

基于多层神经网络的房地产价格批量评估研究

——以深圳市为例

■文 | 熊英楠 种晓丽 赵 静

摘要：房地产价格评估在市场交易、抵押、租赁和财产税征收等领域都起着重要的作用。随着中国房地产市场的不断发展，房地产评估技术也越来越受到关注，批量评估技术也需与最新技术相结合以使评估结果更加准确合理。本研究针对当前商品住房价格评估研究领域中对数据特征挖掘及评估过程需要人为干预的不足，使用改进的神经网络算法对商品住房市场价格评估问题展开系统性研究，并实现端到端的商品房价格评估。实验结果表明，对比多元回归、特征价值法等方法，改进的神经网络模型的评估价格与市场价格更为贴近，整体误差比例更小。

关键词：房地产市场 批量评估 改进神经网络

1 引言

1.1 研究背景

当前，住房问题成为当今社会非常重要的问题之一。在房屋交易中，二手房因配套较为完善以及价格更为低廉的优势，交易中受到人们的青睐。在现实生活中房地产发生转让、租赁、抵押、税收、征收、征用、司法拍卖、分家析产、损害赔偿、保险等活动对房地产估价均有需要。因此作为房地产产业附属业务，房地产估价行业正在快速发展与完善。

发达国家城市化迅速发展，带动了房地产交易市场的繁荣，同时也推动了房地产估价行业的高速发展。我国目前对房地产价格评估的方法众多，评估过程差异较大，因此部分评估结果可信度不是很高。并且使用传统的房地产估价方法进行房地产批量评估，所消耗的财力、物力、人力巨大，更重要的是完成价格的评估效率低，创新型的房地产评估方法如基于模糊实物期权法在房地产价格评估，也取得了较高的评估准确性。但是上述方法完成工作所需要的时间跨度比较长，采用传统的房地产估价方法进行大批量房地产估价研究工作，显然满足不了市场需求。此外现今市场上房地产成交数据比较分散，政府、机构和中介等多方之间各自掌握的数据信息不能很好地共享，已经发生的交易的房地产数据的使用效率很低，造成数据的浪费（李妍，2017）。利用计算机的处理能力将这些房地产交易数据进行储存建立完善的成交数

据系统，利用数据分析功能对数据进行分析整理是未来的评估趋势，研究者使用大数据技术进行评估，得到二手房价格的准确性方面也在不断提升，逐渐获得了更多的市场认可。

现实中的房地产数据存在诸多随机干扰因素，数据分布不明确，数据中蕴含的规律可能随时间变化，因此使用单一的预测技术容易导致不稳定的模型泛化能力，即我们难以训练出一个稳定的且在各个方面表现都较好的模型。

集成学习策略能够有效解决这一问题。集成学习是一种能在各种机器学习任务上提高准确率的强有力技术，其使用一系列相同或相异的模型进行学习，并使用某种规则把各个模型的输出结果进行整合从而获得比单个模型更好的效果。大多数情况下，集成学习能够提高学习系统的泛化能力。近几年国内外的数据挖掘比赛如 Kaggle 和天池大数据比赛中，优秀的解决方案都不约而同地使用了集成学习。目前集成学习多用于离散分类问题研究中，在预测问题上的研究和应用还相对较少。

基于以上背景，本研究拟以神经网络相关算法为基础，力图构造新型、实用且高效的神经网络集成评估技术框架，研究其必要性和优缺点，并用以解决现实中的房地产批量评估问题。本研究对神经网络具有一定的理论意义，在评估领域具有较强的实用价值。



