

文章编号:1000-582X(2008)05-0536-05

组合预测模型在地方财政收入预测中的应用

范敏,石为人,梁勇林,华海玉

(重庆大学自动化学院,重庆400030)

摘要:根据地方财政收入预测受到多因素影响和经济系统具有非线性本质的特点,针对现有预测方法的不足,提出了一种组合预测方法。该方法首先通过灰色关联分析确定影响地方财政收入的主要指标,然后用灰色预测模型分别对各指标进行预测,最后将各指标的预测值作为输入,相应的地方财政收入实际值作为输出,训练并建立神经网络模型。实例分析表明灰色关联分析排除了非主要指标的干扰,灰色预测模型提供了较完善的输入数据,神经网络模型考虑了各主要指标的关联关系。实验结果证实该方法在地方财政收入预测中是有效可行的。

关键词:地方财政收入预测;灰色关联分析;灰色模型;神经网络

中图分类号:TP182

文献标志码:A

Application of a combination forecasting model in local financial revenue forecasting

FAN Min, SHI Wei-ren, LIANG Yong-lin, HUA Hai-yu

(College of Automation, Chongqing University, Chongqing 400030, P. R. China)

Abstract: Local financial revenue indices are affected by many factors and their economic systems thus are characterized by nonlinear properties. Moreover, current forecasting methods have a few disadvantages. A combination forecasting model therefore was proposed. The methods include the followings steps: First, the main factors for local financial revenue were confirmed via gray correlation analysis. Second, the gray forecasting model, GM(1,1), was applied to predict each index. Finally, the results of GM(1,1) were used as inputs and the actual data of relevant local financial revenue was used as outputs, then, a neural network was built. The results suggest that the gray correlation analysis can filter the accidental indices, the gray forecasting model can provide good input data sequences, and the neural network can process the relationships of indices. Experimental results demonstrated the availability and feasibility of the model in local financial revenue forecasting.

Key words: local financial revenue forecasting; gray correlation analysis; gray model; neural networks

为更好地探讨财政经济的发展和财政体制改革中急待解决的问题,对地区财政收支状况进行预测研究非常重要^[1]。地方财政收入体系是一个多层指标体系,按划分标准的不同可分为增值税、营业税、

个人所得税、企业所得税、房产税、农业税、罚款收入、调拨收入及其他收入等。并且各类指标发展趋势也较复杂,易受不确定、不可避免因素的影响,使得地方财政收入的预测难以准确进行。

收稿日期:2008-01-28

基金项目:重庆市科技攻关计划项目(7818-08);重庆市自然科学基金项目(CSTC,2006BB2190)

作者简介:范敏(1975-),女,重庆大学博士研究生,主要从事系统与知识工程、智能控制与智能管理的研究。

石为人(联系人),男,重庆大学教授,博士生导师,(E-mail)wrs@cqu.edu.cn.

欢迎访问重庆大学期刊网 <http://qks.cqu.edu.cn>

财政经济系统既具有明显的非线性、关联性和系统性,又具有一定的随机性,甚至某种混沌性。因此,对地区财政收入预测,试图用一个模型来描述其变化规律,或用一种算法来计算出较好的预测值,都较困难^[2]。所以,需要将各种预测方法综合起来,互相取长补短,以实现不同类型、不同层次信息和知识的集成。目前,灰色理论和人工神经网络都是国内外经济预测研究的热点^[1-8]。灰色系统理论主要研究灰色系统的分析、建模、预测、决策和控制等方面,其灰色预测模型是采用 GM(1,1)、GM(1,M)等进行定量预测分析,它所需建模信息极少,运算方便,建模的精度较高。人工神经网络主要解决非线性系统的建模和预测等问题,它具有有效的学习方式,在预测领域中具有明显优势。

以现行的财政收入指标体系为依据,利用上述方法的组合对财政收入预测问题进行了从局部到整体的分析。在建模过程中,将地方财政收入作为一个整体考虑,利用灰色关联分析方法分析主要影响指标,采用灰色预测模型 GM(1,1)分别对主要指标序列进行预测,削弱数据序列的波动,最后利用 BP 神经网络建立起这些指标变量之间的相互联系,形成一个地方财政收入预测的组合预测模型。

