

基于BP神经网络的房产税税基评估研究

赵愈,许路

(沈阳建筑大学管理学院,辽宁 沈阳 110168)

摘要:2011年,沪渝两市开始对部分居民住房征收房产税,时至今日税改的结果仍未达到预期目标。随着2019年“两会”召开,房产税再次成为全社会关注的热点。沪渝两地房产税改革的经验表明,采取合理的房产税税基评估方法对房产税的征收非常关键。鉴于此,以税基评估和神经网络的基本原理为基础,建立了基于BP神经网络(Back Propagation Neural Network)的房产税税基评估模型,为房产税税基评估研究提供了一种全新的方法,对健全房产税相关配套制度、推动国家税收改革有着重要的意义。

关键词:房产税税基;BP神经网络;批量估计;评估模型

中图分类号:F810.423 **文献标志码:**A

早在2011年1月,上海、重庆试点对个人住房征收房产税,表明我国对住宅房产税的征收已经进入实践阶段。从两会《政府工作报告》可以看出,关于推进房产税立法的描述从2018年的“稳妥”到2019年的“稳步”,虽然只改变了一个字,但表明我国正在逐步落实对个人住房房产税的征收政策。而房产税税基评估作为房产税税制改革的重要配套,是确保房地产课税公平性的前提条件,对促进房产税税收行为规范化、提高政府部门市场监管能力意义重大。从国际经验来看,发达国家或地区大多采用批量评估的方法对需要缴税的房地产价值进行估算,并以此为房产税的计税基础^[1]。相比之下,尽管国内已有部分学者开始将批量评估技术应用于房地产价格评估,但总体而言,我国房产税税基评估研究还不够成熟。因此,就现有研究理论和技术来看,要系统科学地实现房地产税基评估,进而推动房产税的整体改革还

十分困难,积极探索更为有效的评估方法意义重大。鉴于此,笔者借鉴国内外税基评估经验,在税基评估理论和批量评估理论的基础上,利用BP神经网络(Back Propagation Neural Network)构建税基评估模型,以期为我国房产税税基评估的研究提供一种新的方法。

通过文献梳理可以看出,国外学术界基于自动评估模型(Automated Valuation Mode,AVM)进行了大量研究。Robert J Gloudemans^[2]打破了多元线性回归的约束,构建了非线性模型,并将其运用于房地产价格评估;M C William等^[3]提出将地理信息系统(Geographic Information System,GIS)与计算机辅助批量评估相结合,从而提高了批量评估的准确性;R M Kathmann^[4]指出将神经网络技术应用于房地产价格评估,能够大幅提高评估的效率。而在国内,税基批量评估的起步比较滞后,近几年,在国家对房产税

税制改革不断深化的过程中,税基评估制度才受到学术界的广泛关注。纪益成^[5]首次向国内介绍了国际上通用的税基评估理论,并对房地产评估的概念和过程作了详细介绍;陈小悦等^[6]阐述了运用计算机技术进行税基评估的优势,并论述了建立评估系统步骤的构想;耿继进等^[7]在不动产长期趋势法和成本法等基本原理的基础上,以深圳市二手房成交价为计税依据,构建了基于地理信息系统的批量评估模型;刘洪玉等^[8]基于模糊数学方法,提出了房地产分组的思路和应用路径,对科学地提高税基评估准确性具有重要意义。

综上所述,国外一些发达国家的房地产税基评估理论与技术已经比较成熟,相对而言,我国对税基评估的研究仍处于探索阶段。尽管国内一些城市已开始学习国外先进的批

量评估技术,但目前仍无法证明这些理论和方法能否有效用于我国房产税税基评估。因此,对于我国房地产税基评估技术的研究应在借鉴国外成熟经验的前提下,充分结合国内房地产市场的发展现状开展。

