

## 基于位置子市场划分的房价贝叶斯概率模型

秦心静, 章平, 张新杨

安徽工程大学 计算机与信息学院, 安徽 芜湖 241000

**摘要:**针对特征价格模型(HPM)在面对房价与特征复杂关系时容易出现预测精度和可解释性不足的情况,提出一种基于子市场效应的贝叶斯概率模型。在改进算法设计时,首先借鉴子市场聚类思想,引入一个潜在变量表示子市场,依据位置邻近性和可替代性建立子市场标准;其次,将子市场标准和特征价格模型作为贝叶斯网络的概率依赖确定各子市场效应的范围,完成子市场划分;最后,依据房屋所属子市场的概率预测房价,且分析子市场的关键影响因素,以提升预测精度和可解释性。将模型与 5 个现有模型从平均绝对百分比误差、平均绝对误差和均方根误差 3 个方面对比;根据杭州市 2019 年之前的房产数据,分别测试非子市场模型与子市场模型的算法性能。实验表明:该贝叶斯模型对房地产价格预测精度优于对比模型,且具有可解释性的优点。

**关键词:**房价评估;HPM 模型;概率模型;位置特征;子市场

中图分类号:TP391 文献标识码:A doi:10.16055/j.issn.1672-058X.2023.0005.011

### Bayesian Probability Model for Real Estate Price Based on Location Submarket Segmentation

QIN Xinjing, ZHANG Ping, ZHANG Xinyang

School of Computer and Information, Anhui Polytechnic University, Anhui Wuhu 241000, China

**Abstract:** A Bayesian probability model based on sub-market effects was proposed to address the situation that the hedonic price model (HPM) is prone to insufficient prediction accuracy and interpretability in the face of the complex relationship between house prices and characteristics. In improving the algorithm design, the idea of sub-market clustering was borrowed, a latent variable was introduced to represent sub-markets, and sub-market criteria were established based on location proximity and substitutability. Next, sub-market criteria and hedonic price models were used as probabilistic dependencies of Bayesian networks to determine the range of effects in each sub-market, completing the sub-market segmentation. Finally, house prices were predicted based on the probabilities of the submarkets to which the houses belonged, and the key influencing factors of the submarkets were analyzed to improve the prediction accuracy and interpretability. The model was compared with five existing models in terms of mean absolute percentage error, mean absolute error, and root mean square error. The performance of the algorithms of the non-submarket model and the submarket model were tested separately based on property data of Hangzhou City before 2019. The experiments show that the Bayesian model outperforms the comparison models in terms of accuracy in forecasting real estate prices and has the advantage of interpretability.

**Keywords:** house price assessment; Hedonic Price Model (HPM); probability model; location characteristics; submarket

收稿日期:2022-09-20 修回日期:2022-10-26 文章编号:1672-058X(2023)05-0081-08

基金项目:安徽省自然科学基金项目(2108085QF264, 2108085QF268);安徽工程大学校级科研项目(XJKY2022154)。

作者简介:秦心静(1997—),女,安徽宣城人,硕士研究生,从事房产评估研究。

通讯作者:章平(1982—),男,安徽芜湖人,博士,副教授,硕士生导师,从事无线定位、城市计算、物联网研究。Email:pingzhang@ahpu.edu.cn。

引用格式:秦心静,章平,张新杨.基于位置子市场划分的房价贝叶斯概率模型[J].重庆工商大学学报(自然科学版),2023,40(5):81-88.

QIN Xinjing, ZHANG Ping, ZHANG Xinyang. Bayesian probability model for real estate price based on location submarket segmentation[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2023, 40(5): 81-88.

## 1 引言

现代社会,房地产已经成为人民生活和国家经济最重要的话题之一。合理评估房屋特征对房地产价格的影响,对城市规划和社会生活具有重要意义:为住房购买者提供市场价格信息、房市趋势信息,协助其进行购房选择<sup>[1]</sup>;为房地产投资者提供市场划分信息、房价影响因素权重,协助其进行投资决策<sup>[2]</sup>;为城市规划者提供住房周边资源变化所能产生的价值信息,量化不同城市规划决策带来的影响<sup>[3]</sup>。

特征价格模型(Hedonic Price Model, HPM)<sup>[4]</sup>是一种评估房屋特征对房地产价格影响的方法。该模型将房屋的各种特征属性作为自变量,房地产价格作为因变量,建立回归方程,量化各属性对房价的影响。但是,由于房屋特征与房价关系复杂,现有的回归方程在房价预测精度和可解释性方面仍有不足。