1 基本原理

1.1 灰色关联分析

灰色关联分析方法^[8]是用关联度大小来描述事物之间、因素之间关联程度的一种定量方法。即在灰色系统中存在两组随机序列,其中,一个数据序列称为“参考模式”或者“参考序列”,另一个序列为关联序列,灰色关联分析方法就是考察关联序列与参考序列的相关性。

参考序列包含有 n 个特征项(可视为时间序列),表示为

$$X_0 = (x_0(1), x_0(2), \dots, x_0(n)), \quad (1)$$

关联序列也包含有 n 个特征项,表示为

$$X_i = (x_i(1), x_i(2), \dots, x_i(n)), 1 \leq i \leq m, \quad (2)$$

将一系列 X_i 序列设定为 X_0 的影响因素。则 X_i 与 X_0 的灰色关联度定义为

$$\gamma_i = \gamma_i(X_0, X_i) = \frac{1}{n} \sum_{k=1}^n \gamma_{ik}, \quad (3)$$

其中, $\gamma_{ik} = \gamma_{ik}(x_0(k), x_i(k))$ 为关联系数,记为
$$\frac{\min_i \min_k |x_0(k) - x_i(k)| + \zeta \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \zeta \max_i \max_k |x_0(k) - x_i(k)|}, \quad (4)$$

式中 ζ 是关联权重,且 $\zeta \in [0, 1]$,通常取 0.5。

灰色关联度可理解为各关联数列构成的项与参考数列对应项间的相对空间距离。灰色关联度越大,它们之间相对距离越小,说明两序列的趋势越相近,反之亦然。因此,以灰色关联度的大小和差异程度为标准来选取主要影响因素。

1.2 灰色预测模型

灰色模型(Gray Model, GM)^[4-5]是用时间数据序列建立系统的动态模型。它把一组离散的、随机的原始数据列经过 m 次累加生成规律性强的累加生成序列,从而弱化原始序列的随机性;然后对累加生成数列建模,最后进行 m 次累减还原成预测值。

在采用灰色预测模型时,取 $m=1$,作一次累加生成数列建模,即 GM(1,1)灰色模型。模型的计算步骤如下:

Step1:建立第 i 个关联序列 X_i 的初始序列

$$X_i^{(0)} = (x_i(1), \dots, x_i(j), \dots, x_i(n)), \quad (5)$$

其中, $x_i(j)$ 是表示 j 时刻的时序数据。

Step2:基于初始序列 $X_i^{(0)}$ 通过累加产生一个新的序列 $X_i^{(1)}$,表示为

$$X_i^{(1)} = (x_i^{(1)}(1), \dots, x_i^{(1)}(k), \dots, x_i^{(1)}(n)), \quad (6)$$

其中 $x_i^{(1)}(k)$ 可通过下式得到

$$x_i^{(1)}(k) = \sum_{j=1}^k x_i^{(0)}(j), 1 \leq k \leq n, \quad (7)$$

Step3:建立 $X_i^{(1)}$ 的一阶线性微分方程:

$$\frac{dX_i^{(1)}}{dk} + aX_i^{(1)} = u, \quad (8)$$

此式即为 GM(1,1)预测模型。

Step4:求解微分方可得

$$\hat{x}_i^{(1)}(k+1) = \left(x_i^{(0)}(1) - \frac{u}{a}\right)e^{-ak} + \frac{u}{a}, \quad (9)$$

$$\hat{x}_i^{(0)}(k+1) = \hat{x}_i^{(1)}(k+1) - \hat{x}_i^{(1)}(k), \quad (10)$$

其中

$$\hat{a} = (a, u)^T = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{C}, \quad (11)$$

$$\mathbf{B} = \begin{bmatrix} -0.5[x_i^{(1)}(1) + x_i^{(1)}(2)] & 1 \\ -0.5[x_i^{(1)}(2) + x_i^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -0.5[x_i^{(1)}(n-1) + x_i^{(1)}(n)] & 1 \end{bmatrix}, \quad (12)$$

$$\mathbf{C} = \begin{bmatrix} x_i^{(0)}(2) \\ x_i^{(0)}(3) \\ \vdots \\ x_i^{(0)}(n) \end{bmatrix}, \quad (13)$$

