



XXXX

# 融合多源异构数据的城市复杂环境下5G网络覆盖价值评估与优化方法

刘少蔚<sup>1</sup>, 王迎春<sup>2</sup>, 齐勇<sup>1</sup>, 程立杰<sup>1</sup>, 李寅<sup>1</sup>, 黄龙<sup>1</sup>, 李武<sup>1</sup>

(1. 重庆市信息通信咨询设计院有限公司, 重庆市, 400039;

2. 中国电信集团有限公司重庆分公司, 重庆市, 401120)

**摘要:** 5G网络建设进入精细化运营阶段后, 城区居民区的5G覆盖质量成为影响用户体验的关键指标。针对传统评估方法在多维度关联性分析、问题定位精度上的不足, 以及难以适配精细化规划需求等短板, 提出一种基于地理空间大数据融合的城区居民区5G覆盖评估模型。该模型整合楼宇POI、小区级网管数据、栅格级用户感知数据等多源异构信息, 构建四大协同子模型, 实现从场景细分到投资优先级量化的全流程评估; 其中, 引入LambdaMART算法作为核心排序机制, 有效支撑多维指标融合下的精准价值排序。实验结果显示, 模型的弱覆盖识别准确率达92.5%, 决策质量与专家判断高度吻合(NDCG>0.9)。该模型可有效提升覆盖评估的效率与精度, 为5G网络的精细化规划优化提供科学的决策支持。

**关键词:** 5G; 城区居民区; 覆盖评估; 多源异构数据融合; 地理空间大数据; 价值排序; 机器学习

**中图分类号:** TN929.5

**文献标志码:** B

**doi:** 10.11959/j.issn.1000-0801.

## Integration of Multi-Source Heterogeneous Data for Value Assessment and Optimization Methods of 5G Network Coverage in Complex Urban Environments

LIU Shaowei<sup>1</sup>, WANG Yingchun<sup>2</sup>, QI Yong<sup>1</sup>, CHENG Lijie<sup>1</sup>, LI Yin<sup>1</sup>, HUANG Long<sup>1</sup>, LI Wu<sup>1</sup>

1. Chongqing Communication Design Institute Co.,Ltd, Chongqing, 400039

2. China Telecom Corporation Limited Chongqing Branch, Chongqing, 401120

**Abstract:** As 5G network construction progresses into the stage of refined operation, the quality of 5G coverage in urban residential areas has become a critical indicator influencing user experience. To overcome the limitations of traditional evaluation methods—such as insufficient multi-dimensional correlation analysis, low problem localization accuracy, and poor adaptability to refined planning requirements—this paper proposes a 5G coverage evaluation model for urban residential areas based on geospatial big data fusion. The model integrates multi-source heterogeneous information, including building points of interest (POI), cell-level network management data, and grid-level user perception data, and constructs four collaborative sub-models to enable full-process evaluation ranging from scenario segmenta-



tion to investment priority quantification. The LambdaMART algorithm is introduced as the core ranking mechanism, effectively supporting precise value ranking under the fusion of multi-dimensional indicators. Experimental results demonstrate that the model achieves a weak coverage identification accuracy of 92.5%, and its decision quality is highly consistent with expert judgment (NDCG>0.9). The proposed model significantly improves the efficiency and accuracy of coverage evaluation, providing scientific decision support for the refined planning and optimization of 5G networks.

**Key words:** 5G, Urban Residential Areas, Coverage Assessment, Multi-Source Heterogeneous Data Integration, Geospatial Big Data, Value Prioritization, Machine Learning

## 1 引言

随着5G迈入成熟精细化运营阶段，通信运营商战略重心已从网络覆盖广度拓展转向网络投资价值深耕<sup>[1][2]</sup>。但单一数据源存在数据孤岛、评估粒度粗糙、缺乏综合价值判断等问题，难以适配复杂城市环境下精准定位覆盖问题、最大化投资回报的需求。

已有研究证实，融合运营商测量、第三方众包<sup>[3][4]</sup>、高分辨率GIS空间<sup>[5][6]</sup>、市场竞争<sup>[7]</sup>及社会经济<sup>[8]</sup>等多源异构数据，可显著提升评估的空间精度与业务感知度<sup>[9]</sup>。为此，本研究以多源异构数据融合理论为核心方法论，构建了5G网络覆盖评估从“信号度量导向”到“价值运营导向”的完整理论框架与可落地实践路径。

## 2 模型设计

### 2.1 整体架构

模型架构以系统工程理论与JDL（Joint Directors of Laboratories）数据融合模型第四级“过程优化”思想为理论基石<sup>[10]</sup>，遵循“感知-分析-决策”闭环逻辑<sup>[11][12]</sup>，通过地理空间栅格完成多源数据对齐与子模型动态耦合<sup>[13]</sup>。四大子模型构成有机协同体，场景识别、区域问题分析、市场预警子模型分别从场景价值、网络质量、市场竞争维度产出关键特征，共同驱动核心的价值排序子模型完成综合评判。

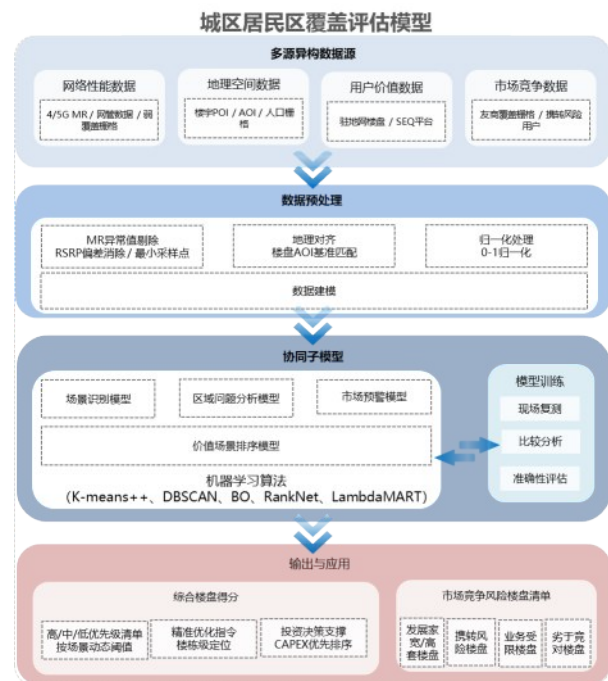


图1 模型整体架构

### 2.2 数据来源与预处理

模型输入数据涵盖四大类：地理空间数据（楼宇兴趣点(point of interest, POI)、楼盘兴趣区域 (Area of Interest, AOI)、第七次人口普查栅格)、网络性能数据：（4/5G测量报告 (Measurement Report, MR)、小区网管数据、弱覆盖栅格)、用户价值数据（驻地网楼盘、客户体验管理使能(Service & Experience Quality, SEQ)平台数据)、市场竞争数据：竞争网络覆盖栅格、携转风险用户分布）。

为实现多源异构数据的有效融合，本研究设计了一套系统化的数据预处理流程。首先，通过

地理对齐操作，以楼盘 AOI 轮廓为基准，将不同来源的栅格、基站小区等数据统一至同一空间基准。

在此基础上，针对无线传播环境中 RSRP 测量值受多径效应、阴影衰落及用户设备 (UE) 差异等因素影响的问题，制定 MR 异常值剔除与质量控制策略。该策略基于“当单位区域采样点充足时，其 RSRP 均值分布近似正态分布<sup>[14][15]</sup>”为核心思想。因此，第一步，通过用户设备 RSRP 差值消除算法，去除因各类移动设备带来的 UE RSRP 偏差<sup>[16]</sup>；第二步，再采用统计学样本量公式进行优化<sup>[17]</sup>，以此确定每个栅格的最小采样点数。公式为  $n = (Z \times \sigma / E)^2$ ，其中  $Z$  为置信水平(取 1.96)， $E$  为可接受的误差范围(取 2dB)， $\sigma$  为该频段 RSRP 的标准差<sup>[18][19]</sup>。为获取稳健的  $\sigma$  值，随机抽取一批采样点丰富的“种子栅格”，分别统计其在不同频段 RSRP 标准差，并采用所有种子栅格标准差的中位数作为该频段  $\sigma$  值。

随后，为消除不同量纲指标对后续模型的影响，所有特征均经过 0-1 归一化处理。最终，基于清洗和对齐后的网管数据，我们构建了结构化的数据模型(如图 2 所示)，为四大子模型的协同工作奠定了坚实的数据基础

表 1 最终阈值(20 m × 20 m 栅格)

频段	经验标准差 $\sigma$ (dB)	样本量公式	最小采样点数
800 MHz	$\sigma \approx 4$	$n = (1.96 \cdot 4/2)^2 \approx 15$	$\geq 15$
1.8 GHz/2.1 GHz	$\sigma \approx 5$	$n = (1.96 \cdot 5/2)^2 \approx 24$	$\geq 24$
3.5 GHz	$\sigma \approx 8$	$n = (1.96 \cdot 8/2)^2 \approx 62$	$\geq 62$

响，所有特征均经过 0-1 归一化处理。最终，基于清洗和对齐后的网管数据，我们构建了结构化的数据模型(如图 2 所示)，为四大子模型的协同工作奠定了坚实的数据基础

### 2.3 子模型设计

#### 2.3.1 场景识别模型

选取高套餐用户占比、宽带用户占比、楼盘户数为核心指标，采用 K-means++ 无监督算法实现场景精准分类<sup>[20]</sup>。采用 Z-score 标准化消除量纲影响，将指标转化为均值 0、标准差 1 的标准分布<sup>[21]</sup>，平衡楼盘户数与占比指标的权重；基于肘部法则确定最优聚类数  $K=3$ ，轮廓系数为 0.78。最终形成高密度高价值、中密度中价值、低密度低价值三类场景标准，为精准资源分配提供数据

