

引用格式:姚尧,任书良,王君毅,等.卷积神经网络和随机森林的城市房价微观尺度制图方法[J].地球信息科学学报,2019,21(2):168-177. [Yao Y, Ren S L, Wang J Y, et al. Mapping the fine-scale housing price distribution by integrating a convolutional neural network and random forest[J]. Journal of Geo-information Science, 2019,21(2):168-177.] DOI:10.12082/dqxxkx.2019.180508

卷积神经网络和随机森林的城市房价微观尺度制图方法

姚尧^{1,2},任书良¹,王君毅¹,关庆锋^{1*}

1. 中国地质大学(武汉)信息工程学院,武汉 430074;2. 阿里巴巴集团,杭州 311121

Mapping the Fine-Scale Housing Price Distribution by Integrating a Convolutional Neural Network and Random Forest

YAO Yao^{1,2}, REN Shuliang¹, WANG Junyi¹, GUAN Qingfeng^{1*}

1. School of Information Engineering, China University of Geosciences, Wuhan 430074, China;

2. Alibaba Group, Hangzhou 311121, China

Abstract: China's rapid urbanization has caused a large number of migrants to move to the city, which has also led to housing shortages. Rapid access to fine-scale house price distribution data plays a very important role in urban housing management, government decision-making, and urban economic model analysis. The availability of data and limitations of existing models make only a few studies involving the mapping of house price distribution at the microscale. By combining house price data with remote sensing images, this study builds a remote sensing image features mining model based on Convolutional Neural Network (CNN) and Random Forest (RF). The proposed CNN-based model in this paper can be applied for accurate and reasonable microscopic mapping of house prices without introducing auxiliary geospatial variables. Only using the house prices data and remote sensing images, we successfully carry out the house prices mapping with the precision of 5 meters in the downtown area of Wuhan city. By comparison with the results generated by the other three traditional mining techniques (including A: using spatial datasets extracted from auxiliary geographic dataset only, B: using original features extracted from high-resolution remote sensing images only, C: using original features extracted from high-resolution remote sensing images and auxiliary geographic dataset), the results show that the proposed CNN-based model has the highest house price simulation accuracy ($R^2=0.805$), at least 23.28% higher than the fitting accuracies of the traditional methods (A: $R^2=0.592$, B: $R^2=0.0434$, C: $R^2=0.653$). Moreover, based on the fine-scale house price map, this study further analyzes the spatial heterogeneity distribution of housing prices in the downtown area of Wuhan city. By comparing the partial and overall similarity of the simulated house price distribution map calculated via the perceptual hash algorithm, the results

收稿日期:2018-10-09;修回日期:2018-11-27.

基金项目:国家重点研发计划项目(2017YFB0503804);国家自然科学基金项目(41671408);国家自然科学基金青年基金项目(41801306);湖北省自然科学基金杰出青年项目(2017CFA041)。[**Foundation items:** National Key Research and Development Program of China, No.2017YFB0503804; National Natural Science Foundation of China, No.41671408, 41801306; Natural Science Fund of Hubei Province, No.2017CFA041.]

作者简介:姚尧(1987-),男,广东梅州人,副教授,研究方向为空间大数据和智慧城市。E-mail: yaoy@cug.edu.cn

*通讯作者:关庆锋(1977-),男,四川绵阳人,博士,教授,研究方向为高性能空间计算。E-mail: guanqf@cug.edu.cn

also demonstrate that the house prices distribution of Wuhan city has remarkable fractal characteristics. The micro-scale house price data obtained in this study can provide a basis for microeconomics and fractal research in the urban economics. Meanwhile, this study also provides a brand-new research method for micro-scale economic analysis and resource optimization of large cities in China.

Key words: Housing price; deep learning; microscale; convolutional neural network; random forest; Wuhan

***Corresponding author:** GUAN Qingfeng, E-mail:guanqf@cug.edu.cn

摘要:随着中国城市化进程的加快,城市人口的大规模集聚带来了住房紧张的问题,房价政策制定的时效性与正确性也时刻吸引着社会的关注,因此在微观尺度下对房价进行精细化制图变得愈发重要。由于数据可获取性和现有模型精度的限制,目前已有研究均较少涉及微观尺度。本研究通过将房价数据和遥感影像相融合,构建了一种基于卷积神经网络(CNN)和随机森林(RF)的遥感影像挖掘模型,以实现在不考虑其他数据的情况下,精确、合理地进行房价的微观尺度制图。本文以武汉市作为研究区,在仅有房价数据和遥感影像的情况下,利用本文所构建的模型成功得到武汉市中心城区5 m精度的精细房价图。此外,还利用其他数据源以及挖掘技术与本文所构模型进行了对比分析。结果显示,本文所构建的模型获得了最高的房价模拟拟合优度($R^2=0.805$),相比传统方法中的最高拟合优度($R^2=0.653$)其精度提升了23.28%,其制图结果可为政府部门规划决策及武汉市经济分布研究提供基础支撑。

关键词:房价;深度学习;微观尺度;卷积神经网络;随机森林;武汉

1 引言

近年来,房价已经成为关乎经济和民生的首要因素。已有研究表明,房价的波动不仅影响区域经济的发展,还与人们的消费水平和健康状况息息相关^[1-2]。而中国房价矛盾突出问题多集中在北京、上海、深圳等特大城市。近2年来,房地产开发市场十分活跃,各项指标均呈现较快增长^[3],对于遏制房价快速上涨的调控政策也相继颁布^[4]。在此形势下,城市房价的精细化制图更加发挥着重要的作用,为相关政府部门和房地产市场提供了极具参考价值的房价信息。例如,通过参考微观尺度房价图,充分考虑房价分布的细空间异质性,采取更加细化的区域房价调节手段以确定合理的住宅补贴水平^[5]。