$\hat{x}_i^{(0)}(k+1)$ 是 $x_i^{(0)}$ 在 $k+1$ 时刻的预测值。最后

得到 $X_i^{(0)}$ 序列的预测值 $\hat{X}_i^{(0)}$ 为

$$\hat{X}_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n)).$$

1.3 BP 神经网络

BP 神经网络是一种多层前馈神经网络,层间神经元实现全连接,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 网络由 3 层构成,即输入层、隐含层和输出层。它以一系列的输入和理想的输出作为训练的样本,采用误差反向传播方法训练网络节点权重和阈值,通过节点权重的调节使误差精度到达允许的范围,从而建立输入输出的映射关系。训练成功的 BP 网具有良好的泛化能力。具体的 BP 算法步骤参考文献[9-10]。

2 组合预测模型

2.1 地方财政收入(预算内)指标体系分析

地方财政收入(预算内)指标体系^[1]具有层次结构,如图 1 所示。

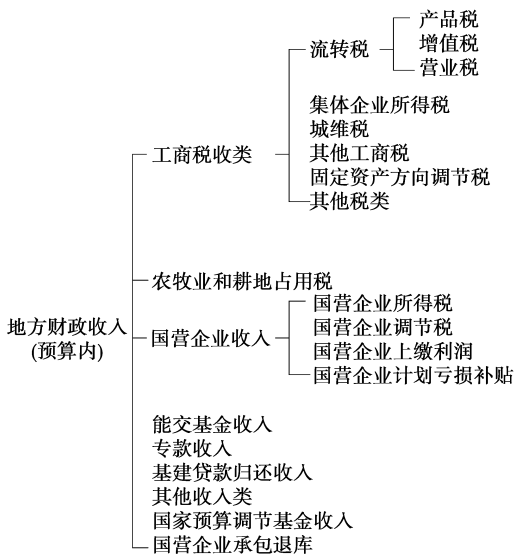


图 1 地方财政收入(预算内)指标体系

从图 1 可以看出,地方财政收入(预算内)受到多层指标影响。由于地方财政对财政收入划分的标准(见图 2)不同,采用分类三进行影响指标项的选择。

2.2 地方财政收入的组合预测模型

根据地方财政收入指标体系的层次性特点,结合上述理论模型的优势,设计了从单一指标预测到多种指标组合预测的地方财政收入组合预测模型。具体模型结构如图 3 所示。模型中将地方财政收入