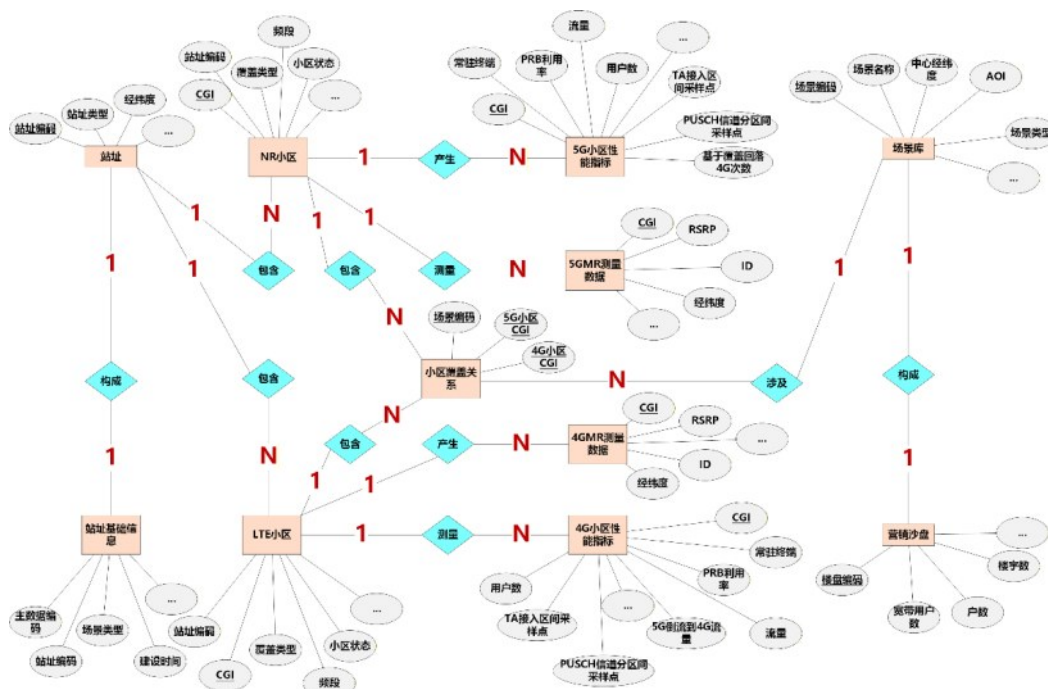


图 2 数据建模 ER 图



支撑。

### 2.3.2 区域问题分析模型

区域问题分析模型核心目标是从海量数据中精准识别并定位网络覆盖与容量两大类核心问

题。在评估指标层面，模型不仅定义了传统的“有效弱覆盖栅格”，还创新性地引入了基于密度的噪声数据空间聚类算法（DBSCAN）的“弱覆盖聚集区”概念，以捕捉局部信号衰减的连片效

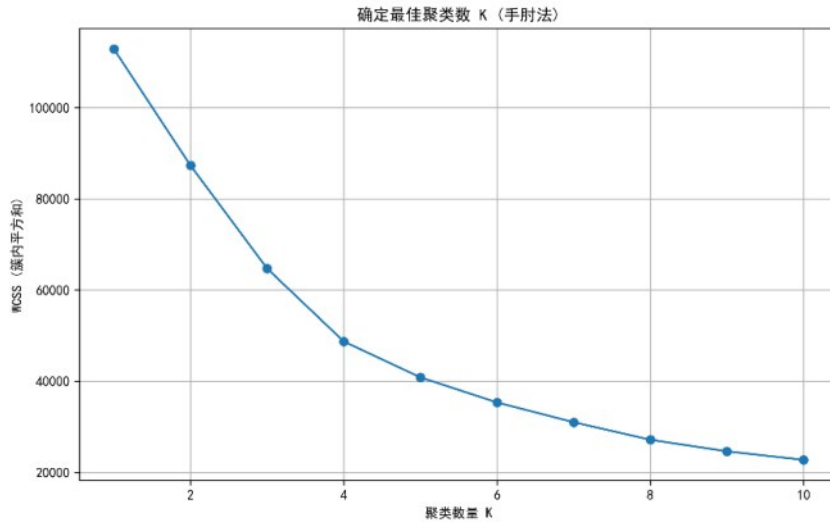


图3 确定K=3的肘部曲线

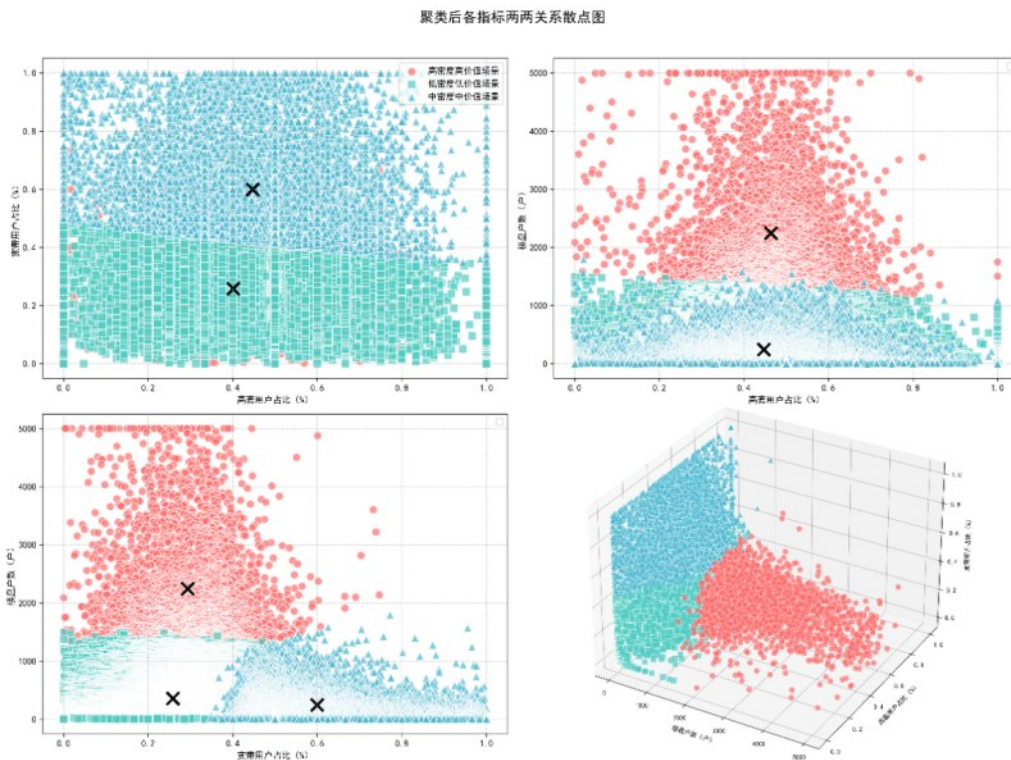


图4 聚类分布,黑色叉号为聚类中心

应。同时，模型也集成了行业通用的基站小区维度覆盖问题扇区定义，包括下行MR弱覆盖<sup>[24]</sup>、TA接入距离过远、高回落及上行功率受限等四类典型场景，从而实现了从宏观到微观的多层次问题刻画。

在问题区域定义层面，模型通过将上述指标与楼盘实体进行关联，实现了问题的精准定位。具体而言，当一个楼盘内5GMR弱覆盖栅格占比超过30%时，即被判定为弱覆盖楼盘；而当其4/5G小区PRB利用率高但5G驻留比低时，则被识别为容量受限楼盘。为进一步提升规划效率，模型还对识别出的问题楼盘，依据地理邻近性进行聚类，从而将孤立的问题点整合为可整体优化的“问题连片区域”。

### 2.3.2.1 弱覆盖聚集区识别：基于自适应DBSCAN的聚类框架

为精准刻画城市复杂环境中弱覆盖信号的连片效应，本模型引入了基于密度的噪声数据空间聚类算法（DBSCAN）。相较于K-means等需预设簇数量的算法，DBSCAN无需先验知识、能识

别任意形状的簇且对噪声鲁棒，天然适用于发现不规则分布的弱覆盖区域。然而，其性能高度依赖于邻域半径（ $\epsilon$ ，以下简称eps）与核心点最小邻居数（MinPts）参数对选择<sup>[22]</sup>。

为此，我们构建了一个“初选-优化-自适应”的三层参数优化框架（如图5所示）：

首先，利用k-距离图进行参数初选，为后续优化提供合理的搜索空间；

其次，将参数选择问题建模为一个黑箱优化问题，并采用（Bayesian Optimization,BO）全局优化技术<sup>[23]</sup>，在极少的评估次数下高效地找到全局最优或近优参数组合；

最后，考虑到城市环境中弱覆盖栅格的空间密度差异巨大，单一全局参数难以普适。因此，在获得最优全局参数并完成初步聚类后，模型会对各密度均匀的子区块进行二次精细聚类，从而实现全域的自适应精准识别。

这一框架确保了弱覆盖聚集区的识别过程既客观又稳定，完全摆脱了主观经验判断。

### 2.3.2.2 贝叶斯优化驱动的DBSCAN参数寻优

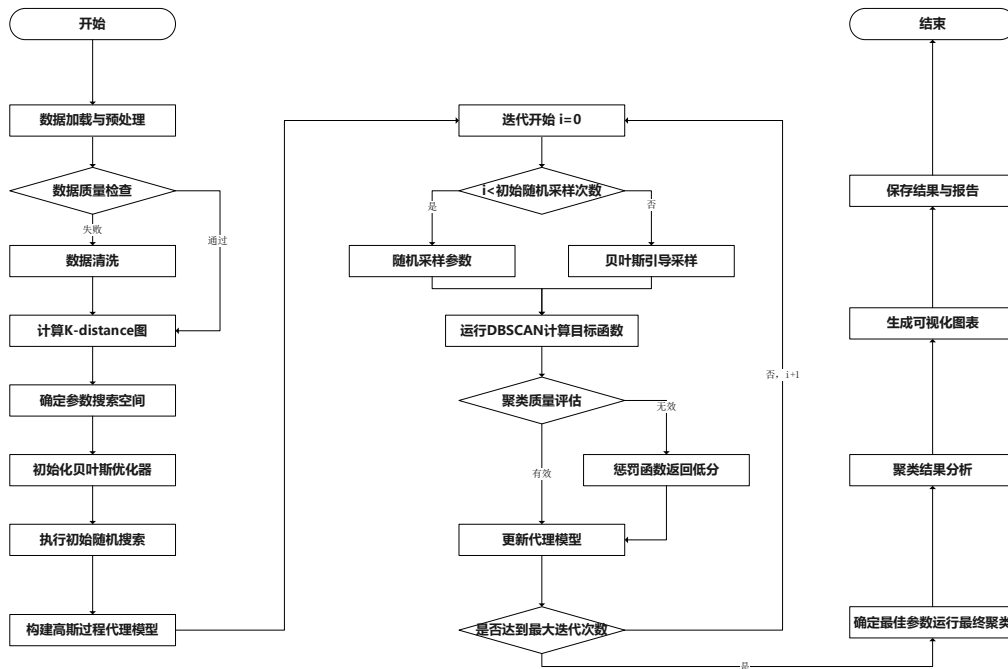


图5 地理空间数据聚类分析流程图



机制

在上述框架中，贝叶斯优化是实现参数智能寻优的核心引擎。其关键在于定义一个能够量化聚类质量的黑箱目标函数。

## 1 目标函数设计原则

由于地理空间聚类属于无监督学习，缺乏真实标签，因此模型融合了多个内部评估<sup>[26]</sup>指标来综合评判聚类效果：

(1)轮廓系数(Silhouette Coefficient):衡量数据点与所属簇的紧密性、与其他簇的分离度，取值[-1,1]，越接近1聚类效果越好<sup>[27]</sup>。

