow \infty} E\{\hat{\theta}(k)\} = \theta_0$

上述结论是可以进行证明, 对(5)式两边同时取数学期望, 得

$$\begin{aligned} E\{\hat{\theta}(k+1)\} &= E\{\hat{\theta}(k) + R(k)x(k)[y(k) - x^T(k)\hat{\theta}(k)]\} \\ &= E\{\hat{\theta}(k)\} + R(k)E\{x(k)y(k)\} - R(k)E\{x(k)x^T(k)\hat{\theta}(k)\} \end{aligned}$$

因为步长间隔 l 的选择必须使输入向量 $x(k)$ 的条件协方差是常数阵, 且与 $\hat{\theta}(k)$ 无关, 所以

$$E\{x(k)\hat{\theta}(k)\} = 0$$

同时利用条件数学期望公式 $E\{f(x)y\} = E\{E\{f(x)|y\}y\}$, 则有(5)式的第三项可以写成

$$E\{x(k)x^T(k)\hat{\theta}(k)\} = E\{E\{x(k)x^T(k)|\hat{\theta}(k)\}\hat{\theta}(k)\}$$

又因为输入是无噪声的, 因此上式为 0。从而有 $E\{\hat{\theta}(k+1)\} = E\{\hat{\theta}(k)\}$, 故得

$$\lim_{k \rightarrow \infty} E\{\hat{\theta}(k)\} = \theta_0$$

上述证明得出的结论说明了该算法具有稳定性。

三、对我国钢铁产量的预测分析

至此, 我们可利用 1952 至 2001 年的钢铁产量数据运用所构造的模型及算法进行预测, 数据来源于 2002 年的中国统计年鉴, 见表 1。

表 1 1952~2001 年我国钢铁年产量 单位: 万吨

年 份	钢铁产量	年 份	钢铁产量	年 份	钢铁产量
1952	135.00	1982	3 716.00	1992	8 094.00
1957	535.00	1983	4 002.00	1993	8 956.00
1962	667.00	1984	4 347.00	1994	9 261.00
1965	1 223.00	1985	4 679.00	1995	9 535.99
1970	1 779.00	1986	5 220.00	1996	10 124.06
1975	2 390.00	1987	5 628.00	1997	10 894.17
1978	3 178.00	1988	5 943.00	1998	11 559.00
1979	3 448.00	1989	6 159.00	1999	12 426.00
1980	3 712.00	1990	6 635.00	2000	12 580.00
1981	3 560.00	1991	7 100.00	2001	15 163.44

将 1952~1997 年中 26 个年产量数据用于建模, 1998~2001 年的 4 个年产量数据用于比较, 采用均方误差对预测的数据和原来的数据进行比较。

在进行预测之前, 首先将数据先进行归一化处理, 即找出所有数据的绝对值的最大值, 然后将所有 30 个数据都除以这个绝对值的最大值, 这样就可以将所有数据的模都小于 1。最后按如下所给的算法进行处理得到新的数据, 再将所得数据还原就可以得到预测值 $\hat{y}_i(k+1|k)$ 。

(a) 初始化权值 $W_{ji}^h = 0, W_j^o = 0$ 选取 m 的大小和 $f_k(\cdot)$, $k=1, 2, \dots, m$ 的次数, 其中 $1 \leq i \leq q, 1 \leq j \leq p$ 。

(b) 最小化目标函数 $L = \sum_{i=1}^q [y_i - O(x_i)]^2$, 求得初步模型 $y_i^{(1)} = O(x_i)$ 。

(c) 令 $r_i(k) = y_i(k+1|k) - \hat{y}_i(k+1|k)$, ($1 \leq i \leq q$), 用 $r_i(k)$ 代替

$y_i(k+1 | k)$ 重复(a)(b), 得到 $y_i^{(2)} = O(x_i)$ 。

(d) 反复进行(c)中的迭代, 即得到 $y_i^{(3)}, \dots, y_i^{(m)}$, 直到某个第 m 步, 残差平方和不大于某个给定值。

这样, 就得到了 $y_i(k+1 | k)$ 的估计 $\hat{y}_i(k+1 | k) = O(x_i)$ 。预测结果如图 1 所示:

