



中国研究生创新实践系列大赛  
“华为杯”第十六届中国研究生  
数学建模竞赛

学 校 厦门大学

---

参赛队号 19103840007

---

1.陈熠

---

队员姓名 2.宋宁宇

---

3.季燕妮

---

中国研究生创新实践系列大赛

“华为杯”第十六届中国研究生

数学建模竞赛

题 目 无线智能传播模型

---

摘 要：

本文针对无线信号的传播问题采用特征工程与深度学习算法开展无线智能传播模型的构建。一个模型好坏不仅取决于算法的好坏，还取决于特征构建的成功与否。我们所构建的特征分为空间几何特征和小区环境统计特征，空间几何特征包括栅格与发射机的平面距离、空间距离以及与信号线所成夹角等，小区环境统计特征则包括小区海拔、建筑高度的均值、方差、极值等统计指标。我们采用 LighGBM 与人工神经网络算法训练模型。最终模型预测结果良好，在交叉验证集上做 RSRP 值预测时两个模型的均方根误差(RMSE)均稳定在 9.0 左右，相对只根据均值预测 RSRP 的 baseline 模型以及一般的线性回归模型都有较大提升。

关键字：RSRP 预测； LightGBM； 特征工程

# 1、问题背景与重述

## 1.1 问题背景

移动通信技术的发展经历一个漫长的阶段，并且使用范围和使用人群正在不断的扩大。1920 年，通信系统仅在船舶和军事方面使用，且频段较少[1]。随后移动通信开始走入公众视野，由于技术的不成熟，人工续接导致网络容量较小。直至 1960 年，美国的改进型移动电话系统使得无线频道能够自动续接到公话网，蜂窝状移动通信网也于随后建成[2]。然而第一代模拟蜂窝网频谱具有供需失衡、设施复杂、通话隐蔽性不强等缺点，经过工程师们孜孜不倦的研究和实践，陆续推出了以 GSM 和窄带 CMDA（Code Division Multiple Access）为代表的第二代移动通信系统、IMT-2000（即 3G）[3]。3G 系统采取码分多址和分组交换，极大的满足了用户对高传输速率、语音通信和多媒体通信等增值服务的需求[4]。基于 3G 技术，3GPP（3<sup>rd</sup> Generation Partnership Project）提出了长期演进系统 LTE（Long Term Evolution）。LTE 因更高的服务质量，更高的宽带和更低的时延被认为是由 3G 向 4G 演进的主流技术[5]。LTE 包括 FDD 方式和 TDD 方式，而后 TDD 方式发展成为 TD-LTE。5G 概念于 2001 年由日本 NTT 公司提出，2013 年 5 月，三星集团宣称以 NoLA 为技术基础研发出首个基于 5G 技术的移动传输网络[6]。5G 网络在实际应用过程中表现出更加强化的功能，并且理论上其传输速度更高，是 4G 移动网络的几百倍。5G 的性能目标是高数据速率、减少延迟、节省能源、降低成本、提高系统容量和大规模设备连接。目前，移动互联网发展的速度迅猛，在用户的日常生活场景中占据重要地位，用户对移动网络流量的要求日益提升。我国已着手 5G 系统的技术研发和战略布局，抢占 5G 市场。

通信技术是信息交互的核心，在这个信息爆炸的时代越来越受人们关注。随着 LTE 技术的逐渐深入，在开展大规模通讯业务的同时也面临着越来越多的挑战。人们对通信的质量提出了更高的要求，要求网络信号覆盖更全，信号质量更高。

## 1.2 问题的提出

在进行网络规划时，需要预测传播损耗以确定基站覆盖范围。传播损耗预测的精度直接影响到后期的网络覆盖效果，若规划出现偏差，将增加后期维护和优化的工作量。因此传播损耗的预测至关重要。在预测传播损耗时，通常采用无线传播模型对目标通信覆盖区域内的无线电波传播特性进行预测，预测的结果可以估算出小区的覆盖范围、小区间网络干扰以及通信速率等指标。影响无线电波传播最重要的因素是无线环境，与有线通信相比，无线环境存在着环境时变以及复杂等特点。无线电波的传播会受到传播路径上各种因素的影响，如平原、山体、建筑物等，使电磁波不再以单一的方式和路径传播而产生复杂的透射、绕射、散射、反射、折射等，如何建立一个准确的无线传播模型，是当下亟待解决的一个问题。