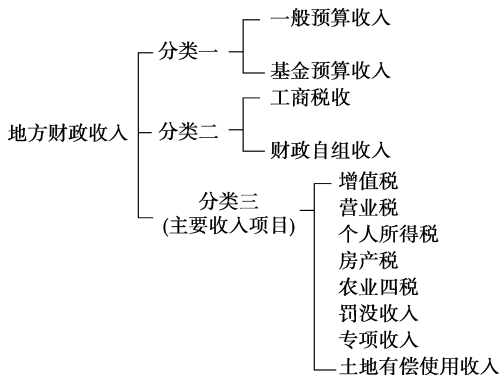


图 2 财政收入划分标准

时间序列设置为参考序列 X_0 ,各影响指标时间序列为关联序列 X_i 。

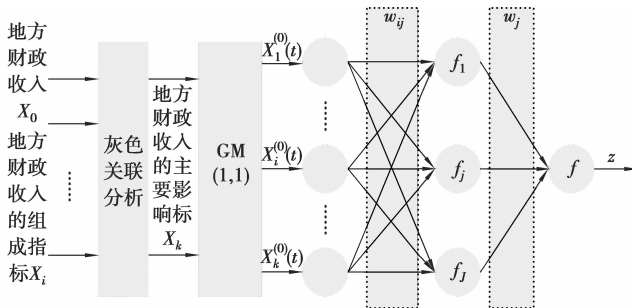


图 3 组合预测模型结构图

组合预测模型首先采用关联分析进行地方财政收入的主要影响因素分析,比较不同指标时间序列与地方财政收入时间序列的灰色关联度,根据关联度大小和预先设置的差值阈值剔除冗余因素;然后对多个主要影响指标序列分别建立 GM(1,1)模型,可以得到对原始序列的一系列预测值;由于预测值与原始值之间一定存在偏差,因此将这些预测值与实际值之间的偏差关系及序列之间的相互关系再综合到神经网络模型中。最后将 GM(1,1)模型的预测值作为神经网络的输入样本,相关地方财政收入的实际值作为输出样本,进行模型训练,调整相应节点的权值与阈值,最终建立财政收入预测模型。

3 实例分析

以重庆某地区地方财政收入数据为例(见表 1)进行分析,主要分析以下收入项目:地方财政收入(X_0),增值税(X_1),营业税(X_2),个人所得税(X_3),企业所得税(X_4),房产税(X_5),农业税(X_6),罚款收入(X_7),调拨收入(X_8)。

1) 选择主要影响因素

将地方财政收入体系视为一个灰色系统,其中

地方财政收入为参照序列 X_0 , 其各项组成指标为关联序列 $X_i (i=1, 2, \dots, 9)$ 。基于表 1 数据, 计算关联序列和参照序列间的灰色相关度 γ_i , 结果见表 2。灰色相关度反映了每一个关联序列 X_i 对参照序列 X_0 的权重, 按照降序排列后, 灰色相关度顺序为:

$$\gamma_2 > \gamma_3 > \gamma_1 > \gamma_4 > \gamma_6 > \gamma_7 > \gamma_8 > \gamma_5$$

通过比较上述权重可以去掉冗余指标项, 保留主要影响因素。从计算结果可以看出 $\gamma_5, \gamma_7, \gamma_8$ 之间的差小于差值阈值 0.01, 说明 (X_5, X_7, X_8) 指标项与 X_0 的灰色相关度是一致的, 且 X_5 和 X_8 指标项与 X_0 相关度较小, 因此选择保留 X_7 , 去掉其余各项。最后由 $(X_1, X_2, X_3, X_4, X_6, X_7)$ 6 个指标组成预测系统的主要影响因素。

表 1 重庆某地区地方财政收入及各指标部分数据样本

年份	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1997	13 024	2 096	3 103	1 857	748	572	862	1 034	496
1998	16 526	2 650	3 738	2 876	928	641	1 059	364	653
1999	20 274	2 541	4 676	3 799	1 616	703	1 094	689	682
2000	25 101	2 939	5 534	5 280	2 706	690	1 322	858	677
2001	31 439	4 402	6 463	6 803	4 038	763	2 057	858	1 138

表 2 各指标灰色相关度

灰色相关度	X_0	X_1	X_2	X_3	X_4
γ_i	0.742 5	0.782 1	0.766 6	0.731 3	0.731 3
灰色相关度	X_5	X_6	X_7	X_8	
γ_i	0.707 9	0.721 4	0.709 7	0.709 2	

2) 对每一指标序列进行 GM(1,1) 预测

根据灰色关联分析的结果和地方财政收入及各指标的样本数据, 采用 GM(1,1) 预测从 1995 到 2005 年的数据, 得到 6 个指标项的预测序列:

$$(X_1^{(0)}, X_2^{(0)}, X_3^{(0)}, X_4^{(0)}, X_6^{(0)}, X_7^{(0)})$$

3) BP 神经网络设计

设计 3 层 BP 神经网络模型, 以 6 个指标项的预测序列 $(X_1^{(0)}, X_2^{(0)}, X_3^{(0)}, X_4^{(0)}, X_6^{(0)}, X_7^{(0)})$ 为输入节点, 地方财政收入 (X_0) 为输出节点, 隐藏节点数由式(14)进行估算

