Operations Research and Fuzziology 运筹与模糊学, 2021, 11(3), 257-267
Published Online August 2021 in Hans. http://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2021.113030

Hans汉斯

基于梯度提升回归树模型的上海市二手房均价分析

汪春丽，刘露萍*

贵州大学数学与统计学院，贵州 贵阳
Email: *llpmath@163.com

收稿日期：2021年6月7日；录用日期：2021年7月12日；发布日期：2021年7月20日

摘　要

本文基于梯度提升回归树集成模型，利用采集的“链家”网站上海市近三年各住宅小区二手房的相关数据，分析影响上海市二手房均价的因素。对各影响因素运用Person相关系数矩阵及热力图进行初步分析，并将收集的数据分为训练集和测试集，训练并测试支持向量机模型、线性回归模型及集成模型。最终实验结果表明，基于梯度提升回归树的集成模型更能准确的预测上海市二手房的均价，且梯度提升回归树的MSE是其中最小，相关系数最大达到0.831，具有较好的拟合效果。

关键词

二手房均价，机器学习，梯度提升回归树，模型对比

Analysis of the Average Price of Second-Hand Houses in Shanghai Based on Gradient Boosting Regression Tree

Chunli Wang, Luping Liu*

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou
Email: *llpmath@163.com

Received: Jun. 7th, 2021; accepted: Jul. 12th, 2021; published: Jul. 20th, 2021

*通讯作者。

文章引用：汪春丽，刘露萍．基于梯度提升回归树模型的上海市二手房均价分析[J]．运筹与模糊学，2021，11(3): 257-267. DOI: 10.12677/orf.2021.113030

Abstract

Based on the gradient boosting regression tree, we analyze the factors affecting the average price of second-hand houses in Shanghai by using the data collected from "HOME LINK" website in recent three years. The Person correlation coefficient matrix and heat map are used for preliminary analysis of each influencing factor. Moreover the collected data are divided into training set and test set, and the support vector machine model, linear regression model and integration model are trained and tested respectively. The final experimental results show that the integrated model based on gradient boosting regression tree can more accurately predict the average price of second-hand houses in Shanghai. And the MSE of gradient boosting regression tree is the smallest, furthermore the correlation coefficient is up to 0.831, which has the best fitting effect.

Keywords

Second-Hand House Average Price, Machine Learning, Gradient Boosting Regression Tree, Model Contrast

1. 引言

房价一直以来都是人们关注的重点民生问题，最近几年随着“房子是用来住的，不是用来炒的”[1]理念已全面贯彻实施。市场已经逐渐熟悉了“房住不炒”的理念，在强调稳地价、稳房价、稳预期的前提下，市场已经逐渐形成了非常稳定的市场预期。

目前，在房屋价格的预测上，许多的学者选用机器学习的算法构建了不同的房价预测模型。如：文献[1]以北京市 2010~2018 年的房屋数据作为研究对象，利用多种单小波分析，证明小波去噪可以很好地保留房价的变动趋势，且用 SVM 模型构造的房价预测模型比原始数据的预测精度更高；文献[2]在传统的 SVR 预测房价的模型中引入蝙蝠算法，对北京市二手房价格指数的变化特征进行分析表明，该改进模型具有更好地泛化能力；文献[3]将关键词关注指数加入了商品房价格预测的回归预测模型，其结果表明加入该指数提高了商品房价格预测能力；文献[4]通过仿真得到影响房价的主要因素，然后在此基础上运用 BP 神经网络的方法构建了房价预测的模型；文献[5]以上海市的房价数据为对象，对数据进行降维处理，然后运用主成分分析的支持向量机方法对上海市房价进行预测，该方法具有较高的泛化能力和较好的预测精度；文献[6]将房屋价格变动的过程视为马尔科夫链，并利用北京商品房销售价格数据作为仿真对象，计算其马尔科夫链转移概率，对北京市房屋的价格走向进行预测分析。

梯度提升回归树模型(Gradient Boosting Regression Trees, GBRT)在建立预测模型的方法中具有广泛的应用。文献[7]运用梯度提升回归模型预测生猪的价格，不仅提高了生猪价格预测的预测精度，而且解决了传统模型速度慢、核函数选择难等问题；文献[8]利用梯度提升回归树模型不仅准确地获取了地面高分辨率、高精度的 O_3 浓度分布数据，而且其构建的预测模型对 O_3 浓度的预测效果显著提升；文献[9]用基于时间序列关系的梯度提升回归树对道路的交通事故进行预测，预测结果表明加入时间序列对交通事