目前已有的无线传播模型可以性质分成：理论模型、经验模型与改进型经验模型。理论模型对实际的现场环境直接应用电磁理论计算，环境描述通常基于几何绕射理论（GTD）、物理光学（PO）的射线跟踪或其他精确方法。经验模型是根据大量测试结果

统计分析后导出的公式，具有代表性的经验模型有 Okumura-Hata 模型，COST231-Hata 模型等。半经验模型是在经验模型的基础上改进而成的，即在拟合公式中引入更多的参数从而可以为更细的分类场景提供计算模型，常见的有 Standard Propagation Model（SPM）模型，COST231 Walfisch - Ikegami 模型，校正之后的 Standard Macrocell 模型等。

为了使预测模型更加精确，在实际传播模型建模中需要利用大量额外实测数据、工程参数以及电子地图对传播模型进行校正，同时大量的用户数据也可以协助构建更为精确的模型。而随着大数据技术在语言、图像处理方面的发展，伴随着并行计算架构的进步，机器学习技术也具备了在线运算的能力，利用大量额外数据修正传播模型成为提升模型预测能力的未来发展趋势。本文的目的是站在设备供应商以及无线运营者的角度，通过合理的运用机器学习模型建立无线传播模型，并利用模型准确预测在新环境下无线信号覆盖强度，从而大大减少网络建设成本，提高网络建设效率。

### 1.3 问题重述

本文所要解决的主要问题是大量实测数据修正不同场景下的经验模型，通过调整传播模型的系数，使得模型预测值更加精确。实测的数据主要分成三类：第一类为工程参数，主要描述基站的相关信息；第二类为地理位置信息，主要记录基站所在地和测量地的地理信息；第三类为信号信息，即实际平均接收功率（Reference Signal Receiving Power,RSRP），是最终信号传播的效果。所有已知的参数信息如下：

表 1-1 工程参数数据及其字段含义

字段名称	含义	单位
<b>Cell Index</b>	小区唯一标识	-
<b>Cell X</b>	小区所属站点的栅格位置，X 坐标	-
<b>Cell Y</b>	小区所示站点的栅格位置，Y 坐标	-
<b>Height</b>	小区发射机相对地面的高度	m
<b>Azimuth</b>	小区发射机水平方向角	Deg
<b>Electrical Downtilt</b>	小区发射机垂直电下倾角	Deg
<b>Mechanical Downtilt</b>	小区发射机垂直机械下倾角	Deg
<b>Frequency Band</b>	小区发射机中心频率	MHz
<b>RS Power</b>	小区发射机发射功率	dBm

表 1-2 地图数据及其字段含义

字段名称	含义	单位
<b>Cell Building Height</b>	小区站点所在栅格 (Cell X,Cell Y) 的建筑物高度	m
<b>Cell Altitude</b>	小区站点所在栅格 (Cell X,Cell Y) 的海拔高度	m
<b>Cell Clutter Index</b>	小区站点所在栅格 (Cell X,Cell Y) 的地物类型索引	-
<b>X</b>	栅格位置, X 坐标	-
<b>Y</b>	栅格位置, Y 坐标	-
<b>Building Height</b>	栅格 (X,Y) 上的建筑物高度	m
<b>Altitude</b>	栅格 (X,Y) 上的海拔高度	m
<b>Clutter Index</b>	栅格 (X,Y) 上的地物索引类型	-

表 1-3 地物类型名称的编号含义

Clutter Index	含义	Clutter Index	含义
<b>1</b>	海洋	<b>11</b>	城区高层建筑 (40m~60m)
<b>2</b>	内陆湖泊	<b>12</b>	城区中高层建筑 (20m~40m)
<b>3</b>	湿地	<b>13</b>	城区<20m 高密度建筑群
<b>4</b>	城郊开阔区域	<b>14</b>	城区<20m 多层建筑
<b>5</b>	市区开阔区域	<b>15</b>	低密度工业建筑区域
<b>6</b>	道路开阔区域	<b>16</b>	高密度工业建筑区域
<b>7</b>	植被区	<b>17</b>	城郊
<b>8</b>	灌木植被	<b>18</b>	发达城郊区域
<b>9</b>	森林植被	<b>19</b>	农村
<b>10</b>	城区超高建筑 (>60m)	<b>20</b>	CBD 商务圈