以往对于房价的空间研究多集中在宏观尺度上,如Rondinelli和Veronese^[6]使用人口普查数据和由房地产开发商提供的住宅租赁价格数据来估算租赁价格的变化。同时,也有研究认为房价的空间自相关性和异质性需要进一步考量,因此Osland^[7]对地理加权回归等模型进行改进并引进了空间计量经济学模型。已有研究表明,遥感影像特征所挖掘的语义信息可作为制图的有效数据,如Duque等^[8]提出基于高分辨率(VHRSR)图像的纹理和结构特征模型,来获得更为合理的城市内部贫民窟指数地图,结果表明基于高分遥感影像获得的居民居住条件同城市居民的生计密切相关,可作为制图有效数据。

在上述研究中,由于数据精度不够精细,且预测模型多采用常规的统计学或简单机器学习模型,难从复杂的地理空间对象中充分挖掘隐藏的语义信息,故导致房价的制图结果分辨率不高,无法体现微观尺度下房价分布的空间异质性和分形性特征,从而较难提供有效的、精细化的指导^[9]。

针对以往的研究的不足和对房价复杂空间异质性、分形性的考虑^[10],本研究将建立一个基于卷积神经网络(CNN)的遥感影像挖掘模型来充分挖掘其空间特征信息,以得到精准、细粒度的武汉市房价分布图。本文首先应用多尺度随机抽样模型,从高分辨率遥感影像和现有房价信息中构建相关空间数据集;然后,本文利用卷积神经网络(CNN)来进行空间数据的特征挖掘;之后将获取所得的特征放入随机森林(RF)拟合模型之中来进行房价分布的拟合;最后,本文以武汉市作为研究区域,基于所提出的CNN模型,结合武汉市中心城区高分辨率遥感影像和武汉市住宅均价数据,精细化拟合以获取武汉市高分辨率房价分布图。通过与多种基于不同空间数据集及挖掘模型的房价制图方法进行对比,对比结果都证明了本研究所提出方法的优势和准确性。

2 基于深度学习的微观尺度房价绘制方法

本文所设计的研究流程图如图1所示,通过多尺度窗口滑动特征采样-卷积神经网络挖掘遥感影

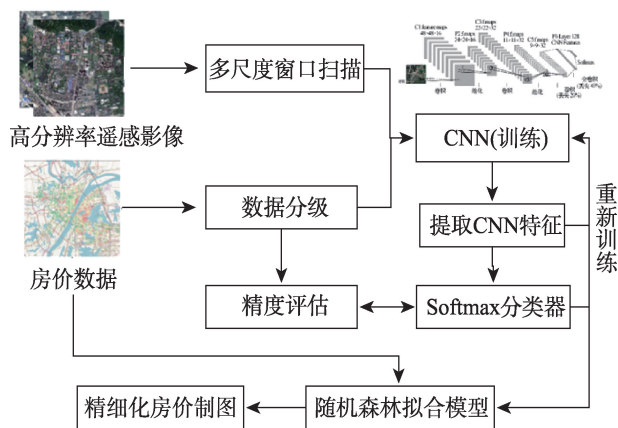


图1 通过CNN挖掘遥感影像数据集来模拟住房价格的流程

Fig. 1 Flow for simulating housing prices by mining remote-sensing image datasets via CNN

像特征-随机森林进行房价拟合3个步骤来获得房价分布图:①通过数据预处理和多尺度滑动窗口采样,构建了城市住房价格影响因子的数据集并训练了一个轻量级卷积神经网络(CNN),根据softmax层计算出的误差进行反向迭代传播,得到最优模型;②去除预训练CNN模型的softmax层,并将其输出的高维特征向量作为训练特征,建立具有原始房价数据的RF拟合模型;③基于拟合模型,计算具有一定窗口大小的每个像元的住房价格,并以精细的规模获得最终的住房价格结果。

2.1 模型构建

本研究设计的卷积神经网络(CNN)结构如图2所示。已有研究表明,UMCNN具有从挖掘对象中获取良好社会信息的巨大潜力,其框架中的浅层卷积神经网络已经应用在模拟城市扩张、城市精细尺度制图等方面,满足空间数据挖掘方面的

需要^[11-12]。同时,考虑到计算机CPU的计算能力并且为了避免出现过拟合现象,本文将训练模型设计为深度较浅、具有较少卷积层的网络结构^[13-14]。此外,在具体的网络结构参数上,本文通过手动调参来确定较好的卷积神经网络结构,参数的设置和选取引用了He等^[11]以及Yao等^[12]利用UMCNN所开展的城市计算的工作时,对UMCNN模型所做的参数调整和选取。

如图2所示,本文最终确定的CNN结构共计7层,包括3层卷积层,2层最大池化层,1层完全连接层和1层softmax层^[15]。其中,在卷积层和完全连接层中使用的激活函数是整流线性单元,这是一种线性激活函数,拥有更好的拟合效果以及更快的训练速度。本文还在第三卷积层和最终完全连接层间设置40%的丢弃率,这会随机禁用某些神经元的权重,从而有效防止模型过度拟合^[16]。已有的研究表明,与较大的卷积核相比,较小的卷积核参数更少,训练效果更好,训练速度也更快^[17]。因此,在卷积核的设置上,本文选择大小为 $3 \times 3 \times 3$ 像元的卷积核。之后,将训练好的卷积神经网络去掉softmax层,形成一个具有高维CNN特征的数据集生成器,并利用其构建拟合模型。在拟合模型的选取上,采用随机森林(RF)作为最终的拟合模型。以往研究指出,随机森林在解决多维特征拟合所带来的多重共线性问题时有很好的效果^[18],因此被广泛应用于多维特征的分类和回归问题^[19]。

此外,为了对抗多尺度效应,解决地面目标的尺度差异问题。本文还应用了多尺度随机土地斑块采样的策略,以房屋样本数据点为中心设置了宽度为 w 的采样窗口。在每个窗口随机抽样 w/s 样