(2)戴维森堡丁指数 (Davies-Bouldin Index, DBI) :计算任意两个簇的簇内平均距离与簇间中心距离的比值，并取所有簇对该比值的最大值的平均。DBI 越小，表示簇内越紧密，簇间越分离，聚类效果越好<sup>[28]</sup>。

(3)Calinski-Harabasz 指数 (Calinski-Harabasz Index, CH) : CH 指数通过类间离散度与类内离散度的比率来评估聚类质量，值越大表示聚类效果越好<sup>[29][30]</sup>。

(4)噪声点比例(Noise Ratio, NR) : 噪声比衡量了DBSCAN无法归入任何聚类的点的比例<sup>[20]</sup>。

## 2 鲁棒性增强策略

为确保优化过程的稳健性，目标函数采用了两项关键策略：

(1)多指标归一化融合：通过表 2 所示的归一化函数，将不同量纲和取值范围的指标统一映射至[0,1]区间，再进行加权融合，形成综合评分。多指标加权综合的数学公式为：

$$S_{(Eps, MinPts)} = (SC)^{\omega_1} \cdot (CH)^{\omega_2} \cdot (1-DB)^{\omega_3} \cdot (NR)^{\omega_4}$$

(2)噪声惩罚与边界处理：若噪声点比例超过 30%<sup>[20]</sup>，目标函数将返回一个显著的低分 (-0.5)，以规避“伪优解”。同时，对于聚类失效

表 2 各指标归一化函数

指标	原始范围	归一化函数	备注	归一化后范围	权重分配
轮廓系数	[-1,1]	$SC = \frac{SC+1}{2}$	/	[0,1]	$\omega_1=0.3$
CH 指数	[0,∞)	$CH = 1 - e^{-\frac{CH}{\tau}}$	$\tau=1000$	[0,1]	$\omega_2=0.3$
DB 指数	[0,∞)	$DBI = e^{-\frac{DB}{\lambda}}$	$\lambda=2$	(0,1]	$\omega_3=0.3$
噪声比	[0,1]	$NR = 1 - NR$	/	[0,1]	$\omega_4=0.1$

(如无有效簇、全样本归为单簇、聚类数 $\geq 1/2$  采样点数)的情况，直接返回惩罚值 (-1)，引导优化器远离无效参数区域。综上，黑箱目标函数为：

$$f_{(Eps, MinPts)} = \begin{cases} -1 & \text{if } k \leq 1 \text{ or } k \geq \frac{n}{2} \\ -0.5 & \text{if } NR > 0.3 \\ S_{(Eps, MinPts)} & \text{otherwise} \end{cases}$$

其中  $k$  为聚类簇数， $n$  为样本总数。

## 3 优化流程

首先，基于  $k$ -距离图进行参数初选：对含  $n$  个弱覆盖栅格的样本集，计算各数据点至其第  $k$  个最近邻的距离并升序排序，绘制  $k$ -距离图。图中曲线拐点对应的距离值，通常指示了数据集中密度变化的边界，可作为 Eps 的候选值<sup>[25]</sup>。通过 BallTree 计算各点第  $k$  个近邻的球面距离。Y 轴是距离值，X 轴是数据点（按距离排序）。X 轴最大值对应的 Y 值即为程序建议的 eps 初始值。贝叶斯优化器会以此为参考，如图 6 所示在其下方进行重点探索。此外 MinPts 的下界设为 3（小于 3 无实际意义），上界取较小整数（如 50）或数据集总点数的小百分比（如 1%），以避免忽略小规模簇。

在此基础上，优化器结合前述多指标融合的目标函数，通过构建高斯过程代理模型和采集函数，迭代地选择最有希望的参数组合进行评估。

最终，当算法收敛至最优解时，得到的 Eps 和 MinPts 参数对将被用于执行 DBSCAN 聚类，

生成精准的弱覆盖聚集区识别结果，生成弱覆盖聚集区结果。

### 2.3 3 市场预警模型

从竞争劣势与用户流失风险两方面构建预警机制（该部分为行业定义或专家经验值，在此不作展开）：

（1）竞争劣势预警：楼盘内 5G 覆盖栅格占比（RSRP $\geq$ -105dBm） $\leq$ 40%，且友商占比  $>$  60%，定义为劣于竞争网络楼盘；

（2）流失风险预警：楼盘内手机用户数 $\geq$ 1 倍家宽开卡用户数或携转风险用户（近 3 个月有离网倾向）占比 $\geq$ 5%，且弱覆盖栅格占比 $\geq$ 20%，定义为携转风险楼盘。

### 2.3 4 价值场景排序模型

本子模型以 LambdaMART 算法为核心，融合网络运营、市场业务与投资效益的 12 个核心指标，其核心创新在于，针对三类建设场景（5G 已覆盖存量区域、5G 未覆盖存量区域、新建场景），定义差异化指标权重策略与优先级判定规则，输出动态匹配场景需求的排序清单，保障有限投资资源优先聚焦于网络问题突出、业务价值

显著或投资回报明确的高价值区域。

#### 2.3.4.1 模型架构

模型采用“数据层->特征层->模型层->应用层”的四层架构，如图 7 所示

（1）数据层：整合多网管 O 域、经分 B 域、资源管理 M 域数据及楼盘 GIS 图层、重点场景标注等空间数据。

（2）特征层：依据既定指标体系，量化计算各楼盘二级指标并归一化，生成多维特征向量。

（3）模型层：以 LambdaMART 算法为核心，输入楼盘特征向量与场景标签，学习不同场景下特征与“投资价值优先级”的复杂非线性关系，输出反映同场景相对价值排序的楼盘综合得分。

（4）应用层：基于量化阈值规则，将楼盘划分为高、中、低三级优先级，生成结构化排序清单，为网络规划与投资决策提供直接支撑。

#### 2.3.4.2 指标体系构建与量化方法

数据预处理是数据挖掘的关键步骤，包括清洗、集成、变换和归约<sup>[31]</sup>。归一化处理对于消除量纲、使不同指标可比至关重要<sup>[32]</sup>。

（1）特征向量构建与预处理：对楼盘 i，经

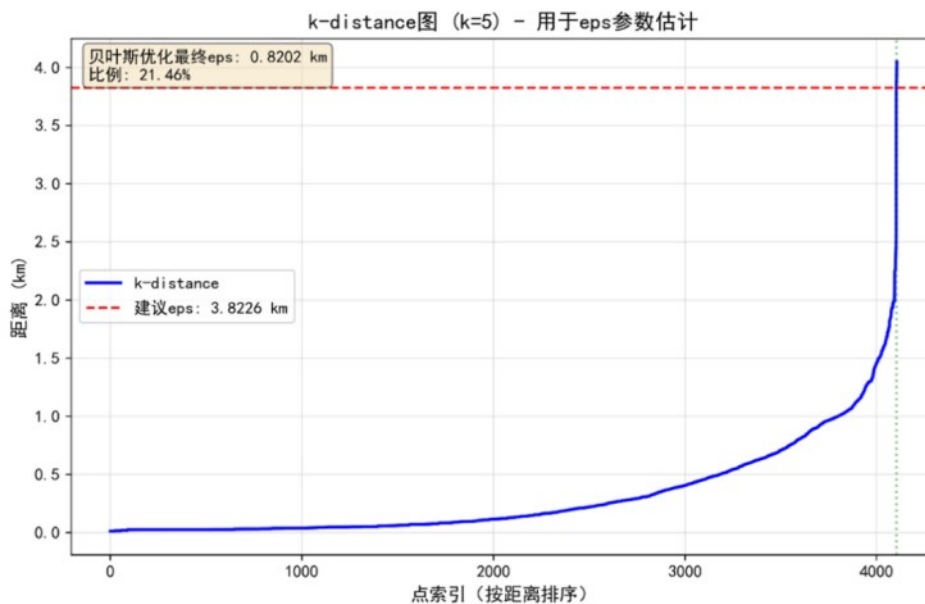


图 6 K-distance 图(K=5)

