

基于演化CatBoost算法的房价预测模型

王成章, 白晓明, 汤文英, 陈书涵

引用本文

王成章, 白晓明, 汤文英, 陈书涵. 基于演化CatBoost算法的房价预测模型[J]. 计算机科学, 2024, 51(11A): 240300180-5.

WANG Chengzhang, BAI Xiaoming, TANG Wenyong, CHEN Shuhan. [Evolutionary CatBoost Based Housing Price Prediction Model](#) [J]. Computer Science, 2024, 51(11A): 240300180-5.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[小样本飞机生产质量偏差数据分析与预测方法研究](#)

Study on Analysis and Prediction Method of Small Sample Aircraft Production Quality Deviation Data
计算机科学, 2024, 51(11A): 240300123-8. <https://doi.org/10.11896/jsjx.240300123>

[基于联邦学习的智能电网AMI入侵检测方法研究](#)

Study on Smart Grid AMI Intrusion Detection Method Based on Federated Learning
计算机科学, 2024, 51(6A): 230700077-8. <https://doi.org/10.11896/jsjx.230700077>

[基于线性规划松弛的移动边缘计算卸载模型](#)

MEC Offloading Model Based on Linear Programming Relaxation
计算机科学, 2023, 50(6A): 211200229-5. <https://doi.org/10.11896/jsjx.211200229>

[基于决策树改进深度交叉网络的推荐模型](#)

Recommendation Model Based on Decision Tree and Improved Deep & Cross Network
计算机科学, 2023, 50(6A): 220300084-7. <https://doi.org/10.11896/jsjx.220300084>

[基于机器学习的SCADE模型组合验证环境假设自动生成方法](#)

Machine Learning Based Environment Assumption Automatic Generation for Compositional Verification of SCADE Models
计算机科学, 2023, 50(6): 297-306. <https://doi.org/10.11896/jsjx.220500207>

基于演化 CatBoost 算法的房价预测模型

王成章¹ 白晓明² 汤文英¹ 陈书涵¹

1 中央财经大学统计与数学学院 北京 100081

2 首都经济贸易大学管理工程学院 北京 100070

摘要 遗传规划算法采用函数变换将原有变量张成的空间映射到新的特征空间,通过遗传算子操作实现目标函数的最优化。影响房价波动的因素有很多,各影响因素与房价之间呈现复杂的非线性关系。文中提出了一种基于演化 CatBoost 算法的房价预测模型,将影响房价波动的各因素变量编码为遗传规划算法的终端变量,采用 CatBoost 算法作为基学习器构建适应度函数,针对房价预测的特点设计合理的遗传算子,在函数映射后的特征空间上实现目标函数的最优化,以提升预测模型的性能。实验结果表明,基于演化 CatBoost 算法的房价预测模型的预测性能优于传统的基于随机森林算法、支持向量机算法、自适应增强算法、极致梯度提升算法等的预测模型,能够更好地实现房价的预测,在相同条件下具有更高的预测准确度。

关键词: 遗传规划; CatBoost 算法; 预测模型; 决策树; 最优化

中图分类号 TP391

Evolutionary CatBoost Based Housing Price Prediction Model

WANG Chengzhang¹, BAI Xiaoming², TANG Wenyong¹ and CHEN Shuhan¹

1 School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China

2 School of Management and Engineering, Capital University of Economics and Business, Beijing 100070, China

Abstract Genetic programming algorithm uses function transformation to map the space formed by the original variables to a new feature space, and optimizes the objective function through genetic operator operations. There are many factors that affect housing price fluctuations, and each influencing factor exhibits a complex nonlinear relationship with housing prices. This paper proposes an evolutionary CatBoost algorithm based housing price prediction model. Various factor variables that affect housing price fluctuations are encoded as terminal variables of the genetic programming algorithm. CatBoost algorithm is employed as the base learner to construct a fitness function, and reasonable genetic operators are designed according to the characteristics of housing price prediction. The objective function is optimized in the feature space after function mapping to improve the performance of the prediction model. Experimental results show that the prediction performance of evolutionary CatBoost algorithm based housing price prediction model is superior to that of traditional prediction models based on random forest algorithm, support vector machine algorithm, adaptive enhancement algorithm, extreme gradient enhancement algorithm, etc. It can predict housing prices more accurately than the rivals under the same conditions.

