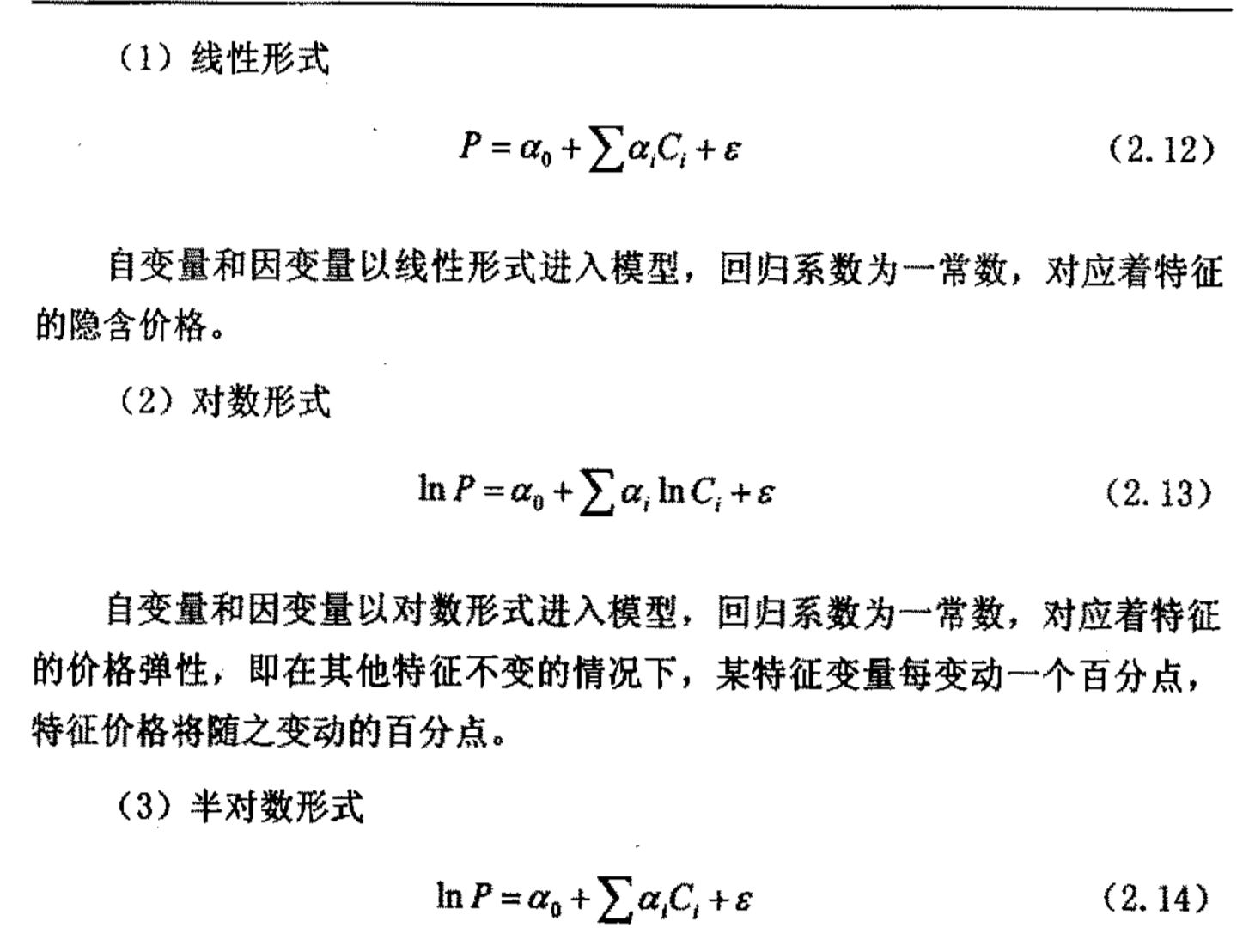
### 传统模型介绍

传统：HEDONIC模型法的基本思路是：将[房地产商品](https://baike.baidu.com/item/%E6%88%BF%E5%9C%B0%E4%BA%A7%E5%95%86%E5%93%81/12751806" \t "/Users/gangzhai/Documents\\x/_blank)的价格分解，以显现出其各项特征的隐含价格，在保持房地产的特征不变的情况下，将[房地产价格](https://baike.baidu.com/item/%E6%88%BF%E5%9C%B0%E4%BA%A7%E4%BB%B7%E6%A0%BC" \t "/Users/gangzhai/Documents\\x/_blank)变动中的特征因素分解，从价格的总变动中逐项剔除特征变动的影响，剩下的便是纯粹由[供求关系](https://baike.baidu.com/item/%E4%BE%9B%E6%B1%82%E5%85%B3%E7%B3%BB" \t "/Users/gangzhai/Documents\\x/_blank)引起的价格变动。

HEDONIC模型法：该理论认为，一种多样性商品具有多方面的不同特征或品质(如房地产商品的面积、楼层、朝向和是否有保安服务等特征)，商品价格则是所有这些特征的综合反映和表现。当商品某一方面的特征改变时，商品的价格也会随之改变。

当商品的某一方面的特征改变时，其价格也会随之发生改变。对函数的各个特征变量分别求[偏导数](https://baike.baidu.com/item/%E5%81%8F%E5%AF%BC%E6%95%B0/5536984" \t "/Users/gangzhai/Documents\\x/_blank)，就取得各特征的变动对商品价格的影响幅度，并假定这种影响的关系在一定的时间内固定不变。这样，在缺乏同质商品的情况下，可以用非同质的房地产在[基期](https://baike.baidu.com/item/%E5%9F%BA%E6%9C%9F" \t "/Users/gangzhai/Documents\\x/_blank)与[报告期](https://baike.baidu.com/item/%E6%8A%A5%E5%91%8A%E6%9C%9F" \t "/Users/gangzhai/Documents\\x/_blank)之间进行比较，从价格的总变动中逐项剔除特征变动的影响，最后剩下的便是纯粹由[供求关系](https://baike.baidu.com/item/%E4%BE%9B%E6%B1%82%E5%85%B3%E7%B3%BB" \t "/Users/gangzhai/Documents\\x/_blank)引起的价格变动了，这样计算的[价格指数](https://baike.baidu.com/item/%E4%BB%B7%E6%A0%BC%E6%8C%87%E6%95%B0" \t "/Users/gangzhai/Documents\\x/_blank)便是基于特征[价格法](https://baike.baidu.com/item/%E4%BB%B7%E6%A0%BC%E6%B3%95/9183383" \t "/Users/gangzhai/Documents\\x/_blank)的[房地产价格指数](https://baike.baidu.com/item/%E6%88%BF%E5%9C%B0%E4%BA%A7%E4%BB%B7%E6%A0%BC%E6%8C%87%E6%95%B0" \t "/Users/gangzhai/Documents\\x/_blank)。



作为传统的统计方法，享乐模型证明了其在假设和估计上的局限性（Park and Bae，2015）。例如，Cebula，R. J.（2009）将享乐主义定价模型应用于佐治亚州萨凡纳市的房地产市场。 Borba和Dentinho（2016）使用享乐价格回归来解释与标定的租金和房屋类型相关的房价，同时评估了特塞拉岛的新房地产价值。这些模型是基于简化的假设而构建的，这些假设是线性的，或者在大多数情况下是房屋租金与解释性因素（它们包括平方项，自然对数等）之间的内在线性关系，而忽略了房地产价格动态的空间性质。 。为了解决这个问题，将考虑协变量和响应变量之间的非线性关系的几种改进方案集成到常规享乐房价模型中。 Giudice，Manganelli和Paola（2017）使用基于惩罚样条平滑的半参数回归估计享乐价格函数，并将价格预测性能与常规参数模型进行了比较。他们的结果表明，半参数模型在住房销售价格的预测上取得了显着改善。类似地，对Montero和Fernández（2018）的研究进一步支持了这样一个论点，即与纯享乐模型相比，考虑到空间异质性的非线性模型是更好的房价预测策略。