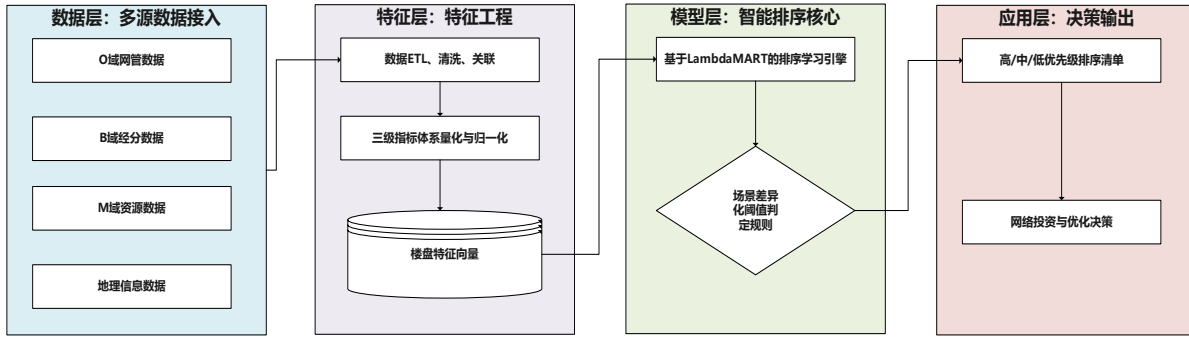


图7 楼盘价值优先级排序模型总体架构图

表3 价值排序核心指标量化方法

序号	一级指标	二级指标	核心定义	量化方法核心逻辑	归一化方式	
1	网络质量 (O级)	覆盖类 TOP 劣化扇区	基站小区维度覆盖类问题高发扇区	综合得分=多维度覆盖问题（弱覆盖/TA过远/高回落/功率受限）分项评分之和，取楼盘所有服务小区的算术平均值	Min-Max 归一化[0,1]	
2		弱覆盖栅格聚集区	楼盘内弱覆盖栅格高聚集区域	综合得分=楼盘内弱覆盖栅格聚集区面积与楼盘总面积的占比	Min-Max 归一化[0,1]	
3	价值区域 (B域)	家宽高价值楼盘	家宽业务渗透潜力高、ARPU 值高的楼盘	综合得分=潜力得分（基于楼盘户数、房价、家宽渗透率综合测算）×该楼盘平均 ARPU 值	Min-Max 归一化[0,1]	
4		营销沙盘重点场景	市场划定的高价值营销区域	综合得分=市场部门提供分级评分或布尔运算（楼盘全部在高价值营销场景内=1，部分=0.5，无=0）	直接映射 [0,1]	
5		集团/省公司重点保障场景	明文要求必须保障覆盖的重点场景	同上，布尔运算/分级评分	直接映射 [0,1]	
6		高价值业务受限楼盘	网络问题导致高价值业务开通/体验差的楼盘	综合得分=50%×(投诉工单数/住户数)+20%×业务开通失败率+30%×容量受限标识（是=1，否=0）	Min-Max 归一化[0,1]	
7		高价值携转风险集中楼盘	高价值用户集中、携转风险高的楼盘	综合得分=劣于竞争网络标识×1+携转风险标识×1+(高风险用户数/总用户数)×高风险用户平均 ARPU	Min-Max 归一化[0,1]	
8		竞品低频5G 深度覆盖场景	竞品低频5G 深度覆盖、我方覆盖劣势的楼盘	综合得分=我方覆盖劣于竞争网络的百分比（竞争网络低频已覆盖区域）	Min-Max 归一化[0,1]	
9		投资效益 (M域)	基站 PRB 预期利用率	投资后基站 PRB 资源预期利用水平	综合得分=预测 PRB 利用率-分公司均值	Min-Max 归一化[0,1]
10			楼盘预期移动数据流量	投资后楼盘产生的移动数据流量规模	综合得分=log10(预测用户数*预测单用户 DOU+1)	Min-Max 归一化[0,1]
11	楼盘常驻5G 终端规模		投资后楼盘内活跃5G 终端数量	综合得分=log10(预测用户数*5G 终端渗透率+1)	Min-Max 归一化[0,1]	
12	投资回收期		投资成本通过新增/挽回收益收回的时间	综合得分=1/(1+总投资成本/年化净增效益)	Min-Max 归一化[0,1]	

量化与归一化后得到 12 维特征向量  $x_i=[x_1,x_2,x_3,\dots,x_{12}]$ ，对应 12 个二级指标的归一化得分。

(2) 数据预处理细则：  
□ 缺失值：采用邻近区域均值填充或剔除对应楼盘；  
□ 异常值：对超出[0,1]范围（如利用率>1）的指标，采取截断或按缺失值处理；  
□ 归一化：除已处于[0,1]区间的指标外，其余均通过 Min-Max 归一化统一量纲；  
□ 特征优化：通过 PCA 降维或手动筛选代表性特征，避免模型不稳定<sup>[33]</sup>；  
□ 场景标签：为楼盘 i 分配标签  $S_i \in \{$ “5G 已覆盖存量区域”：“Cov\_Exist”；“5G 未覆盖存量区域”：“Cov\_None”；“新建场景”：“New\_Build” $\}$ ，作为差异化建模核心输入。

### 2.3.4.3 基于 LambdaMART 核心的排序算法

为实现对城区居民区 5G 网络覆盖价值的精准量化与排序，本研究选取 LambdaMART 作为核心排序框架。其选择依据在于，该算法能够直接优化归一化折损累计增益 (NDCG)<sup>[34][35]</sup>这一与投资规划需求高度契合的排序指标，并具备强大的非线性拟合能力以处理多源异构特征间复杂协同与拮抗效应<sup>[36][37]</sup>，同时对特征尺度不敏感，保证了模型的鲁棒性。

在此框架下，我们将“楼盘价值排序”问题形式化为一个学习排序 (Learning to Rank, LTR) 任务。具体而言，将三类场景 (5G 已覆盖存量区域、5G 未覆盖存量区域、新建场景) 分别视为独立的“查询” (Query)，每个查询包含一组待排序的楼盘，楼盘 i 由其特征向量  $X_i$  和场景标签  $S_i$  表征。模型的目标是，针对每个查询，学习一个函数，该函数能为每个楼盘输出一个实数值分数  $f_i$ ，使得同场景内高投资价值的楼盘获得更高的分数，从而在降序排列后得到最优的投资优先级清单。

为了训练此模型，首先需要定义监督信号——即为每个楼盘 i 赋予“真实投资价值”分数  $Y_i$ 。该分数的构造是模型优化的核心，我们通过

融合专家经验、历史项目库数据及强业务规则公式来综合评定，使其蕴含丰富的业务先验知识。

LambdaMART 的核心创新在于其通过定义“Lambda 梯度”来直接优化 NDCG。NDCG 等排序指标具有非平滑、不可导的特性，传统梯度下降法难以直接应用。LambdaMART 巧妙地绕开了这一难题。其推导始于 RankNet 的概率思想：对于任意一对楼盘(i, j)，模型会计算一个预测概率，表示 i 比 j 更相关（应排在前面）的可能性，并通过交叉熵损失来衡量该预测与理想概率（基于真实价值分数）之间的差异。

具体来说，设楼盘(i,j)预测得分为  $f_i, f_j$ ，相关定义如下：

(1) 预测概率：模型判定 i 比 j 更相关（应排在前面）的概率，由 Sigmoid 函数计算：

$$P_{ij}' = \sigma(f_i - f_j) = \frac{1}{1 + e^{-(f_i - f_j)}}$$

(2) 理想概率：基于真实价值分数  $y_i$  和  $y_j$  定义：

$$P_{ij} = \begin{cases} 1 & \text{if } y_i > y_j \\ 0 & \text{if } y_i < y_j \\ 0.5 & \text{if } y_i = y_j \end{cases}$$

(3) 交叉熵成本函数：衡量预测概率与理想概率的差异。

$$C_{ij} = -P_{ij}' \log(P_{ij}') - (1 - P_{ij}') \log(1 - P_{ij}')$$

(4) 梯度计算：成本函数对模型得分  $f_i$  的梯度为：

$$\frac{\partial C_{ij}}{\partial f_i} = P_{ij}' - P_{ij}$$

该梯度明确了模型调整  $f_i$  以降低配对排序错误损失的方向。然而，RankNet 优化的是交叉熵损失，而非最终的 NDCG 指标。LambdaMART 的关键改进在于，它将任意两个样本 i 和 j 顺序交换所引发的 NDCG 变化量  $\Delta NDCG_{ij}$  作为上述梯度的权重，从而实现对 NDCG 指标的直接优化。

对于含 n 个楼盘查询 q，其前 k 个结果的



NDCG 定义为:

$$NDCG_K = Z_k \sum_{i=1}^k \frac{2^{y_i} - 1}{\log_2(i+1)}$$

其中  $y_i$  为第  $i$  位楼盘的真实价值分数,  $Z_k$  为归一化常数, 确保理想排序的 NDCG 值为 1。

设当前排序中样本  $i$  位于位置  $p$ , 样本  $j$  位于位置  $q$  ( $p < q$ ), 交换两者位置后的 NDCG 变化量为:

$$\Delta NDCG_{ij} = \frac{2^{y_i} - 1}{\log_2(q+1)} + \frac{2^{y_j} - 1}{\log_2(p+1)} - \left( \frac{2^{y_i} - 1}{\log_2(p+1)} + \frac{2^{y_j} - 1}{\log_2(q+1)} \right)$$

简化后得:

$$\Delta NDCG_{ij} = (2^{y_i} - 2^{y_j}) \left( \frac{1}{\log_2(q+1)} - \frac{1}{\log_2(p+1)} \right)$$

由  $p < q$  可知  $\frac{1}{\log_2(q+1)} < \frac{1}{\log_2(p+1)}$ , 故

当  $y_i > y_j$  时,  $\Delta NDCG_{ij} < 0$ , 表明高相关样本前置可提升 NDCG, 反之则降低。基于此, 样本  $i$  的最终梯度  $\lambda_i$  定义为所有相关样本  $j$  的加权梯度之和:

$$\lambda_i = \sum_{j: y_i \neq y_j} (P_{ij}' - P_{ij}) \cdot |\Delta NDCG_{ij}|$$

$$\lambda_j = -\lambda_i$$

若模型误将低相关样本  $i$  排在高相关样本  $j$  之前 ( $y_i < y_j$  但  $f_i > f_j$ ), 则:  $P_{ij}' > P_{ij}$ , 梯度项 ( $P_{ij}' - P_{ij}$ ) 为正且  $\Delta NDCG_{ij} > 0$ ,  $\lambda_i$  获得较大正梯度, 推动模型降低  $f_i$  或提升  $f_j$ , 从而纠正排序错误。

整个训练过程采用梯度提升树 (MART) 框架, 形成一个优化闭环: □ 前向预测: 提升树模型输出样本  $i$  的预测得分  $f_i$ ; □ 梯度计算: 基于  $f_i$  和真实标签  $y_i$ , 利用上述公式计算每个样本的  $\lambda_i$ , ( $P_{ij}' - P_{ij}$ ) 决定梯度方向,  $|\Delta NDCG_{ij}|$  赋予梯度权重; □ 模型更新: 以  $\lambda_i$  为“伪残差”训练新回归树, 叠加后更新  $f_i$ , 间接最大化列表 NDCG。