Keywords Genetic programming, CatBoost algorithm, Prediction model, Decision tree, Optimization

1 引言

近年来,我国商品房价格呈现出“整体上涨,个体异质”的特征^[1],关于房价波动的分析和预测研究一直是学术界关注的热点话题。

房价波动受到错综复杂的诸多因素的影响,大量学者从供给端、需求端、土地政策及宏观经济环境等层面对房价波动的影响因素进行研究。已有研究发现,土地价格和居民可支配收入^[2]、地区 GDP 和大宗商品价格指数^[3]、常住人口数量、房地产开发投资、竣工面积、地区生产总值、货币发行量 M2^[4]、投资者的投机行为^[5]、住房供给弹性^[1]等因素都对商品房价格的变化有着显著的影响。不仅如此,失业率、人均

GDP 和通货膨胀率^[6]、股票价格和利率^[7]、社区社会经济因素^[8]、经济政策不确定性^[9-10]等宏观经济因素也会影响房价的变化,且各种因素对房价变化的影响存在异质性的特点^[11-12]。

传统的房价预测方法多依赖于计量经济模型和统计模型^[3-4],研究者构建了参数回归模型和半参数回归模型^[13]、模糊逻辑模型^[14]、线性自回归模型及非线性平滑过渡自回归模型^[15]来预测房价变化,结果显示非线性模型的预测效果要优于线性模型^[16]。然而,我国房地产市场呈现出复杂的非线性特征,鉴于机器学习算法在有效捕捉数据中的复杂模式和非线性关系方面的优势,越来越多的研究人员采用机器学习算法来捕捉各个影响因素与房价变化之间的非线性关系,以期

基金项目:北京市社会科学基金(21GGB018);中央财经大学教育教学改革基金(2022ZXJG09)

This work was supported by the Social Science Foundation of Beijing(21GGB018) and the Educational Reform Foundation of Central University of Finance and Economics(2022ZXJG09).

通信作者:王成章(czwang@cufe.edu.cn)

提高房价预测的准确性。研究人员在相关工作中提出基于支持向量机(Support Vector Machine, SVM)算法^[17-19]、决策树算法^[20-21]、随机森林算法(Random Forest, RF)^[21]、朴素贝叶斯(Naive Bayesian, NB)算法及自适应增强(Adaptive Boosting, AdaBoost)算法^[20]、极致梯度提升(eXtreme Gradient Boosting, XGBoost)算法^[22]预测未来房价,融合遗传算法(Genetic Algorithm, GA)和支持向量机(SVM)构建GSVM房价预测模型^[23],实验结果表明机器学习算法可以显著提高房价预测的精度。

不同于已有研究在原始特征空间构建机器学习算法,遗传规划(Genetic Programming, GP)^[24]通过遗传编码和遗传算子对原有变量进行函数变换,将由原有变量张成的空间映射到新的特征空间,使得目标函数在新的特征空间上达到最优。在此基础上,演化森林(Evolutionary Forest, EF)算法^[25]以决策树为基学习器构造适应度函数,采用遗传规划算法对原有变量空间进行空间变换,使得随机森林模型在变换后的特征空间上性能达到最优,从而提高算法的预测准确率。值得注意的是,遗传规划只是给出了算法的框架,应用时需要根据具体问题设计特定的适应度函数和遗传算子;而算法中基类学习器的设计对整个预测模型的性能起着至关重要的作用。鉴于此,本文提出了一种新的基于演化CatBoost算法的房价预测模型,采用泛化能力更强、预测性能更优的类别特征梯度提升树(Categorical Boosting, Catboost)算法^[26]作为基学习器,进而基于遗传规划理论,针对房价预测问题设计了合理的适应度函数及遗传算子,将原有的变量空间映射到新的特征空间,提高模型的预测精度。

