

建设需求量预测分析中的人工神经网络和多元回归方法

王颖林, 赖茂宇, 郭丰敏

(福建农林大学交通与土木工程学院, 福建 福州 350001)

摘 要:利用人工神经网络(ANN)和多元回归(MR)预测方法分别基于中国统计年鉴和香港房屋署的相关数据对中国内地和香港地区的建设需求量进行预测,并对两种预测手段得到的预测结果的可信度和离散程度进行对比分析.基于 ANN 和 MR 两种预测手段的不同特性,从预测结果中可以看出,就香港地区的预测情况而言,ANN 方法产生的结果比回归模型更加精确;从内地的预测结果来看,ANN 和 MR 的预测精度几乎一致.对于存在较大波动性的数据而言,ANN 模型建立的非线性关系可以更精确地描述预测结果,反之,两种预测模型的应用均可得出良好结果.同时,经预测得知,两地的建筑需求量都存在上升趋势,有关部门应采取相应措施提前做好规划工作.

关键词:建设需求;人工神经网络;回归分析;预测

中图分类号:F281

文献标识码:A

doi:10.3969/j.issn.1674-2869.2013.11.016

0 引 言

建设需求量的增加会带来社会的发展,然而,污染问题也会随之而来,深入探讨建设需求量对经济增长以及环境保护具有十分重要的意义.一方面,建设需求量的预测可以为开发商、招标企业和政府的规划发展提供建议.另一方面,建筑业是世界上废弃物产生量最多的行业之一.在中国,城市建筑及拆卸废物(CDW)已达到垃圾产生量的30%~40%,城市化进程的加快和城市重建产生的大规模建设和拆迁活动是城市废弃物产生的主要原因^[1].因此,建设需求量的预测可以为相关部门提供未来的建设需求信息,从而有助于提前制定相关的环保策略.

根据建设经济学,建设需求与消费者投资密切相关,社会总的建筑需求主要取决于投资者投资建设市场的意愿和能力.因此,在预测建设总需求时,可以以建筑安装工程总投资作为预测目标^[2].理论分析和统计经验显示,建设安装工程项目投资与国民收入,国民整体消费,商品零售价格指数以及人口数量相关联^[3].

在预测模型建立的过程中,适当的经济指标以及预测技术的选取对预测结果的准确性具有重要的影响.人工神经网络是一种非线性数据统计的建模工具,通常被用来建立具有复杂关系的输

入数据和输出数据之间的关系模型^[3].尽管多元回归分析方法比神经网络更加传统,但是在适当的条件下选用该方法也可以产生相同精度的结果.

欧阳红祥等^[4]借助 BP 神经网络建立了预测材料价格的模型,从而为建筑企业投标报价时正确预测材料价格走势提供了建议.李丽,张海涛^[5]利用 BP 人工神经网络方法对小城镇生态环境质量进行了预测.喻伟^[6]等应用人工神经网络来预测建筑能耗和室内热舒适状况,建立了基于 GA-BP 网络的建筑能耗和室内热舒适状况的预测模型,为建筑师在设计阶段简单准确地获得设计建筑的能耗和室内舒适状况提供了理论依据.王蓉等^[7]以康乐县为研究对象应用多元回归分析对城乡用地面积进行了预测. Victor Kipnis 和 Douglas Midthune 等^[8]对回归分析进行了矫正,并对不同数据结构中的回归方法和两阶段法进行了讨论.杜家龙^[9]对我国 2012 年国内生产总值回归预测进行了实证分析,研究了国内生产总值的主要影响因素及回归预测的新方法,对完善回归预测理论和方法,拓展预测研究思路,增强预测方法的选择性和应用性提出了建议.

1 研究的目标

本文的目的在于为中国香港和中国内地建

收稿日期:2013-09-29

基金项目:国家社会科学基金资助项目(13BGL150)

作者简介:王颖林(1987-),女,山西运城人,博士研究生.研究方向:项目管理.