为了提高房价预测精度,很多学者都进行了深入研究。文献[5,6]采用机器学习与 HPM 相结合的方法来降低评估误差;文献[7,8]通过改变数据特征来提升精度。在结合使用机器学习方面,Yan 等<sup>[5]</sup>在评估中使用遗传算法来优化神经网络模型,通过遗传神经网络选择最合适的权重和阈值,有效降低了评估误差;文献[6]发现很少有应用程序侧重于提高房价预测能力,他们通过对比基于决策树的不同集成方法(如 Bagging、Boosting、随机森林等),为整个西班牙市场开发了一个应用程序,实现完全自动为每个城市提供最佳预测模型。在改变数据特征方面,Chanasit 等<sup>[7]</sup>在改进的 Garson 算法中结合升压策略和输入灵敏度分析来进行特征选择,通过 ANN 模型迭代调整选择的标准,保证每个信息特征的灵敏度系数;Xu 等<sup>[8]</sup>利用信息增益比方法对特征进行加权、排序,在有限混合模型的基础上建立回归模型。

但在提高房价预测精度时,部分评估方法往往会存在一些可解释性方面的不足。针对这个情况,Wu 等<sup>[9]</sup>通过探讨空间连续性对住房子市场分类的作用,提出一种基于主成分分析和聚类分析的数据驱动模型,利用房屋地理位置,细分整个空间的子市场;在此基础上,Liu 等<sup>[10]</sup>基于建筑环境和房屋基础特征对整个子市场建模,利用子市场效应提升房价预测精度,并提出概率层次聚类方法推断市场层次结构。

虽然已有研究对预测精度和可解释性进行提升,但在特征选择和预测结果分析方面仍有改进空间。对此,基于子市场效应的研究,本文采用贝叶斯概率框架,根据房屋位置划分子市场,建立各子市场内的房地产特征价格模型,最终利用贝叶斯估计提升房价预测精度。主要贡献如下:

(1) 引入表示子市场划分的潜在变量,依据空间邻近性和可替代性迭代优化子市场划分结果,获得空间任一位置属于不同子市场的概率分布。

(2) 分析子市场内部影响房价因素排名,对比了不同子市场房价形成机制差异性,提高模型可解释性。

(3) 对比子市场的划分结果和地理行政区划的关系,证实地理行政区划对房价形成机制的重大影响。

## 2 房价影响因素

房地产价格主要取决于自身地理特征、附近房产价格及潜在商业区价值<sup>[11]</sup>。国内外学者从房屋基础特征<sup>[12]</sup>、环境特征<sup>[13]</sup>、区位特征<sup>[14]</sup>、社会经济因素<sup>[15]</sup>等方面研究房地产价格的影响因素。本文主要从房屋基础特征和环境特征来构造房屋特征,如表 1 所示。

表 1 房屋特征

Table 1 Attributes of House

类型	特征
房屋位置	经纬度坐标
基础特征	面积、楼层、装修类型、朝向、年份、客厅数量、卧室数量、浴室数量
周边建筑类型	1 km 内 20 个类别的 POI 数量
学校	1.5 km 内公立普通中小学数量
	1.5 km 内公立重点中小学数量
	1.5 km 内私立普通中小学数量
	1.5 km 内私立重点中小学数量
交通设施	1 km 内地铁站数量
	0.5 km 内公交车站数量 0.5 km 内停车场数量

房屋基础特征主要研究房屋自身建筑属性对房屋价格的影响。本文选用的房屋基础特征包括面积、楼层、年份、装修类型、房屋朝向、客厅数量、卧室数量、浴室数量共 8 种。环境特征反映房屋与学校、医院、商场等周边设施的距离对房价的影响,其中,学区房是近年来较为突出的影响因素,为了孩子能接受更好的教育,越来越多的家庭对周边优质学区房产生极高的需求。

除此之外,人们对公共交通出行需求的增加,使得周边交通设施对房价也有一定影响。

本文选用周边建筑类型、学校和附近交通设施等特征刻画环境特征。通过建立特征索引树,计算房屋经纬度坐标与周边建筑经纬度坐标的差值是否小于阈值,以此量化环境特征。所选用的周边建筑类型包括商务办公职能、零售服务职能、餐饮服务职能、教育科研职能、医疗卫生职能等 20 种 POI 类型;学校分为公立、私立、重点、普通 4 种;交通设施包括地铁站、公交站、停车场。

将房屋特征记作  $R = [r_1, r_2, \dots, r_N] \in R^{35 \times N}$ ,  $r_n \in R^{35}$  代表第  $n$  套房屋的基础特征与环境特征,  $N$  为总房屋数量。房屋位置由经纬度坐标表示,记作  $P = [p_1, p_2, \dots, p_N] \in R^{2 \times N}$ ,其中  $p_n \in R^2$  代表第  $n$  套房屋的经纬度位置,房屋观测价格记作  $Y = [y_1, y_2, \dots, y_N]^T \in R^N$ 。

