

基于 BP 神经网络自贡房地产价格走势预测

杜道渊, 柏宏斌, 周 锋

(四川理工学院理学院, 四川 自贡 643000)

摘 要:文章通过分析调查影响自贡房地产市场的主要因素,基于 BP 神经网络,结合自贡住宅市场的实际情况,建立两类 BP 神经网络预测模型:基于时间序列的趋势预测模型、基于影响因素的回归预测模型,预测了自贡房地产市场价格走势。模拟预测 2010 年的结果证明了 2011 年房价预测的有效性,可为自贡城市建设的可持续发展提供有价值的指导意见。

关键词:神经网络; 房价预测; 响应函数; 训练函数

中图分类号: TP183

文献标识码: A

随着经济发展,全国多数大中城市房价一直呈上升趋势,自贡房价自 1999 年起也是节节攀升,房价问题引起社会的普遍关注。针对自贡的实际情况进行分析,建立模型预测 2010 年和 2011 年一、二、三、四季度的房价走势,通过比对 2010 年的统计结果,平均误差为 4.85%,所以 2011 年的预测具有一定指导性,可以在一定程度上帮助消费者和开发商正确的面对市场,对自贡大城市建设的可持续发展提供有价值的指导意见。

1 问题假设

对于房价的分析和预测,有很多的确定和不确定因素,在建立预测模型时需要对一些条件进行假设。设固定相关税收、土地价格、建材价格、劳动力薪酬和房产商做宣传的成本。主要考虑人均居住面积、人口总数、人均可支配收入和房地产开发投资等主要因素,并根据已有数据来分析与预测房价走势。

2 符号说明

符号说明见表 1。

3 BP 神经网络进行预测和分析

3.1 BP 神经网络预测的方法步骤

在现有的数据条件下,采用 Matlab7.0 提供的 BP 神经网络趋势预测和回归预测的思路,把自贡市房地产住宅市场的供给、需求与房价的历年数据以及其影响因素的数据分别作为学习样本,通过 BP 神经网络分析自

贡市房地产住宅市场的供给、需求、房价与其影响因素之间的关联关系,然后预测自贡市房地产住宅市场未来的发展情况^[1-3]。具体步骤如下:

表 1 符号说明

序号	符号	符号说明
1	z_i	神经元的第 i 个输入
2	θ	神经元的阈值
3	y_1	人均居住使用面积(平方米)
4	y_2	市区人口总数(万人)
5	y_3	人均可支配收入(元)
6	y_4	人均消费性支出(元)
7	y_5	住宅平均售价(元/平方米)
8	y_6	房地产开发投资(亿元)
9	y_7	地区生产总值(亿元)
10	y_8	贷款利率(%)
11	y_9	住宅销售面积(平方米)
12	E	测试集网络输出和目标误差
13	w^2	输出层第 k 个神经元与隐含层第 j 个关联度
14	w^1	输出层第 j 个神经元与隐含层第 i 个关联度
15	y_n	输入量
16	y_n	输出量
17	N	隐含层节点数
18	n	输入层节点数
19	M	输出层节点数
20	$f_j^o(x)$	隐含层激活函数
21	Mse	均方误差
22	h	学习率
23	a	动量因子
24	goal	收敛误差界值

第一步 输入变量的分析与预处理

为使网络训练更加有效,对神经网络的输入、输出

数据进行一定的预处理可以加快网络的训练速度,对数据进行归一化处理,即将输入、输出数据映射到 [-1, 1] 范围内,训练结束后再映射到原数据范围。

第二步 BP 神经网络结构设计

(1) 网络层数和隐层的确定

预测模型采用三层神经网络,即输入层—隐层—输出层结构和若干个隐层。

(2) 初始权值的选取

置隐层节点的初始值为均为分布在零附近的很小的随机值。置输出层节点所连的权值数一半为 +1,另一半为 -1。网络节点的偏置(θ)统一设置为 0。

(3) 响应函数(传输函数)的选取

采用 Sigmoid 激活函数。

$$w^1 = (w_{11} \mu_{v_2} \dots \mu_{v_{i-1}} \mu_{v_i})^T = \begin{pmatrix} w^1 & \dots & w^1 \\ \vdots & \ddots & \vdots \\ w^1 & \dots & w^1 \end{pmatrix}$$

$$w^2 = (w_{11} \mu_{v_2} \dots \mu_{v_{n-1}} \mu_{v_n})^T = \begin{pmatrix} w^2 & \dots & w^2 \\ \vdots & \ddots & \vdots \\ w^2 & \dots & w^2 \end{pmatrix}$$