2 基于演化CatBoost算法的房价预测模型

借鉴已有文献的研究,本文选择地区生产总值、居民可支配收入、居民消费价格指数等房价变动影响因素为自变量,商品房平均销售价格为因变量。各变量的详细说明如表1所列。

表1 变量说明

Table 1 Variable descriptions

变量	变量含义
y	商品房平均销售价格
x_0	居民人均可支配收入
x_1	地区生产总值
x_2	居民消费价格指数
x_3	年末常住人口数
x_4	城镇居民人均可支配收入
x_5	房地产开发企业住宅新开工房屋面积
x_6	房地产开发企业住宅竣工套数
x_7	房地产开发企业土地成交价款
x_8	房地产开发企业土地购置费用
x_9	房地产开发住宅投资额
x_{10}	地域省份标识
x_{11}	时间年份标识

遗传规划算法通过对原有变量的函数变换实现由原始特征空间到新的特征空间的映射,算法中包含两个集合:变换函数集合 \mathcal{F} 和终端变量集合 \mathcal{X} 。集合 \mathcal{F} 由支持线性和非线性运算的函数组成,包括加法、减法、乘法、除法、正弦、余弦、平方、最大化、最小化、平方根等函数,即:

$$\mathcal{F} = \{+, -, \times, \div, \sin, \cos, \wedge 2, \max, \min, \text{sqrt}, \dots\}$$

集合 \mathcal{X} 由原有变量和一些常数数值构成,即:

$$\mathcal{X} = \{x_1, x_2, \dots, x_n, \pi, 5, \text{rand}, \dots\}$$

其中,rand表示随机生成的常数。在遗传规划算法中,种群中的个体由终端变量和变量之间的运算函数组合而成,表示为树状结构形式。例如,两个终端变量 $x_1 \in \mathcal{X}$ 和 $x_2 \in \mathcal{X}$ 在加法运算 $f_+ \in \mathcal{F}$ (标记为“+”)函数下组合而成的个体 $\varphi = x_1 + x_2$ 的树状结构如图1所示。

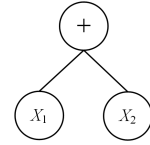


图1 遗传规划中的个体表示

Fig. 1 Individual representation in genetic programming

遗传规划算法根据适应度函数对个体的优劣进行评价。为了更加准确地预测房价,本文设定衡量基学习器预测性能的判定系数 R^2 为适应度函数。作为演化森林算法^[25]基学习器的决策树可以在一定程度上有效刻画变量间的非线性关系,但是决策树算法本身存在稳定性差、泛化能力不足的问题。CatBoost算法^[26]采用梯度提升机制来解决学习过程中的过拟合及预测偏移问题,可有效提升算法的准确性和泛化能力。鉴于此,本文采用性能更优的CatBoost算法^[26]作为基学习器构建基于演化CatBoost算法(Evolutionary CatBoost, ECatboost)的房价预测模型。

具体来讲,本文设计了如下的遗传算子。

1)选择算子:选择算子是从父代个体中选择适应度高的个体进入下一代。为了提高房价预测模型的泛化能力,我们采用 K 折交叉验证的方法估计每个个体的适应度,因此个体 φ 会计算得到由 K 个适应度函数值构成的适应度向量 Fit_φ ,即:

$$Fit_\varphi = (fit_\varphi^{(1)}, fit_\varphi^{(2)}, \dots, fit_\varphi^{(K)})$$

为保持种群的多样性,本文针对适应度向量设计与文献^[25]类似的词典库选择算子。首先引入弹性因子 ϵ ,当个体 φ_1 和 φ_2 的适应度之差小于 ϵ 时,认为两个个体是等价的。然后生成 $1 \sim K$ 之间的随机数 ind ,比较两个个体适应度向量中第 ind 个分量的大小,选择适应度值高的个体进入下一代。