为了对子市场效应建模,引入潜在变量  $s_n \in \{1, 2, \dots, K\}$  指示第  $n$  套房屋所属子市场,即  $s_n = k$  代表第  $n$  套房屋属于第  $k$  个子市场,  $K$  为子市场总个数。为方便表示,记  $S = [s_1, s_2, \dots, s_N]^T \in Z^N$  表示所有房屋子市场归属。本文所使用的符号定义如表 2 所示。

表 2 符号定义

Table 2 Notations definition

符号	维数	描述
$R$	$35 \times N$	房屋特征
$Y$	$N$	房屋价格
$P$	$2 \times N$	房屋位置
$S$	$N$	所属子市场
$K$	1	子市场数量

### 3 贝叶斯概率模型

如图 1 所示,通过房地产交易网站获取房屋特征及位置,包括面积、楼层、装修类型、朝向等基础特征,基于房屋位置利用高德地图兴趣点(Point of Interest, POI)数据获取周边建筑类型、学校和交通设施等环境特征。利用房屋位置与房屋特征构建基于子市场划分的贝叶斯概率模型,最终实现房地产价格预测和子市场内部影响房价因素的分析。

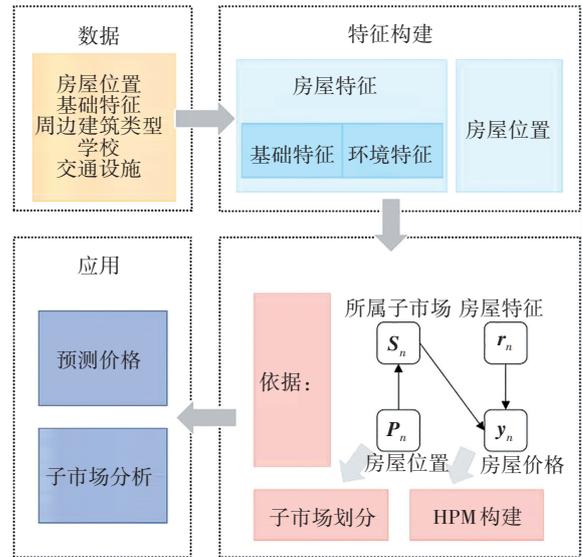


图 1 价格预测与子市场分析流程

Fig. 1 Price forecast and submarket analysis process

#### 3.1 模型假设

本模型提出的贝叶斯网络中,房屋特征、房屋位置以及房屋价格是可观测的,而房屋价格与房屋特征和位置的关系都依赖于子市场。模型流程如图 2 所示。

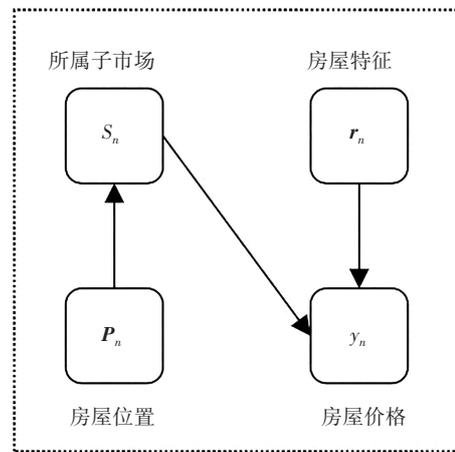


图 2 概率模型关系示意图

Fig. 2 Diagram of probability model relationship

把房屋所属子市场看作随机变量,为了表示房屋属于每个子市场的先验概率,所属子市场  $s_n$  服从参数为  $a$  的多点分布概率向量,  $a \in R^k$  刻画了属于不同子市场房产数量的先验比例。房屋位置与子市场的关系用高斯分布来表示,参数  $\mu_{s_n}$  代表子市场的位置中心点,参数  $\Sigma_{s_n}$  刻画范围大小。房屋价格由所属子市场和房屋特征共同确定,函数  $f_{s_n}(r_n)$  则是子市场内部的特征价格模型,  $\sigma_{s_n}^2$  表示预测价格的浮动范围。

对第  $n$  套住房,  $n \in \{1, 2, \dots, N\}$ , 如式(1)所示:

$$\begin{aligned} s_n &\sim \text{Categorical}(\mathbf{a}) \\ \mathbf{p}_n &\sim N(\mathbf{p} \mid \boldsymbol{\mu}_{s_n}, \boldsymbol{\Sigma}_{s_n}) \\ y_n &\sim N(y \mid f_{s_n}(\mathbf{r}_n), \sigma_{s_n}^2) \end{aligned} \quad (1)$$