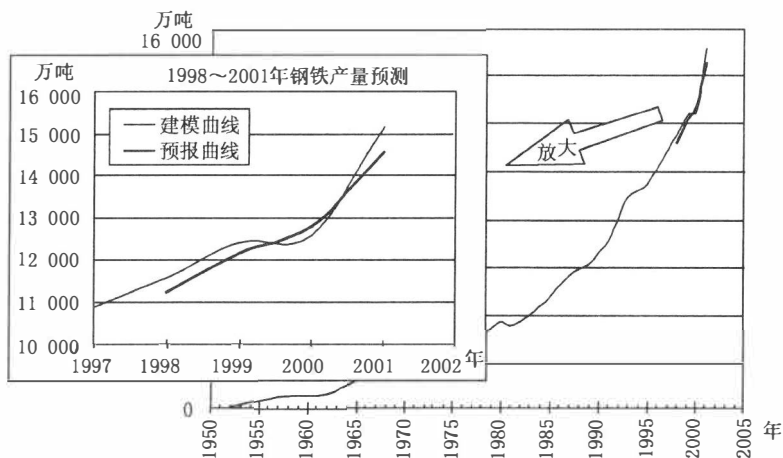


图 1 钢铁的年产量预测

从图 1 中可以看到, 预测线与实际线非常接近, 两者的均方误差为 0.024, 用人工神经网络和时间序列分析方法的均方误差分别为 0.021 和 0.028。比较的结果说明: 人工神经网络的预测精度要高于递归神经网络, 因为它可以以任意精度逼近它所要达到的结果, 但是在该实例中如果网络的初始权值接近于 0 的话, 将会出现陷入局部极小的情况。时间序列分析法的结果要低于递归神经网络的结果, 主要原因是拟合的模型需要不断地进行修正。比较结果说明递归神经网络算法是可行的。

四、结 论

预测问题是基于数据挖掘和知识发现的初步应用, 现在还没有一个通用的理论用来确定解决某一给定问题的神经网络的类型、层数、节点数或者所采用的学习算法。为了解决预测问题, 本文通过已有的神经网络的文献得到一些规则, 并进行多次论证, 将之应用于预测问题, 建立了解决预测问题的递归神经网络技术。通过对钢铁产量的预测结果说明这种算法的可行性, 当然这种算法仍存在某些缺陷, 例如如何确定最合理的节点个数还没有理论基础来支持, 目前只能依靠经验来获取。但是, 随着基于数据挖掘和知识发现的神经网络技术的不断发展, 以及计算机计算能力的发展和业务复杂性的提高, 数据的类型会越来越多、越来越复杂, 数据挖掘将发挥出越来越大的作用。在不远的将来, 基于数据

挖掘和知识发现的预测研究必然可以取得更加满意的效果。

参考文献:

- [1]Chow W S, Fang Y. A recurrent neural-network-based real-time learning control strategy applying to nonlinear systems with unknown dynamics[J]. IEEE Trans on Neural Network, 1998, 45(1): 151~161.
- [2]韦巍. 一种回归神经网络的快速在线学习算法[J]. 自动化学报, 1998, (5).
- [3]K Hornik. Some new results on neural networks approximation[J]. Neural Networks, 1993, Vol. 6: 1069~1072.
- [4]Leontaritis I J, Billings S A. Input-output parametric models for nonlinear systems, Part I; Deterministic nonlinear systems[J]. Int. J. Control, 1985, Vol. 41, No. 2: 303~328, 329~344.
- [5]Ku C C, Lee K Y. Diagonal recurrent neural networks for dynamic systems control [J]. IEEE Trans on Neural Network, 1995, 6, (1): 144~156.
- [6]国家统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2002.
- [7]邓祖新, SAS系统和数据分析[M]. 北京: 电子工业出版社, 2002.

Prediction Research of the Recursive Neural Network of Nonlinear Dynamic System Prediction Analysis of the Iron and Steel Output in China

LIU Lan-juan, XIE Mei-ping

(Information Department, Shanghai University of Finance and Economics, Shanghai 200433, China)

Abstract: The iron and steel industry is one of the pillar industries of the national economy in China and makes positive contribution to the sustainable development of the national economy. So, the predicting study of iron and steel industry has been a very important research topic. This paper employs the neural network method, time series analysis method and recursive neural networks technology based on data mining and knowledge discovery to predict the iron and steel output. In this paper, the emphasis is laid on recursive neural networks technology. It makes comparison between its predicted results and those of the above two methods. The comparison result shows that recursive neural networks technology is feasible.

Key words: iron and steel output; neural network; time series analysis; recursive neural network technology