对于每个节点:

$$Y^2 = (y_1 \ y_2 \ \dots \ y_{m-1} \ y_m)^T$$

隐含层激活水平:

$$S^1 = (s_1 \ s_2 \ \dots \ s_{m-1} \ s_m)^T$$

隐层输出:

$$Y^1 = (y_1 \ y_2 \ \dots \ y_{m-1} \ y_m)^T$$

输出层激活水平:

$$S^2 = (s_1 \ s_2 \ \dots \ s_{m-1} \ s_m)^T$$

输出层输出:

$$Y^2 = (y_1 \ y_2 \ \dots \ y_{m-1} \ y_m)^T$$

则:

$$Y^1 = F^1(S^1) = F^1(W^1 X)$$

$$Y^2 = F^2(S^2) = F^2(W^2 F^1(W^1 X))$$

第三步 训练算法及训练参数的选择

针对不同的应用, BP 网络提供了多种训练、学习方法,通常对于包含数百个权值的函数逼近网络,训练函数 Trainlm 收敛速度最快。将 RPROP 算法的训练函数 Trainrnp 应用于模式识别时,其速度是最快的。使用变梯度算法的训练函数 Traincgf,在网络规模比较大的场合,其性能都很好。

本研究采用了 LM 算法,训练模式选择批变模式,以批变模式来训练网络的函数是 Train; BP 算法中几个常用参数的选取,包括学习率 η,动量因子 a 及收敛误差差值 goal 等,这些参数对训练速度的影响很关键。

第四步 网络训练与预测

把输入样本输入到神经网络,计算网络输出值,然后与实际输出相比较,使用选定的网络训练算法,以一定的规则修改网络的连接权值。反复计算误差和修改权值,直到误差达到一定的范围以内,输入检验样本,判断检验结果;还原处理及结果分析,对样本结果进行还原处理得到实际值,如果训练误差在允许范围内,而且网络泛化能力较好,就可以利用训练好的 BP 神经网络来预测房地产住宅市场的未来发展情况。

3.2 建立 BP 网络预测模型

3.2.1 预测步骤

由于激活函数一般为 Sigmoid 函数,当自变量大于一定值后,函数将趋近于 1 或者 0,随着自变量的增大或减小,函数值将变化不大。如果直接将实际数据作为神经网络的输入参数计算,将会使学习一开始就处于饱和状态,看不出不同样本的差别,无法进行调整^[4-6]。

为了使输入样本对神经元的刺激有可比性,在学习之前,需要对样本数据作预处理,通常是把样本值换成 [0, 1] 或者 [-1, 1] 范围内的数据,这一过程就是前面讲的归一化处理。使 BP 网络的输入输出值限制在区间 [0.02, 0.09] 之间,然后在输入网络进行计算。原始数据见表 2 左边内容,归一化后的数据见表 2 右边内容。

表 2 自贡市 2009 年住宅实际平均每季售价数据及归一后数据表

季度	平均售价	季度	平均售价
1	2642	1	0.02
2	2790	2	0.04987
3	2842.57	3	0.06047
4	2988.89	4	0.09

3.2.2 住宅平均售价的预测与结果分析

(1) 预测网络输入节点的确定

依据第一季度到第四季度获得的住宅实际平均售价数据,拟建一个有 2 个输入节点的三层网络。这 2 个节点依次为历年住宅实际平均售价。

(2) 隐含层的考虑

根据 hornik 提出的公式: $N = [\sqrt{2n + m} \ 2n + m]$, 其中 n 为输入层节点数, m 为输出层节点数。由公式确定隐层节点数的取值范围为 2-5,但是最佳的隐节点数要通过实验测试。本文采用 Matlab7.0 的 BP 神经网络工具箱进行仿真实验,实验中参数取值: goal = 0.1, epochs = 5000, η = 0.6。通过多次实验确定隐含层最佳个数为 3。

(3) 训练及结果分析

训练函数设为 trainlm,学习函数设为 learnngdm,传递函数设为 purelin,训练步数为 5000 步,训练精度为 0.1。训练收敛的极快,多次训练过程中保持网络训练函数、学习函数、传递函数均不变。训练的最佳效果收敛图以及预测图如图 1、图 2 所示。