此模型中,房屋位置划分子市场的标准主要基于异质性的相关标准,基于贝叶斯网络结构,子市场位置和范围由空间邻近性<sup>[16]</sup>和可替代性决定。空间邻近性表示子市场和房屋位置之间的依赖关系,要求同一子市场内的房屋位置临近,假设每个子市场都存在一个中心点,则高斯分布的特点让属于同一子市场的房屋围绕该中心点分布。可替代性指相同子市场中房屋特征转售机制的相似性,使用 HPM 回归价格与观测价格之间的方差量化可替代性。

总体来讲,房屋价格由房屋特征和所属子市场决定,如图 2 所示。各子市场都有唯一的 HPM 与其对应,本文所采用的 HPM 是梯度推进回归树(GBDT)<sup>[17]</sup>。与线性模型相比,此模型具有相对较低的偏差,并且使用部分依赖的概念,具有良好的解释性。

### 3.2 超参数估计

完成模型构造后,采用经验贝叶斯方法<sup>[18]</sup>进行超参数估计,超参数表示为  $\theta = \{a_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, f_k, \sigma_k^2 \mid k=1, \dots, K\}$ 、参数(潜在变量)  $S$ 、观测数据  $D = \{(\mathbf{p}_n, y_n)\}$ 。给定观测数据和子市场数量  $K$  的情况下,使用最大似然原理估计  $\theta$ 。由式(1)得  $D$  的对数似然如式(2)所示:

$$\begin{aligned} L(\theta) &= \log Pr(D \mid \theta, K) = \\ &\sum_{n=1}^N \log \left( \sum_{k=1}^K Pr(\mathbf{p}_n, y_n \mid s_n = k, \theta, K) \times \right. \\ &\quad \left. Pr(s_n = k \mid \theta, K) \right) \end{aligned} \quad (2)$$

为了最大化似然函数,使用期望最大化算法(EM 算法)评估超参数  $\theta$ ,推导过程如式(3)和式(4)所示:

$$Q(\theta \mid \hat{\theta}^{(t)}) = E_s [\log Pr(D, S \mid \theta) \mid D, \hat{\theta}^{(t)}] = \sum_s Pr(S \mid D, \hat{\theta}^{(t)}) \log Pr(D, S \mid \theta) \quad (3)$$

$$\hat{\theta}^{(t+1)} = \operatorname{argmax}_{\theta} Q(\theta \mid \hat{\theta}^{(t)}) \quad (4)$$

EM 算法的详细推导过程:算法分为两步,期望(E 步)如式(5)所示,利用对超参数的现有估计值,评估所属子市场的后验概率。

$$\gamma(s_n) = Pr(k \mid \mathbf{p}_n, y_n, \theta^{(t)}, K) =$$

$$\frac{Pr(k \mid \theta^{(t)}, K) Pr(\mathbf{p}_n, y_n \mid k, \theta^{(t)}, K)}{\sum_{i=1}^K a_i^{(t)} Pr(y_n \mid i, \theta^{(t)}, K) Pr(\mathbf{p}_n \mid i, \theta^{(t)}, K)} \quad (5)$$

最大化(M 步)如式(6)所示,极大化 E 步求得后验概率更新超参数的值,并用于下一次迭代的 E 步。

$$T_n = \operatorname{argmax}_k \gamma(s_n) \quad (6)$$

再整合第  $k$  个子市场,根据房屋所属子市场,对 HPM 回归更新,即更新  $f_k^{(t+1)}$ 。计算出有效成员数量  $N_k^{(t+1)} = \sum_{n=1}^N \gamma(s_n)$ ,并最大化  $Q(\theta \mid \hat{\theta}^{(t)})$ ,令其求导为零,求得超参数  $\theta$ ,具体的超参数值如式(7)所示。

$$\begin{aligned} a_k^{(t+1)} &= \frac{N_k^{(t+1)}}{N} \\ \boldsymbol{\mu}_k^{(t+1)} &= \frac{1}{N_k^{(t+1)}} \sum_{n=1}^N \gamma(s_n) \mathbf{p}_n \\ \boldsymbol{\Sigma}_k^{(t+1)} &= \frac{1}{N_k^{(t+1)}} \sum_{n=1}^N \gamma(s_n) (\mathbf{p}_n - \boldsymbol{\mu}_k^{(t+1)}) (\mathbf{p}_n - \boldsymbol{\mu}_k^{(t+1)})^T \\ \sigma_{y,k}^{2(t+1)} &= \frac{1}{N_k^{(t+1)}} \sum_{n=1}^N \gamma(s_n) (y_n - f_k^{(t+1)}(\mathbf{r}_n))^2 \end{aligned} \quad (7)$$