设需求量的预测选取合适的变量及预测方法,并分析了这两个地区近年来建筑需求量的发展趋势.研究方法如下:第一,为两地的建设需求量预测分别选择适当的经济变量;第二,利用回归分析的方法确定适合的回归预测模型;第三,运用人工神经网络建立模型.最后,对两种预测方法产生的中国香港和内地的 4 组预测模型分别进行比较.

2 数据选择

2.1 数据选择——中国香港

由于建设需求和社会经济指标有密切的关联,因此住宅和非住宅建筑的需求可以通过相关的社会经济指标来反映.

对于中国香港而言,与建筑需求量有关的经济指标可以从香港有关部门的网站(如香港房屋署)得到.其中可以用来描述建设需求的经济指标包括人口规模(POP),人均国民收入(NPCI),相对价格指数(PI),它等于居民消费价格指数除以建造成本指数,家庭构成率(HF),工业生产指数(MPI),出口值(EXI),国内生产总值(GDP),政府收入与建设成本指数的比率(GRI),政府开支(GEI),国民生产总值(GNP),个人开支(PEI),国内固定资本形成总额(GFCF),建筑成本指数(CCI),居民消费价格指数(CPI),政府财政收入(GR)和政府总支出(GTS).除了人口规模以外,其他经济指标的单位为百万港币.

2.2 数据选择——中国内地

对于不同地区而言数据收集的方式有所区别,中国内地的建设经济指标来源于中国统计年鉴.根据统计经验和理论分析可知,建设工程安装投资与宏观经济系统有关,其中包括企业,银行和个人的社会行为.因此,一些经济数据,如国民收入,国民整体消费,商品零售价格指数,以及人口等数据可以作为因变量来分析建设需求量^[3].

对内地而言,有 13 种初始变量可供选择,其中包括人口规模(POP),国内贷款(DL),外国投资(FI),自筹资金(SF),人均国民收入(NPCI),相对价格指数(PI),等于居民消费价格指数与建设成本指数的比值,利率(INT),工业生产指数(MPI),出口值(EXI),国内生产总值(GDP),政府收入与建筑成本指数的比值(GRI),政府开支(GEI),国民生产总值(GNP).除了人口规模以外,其他经济指标的单位为亿元.

3 多元回归分析方法

3.1 应用多元回归法对香港建设需求的预测

如第二部分所述,在进行多元回归模型(MR)建立时对中国香港而言有 15 种经济变量可供选择,但需要通过逐步回归模型对其中与预测模型关联度较小的变量进行剔除.确定变量的 R^2 是为了使得该值足够大从而保证该变量能够准确描述模型的发展趋势.T 检验则是用于判断该变量是否与模型具有足够的相关性,从而准确地描述预测模型.对于不能准确描述模型走势以及相关性强弱的变量应予以剔除.

首先,用 SPSS 软件对上述 15 个变量进行相关性分析,从而可以剔除 8 个相关性较差的经济指标,它们分别是 MPI, GNP, PEI, GFCF, CCI, CPI, GR, GTS.

在剩余的 7 个经济指标中,首先应用 POP, NCPI, PI 和 GRI 来构建第一个模型,其中 Y 是建筑总需求的预测值.在这个模型中,可以得到 $R^2 = 0.858$,这意味着 85.8% 的总变异可以在这个回归模型中得到解释.当国内生产总值和政府总支出加入模型一之后可以得到第二个模型的 $R^2 = 0.968$,显然这里的 R^2 大于第一个模型.

HF 和 GEI 加入第二个模型后,不仅使得 R^2 增加,T 检验的显著性也有所提高,这意味着该模型更适合描述预测数据的走势.因此,POP, GDP, NCPI, GEI, PI, GRI 以及 HF 作为预测模型的经济变量.