[1]2016 年 12 月 14 日至 16 日中央经济工作会议明确：“房子是用来住的，不是用来炒的”。

故预测的精度大大提升；文献[10]在集成学习的构架下，用优化的梯度提升树算法对旅游流量构建模型进行预测，其构建的预测模型对桂林市的旅游客流量有较好的预测效果。基于以上对梯度提升树模型应用的研读可以发现：梯度提升回归树算法建立的预测模型有较好的预测精度。因此，本文提出基于梯度提升回归树模型对二手房均价进行预测和分析。

2. 研究方法

2.1. 梯度提升回归树算法简介

梯度提升是建立预测模型最好的机器学习方法之一。梯度提升回归树(Gradient Boosting Regression Trees, GBRT)算法[5]是有监督的集成学习方法[11]。提升树是利用加法模型和前向分布算法实现模型的优化，即通过迭代一系列的弱学习器，并通过不同的组合方法得到相应的强学习器。只要得到的强分类器的预测效果比之前一系列弱学习器的预测效果好，其学习就能达到比较好的预测精度。而迭代的目的是找到一系列的弱分类器使得损失函数达到最小，因此对于损失函数的度量就是问题的关键所在，但对于一般的损失函数其优化比较困难，且没有通用的拟合方法。针对这一问题，Freidman 等提出了梯度提升算法，即利用损失函数的负梯度在当前模型的值作为回归问题中提升树算法的残差近似值，这就是 GBRT 算法。

2.2. 梯度提升回归树原始模型

给定数据训练集 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$，最大的迭代次数为树的个数 M，损失函数为 L。

(1) 初始化弱分类器，根据下列公式估计使损失函数极小化的常数 c，

$$f_0(x)=\arg\min_c\sum_{i=1}^{N}L(y_i,c).$$

(2) 迭代次数为 m，$m=1,2,\cdots,M$；

a) 对于样本 $n=1,2,\cdots,N$，根据下列公式计算损失函数的负梯度在当前模型的值：

$$r_{mi}=-\left[\frac{\partial L(y,f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}.$$

b) 对 r_{mi} 拟合一个回归树，得到第 t 棵数的叶子结点的区域为：R_{mj}，$j=1,2,\cdots,J$，J 为第 t 棵回归树的叶子结点的个数；

c) 对第 t 棵回归树的叶子结点区域 $j=1,2,\cdots,J$ 计算，使得损失函数极小化：

$$c_{mj}=\arg\min_c\sum_{x_i\in R_{mj}}L(y_i,f_{m-1}(x_i)+c),$$

c_{mj} 是 R_{mj} 的平方损失最小值。

d) 更新回归树

$$f_m(x)=f_{m-1}(x)+\sum_{j=1}^{J}c_{mj}I\ (x\in R_{mj}),$$

上式中：I 为指示函数，当 x 属于叶子区域 R_{mj} 时为 1, 否则为 0；

(3) 综上所述，得到梯度提升树的最终模型：

$$\hat{f}(x)=f_M(x)=f_0(x)+\rho\sum_{m=1}^{M}\sum_{j=1}^{J}c_{mj}I\ (x\in R_{mj}),$$

ρ 为学习率。

3. 案例分析

随着国家对房地产行业的调控，二手房的市场占比也在不断攀升，房屋价格的变动越来越受到人们的关注，因此采用不同的算法展开对房屋价格预测的研究非常有必要。而上海市房地产市场作为我国制度最完善、最具有代表性的房地产市场之一，对于我国目前房地产市场的合理性，具有极大的代表性。

本文通过收集“链家”上海市近三年二手房的售卖信息，分析影响上海市二手房价格的原因[12]。影响房屋价格的原因有许多方面诸如：位置、交通、配套设施、空间布局、建筑面积、朝向、梯户等，探究二手房房价有助于解决市场信息不对称，使二手房定价趋于合理价位，降低交易成本。因此本文基于上海市二手房数据利用梯度提升回算法进行建模，随机的将数据集的75%划分为训练集，剩下的25%为测试集，对二手房均价进行预测。

3.1. 数据来源、预处理

本文以上海市二手房数据作为分析对象[13]，通过 Python 网络爬虫的方法采集了上海市 2017~2020 年 15 个行政区共 38,900 个小区的二手房数据。将收集到的数据中缺失和有明显的错误的数据进行剔除和填补之后，剩下共 37,483 条有效数据。收集到的数据包括总价、户型、建筑面积、建筑类型、户型结构、均价、配备电梯、装修情况等共 16 个因素。预处理后的数据变量还是比较多，且很多变量之间存在严重的多重共线性，直接用预处理之后的数据进行模型的构建难以产生满意的模型。因此，在建模之前需要对多个变量进行筛选，将二手房房屋均价作为因变量 y，将剩余的其余变量作为自变量 x。