表 1-4 RSRP 标签及其字段含义

字段名称	含义	单位
<b>RSRP</b>	栅格 (X,Y) 的平均信号接收功率	dBm

在工程参数、地理位置以及对应的 RSRP 已知的情形下, 可以通过训练一套合适的机器学习模型对不同场景下信道传播路径损耗进行准确预测。

问题一给出了明确的数据集, 限定在 Cost 231-Hata 模型的框架下设计合适的特征, 并阐述选择该特征的原因, 其给出的数据集信息如下:

工程参数信息								
Cell Index	Cell X	Cell Y	Height	Azimuth	Electric Downtilt	Mechanical Downtilt	Frequency Band	RS Power
2	100	100	49m	45°	2°	2°	1800MHz	18.2dBm
地图数据								
Cell Altitude	Cell Building Height	Cell Clutter Index	X	Y	Altitude	Building Height	Clutter Index	
47m	9m	11	500	500	9m	0m	1	
RSRP 标签数据								
RSRP								
-100dBm								

问题二承继了问题一的思想，并进一步放开模型限制且扩大数据集，要求选择合适的特征并对特征与目标的相关性进行排序。第二问与第一问的差别在于：第一问中的仅给出一条数据，同时传播模型限定为 **Cost 231-Hata** 模型，而在第二问中，数据集范围扩大至所有小区，同时在模型要求上涵盖所有机器学习模型。问题二最终要求阐明设计所选特征的原因及用于排序的量化数值的方法，最后形成特征名称及其余目标相关性的表格。

问题三则在问题一、二的基础上构建预测模型对 **RSRP** 进行预测。根据问题二中已有的特征集及问题所提供的训练数据集，建立基于 **AI** 的无线传播模型对不同地理位置的 **RPRP** 进行预测。预测的评判标准为在达到 **PCRR** 精度的基础上使 **RMSE** 最小。

## 2、问题假设

- 1) 只有直射信号能被检测到。即不考虑基站发射信号的折射等，但考虑反射信号。
- 2) 不考虑基站间的影响。数据中发现部分基站距离极近，我们假设相邻基站测量信号是相互正交的，不考虑它们之间的相互影响。
- 3) 一个小区只有一个发射天线。
- 4) 不存在天线增益效应。

### 3、符号说明

表 3-1 模型符号说明

符号名称	符号含义	单位
<b>PL</b>	传播路径损耗	dB
<b>f</b>	载波频率	MHz
<b><math>h_b</math></b>	基站天线有效高度	m
<b><math>\alpha</math></b>	用户天线高度纠正项	dB
<b>d</b>	链路距离	m
<b>Cm</b>	场景纠正常数	dB
<b><math>P_t</math></b>	小区发射机发射功率	dBm
<b>distance</b>	测量点与发射机的直线距离	m
<b><math>\Delta X</math></b>	测量点与发射机的横坐标相对值	m
<b><math>\Delta Y</math></b>	测量点与发射机的纵坐标相对值	m
<b><math>\Delta \text{Altitude}</math></b>	测量点与发射机的海拔高度相对值	m
<b><math>\Delta \text{Building Height}</math></b>	测量点与发射机的建筑物高度相对值	m
<b><math>\Delta \text{Height}</math></b>	栅格与信号线的相对高度	m
<b>Angle North</b>	栅格与发射机中心点构成的射线与北方向夹角	°
<b>Angle Line</b>	栅格与发射机中心点构成的射线与射线发射方向所成的两个角的较小值	°
<b>number points</b>	小区内栅格总数	-
<b>number buildings</b>	建筑物总数	-
<b>number centers</b>	发射机发射射线数量	-



## 4、问题分析

### 4.1 特征工程中的特征设计

问题一是构建预测模型的第一步，即根据已有信息设计合适的特征。该问题把对特征的考虑范围限制在 Cost 231-Hata 模型以及一条数据集的信息内，即通过这两条信息构建合适的特征。在问题一中，需要考虑 Cost 231-Hata 模型中影响平均信号接收功率的因素，同时也需要根据数据集信息设计适合的特征，不仅包括数据集本身的特征，还涵盖数据集衍生出的几何特征、统计特征等。