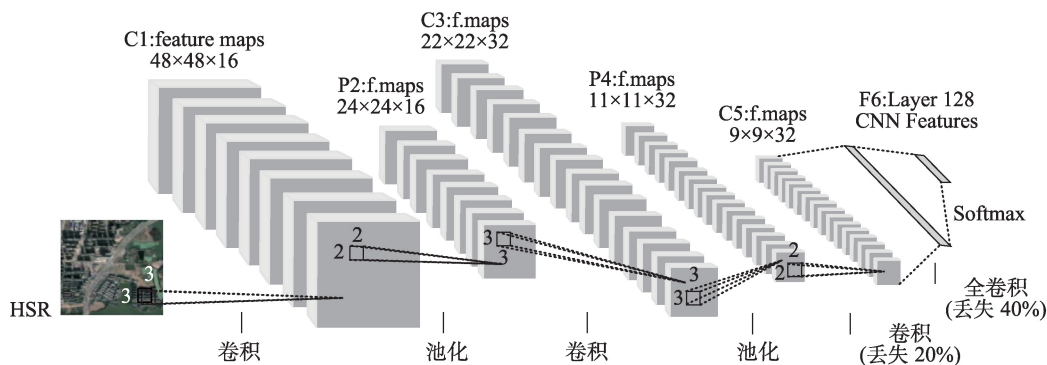


图2 用于进行数据挖掘特征提取的CNN计算框架

Fig. 2 The computational framework of proposed CNN used to feature extraction

本,并保证 s 小于 w ,逐步增加直到和 w 相等。通过多尺度随机斑块采样,可以扩大CNN训练数据的数量,有助于避免过拟合问题,进一步提高CNN模型分类准确性^[13]。

本研究将通过多尺度随机斑块采样得到的样本数据集,按照80%和20%进行划分并分别用于CNN模型的训练以及用于验证与误差的传播。为了检验最终的房价分布模拟精度,将预处理后的原始房价数据集随机分为60%的训练集和40%的测试集,用于训练和验证本文所设计的遥感影像特征挖掘模型拟合部分的精度。在随机森林对房价分布进行精细化拟合时,通过装袋方法从原始数据集中随机提取样本生成新的数据集。在随机特征选择的过程中,每个训练子数据集构建单个决策树,在生长过程中决策树不被修剪,可以获得袋外(OOB)的估计错误报告。以往研究已经指出,OOB估计的拟合误差被证明要优于交叉验证方法^[20]。同时,为了确保结果的可靠性,本研究将采用Bagging方法在训练数据集中随机划分出部分数据作为OOB数据,通过20次随机训练和预测,得到其预测的平均精度。

2.2 模型准确性评估

本文采用若干精度评价指标,将模拟的房价结果和真实结果进行对比来定量评估本文提出的模型最终房价制图的准确性。这些精度评价指标包括:皮尔森相关系数(Pearson R),拟合优度(Standard R^2),均方根误差(RMSE),平均绝对误差(MAE)。

$$Pearson R = \frac{\sum_{i=1}^n (h_{i,s} - \bar{h}_s)(h_{i,o} - \bar{h}_o)}{\sqrt{\sum_{i=1}^n (h_{i,o} - \bar{h}_o)^2} \sqrt{\sum_{i=1}^n (h_{i,s} - \bar{h}_s)^2}} \quad (1)$$

$$Standard R^2 = \frac{n \sum_{i=1}^n h_{i,o} h_{i,s} - \sum_{i=1}^n h_{i,o} \sum_{i=1}^n h_{i,s}}{\sqrt{n \sum_{i=1}^n h_{i,o}^2 - \left(\sum_{i=1}^n h_{i,o} \right)^2} \sqrt{n \sum_{i=1}^n h_{i,s}^2 - \left(\sum_{i=1}^n h_{i,s} \right)^2}} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (h_{i,o} - h_{i,s})^2}{n}} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |h_{i,o} - h_{i,s}| \quad (4)$$

式中: $h_{i,o}$ 和 $h_{i,s}$ 分别是第 i 个样本的实际和模拟房价; n 是研究区域内的总样本。

3 武汉市中心城区微观房价制图

3.1 研究区概况以及数据准备

武汉市是中国中部地区的中心城市,2017年生产总值为13 410.34亿元,人均生产总值为123 831元人民币,其发达的经济基础和便利的交通位置使得住宅价格稳定提升。如图3所示,本文的研究区域为武汉市内中心城区,共涉及主城区6个行政区划(江岸区、江汉区、汉阳区、武昌区、洪山区、青山区),远城区4个行政区划(东西湖区、蔡甸区、江夏区、黄陂区),总面积约为9147 km²。主城区和远城区的划分参考自武汉市最新颁布的买房落户政策,不同的政策要求下使得各行政区域的人口流动格局和经济结构发生变化^[21],这也使不同行政区之间的空间异质性更加复杂^[22-24]。