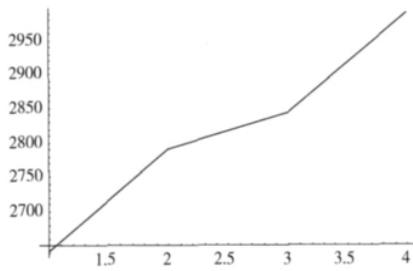


图1 自贡市房屋平均售价

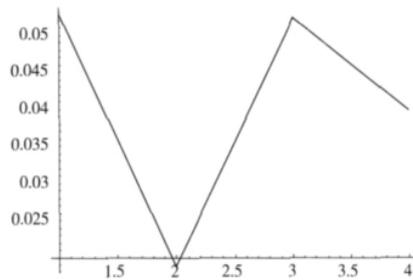


图2 2009年1—4季度的房价变化趋势

训练多个网络后,从图2可以看到能够达到预期的预测效果。当训练好一个网络,该网络能较准确的预测

平均售价。

通过表3可看出,误差都在允许的误差范围内,这样的网络结构获得了较好的拟合效果,预测值和实际值之间的误差也较小。从某种程度上依然可以预测住宅平均售价情况。

表3 真实值与预测值的比较

季度	预测值	真实值	误差	误差率
1	2530.81	2642	111.19	0.042
2	2705.34	2790	84.66	0.030
3	2789.26	2842.57	53.31	0.019
4	2895.08	2988.89	93.81	0.031

3.3 基于影响自贡住宅销售平均价格因素的BP网络预测模型

3.3.1 预测步骤

选用的影响因素指标为:人均居住使用面积 Y1,市区人口总数 Y2,人均可支配收入 Y3,人均消费性支出 Y4,住宅平均售价 Y5,房地产开发投资 Y6,地区生产总值 Y7,贷款利率 Y8,住宅销售面积 Y9,它们的关系如图3所示,具体数据见表4。对表4进行归一化处理,其结果见表5。

表4 原始数据

年份	2002	2003	2004	2005	2006	2007	2008	2009
Y5	1356	1427	1568	1683	1942	2120	2566	2858
Y1	17.93	18.2	18.63	18.78	19.3	19.55	20.1	21.25
Y2	105.32	110.79	112.55	118.79	119.21	120.32	124.35	128.3
Y3	5728.95	5919.84	6553.45	7434.97	8576.48	10244.03	11706.44	12888
Y4	4819.23	5354.29	5981.34	6358.74	6843	7021.56	7956.48	8521.24
Y6	12.48	13.16	15.27	16.82	20.49	26.57	31.82	34.28
Y7	167.82	197.92	221.67	249.35	321.99	373.24	434.82	541.05
Y8	6.21	5.76	5.76	5.76	5.76	5.76	5.49	5.49
Y9	468203	429381	478735	510892	594306	638427	683700	752600

表5 归一化后的数据

年份	2002	2003	2004	2005	2006	2007	2008	2009
Y5	0.1	0.1378	0.2129	0.2742	0.4121	0.5069	0.7445	0.9
Y1	0.1	0.2904	0.3517	0.5689	0.5836	0.6222	0.7625	0.9
Y2	0.1	0.1213	0.1921	0.2906	0.4182	0.6045	0.768	0.9
Y3	0.1	0.2156	0.3511	0.4327	0.5373	0.5759	0.778	0.9
Y4	0.1	0.125	0.2024	0.2593	0.3939	0.6171	0.8097	0.9
Y6	0.1	0.1645	0.2154	0.2748	0.4305	0.5403	0.6723	0.9
Y7	0.9	0.4	0.4	0.4	0.4	0.4	0.1	0.1
Y8	0.1961	0.1	0.2222	0.3017	0.5082	0.6174	0.7295	0.9
Y9	468203	429381	478735	510892	594306	638427	683700	752600

3.3.2 自贡住宅销售平均价格预测与结果分析

(1) 输入层和输出层的设计

在本模型中,取输入数据的维数为8,可以表示为一个8维的向量,即 $Y = (\text{人均居住使用面积}, \text{市区人口总数}, \text{人均可支配收入}, \text{一年贷款利率}, \text{人均消费性支出}, \text{房地产开发投资}, \text{地区生产总值}, \text{住宅销售面积})$,因而网络的输入节点数确定为8,输出层的维数可根据使用者的要求,在本文中输出层的维数为1。

(2) 隐含层层数确定

根据 hornik 提出的公式: $N = [\sqrt{2n + m}, 2n + m]$ 求得。其中, n 为输入层节点数, m 为输出层节点数。根据公式,确定隐层节点数的取值范围为 4 - 17,但是最佳的隐节点数要通过实验测试。本文采用 Matlab7.0 的 BP 神经网络工具箱进行仿真实验,实验中参数取值: $goal = 1e - 4$, $epochs = 500$, $\alpha = 0.9$, $\eta = 0.05$,通过多次实验确定隐含层最佳个数为 12。