其中，Cost 231-Hata 模型定义了传播路径损耗与载波频率、基站天线有效高度、链路距离等特征的关系，其表达式为：

$$PL = 46.3 + 33.9\log_{10}f - 13.82\log_{10}h_b - \alpha + (44.9 - 6.55\log_{10}h_b)\log_{10}d + C_m$$

而平均信号接收功率 RSRP 与传播路径损耗 PL 之间反向关系为：

$$RSRP = P_t - PL$$

由此，在 Cost 231-Hata 模型下，平均信号接收功率 RSRP 的表达式为：

$$RSRP = P_t - 46.3 - 33.9\log_{10}f + 13.82\log_{10}h_b + \alpha - (44.9 - 6.55\log_{10}h_b)\log_{10}d - C_m$$

### 4.2 特征工程中的特征选择

继问题一初步的探讨如何设计特征，问题二放开了模型的限制，即可用机器学习模型进行训练，同时把数据集范围扩大至各小区。要求基于各小区数据集设计多个合适的特征，并且计算这些特征与目标的相关性最后排序。

在问题一中，我们已经找到了一个合适的特征，因此本问的关键在于其他有效特征的构造，具体的特征构造情况请见第五章 5.1 和 5.2 节。

本问题中主要采用的机器学习模型为 lightGBM（基于决策树算法的分布式梯度提升框架）。LightGBM 全称为 Light Gradient Boosting Machine，由微软开源，一个基于决策树，并且具有快速、分布式、高性能特点的梯度提升框架。它可被用于排行、分类以及其他许多机器学习任务中。lightGBM 有两大特点：采用 Histogram 算法及 leaf-wise 的生长策略。

LightGBM 采用 Histogram 算法，其思想是将连续的浮点特征离散成 k 个离散值，并构造宽度为 k 的 Histogram。然后遍历训练数据，统计每个离散值在直方图中的累计统计量。在进行特征选择时，只需要根据直方图的离散值，遍历寻找最优的分割点。

LightGBM 的 leaf-wise 的生长策略：它摒弃了现在大部分 GBDT 使用的按层生长（level-wise）的决策树生长策略，使用带有深度限制的按叶子生长（leaf-wise）的策略。level-wise 过一次数据可以同时分裂同一层的叶子，容易进行多线程优化，也好控制模型复杂度，不容易过拟合。但实际上 level-wise 是一种低效的算法，因为它不加区分的对待同一层的叶子，带来了许多没必要的开销，因为实际上很多叶子的分裂增益较低，没必要进行搜索和分裂。Leaf-wise 则是一种更为高效的策略，每次从当前所有叶子中，找到分裂增益最大的一个叶子，然后分裂，如此循环。因此同 Level-wise 相比，在分裂次数相同的情况下，Leaf-wise 可以降低更多的误差，得到更好的精度。Leaf-wise 的缺点是可能会长出比较深的决

策树，产生过拟合。因此 **LightGBM** 在 **Leaf-wise** 之上增加了一个最大深度的限制，在保证高效率的同时防止过拟合。

## 5、模型的建立与求解

### 5.1 设计特征工程中特征

根据问题一所提供的数据集信息，本文把设计的特征分成两大类：第一类为原始特征，即通过数据集获得的相关信息；第二类为设计特征，即根据已有的数据集信息衍生构造出的特征。

#### 5.1.1 原始特征

原始特征包括数据集提供的所有信息，由工程参数特征、地图特征构成。由于这些均为问题一数据集提供的基础信息，因此有必要作为特征考虑在内。

工程参数特征有 8 个，主要为发射机的位置及相关信息，包括：小区所属站点的栅格位置 X 坐标、小区所属站点的栅格位置 Y 坐标、小区发射机相对地面的高度、小区发射机水平方向角、小区发射机垂直电下倾角、小区发射机垂直机械下倾角、小区发射机中心频率以及发射功率。

地图特征有 8 个，主要为测量地和发射机所属栅格的地形特征，包括：栅格位置 X 坐标、栅格位置 Y 坐标、测量地所在栅格的建筑物高度、测量地所在栅格的海拔高度、测量地所在栅格的地物类型索引、发射机所在栅格的建筑物高度、发射机所在栅格的海拔高度、发射机所在栅格的地物类型索引。