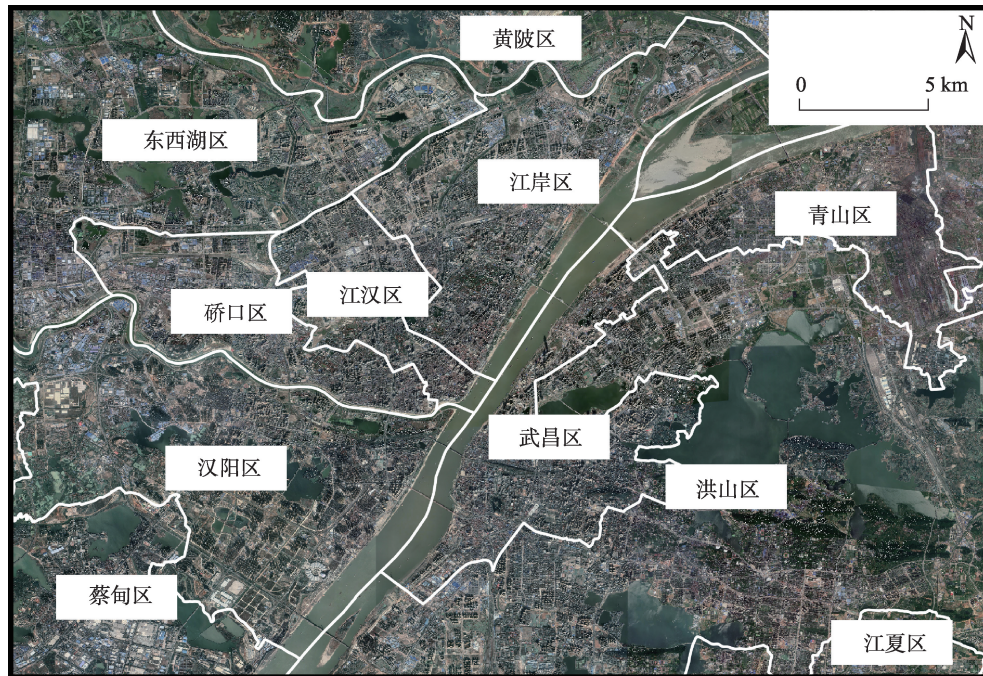
本文从中国最大的在线住宅市场网站Fang.com获取了研究区域内带有编号、经度、维度、价格/平方米等信息的房价数据。预处理,共获得3016组有效数据(图4)。

由于采集的房价数据是连续数值,本文通过对它们进行标准偏差等级划分来进行离散化^[25]。设研究区每平方米住房价格的平均值和标准差分别为 μ_h 和 σ_h ,在数据预处理中,为保证数据的可靠性,删除了低于2000元/m²或者高于 $\mu_h + 3 \cdot \sigma_h$ 的房价样本。然后,对原始房价数据进行评级,步长为 $0.25 \cdot \sigma_h$,提取出的数据作为CNN的输入数据^[12]。研究中部分试验也应用了辅助地理空间数据(表1和图5,均见第173页),数据的选择覆盖了影响武汉城市房价的几个主要要素,包括生活环境、交通条件、生活便利、道路密度等。

3.2 对比实验

为对比证明本文设计的房价制图模型在进行精细房价制图时的可靠性和优越性,本研究一共进行了4组不同的对比方法和实验。如表2(见第174页)所示,A组为仅使用辅助地理数据特征,B组为仅使用遥感图像原始特征,C组为使用辅助地理数据和遥感图像原始特征组合,D组为使用组合数据并且通过CNN挖掘遥感影像特征。经过实验,各类别的估测结果和误差分布如表3(见第174页)所示。

通过对比4组实验(A组、B组、C组、D组),实验D均方根误差(RMSE)、平均绝对误差(MAE)均为最小,表明实验D所模拟的预测房价与真实房价之



注:背景数据是 Google Earth 提供的湖北省武汉市遥感图像,空间分辨率为 5 m。

图3 研究区域(武汉市)

Fig. 3 Study area (Wuhan city)

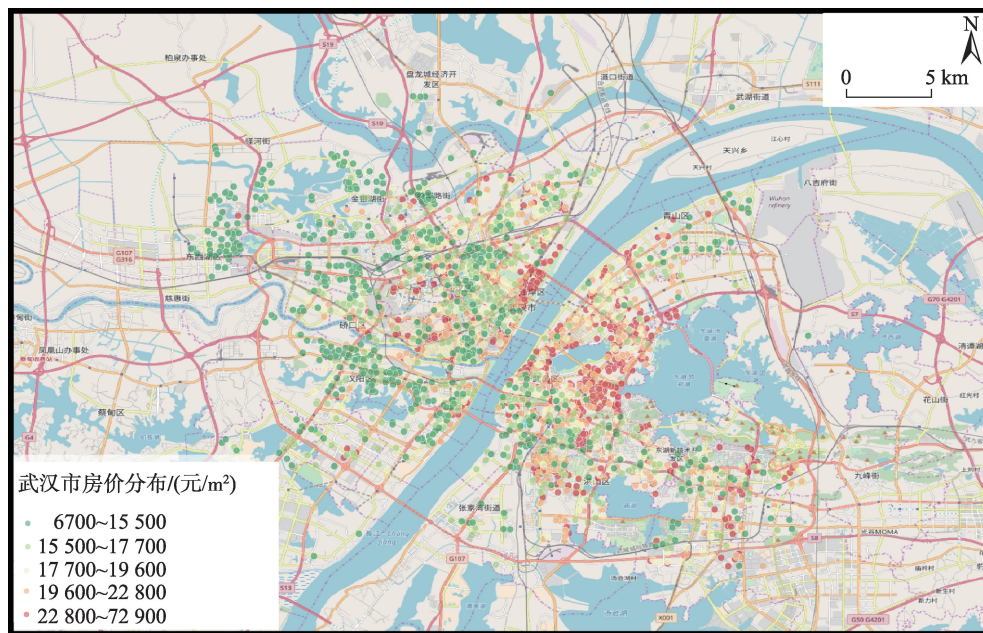


图4 中国最大的在线住宅市场网站Fang.com收录的房价数据

Fig. 4 The acquired housing price data from Fang.com, China's biggest online housing market website

间误差最小,结果最为准确。实验D中拟合模型获得了最高的拟合优度($R^2=0.805$),相比3种传统方法(A: $R^2=0.592$;B: $R^2=0.434$;C: $R^2=0.653$)其精度提升了至少23.28%。这表明当采用CNN进行遥感影

像高级语义特征提取时(D组),比仅使用单一类型的输入数据(A组、B组)的拟合模型或通过多源数据简单地应用特征组合(C组)取得的精度更高。这表明仅仅使用遥感影像的原始纹理和结构特征或

