

1 内部因素（标准化）

- 面积修正
- 层次修正
- 空间布局修正
 - 是否合法分层
- 平面形状修正
 - 正方形/短长方形/长长方形/倒梯形/梯形/扇形/倒三角形/其他
- 临街状况修正
 - 单面临街/双面临街/三面临街等
- 门面特殊因素修正
 - 有无绿化带或人行道/有其他遮挡物/是主要人流入口
- 开间、进深修正
 - 土地修正
 - 按照土地占总租金的65%计
 - 按照4-3-2-1以及9-8-7-6修正法则来修正
 - 选取标准开间进深为9m * 9m

2 外部因素（模型化）

2.1 人流量

- 人流量时间限制
 - 工作日: 06点~20点
 - 双休日: 10点~22点
- 人口密度
 - 内涵: 待估点所在栅格的人流量密度
 - 方法: 按照 $0.4 * q(\text{工作日}) + 0.6 * q(\text{双休日})$ 的比例
- 购买力
 - 内涵: 通过人的收入/职业/阶层反映
 - 方法: 通过绑定周边物业, 间接反映人的购买力, 包括:
 - 绑定周边住宅的价格或档次分布(权重: 0.5)
 - 方法: $0.5 * \text{内环内监测价格标准化(由低到高)} + 0.5 * \text{小区档次量化分值(低\中\高\超高档: 0.4, 0.7, 0.9, 1)}$
 - 绑定周边不同商业业态(权重: 0.5)
 - 方法: $0.4 * \text{内环内综合商场租金标准化(由低到高)} + 0.4 * \text{综合商场档次量化分值(普通\中\高: 0.3, 0.6, 0.9)} + 0.2 * \text{空置率标准化(由高到低)}$
 - 备注: 由低到高 指分值越低标准化后分值越低, 由高到低 指分值越低标准化后分值越高
- 消费方向
 - 内涵: 从房地产功能多样性以及消费频次(低频、高频、单次偶然、常规多次)两个维度分析
 - 方法: 模型计算暂不考虑
- 人流量的变化
 - 内涵: 偏度统计量的17个样本估计值(06点~21点)
 - 偏度是相对于平均值不对称程度的度量;一般来说,右偏(右侧长尾)时算术平均数>中位数>众数, 左偏时相反, 即众数>中位数>平均数



- 对于分小时的人流量序列,右偏说明有偶发的人流量增大的时间段,左偏说明人流量总体偏向均值,但有偶发的人流量减少的时间段.
 - 因此,左偏的序列标准化后分值偏高、右偏序列标准化后分值偏低;而无偏序列,没有体现明显的人流量变化趋势,较中庸,标准化后分值也位于中间。
- 方法:
 - 计算每个栅格的偏度统计量,按照 $1 - (Skewness - SkMn) * 1.0 / (SkMx - SkMn)$ 的方法标准化
- 消费时间
 - 内涵:
 - 全天/偶然/定时(日/夜)
 - 同上,通过人流量变化判断人的消费时间
 - 方法:根据时间段分早中晚('1:Time_M'/2:Time_N'/3:Time_E')三类分别统计人流量,根据规则判断属于三类中的某一类或某两类。
- 到达性
 - 人流是通过周边临近区域漫延而来,还是通过其他区域跳跃过来
 - 通过绑定周边公交、轨交及距离,一定程度上判断人流聚集是属于漫延还是跳跃类型。
 - 方法:
 - 先绑定周边8个栅格,以及每个栅格周末相对于工作日的人流密度涨跌幅
 - 取8个栅格的涨跌幅众数方向,计算众数方向的平均值,记为q;目标栅格涨跌幅记为q0
 - 如果 $q > 0$, $q0 > 0$ 且 $q - q0 > 0$,则为跳跃,打分1
 - 如果 $q < 0$, $q0 > 0$ 且 $abs(q) - q0 > 0$,则为漫延,打分0.5
 - 其它情况视为既有跳跃又有漫延,打分0.7
 - 方法2:
 - 经过测算,栅格人流量均为周末小于工作日,那么认为周末减少的幅度越大,越倾向于人流跳跃而来,其分值就越高