#### 5.1.2 设计特征

设计特征是在原始数据集的基础上，考虑经验模型、几何位置及统计特性等因素构造出来的特征。设计特征一共有 32 个，根据数据集描述的对象不同，可以分成工程参数特征、相对特征以及小区栅格分布情况特征。

工程参数特征为：下倾角，即机械下倾角与电下倾角之和。机械下倾角是一种物理信号下倾，可以通过调整天线面板后面的支架实现，电下倾角通过调整天线内部的线圈实现，是一种电信号下倾，实际的信号线下倾是两者之和。由于实际信号线下倾角决定信号传播的方向角度，我们认为它信号传播模型至关重要。

相对特征考虑测量点与发射机的相对长度和角度。显而易见，发射地与测量地的相对距离和角度都能影响信号的传播效果，因此也是本文选择的重要特征之一。长度特征中包括：测量点与发射机的直线距离(distance)、测量点与发射机的横坐标相对值( $\Delta X$ )、测量点与发射机的纵坐标相对值( $\Delta Y$ )、测量点与发射机的海拔高度相对值( $\Delta \text{Altitude}$ )、测量点与发射机的建筑物高度相对值( $\Delta \text{Building Height}$ )、栅格与信号线的相对高度( $\Delta \text{Height}$ ，几何特征)。而角度特征包括：栅格与发射机中心点构成的射线与北方向夹角(Angle North，范围 0-360)、栅格与发射机中心点构成的射线与射线发射方向所成的两个角的较小值(Angle Line，范围 0-180)。其中两个角度特征的详细示意图如下：

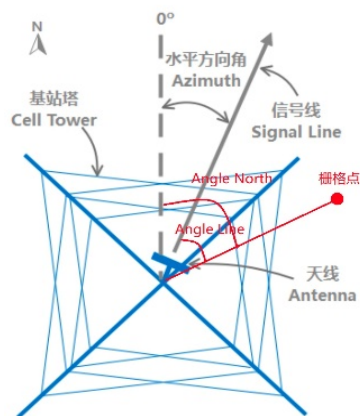


图 1 自定义角度变量示意图

小区栅格分布情况特征主要考虑与测量地相关特征的计数特征与统计特征。由于测量地是信号接收地，测量地的地形特点以及与发射地的相关信息都是影响最终信号接收效果的重要指标，故而把小区栅格分布情况特征考虑在内。根据这些指标的特点，分为计数特征和统计特征。其中，计数特征包括小区内栅格总数(number points)、建筑物总数(number buildings)、发射机发射射线数量(number centers)；

统计特征包括：小区内所有栅格相应特征(即海拔高度、建筑高度、与发射机距离、与发射机中心点构成的射线与北方向夹角、与发射机中心点构成的射线与发射方向夹角)的平均值、标准差、最大值、最小值。

