

机器学习在房地产批量评估中的应用研究综述

陈超 马鑫晟

摘要:房地产批量评估新方法不断涌现,以新的视角对机器学习在国内外房地产批量评估中的应用进行综述,发现已有的研究成果总体上缺乏系统性,各位学者对不同类型批量评估方法的可靠性莫衷一是,针对我国本土的房地产批量评估方法创新相对较少。未来的研究工作可重点在智能化批量评估系统设计、房地产批量评估分区升级和房地产批量评估方法应用创新等方面寻求突破。梳理与分析研究文献可以为同行开展进一步研究提供参考,助力建立具有中国特色的房地产批量评估方法体系。

关键词:房地产;批量评估;机器学习;研究现状

一、引言

国际税收评估人员协会和国际评估准则在对批量评估的定义中都强调了要利用统计检验方法确定财产价值。批量评估方法凭借速度快、成本低、精度高的优势被大多数发达国家应用于税基评估、抵押贷款、拆迁补偿等事项中,且仍有大量不同领域的专家学者正在直接或间接地为批量评估方法改进做着贡献。各国房地产市场在完善程度、法律制度、税制改革等方面都存在较大的区别,我国房地产评估应充分借鉴成熟经验,但又要特别重视发展具有中国特色的房地产评估体系。

鉴于此,本文梳理了近十年关于房地产批量评估方法的国内外研究成果,以房地产评估过程为主线,试图让读者对评估方法体系有个初步的认识,起到提纲挈领的作用。更重要的是通过厘清房地产批量评估方法体系,探索适合中国的房地产批量评估技术并指出未来房地产批量评估的发展方向,为科研工作者进一步开展房地产评估研究提供一些便利,也为实际评估工作提供一些参考。同时房地产批量评估这一技术性难题的解决,不仅有利于大幅提高房地产评估工作效率,还有助于推动房地产税改革,提升征税信息化水平,强化征税公平性。

二、房地产批量评估的前期准备

(一) 样本选取

房地产批量评估是一种计算机辅助批量评估技术,高质量、完备和翔实的房地产特征信息和交易数据是采用计算机辅助批量评估的前提,如果没有完整、可靠、可信任的数据,再好的评估方法也得不到满意的结果。耿继进和张晖(2012)曾提出并证明利用地理信息系统(geographic information systems, GIS)构建我国房地产批量评估基础数据库的可行性,该方法能有效解决不同类别数据的集成问题,提高数据

作者简介 陈超:辽宁工业大学,讲师;

马鑫晟:辽宁工业大学,硕士研究生。

基金项目:辽宁省教育厅青年项目(LJKQR2021067)。

应用的效率。不过对于房地产基础信息数据的采集和集成不是本文综述的重点,我们更关心在批量评估过程中如何选出高度可比性样本以及如何处理这些样本。

不同类型的房地产价格影响因素和评估方法差异巨大,同一类型的房地产个体在区位特征、建筑特征上也存在很大的差异。批量评估涉及的样本数据数量多、范围广,样本个体间较大的差异使得运用单一的评估模型不利于评估准确性,将样本按一定规则划分后分别评估似乎是一个不错的办法。从经济学供求理论的角度来看,供求平衡的自由市场中,同类产品价格趋同且有内部关联,将样本划分后分别评估存在可行性。最初学者们大多简单按行政区域划分样本,但样本间的差异性与行政区划关联性并不强,显然这样简单划分并不能取得较好的效果。为此有学者提出在考虑行政归属的基础上按区位因素相同或相近分区,以减少特征价格模型中的特征变量,从而提高评估精度(金桂荣,2012)。张然等(2013)在评估商业地产价值之前,依据类别相同、档次相当的原则,构建了以商圈级别和繁华程度为指标的样本区位级别定量评价体系,并参照实际交通情况等其他地理信息对样本进行集合划分,实际应用中效果良好。刘洪玉和李妍(2017)则将房地产样本的划分分为分类、分区、分组三步,在分类和分区完成后,引入模糊数学,采用贴近度方法测度不同房地产之间的相似程度,进而提出房地产分组的思路和应用路径,符合房地产价值分析具有一定模糊性,且同时具有定量分析科学性的特点。随着研究的深入,还有一些统计分析方法也被用于样本划分。Greiner和Thomas(2012)在使用上游主成分分析和聚类分析研究基于样本的住宅房地产组合估值的可能性和边界时,为了寻求在不增加总样本容量的情况下提高估计精度,他们在统计推断房地产组合价值之前讨论了不同的样本分层方法,发现使用分层的简单随机样本比使用非分层的简单随机样本性能更好。崔占如(2014)在批量评估样本的选择中引入系统聚类法,能够提高样本选择的便利性和科学性。目前还不断有新的样本分区方法出现,Hozer等(2019)在现有调查区、分区图、背景图、专题地图、房地产市场分析和统计方法的基础上,提出了基本地形边界的界定方法,从而帮助房地产批量估价确定了给定区域内需要自动化评估的不动产的数量。但当前针对样本分区的方法并没有一个统一的定论,其划分效果和适用性还有待研究。