2)交叉算子:交叉算子是遗传规划算法中用于生成新个体的重要操作,通过结合父代个体的部分遗传信息生成具有新特征的子代个体。本文设计的交叉算子为:首先随机选择两个或多个父代个体,并在父代个体中随机定位交叉位置。例如,以2个父代个体 $\varphi_1 = x_1 / \sqrt{\text{rand}^2 + x_4}$ 和 $\varphi_2 = x_2 / (x_6 + 1)$ 为例,随机定位其交叉位置,如图2所示。

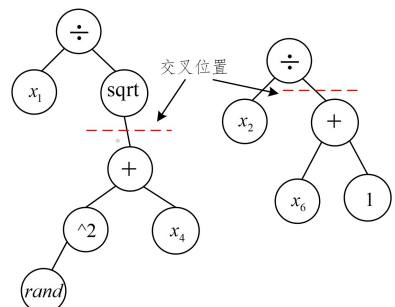


图2 交叉操作父代个体

Fig. 2 Parent individual of crossover operation

然后,生成 0~1 之间的随机数 p_1 ,当 p_1 小于事先设定的交叉概率 p_c 时,对选定的父代个体执行交叉操作。将父代个体 φ_1 和 φ_2 中位于交叉位置的子树进行互换,生成下一代的两个新个体 $\varphi_1' = x_1 / \sqrt{x_6 + 1}$ 和 $\varphi_2' = x_2 / (rand^2 + x_4)$,如图 3 所示。

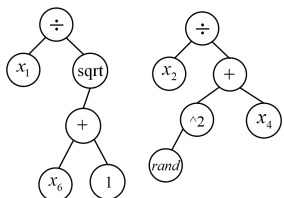


图 3 交叉操作子代个体

Fig. 3 Offspring individual of crossover operation

3)变异算子:变异算子是遗传规划算法中增加种群多样性的重要操作,通过随机修改父代个体中的部分遗传信息来生成变异后的新个体。本文设计的变异算子为:首先随机选择 1 个或多个父代个体,并在父代个体中随机确定变异位置。例如,以 1 个父代个体 $\varphi_3 = x_1 / \sqrt{x_2^2 + x_3}$ 为例,随机确定变异位置如图 4 所示。

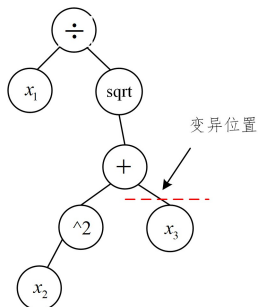


图 4 变异操作父代个体

Fig. 4 Parent individual of mutation operation

然后,生成 0~1 之间的随机数 p_2 ,当 p_2 小于事先设定的变异概率 p_m 时,对选定的父代个体执行变异操作。将父代个体 φ_3 中位于变异位置的子树进行随机变换,生成下一代的新个体 $\varphi_3' = x_1 / \sqrt{x_2^2 + 5}$,如图 5 所示。

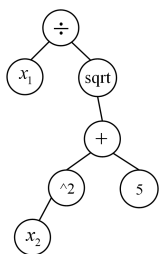


图 5 变异操作子代个体

Fig. 5 Offspring individual of mutation operation

4)精英保留策略:为了防止最优个体在进化过程中丢失,本文设计了最优个体保留的精英保留策略,对进化过程中每一代适应度最高的最优个体执行保留策略,构成最优个体集合,进化结束后,模型的最优解最终由最优个体集合产生。