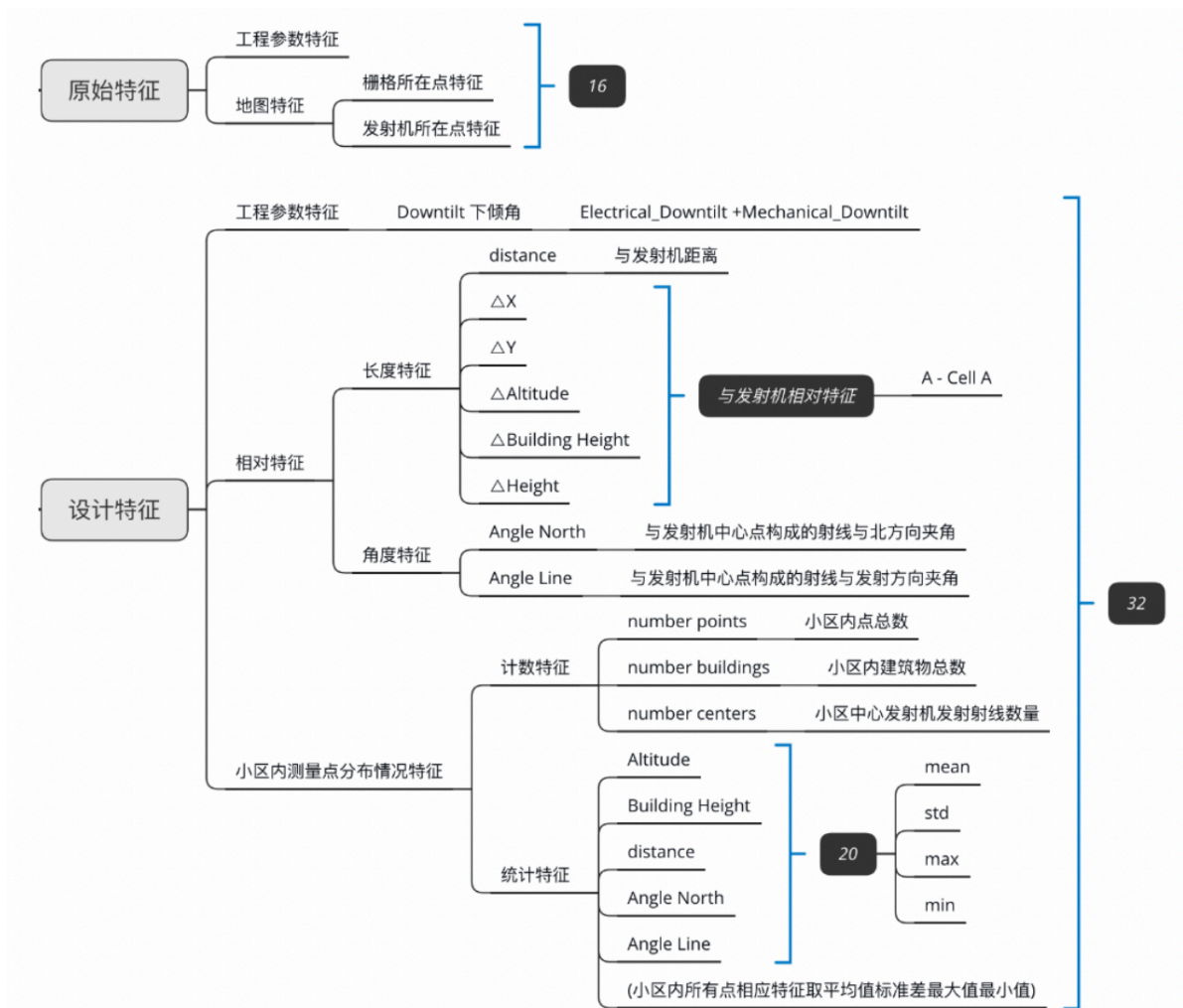


图 2 特征工程

## 5.2 特征工程中的特征选择

将所有小区数据合并，并计算出全部原始特征与设计特征，将数据集按 4:1 切分为 3200 个小区和 800 个小区，将其作为训练集和测试集，运用微软开源的 lightgbm 算法进行计算(基于梯度提升树的算法)，损失函数为 RMSE，测试集 RMSE 在 1000 轮左右后达到 8.87，选取 1113 轮结果，树模型 48 个特征重要性表格如下表。其中重要性衡量指标为 gain，即特征在所有模型中运用的次数，gain 越大则意味着变量与 RSRP 的联系越紧密，由于 gain 是决策树模型中的概念因此它也同时衡量了变量与 RSRP 之间非线性关系的强度。而决策树在每个树节点使用信息熵(entropy)来决定对哪个变量进行分割。

表 5-1 基于树模型 gain 的变量重要性排序

特征名称	gain	特征名称	gain
BuildingHeight_min	0	Downtilt	260
AngleNorth_min	25	RS_Power	301
Frequency_Band	35	Mechanical_Downtilt	312
Electrical_Downtilt	133	center_num	320
delta_Building_Height	133	Altitude_min	323
Building_Height	189	distance_min	342
Cell_Building_Height	240	Cell_Altitude	360
Clutter_Index	385	Altitude	754
Altitude_max	426	num_point	758
Altitude_mean	426	distance_mean	797
Cell_Y	478	num_building	839
Azimuth	533	AngleNorth_std	852
Angleline_max	536	Angleline_std	884
Cell_X	545	delta_Height	911
delta_Altitude	593	Angleline_mean	937
Angleline_min	594	AngleNorth_mean	1018
BuildingHeight_std	639	Cell_Clutter_Index	1073
BuildingHeight_mean	640	distance	1370
distance_std	644	delta_X	1554
BuildingHeight_max	662	Angle_North	1626
Height	688	delta_Y	1639
AngleNorth_max	715	Angle_line	1680
distance_max	718	Y	1855
Altitude_std	748	X	1900