表1 辅助地理空间数据选择表

Tab. 1 Selection table of auxiliary geospatial data

参数类别	二级分类
区位特征	政府机构
教育资源	幼儿园、小学、初中 高中、大学
交通便捷	地铁站 公交站
医疗资源	大型医院 小型门诊
便民服务	超市 商场 美食 公园 生活服务 银行
休闲娱乐	休闲广场 宾馆 娱乐场所 游乐园
道路密度	快速路 主干路 次干路 支路

辅助地理数据并不能实现高精度的模拟预测,在对数据的挖掘过程中会丢失部分特征信息。相比之下,CNN则更适用于对遥感影像高级语义和背景信息的挖掘与提取^[13,17,26]。

3.3 武汉市精细房价制图结果分析

武汉市主要城区房价的 Moran's I 指数的计算结果为 0.109, z 得分为 126.054, p 得分为 0。武汉市主城区的房价已经通过了空间自相关检验且 Moran's I 指数为正值,表明武汉市主城区的聚类分布特征已经非常明显,存在高值聚类、低值聚类的房价分布特点^[27]。从图 6 可看出,高房价区域($\geq 25\ 000$ 元/ m^2)主要位于江岸、江汉、武昌、洪山的沿江地带,低房价区域($<10\ 000$ 元/ m^2)多集中在东西湖区、汉阳区的南部、黄陂等较远区域。此外,由表 4(见第 175 页)可知,相对于武汉中心主城区(江汉区,江岸区,洪山区,武昌区,汉阳区,青山区),远城区(江夏,蔡甸区,东西湖区,黄陂区)的房价模拟准确度较低。这是由于同主城区相比,远城区划入武汉市行政区的时间较晚,城市规划和基础设施配套建设虽然比较完善,但是住宅和商业地产的发展却很缓慢,大部分地区仍为农村住宅社区和村庄,这导致 CNN 提取遥感影像特征后进行房价模拟时容易出现误差。同时,本文还发现主城区的房价模拟精度与各区生产总值排名相似,远城区的精度则与各区房屋均价排名一致。这表明在主城区武汉房价已基本稳定,区域经济发展和房价基本相匹配,而远城区则还处于房价的增长阶段。

由模拟房价分布与真实房价分布的对比(图 7,见第 175 页)可看出,模拟房价直方图与真实房价直方图具有较高的相似性,图像特征和形状相近,这

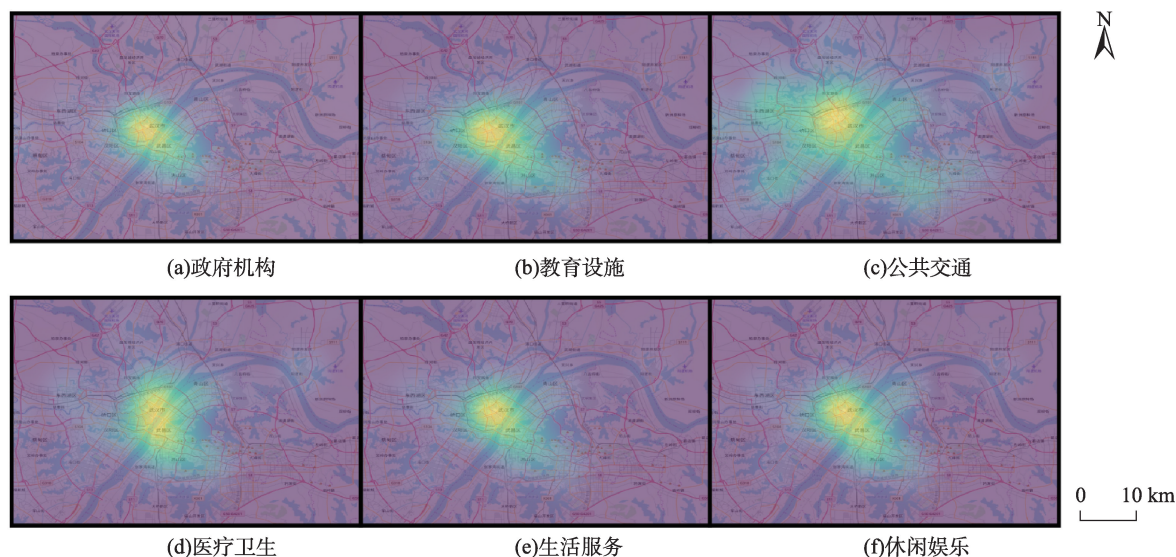


图5 POI辅助地理空间数据集空间分布密度

Fig. 5 The spatial distribution density of Baidu POIs and the auxiliary geospatial datasets

表2 基于不同空间数据挖掘模型的空间房价分布
精细制图方法

Tab. 2 The methods of mapping fine-scale housing prices
via different mining model

实验编号	实验描述	实验标签
A	仅使用辅助空间数据集	RF(SD)
B	仅使用遥感影像原始特征	RF(HSR)
C	使用辅助空间数据集和遥感 影像原始特征	RF(HSR&SD)
D	通过CNN从遥感影像中提特征	CNN(HSR)

表3 不同房价分布制图方法得到的结果精度对比结果

Tab. 3 The accuracy assessment results of different
mapping fine-scale housing prices methods

精度评价指标	实验编号			
	A	B	C	D
Pearson R	0.775	0.655	0.809	0.818
Standard R ²	0.592	0.434	0.653	0.805
RMSE	3915.534	4650.235	3614.826	3462.558
MAE	2884.581	2594.251	2535.015	2168.949

表明利用CNN模拟的武汉市房价总体是精准的。另外,由图8可看出,洪山区鲁磨路南区域周边房价在15 000~22 000元/m²之间,且房屋较聚集。洪山区鲁磨路北段区域周边房价多在10 000~18 000元/m²之间,且房屋分布较分散。随后通过遥感图像(图8)

对比,发现鲁磨路南段靠近光谷商业圈,而且处于大学生活动辐射周边区域,人流量密集,经济活跃,房屋价格较高,且价格区域较为集中。鲁磨路北段区域临近东湖风景区,政府实行环境保护政策,禁止开发和拆迁,此地多为乡村住宅因此价格较低。