算法的伪代码如算法 1 所列。

算法 1 算法的伪代码

1. 随机初始化种群 $P = \{\varphi_1, \dots, \varphi_L\}$
2. 初始化最优个体集合 $A \leftarrow \emptyset$
3. 初始化进化代数 $g \leftarrow 0$
4. While $g \leq g_{max}$ do # 主循环
5. for $i = 1, \dots, L$ do
6. if $i \% 2 = 1$ & $rand() < p_c$ then # 交叉
7. $\{\varphi_i, \varphi_{i+1}\} \leftarrow crossover(\varphi_i, \varphi_{i+1})$
8. if $rand() < p_m$ then # 变异
9. $\varphi_i \leftarrow mutation(\varphi_i)$
10. for $\varphi \in P$ do
11. $f_\varphi = cross\ validation(\varphi)$
12. $P' \leftarrow selection(P)$ # 选择
13. $A \leftarrow update(A, P)$
14. $P \leftarrow P'; g \leftarrow g + 1$
15. return A

3 实验结果

本文中采用的我国房价数据来自国家统计局官方网站,数据集包含了 2003—2022 年我国除港澳台地区以外的 31 个省级行政区的房价及影响因素的相关数据,删除缺失值数据后样本量大小为 526。除此之外,我们还采用了来自美国卡内基梅隆大学统计学院发布的波士顿房价数据集和扩展后的波士顿房价数据集来进行房价预测实验¹⁾,样本量大小为 506。

为了验证本文提出的基于演化 CatBoost 算法的房价预测模型(ECatBoost)的有效性,我们还在相同条件下分别采用随机森林算法(RF)、支持向量机算法(SVM)、自适应增强算法(AdaBoost)、极致随机树算法(Extremely Randomized Trees,ET)、梯度提升树算法(Gradient Boosted Decision Trees,GBDT)、极致梯度提升算法(XGBoost)、轻量级梯度提升学习机算法(Light Gradient Boosting Machine, Light-GBM)、丢失机制多加性回归树算法(Dropouts meet Multiple Additive Regression Trees,DART)、类别特征梯度提升树算法(Catboost)进行了房价预测实验,并对实验结果进行了比较分析。

本文分别统计了相同条件下各个预测模型的判定系数 R^2 ,预测结果的均方根误差 MSE,预测结果的平均绝对误差 MAE,计算式如式(1)一式(3)所示:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \tag{1}$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \tag{2}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \tag{3}$$

其中, y_i 为真实值, \hat{y}_i 为预测值, \bar{y} 为真实值的平均值。我国房价数据集上的实验统计结果如表 2 所列。

实验结果表明,本文提出的基于演化 CatBoost 算法的房价预测模型(ECatBoost)的预测效果在所有同类比较算法中是最优的,其次是 XGBoost 算法,而 SVM 算法的预测性能最差。在相同实验条件下,ECatBoost 算法的判定系数 R^2 最大,

¹⁾ <https://lib.stat.cmu.edu/datasets/>

达到了 0.97, 比 SVM 算法提高了 14.1 个百分点。ECatBoost 算法的预测精度也是最高的, 均方根误差 MSE 为 0.01, 比 SVM 算法的 MSE 降低了 5.4 个百分点; ECatBoost 算法的平均绝对误差 MAE 为 0.077, 比 SVM 算法降低了 10.7 个百分点。

美国波士顿房价数据集和扩展后的波士顿房价数据集上的实验统计结果如表 3 和表 4 所列。

表 2 我国房价数据集上的预测性能比较

Table 2 Comparison of predictive performance on Chinese housing price datasets

算法	R^2	MSE	MAE
AdaBoost	0.884	0.046	0.167
CatBoost	0.946	0.022	0.114
DART	0.882	0.048	0.163
ET	0.922	0.031	0.126
GBDT	0.888	0.045	0.162
LightGBM	0.891	0.044	0.156
RF	0.921	0.032	0.131
SVM	0.839	0.064	0.184
XGBoost	0.946	0.022	0.108
ECatBoost	0.970	0.010	0.077