当连续两次迭代得到的对数似然值之差小于阈值时,EM 算法迭代终止。算法流程如图 3 所示。

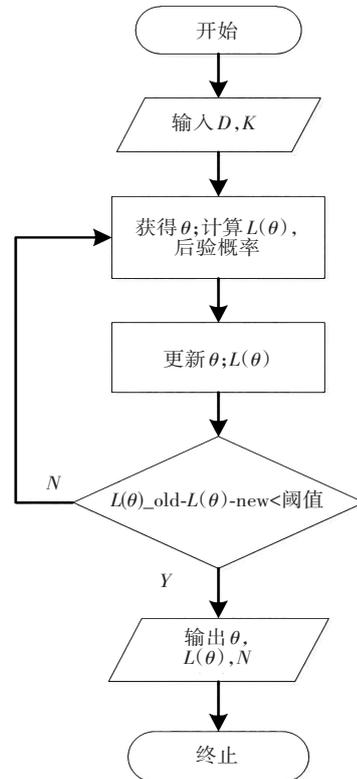


图 3 算法流程图

Fig. 3 Algorithm flowchart

### 3.3 子市场划分与价格预测

求解出最优超参数后,使用贝叶斯平均值作为最终预测价格。具体来说,利用 EM 算法求得的参数  $\theta$ , 计算房屋在每个子市场的后验概率,如式(8)所示:

$$Pr(s = k | p, r, \hat{\theta}) = \frac{Pr(s = k | \hat{\theta}) Pr(p, r | s = k, \hat{\theta})}{\sum_{i=1}^K Pr(s = i | \hat{\theta}) Pr(p, r | s = i, \hat{\theta})} \quad (8)$$

求出该房屋在  $k$  个 HPM 中的  $k$  个子市场的预测价格平均值,如式(9)所示,即为最终预测价格:

$$\tilde{y} = \sum_{k=1}^K \tilde{y}_k Pr(s = k | p, r, \hat{\theta}) \quad (9)$$

再通过该房屋的后验概率值,导出其所属子市场的标签,如式(10)所示:

$$T = \operatorname{argmax}_k Pr(s = k | p, r, \hat{\theta}) \quad (10)$$

## 4 实验

### 4.1 数据获取

选用杭州市二手房房地产市场数据来评估模型, 实验数据集的统计数据如表 3 所示。该数据来自中国最大的在线房地产交易网络之一的 Fang. com, 包括二手房价和 8 种房屋特征。

表 3 数据集、属性、统计

Table 3 Data sets, attributes, statistics

数据集	属性	统计量
二手房数据	房屋数量	42 064
	房屋特征数量	9
POI 数据	类型数量	20
	POI 总数	633 813
交通设施数据	地铁站数量	257
	公交站数量	19 648
	停车场数量	35 704
中小学数据	公立普通学校数量	716
	公立重点学校数量	75
	私立普通学校数量	10
	私立重点学校数量	6

POI 数据、交通设施数据和中小学数据从中国最大的在线定位服务提供商之一的 Amap. com 获取。采集杭州 2019 年及此之前的房地产数据分析,将房地产数据分为训练集、验证集和测试集。随机抽选 60%的数

据为训练集,20%数据作为验证集,剩下 20%数据作为测试集。

### 4.2 子市场数量选择

为了确定子市场数量,合并验证集和训练集进行交叉验证,子市场数量从 2 开始增加到足够大,观察平均绝对百分比误差( $\epsilon_{MAPD}$ )的结果如图 4 所示。结果显示:当子市场数量从 2 增加时,验证集上的  $\epsilon_{MAPD}$  不断减少,直到子市场的数量达到 49 时为肘点。因此,将子市场的数量设置为 49。子市场划分结果如图 5 所示。