筛掉全特征中 lightgbm 重要性较低的一些特征，以及栅栏坐标、中心点坐标重新进行 lightgbm 计算，损失函数为 RMSE，测试集 RMSE 在 600 轮左右后达到 9.01，选取 600 轮结果，树模型 36 个特征重要性表格如下：

表 5-2 筛选后的变量重要性排序

特征符号	gain	特征符号	gain
Frequency_Band	31	BuildingHeight_max	498
Building_Height	133	AngleNorth_max	527
Clutter_Index	182	distance_max	532
RS_Power	221	Height	535
center_num	237	num_building	536
Cell_Building_Height	262	distance_mean	545
Downtilt	289	Angleline_std	548
delta_Altitude	296	Altitude_std	557
Azimuth	330	num_point	563
Angleline_max	353	Angleline_mean	662
Cell_Altitude	382	AngleNorth_std	673
Angleline_min	436	AngleNorth_mean	690
distance_std	467	Cell_Clutter_Index	749
Altitude	471	distance	808
delta_Height	476	Angle_North	826
BuildingHeight_std	480	delta_Y	887
BuildingHeight_mean	486	Angle_line	916
Altitude_mean	493	delta_X	923

由上述结果可见，比较重要的特征为：相对位置、栅格与发射机中心点构成的射线与北方向夹角、栅格与发射机中心点构成的射线与射线发射方向所成的两个角的较小值、与发射机距离、栅格地物类型。与此同时，小区栅格分布情况的计数特征、统计特征都对预测结果准确性有一定提升。

## 5.3 预测 RSRP 值

针对本题，我们利用了 LightGBM 框架和 tensorflow.keras 框架分布搭建了梯度提升树模型和神经网络模型。这两类型非线性预测模型具有十分强大的适应能力，能够自动学习复杂逻辑，但也存在过拟合的风险，过拟合即模型在训练集表现良好而在测试集表现极差的现象。为了防止过拟合，我们在训练时尽可能减少神经网络层数与树模型复杂度，并将原始训练数据进一步按 8:2 拆分成训练集与交叉验证集，通过控制模型在两个数据集上的损失函数的发散度来控制过拟合。

### 5.3.1 数据预处理

将所有小区数据合并，一共 12011833 条数据，并计算出全部原始特征与设计特征，将数据集按 4:1 切分为 3200 个小区和 800 个小区，将其作为训练集和测试集，其中训练集一共 9849152 条数据，将其中 RSRP 偏离均值三倍标准差的异常点出去后，训练集一共

9828166 条数据，训练集一共 2162681 条数据，不做处理。

### 5.3.2 LightGBM 树模型的建立及预测结果

对于 LightGBM 梯度提升树模型，将所有特征全部放入模型，其中包括 2 个类别特征，46 个连续特征，共 48 个特征，其中原始特征为 16 个，新设计特征 32 个。运用谷歌开源的 lightgbm 算法进行计算，损失函数为 RMSE，训练轮数为 5000，采取早停策略，当再进行 100 轮梯度提升算法后，测试集 RMSE 不再降低时，训练停止。测试集 RMSE 在 1000 轮左右后达到 8.87，其中 1113 轮结果 RMSE 最低，测试集 RMSE 为 8.875，同时测试集 RMSE 与训练集 RMSE 之间的差距既不明显发散也不明显减小，这说明模型训练到位且不存在过拟合现象。

### 5.3.3 tensorflow.keras 深度学习模型简介及预测结果

- 选择特征

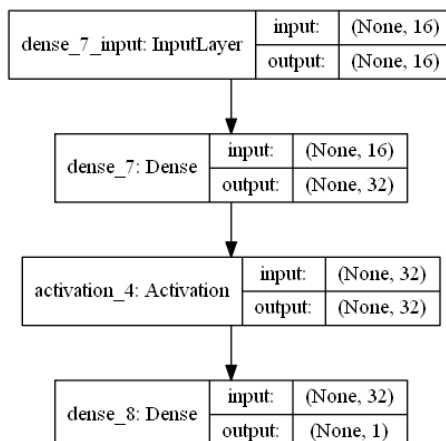
由于神经网络模型本身没有变量筛选功能，为了防止引入过多噪声，我们在训练神经网络模型前经过 lgb 特征重要性、经验主义筛选，在所有特征中选取如下 16 个特征进行神经网络的训练，传入神经网络前进行标准化预处理，所采用的特征见下表。