(二) 评估指标的选择及测度

按照大部分主流的批量评估方法,样本分区完成后进入评估指标的选择及量化环节。房地产价格受到诸多特征因素的影响,理论上选取的特征因素越全面,越能完整描述房地产的真实价格(金桂荣,2012)。但实际选取的特征因素中可能存在相关性和冗余信息,这会影响评估模型的精度和泛化能力。为此,郑江晖(2014)将聚类与判别分析方法引入房地产特征因素的系统性识别与分类,该方法能够明晰特征指标间、特征指标与房地产价格间的相互关系,提高多元回归特征价格模型的适用性。王吓忠和邱岳(2015)在传统特征价格模型基础上,引入粗糙集方法来根据属性重要度约简评价指标,评估结果显示,改进后的模型准确率更高,粗糙集方法可以提高房地产批量评估模型的易用性和准确度。之后王阿忠和李倩(2019)又在传统神经网络上引入粗糙集方法来精简变量,结果表明变量精简后的反向传播(back propagation, BP)神经网络模型在住宅估价上有很好的可靠性,且比普通BP神经网络模型的性能更好、批量评估精度更高。Doszyń(2019)则应用肯德尔等级相关系数确定变量之间关系的强度和方向,在此基础上提出了一种基于Hellwig信息载体个体容量的确定房地产属性影响的统计方法,用来选择计量经济模型中的解释变量,实际评估结果与人工评估结果非常相近。从现有的研究来看,探究评估指标间的关系并对其进行筛选,似乎有助于提高房地产批量评估精度。

评估指标的量化上,单宗房地产评估指标的量化研究对批量评估仍具有一定的借鉴意义。如张协奎等(2001)用隶属函数值来描述房地产价格特征因素,吴红华和赖华勇(2012)使用区间数来度量房地产价格的影响因素等等。大部分学者使用了数值型指标的原始数据,对定性指标则采取了分级打分或虚拟值的方式作量化处理。但人工量化难免存在效率低和主观随意性的问题,为此也有学者预先对指

标的测度做了进一步深化分析。Tian 和 Yang(2015)探索了 GIS 与特征价格模型的结合,利用 GIS 空间分析函数对房地产价格的影响因素进行定量分析并进行房地产批量评估,以西安市为例的实证研究表明,将 GIS 应用于城市住宅房地产批量评估是可行和有效的。黄梦吟和郭化林(2012)、李恒凯等(2018)利用 GIS 中缓冲区分析及网络分析的功能,对房地产交易案例库及待估房地产的价格影响因子进行批量自动化量化,这种方法能够有效避免以往人工量化的主观随意性,为模型的评估精度提供必要的保障。GIS 技术已成为当前房地产批量评估指标精准量化的常用手段。