尽管基本的统计方法简单明了，但仍存在一些主要限制，包括较弱的预测能力和对先验知识的高度依赖，尤其是在将其用于处理大型数据集和多尺度效应时（Lam，Yu＆Lam，2008）。 。 另一方面，事实证明机器学习模型在处理海量数据方面表现出色，占据了主导地位（de Aquino et al，2019）。 机器学习技术的改进和可用数据的扩散为房地产研究铺平了道路，人们进行了更多尝试，以将传统模型与机器学习方法（例如人工神经网络，SVM和随机森林）进行显式比较。

|  |  |  |
| --- | --- | --- |
| 描述  属性 | 详情 | 备注 |
| 文章信息 | **Chiarazzo, V. , Caggiani, L. , Marinelli, M. and Ottomanelli, M. 2014. A neural network based model for real estate price estimation considering environmental quality of property location. Transportation Research Procedia, 810-817.** |  |
| 输入 | **运输系统和与环境质量有关的属性，还考虑了其他属性，例如建筑物特征和当地土地利用属性** |  |
| 模型 | **ANN，三个隐藏层，1、2层有20个神经元，最后一层一个神经元** |  |
| **输入数据** | **环境属性与土地属性s组成的矩阵** |  |
| **数据集** | **193条，70%训练集、20%验证集，10%测试集** |  |
| **为了得到重要性序列（敏感性分析）** | **重复训练42次，每次减少一个特征，再比较准确度，减少比较小的影响较小，多的反之。** |  |
| **测试数据精度** | **82%** |  |
| **提升猜想** | **提出用聚类的方法提升模型** |  |
| **房屋相关特征** | **PAP是物业要价；**  **SQM是物业的表面积，以平方米为单位；**  **ROOMS是酒店的卧室数；**  **浴室是酒店的浴室数量；**  **如果属性需要重大改进，则IMPR是值为1的虚拟变量；**  **楼层是物业在建筑物中所在的楼层；**  **LIFT是一个虚拟变量，如果该属性所在的建筑物具有电梯（电梯），则其值为1；**  **TAP是一个虚拟变量，如果属性为平面，则值为1；**  **TSAP是一个虚拟变量，如果该属性是一个单一家庭公寓，则值为1；**  **TRH是一个虚拟变量，如果该属性是农村房屋，则值为1；**  **如果属性是两层楼的公寓，TTF是一个虚拟变量，其值为1；**  **如果该物业是单间公寓，则TSF是一个虚拟变量，其值为1；**  **如果属性具有平台，则TE是一个虚拟变量，其值为1；**  **ILVA是使用道路网络从物业到达工业中心的距离（以公里为单位）；**  **GA是一个虚拟变量，如果该属性具有花园，则其值为1；**  **BCH是一个虚拟变量，如果该属性位于距离海滩不到800 m的海滩上，则其值为1。**  **海滩;**  **GAR是一个虚拟变量，如果该属性具有车库，则其值为1；**  **NC是一个虚拟变量，如果该属性是新构造，则其值为1；**  **LTI是服务于该区域的内部线路总线的数量；**  **FTV是一个虚拟变量，如果该属性的公交车站距离其不到400 m，则该变量的值为1**  **该公交车站的服务线路数量；**  **TACCT是在高峰时间从物业到达塔兰托中央商务区的时间，以分钟为单位**  **使用公路网，考虑交通拥堵；**  **TRAIN是一个虚拟变量，如果该属性距离郊区火车站不到500 m，则其值为1；**  **CE是一个虚拟变量，如果该属性位于市中心，则其值为1；**  **DEN是区域内人口密度的度量单位（每单位面积的居民）；**  **HOU是区域家庭数；**  **ENT是区域企业的数量；**  **EMP是区域员工人数；**  **跨界动植物病虫害紧急预防系统是一个区域中的就业人数；**  **INH是一个区域中的居民人数；**  **INHAGG是聚集区内居民的数量；**  **AREA是区域面积；**  **SO2，NOX，NO，NO2，CO，PM10是测得的环境污染者的值（μg/ m3 293K）;**  **SO2\_MAX，NOX\_MAX，NO\_MAX，NO2\_MAX，CO\_MAX，PM10\_MAX是测得的最大值**  **环境污染者（μg/ m3 293°K）。污染物由研究区域内的11个固定空气质量站测量，并给出了污染物的最大值。** |  |