$$n_1 = \sqrt{n + m} + a, \quad (14)$$

其中, n_1 是隐藏节点数, n 是输入的神经元个数, m 是输出的神经元个数, $a \in [1, 10]$ 。

在 MATLAB 中进行 BP 神经网络模型训练, 以 1995~2001 年的数据为训练样本, 2001~2005 年的数据为测试样本, 设置误差精度为 $\epsilon = 0.001$ 。模型通过增加 n_1 的个数来测试误差的收敛速度。在期

望的精度范围内, 隐藏节点为 9 个时, 模型可达到最好的收敛率。所以最后确定 BP 神经网络为 6:9:1 的拓扑结构。神经网络模型经过上百步迭代后形成最终的预测模型。采用该模型预测的结果与单独采用 GM(1,1) 模型结果进行了对比, 结果见表 3。

表 3 组合模型与 GM(1,1) 模型预测误差对比表

年份	2001	2002	2003	2004	2005
组合模型	0.002 7	0.053 5	0.021 4	0.040 0	0.021 2
GM(1,1)模型	0.016 8	0.011 4	0.042 3	0.089 3	0.045 9

从实验结果分析, 组合模型的预测结果除了 2002 年的误差高于 GM(1,1) 模型外, 其余年份预测结果误差均低于 GM(1,1) 模型; 并且组合模型的预测误差最大为 5.35%, GM(1,1) 模型的预测误差最大为 8.93%, 显然前者低于后者。说明文中提出的组合预测模型充分考虑了多指标的共同作用, 较好地改善了综合指标的预测精度, 在地方财政收入预测中是切实可行的。另外, 在计算中即使缺乏某年的指标项数据, 通过 GM(1,1) 单指标预测所产生的预测序列, 能弥补神经网络输入数据的不足, 使组合模型仍能较准确预测出未来年份的综合指标趋势。

4 结 论

组合预测方法通过实验论证可以得出 3 点结论。第一, 利用灰色关联分析可以分析主要影响因素, 有效减少了输入神经元的个数。第二, 采用灰色预测模型 GM(1,1), 能对单因素进行预测, 并且削弱了数据序列的波动, 为神经网络提供满意的输入样本。第三, 神经网络能建立非线性系统多变量之间的关联关系。同时, 由于灰色系统原理和神经网络模型的结合, 实现了优势互补, 有效地改善了模型的预测精度, 在地方财政收入预测中取得了较好的效果。但是随着样本量的增加, 为了保证模型的预测精度, 下一步的研究主要考虑解决模型的动态学习问题。

参考文献:

[1] 李洪心, 易允文. 财政预测模型研究及应用[J]. 信息与控制, 1997, 26(3): 215-220.
LI HONG-XIN, YI YUN-WEN. Research and application of financial forecasting model [J]. Information and Control, 1997, 26(3): 215-220.
[2] 郭崇慧, 唐焕文. 宏观经济预测模型体系研究[J]. 运筹与管理, 2001, 10(4): 1-8.