此外,已有研究表明,武汉市城市空间结构在不断发展的过程中呈现出分形学特征^[28]。为了进一步验证该结论,本文利用感知哈希算法^[29],计算得到光谷附近区域模拟房价分布图(图8)与武汉市整体模拟房价分布图(图6)的相似度为0.828。并且,通过局部细节观察可知,鲁磨路南起光谷广场向北延伸。而光谷广场作为大型商业中心,其附近区域房价呈现由中心向四周下降的特点,并呈现出商业区、高新产业区、商住混合区房价较高,风景保护区、市郊区房价较低的分布规律,这与武汉市整体房价分布规律保持高度一致。因此,光谷区域的房价分布可以作为武汉市房价分布的自相似分形,并且也验证了武汉市房价分布有一定的分形学特征^[30]。

综上所述,本文所提出的精细化制图方法不仅可以较高精度地模拟武汉市中心城区微观尺度下的房价分布,而且可以较好地显示出房价分布的空间异质性和城市分形特征。

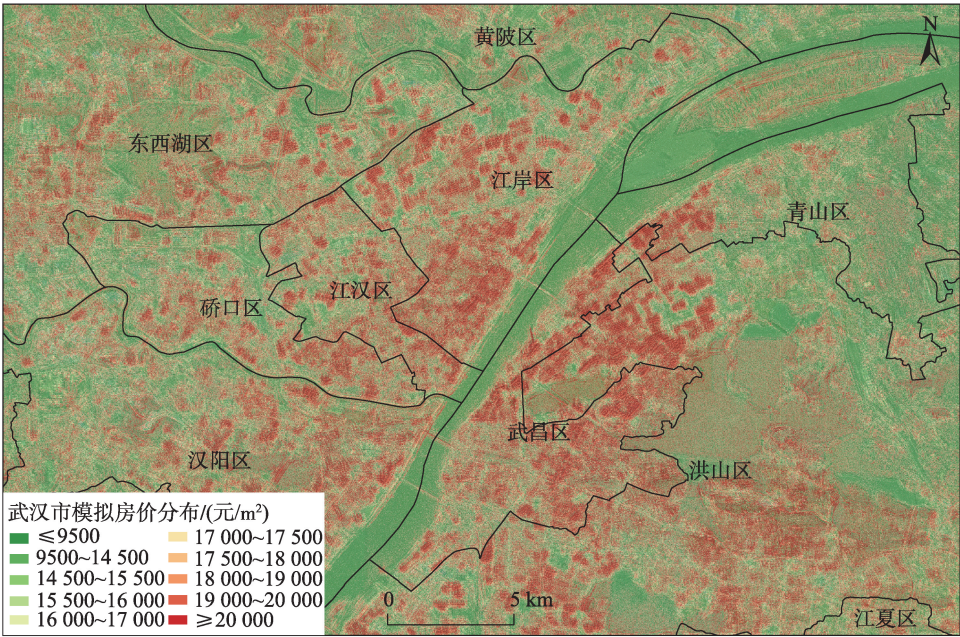


图6 CNN(HSR)模拟的武汉市房价分布

Fig. 6 Wuhan's housing prices simulated via CNN(HSR)

表4 武汉中心区域不同地区房价的平均值、标准差和总体准确度

Tab. 4 Average values, standard deviations, and overall accuracies of housing prices in different districts in Wuhan central area

区域类别	区域	真实/(元/m ²)		预测/(元/m ²)		准确度/%
		平均值	标准差	平均值	标准差	
主城区	江汉区	19 577.900	4612.090	19 261.333	3913.542	98.38
	江岸区	20 845.530	7071.716	20 337.013	5188.628	97.56
	洪山区	20 384.442	4577.304	19 654.007	3639.945	96.42
	武昌区	22 129.308	6297.948	20 964.847	3467.340	94.74
	汉阳区	16 674.734	3868.972	17 708.541	3544.656	93.80
	青山区	15 822.054	4327.896	16 813.957	3255.780	93.73
远城区	江夏区	18 222.980	4833.506	18 423.225	3861.028	98.90
	蔡甸区	15 347.774	4090.231	17 043.074	2801.621	88.95
	东西湖区	14 127.650	3076.744	16 367.463	2822.753	84.15
	黄陂区	12 825.273	2654.790	16 298.700	2303.813	72.92

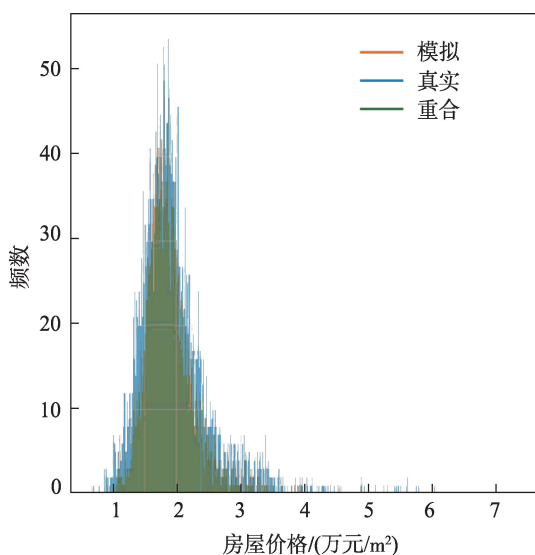


图7 模拟房价和真实房价对比直方图

Fig. 7 Contrast histogram of simulated house prices and real house prices

4 结论

中国大城市的房价过高现象^[31],是目前中国城市经济发展面临的最紧迫的问题之一,最近越来越受到学术界的关注。然而,之前对于房价的研究主要集中于房价定价的驱动因素^[32],或房价变化和经济波动之间的相关性的大尺度研究^[33-36]。由于缺乏有效的数据和模型,在精细化刻画城市房价空间异质性分布的研究尚存空缺。