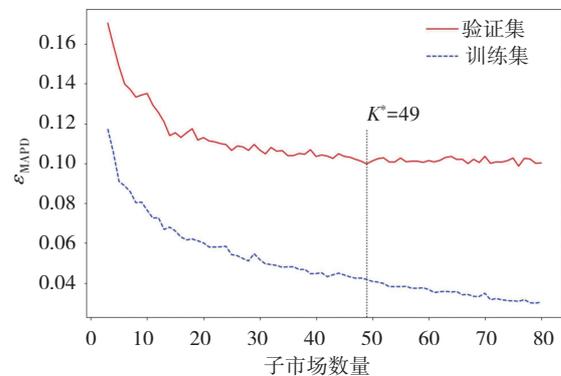


图 4 交叉验证结果

Fig. 4 Cross-verification results

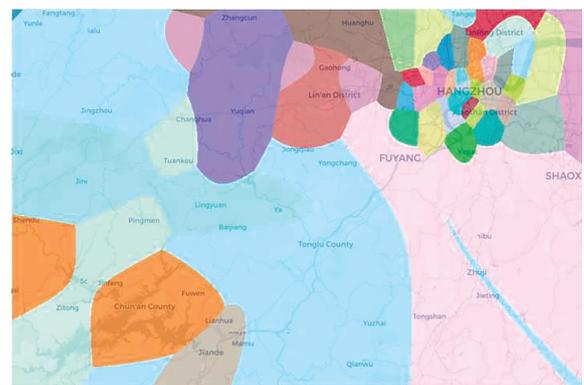


图 5 子市场划分结果展示

Fig. 5 Display of submarket division results

### 4.3 算法性能评估

本文选取平均绝对百分比误差( $\epsilon_{MAPD}$ )作为主要指标评估房价预测的准确性,将原始误差标准化,消除数值过大存在的偏差;平均绝对误差( $\epsilon_{MAE}$ )提供直观的误差度量,反映预测价格与实际价格的误差大小;均方根误差( $\epsilon_{RMSE}$ )反应数据集的离散程度。定义如式(11)所示:

$$\begin{aligned}\varepsilon_{\text{MAPD}} &= \frac{100\%}{n} \sum_{j=1}^n \left| \frac{\hat{y}_j - y_j}{y_j} \right| \\ \varepsilon_{\text{MAE}} &= \frac{1}{n} \sum_{j=1}^n |\hat{y}_j - y_j| \\ \varepsilon_{\text{RMSE}} &= \sqrt{\frac{1}{n} \sum_{j=1}^n (\hat{y}_j - y_j)^2}\end{aligned}\quad (11)$$

为了证明子市场的有效性,选用非子市场模型和子市场模型进行估值比较,评估结果如表 4 所示。

表 4 性能评估对比

类 型	模 型	$\varepsilon_{\text{MAPD}}$	$\varepsilon_{\text{MAE}}$	$\varepsilon_{\text{RMSE}}$	%
非子市场 模型	GBDT	18.13	5 055	7 184	
	SVM	20.99	6 707	11 354	
	ANN	24.30	6 901	9 878	
	OLS	28.30	8 017	11 067	
子市场 模型	K-means+GBDT 本文模型	13.00 10.00	3 807 3 154	6 180 5 272	

梯度提升决策树(GBDT)是一种迭代的决策树算法,由多棵决策树组成,所有树的结论累加起来做最终答案,可以发现多种有区分性的特征以及特征组合。

支持向量机(SVM)是一种二分类模型,目标是找到一个超平面,使用两类数据离超平面越远越好,从而对新的数据分类更准确,可以使用带有径向基函数的SVM来捕捉非线性。

人工神经网络(ANN)是一种非线性模型,可以自动学习有效的特征表示,并提供准确的预测。

普通最小二乘法(OLS)主要用于线性回归的参数估计,通过最小化误差的平方和寻找数据的最佳函数匹配。

K-means 是一种广泛用于识别子市场的聚类方法。位置和房屋属性的连接被用作 K-means 聚类的输入。对于每个已识别的子市场,使用 GBDT 作为模型输入来构建唯一的 HPM。

#### 4.4 子市场影响因素分析

为了证明模型可以有效评估房屋特征对房价的影响,通过 HPM 模型的特征重要性反映房屋特征在房地产市场中的影响程度,本文选取典型子市场和旅游地产进行影响因素分析。

#### 4.4.1 典型子市场分析

依据所划分的 49 个子市场,结合杭州城市规划现状,选取 4 个典型子市场进行分析,区域范围如图 6 所示。表 5 列出了这些子市场的重要影响特征。研究表明,4 个子市场中年份和面积都具有较大作用。

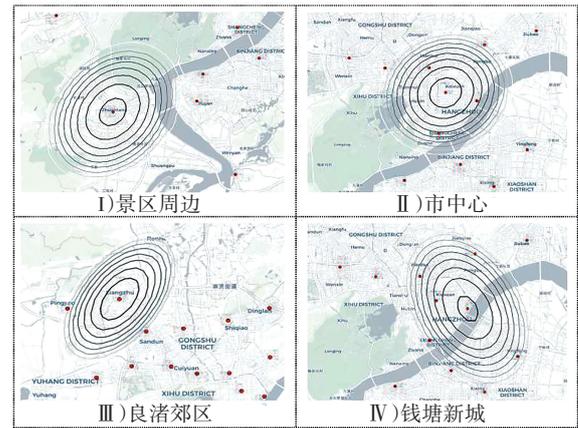


图 6 典型子市场

Fig. 6 Typical submarket

表 5 典型子市场影响特征

子市场	类 型	重要特征
I	转塘	面积、公交数量、年份、浴室数量、房间数量、装修-中等
II	市中心	面积、年份、住宿服务职能、专业市场职能、公交数量、非制造业公司
III	良渚	面积、年份、社会福利职能、浴室数量、教育科研职能、住宿服务职能
IV	钱塘新城	面积、年份、停车场、浴室数量、房间数量、住宿服务职能