- GUO CHON-GHUI, TANG HUAN-WEN. Research on Macroeconomic Forecasting Model System[J]. Operations Research and Management Science, 2001, 10(4):1-8.
- [3] 邓聚龙. 灰预测与灰决策(修订版)[M]. 武汉: 华中科技大学出版社, 2002.
- DENG JULONG. Grey Forecasting and Decision [M]. Wuhan: Huazhong University of Science and Technology Press, 2002.
- [4] 邢棉. 季节性预测的组合灰色神经网络模型研究[J]. 系统工程理论与实践, 2001, 21(1):31-35.
- XING MIAN. Research on Combined Grey Neural Network Model of Seasonal Forecast [J]. System Engineer Theory and Practice, 2001, 21(1):31-35.
- [5] 刘斌, 刘思峰, 党耀国. 基于灰色系统理论的时序数据挖掘技术[J]. 中国工程科学, 2003, 15(9): 32-35.
- LIU BIN, LIU S-IFENG, DANG YAO-GUO. The Time Sequence Data Mining Techniques Based on Grey System Theory [J]. Engineering Science, 2003, 15(9): 32-35.
- [6] 秦四强, 魏砾宏. 基于灰色关联度和BP神经网络研究生物质热解产气特性[J]. 太阳能学报, 2003, 24(6): 776-780.
- QIN SI-QIANG, WEI LI-HONG. Study on Characteristics of Biomass Pyrolysis Gas in View of Grey Relation Analysis and B PNeural Network [J]. Acta Energetica Solaris Sinica, 2003, 24(6): 776-780.
- [7] 徐建华, 李建霞, 吴玉鸣. 中国粮食多因子灰色关联神经网络预测研究[J]. 华中师范大学学报:自然科学版, 2002, 36(4): 419-423.
- XU JIAN-HUA, LI JIAN-XIA, WU YU-MING. A Grey Relational Analysis and Artificial Neural Networks of Corn Production Prediction in China[J]. Journal of Central China Normal University: Nat. Sci., 2002, 36(4): 419-423.
- [8] 邓宏贵, 罗安. 关联度分析在变压器故障诊断中的应用[J]. 电力系统自动化, 2005, 29(18):73-75.
- DENG HONG-GUI, LUO AN. Application of Three-level Dependent Degree for Transformer Fault Diagnosis[J]. Automation of Electric Power System, 2005, 29(18):73-75.
- [9] 蒋宗礼. 神经网络导论[M]. 北京: 高等教育出版社, 2001.
- [10] 飞思科技产品研发中心. MATLAB6.5 辅助神经网络分析与设计[M]. 北京: 电子工业出版社, 2003.

(编辑 陈移峰)

(上接第 535 页)

参考文献:

- [1] 王经瑾. 核电子学(上、下册)[M]. 北京: 原子能出版社, 1985.
- [2] 张钦玉, 周清华, 罗小兵, 等. 核信号数值仿真方法的研究及应用[J]. 核电子学与探测技术, 2006, 26(4):421.
- ZHANG RUAN-YU, ZHOU QING-HUA, LUO XIAO-BING, et al. The research and application of digital simulation in nuclear signal [J]. Nuclear Electronics & Detection Technology, 2006, 26(4):421.
- [3] KNOLL G F. Radiation detection and measurement [M]. New York: John Wiley & Sons Press, 1979.
- [4] 王芝英. 核电子技术原理[M]. 北京: 原子能出版社, 1989.
- [5] VALENTINE T E. Review of subcritical source-driven noise analysis measurements[R]. the U. S. Department of Energy Report: ORNL/TM-1999/288, 1999.
- [6] MIHALCZO J T, VALENTINE T E, MATTINGLY J K, et al. Active neutron interrogation for verification of storage of weapons components at the oak ridge Y-12 plant[R]. the U. S. Department of Energy Report Y/LB-15.964, 23, 1998.
- [7] MATTINGLY J K, MARCH-LEUBA J A, MIHALCZO J. T, et al. Passive NWIS measurements to estimate shape of plutonium assemblies (slide presentation) [R]. the U. S. Department of Energy Report Y/LB-15.988, 25, 1998.
- [8] 李鹏宇, 薛志华, 刘松秋. 核随机脉冲快速测量及统计特性研究[J]. 核电子学与探测技术, 2006, 26(6):886.
- LI YU-PENG, XUE ZHI-HUA, LIU SONG-QIU. Fast measurement for the statistical distribution of nuclear random pulse [J]. Nuclear Electronics & Detection Technology, 2006, 26(6):886.
- [9] 罗璋琳, 罗安仁. 实验堆物理[M]. 北京: 原子能出版社, 1987:208-215.
- [10] 董长虹, 高志. 小波分析工具箱原理与应用[M]. 北京: 国防工业出版社, 2004.

(编辑 陈移峰)