本研究通过卷积神经网络(CNN)实现对遥感影像高级语义特征的挖掘提取并利用随机森林(RF)进行拟合,以获得高精度的微观尺度房价分布图。通过在武汉市中心城区进行实验,并与其他不同的特征挖掘和模型相比,本文所提出的模型精度为最高,比传统方法提升至少23.28%。本研究得到的精细化的房价分布制图,除尺度更加精细,总体

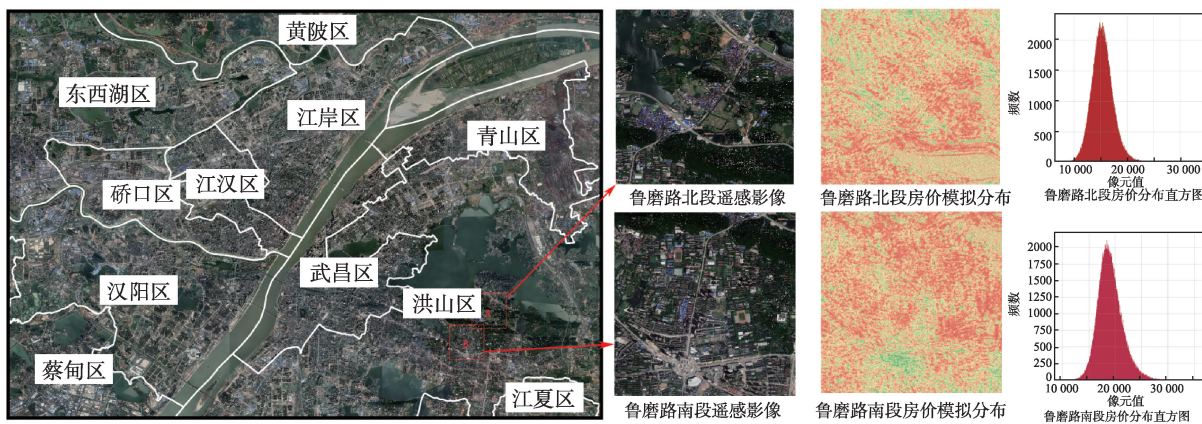


图8 CNN房价模拟图细节

Fig. 8 Some details of the spatial distribution of housing prices simulated via CNN

准确度更高外,还可以较好地显示出房价分布的空间异质性以及分形特征。在挖掘模型的训练中,数据是影响制图精度的重要因素之一^[37],而本研究采用的CNN房价制图模型,对数据要求较低,数据组织和获取较容易,只需要高分辨率遥感影像和基础房价数据即可。此外,由于房价的分布可以在一定程度上代表居民的经济状况^[38],所以此模型还可以对许多城市经济进行评估,为未来经济学和城市分形研究提供基础数据。例如,人类活动模式识别和城市土地价格分类以及城镇体系结构提取等。

然而,本研究所提出的模型没有将时间因素考虑在内。在现实的城市规划和建设中,往往存在基础配套设施建设和住房规划不同步的问题,而配套设施和住房价格又相互影响^[39-40]。这就造成了在对遥感影像挖掘过程中,难免会因为此问题造成部分区域精度不高的问题。另外,如何确定并调整深度学习的参数,包括层数和卷积核大小,仍然是机器学习领域一个引人深思的问题^[17,41]。本文所采用的模型主要依赖于对已有研究较好模型参数的参考^[11-12],今后也将对此开展进一步的研究,力求得到更高精度的模型和更细致的模拟房价分布图。

下一步将从3个方面开展研究工作:①提高城市空间变异映射中CNN模型的可解释性;②考虑将时间语义量化,并将其输入CNN模型,以实现更加精确的房价绘制和分析;③提高模型的学习能力,以便在更广泛的范围内(如国家和全球范围)绘制住房价格。

参考文献(References):

- [1] Atalay K, Edwards R, Liu B Y J. Effects of house prices on health: New evidence from Australia[J]. *Social Science & Medicine*, 2017,192(21):36-48.
- [2] Fan Q. In: Impact of Fluctuation in House Price on Urban Household Consumption[C]. Singapore: 2nd International Conference on Economic, Business Management and Education Innovation, 2014:6.
- [3] Ge T, Wu T. Urbanization, inequality and property prices: Equilibrium pricing and transaction in the Chinese housing market[J]. *China Economic Review*, 2016,45(4):310-328.
- [4] Target Range of Chinese House Price Regulation[C]. Melbourne: 2011 International Conference on Mechanical, Industrial, and Manufacturing Engineering, 2011:4.
- [5] Liu F Y. Drastic measures for curbing housing price rising [J]. *China Opening Herald*, 2006,131(3):11-13,29.
- [6] Rondinelli C, Veronese G. Housing rent dynamics in Italy[J]. *Questioni Di Economia E Finanza*, 2010,28(62):540-548.
- [7] Osland L. An application of spatial econometrics in relation to hedonic house price modeling[J]. *Journal of Real Estate Research*, 2010,32(3):289-320.
- [8] Duque J C, Patino J E, Ruiz L A, et al. Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data[J]. *Landscape & Urban Planning*, 2015,135(3):11-21.
- [9] Huang X, Zhang L. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2012,51(1):257-272.
- [10] Razavian A S, Azizpour H, Sullivan J, et al. CNN Features off-the-shelf: An astounding baseline for recognition [C]. Stockholm: Computer Vision and Pattern Recognition Workshops, 2014:512-519.
- [11] He J, Xia L, Yao Y, et al. Mining transition rules of cellular automata for simulating urban expansion by using the deep learning techniques[J]. *International Journal of Geographical Information Science*, 2018,32(10):1-22.
- [12] Yao Y, Zhang J, Hong Y, et al. Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data[J]. *Transactions in GIS*, 2018,22(2):561-581.
- [13] Zhong Y, Fei F, Zhang L. Large patch convolutional neural networks for the scene classification of high spatial resolution imagery[J]. *Journal of Applied Remote Sensing*, 2016,10(2):25006.
- [14] Tong S, Ling H, Fu Y, et al. Cross-view gait identification with embedded learning[C]. New York: Thematic Workshops of Acm Multimedia. ACM, 2017.
- [15] Krizhevsky A, Sutskever I, Hinton G E. Jeju Island, Korea: ImageNet classification with deep convolutional neural networks[C]. Lake Tahoe: Advances in neural information processing systems, 2012.
- [16] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. *Computer Science*, 2012,3(4):212-223.
- [17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- [18] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1):5-32.
- [19] 詹国旗,杨国东,王凤艳,等.基于特征空间优化的随机森林算法在GF-2影像湿地分类中的研究[J].*地球信息科学学报*,2018,20(10):1520-1528. [Zhang G Q, Yang G D, Wang F, et al. The random forest classification of wetland from