表 3 波士顿房价数据集上的预测性能比较

Table 3 Comparison of predictive performance on Boston housing price datasets

算法	R^2	MSE	MAE
AdaBoost	0.760	19.131	2.875
CatBoost	0.831	13.314	2.456
DART	0.778	18.003	2.739
ET	0.848	12.073	2.220
GBDT	0.795	16.374	2.806
LightGBM	0.789	17.072	2.647
RF	0.810	15.166	2.416
SVM	0.630	31.919	3.396
XGBoost	0.808	15.246	2.367
ECatBoost	0.928	6.053	1.660

表 4 扩展后波士顿房价数据集上的预测性能比较

Table 4 Comparison of predictive performance on extended Boston housing price datasets

算法	R^2	MSE	MAE
AdaBoost	0.766	18.703	2.908
CatBoost	0.822	14.563	2.392
DART	0.784	17.555	2.641
ET	0.857	11.387	2.085
GBDT	0.801	15.841	2.723
LightGBM	0.796	16.598	2.544
RF	0.816	14.614	2.286
SVM	0.618	32.998	3.419
XGBoost	0.814	14.918	2.314
ECatBoost	0.935	5.464	1.638

在美国波士顿房价数据集及扩展后的波士顿房价数据集上的实验结果也验证了本文提出的房价预测模型的有效性, 相同条件下本文提出的 ECatBoost 预测模型的预测性能要明显优于其他算法。CatBoost 算法可以更好地刻画房价影响因素与房价波动之间的非线性复杂关系, 结合遗传规划则可以在变换后的特征空间上进一步提升模型的预测准确性。

结束语 本文提出了一种新的基于演化 CatBoost 算法的房价预测模型 (ECatBoost), 基于遗传规划理论, 设计了合理的遗传算子将原始的变量空间映射到新的特征空间, 采用泛化能力更强、预测性能更优的 Catboost 算法作为基学习

器, 在映射后的特征空间上构建预测模型。多个数据集上的实验结果均表明, 本文提出的 ECatBoost 预测模型在房价预测方面的性能要优于比较算法 RF, ET, AdaBoost, GBDT, SVM, DART, XGBoost, LightGBM, Catboost。ECatBoost 预测模型显著提高了房价预测的性能和精确度。

目前 ECatBoost 预测模型的实现还只是基于单线程的做法, 如何采用多线程并行策略进一步改进算法, 以提高算法的运行效率是今后需要进一步研究的重要方面。

参考文献

- [1] LIU X Y, DU C, LI S L. Natural geographical constraints, land use regulations, and China's housing supply elasticity[J]. *Economic Research*, 2019, 54(4): 99-115.
- [2] LIN Z L, LI X Y. Empirical study on the influencing factors of commodity housing prices based on panel model[J]. *Economic Mathematics*, 2017, 34(4): 73-78.
- [3] XU J, YE Z Q. Analysis of Factors Influencing the Price of Commercial Housing Based on VAR Model[J]. *Statistics and Decision Making*, 2017(11): 93-97.
- [4] GAO T. Empirical Study on Factors Influencing Housing Prices in Nanjing City[J]. *Economic Research Guide*, 2019(12): 125-127, 172.
- [5] ZHENG M, WANG H F, WANG C Z, et al. Speculative behavior in a housing market: Boom and bust[J]. *Economic Modelling*, 2017(61): 50-64.
- [6] ALFREDAS L, ANTANAS L, ALGIMANTAS L. Macroeconomic Variables Influencing Housing Prices in Vilnius[J]. *International Journal of Strategic Property Management*, 2022, 26(1): 24-34.
- [7] YANG H, LI C. Research on the influencing factors and contribution of housing prices in Chinese cities: a relative importance decomposition based on $R \sim 2$ [J]. *Exploration of Economic Issues*, 2019(11): 49-62.
- [8] ZHANG Y, ZHANG D, MILLER E J. Spatial Autoregressive Analysis and Modeling of Housing Prices in City of Toronto[J]. *Journal of Urban Planning and Development*, 2021, 147(1): 05021003.
- [9] WANG S, ZENG Y N, YAO J Y, et al. Economic policy uncertainty, monetary policy, and housing price in China[J]. *Journal of Applied Economics*, 2020, 23(1): 235-252.
- [10] SU C W, LI X, TAO R. How does economic policy uncertainty affect prices of housing? Evidence from Germany[J]. *Argumenta Oeconomica*, 2019(1): 131-153.
- [11] CUI Z, ZHOU M Q, KONG L Z. Research on Heterogeneity of Factors Influencing Urban Housing Prices in China[J]. *Taxation and Economics*, 2022(6): 65-74.
- [12] LI C Q. Research on the Regional Heterogeneity of the Impact of Monetary Policy on the Real Economy and Housing Prices: A GVAR Model Based on the Construction of Payment Data Weight Matrix[J]. *Shanghai Finance*, 2023(6): 56-70.
- [13] BIN O. A prediction comparison of housing sales prices by parametric versus semi-parametric regressions[J]. *Journal of Housing Economics*, 2004, 13(1): 68-84.
- [14] KUSAN H, AYTEKIN O, ÖZDEMİR I. The use of fuzzy logic in predicting house selling price[J]. *Expert Systems with Applications*, 2010(37): 1808-1813.