2.2 POI因素

- 数量的分值-相对沿街店铺有效
 - 包括:原始POI和住宅、办公、商场、沿街店铺等广义的POI
 - 原始POI:按照之前德尔菲打分法计算的分值
 - 住宅、办公、商场、沿街店铺
 - 取周边500m内的POI,根据距离进行衰减 $\sum exp(-d^2/2)$
 - 栅格汇总
 - 对所有样本点的分值直接求均值(只考虑栅格内已有的POI类型,没有的POI类型不参与均值计算)
- 数量的分值-相对沿街店铺无效
 - 对栅格里面每个样本点,及其对应的每一类POI,分别计算其有效半径之外的POI点的个数,记为与该样本点对应的无效POI个数
 - 对栅格内的所有样本点,按照其有效缓冲区面积作为权重,计算栅格的加权无效POI个数;无样本点的栅格,其无效POI个数等于其内的POI总数
 - 栅格对第j类POI的无效数量为: $N(栅格, j) = AVG_i(N(样本点_i, POI_j))$
 - 对所有的栅格的 $N(栅格, j)$,按照数量少分值大的方向进行标准化,记为 $NStd(栅格, j)$.
 - 栅格的无效POI数量 $N(栅格)$ 的计算: $N(栅格) = \sum_j N(栅格, j)$,然后标准化,记为 $NStd(栅格)$.
- 多样性
 - 分别统计每个栅格内的POI的种类的数量,此处数据应包括参与和没有参与距离打分的所有POI
 - 按照数量多分值大进行标准化

FID	VfStd	ProjectLevelScore_C	RentPriceStd	VacancyRateStd	ProjectLevelScore_R	BasePriceStd	BindResidentScore	BindCommercialScore	PurchPower	Skewness	ArrivalProperty	IsTime_M	IsTime_N	HospitalNStd	SportsVenuesNStd	BusStopNStd	HighSchoolNStd
0	0.213548	0.600000	0.688259	0.133333	0.582477	0.669913	0.626195	0.541970	0.584083	0.677032	0.381317	0	1	0.262846	0.000000	0.580296	0.887787
1	0.299715	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.657010	0.506471	1	0	0.000000	0.000000	0.000000	0.000000
2	0.257452	0.600000	0.890688	0.026667	0.434897	0.708603	0.571750	0.601609	0.586679	0.628735	0.445542	1	0	0.291008	0.815192	0.525121	0.160031
3	0.014568	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.642642	0.460567	0	1	0.000000	0.000000	0.000000	0.000000
4	0.045232	0.000000	0.000000	0.000000	0.535242	0.741525	0.638383	0.000000	0.319192	0.506444	0.095397	1	0	0.394269	0.000000	0.133183	0.000000
5	0.479868	0.719244	0.667452	0.164679	0.464863	0.669421	0.567142	0.587614	0.577378	0.564993	0.322649	1	0	0.378623	0.000000	0.448699	0.118892
6	0.794802	0.761895	0.722494	0.159348	0.506042	0.633729	0.569886	0.625625	0.597755	0.719666	0.498046	0	1	0.356719	0.364624	0.770816	0.756408
7	0.470187	0.502021	0.890669	0.084906	0.526379	0.666063	0.596221	0.574057	0.585139	0.488621	0.454933	1	1	0.338883	0.545050	0.553479	0.330966

FID	MarketWholesaleNStd	CommunityServiceNStd	DepartStoreNStd	ShoppingMallNStd	RailStationNStd	ResidentNStd	OfficeNStd	BusinessNStd	POINumStd	POITypeNumStd	POINumOStd	LngStd	LatStd	PriceRentDSTD
0	0.893778	0.249274	0.601680	0.565205	0.565205	0.380245	0.730959	0.086014	0.649975	0.588235	0.097561	0.07617034753326999	0.01565635512297267	25.880400
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.588235	1.000000	0.13254351437121491	0.000000000000000000	{null}
2	0.601810	0.417002	0.535830	0.543857	0.543857	0.321482	0.354886	0.283285	0.595924	0.588235	0.018699	0.18895546119093324	0.00643704012486306	12.570028
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.058824	0.012195	0.24281439254390217	0.03083768770661056	{null}
4	0.000000	0.268735	0.626374	0.565205	0.565205	0.505791	0.224195	0.137399	0.335774	0.352941	0.000000	0.01979071736502950	0.10859596909445193	20.858800
5	0.311432	0.132565	0.590774	0.556652	0.556652	0.512098	0.281403	0.196094	0.423264	0.529412	0.074526	0.06425196646824242	0.08423409886285420	27.592075
6	0.435142	0.432755	0.876177	0.873098	0.873098	0.476930	0.557019	0.144795	0.796900	0.823529	0.165505	0.13078548853082039	0.08168448809050633	24.214333
7	0.519583	0.699627	0.734932	0.685735	0.685735	0.597128	0.486138	0.426360	0.755502	0.705882	0.053659	0.19857289667073856	0.08054055626108789	26.895700