为确保模型不仅在数学上最优, 更在业务上可用, 该模型深度集成了业务规则。一方面, 在

构造真实价值分数  $y_i$  融入业务先验 (例如, 重点场景楼盘设置更高基础分); 另一方面, 在 Lambda 梯度的计算中引入惩罚项, 当模型输出违反了核心业务规则 (例如, 将一个非重点场景楼盘排在所有重点场景之前) 时, 施加额外的梯度惩罚, 强制模型进行纠偏。

此外, 完备的特征工程是模型性能的基石。除了基础的 12 维归一化特征向量外, 还补充了具有明确业务意义的交叉特征, 如“弱覆盖聚集区面积占比 × 家宽价值得分” (高价值区域覆盖短板)、“流量预测 × (1 - 利用率)” (高流量低容量充裕) 等, 并通过场景标签进行 One-Hot 编码, 使模型能够自适应地学习不同场景下的差异化排序偏好。

具体的迭代训练流程如下: (1) 基于当前模型  $F_{t-1}$  计算所有样本的  $\lambda_i$ ; (2) 以这些  $\lambda_i$  梯度为“伪残差”, 训练一棵新的回归树  $h_t$  进行拟合; (3) 通过  $F_t = F_{t-1} + \eta \cdot h_t$  更新模型, 其中  $\eta$  为学习率; (4) 重复上述步骤直至达到指定轮数或验证集 NDCG 不再提升。

通过上述设计, LambdaMART 模型成功地将复杂的多维度业务逻辑转化为可学习的排序任务, 为生成科学、可靠的投资优先级清单提供了坚实的算法支撑。

#### 2.3.4.4 训练数据构造

“楼盘-真实投资价值”配对数据的获取是排序学习的核心挑战。提出以下三个构造方法:

(1) 基于项目库的优先级映射: 构建 1-2 年网络建设项目库, 含楼盘 ID、投资类型、金额及时间等信息; 若楼盘  $i$  投资时间早于  $j$  且投资类型紧急度更高 (如扩容拥塞解决 vs 常规补盲), 则定义  $y_i > y_j$ , 进一步将  $y_i$  量化为投资优先级系数, 结合投资决策纪要、问题紧急程度 (如影响用户数量) 及投资效益后评估报告等, 综合评定为 1-10 分的量化值。

(2) 专家经验排序标注：汇总网络规划、建设、优化、市场等部门专家，对覆盖各类场景与特征组合的代表性楼盘进行独立排序或打分；基于贝叶斯个性化排序 (BPR) 方法融合专家排序结果，或直接以平均序位作为  $y_i$ 。同步记录排序核心业务规则。专家仅依据楼盘基础信息与经验完成场景内排序，其结果构成高质量真实排序数据。

(3) 业务规则驱动弱监督构造：设计强业务规则公式计算楼盘初始价值分  $y_{synth}$ ，作为初始训练目标训练第一版模型，公式示例：

5G 已覆盖存量区域： $y_{synth}=0.4*弱覆盖聚集区+0.3*业务受限+0.2*TOP 扇区负荷+0.1*携转风险$

5G 未覆盖存量区域： $y_{synth}=0.3*家宽价值+0.3*流量预测+0.2*投资回报率(倒数)+0.1*营销沙盘+0.1*友商覆盖$

新建场景： $y_{synth}=0.4*重点场景+0.3*营销沙盘+0.2*流量预测+0.1*常驻终端$

训练数据最终形式：一个数据集，每条记录为：(楼盘 ID, 场景标签, 特征向量  $x$ , 真实价值分数  $y$ )。该数据集按场景分组，用于训练 LambdaMART 模型。

### 2.3.4.5 差异化排序规则

LambdaMART 模型以场景为查询维度自适应学习排序偏好，为贴合业务战略并提升结果可解释性，需在数据构造、特征分析、阈值后处理环节嵌入差异化规则。

在 5G 已覆盖存量区域，核心目标是破解资源有限与网络问题广泛的矛盾，精准定位短板与高价值用户区重叠点，深挖存量价值并优化用户体验。指标设定上，O 域携转风险为最高优先级，弱覆盖聚集区、质差 TOP 扇区权重占优；B 域紧迫性指标辅助区分同质楼盘优先级；M 域利用率、流量验证优化必要性，投资回报率权重下调。模型引导遵循“网络质量与用户需求优先”的

排序信号。

在 5G 未覆盖存量区域，核心目标是权衡建设成本与收益风险，评估覆盖价值和投资效率，实现低成本有效覆盖并应对市场竞争。指标优先级设定为：M 域投资价值、B 域价值区域指标并重，高赋权价值楼盘、流量预测、投资回收期等指标；B 域竞争风险为核心考量，友商覆盖、高携转风险楼盘优先布局；O 域指标因取值无效或无参考性，赋予低权重或零权重。模型引导学习“高收益、高效率、防风险”排序逻辑，平衡投资回报与收益指标。

在新建场景，核心目标是平衡远期战略与近期投资不确定性，识别战略高地和高潜力增长点，落实前瞻性布局。指标设定以 B 域战略指标为核心，重点场景优先级最高，营销沙盘场景次之；M 域流量、终端预测为中长期评估依据，投资回报率权重低于未覆盖存量区；O 域指标不适用。模型引导遵循“战略优先、潜力评估”原则，重点场景可赋固定高  $y$  值保障优先级。

### 2.3.4.6 动态优先级判定阈值

模型输出  $[0,100]$  综合得分后，结合场景特性、预算及项目需求，动态划定高、中、低三级阈值。

在 5G 已覆盖存量区域，高优先级为质量驱动，高价值区存严重弱覆盖/容量瓶颈，需即刻优化、扩容或建站；中优先级为价值驱动，轻量网络问题或高价值/高风险用户集聚，纳入常规运维，按需补盲；低优先级为网络达标或价值密度低，归入长期问题库，周期监测滚动评估。

在 5G 未覆盖存量区域，高优先级为效益风险双驱动，回报期短、收益高或竞争激烈，优先建宏站抢占市场；中优先级为风险/潜力驱动，效益中等但存竞争风险或潜力，采用共建共享或纳入二批建设；低优先级为效益低、风险小、需求弱，暂缓建设并跟踪区域发展。

在新建场景，高优先级为战略驱动，属重点/



高价值沙盘且业务密度高，同步建宏站/室分，楼起网通；中优先级为潜力驱动，沙盘价值场景为主、流量预测适中，按预算与市场节奏部署基础覆盖；低优先级为战略与业务潜力不足，暂不纳入同步建设，待区域成熟后复评。

### 3 模型对比验证

为客观全面评估所提模型的实际性能与先进性，本章设计并开展系列对比验证实验，将其与传统5G网络覆盖评估方法在准确性、精细化程度、决策价值及评估效率等维度进行深度对比，通过量化与定性结合的方式，验证模型相较于传统方法的显著优势。

#### 3.1 实验环境及数据

##### 3.1.1 实验环境

实验环境配置如表4所示：

表4 实验环境配置

序号	名称	配置信息
1	操作系统	Windows 11
2	开发环境	conda 24.11.3
3	开发语言	Python 3.12.7
4	集成开发环境	Spyder 5.5.1
5	CPU	Intel(R)Core(TM)i5-12500H(2.5GHz)
6	GPU	GeForceRTX4060 (8G)
7	内存	16GB

Python关键依赖库如表5所示：

表5 Python关键依赖库

序号	名称	版本号	备注
1	Geopandas	1.0.1	地理空间数据处理
2	geometry	0.0.23	几何计算
3	SciPy	1.13.1	几何计算
4	Scikit-learn	1.5.1	评估指标计算
5	bayesian-optimization	1.5.1	贝叶斯优化实现
6	XGBoost	3.1.2	价值排序模型训练
7	LightGBM	4.6.0	梯度提升树建模
8	Scikit-plot	0.3.7	机器学习模型

##### 3.1.2 实验数据

以某地区3.7万个居民区楼盘为研究对象，采用2025年6-8月的地理、网络、用户、市场四类数据。

### 3.2 对比方法设计

#### 3.2.1 对比实验真值构建

通过三重交叉验证构建高可信度真值集：□融合高频用户投诉数据；□基于DT路测获取米级精度的信号强度与质量数据；□专家团队结合历史经验与区域认知，人工审核疑似弱覆盖点。经三者共同验证的楼盘标记为“真实弱覆盖楼盘”，作为评估基准。评估指标聚焦于识别弱覆盖区域与真值的吻合程度。

#### 3.2.2 核心评估指标定义

##### (1) 弱覆盖识别准确率

精确率(Precision)：模型判定为弱覆盖的楼盘中真实弱覆盖楼盘的占比，计算公式为：

$$Precision = \frac{TP}{(TP + FP)}$$

其中，TP为正确识别的弱覆盖楼盘数，FP为误判为弱覆盖的楼盘数，精确率越高，模型误报率越低。

召回率(Recall)：真实弱覆盖楼盘中被模型成功识别的占比，计算公式为：

$$Recall = \frac{TP}{(TP + FN)}$$

其中，FN为漏检的真实弱覆盖楼盘数，召回率越高，模型漏报率越低。

F1分数(F1<sub>Score</sub>)：精确率与召回率的调和均值，用于综合评价模型性能，计算公式为：

$$F1_{score} = \frac{2(Precision * Recall)}{(Precision + Recall)}$$

##### (2) 价值排序一致性

通过市场、营销及网络规划专家对真值弱覆盖楼盘进行价值评分与排序，构建专家基准排序。分别采用传统单指标与本文LambdaMART

算法综合评分，生成对应排序列表。以  $NDCG@k$  衡量前  $k$  个结果与专家排序一致性，取值范围为  $[0,1]$ ，数值越接近 1 表示排序效果越优。

### (3) 问题定位粒度与空间精度

选取同一区域，分别对传统方法与所提模型的评估结果进行可视化呈现与对比分析。

## 3.3 对比实验结果与分析

### 3.3.1 准确性对比分析

在包含 500 个经多方验证的“真实弱覆盖楼盘”的测试集上，对两种方法的弱覆盖识别能力进行了严格的量化评估。

表 6 准确性对比分析结果

估指标	传统覆盖评估方法	本研究模型	性能提升
精确率	35.2%	92.5%	57.3PP
召回率	41.6%	90.8%	49.2PP
F1 分数	38.1%	91.6%	53.5PP