一下是论文阅读笔记：

|  |  |  |
| --- | --- | --- |
| 描述  属性 | 详情 | 备注 |
| 文章信息 | Ziyue Yan.Lu Zong.2020.Spatial Prediction of Housing Prices in Beijing Using Machine Learning Algorithms.ACM ISBN 978-1-4503-7560-3/20/07...$15.00 https://doi.org/10.1145/3409501.3409543 | ACM |
| 目的 | 旨在探索重要的解释特征，并通过结合一系列机器学习技术（包括XGBoost，线性回归，Random Forest Regression，Ridge和Lasso模型，bagging and boosting）来确定在北京实施房价空间预测的精确机制； |  |
| 模型1 | *Hedonic Model* Rosen (1974) |  |
| 模型2 | *Random Forest /* |  |
| 模型3、4 | *Ridge and Lasso Regressions https://www.cnblogs.com/wuliytTaotao/archive/2019/05/11/10837533.html* |  |
| 模型5 | *XGBoost* |  |
| **输入数据** | **36个特征的矩阵** |  |
| **数据集** | **原始房地产数据包括北京各地区房屋的价格，结构和环境特征均来自中国最大的房地产中介机构联嘉有限公司的官方网站。包含26个变量，包括13个结构变量和13个环境变量。除去异常值后，所使用的数据集的样本大小为14,758。** |  |
| **房屋价格数据分布及数据分析** | **成正太分布、用热图分析出重要特征。** |  |
| **结论** | LASSO 分析出了比较重要的特征（用系数值判断）XGBoost与其类似；  发现XGBoost优于其他机器学习模型，这是由于尽管功能稍有不足，但其在调整和预测方面的强大和强大性能所反映。 结果还证明，传统的线性回归方法（即Hedonic模型）处理方法较弱 |  |
| **测试数据精度** |  |  |
| **提升猜想** | **房地产市场具有重要的区域差异（Eichholtz和Schweitzer，2000）**  **结合其他机器学习算法（如人工神经网络（ANN）和支持向量机（SVM））的性能比较，以进一步提高建模性能** |  |
| **房屋相关特征** | |  |  |  |  | | --- | --- | --- | --- | | **Category** | **Variable Name** | **Description** | **Data Type** | | Structural | price | Average price (¥) by square | Numerical | | followers | Number of people follow the transaction | Numerical | | livingRoom | Number of living room | Numerical | | bathroom | Number of bathroom | Numerical | | drawingRo om | Number of drawing Room | Numerical | | floor | Height of the house (meter) | Numerical | | buildingTy pe | Including tower (1) , bungalow (2)， combination of plate and tower (3), plate (4) | Categorical | | constructio nTime | Year of construction | Numerical | | renovation Condition | Including other (1), rough (2),Simplicity (3), hardcover (4) | Categorical | | buildingStr ucture | Including unknown (1), mixed (2), brick and wood (3), brick and concrete (4) ,steel (5) and steel-concrete composite (6) | Categorical | | ladderRatio | Proportion between number of residents on the same floor and number of elevator of ladder. It describes how many ladders a resident have on average. | Numerical | | elevator | Have (1) or not have elevator (0) | Categorical | | total\_floor | Total floors of the house | Numerical | | Locational | subway | whether near the subway near(1) | Categorical | | is\_in\_busin ess\_area | In business area(1) | Categorical | | 500m\_uni\_ cnt | Number of university within 500 meters | Numerical | |  |
|  | |  |  |  |  | | --- | --- | --- | --- | |  | 500m\_midd le\_cnt | Number of middle school within 500 meters | Numerical | | 500m\_mall \_cnt | Number of shopping mall within 500 meters | Numerical | | 500m\_hosp ital\_cnt | Number of shopping mall within 500 meters | Numerical | | 500m\_busi ness\_build\_ cnt | Number of business building within 500 meters | Numerical | | 500m\_attra ction\_cnt | Number of attraction within 500 meters | Numerical | | 250m\_bus\_ cnt | Number of bus station within 250 meters | Numerical | | 500m\_supe rmarket\_cnt | Number of supermarket within 500 meters | Numerical | | 500m\_prim ary\_cnt | Number of primary school within 500 meters | Numerical | | 500m\_kind ergarten\_cn t | Number of kindergarten within 500 meters | Numerical | | 500m\_finan cial\_firm\_c nt | Number of financial firm within 500 meters | Numerical | |  |