三、房地产批量评估模型及求解

目前已经涌现出了大量关于房地产批量评估方法的研究成果,如经典回归分析、空间回归分析、机器学习算法等,本文着重探讨基于机器学习算法的房地产批量评估模型。房地产批量评估中,机器学习模型能将房地产特征对价格的非线性影响和交互影响考虑在内,对特征间和特征与价格间的复杂关系进行深度学习,从而提升估值准确性(公云龙和杨雨涵,2022)。诸多学者将支持向量机(support vector machine, SVM)、集成学习、人工神经网络等机器学习算法用于批量评估,出现了大量研究文献。Zurada 等(2011)的研究很具有代表性,他们比较了多种回归分析和人工智能批量评估方法,包括传统的多元回归分析、非传统的 SVM、可加性回归、模型树以及基于人工智能的人工神经网络、径向基函数神经网络、记忆基础推理法,在某些模拟场景下,非传统的回归分析方法在数据集同质性较高的情况下表现的更好,基于人工智能的方法则在数据集同质性较低的情况下表现良好。

(一) 支持向量机

同时期,我国学者也将机器学习算法应用于房地产批量评估,如龚科等(2010)构建了一个基于最小二乘支持向量机(least squares support vector machine, LS-SVM)的物业税税基批量评估模型,代入单个小区的交易数据验证发现,样本量较小的情况下,该模型仍具有较高精度,但在选用的样本量增加、样本分布区域扩大之后,必须充分考虑影响房产价格的各类因素,否则模型精度将降低。之后还有学者将 SVM 用于房地产整体估价模型的改进,SVM 凭借小样本评估时较高的准确性,在整体估价模型比价系数的预测上取得了不错的效果,最后评估模型在训练集和测试集上取得相近且良好的结果(刘璐和施昆昊,2018)。潘巍和晋松(2019)则比较了基于 LS-SVM、SVM 和人工神经网络的房地产估价模型,发现基于 LS-SVM 的评估模型准确率最高且模型训练最快。

相较于传统的多元线性回归模型,SVM 的使用不存在诸多基本假定,进而模型设定较为简单,所需的样本容量也较小。对比人工神经网络,SVM 收敛速度快且不会出现陷入局部最优解和过拟合的问题,LS-SVM 则是 SVM 的改进,能进一步加快收敛速度。但是 Zurada 等(2011)的研究发现仍不能忽视,数据集较低的同质性不利于基于 SVM 构建的房地产估价模型的评估准确性,筛选样本时应尽量选择与评估对象相似的交易案例。因此,基于 SVM 的评估模型更适合于小范围同质房地产的批量评估。

(二) 集成学习

集成学习算法主要可分为 Bagging 系列的并行算法和 Boosting 系列的序列式算法,在房地产批量评估领域都有应用。

回归树是一种应用于因变量为连续变量的决策树,随机森林则是一种由 Bagging 算法改进的基于回归树的组合算法,其对多元线性不敏感、对缺失数据和非平衡的数据比较也不敏感,可以有效解决回归树模型因节点分裂深度过大而产生过拟合的问题。随机森林最早由 Antipov 和 Pokryshevskaya(2012)应用于房地产批量评估,观察其使用住宅公寓数据的实证研究结果发现,该方法无论是在训练集还是测试集都能保持较高的准确性,且比多元回归分析、人工神经网络等方法表现更好,同时还证明了该方法在

样本中存在异常值、缺失值时的稳定性,和正确处理具有多个层次的分类变量的能力。陈诗沁和王洪伟(2020)基于特征价格理论和上海市二手房市场交易数据,从信息增益角度运用随机森林评价每个特征的重要性并据此筛选后,构建多元回归、人工神经网络和随机森林批量评估模型,对比各模型的泛化结果发现,随机森林模型评估误差最小,拟合优度则最高,并提出了评估模型中分类变量较多致使随机森林效果好于人工神经网络的观点。Yilmazer 和 Kocaman(2020)根据土耳其土地登记和地籍总局汇编的现代化地籍数据,比较了多元回归和随机森林模型在房地产批量评估中的应用,并取用训练数据的子集作随机测试,发现随机森林模型的拟合优度略高。公云龙和杨雨涵(2022)从评估精度、泛化能力和经济解释三方面对比了随机森林模型、特征价格模型、支持向量回归模型和BP神经网络模型,发现随机森林模型在房地产评估中的精度高于其他模型,且在测试集上的表现稍好于在训练集上的表现,显示出稳定的泛化能力,另外随机森林对变量重要性的评价还能揭示房地产价格的构成特征,适合于实践应用。可见在房地产批量评估中,随机森林模型在评估精度和泛化能力上具有一定优越性,并且凭借较少的超参数和无需反复迭代运算,其建模、调参和运算速度相较于人工神经网络也具有优势。