由表中数据可见，模型在弱覆盖识别各准确性指标上显著优于传统方法。传统方法以小区为评估粒度，对比时需降维至楼盘层面：若判定某小区为弱覆盖，其覆盖范围内所有楼盘均标记为弱覆盖，直观体现其粒度粗糙性。传统方法性能欠佳的核心原因：高误报率（小区内少数楼盘信号衰减则全小区被判弱覆盖）、高漏报率（小区整体信号达标时，局部弱覆盖楼盘被平均化效应掩盖）。模型以楼盘为粒度进行特征提取与预测，融合楼宇轮廓、高度、周边环境等地理空间信息及精细化 MR 采样数据，有效规避平均化效应，显著降低 FP 与 FN，提升综合 F1 分数。

### 3.3.2 价值排序一致性对比分析

将两种方法产出的弱覆盖楼盘排序列表，与专家基准排序进行比较，计算其  $NDCG@50$ （即评估排名前 50 位的结果）得分。

$NDCG$  得分差异显著，体现模型核心价值——兼具问题识别与精准决策能力。运营商可基于模型输出的 TopN 列表，将有限资本性支出

表 7 价值排序一致性分析结果

评估指标	传统覆盖评估方法	本研究模型
5G 已覆盖存量区域	0.28	0.96
5G 未覆盖存量区域	0.29	0.95
新建场景	0.29	0.96

（CAPEX）优先配置于用户价值、市场竞争优势及长期收益最优区域。

### 3.3.3 精细化与效率对比分析

为了直观展示两种方法在定位粒度上的差异，选取城市中心一个典型高密度住宅区进行案例分析。

传统方法输出小区级大面积高亮区域，优化目标模糊，仅能给出小区级优化指令，需大量现场勘察；模型可精确定位至楼栋级点/面目标，直接输出“优化 XX 小区 A 栋、B 栋”的明确指令，为微站及室内分布系统精准部署提供直接指导，显著提升优化效率与资源利用率。

## 3.4 关键算法敏感性分析

为保障模型所用机器学习算法的鲁棒性、稳定性与可解释性，本节对聚类和排序两大核心算法开展系统敏感性分析，通过实验设计与统计检验量化关键超参数对输出性能的影响，为实际部署的参数调优提供科学依据。

### 3.4.1 基于贝叶斯优化的 DBSCAN 聚类算法敏感性分析

在区域问题分析模型中，DBSCAN 结合贝叶斯优化用于弱覆盖栅格聚类。本分析旨在评估优化配置对聚类质量与收敛速度的影响。

#### (1) 目标与响应变量

聚类质量：DBC<sub>V</sub> 指数（Density-Based Clustering Validation Index, DBC<sub>V</sub>），为密度聚类专用的质量量化指标，取值范围为  $[0,1]$ ，值越接近 1 表示聚类质量越优（簇内紧密性与簇间分离度越好）。

优化效率：收敛轮数、优化耗时。

输入参数：采集函数（Expected Improve-

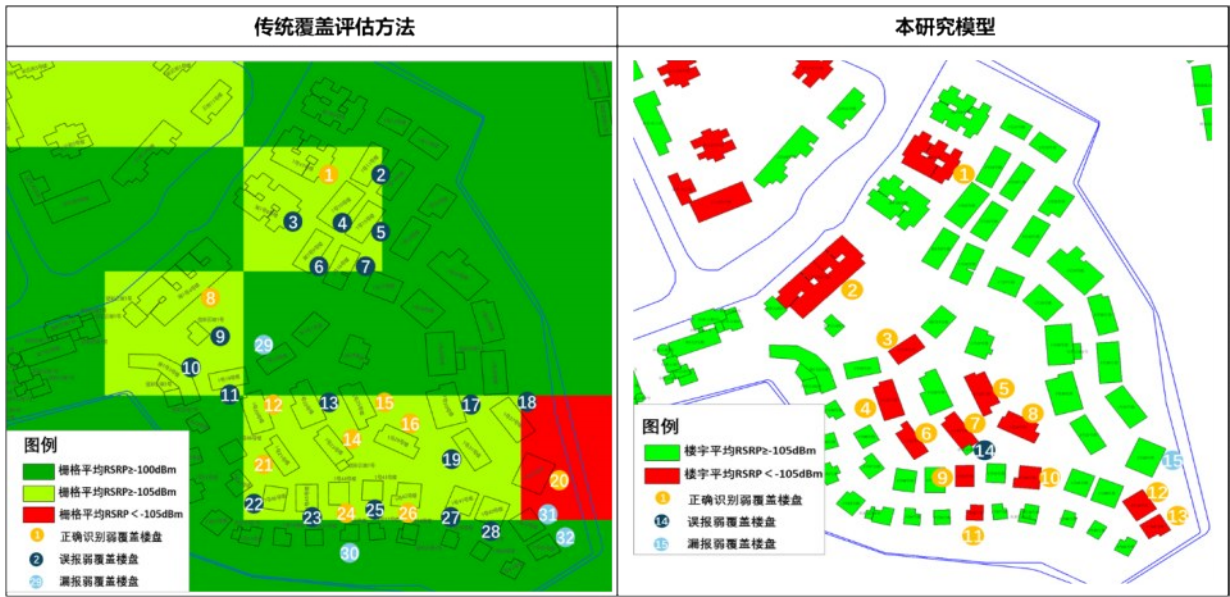


图8 模型输出结果精细化与效率对比分析图

ment, EI; Upper Confidence Bound, UCB; Probability of Improvement, PI)、初始点数 (5/10/20)、迭代次数上限 (30/50/100)。

(2) 分层实验设计

采用3因素3水平全因子设计, 在3类场景(平原、丘陵、山地)下运行, 共27组实验。使用ANOVA元分析跨场景稳健性。

(3) 执行分析与可视化

ANOVA元分析表明, 采集函数类型对DBCW影响最显著 (p<0.001), 初始点数对收敛轮数影响显著。

平行坐标图: 展示多参数组合与多输出响应的关系, 如图9所示。

(4) 敏感性分析小结

- 采集函数EI在多数场景下获得最高

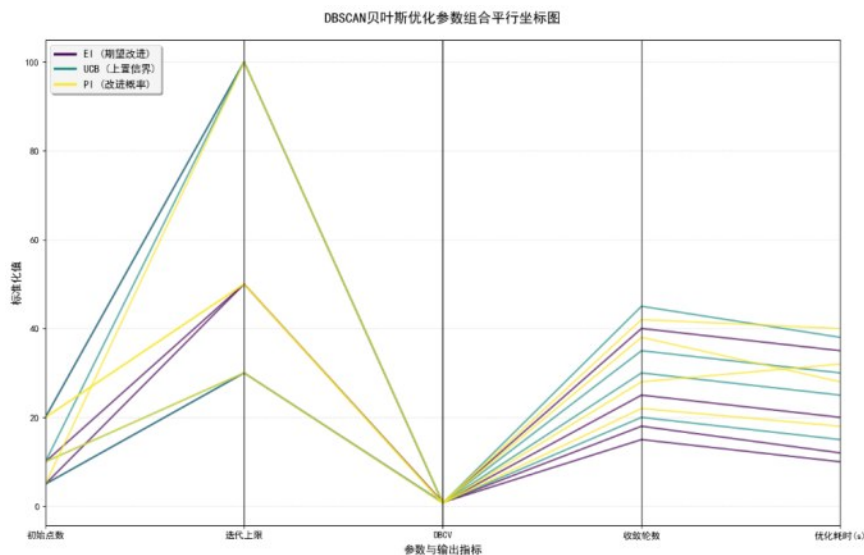


图9 DBSCAN 贝叶斯优化参数组合平行坐标图

表 8 基于贝叶斯优化的DBSCAN 聚类算法 ANOVA 方差分析

因素	DBCVC		收敛轮数	
	F 值	p 值	F 值	p 值
采集函数	25.67	<0.001	8.91	0.001
初始点数	6.34	0.005	15.23	<0.001
迭代上限	2.11	0.136	3.45	0.045
数据场景	4.56	0.018	5.78	0.007

DBCVC (0.85~0.88)，推荐作为默认选择。

□ 初始点数 10 为最优折中，过少易陷入局部最优，过多增加计算开销。

□ 迭代上限 50 轮可满足大多数场景收敛，继续增加收益递减。

### (5) 复杂度分析

DBSCAN 算法的复杂度高度依赖空间索引机制与参数配置，结合贝叶斯优化后整体开销需综合评估。

□ 时间复杂度：

原始 DBSCAN：采用 BallTree 空间索引，单次执行时间复杂度为  $O(N \log N)$ ，其中 N 为弱覆盖栅格数量。

贝叶斯优化：每次迭代需训练高斯过程代理模型。训练复杂度为  $O(T^3)$ ，T 为迭代次数，由  $T \ll N$ ，且仅作用于超参空间，其开销远小于 DBSCAN。预测阶段为  $O(T^2)$ ，可忽略不计。需二次精细聚类。整体时间复杂度为  $O(2 \times (T \cdot N \log N + T^3))$ 。

模型实验数据中  $N \approx 20$  万， $T=50$ ，理论浮点运算量约  $1.06 \times 10^8$  次，现有硬件环境下，单次聚类耗时 <2 秒，单区域 BO 耗时 <3 分钟，经多区域并行计算，实测整体流程耗时 10 分钟。

□ 空间复杂度：需存储样本点坐标 (2N)、核心距离、可达距离及聚类标签、BallTree 索引额外开销，空间复杂度  $O(6N)$ 。内存占用约  $(2 \times 10^5 \times 6) \times 8 \text{Bytes} \approx 9.16 \text{MB}$

□ 综上，实际部署中可通过区域划分并行计算进一步压缩耗时，算法在内存与时间消耗上均具备良好的工程可行性。

### 3.4.2 基于 LambdaMART 排序算法敏感性分析

在价值排序模型中，LambdaMART 算法用于楼盘优先级排序。本分析采用两阶段实验，识别关键参数并建模优化。

#### (1) 目标与响应变量

排序质量：NDCG@50、平均精度均值 (Mean Average Precision, MAP)。

输入参数 (候选)：决策树数量 ( $n\_estimators$ )、学习率 ( $learning\_rate$ )、决策树最大深度 ( $max\_depth$ )、子节点最小样本数 ( $min\_child\_samples$ )、特征采样比例 ( $colsample\_bytree$ )。