对变量进行相关性分析，计算各变量之间的 Pearson 相关系数，其计算公式如下：

$$r_{xy}=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-\left(\sum x_i\right)^2}\sqrt{n\sum y_i^2-\left(\sum y_i\right)^2}},$$

n 为观测对象的数量，x_i 为 x 的第 i 个观测值，y_i 为 y 的第 i 个观测值，r_{xy} 为变量 x、y 的 Pearson 相关系数。r 的取值范围为[−1, 1]，若 $0<r\le 1$，变量之间存在正线性相关关系；若 $-1\le r<0$，变量间存在负线性相关关系；若 $r=0$ 则变量之间不存在线性相关关系。

3.2. 变量的选择

本文变量比较多，首先挑选可能存在多重共线性的变量计算相关性系数矩阵，然后根据相关性系数矩阵作出热力图。多次绘制热力图，找出各热力图中存在高度相关性的变量将其从变量之中剔除，热力图如图 1 所示。

根据 Pearson 相关系数原理知道，相关系数绝对值越大相关性越强。通过对比各热力图发现房屋的建筑面积和房屋的房间分布存在严重的多重共线性，即只需将房屋的建筑面积纳入模型；房屋的总层数和电梯配备之间存在严重的多重共线性，只用将房屋的总层数纳入本文的模型之中。依此类推，最终挑选出了：房屋建筑面积、小区所在行政区划、户型结构、所处楼层，总层数作为二手房房屋均价的影响因素。通过热力图及相关系数矩阵的计算可以发现房屋建筑面积、小区所在行政区划这两个变量对房屋均价的影响是最大的两个变量，户型结构、所处楼层、总层数是相对于其他变量来说较大的影响变量。另外，从小区所在行政疗、商区划还可以看到房屋周边的环境以及交通、医业等基础配套设施情况。

3.3. 房屋均价的分布情况

首先，先对上海市 15 个行政区的二手房房屋均价进行排序，其结果如下表 1 所示。从表 1 可以看到黄浦区的均价最高为：98,251.1382 元/平方米，最低的均价为金山区：20,291.667 元/平方米。这与上海公

布的中心城区名单：黄浦、静安、徐汇、长宁、虹口、杨浦、普陀以及浦东新区的外环内城区相符合，即房屋均价与中心城区的分布十分的吻合。

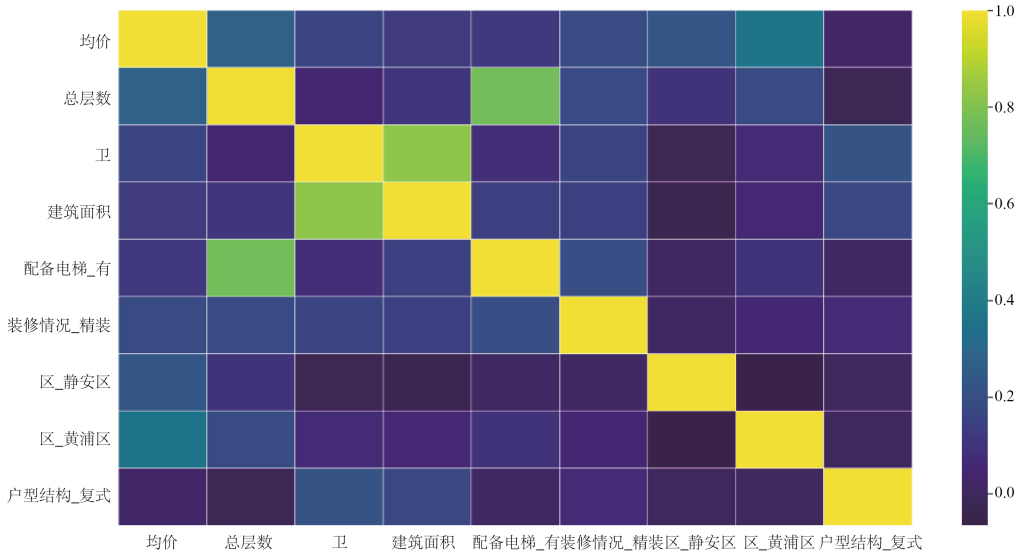

Figure 1. Heatmap
图 1. 热力图

Table 1. The average price of each administrative region and its ranking table
表 1. 各行政区内房价的平均值及其排序表