3 模型构建过程

3.1 栅格的平均租金模型

3.1.1 训练集、测试集、泛化集的拆分

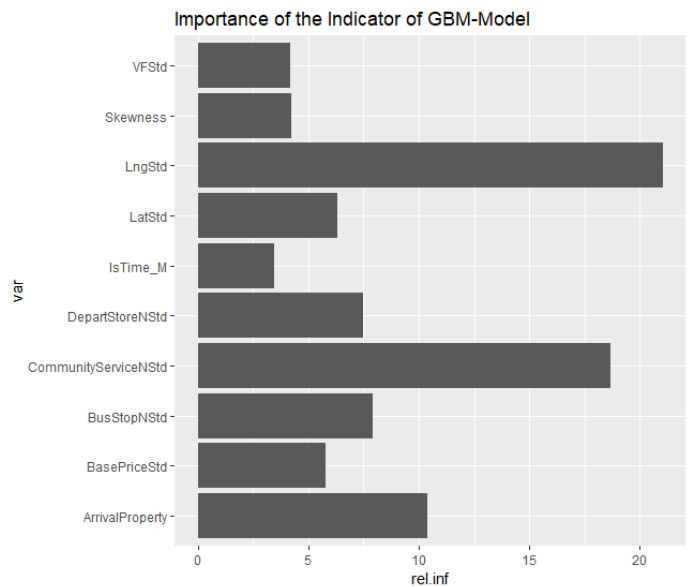
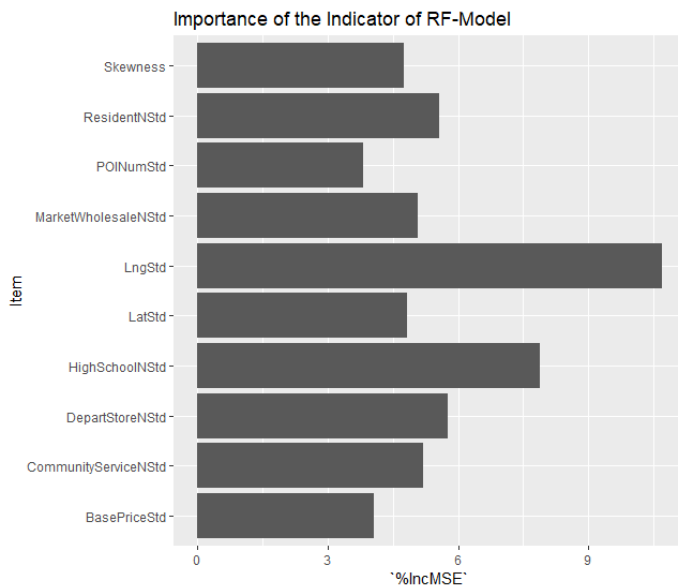
- 泛化集（无样本点栅格）占比： 20%
- 训练集占比： 68%
- 测试集占比： 12%