(3) 训练及结果分析

训练时训练函数设为 trainlm,学习函数设为 learnf-

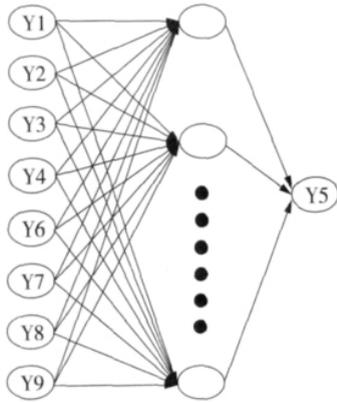


图 3 预测模型网络结构图

度为 0.000 1。2010 年的预测平均误差为 4.85% 2010 年和 2011 年的训练收敛结果和预测如图 4 所示。

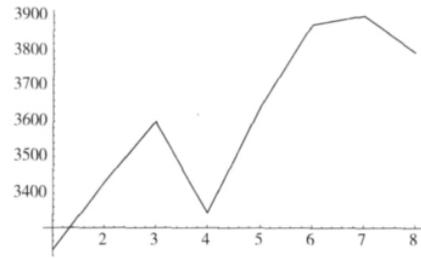


图 4 2010-2011 年 8 个季度房价预测

dm, 传递函数设为 purelin。训练步数为 500 步, 训练精

将预测的样本值作为网络的输入, 并和以前的样本合并再重新训练更新网络权值, 得到 2010 年和 2011 年第一至第四季度的住宅销售平均价格的预测值(表 6)。

表 6 2010-2011 预测值

年度 季度	2010				2011			
预测值 (元/平方米)	1	2	3	4	1	2	3	4
	3234.53	3427.62	3600.48	3342.84	3639.81	3875.59	3903.08	3795.66

5 结束语

自贡市作为中国中等城市, 从 2000 年到 2009 年, 人均可支配收入明显增加, 全市对房地产投资也在大幅度加大。根据 2010 年各季度的房屋销售价格的预测与真实值分析, 预测出在 2011 年的房屋销售价格。从总体来看, 整年的平均价格仍然处于上升趋势^[7-8]。一至二季度快步上升, 二至三季度的增长速度有所缓和, 第四季度的价格处于下降趋势。2011 年自贡市房价的预测结果, 推其主要原因就是国家宏观调控的影响, 使自贡房价的增长速度有所下降。全市的经济发展, 人均收入的增加, 人均居住面积的增多是让房价增长的直接原因。

参考文献:

[1] 李志林, 欧宜贵. 数学建模及典型案例分析 [M]. 北京: 化学工业出版社, 2007.

[2] 陈东彦, 李冬梅, 王树忠. 数学建模 [M]. 北京: 科学出版社, 2007.

[3] 徐瑞, 黄兆东, 阎凤玉. MATLAB2007 科学计算与工程分析 [M]. 北京: 科学出版社, 2008.

[4] 王瑜, 李天增, 付磊等. 关于卷积码编码方案的研究 [J]. 四川理工学院学报: 自然科学版, 2009, 22(3): 35-38.

[5] 雷远明, 刘自山. 弹性动力学方程解的带权范数衰减估计 [J]. 四川理工学院学报: 自然科学版, 2008, 21(6): 17-20.

[6] 李柳芬. 度量空间中距离算子的延拓问题 [J]. 四川理工学院学报: 自然科学版, 2008, 21(6): 21-22.

[7] 蔺富明. 强相依非平稳序列位置和高度联合极限分布 [J]. 四川理工学院学报: 自然科学版, 2008, 21(6): 26-28.

[8] 叶俊, 张先君. 最优决策模型及应用 [J]. 四川理工学院学报: 自然科学版, 2008, 21(6): 23-25.

Trends Forecast of the Zigong House Price Based on the BP Network

DU Dao-yuan, BAI Hong-bin, ZHOU Feng

(School of Science, Sichuan University of Science & Engineering, Zigong 643000, China)

Abstract: In this paper, two kinds of BP network forecast model are given by analyzed the main factors which influenced the Zigong real estate, and combined the actual situation of the Zigong real estate which based on the BP network: one is the trends forecast model that is based on the time series; another is the model of regression forecast model which is based on the influence factors, and forecasts the trends of the Zigong house price. The results of the simulated situation for 2010 proved the effectiveness of our methods in the prediction for the forecast of the Zigong house price for 2011, and this can give a valued direction suggestion for the continued develop of the construction of Zigong city.

Key words: BP network; the forecast of the house price; response function; training function