|  |  |  |
| --- | --- | --- |
| 描述  属性 | 详情 | 备注 |
| 文章信息 | **Housing Market Prediction Problem using Different Machine Learning Algorithms: A Case Study.2020**  **arXiv** | 处理数据比较好 |
| 输入 | **运输系统和与环境质量有关的属性，还考虑了其他属性，例如建筑物特征和当地土地利用属性** |  |
| 模型 | **XGBoost，CatBoost，Lasso，投票回归，随机森林，决策树，线性回归和支持向量回归** |  |
| **输入数据** |  |  |
| **数据集** | **62,723条带有19个变量** |  |
| **为了得到重要性序列（敏感性分析）** | **重复训练42次，每次减少一个特征，再比较准确度，减少比较小的影响较小，多的反之。** |  |
| **测试数据精度** | **与其他现代机器学习技术相比，模型的准确性提高了10％，**   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  |  | **Before Target Binning** | | | **After Target Binning** | | | | ***R*2** | ***MSE*** | ***MAE*** | ***R*2** | ***MSE*** | ***MAE*** | | 1 | VCPA Model | 0.74 | 9.36 | 4.51 | - | - | - | | 2 | Linear Regression | 0.8 | 9.98 | 4.33 | 0.97 | 9.87 | 4.53 | | 3 | SVR | 0.78 | 9.83 | 4.36 | 0.87 | 9.29 | 4.47 | | 4 | Decision Tree | 0.84 | 9.1 | 4.36 | 0.93 | 8.57 | 4.13 | | 5 | Random Forest | 0.91 | 8.79 | 4.2 | 0.97 | 8.2 | 4.07 | | 6 | XGBoost | 0.92 | 8.59 | 4.1 | 0.97 | 8.17 | 4.03 | | 7 | Lasso | 0.81 | 9.18 | 4.33 | 0.97 | 8.19 | 4.17 | | 8 | Voting Regressor | 0.9 | 8.87 | 4.36 | 0.97 | 8.37 | 4.04 | | 9 | CatBoost | 0.92 | 8.81 | 4.22 | 0.97 | 8.09 | 4.07 | |  |
| **提升猜想** | **可以通过增加数据集的大小在后续研究中扩大研究范围，从而可以解决数据集和本研究潜在的未发现细节和特征** |  |
| **房屋相关特征** | |  |  |  |  | | --- | --- | --- | --- | | **Variable Name** | **Description** | **Mean** | **Std. Deviation** | | ***Parid*** | Parcel ID (property ID) | 4386556 | 1780110 | | ***Aprland*** | Total Land Just Value | 31512.52238 | 22474.45819 | | ***Aprbldg*** | Total Building(s) Just Value | 153352.0601 | 65863.09319 | | ***Aprtot*** | Just Value at time of Sale or Total Just Value | 184864.5824 | 78255.48143 | | ***Nbhd*** | Neighborhood Code | 3553.249727 | 1617.738503 | | ***Rmbed*** | Number of bedrooms | 2.940746 | 0.747859 | | ***Sfla*** | Square Footage of Living Area | 1670.539912 | 590.574028 | | ***total\_area*** | Total Building Square Footage | 2409.546287 | 817.763307 | | ***yrblt*** | Year Built | 1988 | 21 | | ***misc\_area*** | Miscellaneous area that includes the gym, swimming pool, parking | 141.641915 | 236.598402 | | ***ZIP21*** | zip code of area | 32394.25951 | 289.84512 | | ***sale\_date*** | sale date of parcel | - | - | | ***sasd*** | School Assessed Value | 172068.0029 | 75437.46026 | | ***nsasd*** | Non-School Assessed Value | 171783.2722 | 75619.59938 | | ***stxbl*** | School Taxable | 154036.8721 | 78289.80939 | | ***nstxbl*** | Non-School Taxable Value | 139935.5139 | 79855.06776 | | ***cotxbl*** | County Taxable Value | 138761.7173 | 80899.04927 | | ***citxbl*** | Sale Price of House City Taxable Value | 109971.8981 | 88619.63572 | | ***Price*** | Sale Price of House | 197912.7255 | 94021.27873 | |  |