#### (2) 两阶段实验设计

阶段一 (筛选)：采用 12 次试验的 Plackett-Burman 设计，旨在从 5 个候选输入参数中筛选出 3 个关键影响参数，最终确定  $n\_estimators$ 、 $learning\_rate$ 、 $max\_depth$  为核心参数；

阶段二 (建模与优化)：针对上述关键参数采用中心复合设计，共设置 20 组验证试验，进而构建参数与排序质量指标间的响应面模型。

#### (3) 执行分析与可视化

响应面分析：构建二次回归模型，ANOVA 显示模型显著 ( $p < 0.001$ )， $R^2=0.92$ 。等高线图如图 10 所示。

交互效应图：如图 11 所示， $n\_estimators$  与  $learning\_rate$  存在明显交互作用，低学习率需更多树才能收敛。

#### (4) 敏感性分析小结

最优参数组合： $n\_estimators=500$ 、 $learning\_rate=0.05$ 、 $max\_depth=6$ ，此时  $NDCG@50=0.95$ 。

学习率对排序质量影响最敏感，推荐使用较小值 (0.05~0.1) 配合足够树的数量。

$max\_depth$  在 6~8 之间效果最佳，过深易过

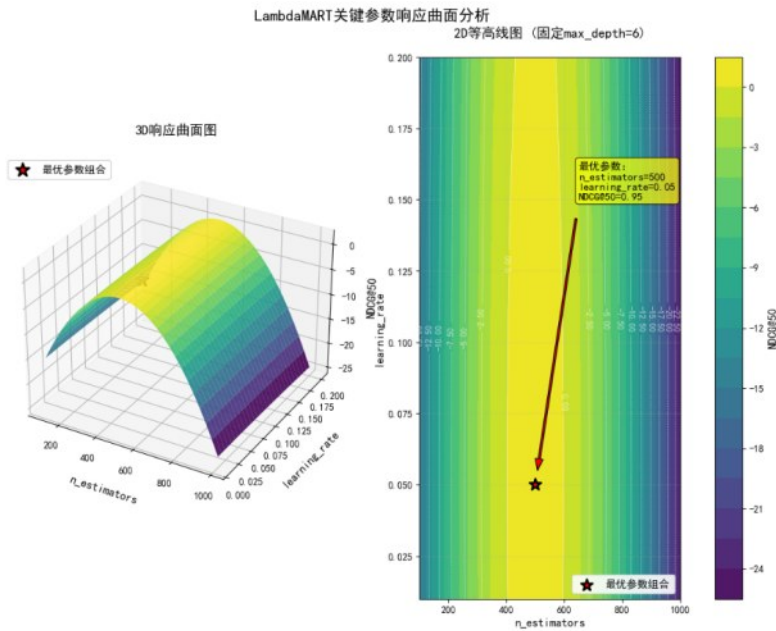


图 10 LambdaMART 关键参数响应曲面图

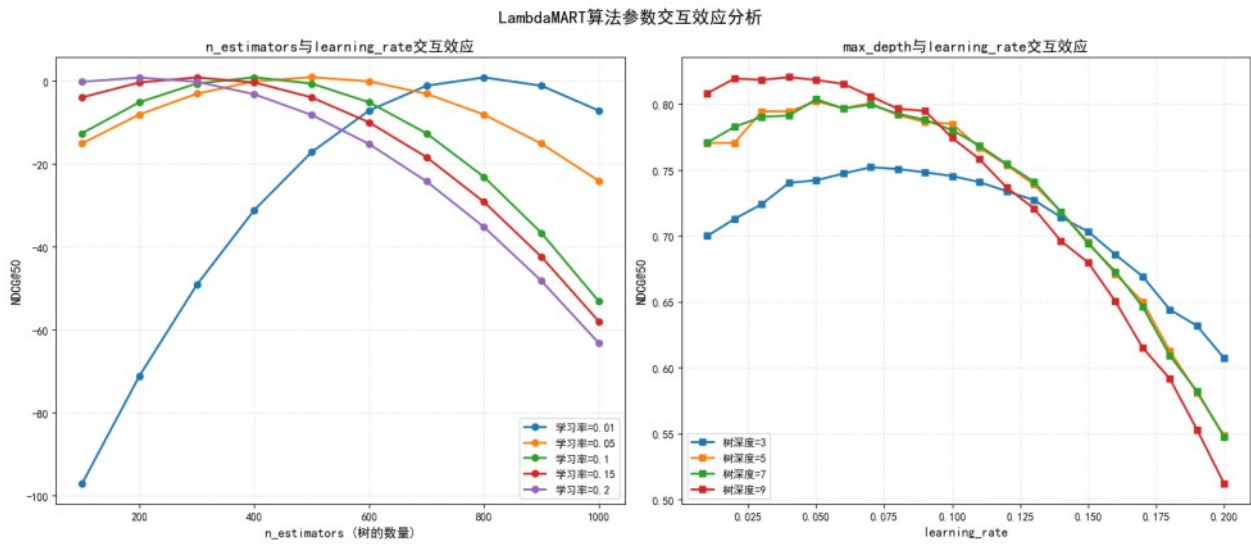


图 11 LambdaMART 算法参数交互效应图

拟合，过浅则拟合不足。

### (5) 复杂度分析

LambdaMART 作为基于 GBDT 的排序算法，其复杂度取决于树的数量、深度及样本规模。

训练阶段：

设训练样本数为  $N$ ，特征维度为  $d$ ，树的数量为  $M$ ，每棵树最大深度为  $D$ 。每轮构建一棵树需遍历所有特征与分割点，模型采用 LightGBM

通过直方图近似将单棵树构建复杂度降至  $O(N \cdot d \cdot \log N)$ ，故总训练时间复杂度为  $O(M \cdot N \cdot d \cdot \log N)$ 。

空间复杂度主要包括存储样本、梯度及树结构，约为  $O(N \cdot d + M \cdot 2^D)$ 。

本实验  $N=37000$ ， $d=12$ ， $M=500$ ， $D=6$ ，理论计算量约  $1.014 \times 10^9$  次比较操作。使用 LightGBM 优化后，实际训练耗时约 5 分钟。内存占用

约  $(37000 \times 12 + 500 \times 2^6) \times 8\text{Bytes} \approx 9.16\text{MB}$

推理阶段:

对单个样本, 需遍历  $M$  棵树的路径, 每棵树最多访问  $D$  个节点, 故单样本推理时间为  $O(M \cdot D)$ 。批量推理 (batch size= $B$ ) 时间复杂度为  $O(B \cdot M \cdot D)$ , 空间复杂度为  $O(B \cdot d + M \cdot 2^D)$  (模型常驻内存)。

本实验  $B=5000$ , 理论计算量约  $1.5 \times 10^7$  次比较操作。10 万楼盘批量预测 < 1 秒。内存占用约  $(5000 \times 12 + 500 \times 2^6) \times 8\text{Bytes} \approx 0.7\text{MB}$

综上, LambdaMART 算法在合理配置下具备快速训练与极速推理能力, 模型轻量化, 非常适用于运营商日常滚动评估与即时决策场景。

### 3.5 模型稳定性与普适性验证

为评估模型在异质环境下的稳健性与适用性, 选取跨地形、不同规模等级城市为测试场景, 验证其对城市社会经济属性差异引发的数据分布、场景复杂度及业务目标变化的适配能力。重庆下辖各区县城区地貌与城市形态丰富, 为模型普适性验证提供天然试验场。选取平原、丘陵、山地各 5 个城市楼盘开展验证, 结果如下:

(1) 弱覆盖识别准确率在三类地形城市中分别为 90.7%、92.2%、93.3%, 单类城市内波动  $\leq 0.05\%$ , 类间波动  $\leq 2.6\%$ ;

(2) 价值排序一致性方面, 模型预测结果与

人工专家评估分数的 Pearson 相关系数  $r \geq 0.92$  ( $P < 0.01$ ), 表明模型投资价值评估逻辑在不同地形场景下具有良好稳定性与统计显著性;

(3) 算法收敛性优良, 同一批数据重复计算 5 次后, 弱覆盖聚集区边界重合度  $\geq 95\%$ , 价值评分标准差  $\leq 0.03$ , 证明模型输出结果可复现。

### 3.6 小结

对比验证结果表明, 模型在评估精度、效率、资源优化三个维度均显著优于传统方法, 且在不同地形、规模城市中具有良好的稳定性与普适性, 验证了模型的科学与实用性。

## 4 结论

本文构建的城区居民区 5G 覆盖评估模型, 基于多源异构数据融合与 LambdaMART 价值排序机制, 实现 92.5% 的弱覆盖识别准确率、 $\text{NDCG} > 0.9$  的决策效能, 为网络精细化运营提供可复用的方法论支撑。当前模型仅经重庆地区验证, 缺乏超大型城市核心区应用数据, 且单运营商数据架构在跨运营商、IoT 终端等多源数据融合适配方面仍需拓展。

未来研究将聚焦三方面: (1) 引入数字孪生与轻量化联邦学习框架, 提升高密度场景建模精度与多主体数据协同效率; (2) 适配 6G 太赫兹频段传播特性与“通信-感知-计算”一体化需求,