排序	行政区	房屋均价的平均值(元/平方米)	排序	行政区	房屋均价的平均值(元/平方米)
1	黄浦区	98,251.1382	9	闵行区	51,128.7002
2	徐汇区	76,707.3964	10	宝山区	44,035.8198
3	静安区	76,512.7462	11	松江区	37,983.9645
4	长宁区	71,410.9732	12	青浦区	36,679.5868
5	虹口区	66,474.3184	13	嘉定区	33,461.1735
6	杨浦区	63,642.8784	14	奉贤区	23,494.7114
7	普陀区	60,907.6722	15	金山区	20,291.667
8	浦东新区	56,708.9189			

由表 1 可以得出上海市二手房最高均价为：319,960.62 元/平方米，最低均价为：10,000.00 元/平方米，总体的平均均价为：56,466.26 元/平方米，中位数均价为：53,290.82 元/平方米，上海二手房最高价格和最低价格极差较大。作出上海市二手房房屋均价的整体分布图形，其结果如下图 2 所示。从图 2 可以看出数据分布类似于正态分布，呈现单峰特征，平均值和中位数十分接近峰值。

3.4. 单个因素对于房屋均价的影响

本文挑选的五个变量：房屋建筑面积、小区所在行政区划、户型结构、所处楼层、总层数中，房屋建筑面积、小区所在行政区划对上海市二手房均价的影响是最大的，先单独分析这两个因素的影响以便于后期模型的构建。

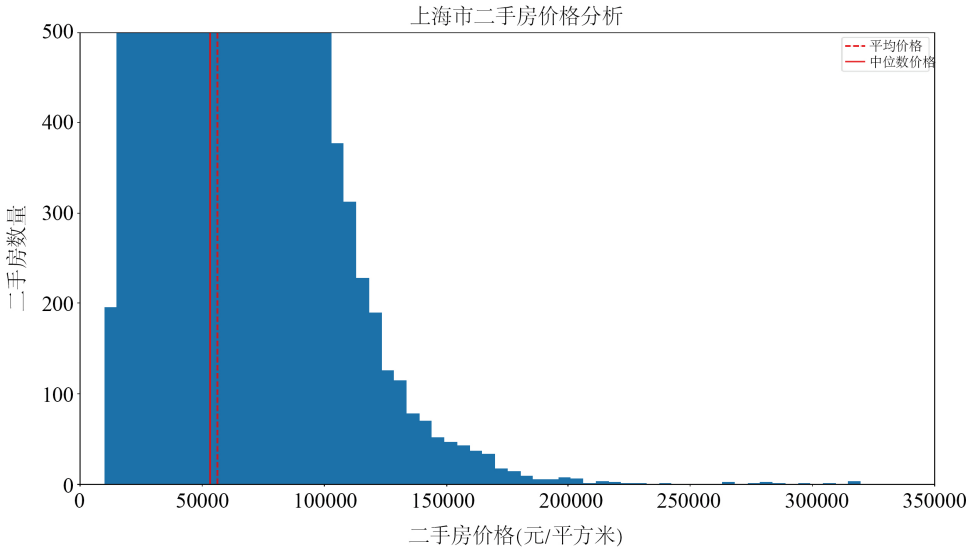

Figure 2. Price analysis of secondary housing in Shanghai
图 2. 上海市二手房价格分析

3.4.1. 行政区划对房屋均价的影响

首先分析各个行政区划内房屋均价的分布情况，作出各个行政区划内的房屋均价分布的箱形图，如下图 3 所示。箱形图的优点是：不受异常值的影响，可以以一种相对稳定的方式描述数据的分布情况；能够观察到数据异常值的分布。从下图 3 可以看到各个行政区划房屋均价的最小值、最大值、上下四分位值以及中位数的值。从下图 3 中看到除了上海的嘉定、金山区以外的其他区域，均具有较多的上侧异常值。这是因为嘉定和金山区是属于上海的周边区域，房价分布区域较小且异常值较少。而其他区域所处的位置比较接近上海市的中心城区，因此有较多的异常值存在。