数据选择的过程就是将不同的变量分别加入模型,并进行准确性分析,从而得出可选择的适当的变量.对不同变量组合的模型比较后,图 1 模型对于描述预测值而言具有相对较高的准确度:

$$Y = 66.13X_{\text{POP}} - 0.52X_{\text{GDP}} + 0.66X_{\text{NCPI}} + 0.95X_{\text{GEI}} + 1.07X_{\text{GTS}} + 0.25X_{\text{PI}} - 0.52X_{\text{GRI}} - 72.22X_{\text{HF}}$$

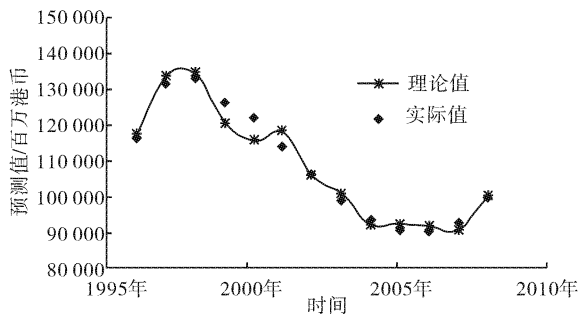


图 1 中国香港 MR 预测模型

Fig. 1 MR forecasting model for HK

3.2 应用多元回归法对内地建设需求的预测

与对中国香港地区进行建模时所采用的方法

一样,第一步是选择与预测值有较强相关性的影响因子.通过 SPSS 软件对 13 个经济变量进行分析后,3 个关联性较弱的指标(FI,GDP 和 GNP)被剔除.

在第一阶段中,选择 5 个经济指标(POP, NCPI,PI,INT 和 GRI)进行建模,并得到 $R^2 = 0.8594$,这一数值不具有较强的模型解释力度.将剩余的经济指数进行代入,将建模过程重复 3 次,并对这些模型进行比较,得到的相对最合适的预测模型如图 2.

$$Y = -397.24 + 33.84X_{\text{POP}} - 0.188X_{\text{DL}} - 0.014X_{\text{NCPI}} + 4.759X_{\text{PI}} + 0.92X_{\text{INT}} - 3.4X_{\text{MPI}} + 1.1X_{\text{EXI}} - 2X_{\text{GRI}} + 0.87X_{\text{GEI}}$$

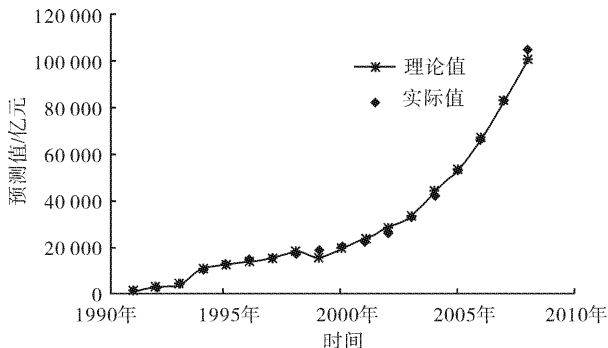


图 2 中国内地 MR 预测模型

Fig. 2 MR forecasting model for Mainland China

4 人工神经网络分析法

4.1 利用 ANN 在中国香港的建设需求预测

确定隐藏层节点的数量是人工神经网络预测模型建立过程中的难点之一^[4].如果隐藏节点太少,那么模型的准确性会受到影响,如果节点过多,则会导致网络训练过度并影响模型的特性.

在运用 ANN 对中国香港的数据进行预测的过程中发现,设置 5 个输入变量,每个隐藏层建立 4 个节点的模型可以达到预测目标的准确度.通过训练可知,对于香港地区的数据而言,最佳的网络结构是 5-4-1,即网络的输入节点数目是 5,隐藏节点的数量是 4,同时只有一个输出值.

可得 $R^2 = 0.99$,因此可以明显看出人工神经网络对数据的训练与实际数据的贴合度较高,能够很好地解释数据的发展趋势,见图 3.