Catboost 算法是 Boosting 算法中的一种,它是在梯度提升决策树(gradient boosting decision tree, GB-DT)框架上改进而来的,解决了GBDT的梯度偏差和预测偏移问题,减少了模型过拟合的发生,并且Catboost 算法能高效处理类别型特征,所需调节的参数少而准确率极高,应用领域在逐渐拓宽,现已出现在房地产批量评估研究中。杨智璇和张继瑜(2022)比较了 Catboost 模型、多元线性回归模型、随机森林模型在房地产批量评估中的效果,从测试集的结果可以看出,Catboost 模型的预测精度最高,而误差则最小,与此同时,Catboost 模型相较于另外两个模型准确率提升很大,该模型在保持明显的非线性拟合优势的同时建模工作量则最低。杨艺璇和刘超(2022)同样证实了基于 Boosting 算法的模型在房地产估价中的评估精度高于线性回归、随机森林、神经网络等机器学习模型,他们还进一步利用贝叶斯优化算法改进 Xgboost 模型、Lightgbm 模型和 Catboost 模型的超参数,从测试集的结果看,三个模型的平均绝对误差和平均相对误差都有所降低,改进的效果良好。但是基于 Boosting 算法的评估模型存在与人工神经网络模型类似的问题,作为一种复杂的非线性模型,其不具备类似于多元线性回归模型的显性表达式输出。

(三) 人工神经网络

人工神经网络能够很好地处理非线性问题,相比于前几种机器学习算法,其在房地产批量评估领域的应用得到更多专家和学者的关注。与 Zurada 等(2011)的工作类似,Chun Lin 和 Mohan(2011)研究了当时在美国使用最广泛的多元线性回归模型、非参数可加回归模型和人工神经网络模型的预测精确度,两位作者发现三种方法在评估中低价位住宅时精确度比较接近,并且具有可靠性和很好的成本效益,但在评估高价位住宅时人工神经网络模型的表现更好。不过能否成功地应用于房地产批量评估与该技术具有优越的预测能力同等重要,人工神经网络经常由于“黑箱”特性被质疑。还有研究指出人工神经网络模型的预测精度不及非线性回归模型,加之人工神经网络的输出不够透明,无法提供一个明确的评估模型,因而认定其在房地产批量评估中并不具备绝对优势(Mccluskey 等,2012)。此外,虽然人工神经网络具有较好的自组织性、自适应性及鲁棒性,但其易陷入局部最优,且存在过度适应数据的倾向,有可能导致模型应用于新数据时迅速恶化。

人工神经网络的训练大多采用BP算法,但BP神经网络模型存在收敛速度慢、易陷入极值陷阱等问题。为此有学者尝试在训练人工神经网络时将两种或两种以上的算法相结合,起到优势互补的作用,从而提升人工神经网络的性能。吕霁(2014)提出了一种结合遗传算法(genetic algorithm, GA)和BP神经网络的房地产估价模型,该模型运用BP神经网络确定房地产价格影响因子与评估价格间的函数映射关系,再用GA算法优化BP神经网络的权值和阈值,从测试集的预测结果可以看出,相较于未改进前,改进后的模型拟合效果更好,同时该模型收敛速度快、学习过程中不易陷入局部最优。Yacim 等(2016)则

引入了布谷鸟搜索算法(cuckoo search, CS)来帮助优化人工神经网络,并将其与BP算法、莱文伯格-马夸尔特算法(levenberg-marquardt, LM)结合,从研究结果可以看出,基于CSLM算法、CSBP算法的房地产评估模型无论是准确率还是均方误差、均方根误差都优于单独的BP算法,自此CS算法作为一种元启发式算法被引入房地产批量评估行业。随后,Yacim和Boshoff(2017)又探讨了粒子群优化算法(particle swarm optimization, PSO)和BP算法在人工神经网络权值优化和训练中的应用,研究发现将PSO算法与BP算法结合,在全局和局部搜索属性权重,能够提高人工神经网络的预测精度,并且由于其显示了属性的相对重要性,模型的透明性得到增强,两种算法结合在房地产批量评估中的有效性得到证实。此外,蚁群算法和萤火虫算法等多种算法也可以用于人工神经网络的优化,但还尚未应用在房地产批量评估领域。