- gf-2 imagery based on the optimized feature space[J]. Journal of Geo-information Science, 2018,20(10):1520-1528.]
- [20] Biau G. Analysis of a Random Forests Model[J]. Journal of Machine Learning Research, 2010,13(2):1063-1095.
- [21] 单卓然,黄亚平,张衍春.中部典型特大城市人口密度空间分布格局——以武汉为例[J].经济地理,2015,35(9):33-39. [Shan Z R,Huang Y P,Zhang X C. Spatial distribution of population density of typical megacities in central China: A case study of Wuhan[J]. Economic Geography, 2015,35(9):33-39.]
- [22] 张俊峰,张安录.土地资源空间异质性与经济发展水平的关系研究——以武汉城市圈为例[J].自然资源学报,2015,30(5):725-735. [Zhang J F, Zhang A L. Study on relationship between spatial heterogeneity of land resources and economic development: A case study of Wuhan urban agglomeration[J]. Journal of Natural Resources, 2015, 30(5):725-735.]
- [23] 龙莹.空间异质性与区域房地产价格波动的差异——基于地理加权回归的实证研究[J].中央财经大学学报,2010,10(11):80-85. [Long Y. Spatial heterogeneity and the difference of regional fluctuation in housing price: Based on the geographically weighted regression[J]. Journal of Central University of Finance & Economics, 2010, 10(11):80-85.]
- [24] 何鑫,田丽慧,楚尔鸣.人口流动视角下中国房价波动的空间异质性[J].人口与经济,2017,181(6):43-57. [He X, Tian L H, Chu E M. The spatial heterogeneity of house price volatility in China from the perspective of migration [J]. Population & Economics, 2017,181(6):43-57.]
- [25] Jean N, Burke M, Xie M, et al. Combining satellite imagery and machine learning to predict poverty[J]. Science, 2016,353(6301):790-794.
- [26] Chen X W, Lin X. Big data deep learning: Challenges and perspectives[J]. IEEE Access, 2014,2:514-525.
- [27] Getis A, Ord J K. The analysis of spatial association by use of distance statistics[J]. Geographical Analysis, 1992, 24(3):189-206.
- [28] 陈涛,刘继生.城市体系分形特征的初步研究[J].人文地理,1994,233(1):25-30. [Chen T, Liu J. Preliminary studies on the fractal properties of systems of towns[J]. Human Geography, 1994,233(1):25-30.]
- [29] Li W, Preneel B. A secure perceptual hash algorithm for image content authentication[J]. Lecture Notes in Computer Science, 2011,7025(1):108-121.
- [30] 陈彦光.分形城市与城市规划[J].城市规划,2005,33(2):33-40. [Chen Y G. Fractal cities and city planning[J]. City Planning Review, 2005,33(2):33-40.]
- [31] Wu B, Li R, Huang B. A geographically and temporally weighted autoregressive model with application to housing prices[J]. International Journal of Geographical Information Science, 2014,28(5):1186-1204.
- [32] He C, Wang Z, Guo H, et al. Driving forces analysis for residential housing price in Beijing[J]. Procedia Environmental Sciences, 2010,2(1):925-936.
- [33] Wang L. Empirical study on the impact of urban rail Transit on house price: Based on hedonic price model[J]. Areal Research & Development, 2009,28(2):57-61,71.
- [34] Wang Y, Wang D L, Liu L H, et al. Spatial differentiation of urban housing prices and its impacts on land market in China[J]. China Land Sciences, 2015,29(6):33-40.
- [35] An H, Yu L, Gupta R. Capital inflows and house prices: Aggregate and regional evidence from China[J]. Australian Economic Papers, 2016,55(4):451-475.
- [36] 尹上岗,李在军,宋伟轩,等.基于地理探测器的南京市住宅租金空间分异格局及驱动因素研究[J].地球信息科学学报,2018,20(8):1139. [Yin S G, Li Z J, Song W X, et al. Spatial differentiation and influence factors of residential rent in Nanjing based on geographical detector. Journal of Geo-information Science, 2018,20(8):1139-1149.]
- [37] 郭志懋,周傲英.数据质量和数据清洗研究综述[J].软件学报,2002,13(11):2076-2082. [Guo Z M, Zhou A Y. Research on data quality and data cleaning: A Survey[J]. Journal of Software, 2002,13(11):2076-2082.]
- [38] Chan S. Spatial lock-in: Do falling house prices constrain residential mobility?[J]. Journal of Urban Economics, 2001,49(3):567-586.
- [39] Wen H, Zhang Y, Zhang L. Do educational facilities affect housing price? An empirical study in Hangzhou, China[J]. Habitat International, 2014,42(42):155-163.
- [40] 王芳,高晓路,颜秉秋.基于住宅价格的北京城市空间结构研究[J].地理科学进展,2014,33(10):1322-1331. [Wang F, Gao X L, Yan B Q. Research on urban spatial structure in Beijing based on housing prices[J]. Progress in Geography, 2014,33(10):1322-1331.]
- [41] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous distributed systems[J]. ArXiv preprint, arXiv:1603.04467,2016-03-14.