4.2 利用 ANN 在中国内地的建设需求预测

对于中国内地的数据而言,有 9 个输入指标,因此输入层具有 9 个节点.经过反复训练可知,一个隐藏层中设计 6 个节点得出的计算结果具有较高的准确性.因此,内地预测数据的最佳网络结构是 9-6-1,表示输入节点的数目为 9,隐藏节点的数量为 6,且只有一个输出.

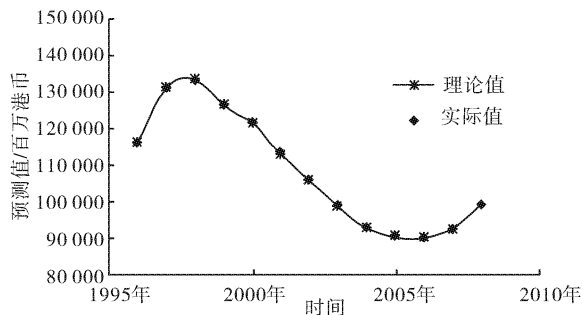


图 3 中国香港 ANN 预测模型

Fig. 3 ANN forecasting model for HK

目是 6,且只有一个输出.

经过网络训练可得判定系数 $R^2 = 0.96$,这意味着 ANN 建立的模型可以解释预测数据 96% 的总变异.虽然这里的 $R^2 = 0.96$ 与回归模型计算的 $R^2 = 0.958$ 几乎相同,但 ANN 模型所产生的数据趋势图大大优于回归绘制的结果,见图 4.

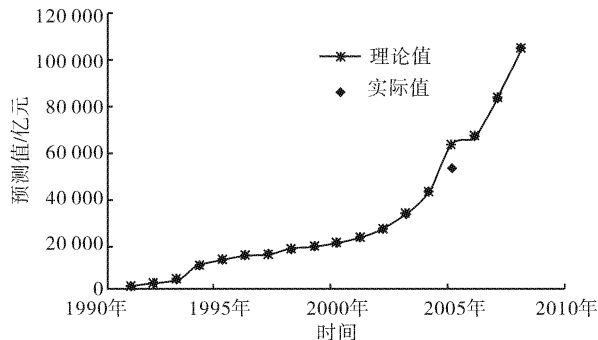


图 4 中国大陆 ANN 预测模型

Fig. 4 ANN forecasting model for Mainland China

5 模型比较

首先,通过多元回归模式,将中国香港和中国内地两者之间的预测结果进行比较.正如上述步骤计算出的结果,中国香港和中国内地的 R^2 分别为 0.97 和 0.958,相差仅为 0.012.这表明通过回归模式,在发达地区和发展中地区得出的结果几乎是一样精确的.

第二,使用人工神经网络比较中国香港和中国内地的预测结果.通过人工神经网络模型得出两个地区具有不同的 R^2 ,中国香港和中国内地分别是 0.99 和 0.96.我们可以看出,虽然对香港预测时采用的数据量少于内地,但是模型偏差比内地更小.

第三,在多元回归模式和人工神经网络模型基础上对香港的预测结果进行比较.对于中国香港来说,很明显可以看出由人工神经网络的拟合模型远远优于回归方法.这两个模型的预测准确度可以通过偏差值反映出来.人工神经网络和回归模式的 R^2 分别为 0.99 和 0.97.虽然这两种模

式都能够产生准确的预测结果,由于中国香港地区的房地产市场处于不稳定波动状态,人工神经网络模式在描述具有非线性关系的数据时具有明显优势,因此 ANN 产生的预测数据走势比回归分析模型具有更好的拟合度。

最后,将多元回归模式和人工神经网络预测对中国内地的预测结果进行比较。对于中国内地而言,无论是通过回归分析模型或人工神经网络方法,所产生的偏差几乎是一样的。由于中国的房地产市场相对稳定,呈现出了逐渐上升的趋势。因而,两种方法在预测时都可以得出较准确的结论。

6 结 语

在预测的过程中,合适的经济指标的选择对于预测结果的准确度具有至关重要的影响。例如,对于中国香港而言,约有 15 个与建设需求相关的经济指标可供选择,然而,根据回归分析法和人工神经网络的特性,一些关联性不高的变量要予以适当剔除。