子市场 I 为杭州西湖区转塘街道,房价与附近公交数量、浴室数量、房间数量、中等装修有较强相关性,这是由于该子市场位于市区,交通便利且邻近西湖、灵山、钱塘江等诸多景区,居民对住宅舒适性有较强的需求。子市场 II 为杭州的上城区,主要影响因素为面积、年份、住宿服务职能、专业市场职能、公交数量、非制造业公司,这是因为市中心居民对出行和办公的相关职能需求较大。子市场 III 为杭州的良渚镇,位于杭州余杭区的郊区,其作为良渚文化旅游中心,将教育科研职能和住宿服务职能视为重要特征。子市场 IV 为杭州钱塘新城新区,是政府政策支持的一个新商务区,与其一

致,商务区住房将停车场数量、浴室数量、房间数量及住宿服务职能视为重要特征,因为位于城区的办公区域,对停车场数量的需求较大,且周边学校数量以及环境设施较为齐全,所以影响该子市场的因素主要为房屋基础特征。

#### 4.4.2 旅游地产分析

杭州拥有数量众多的创业企业,经济发展快速,也是中国著名的风景旅游城市,拥有“三江两湖一山”的旅游资源,有较好的经济文化环境发展旅游地产<sup>[19]</sup>。因此,本文选取 4 个旅游区进行子市场分析,区域范围如图 7 所示。表 6 列出了这些子市场的重要影响特征。

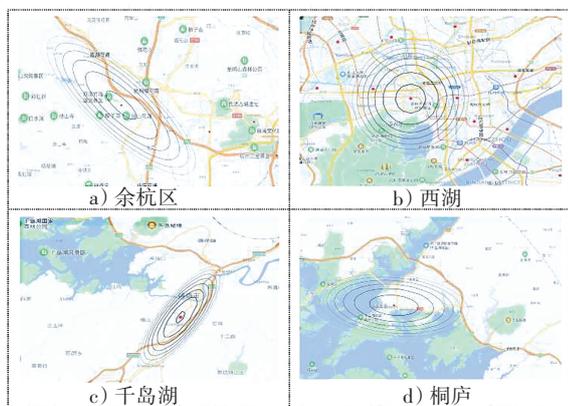


图 7 旅游区子市场

Fig. 7 Tourist area market

表 6 旅游区子市场影响特征

Table 6 Travel area market influence characteristics

子市场	类型	重要特征
a	余杭区	面积、年份、公交、教育科研职能、公立普通学校数量、朝向-南
b	西湖	面积、教育科研职能、年份、娱乐康体职能、住宿服务职能、停车场数量
c	千岛湖	面积、年份、浴室数量、客厅数量、产业园区、装修-精装
d	瑶林	零售服务职能、住宿服务职能、餐饮服务职能、其他生活服务、文化艺术职能、金融保险职能

子市场 a 为杭州余杭郊区,包括径山、双溪竹海、良渚文化等诸多景点,主要影响因素为面积、年份、公交数量、教育科研职能、公立普通学校数量,说明该景点周边的房屋对附近教育资源需求更大。子市场 b 为

位于杭州城区内的西湖景区,其房价与教育科研职能、娱乐康体职能、住宿服务职能以及停车场数量有较强相关性,是因为城区内更注重生活娱乐项目的丰富便利。子市场 c 为千岛湖景区,可以看到该地房价受房屋基础特征影响较大,将浴室数量、客厅数量、装修风格视为重要特征,是因为其位于淳安县,周边具有学校、医疗等环境资源,居民更注重房屋的属性。子市场 d 为桐庐县瑶琳仙境景点,其房价与零售服务职能、住宿服务职能、餐饮服务职能、其他生活服务、文化艺术职能、金融保险职能有很强的相关性,是因为该景点位置偏离县区,居民购房更看重周边建筑职能是否能满足日常生活需要。

研究表明:杭州市景点众多且分布较广,房地产市场受景点及周边环境影响,会形成以旅游区为中心的旅游地产。通过分析典型子市场与旅游区子市场发现,本文子市场划分符合杭州城市区块划分。

## 5 结论

提出一个概率框架评估房地产特征价格的子市场效应,将子市场视为潜在变量,在贝叶斯网络上捕获子市场标准,包括位置邻近性和可替代性;根据子市场标准和特征价格模型作为贝叶斯网络的概率依赖确定子市场效应的范围,以此完成子市场划分,并预测房屋价格;最后,以杭州房地产市场验证模型。实验结果证明:本文概率模型在房地产市场估价上优于其他模型。此外,通过对各子市场影响因素的分析,还证明了本文概率模型量化住房属性对房价影响程度的有效性。

### 参考文献(References):

[1] KUMAR S, TALASILA V, PASUMARTHY R. A novel architecture to identify locations for real estate investment[J]. International Journal of Information Management, 2021, 56: 102012—102018.