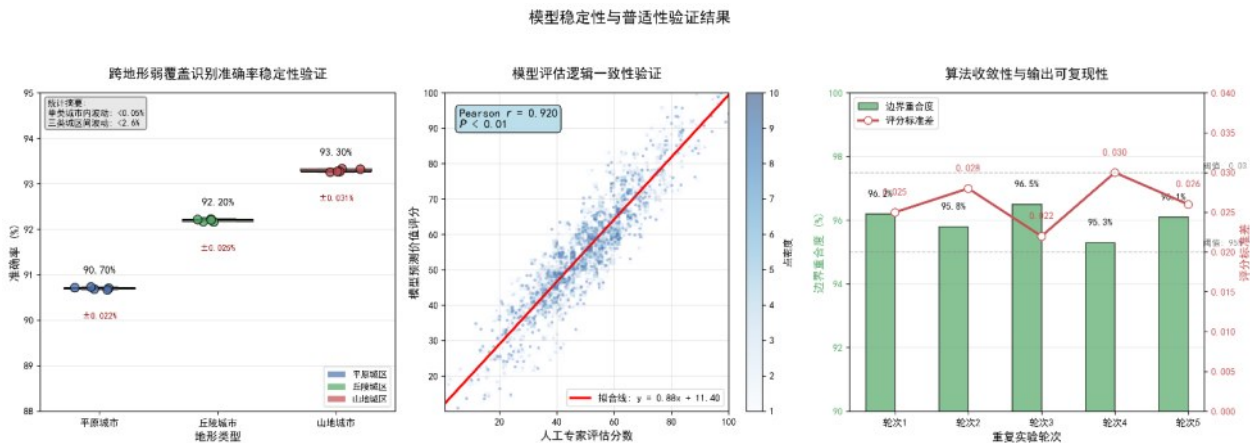


图 12 模型稳定性与普适性验证



升级模型为早期规划工具，通过虚拟网络仿真预判覆盖盲区，结合AI生成式设计完成资源预配置，支撑6G网络“数字孪生先行”规划<sup>[38]</sup>；（3）构建动态时空评估机制，强化模型对用户潮汐效应、突发流量等场景的自适应响应能力。本研究框架兼具理论延展性与工程迁移性，为未来移动通信网络智能化规划提供基础支撑。

## 5 参考文献：

- [1] Hossain E, Hasan M. 5G cellular: key enabling technologies and research challenges[J]. *IEEE Instrumentation & Measurement Magazine*, 2015, 18(3): 11-21. [LinkOut]
- [2] Suryanto S, Ningsih Y K. Overview of 5G services and spectrum deployment in urban regions[J]. *Jurnal Sains, Teknologi dan Informatika*, 2024, 11(1): 91-102. [LinkOut]
- [3] 周灿, 史文祥, 王森, 等. 基于OTT MR数据的规建维优营分析体系研究[J]. *邮电设计技术*, 2024(10): 42-48.  
Zhou C, Shi W X, Wang S, et al. Research on analysis system of PCMOM(planning, construction, maintenance, optimization, marketing)based on OTT MR data[J]. *Designing Techniques of Posts and Telecommunications*, 2024(10): 42-48. [万方]
- [4] 陈玲, 张鹏, 龙青良, 等. 基于多维价值评估的楼宇分类体系及场景化解决方案[J]. *邮电设计技术*, 2023(2): 54-57.  
Chen L, Zhang P, Long Q L, et al. Building classification system and scenario solution based on multi-dimensional valuation [J]. *Designing Techniques of Posts and Telecommunications*, 2023(2): 54-57. [万方]
- [5] Usa D E, Vasdev K. GIS for 5G network deployment: optimizing coverage and capacity with spatial analysis[J]. *Journal of Artificial Intelligence & Cloud Computing*, 2022, 1(3): 1-3. [LinkOut]
- [6] Wen L, Tan H, Li W H, et al. Evaluating the coverage of 5G signals: coupling spatial autocorrelation and 3D sight lines based on GIS[C]//*Mobile Networks and Management*. Cham: Springer, 2023: 255-264. [LinkOut]
- [7] 孟炜谱, 刘旭东, 苗二龙, 等. 基于众筹模式的5G网络评估与优化[J]. *邮电设计技术*, 2023(10): 50-53.  
Meng W P, Liu X D, Miao E L, et al. 5G network evaluation and optimization based on crowdfunding model[J]. *Designing Techniques of Posts and Telecommunications*, 2023(10): 50-53. [知网][知网英文]
- [8] Martínez Abuabara R, Robles C A, Camargo L L. A multicriteria model for prioritizing 5G network deployment with Monte Carlo stability analysis: a case study in Magdalena, Colombia [J]. *PLoS One*, 2025, 20(10): e0334781. [LinkOut]
- [9] Liu B L, Li Q, Zheng Z H, et al. A review of multi-source data fusion and analysis algorithms in smart city construction: facilitating real estate management and urban optimization[J]. *Algorithms*, 2025, 18(1): 30. [LinkOut]
- [10] Heintz F . DyKnow: A Stream-Based Knowledge Processing Middleware Framework[J]. Linköping University Electronic Press, 2009.
- [11] Hoffman D D, Singh M, Prakash C. The interface theory of perception[J]. *Psychonomic Bulletin & Review*, 2015, 22(6): 1480-1506. [LinkOut]
- [12] Zhang W Q, Zhao K, Li P, et al. A closed-loop perception, decision-making and reasoning mechanism for human-like navigation[C]//*Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 2022: 4717-4724. [LinkOut]
- [13] Mugglin A S, Carlin B P, Gelfand A E. Fully model-based approaches for spatially misaligned data[J]. *Journal of the American Statistical Association*, 2000, 95(451): 877-887. [LinkOut]
- [14] Chaves-Villota A, Viteri-Mera C A. DeepREM: deep-learning-based radio environment map estimation from sparse measurements[J]. *IEEE Access*, 2023, 11: 48697-48714. [LinkOut]
- [15] Li P Z, Wang X Y, Piechocki R, et al. Variational autoencoder assisted neural network likelihood RSRP prediction model[C]//*Proceedings of the 2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. Piscataway: IEEE Press, 2022: 554-559. [LinkOut]
- [16] Ma L, Jin N D, Zhang Y L, et al. RSRP difference elimination and motion state classification for fingerprint-based cellular network positioning system[J]. *Telecommunication Systems*, 2019, 71(2): 191-203. [LinkOut]
- [17] Wegner T . *Applied Business Statistics: Methods and Excel-Based Applications*[J]. 2010.
- [18] Kupper L L, Hafner K B. How appropriate are popular sample size formulas?[J]. *The American Statistician*, 1989, 43(2): 101-105. [LinkOut]
- [19] Greenland S. On sample-size and power calculations for studies using confidence intervals[J]. *American Journal of Epidemiology*, 1988, 128(1): 231-237. [LinkOut]
- [20] Kyrgyzov I O . Mining satellite image database of landscapes and application for urban zones: clustering, consensus and categorisation[J]. 2008.
- [21] Zhang S J, Xia Y, Li Z J, et al. An assessment of urban residen-

- tial environment quality based on multi-source geospatial data: a case study of Beijing, China[J]. *Land*, 2024, 13(6): 823. [LinkOut]
- [22] Xia Z X, Chong S. WiFi-based indoor passive fall detection for medical Internet of Things[J]. *Computers and Electrical Engineering*, 2023, 109: 108763. [LinkOut]
- [23] Bergstra J, Bardenet R, Bengio Y, et al. Algorithms for Hyper-Parameter Optimization[J]. *Advances in Neural Information Processing Systems*, 2011:2546-2554.
- [24] Li X, Guo H, Xie W L, et al. A 5G coverage calculation optimization algorithm based on multifrequency collaboration[J]. *Electronics*, 2023, 12(19): 4044. [LinkOut]
- [25] Fauzan A, Fadillah G, Fitria A, et al. Cluster mapping of waste exposure using DBSCAN approach: study of spatial patterns and potential distribution in bantul regency[J]. *International Journal on Informatics Visualization*, 2024, 8(2): 751. [LinkOut]
- [26] 夏鲁宁, 荆继武. SA-DBSCAN: 一种自适应基于密度聚类算法[J]. *中国科学院研究生院学报*, 2009, 26(4): 530-538.  
Xia L N, Jing J W. SA-DBSCAN: a self-adaptive density-based clustering algorithm[J]. *Journal of the Graduate School of the Chinese Academy of Sciences*, 2009, 26(4): 530-538. [万方]
- [27] Monko G J, Kimura M. Optimized DBSCAN parameter selection: stratified sampling for epsilon and gridsearch for minimum samples[C]//*Proceedings of the Software Engineering and Automation. Academy & Industry Research Collaboration Center*, 2023: 43-61. [LinkOut]
- [28] Karami A, Johansson R. Choosing DBSCAN parameters automatically using differential evolution[J]. *International Journal of Computer Applications*, 2014, 91(7): 1-11. [LinkOut]
- [29] Matsumoto N, Hamakawa Y, Tatsumura K, et al. Distance-based clustering using QUBO formulations[J]. *Scientific Reports*, 2022, 12: 2669. [LinkOut]
- [30] Zhang X L, Zhou S B. WOA-DBSCAN: application of whale optimization algorithm in DBSCAN parameter adaption[J]. *IEEE Access*, 2023, 11: 91861-91878. [LinkOut]
- [31] 方洪鹰. 数据挖掘中数据预处理的方法研究[D]. 重庆: 西南大学, 2009. [知网硕士中文]
- [32] 王满, 徐朝军. 网络课程资源自动量化评价研究[J]. *现代图书情报技术*, 2010(1): 88-93.  
Wang M, Xu C J. Study on automatic quantitative evaluation of web course resources[J]. *New Technology of Library and Information Service*, 2010(1): 88-93. [知网][知网英文]
- [33] 唐懿芳, 钟达夫. 主成分分析方法对数据进行预处理[J]. *广西师范大学学报(哲学社会科学版)*, 2002, 38(S1): 223-225.  
Tang Y F, Zhong D F. Principal component analysis is used to preprocess the data[J]. *Journal of Guangxi Normal University (Philosophy and Social Sciences Edition)*, 2002, 38(S1): 223-225. [知网][Sogou 标题中译英]
- [34] Burges C J C. From RankNet to LambdaRank to LambdaMART: An Overview[J]. *learning*, 2010.
- [35] Haenelt K. Ranking with Boosted Decision Trees[J]. 2012.
- [36] Fu Y J, Ge Y, Zheng Y, et al. Sparse real estate ranking with online user reviews and offline moving behaviors[C]//*Proceedings of the 2014 IEEE International Conference on Data Mining*. Piscataway: IEEE Press, 2014: 120-129. [LinkOut]
- [37] Zhang X, Ren Z, Li X, et al. Phocus: Picking Valuable Research from a Sea of Citations[J]. 2022
- [38] Munir M S, Kim K T, Adhikary A, et al. Neuro-symbolic explainable artificial intelligence twin for zero-touch IoE in wireless network[J]. *IEEE Internet of Things Journal*, 2023, 10(24): 22451-22468. [LinkOut]