人工神经网络算法因“黑箱”特性被质疑,而进化多项式回归(evolutionary polynomial regression, EPR)使用多目标遗传算法来搜索那些同时实现最大化数据准确性和数学函数简约性的模型表达式,相对于人工神经网络的黑箱模型在测度变量关系方面优势明显。Morano等(2018)测试和比较了效用可加方法和EPR在住宅物业批量评估领域的应用,分析阐述了两种方法实施的潜力和机会,以及联合应用它们来解释和预测房地产现象的可能性。随后Morano等(2019)用EPR对三组意大利住宅样本同时实施了评估,最终发现该模型有能力为所考虑的三种不同的地域环境分别生成一种“独特”的函数形式。EPR虽然具有极强的鲁棒性,但其对样本中的异常值比较敏感,可能会影响到分析结果。

四、结论与展望

通过对近十年中外文研究文献的整理和分析发现:首先,现有的房地产批量评估研究大多侧重评估的某个方面,对评估工作前后之间的联系关注较少,暂时还没有能覆盖房地产批量评估各个环节的、系统性较强的研究成果;其次,房地产批量评估方法已较为丰富,但还没有关于批量评估方法适用性与可靠性能够得到普遍认可的研究结论;再次,房地产批量评估方法日新月异,但针对我国本土的批量评估方法研究相对较少。综上,未来我国房地产批量评估研究工作若想实现跨越,应重点关注以下三个方面:

(一) 智能化批量评估系统设计

未来房地产批量评估必将实现高度智能化,批量评估智能化不仅是大幅提升评估效率的保证,也是顺应数字经济时代的必然要求。智能化批量评估系统由多个子系统组成,各子系统具有相对独立性又彼此密切关联,贯穿于房地产批量评估分区、评估指标筛选及量化、房地产批量评估建模及求解整个过程,其智能化表现为不同工作内容的子系统自助完成任务。当然这需要以房地产基础数据库完成开发为前提,并借助数据挖掘、统计分析方法以及计算机技术,自动识别、归类、整理、分析和检验,且能够随着外部环境的变化进行自动调整。更重要的,在研究房地产批量评估方法的过程中,要搭建不同工作环节反复迭代的通道,让各评估工作环节的误差修正更加智能。例如,可以针对批量评估工作的多个环节以及房地产特征的区域异质性,通过贝叶斯优化、强化学习等多种类自动机器学习让一些通用步骤如数据预处理、模型选择和调整超参数等工作实现自动化,来简化机器学习中生成模型的过程。最后,智能化批量评估系统在终端操作上还应比较简便。

(二) 房地产批量评估分区升级

分区域批量评估是业界共识,合理的样本集合选择能够最大限度降低随机扰动因素的影响,这也是房地产批量评估后续工作的基础。传统的按行政区域或使用高分辨率遥感影像采样确定样本集合的方法无法助力评估模型质量的提升,需要充分吸收多核心城市空间结构模式、层次线性回归分析以及基于GIS的空间分析技术等理论和方法的基本思想,探究如何准确体现房地产特征属性并科学分类,如何随

着外部环境变化及时调整分区。为了验证样本划分的合理性,需要通过模型构建、校准和检测等工作反复迭代以使样本分区达到最优,实际上也是要求每个批量评估环节都要把误差控制在最小范围,才能保证最后的评估结果是可信和可靠的。并嵌入新增样本分区搜索机制,使得只要房地产基础数据库有更新,批量评估系统就能够自动识别分区并开展后续工作以及实现通过智能化反复迭代降低误差的效果。目前该项工作研究进展缓慢,主因在于探寻合适的批量评估分区方法需要极大的耐心,很可能要承担在耗费大量时间和精力后仍无法推进研究工作的风险,房地产批量评估亟需在评估样本分区依据、视角、技巧和方法上取得突破。