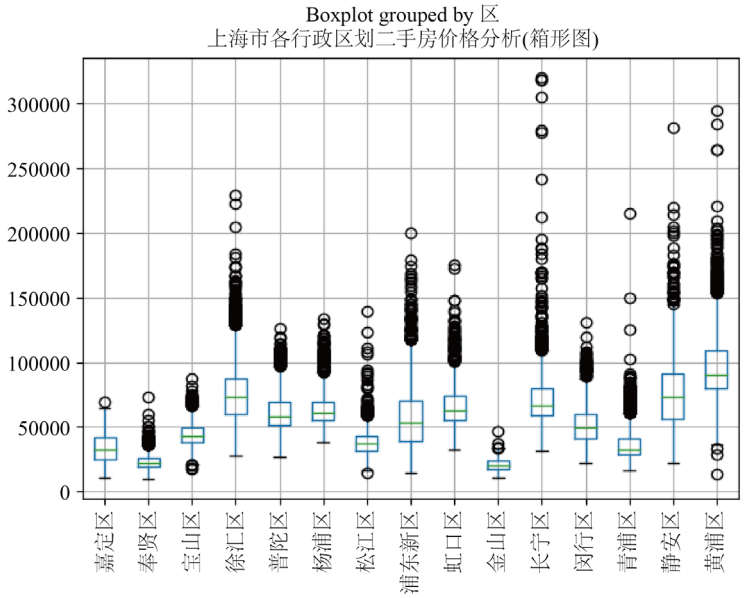

Figure 3. Box plot about Shanghai' each region of second-hand house average price
图 3. 按上海市各行政区划分的二手房均价分析箱形图

3.4.2. 房屋建筑面积对房屋均价的影响

最直观的观察房屋建筑面积对房屋均价的影响是将处理后的数据作散点图，从下图 4 可以大致的观察二者之间的关系。图 4 显示房屋的建筑面积大都集中在 500 平方米以内，数据的分布比较集中，难以看出明显的数据分布规律。

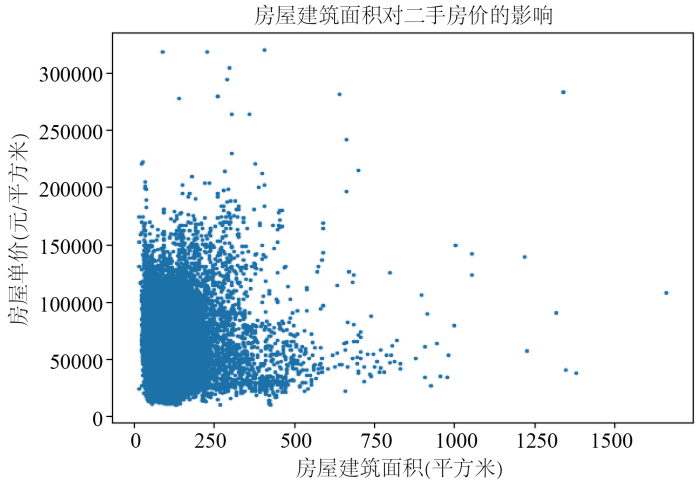

Figure 4. Scatter plot between floor area and house price
图 4. 房屋建筑面积与房价之间的散点图

我们考虑将行政区划和房屋建筑面积结合起来观察，考虑各个行政区划内不同房屋建筑面积对于房屋均价的影响是如何变化的。

3.5. 两个变量联合对于房屋均价的影响

在不同的行政区划内，不同的房屋建筑面积对于房屋均价的影响应该是不一样的。首先作出散点图观察二者对于房屋均价的影响，在作散点图之前，我们要先给不同的行政区划做不同的分类。由上表 1 可以将行政区划分为四个梯队：第一个红色梯队对应房屋均价格最高的四个行政区划；第二个蓝色梯队对应房屋均价次之的四个行政区划；第三个绿色梯队对应房屋均价再次的四个行政区划；最后剩下的最低的房屋均价的三个行政区划作为最后一个灰色梯队。作出的散点图如下图 5 所示。

从下图 5 可以看到，第一个梯队的红色散点图偏向左上方，价格明显比剩下的三个梯队更高一点，每一个梯队相对于上一个梯队都整体稍向下偏移，这与图 4 的散点图分布是相吻合的。

为了更加直观地观察两个联合因素对于房屋均价的影响，考虑对各区的散点图进行最小二乘拟合。依然用前面对各个行政区划的梯度划分，其结果如下图 6 所示。

从下图 6 分析来看，房屋价格随房屋建筑面积的变化规律与预期的相吻合，即房屋均价越高的行政区划内，房屋均价随房屋建筑面积变化的规律越明显，其回归的斜率也更高。

4. 模型测试及结果分析对比

4.1. 变量转换及性能评价指标

在构建模型之前，先对为文字描述的行政区划进行处理。本文使用 One-Hot 编码，是将分类变量作为二进制向量表示，对行政区划的特征进行转化。One-Hot 编码，又称为一位有效编码，主要是采用 N 位状态寄存器来对 N 个状态进行编码，每个状态都由他独立的寄存器位，并且在任意时候只有一位有效。