3.1.2 模型训练及模型评价

使用随机森林、SVR、GBDT三类模型分别建模，再用Stacking方式进行模型融合，得到最终的融合模型。

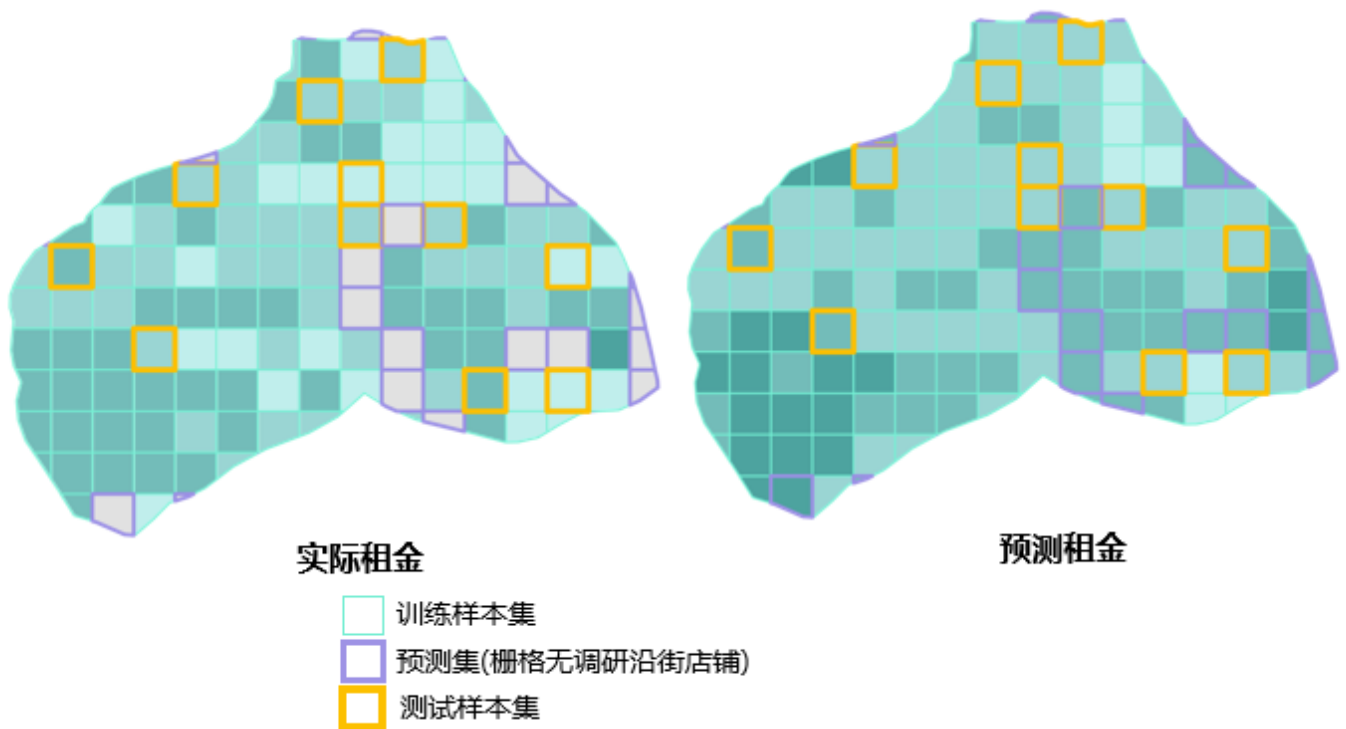
- 评价指标如下：
 - R2(R方)—0.62,还有需要提升的空间
 - RMSE(均方误差)- 2.2890, MAE(平均绝对误差)- 1.98890；说明模型表现能力使得栅格平均租金的模型拟合值与实际值相差2元左右
 - MAPE(平均绝对百分比误差)-10.704%；说明栅格的平均租金的模型拟合值与实际值平均偏离10.704%
- 模型偏离：
 - 误差10%及以内的样本栅格占比为62%
 - 误差15%及以内的样本栅格占比为79%

指标的重要性度量



◦ 同档次租金水平划分后的模型性能评价：

- 把栅格的实际租金和模型预测租金按照相同规则划分4个档次水平
- 然后根据预测租金的水平与实际租金水平是否相等，作为预测结果是否准确的度量
- 模型在测试集上泛化准确率为8/11 (73%)