一、房产税税基及批量评估理论

1. 房产税税基评估和影响因素

房产税税基是指房屋在课税时能参考的标准或依据。税基评估指具备相应资格的评估机构运用适当的估价方法,客观公正地进行评估价格评估的过程^[9]。就国外来看,房地产保有环节税的税基主要取决于房地产的市场价格,由于房地产本身的特殊性,其价值影响因素较多且不同环境下考虑的因素差别较大。因此,影响税基价值的因素大致可以分为一般因素、区域因素和个别因素(见图 1)。

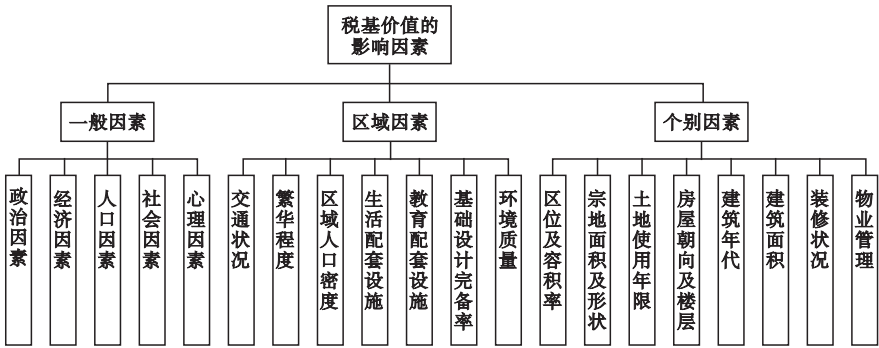


图 1 房产税税基价值的影响因素

2. 批量评估理论

批量评估是指利用共同的数据、标准化的方法和统计检验技术评估一组财产在某一确定时间价值的过程^[10]。批量评估方法有市场比较法、收益法和成本法。鉴于笔者研究对象为个人住宅,主要以市场比较法为评估的基础。

二、BP 神经网络的基本原理

BP 神经网络是以 Rumelhart 和 McClelland 为首的科学团队于 1986 年提出的概念,是目前应用最广泛的神经网络^[11]。该网络由输入层、隐含层和输出层构成,典型的 BP 神经网络拓扑结构如图 2 所示,其模

型处理信息的基本原理如图 3 所示。

现有的数学理论可以证明 3 层的 BP 神经网络能够以任意精度拟合任意复杂度的映射关系^[12],因而笔者构建的房地产税基评估模型网络为 3 层。假设输入层为 X ,有 M 个神经元,其中,第 m 个记为 x_m ;隐含层为 K ,

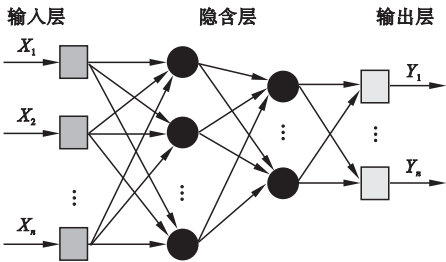


图 2 BP 神经网络拓扑结构

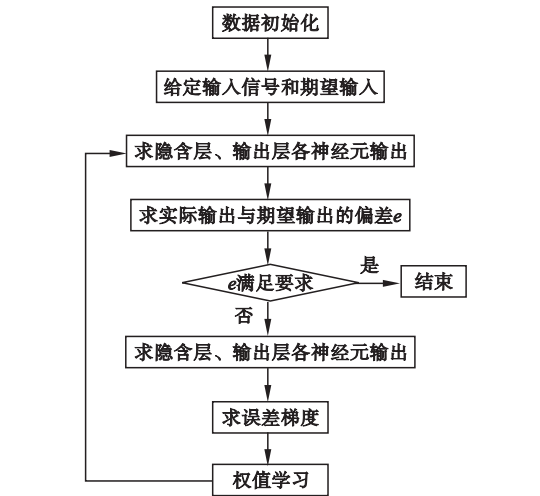


图3 BP神经网络中模型处理信息的基本原理

有 P 个神经元,第 p 个记为 k_p ,激活函数为 f_1 ;输出层为 Y ,有 J 个神经元,第 j 个记为 y_j ,激活函数为 $f_2^{[13]}$ 。隐含层节点输出为