处理行政区划变量之后，对数据进行分割，随机采样 25%作为测试样本，75%作为训练样本。

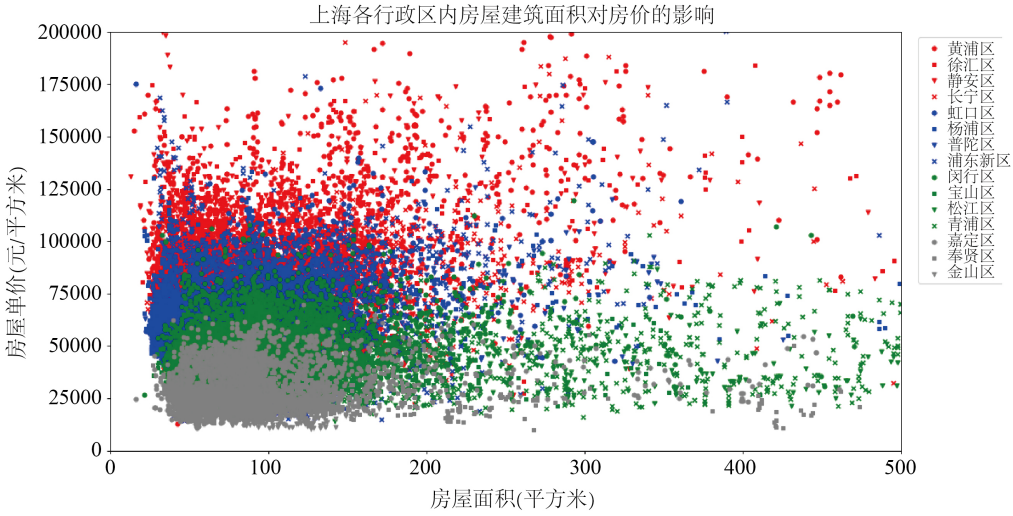

Figure 5. Scatter chart of building floor area on house price in each district of Shanghai
图 5. 上海各行政区内房屋建筑面积对房价的影响的散点图

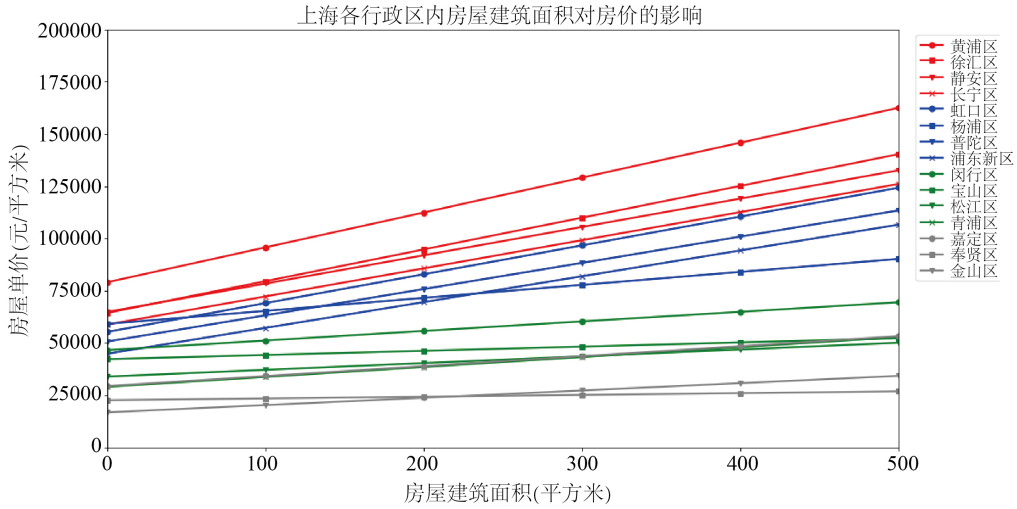

Figure 6. The linear fitting chart of building floor area on house price in each region of Shanghai
图 6. 上海各行政区内房屋建筑面积对房价的影响的线性拟合图

本文选取的模型评价指标为：均方误差 MSE (Mean Squared Error) [14]和拟合优度 R^2 (R-squared) [14]，指标计算如下公式所示：