1.2 国内外研究现状

批量评估方法是一种相对于个案评估的技术方法，它是在传统评估理论的基础上，融入了统计学、地理信息系统、数据库技术等多方面现代科学技术而形成的适应时代发展的产物。

批量评估方法是一种相对于个案评估的技术方法，它是在传统评估理论的基础上，融入了统计学、地理信息系统、数据库技术等多方面现代科学技术而形成的适应时代发展的产物。事实上，批量评估技术在国外已有多年的研究历史，其思想最早可以追溯到 1919 年，部分西方学者将统计学中的多元回归和房地产批量评估相结合。1966 年，Lancaster 提出了基于多元回归的特征价格模型 (Hedonic Price Model, HPM)，这也是至今被使用最广泛的模型。特征价格模型的思想是通过匹配与房地产价格相关的房地产特征因素，建立起房地产价格与这些特征因素的映射关系，并根据这种拟合出来的关系对待评估房地产的价格进行估计。Rosen(1974) 将特征价格模型首次用于房地产市场的分析，特征价格模型开始被广泛地应用于城市经济学与房地产领域。

20 世纪 90 年代，随着计算机技术逐步替代人工计算，批量评估应用迅猛发展，国外学者也开始对批量评估进行了大量的理论研究和实证分析。1977 年，Carbone 和 Longini 对现代批量评估理论进行完善，首次在《反馈模型在房地产自动评估中的应用》中提出自动评估的概念，即通过运用统一的数理统计模型，进行计算机的自动评估，减少人为因素的影响，不仅提高了评估过程的科学性和评估结果的准确性，也促进了税收的公平，为后续评估系统的建立及大量评估模型的探索奠定了基础。这个概念的提出为后续评估系统的建立及大量评估模型的探索奠定了基础。Gonzalez 和 Laureano-Ortiz (1992) 认为房地产评估仅有价格一个解释变量，在批量评估中全面考虑各种解释变量，可以得出合理的评估结果，并通过案例成功验证了批量评估结果的可行性。Benjamin 等

人(2004)系统地介绍了多元回归模型如何在房地产批量评估中应用,并以住宅房地产为例进行了实践操作。同年,d'Amato将粗糙集理论(Rough Set Theory)应用到房地产批量评估中,解决多元回归模型在市场数据数不足时的数据缺陷问题。为了解决位于两个市场分区交界线上的房产的估价问题,Gonzalez 和 Formoso(2006)又将模糊数学理论(Genetic Fuzzy Rule-based System)引入到房地产批量评估当中,使得传统的多元回归模型得到了进一步的丰富和完善。

随着信息技术的发展,人们逐渐开始研究地理信息系统(GIS)的房地产评估技术。Castle(1995)研究表明, GIS能够获取全面的空间区位信息,包括二维和三维图形的显示、空间区位变化的快速提取、可视化分析结果的输出,并且能够降低评估成本。Wyatt(1997)指出房地产价格是区位因素、自然因素和经济因素等的函数,传统的房地产评估方法依靠评估人员经验获取区位因素数据,主观性太强,而基于GIS技术对区位因素进行空间测度,可提高评估的客观性。鉴于GIS优势,Albert 和 Golledge(1999)将批量评估与GIS技术结合,为后来GIS技术在房地产批量评估中的广泛应用奠定了技术基础。同年, Figueiroa 借助GIS技术,建立批量评估模型,采用实例验证了GIS在辅助估计空间位置的优势。2014年,Zrobek 等人采用GIS的空间分析方法,探讨通过建立地价图辅助实施批量评估的技术路径。

随着机器学习的发展,研究人员将机器学习中的回归方法应用于房价批量评估问题。Torres-Pruñonosa 等(2021)比较了人工神经网络(Artificial Neural Network, ANN),分位数回归和半对数回归在房价评估中的作用。事实证明,在房地产质量评估中, ANN 优于分位数回归和半对数回归方法。Xu 和 Li(2021)以多种机器学习模型为基础,构建了基于双层堆叠框架的二手房评估模型。模型训练结果表明,与线性多元回归和空间计量经济学模型相比,机器学习模型的准确性显著提高,并且堆叠模型的性能优于独立机器学习模型。Hong 等(2020)比较了Hedonic 和随机森林(Random Forest, RF)方法,结果表明,随机森林可以拟合房价和属性之间的非线性关系,这是对Hedonic 方法的有效补充。相较于传统回归方法,机器学习模型在房价批量评估任务上的优势已经得到证明。