|  |  |  |
| --- | --- | --- |
| 描述  属性 | 详情 | 备注 |
| 文章信息 | **An Intelligent Automatic Valuation System for Real Estate Based on Machine Learning.AIIPCC '19, December 19–21, 2019, Sanya, China © 2019 Association for Computing Machinery. ACM ISBN 978-1-4503-7633-4/19/12...$15.00 https://doi.org/10.1145/3371425.3371454** | **AIIPCC** |
| **特点** | 1. **剔除重复数据（找到相似特征，基于神经网络的语义相似度算法）；** 2. **梯度提升决策树（GBDT）模型，随机森林（RF）模型和反向传播（BP）神经网络模型；** 3. **加权平均法结合了GBDT，RF和BP神经网络模型** |  |
| 输入 |  |  |
| 模型 | **梯度提升决策树（GBDT）模型，随机森林（RF）模型和反向传播（BP）神经网络模型** |  |
| **输入数据** |  |  |
| **数据集** |  |  |
| **为了得到重要性序列（敏感性分析）** |  |  |
| **测试数据精度** | **82%** |  |
| **提升猜想** | **提出用聚类的方法提升模型** |  |
| **房屋相关特征** | **社区级功能包括位置，容积率，绿化率，建筑年限，管理费用，停车位，交通状况，生活设施和学区住房。 建筑物级别的功能包括建造时间，建筑物类型，建筑，物业类型，层数和电梯状态。 房屋级别特征由房屋结构，面积，楼层，方向和装饰状态组成** |  |

|  |  |  |
| --- | --- | --- |
| 描述  属性 | 详情 | 备注 |
| 文章信息 | **Boosting House Price Predictions using Geo-Spatial Network Embedding.[arXiv](https://arxiv.org/abs/2009.00254).9.2019.Yuan-Fang Li · Yong-Bin Kang · Timos Sellis** | **[arXiv](https://arxiv.org/abs/2009.00254)** |
| 目的 | 虑到邻里便利设施的地理空间环境对房价的影响,例如房屋离火车站，高级学校或购物中心有多近。例如“这所房子离火车站有多近？”或“从这所房子步行可到达的学校是否很好” ？”。这些基于邻里设施的地理空间环境会极大地影响用户购买房屋的偏好，因此可能成为决定房屋价格的关键因素（POIs） |  |
| 模型1 | *图神经网络，地理空间网络嵌入（GSNE），GSNE采用基于高斯的嵌入方法* |  |
| **数据集** | **澳大利亚墨尔本的大型房地产和POI数据集，提取了2013年至2015年间总计52,851笔房屋交易记录，数据集包含13340个地区，709所学校和218个火车站的信息** |  |
| **房屋价格数据分布及数据分析** |  |  |
| **结论** |  |  |
| **亮点** | **–提出了一种新颖的地理空间网络嵌入（GSNE）框架，该框架可以根据不同类型的POI及其特征以及加权，归因的多部分图中这些POI之间的关系来准确捕获地理空间邻域上下文。**  **–我们采用并扩展了高斯嵌入方法，以实现我们的GSNE框架，该框架效率很高，并且可以处理异构类型的节点和特征。**  **–我们对大型房地产数据集的综合评估表明，对于房屋预测任务，无论下游回归模型如何，将GSNE获悉的地理空间嵌入矢量与房屋特征相结合，始终比仅原始特征具有更好的预测性能。** |  |
| **提升** |  |  |
| **提升猜想** |  |  |
| **房屋相关特征** |  |  |
|  |  |  |
| 描述  属性 | 详情 | 备注 |
| 文章信息 | **Piao, Y., Chen, A., Shang, Z.: Housing price prediction based on cnn. In: 2019 9th International Conference on Information Science and Technology (ICIST), pp. 491–495. IEEE** | **IEEE** |
| 目的 | 考虑到各种因素可能会对房价产生意想不到的相互影响，引入了基于CNN的模型来处理这种复杂性） |  |
| 模型 | CNN |  |
| **数据集** | **中国大连市国土资源和房屋信息中心的数据集。 该数据集包含2013年至2017年的171,155笔房地产交易记录**  每次迭代随机选择80％的数据作为训练集，并选择20％的数据作为测试集  2000次迭代，CNN的损失函数收敛，最终误差最终到达相对较低的点 |  |
| **房屋价格数据分布及数据分析** | **利用**XGBoos 得出特征重要性，利用准确性与MSE来评估 |  |
| **结论** | **CNN模型的误差收敛到0.01057，精度接近98.68％，表明CNN模型的预测精度和均方差均优于GM模型和XGBoost模型。 GM模型对数据的要求最低，但其中的预测精度也最低。 XGBoost模型需要进行更多的预处理以进行特征工程设计，但准确性往往比GM高。 CNN具有最高的预测精度，并且通过将数据转换为二维格式在房地产数据处理中具有巨大优势** |  |
| **亮点** | **GDP，房地产开发投资，人均收入等重要因素综合到预测模型中** |  |
| **提升** |  |  |
| **提升猜想** |  |  |
| **房屋相关特征** |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| 描述  属性 | 详情 | 备注 |
| 文章信息 | Learning Shared Vertex Representation in Heterogeneous Graphs with Convolutional Networks for Recommendation |  |
| 目的 | 旨在从交互数据中挖掘三种信息，即用户偏好，项目依赖性和用户行为的相似性 |  |
| 输入 |  |  |
| 模型 | **三个图，包括用户项目图，项目项目图和用户子图，以分别表示用户对项目的偏好，项目依赖性和用户相似性：PGCN的图卷积网络** |  |
| **输入数据** |  |  |
| **数据集** |  |  |
| **为了得到重要性序列（敏感性分析）** |  |  |
| **测试数据精度** |  |  |
| **提升猜想** |  |  |
| **房屋相关特征** |  |  |