- [15] BALCILAR M, GUPTA R, MILLER S M. The out-of-sample forecasting performance of nonlinear models of regional housing prices in the US[J]. *Applied Economics*, 2015(47):2259-2277.
- [16] JIANG X, JIA Z, LI L, et al. Understanding Housing Prices Using Geographic Big Data: A Case Study in Shenzhen[J]. *Sustainability*, 2022, 14(9):5307.
- [17] CHEN J H, ONG C F, ZHENG L, et al. Forecasting spatial dynamics of the housing market using Support Vector Machine [J]. *International Journal of Strategic Property Management*, 2017(4):273-283.
- [18] WU J Y, WANG S Y, SHI H W, et al. Analysis and prediction of housing market prices in Beijing based on multi wavelet analysis[J]. *Journal of Beijing University of Chemical Technology: Natural Science Edition*, 2019, 46(5):101-106.
- [19] ZHOU L J, ZHAO M Y. Analysis of house price prediction based on several types of machine learning models[J]. *National Circulation Economy*, 2022(6):111-116.
- [20] PARK B, BAE J K. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data[J]. *Expert Systems with Applications*, 2015, 42(6):2928-2934.
- [21] TRAWIŃSKI B, et al. Comparison of expert algorithms with machine learning models for real estate appraisal[C]// *IEEE International Conference on Innovations in Intelligent Systems and Applications*. 2017:51-54.
- [22] ZHU H Y, WANG Z J, YE C C. Prediction of Housing Prices in Urban Hotspot Areas Based on XGBoost Algorithm: A Case Study of Jiangbei New Area in Nanjing[J]. *Building Economy*, 2022, 43(S2):433-437.
- [23] GU J R, ZHU M C, JIANG L G Y. Housing price forecasting based on genetic algorithm and support vector machine[J]. *Expert Systems with Applications*, 2011(38):3383-3386.
- [24] AGAPITOS R, LOUGHRAN M, NICOLAU S, et al. A Survey of Statistical Machine Learning Elements in Genetic Programming[J]. *IEEE Transactions on Evolutionary Computation*, 2019, 23(6):1029-1048.
- [25] ZHANG H, ZHOU A, ZHANG H. An Evolutionary Forest for Regression[J]. *IEEE Transactions on Evolutionary Computation*, 2022, 26(4):735-749.
- [26] PROKHORENKOVA L, GUSEV G, VOROBEV A, et al. Catboost: unbiased boosting with categorical features[J]. *Advances in Neural Information Processing Systems*, 2018(31):6638-6648.



WANG Chengzhang, born in 1977, Ph.D, associate professor, master supervisor. His main research interests include machine learning, pattern recognition and big data analysis.