基于我国房地产的具体国情,耿继进等人在2012年创

新性的提出了房地产整体估价模型概念,并以深圳市住宅为例,开展了城市内住宅类房地产的批量评估。与前者采用深圳住房市场作为研究对象不同,唐勇等(2012)以深圳市商业地产作为研究对象,将适应估价技术作为理论基础,初步探讨城市商业地产的批量评估。除此之外,部分专家学者还对房地产批量评估在房地产抵押估价中的应用(刘海虹,2014),房地产批量评估中的市场细分等方面,进行了具有积极意义的技术探索。王吓忠(2015)根据房产特征属性的重要性,以筛选后的特征变量为基础,建立特征价格模型进行实证分析,研究发展特征价格理论框架下的评估模型效果较好。刘洪玉和李妍(2017)通过引入模糊数学方法,采用贴进度方法测度不同房地产之间的相似程度,相似的房产划归同组,从而提出房地产分组的思路和应用路径。

从国内评估研究现状来看,不仅有学者通过引入聚类分析法、特征价格理论、模糊数学理论等技术手段来改进评估方法,还有学者引入其他技术来改进评估方法。近年来,国内陆续有学者将GIS技术引入到评估中进行扩展研究。李伟(2009)指出房地产空间特性决定了房地产评估实质上是空间数据的基础上进行的空间数据挖掘、分析决策的过程。因此,在房地产评估中,选择GIS与空间数据挖掘技术的集成来实现房地产评估所需空间数据的相关分析,可以提高评估结果的有效性。2010年,王秀丽和李恒凯在《基于GIS和特征价格的住宅房产价格评估》中提出GIS技术在批量评估住宅类商品时,具备很强实用性, GIS技术具有重要的应用推广价值。同样借助GIS技术,王倩和潘郁(2010)建立空间维度、时间维度等多维度立体评估系统,从而提高了评估效率。宁



新稳(2016)将铁路线路、房屋、土地等信息进行整合,创建房地产信息数据库,使得房地产的管理可以实现三维可视化。王炜昱(2020)通过分析目前房地产评估实务中存在的主要技术问题,进而引入基于大数据的地理信息系统的优势,具体阐述基于大数据的地理信息系统如何解决房产评估中存在的技术问题。研究发现,在房地产评估行业中融入地理信息系统的应用,不仅可以解决因获取地理信息而耗费大量资源的问题,同时能够提升房地产评估行业的工作效率和质量。因此,地理信息系统在我国房地产评估中的应用具有极大的潜力和价值,对于房地产行业的发展具有一定的促进作用。

1.3 研究意义

传统的二手房价格个案评估程序复杂且劳动强度大,通常不可能批量评估二手住宅房的价格,基于节省人员耗费和提高估价效率的重要性,本研究以深圳的二手房估价为例,参考国内外文献,采用新的二手房估价方法,对二手住宅房地进行估价。总体说来,新方法是一种实用、方便、高效和准确的方法,可以用来评估二手住宅房地产的价格。

2 模型建立与数据处理

2.1 Non-linear Deep Neural Network(NLDNN) 模型建立

多元线性回归模型的回归方程式为:

$$y = \sum_{i=1}^m \theta_i x_i \quad (1)$$

而其最终目标在于最小化以下损失函数,基本思想为最小二乘法:

$$\min \left(\frac{1}{2m} \sum_{i=1}^m (\theta^T x^{(i)} - y^{(i)})^2 \right) \quad (2)$$