(三) 房地产批量评估方法应用创新

批量评估模型的构建是房地产评估工作的核心,然而许多研究工作仍旧停留在不同类型方法的比较上且结论莫衷一是,亟需从房地产批量评估建模系统化和本土化、模型算法应用创新上取得突破。建模工作就如同做物理化学实验,每一项实验内容均涉及到不断改变的外部环境条件、模型设定形式的调整、参数的调节及算法的优化,以探寻纵向比较中达到更高精确度的批量评估模型。在保证模型批量评估精度的前提下,亦不能牺牲模型的鲁棒性,建模过程中的过拟合问题也应充分考虑。调节参数或优化算法不是对基本算法生搬硬套,而是要根据具体问题进行设计,包括如何将特征变量的表示参数化为一个函数族,隐藏单元的设计等,新的模型中可能包括多个算法的复合计算。与此同时,房地产批量评估方法的应用创新还体现在如何获取批量评估模型求解方法,模型的全局最优解往往需要通过反复测试并改进智能优化算法来搜寻超参数值。此外,模型的可解释性越高就越容易理解评估结果是如何预测的,在建模阶段模型可解释性辅助开发人员理解模型,在投入运行阶段通过解释模型内部机制对结果给予论证支持,在房地产批量评估方法应用创新中对模型的可解释性也要高度重视。

参考文献

- [1] 耿继进, 张晖. 基于 GIS 的房地产批量评估数据库构建研究——以深圳市为例[J]. 遥感技术与应用, 2012, 27(3): 479-486.
- [2] 金桂荣. 房地产税征管中的房地产批量评估问题研究[J]. 中国注册会计师, 2012(11): 102-106.
- [3] 张然, 耿继进, 李妍, 等. 基于空间位置分析的商业房地产适应性估价技术[J]. 武汉大学学报(工学版), 2013, 46(4): 508-513.
- [4] 刘洪玉, 李妍. 基于模糊数学的房地产批量评估[J/OL]. 清华大学学报(自然科学版), 2017, 57(11): 1202-1206.
- [5] Greiner M, Thomas M. Mass Appraisal of Residential Real Estate Portfolios with Stratified Sampling: A Case Study[J/OL]. Journal of Real Estate Portfolio Management, 2012, 18(3): 305-321.
- [6] 崔占如. 聚类分析法在房产税基批量评估中的应用[J/OL]. 河北金融, 2014(8): 24-26.
- [7] Hozer J, Gnat S, Kokot S, et al. The Problem of Designating Elementary Terrains for the Purpose of Szczecin Algorithm of Real Estate Mass Appraisal[J/OL]. Real Estate Management and Valuation, 2019, 27(3): 42-58.
- [8] 郑江晖. 用聚类与判别分析的房地产特征因素识别与分类方法研究[J]. 福建论坛(人文社会科学版), 2014(10): 36-41.
- [9] 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(6): 26-34.
- [10] 王阿忠, 李倩. 基于粗糙集神经网络的房产税基批量评估研究[J]. 福州大学学报(哲学社会科学版), 2019, 33(5): 30-37.
- [11] Doszyn M. Individual Capacities of Hellwig's Information Carriers and the Impact of Attributes in the Szczecin Algorithm of Real Estate Mass Appraisal[J/OL]. Real Estate Management and Valuation, 2019, 27(1): 15-24.
- [12] 张协奎, 陈伟清, 成文山, 等. 基于模糊数学的市场比较法研究[J/OL]. 中国管理科学, 2001(3): 38-43.
- [13] 吴红华, 赖华勇. 房地产估价的区间数灰色模糊法[J]. 湖南大学学报(自然科学版), 2012, 39(10): 27-30.
- [14] Tian Y, Yang J P. Application of Geographic Information System on Urban Residential Real Estate Mass Appraisal[J/OL].