搜索关键词：

House price prediction neural network

real estate price prediction

Housing Price Prediction Using Machine Learning Algorithms

real estate price neural network

real estate price graph

思考：

1、已经提出的房价预测模型：回归、基于决策树的相关模型、CNN、ANN，少有用图模型；

2、提出一种开发商+房屋+房屋特征的预测，每个开发商都拥有自己的定价策略，定价的过程类似于推荐系统中用户选择商品（用户选择商品也是根据商品的特征进行定价与系统展示的价格对比过后的选择）是否能结合推荐系统做一个预测模型；

3、第二点存在的问题，是否准确性比已经提出的模型更好，如果没已经提出的模型准确性高，有没有可行性；

4、想试试第2点的原因：

1）因为已经获得了1000项目的数据，每个项目都获得了其建设的开发商（包括联建开发商），每个项目包含多个楼盘，并其中每个楼盘超过100套房子，这些房子全部是商品房，可以借鉴模仿推荐系统中推荐用户选择商品模型。以下是房屋数据（约计：1000\*5\*100=10万条房屋数据）：

**2）根据自己读的论文，已经提出的房价预测模型中没有采用商品房的数据，都是采用二手房数据集；**

**3）根据自己拥有的经验，购房人购房过程中商品房的交易量远远大于二手房交易数据，用商品房成交叫个进行预测更有意义；**

**4）用图的模型没看到；**

1. **若采用第2点建造模型，会增加其他特征数据，比如：房屋周围信息，学校、地铁、公交等特征；**
2. **听听老师的建议。**

**期待读的论文：**

**He, Y., & Xia, F. (2020). Heterogeneous traders , house prices and healthy urban housing market : A DSGE model based on behavioral economics. Habitat International Journal, 96(March 2019). 研究投资对房价的影响**

**Bin J, Tang S, Liu Y, et al. (2017). Regression model for appraisal of real estate**

**using recurrent neural network and boosting tree. 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), IEEE, 209-213.LSTM的性能**

**Poursaeed O, Matera T and Belongie S (2018). Vision-based real estate price estimation. Machine Vision and Applications, 29(4), 667-676.CNN房屋视觉对房价的影响**