生活中,很多现象之间的关系往往不是线性关系。在房地产批量评估问题中,房地产属性涉及“地一楼一房”三个维度,使用线性回归模型会造成属性层次在同一纬度,无法得到精确的结果。使用神经网络模型可更好针对不同属性维度进行分析。因此,在构建比价关系矩阵时应尽量使用神经网络。本研究中,神经网络的公式表达如下:

$$y = f_3(f_2(f_1(x_1, x_2, \dots, x_{k_1}), \dots, x_{k_2}), x_{k_2+1}, \dots, x_m) \quad (3)$$

$$f_j(x) = \sum_{i=1}^{k_1} \omega_i x_i \quad j = 1, 2, 3 \quad (4)$$

其中, $f_j(\cdot)$ 代表每个层级的回归函数。详细来说, $f_1(\cdot)$

表示土地/小区层级属性对价格影响的回归, $f_2(\cdot)$ 表示楼栋层级, $f_3(\cdot)$ 表示房屋层级,由三个层级的回归构成的非线性函数可良好地反映批量评估中对于三个维度的价格影响,具体网络模型如图 1。

借鉴 ElasticNet 思想,其损失函数为:

$$\min \left(\frac{1}{2k} \left(\sum_{i=1}^k f_3 \left(f_2 \left(f_1 \left(\mathbf{X}^{(i)} \right) \right) \right) - y^{(i)} \right)^2 + \lambda_1 \sum_{j=1}^n \|\theta\|_1 + \lambda_2 \sum_{j=1}^n \|\theta\|_2^2 \right) \quad (5)$$

其中,第一项代表正常的损失函数,第二项和第三项表示 L1 和 L2 正则化,使模型变得更加稀疏的情况下参数可以更趋向于 0,提高泛化能力。

2.2 数据选择与处理

2.2.1 数据选择

本研究数据来源于深圳市房地产和城市建设发展研究中心。中心作为深圳市住房和建设局直属机构,在住房与房地产研究、房地产评估、地质环境、立法研究等领域承担了大量国家、省、市级研究项目和课题,能够为课题的开展提供必要的研究设备和前期调研数据,遥感数据等,为本项目的顺利实施提供良好的数据基础。

本节所讨论的房地产价格的影响因素可以划分为小区属性、小区内楼栋属性和楼栋内房屋属性三大部分,以下归纳房地产价格的各级各类因素,是开展房地产价格评估的重要基础。具体因素见图 2。

2.2.2 数据处理

数据预处理包括数据清洗、特征组合及数据归一化。

(1) 数据清洗。计算单“商品房单价”(“总价格”除以“建筑面积”)并排除重复值、登记错误信息等记录。将“商品房单价”聚合入房屋属性层面,作为该属性值的因变量。共记录从 2019 年—2022 年共 10 万条小区层面数据。