- Applied Mechanics and Materials, 2015, 744-746: 1665-1668.
- [15] 黄梦吟, 郭化林. 基于 GIS 的房地产税税基批量评估[J/OL]. 财会月刊, 2012(18): 49-53.
- [16] 李恒凯, 柯江晨, 王秀丽. 融 GIS 和 BP 神经网络的住宅房产评估模型[J/OL]. 测绘科学, 2018, 43(8): 104-109.
- [17] 公云龙, 杨雨涵. 随机森林房地产自动评估模型构建及其比较研究[J]. 中国资产评估, 2022(1): 32-41.
- [18] Zurada J, Levitan A, Guan J. A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context[J/OL]. Journal of Real Estate Research, 2011, 33(3): 349-388.
- [19] 龚科, 肖智, 李丹. 基于支持向量机的物业税税基批量评估[J/OL]. 涉外税务, 2010(5): 35-39.
- [20] 刘璐, 施昆昊. 基于支持向量机回归的房屋比价关系研究[J]. 特区经济, 2018(7): 126-128.
- [21] 潘巍, 晋松. 基于回归 LS-SVM 的房地产估价方法[J]. 电子技术与软件工程, 2019(19): 154-155.
- [22] Antipov E A, Pokryshevskaya E B. Mass Appraisal of Residential Apartments: An Application of Random Forest for Valuation and a CART-based Approach for Model Diagnostics[J/OL]. Expert Systems with Applications, 2012, 39(2): 1772-1778.
- [23] 陈诗沁, 王洪伟. 基于机器学习的房地产批量评估模型[J/OL]. 统计与决策, 2020, 36(9): 181-185.
- [24] Yilmazer S, Kocaman S. A Mass Appraisal Assessment Study Using Machine Learning Based on Multiple Regression and Random Forest[J/OL]. Land Use Policy, 2020, 99: 104889.
- [25] 杨智璇, 张继瑜. 数据驱动下房地产批量评估方法研究——以大连市为例[J/OL]. 中国房地产, 2022(15): 54-59.
- [26] 杨艺璇, 刘超. 基于 Boosting 算法的房产估价模型[C/OL]. 香港新世纪文化出版社, 2022: 288-293.
- [27] Chun LIN C, Mohan S B. Effectiveness Comparison of the Residential Property Mass Appraisal Methodologies in the USA [J/OL]. International Journal of Housing Markets and Analysis, 2011, 4(3): 224-243.
- [28] Mccluskey W, Davis P, Haran M, et al. The Potential of Artificial Neural Networks in Mass Appraisal: the Case Revisited[J/OL]. Journal of Financial Management of Property and Construction, 2012, 17(3): 274-292.
- [29] 吕零. 基于遗传算法优化神经网络的房地产评估模型及实证研究[J]. 计算机科学, 2014, 41(S2): 75-77+87.
- [30] Yacim J A, Boshoff D G B, Khan A. Hybridizing Cuckoo Search with Levenberg-Marquardt Algorithms in Optimization and Training Of ANNs for Mass Appraisal of Properties[J/OL]. Journal of Real Estate Literature, 2016, 24(2): 473-492.
- [31] Yacim J A, Boshoff D G B. Combining BP with PSO Algorithms in Weights Optimization and ANNs Training for Mass Appraisal of Properties[J/OL]. International Journal of Housing Markets and Analysis, 2018, 11(2): 290-314.
- [32] Morano P, Tajani F, Locurcio M. Multicriteria Analysis and Genetic Algorithms for Mass Appraisals in the Italian Property Market[J/OL]. International Journal of Housing Markets and Analysis, 2018, 11(2): 229-262.
- [33] Morano, Rosato, Tajani, et al. Contextualized Property Market Models vs. Generalized Mass Appraisals: An Innovative Approach[J/OL]. Sustainability, 2019, 11(18): 4896.

Abstract: New methods of real estate mass appraisal are emerging. This paper summarizes the application of machine learning in real estate mass appraisal at home and abroad from a new view angle, and finds that the existing research results are generally not systematic, scholars differ on the reliability of different types of mass appraisal methods, and there is a lack of innovation for the methods of real estate mass appraisal in our country. Future research work can focus on the design of intelligent mass appraisal system, the upgrading of real estate mass appraisal zoning and the application innovation of real estate mass appraisal method. Sorting out and analyzing research literature can provide references for peers to carry out further research and help establish a real estate mass appraisal method system with Chinese characteristics.

Key words: Real Estate; Mass Appraisal; Machine Learning; Research Status