$$k_i(n)=f_1\left(\sum_{m=1}^Mx_{mi}(n)\omega_{mi}(n)+b_i(n)\right)$$

式中: ω 为权值; b 为阈值; n 为迭代次数。

表1 样本指标及其量化标准

指标名称	具体情况	量化标准
房间	直接用房间(不包含厨房和洗手间)数量来表示	房间总数
洗手间	洗手间个数	数量
楼层	住宅层数1/3以内的楼层为低层,1/3~2/3的楼层为中层,2/3以上的楼层为高层	高层为3;中层为2;低层为1
已使用年限	房屋建成到交易的时间	2010—2011年楼龄为1;2012—2013年楼龄为2;2014—2015年楼龄为3;2016—2017年楼龄为4;2018—2019年楼龄为5
朝向	通常朝南为主导,东南、西南为次方向,朝北为最次	朝南为5;朝东南、西南为4;朝东为3;朝西为2;朝北为1
面积	实际面积数/ m^2	180 m^2 及以上为5;141~179 m^2 为4;110~139 m^2 为3;80~109 m^2 为2;50~79 m^2 为1
装修	可按装修费用、材料档次、设计效果、家俱设施齐全程度等分级:豪华装修、精装修、普通装修、毛坯	豪华装修为4;精致装修为3;普通装修为2;毛坯为1
商业	附近(3 km范围内)是否有大型市场或商业区	大型市场或商业区的数量
交通	附近(3 km范围内)是否有地铁口、公交站台	地铁口、公交站台的数量
区域环境	附近(3 km范围内)是否有公园、公共绿地、综合性花园	公园、公共绿地、综合性花园的数量
教育	附近(3 km范围内)是否有知名学校和幼儿园	知名学校、幼儿园的数量

2. 学习参数确定

- (1)输入层神经元个数:表1选取了11个影响税基价格的量化指标,输入层神经元个数为11。
- (2)输出层神经元个数:输出节点选择待估房产单位平方米的价格,所以输出层神经元个数为1。

输出层节点输出模型为

$$y_i(n)=f_2\left(\sum_{i=1}^Pk_{ij}(n)\omega_{ij}(n)\right)$$

误差函数为

$$e_j(n)=\frac{1}{2}\sum_{j=1}^J(t_j(n)-y_j(n))^2$$

式中: t 为期望输出。

如果偏差 e 符合条件要求,计算终止;否则网络进入逆向传播,则其权值修正量为

$$\Delta\omega=-\eta\frac{\partial e(n)}{\partial\omega(n)}$$

式中: η 为学习率。

上述过程就是BP神经网络处理信息的基本原理。

三、基于BP神经网络的税基评估模型

1. 指标参数确定

根据我国房地产市场现状以及国内外税基评估的特征因素在文献中出现的情况,笔者选取了11个高频影响因素为评估指标(见表1)。

(3)隐含层节点数量:隐含层节点的数量对网络模型训练准确性有直接影响,通常情况下确定隐含层节点数的经验公式为

$$r<\sqrt{(s+l)}+a$$

式中: r 为隐含层节点数; s 为输入层节点数; l 为输出层节点数; a 为0~10的常数。则

$$r<\sqrt{(11+1)}+a\approx3+a$$

r 可以取 $[3, 13]$ 的整数, 选取 r 取值中误差最小时对应的节点为最终隐含层节点个数。

(4) 初始权值: BP 神经网络算法对初始权值的选取较为敏感。一般来说, 希望每个初始权值都能出现在 0 点附近, 因此, 将网络初始权值定为 $(0, 1)$ 的随机数。