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2.$$

$$R^2 = \frac{SSR}{SST} = \frac{\sum_i \left(\hat{y}_i - \bar{y}\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2}.$$

拟合优度的取值范围为[0, 1]，其值越接近 1 说明自变量对因变量方差变化的解释越好，值越接近于

0 效果越差。均方误差 MSE (Mean Squared Error)，用于评估预测的结果和真实的数据集之间的接近程度，其值越小拟合效果越好。

4.2. 参数优化

GBRT 算法中有两个重要的参数：一个是学习率(learning_rate)，即弱学习器的权重缩减系数，也称作步长。若学习率太小，意味着需要更多的弱学习器进行迭代，会导致迭代次数大幅度增加，学习的时间也会增大；另外一个是树深度(n_estimators)即迭代次数，一般树深度太小，越容易欠拟合，太大容易过拟合。且学习率和迭代次数相互作用，因此在调参的过程中，将学习率和树深度结合考虑。当学习率的值较小的时候，需要更高数量的树深度，使训练误差收敛。本文对通过对两个参数的排列组合，最后将学习率和树深度分别设置为 0.2 和 1000，其他参数为默认参数设置，得到最终的 GBRT 训练模型。

4.3. 结果分析对比

为了更客观的评估 GBRT 模型[15]，本文将基于集成模型的 GBRT 模型预测结果与线性回归模型和基于支持向量机(SVM)模型的预测结果进行对比分析，得到各模型在测试集上的评价指标如表 2 所示。表 2 显示基于集成模型的预测结果优于基于支持向量机(SVM)和线性回归模型，且基于集成模型的 GBRT 的各项评价指标达到最好的结果，其拟合优度最大达到 0.831，其均方误差 MSE (Mean Squared Error)最小。

Table 2. The prediction accuracy of each model in the test set
表 2. 各模型在测试集上的预测精度

线性回归模型	支持向量机			集成模型		
	线性核	多项式核	径向基核	普通随机森林	终端随机森林	梯度提升树
1.60E+05	1.86E+05	7.18E+04	2.11E+05	6.87E+04	6.82E+04	3.34E+04
0.753	0.712	0.889	0.674	0.895	0.895	0.948

由表 2 中可以看到基于支持向量机(SVM)的径向基核函数的预测模型的效果是最差的，其拟合优度值是最小的且其均方误差(MSE)是最大的。对预测效果最好的提升树集成模型进行拟合，随机挑选 100 个数据，查看其拟合效果。结果如下图 7 所示：

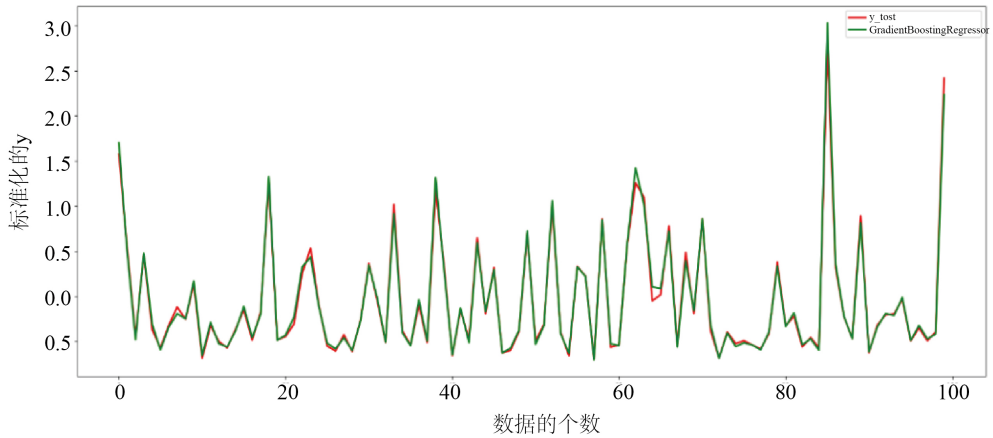

Figure 7. The fitting effect of Gradient Boosting Regression Tree
图 7. 梯度提升回归树的拟合效果