(2) 由于时间跨度较大,通过 70 指数、各行政区指数及循环神经网络将案例价格修正至相似时间节点。

(3) 将属性数据通过公式(6)归一化,消除不同量纲的影响,同时提高模型预测精度。

$$X = \frac{X - E(X)}{\sqrt{D(X)}} \quad (6)$$

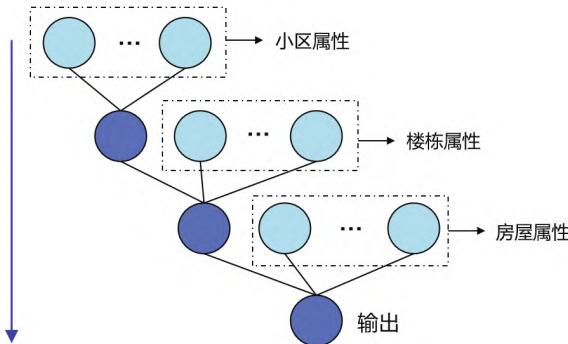


图 1: 神经网络模型正向传播图

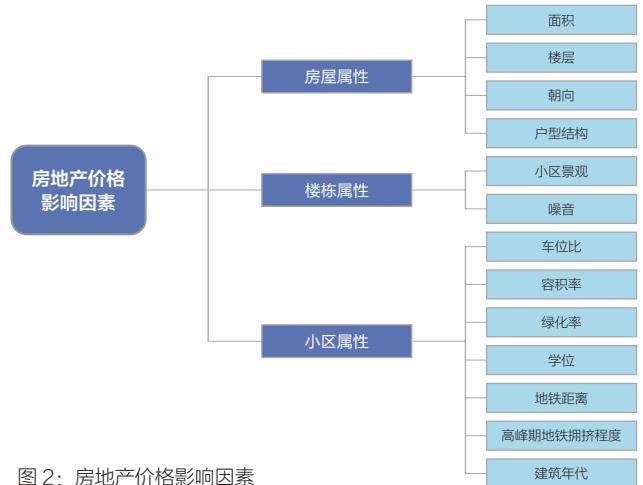


图 2: 房地产价格影响因素

其中, $E(X)$ 代表属性 X 的期望, $D(X)$ 表示属性 X 的方差。

3 基于 NLDNN 的实验参数及结果

3.1 实验参数

上述模型中, 每一层函数的神经元个数等于每层属性个数, 隐藏层的激活函数为 ReLU 函数, 权值优化采用 Adam 算法, 每次迭代优化步长为 10^{-3} 。实验采用十字交叉验证法, 即随机选择 80% 数据作为训练集, 剩下 20% 为预测集, 通过 10 次实验得出平均误差, 误差计算使用平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 作为度量标准, MAPE 公式如下:

$$MAPE = \frac{\sum |y_i - y_{i,pred}| / y_i}{N} \quad (7)$$

其中, y_i 表示预测集真实结果, $y_{i,pred}$ 表示预测结果, N 表示个数。

3.2 实验结果

本节选取线性回归、Lasso、岭回归及本文模型进行结果对比, 综合结果如表 1 所示。

表 1: 实验测试集综合误差

指标	NLDNN	线性回归	Lasso	岭回归
MAPE(%)	4.65	9.36	7.13	7.06

结果表明, NLDNN 模型房价预测综合误差为 4.65%, 相较于其他三种模型的 9.36%, 7.13% 和 7.06% 误差明显降低, 说明模型拟合程度较高, 能较好匹配所使用的模型。对比部分小区的预测集误差, 结果如表 2 所示。

表 2: 部分小区结果误差表

项目名称 / 算法 MAPE(%)	NLDNN	线性回归	Lasso	岭回归
民 * 厦	5.71	7.57	12.05	7.95
国 * 园	8.03	10.22	6.38	10.91
景 * 苑	6.52	12.91	9.27	9.30
宝 * 厦	4.12	11.14	6.08	5.72
幸 * 园	9.26	17.18	15.20	7.55
锦 * 阁	6.26	11.76	6.60	8.78
景 * 厦	2.08	5.01	5.15	3.36
深 * 苑	9.36	10.39	11.64	15.46
天 * 居	5.53	6.99	13.33	8.85
华 * 园	3.40	10.24	1.64	1.83
彩 * 园	7.88	9.81	15.73	9.07

上述结果表明，模型的单个小区误差在绝大多数情况下低于其他对比算法，说明本文设计的模型不仅在全局效果较好，对于单个小区也有良好的拟合效果，具有较好的泛化性。

4 总结

为了提高房地产市场价格预测的准确性，为市场交易、抵押、租赁和财产税征收等领域提供帮助，本文提出了一种基于“地一楼一房”属性的三层神经网络模型——NLDNN。传统线性回归模型对于同类别属性的计算有较好的拟合效果，但对于有层次的多类别属性和非线性属性拟合效果较差，导致预测准确性降低。本文提出的 NLDNN 模型可对三层属性分别建立属性层，使得不同类别属性可以分别计算，并在对应层中联合计算，使得房地产价格预测具有较好的可解释性。实验以深圳市二手住宅小区为对象进行实证检验，对比线性回归、Lasso 和岭回归模型，结果表明，NLDNN 模型预测精准度在总体上高于任何一种对比模型，且在单个小区的检验中也具有较好的精度，证明该模型在房价预测中具有较强的优势。