四、BP 神经网络税基评估模型案例分析

1. 数据来源

我国住宅用房结构类型相似度较高, 因

此, 选取 2019 年 12 月江苏省徐州市 65 套个人住宅为样本, 基于批量评估理论以及 BP 神经网络原理构建房产税税基评估模型。数据样本主要通过查询专业第三方挂牌信息以及咨询相关机构而得。为确保样本的选择具有代表性和全面性, 笔者选取徐州市云龙区、鼓楼区等 6 个主要行政区域, 案例尽可能具有该区域代表性特征, 每个区域选取的成交案例都尽可能相等。根据上文所述量化指标, 在 65 套二手房住宅交易样本中选取 1 ~ 50 套作为训练样本, 51 ~ 65 套作为检验模型可行性的测试样本, 具体样本如表 2 所示。

表 2 具体样本数据表

序号	房间	洗手间	楼层	已使用年限	朝向	建筑面积	装修	商业	交通	区域环境	教育	楼价/(元·m ⁻²)
1	4	1	3	3	5	3	3	3	3	2	2	13 207
2	3	1	3	2	5	2	2	3	3	4	1	10 526
3	5	2	3	2	5	5	3	3	3	2	3	13 698
4	3	1	1	1	5	3	3	4	4	1	4	10 000
5	4	1	2	4	5	4	4	2	3	2	2	15 070
6	3	1	1	1	5	2	2	3	4	3	4	9 005
7	5	2	3	3	5	5	1	2	1	2	2	14 772
8	2	1	1	2	5	1	3	2	3	2	3	10 216
9	4	1	3	5	5	3	1	2	2	2	1	13 011
10	3	1	3	1	5	2	2	3	4	3	4	7 878
11	5	1	3	4	5	4	4	2	2	4	3	11 165
12	3	1	3	1	5	2	3	4	5	2	4	9 701
13	4	1	1	5	5	3	1	4	4	4	3	11 445
14	4	1	2	3	5	5	4	3	5	5	2	13 878
15	7	2	3	1	5	3	4	2	2	4	2	10 668
16	5	1	2	2	3	3	4	4	4	3	2	12 344
17	5	1	2	5	5	3	1	1	1	1	1	7 079
18	5	1	2	5	5	2	1	1	2	2	1	8 364
19	4	1	1	4	5	3	1	2	3	3	2	12 209
20	6	2	3	2	5	3	1	2	2	5	2	12 673
21	4	1	2	3	5	4	4	4	5	1	2	13 954
22	3	1	1	2	5	2	1	3	3	5	2	11 724
23	5	2	3	4	3	4	2	5	5	3	3	16 364
24	1	1	1	2	1	1	3	3	3	1	2	10 028
25	5	1	3	4	3	3	2	2	3	1	2	12 935
26	5	2	2	3	5	5	3	4	4	5	4	25 975
27	5	1	1	2	5	3	2	3	3	2	3	11 194
28	4	1	2	4	5	4	4	4	4	2	2	15 384
29	5	2	1	3	5	5	3	3	4	4	2	13 703
30	5	2	3	3	5	3	1	3	3	2	2	10 495
31	2	1	1	3	5	1	2	3	4	1	5	10 487
32	5	1	3	3	5	3	1	3	4	3	3	11 566
33	3	1	2	3	5	2	1	3	4	3	4	10 562
34	4	1	2	2	5	2	2	3	3	2	3	10 857
35	5	1	2	5	5	3	4	1	1	1	1	8 478
36	5	2	1	3	3	4	4	3	4	3	2	14 231
37	7	3	2	2	5	5	1	2	2	3	2	15 449
38	4	1	3	3	5	3	4	4	2	2	1	14 706
39	8	3	3	4	5	3	4	2	2	4	1	9 035
40	5	1	2	2	5	4	4	2	2	4	2	13 673
41	2	1	2	3	1	2	2	3	4	2	3	8 587
42	7	3	2	2	5	5	1	2	2	3	2	15 449