**Shashi Bhushan Jha, Radu F. Babiceanu, Vijay Pandey, Rajesh Kumar Jha.Housing Market Prediction Problem using Different Machine Learning Algorithms: A Case Study.2020**

**Piao, Y., Chen, A., Shang, Z.: Housing price prediction based on cnn. In: 2019 9th International Conference on Information Science and Technology (ICIST), pp. 491–495. IEEE** (2019) （已读）

**Zhao, Y., Chetty, G., Tran, D.: Deep learning with xgboost for real estate appraisal. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1396–1401. IEEE (2019) (房屋视觉)**

**Chen, X., Wei, L., Xu, J.: House price prediction using lstm. arXiv preprint arXiv:1709.08432 (2017)**

**Wang, X., Wen, J., Zhang, Y., Wang, Y.: Real estate price forecasting based on svm optimized by pso. Optik125(3), 1439–1443 (2014)**

**Jenkins, P., Farag, A., Wang, S., Li, Z.: Unsupervised representation learning of spatial data via multimodal embedding. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Manage-ment, pp. 1993–2002 (2019)**

Cai, H., Zheng, V.W., Chang, K.C.C.: A comprehensive survey of graph embedding: Problems, techniques, and

applications. IEEE Transactions on Knowledge and Data Engineering 30(9), 1616–1637 (2018)

Bojchevski, A., Gu ̈nnemann, S.: Deep gaussian embedding of attributed graphs: Unsupervised inductive learning

via ranking. arXiv preprint arXiv:1707.03815 (2017)高斯嵌入

Bae SW, “Predicting the real estate price index using deep learing,” Korea Real Estate Review, vol. 27, no. 3, pp. 71-86, 2017.

RN Shen, and CJ Fan, “Prediction of Shanghai housing price based on support vector machine model of principal component analysis,” Practice and Understanding of Mathematics, vol. 43, no. 23, pp. 11- 16, 2013.

GL Wei, XY Deng, and Q Zhang, “Forecast and analysis of Kunming house price based on Markov chain forecasting model,” Chinese Market, vol. 21, pp. 86-87, 2015.

YJ Liu, “Self-learning and application of neural networks”, University of science and technology of China, 2018.（侧重于处理文本数据以预测房价）

[Cheng et al., 2013] Chen Cheng, Haiqin Yang, Michael R Lyu, and Irwin King. Where you like to go next: Succes- sive point-of-interest recommendation. In IJCAI, pages 2605–2611, 2013.

[Cheng et al., 2017] Zhiyong Cheng, Jialie Shen, Lei Zhu, Mohan Kankanhalli, and Liqiang Nie. Exploiting music play sequence for music recommendation. In IJCAI, pages 3654–3660, 2017.（用户推荐）

[He et al., 2018] Xiangnan He, Zhankui He, Jingkuan Song, Zhenguang Liu, Yu-Gang Jiang, and Tat-Seng Chua. Nais: Neural attentive item similarity model for recommenda- tion. TKDE, 30(12):2354–2366, 2018.（用户推荐）

**Ma, J. et al. 2020, Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques, Land Use Policy, vol. 94, DOI= https://doi- org.ez.xjtlu.edu.cn/10.1016/j.landusepol.2020.104537.选择了六种ML方法，包括随机森林（RF），梯度提升决策树（GBDT），多线性回归（MLR），线性支持向量回归（SVR），多层感知器（MLP）回归和K最近邻（KNN）基于纽约市房地产市场的大型数据集进行回归以评估和比较其准确性，并得出具有最高R平方值的RM的主导性能。 Park and Bae（2015）的研究基于C4.5，RIPPER，NaïveBayesian和AdaBoost等算法开发了房价预测模型，并比较了其分类性能。结果表明，在预测房价方面，RIPPER始终优于其他模型。此外，Allan，Martin和Andre（2001）证明，基于序数的传统享乐模型在解释环境变量对房地产估值的影响方面产生的效果与非线性人工神经网络（ANN）模型相似。在对中国的房屋价格进行建模时，龚（2018）运用XGBoost和LASSO来预测武汉的房屋价格，并发现XGBoost具有显着的优势。**