本文用两种不同的预测方法对中国香港和中国内地的建筑需求量进行分析,产生了 4 个不同的模型。从理论上讲,由于神经网络模型内在特性,输入和输出变量之间的非线性关系可以自动获得,因此,神经网络模型与回归模型相比可能产生更加精确的结果。中国香港的数据分析的例子便可证明这一判断。但对于中国内地而言,两种方法产生了精确度几乎相同的结果。从中国香港及中国内地的预测结果中可以看出,在未来几年里,建设需求量呈增长趋势。这意味着虽然未来的经济发展趋势良好,但也将会有更多的建筑垃圾产生。因此,有关当局应采取一些有效的措施来处理和回收不断增长的建筑垃圾,从而保护环境、节约能源。

致 谢

感谢国家社会科学基金委员会的资金支持。

参考文献:

- [1] 赵利,鹿吉祥,顾洪滨. 建筑垃圾综合治理产业化运作与对策研究[J]. 建筑经济,2011(5):16-20.
ZHAO Li, LU Ji-xiang, GU Hong-bin. Comprehensive management of construction waste Industrial Operation and Countermeasures [J]. Construction Economics,2011(5):16-20. (in Chinese)
- [2] 赵黎明,沈艳丰,宫晓玲. 建筑市场需求预测中的神经网络方法[J]. 系统工程理论与实践,2000(10):121-125.
ZHAO Li-ming, SHEN Yan-feng, GONG Xiao-ling. Neural Networks Using in Construction Market Demand Forecasting [J]. Systems Engineering Theory and Practice, 2000 (10): 121-125. (in Chinese)
- [3] Khanna T. Foundation of Neural Networks [M]. Boston: Addison-Wesley, 1990.
- [4] 欧阳红祥,李欣,张信娟. 人工神经网络在建筑材料价格预测中的应用[J]. 武汉理工大学学报:信息与管理工程版,2013(1):115-118.
OUYANG Hong-xiang, LI Xin, ZHANG Xin-juan. Construction Materials Price Prediction by using Artificial Neural Networks[J]. Wuhan University of Technology: Information & Management Engineering,2013(1):115-118. (in Chinese)
- [5] 李丽,张海涛. 基于 BP 人工神经网络的小城镇生态环境质量评价模型[J]. 应用生态学报,2008(12):2693-2698.
LI Li, ZHANG Hai-tao. Small Towns Ecological Environment Quality Evaluation Model based on BP Artificial Neural Network [J]. Journal of Applied Ecology,2008(12):2693-2698. (in Chinese)
- [6] 喻伟,李百战,杨明宇,等. 基于人工神经网络的建筑多目标预测模型[J]. 中南大学学报:自然科学版,2012(12):4949-4955.
YU Wei, LI Bai-zhan, YANG Ming-yu, et al. Architecture multi-target prediction model based on artificial neural network [J]. Central South University: Natural Science Edition, 2012 (12): 4949-4955.
- [7] 王蓉,张仁陟,陈英. 基于多元回归分析和灰色模型的康乐县城乡建设用地预测[J]. 甘肃农业大学学报,2012(1):134-139.
WANG Rong, ZHANG Ren-bu, CHEN Ying. Kang Le County Rural Construction Land Prediction based on Multivariate Regression Analysis and Gray Model [J]. Journal of Gansu Agricultural University, 2012 (1):134-139. (in Chinese)
- [8] Victor Kipnis, Douglas Midthune, Laurence S Freedman, et al. Regression calibration with more surrogates than mismeasured variables [J]. Statist Med, 2012, 31:23.
- [9] 杜家龙. 国内生产总值回归预测新探[J]. 统计与决策,2013(9):9-14.
DU Jia-long. GDP regression Exploration [J]. Statistics and Decision,2013(9):9-14. (in Chinese)

(下转第 86 页)