续表

序号	房间	洗手间	楼层	已使用年限	朝向	建筑面积	装修	商业	交通	区域环境	教育	楼价/(元·m ⁻²)
43	5	1	2	1	3	3	3	4	3	3	3	12 245
44	4	1	3	3	5	3	2	2	3	1	2	10 392
45	5	1	3	3	5	5	3	3	4	3	2	14 727
46	2	1	1	2	3	1	2	2	3	2	3	11 478
47	5	2	3	2	3	5	4	4	4	5	4	23 211
48	5	1	1	2	3	3	4	1	1	4	1	12 829
49	4	1	3	1	5	4	3	3	3	4	4	15 544
50	5	1	2	2	5	5	3	4	4	5	3	12 495
51	3	1	2	3	5	2	2	2	4	3	2	9 077
52	5	1	2	1	5	5	3	4	3	3	2	13 511
53	4	2	1	3	5	3	3	4	4	3	3	13 346
54	5	1	1	3	5	3	3	2	2	4	2	12 889
55	4	1	3	3	5	5	4	5	4	4	3	12 144
56	5	1	1	2	5	4	2	4	4	3	5	11 314
57	5	2	2	4	5	4	1	2	2	5	2	8 970
58	5	1	3	2	5	4	2	5	5	4	4	12 680
59	6	2	1	2	5	5	2	5	5	4	4	11 790
60	2	1	1	2	4	2	3	5	5	4	4	12 715
61	6	2	3	1	5	4	4	3	3	4	4	10 903
62	4	1	1	4	3	3	4	4	4	4	5	11 147
63	4	2	2	2	5	4	3	4	4	4	5	11 694
64	5	2	3	2	5	5	4	4	4	4	3	12 152
65	4	1	3	4	2	5	4	5	5	3	5	16 829

2. 数据处理

利用 Matlab 软件进行程序的编写与调试,运用 `traibfg` 函数构建税基评估模型,进而实现模型的训练和仿真。根据具体样本数据能够确定神经网络模型的输入层节点为 11 个、输出层节点为 1 个。为消除数量级差异对预测精度造成的影响,笔者利用 `mapminmax` 函数对输入和输出数据做归一化处理。

3. 训练过程和仿真模拟

通过 BP 神经网络对训练样本(序号 1~50)进行迭代,训练过程与结果如图 4~图 7 所示。

由图 4 中可知,当隐含层节点为 11 时,均方误差(mse)初始值为 1.49,结束值为 $9.68e^{-07}$,网络耗时 3 s,经过 368 次训练后,模型精度达到预设值 $1.00e^{-06}$ 。由图 5 可知,均方误差的收敛效果可以达到预期目标值。

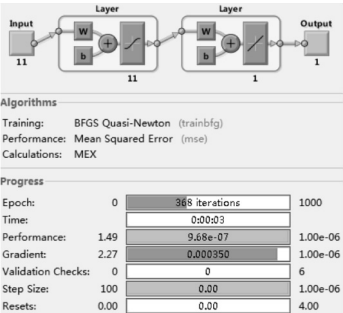


图 4 BP 神经网络训练过程与结果(训练样本)

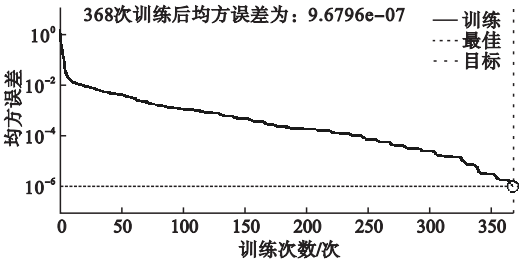


图 5 均方误差图(训练样本)

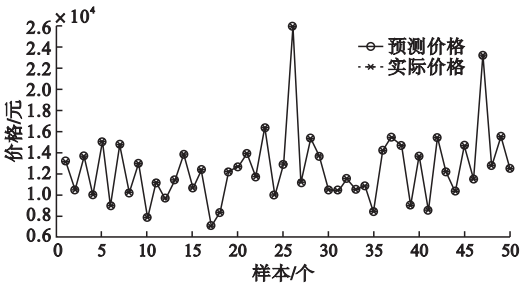


图 6 预测价格与实际价格对比(训练样本)