参考文献

- [1]Albert W S, Golledge R G. The Use of Spatial Cognitive Abilities in Geographical Information Systems: The Map Overlay Operation[J]. *Transactions in GIS*, 1999, 3(1): 7-21.
- [2]Benjamin J, Guttery R, Sirmans C F. Mass Appraisal: An Introduction to Multiple Regression Analysis for Real Estate Valuation[J]. *Journal of Real Estate Practice and Education*, 2004, 7(1): 65-77.
- [3]Castle G. Geographic Information Systems (GIS): The Location[J]. *Location, Location Technology, Real Estate Issues*, 1995, 20: 2-52.
- [4]Gonzalez A J, Laureano-Ortiz R. A Case-based Reasoning Approach to Real Estate Property Appraisal[J]. *Expert Systems with Applications*, 1992, 4(2): 229-246.
- [5]González M A S, Formoso C T. Mass Appraisal with Genetic Fuzzy Rule-based Systems[J]. *Property Management*, 2006.
- [6]Hong J, Choi H and Kim W. A House Price Valuation Based on the Random Forest Approach: The Mass appraisal of Residential Property in South Korea [J]. *International Journal of Strategic Property Management*, 2020, 24(3): 140-152.
- [7]Rosen S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition[J]. *Journal of Political Economy*, 1974, 82(1): 34-55.
- [8]Torres-Pruñonosa J, García-Estevez P and Prado-Román C. Artificial Neural Network, Quantile and Semi-log Regression Modelling of Mass Appraisal in Housing [J]. *Mathematics*, 2021, 9(7): 783.
- [9]Xu L and Li Z. A New Appraisal Model of Second-hand Housing Prices in China's First-Tier Cities Based on Machine Learning Algorithms[J]. *Computational Economics*, 2021, 57(2): 617-637.
- [10]李伟. 基于空间数据挖掘技术的房地产估价系统设计与研究 [J]. *现代测绘*, 2009, 32(06): 28-29.
- [11]刘海虹. 房地产抵押估价中批量评估法的应用研究 [D]. 首都经济贸易大学, 2014.
- [12]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.
- [13]宁新稳. 基于 3D GIS 的铁路房地产信息管理系统研究与开发 [J]. *铁路计算机应用*, 2016, 25(12): 1-3+7.
- [14]唐勇, 耿继进, 朱葵花. 基于适应估计技术的存量住房价格批量评估模型——以深圳市为例 [C]// 中国房地产估价与经纪 2012 年第 1 期 (总第 92 期). 2012: 12-16.
- [15]王倩, 潘郁. 基于 GIS 的房地产智能评估系统研究 [J]. *中国管理信息化*, 2010, 13(09): 75-78.
- [16]王炜呈. 基于大数据的地理信息系统在我国房地产评估中的应用潜力 [J]. *中国资产评估*, 2020(10): 51-56.
- [17]王吓忠, 邱岳. 基于粗糙集的房地产税基批量评估应用 [J]. *福州大学学报 (哲学社会科学版)*, 2015, 29(06): 26-34.
- [18]王秀丽, 李恒凯. 基于 GIS 和特征价格的住宅房产价格评估 [J]. *城市勘测*, 2010(05): 57-60+64.

作者 / 熊英楠

博士后 研究方向为房地产价格批量评估
单位中国科学院深圳先进技术研究院 /
深圳市房地产和城市建设发展研究中心

作者 / 种晓丽

高级工程师
研究方向为房地产评估
单位深圳市房地产和城市建设发展研究中心

作者 / 赵静

高级工程师
研究方向为房地产评估
单位深圳市房地产和城市建设发展研究中心