[2] SILVA J, FILHO A. Performing hierarchical Bayesian regression to assess the best districts for building new residential real estate developments[C]//2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2020: 2411—2416.

[3] 袁韶华, 汪应宏, 左晓宝, 等. 徐州房地产空间异质分析及对城市规划的启示[J]. 工程经济, 2017, 27(12): 46—51.

YUAN Shao-hua, WANG Ying-hong, ZUO Xiao-bao, et al.

- Spatial heterogeneity analysis of real estate in Xuzhou and the Inspiration for urban planning [J]. *Engineering Economy*, 2017, 27(12): 46—51.
- [4] STEVENSON S. New empirical evidence on heteroscedasticity in hedonic housing models[J]. *Journal of Housing Economics*, 2004, 13(2): 136—153.
- [5] YAN S. Real estate evaluation model based on genetic algorithm optimized neural network[J]. *Data Science Journal*, 2019(18): 1—9.
- [6] ALFARO-NAVARRO J L, CANO E L, ALFARO-CORTES E, et al. A fully automated adjustment of ensemble methods in machine learning for modeling complex real estate systems[J]. *Complexity*, 2020, 2020: 5287263—5287265.
- [7] CHANASIT K, CHUANGSUWANICH E, SUCHATO A, et al. A real estate valuation model using boosted feature selection[J]. *IEEE Access*, 2021(99): 1—10.
- [8] XU X, HUANG Z Y, WU J Y, et al. Finding the key influences on the house price by finite mixture model based on the real estate data in Changchun[C]//*Database Systems for Advanced Applications DASFAA 2019*. Cham: Springer, 2019: 378—382.
- [9] WU C S, SHARMA R. Housing submarket classification: the role of spatial contiguity[J]. *Applied Geography*, 2012, 32(2): 746—756.
- [10] LIU Z C, CAO J, XIE R J, et al. Modeling submarket effect for real estate hedonic valuation: a probabilistic approach[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(7): 2943—2955.
- [11] FU Y J, XIONG H, GE Y, et al. Modeling of geographic dependencies for real estate ranking on site selection[J]. *Acm Transactions on Knowledge Discovery from Data*, 2016, 11(1): 11—20.
- [12] 薛冰, 肖骁, 李京忠, 等. 基于 POI 大数据的老工业区房价影响因素空间分异与实证[J]. *人文地理*, 2019, 34(4): 106—114.
- XUE Bing, XIAO Xiao, LI Jing-zhong, et al. POI-Based analysis on the affecting factors of property price' spatial distribution in the traditional industrial area [J]. *Human Geography*, 2019, 34(4): 106—114.
- [13] BERAWI M A, MIRAJ P, SAROJI P, et al. Impact of rail transit station proximity to commercial property prices: utilizing big data in urban real estate[J]. *Journal of Big Data*, 2020, 7(1): 71—75.
- [14] 钟海玥, 张安录, 蔡银莺. 武汉市南湖景观对周边住宅价值的影响——基于 Hedonic 模型的实证研究[J]. *中国土地科学*, 2009, 23(12): 63—68.
- ZHONG Hai-yue, ZHANG An-lu, CAI Yin-ying. Impacts of the Nanhu Lake in Wuhan city on the price of peripheral houses: empirical research based on hedonic model[J]. *China Land Science*, 2009, 23(12): 63—68.
- [15] YASNITSKY L N, YASNITSKY V L, ALEKSEEV A O. The complex neural network model for mass appraisal and scenario forecasting of the urban real estate market value that adapts itself to space and time [J]. *Complexity*, 2021, 2021: 5392170—5392178.
- [16] GOODMAN A C, THIBODEAU T G. The spatial proximity of metropolitan area housing submarkets[J]. *Real Estate Economics*, 2007, 35(2): 209—232.
- [17] FRIEDMAN J. Greedy function approximation: a gradient boosting machine [J]. *Annals of Statistics*, 2001, 29(5): 1189—1232.
- [18] 李勇. 基于先验的贝叶斯先验选择方法[J]. *重庆工商大学学报(自然科学版)*, 2006, 23(6): 548—550.
- LI Yong. On Bayesian prior selection method based on prior[J]. *Journal of Chongqing Technology and Business University(Natural Science Edition)*, 2006, 23(6): 548—550.
- [19] 汤琦. 旅游地产购买行为影响因素研究——以杭州旅游地产购买者为对象[D]. 杭州: 浙江工业大学, 2013.
- TANG Qi. The influence factors of tourist real estate purchase behavior: the buyers of tourist real estate in Hangzhou as an example[D]. Hangzhou: Zhejiang University of Technology, 2013.

责任编辑:李翠薇