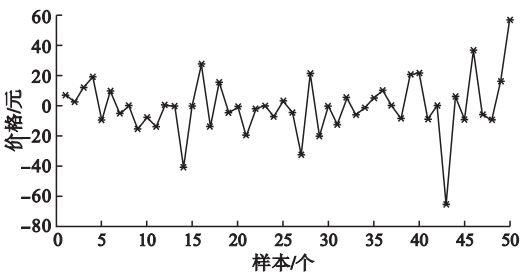


图 7 预测价格与实际价格的误差(训练样本)

图 6 中节点为“○”的线表示神经网络预测出的价格,节点为“×”的线表示房地产实际价格。可以看出,数据经过 368 次训练后可以达到预先设定的目标值。由图 7 可知,经过训练后,得到的预测值和实际值能够拟合,且误差控制在 100 元以内。图 4~图 7 证明了 BP 神经网络税基评估模型网络训练成功,训练结果实现了合理的线性拟合。此外,选取表 2 数据(序号 51~65)进行神经网络准确性的测试,结果如图 8、图 9 所示。

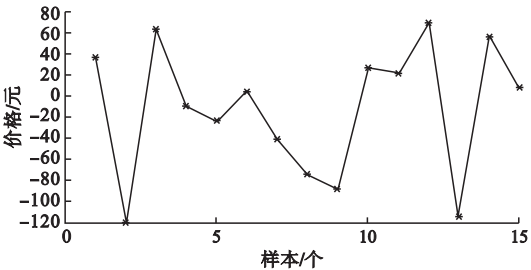


图 8 预测价格与实际价格的误差(测试样本)

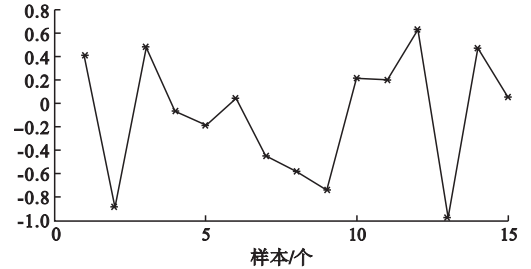


图 9 误差百分比(测试样本)

由图 8、图 9 可知,虽然 15 组测试样本经过网络训练后有误差产生,但误差不超过 $\pm 1\%$,这表明训练后的 BP 神经网络模拟房地产价格与其影响因素之间关系时性能突出,同时,预测输出与期望输出比较接近也说明了 BP 网络具有良好的函数逼近功能。

4. 结果分析

将测试结果用 mapminmax 函数进行反归一化处理,用得到的数据与期望价格进行比对,并进行误差分析(见表 3)。

通过上述实验能够看出,经过训练的 BP 神经网络可以较好地模拟房产税税基与其影响因素之间的关系,这意味着只要样本数量足够多,并且训练函数选取得当,BP 神经网络模型就能很好地对房产税税基进行评估。研究证明 BP 神经网络在税基评估方面性能

表 3 预测价格与实际价格误差表(测试样本)

序号	预测价格/ (元·m ⁻²)	实际价格/ (元·m ⁻²)	误差/ (元·m ⁻²)	相对误差/ %
51	9 114. 314	9 077	37. 314	0. 411
52	13 392. 082	13 511	-118. 918	-0. 880
53	13 410. 266	13 346	64. 266	0. 482
54	12 879. 942	12 889	-9. 058	-0. 070
55	12 121. 115	12 144	-22. 885	-0. 188
56	11 318. 483	11 314	4. 483	0. 040
57	8 929. 916	8 970	-40. 084	-0. 447
58	12 606. 886	12 680	-73. 114	-0. 577
59	11 702. 877	11 790	-87. 123	-0. 739
60	12 742. 108	12 715	27. 108	0. 213
61	10 924. 475	10 903	21. 475	0. 197
62	11 216. 996	11 147	69. 996	0. 628
63	11 580. 417	11 694	-113. 583	-0. 971
64	12 209. 102	12 152	57. 102	0. 470
65	16 837. 911	16 829	8. 911	0. 053