由梯度提升树集成模型的拟合效果图可以看到，其拟合曲线基本拟合真实值曲线。证明梯度提升回归模型建立合理，可以对上海市二手房均价进行预测。

5. 结论

近十年以来房价保持着不断增长的趋势，为购房者带来了巨大的购房压力。本文对爬取的上海市二手房数据进行了分析，采用梯度提升树回归集成模型对二手房均价进行建模，对比机器学习的其他模型，本文运用的梯度提升回归树模型对上海市二手房均价有较好的预测效果。通过对数据集分割为训练集和测试集，并测试集对预测效果最好的梯度提升树回归集成模型，进行拟合评估分析。评估的结果表明，GBRT 模型的预测效果较好[16]，有一定的稳定性，且对上海市二手房均价的预测具有一定的实用性，可以为购买二手房的购房者提供房价影响因素的分析。但是基于梯度提升回归树模型对于上海市二手房均价的分析中还有许多需要改进和完善的地方，首先是模型的泛化能力并不高，因为不同城市、不同时间其价格的变动都不能用一个通用的模型来表示，因此需要对每个特定的城市及历史阶段构建不同的模型来进行预测；其次是本文仅考虑微观层面对于房屋均价的影响，并没有将宏观层次的相关政策措施纳入模型，可以在之后的研究中将宏观层面的政策加入到研究当中进行分析。

基金项目

本文获得国家自然科学基金资助项目(No. 713713516；[12061020]，[71961003])；贵州省科技基金项目(No. 20201y284，No. 20205016，No. 2021088)；贵州大学基金项目(No. 201405，No. 201811)资助。

参考文献

[1] 邬嘉怡, 王思玉, 史宏炜, 李虎森, 楼凯达, 崔丽鸿. 基于多小波的北京市房屋市场价格的分析预测[J]. 北京化工大学学报(自然科学版), 2019, 46(5): 101-106.

[2] 唐晓彬, 张瑞, 刘立新. 基于蝙蝠算法 SVR 模型的北京市二手房价预测研究[J]. 统计研究, 2018, 35(11): 71-81.

[3] 白丽娟, 闫相斌, 金家华. 基于搜索关键词关注度的商品房价格指数预测[J]. 预测, 2015, 34(4): 65-70.

[4] 陆丽丽, 胡斌, 李辉, 端木怡婷. 中国房价构成与预测的仿真分析[J]. 计算机仿真, 2014, 31(3): 230-238.

[5] 申瑞娜, 曹昶, 樊重俊. 基于主成分分析的支持向量机模型对上海房价的预测研究[J]. 数学的实践与认识, 2013, 43(23): 11-16.

[6] 谷秀娟, 李超. 基于马尔科夫链的房价预测研究[J]. 消费经济, 2012, 28(5): 40-42+48.

[7] 付莲莲, 伍健. 基于梯度提升回归模型的生猪价格预测[J]. 计算机仿真, 2020, 37(1): 347-350.

[8] 李一蜚, 秦凯, 李丁, 樊文智, 何秦. 基于梯度提升回归树算法的地面臭氧浓度估算[J]. 中国环境科学, 2020, 40(3): 997-1007.

[9] 杨文忠, 张志豪, 吾守尔·斯拉木, 温杰彬, 富雅玲, 王丽花, 王婷. 基于时间序列关系的 GBRT 交通事故预测模型[J]. 电子科技大学学报, 2020, 49(4): 615-621.

[10] 康传利, 顾峻峰, 刘兆威. 梯度提升回归树的旅游流量预测模型[J]. 数学的实践与认识, 2019, 49(15): 251-261.

[11] Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, **29**, 1189-1232. https://doi.org/10.1214/aos/1013203451

[12] Seon, J. and Yang, H.-S. (2019) A Study on Prediction of Housing Price Using Deep Learning. *Residential Environment*: *Journal of the Residential Environment Institute of Korea*, **17**, 37-49. https://doi.org/10.22313/reik.2019.17.2.37

[13] Daradi, S.A.M., Yusof, U.K. and Kader, N.I.B.A. (2018) Prediction of Housing Price Index in Malaysia Using Optimized Artificial Neural Network. *Advanced Science Letters*, **24**, 1307-1311. https://doi.org/10.1166/asl.2018.10738

[14] 贾俊平, 何晓群, 金勇进. 统计学(第六版) [M]. 北京: 中国人民大学出版社, 2015.

[15] 王琴英. 北京房价与 CPI 的波动特性分析及趋势预测——基于协整关系的 GARCH 族模型分析[J]. 价格理论与实践, 2011(7): 57-58.

[16] Hashem, S.S., Barat, G. and Mohsen, N. (2021) Prediction of Higher Heating Value of Biomass materials Based on Proximate Analysis Using Gradient Boosted Regression Trees Method. *Energy Sources*, *Part A*: *Recovery*, *Utilization*, *and Environmental Effects*, **43**, 672-681. https://doi.org/10.1080/15567036.2019.1630521