◦ 模型不足

- 影响因子考虑过少
- 样本栅格量过少
- 模型可能出现一定程度的过拟合

3.2 栅格内沿街店铺的空间插值模型

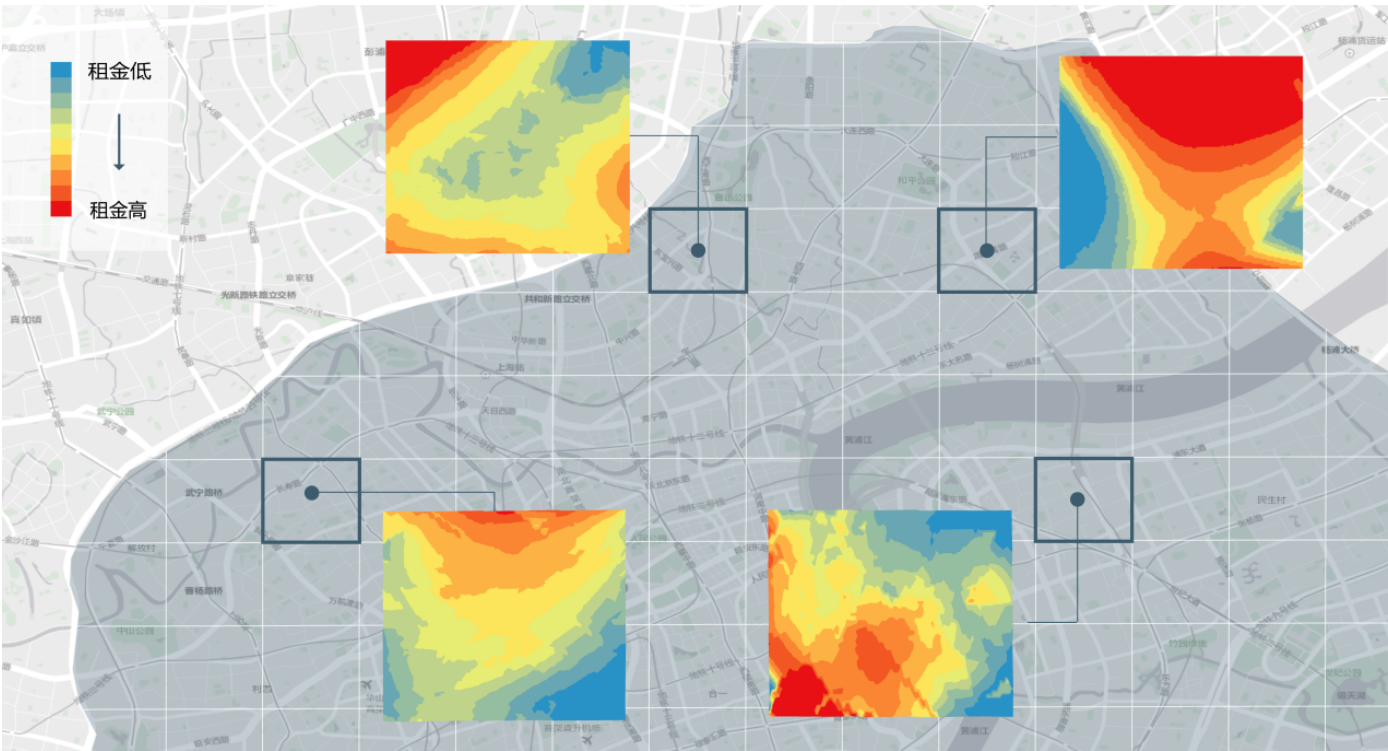
3.2.1 适用条件

- 基于地统计的插值算法，是一种不确定插值方法，依赖于数学模型和统计模型
- 样本点数量和分布有要求。样本点随机分布且样本量较大-平均1公里网格至少具有6个样本点
- 有一定的异常值识别功能。能反应价格数据随空间位置变异的趋势；对于样本量较大的情况，一方面识别并修正异常值，另一方面给出非样本点符合分布趋势的拟合值

3.2.2 模型测算

- 从所有栅格中筛选出满足Kriging插值条件的所有栅格（约50个）
- 再随机选取了其中的四个
- 以其中的沿街店铺的标准租金作为样本点，分别进行Kriging空间插值测算

示例栅格内的沿街店铺标准租金Kriging空间插值效果展示



3.2.3 模型评价

对上面筛选的4个栅格的Kriging空间插值拟合值与沿街店铺的标准租金进行对比，平均效果如下：

- 误差10%以内的占比为40.5%
- 误差15%以内占比为61.5%

4 模型优化方向

- 扩大样本点范围。由内环内扩展到外环内

- 人流量数据的完善。更详实的人流量数据
- POI的影响量化规则的优化。