优越,能够满足低成本、高效率的批量评估要求。

五、结 语

传统的多元线性回归模型都是基于线性假设的,而影响房产税税基的因素有很多,涉及的特征变量也比较多。笔者重点研究了基于 BP 神经网络进行房产税税基评估的可行性,并对现有结论展开了进一步研究。目前,我国正处于推进房产税立法的重要阶段,研究可以为房产税征收提供理论和实践指导,从而能够在一定程度上推动我国顺利征收房产税。笔者应用 BP 神经网络对税基进行了评估,期望输出和实际输出的误差范围基本控制在 $\pm 4\%$ 以内,模型训练结果较好,为房产税税基评估研究提供了一种全新的方法和工具。BP 神经网络具有一定的自学习能力,税基评估模型的预测过程并不是一成不变的。在实际评估过程中,随着经验数据不断改善和完备,税基评估模型的训练能力也可以不断加强,从而在税基评估过程中进一步提高预测精度,逐步得到更加完善的评估系统。

参考文献:

[1] 纪益成,王诚军,傅传锐. 国外 AVM 技术在批量评估中的应用[J]. 中国资产评估,2006 (3):13-18.

[2] ROBERT J G. Comparison of three residential

regression models: additive, multiplicative and nonlinear[J]. Assessment journal, 2002(21): 31-35.

[3] WILLIAM M, WILLIAM D, ADAM M. Development of a geography information systems for mass appraisal of residential property[J]. Journal of property valuation & investment, 1997(15): 448-465.

[4] KATHMANN R M. Neural networks for the mass appraisal of real estate [J]. Comput environ and urban systems, 1993(17): 373-384.

[5] 纪益成. 税基评估有关问题研究[J]. 国有资产研究, 1999(6): 26-28.

[6] 陈小悦, 孙力强. 关于建立中国房地产税批量评估系统的几点思考[J]. 财政研究, 2007(12): 48-51.

[7] 耿继进, 何素芳. 房地产计税价格批量评估实证研究[J]. 地理空间信息, 2011, 9(3): 27-31.

[8] 刘洪玉, 李妍. 基于模糊数学的房地产批量评估[J]. 清华大学学报(自然科学版), 2017, 57(11): 1202-1206.

[9] 纪益成, 傅传锐. 批量评估: 从价税的税基评估方法[J]. 中国资产评估, 2005(11): 5-9.

[10] 田珊. 批量评估技术在房产税税基评估中的应用研究[J]. 劳动保障世界(理论版), 2013(7): 76-77.

[11] 蔡宛睿, 夏虹, 杨波. 基于 BP 神经网络的堆芯三维功率重构方法研究[J]. 原子能科学技术, 2018, 52(12): 2130-2135.

[12] 吴翼凡, 刘清倦, 杨杰, 等. 基于 BP 神经网络的温度传感器辐射误差修正[J]. 现代电子技术, 2018, 41(24): 43-47.

[13] 杨继星, 余笑梅, 黄玉钊, 等. 基于 BP 神经网络的苯储罐泄漏事故风险评价模型研究[J]. 中国安全生产科学技术, 2019, 15(1): 157-162.

Research on BP Neural Networks for the Tax Base Assessment

ZHAO Yu, XU Lu
(School of Management, Shenyang Jianzhu University, Shenyang 110168, China)

Abstract: In 2011, Shanghai and Chongqing began to levy real estate tax on some residents' houses. Up to now, the result of tax reform has not reached the expected goal. With the "two sessions" held in 2019, the real estate tax has once again become a hot topic of the whole society. The experience of real estate tax reform in Shanghai and Chongqing shows that it is very important to adopt a reasonable tax base assessment method for real estate tax collection. In view of this, based on the basic principles of tax base assessment and neural network, the author establishes a tax base assessment model of real estate based on BP neural network, which provides a new method for the study of tax base assessment of real estate, and is of great significance for improving the relevant supporting system of real estate tax and promoting the reform of national tax revenue.

Key words: the tax base of property; BP neural network; mass appraisal; evaluation model
(责任编辑: 郝雪 英文审校: 林昊)