表 5-3 神经网络模型所采用变量汇总

特征名称	特征符号
小区发射机中心频率	Frequency_Band
栅格上建筑物高度	Building_Height
小区发射机功率	RS_Power
小区站点所在栅格的建筑物高度	Cell_Building_Height
发射机下倾角	Downtilt
小区发射机水平方向角	Azimuth
小区所在栅格海拔高度	Cell_Altitude
栅格海拔高度	Altitude
小区发射机相对地面高度	Height
栅格与发射机的 2D 距离(由两者 X、Y 计算)	distance
栅格相对水平方向所成夹角	Angle_North
Cell Y-Y	delta_Y
栅格相对信号线发射方向所成夹角	Angle_line
Cell X-X	delta_X
栅格与信号线的相对高度	delta_Height
栅格与发射机的 3D 距离(由两者 X、Y 和海拔计算)	dis_3d



- 训练方式及结果



- 训练方式

损失函数为均方误差(下称 mse), 优化器为批随机梯度下降(下称 batch-SGD), 每批传入 163842(2^14)个样本点进行梯度下降算法, 训练轮数为 10 轮。

- 结果

由于 lgb 框架下的梯度提升树在 tensorflow 上实现起来较困难, 因此在 ModelArts 上我们只上传神经网络模型。

Lgb 模型的测试集的 MSE 在第 5 轮训练就已经收敛并达到最低的 87.2, RMSE 为 9.338。神经网络模型的测试集 MSE 在训练的第 5 轮达到 87.24, RMSE 为 9.34, 两者差距不大。

## 6、模型的优缺点及其改进

经过多次交叉验证与超参数寻优，我们的模型给出了一个相对合理的预测结果，且模型预测时具有一定稳健性，即不随着测试集的更换而出现预测效果剧烈波动的情况。这与样本量大、模型未过拟合以及参数最优化都有关系，这也正是本模型的有点所在。

但也应该看到模型的不足：下图为某一小区的真实 RSRP(左)与 lgb 模型预测 RSRP(右)，从预测效果来看，模型对弱 RSRP 区域（颜色深的区域）的预测能力较低，我们认为只是特征工程没有做到的位的结果，改进方向应该考虑添建有针对性的特征变量(比如，被测栅格与发射基站中间的遮挡物数量等)。

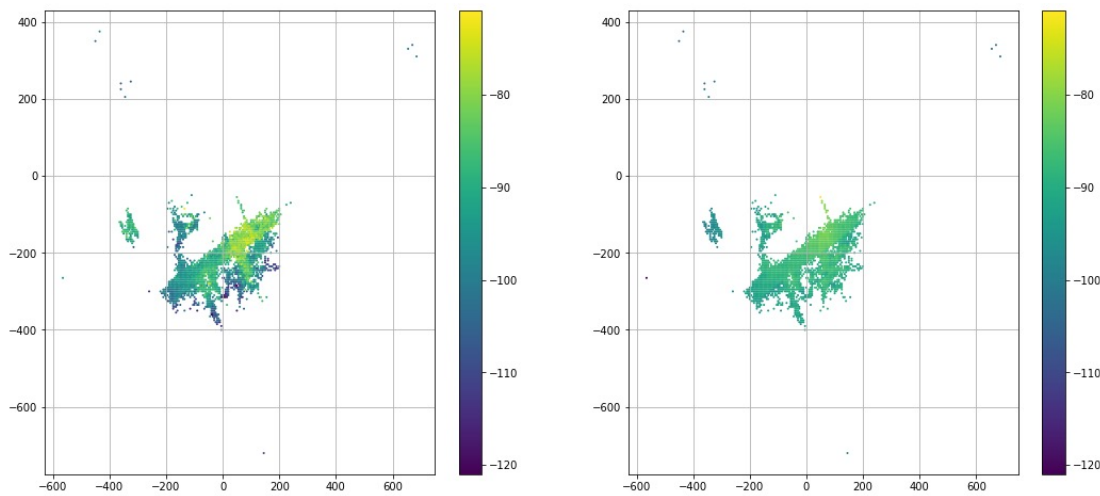


图 4 某小区真实 RSRP(左)与预测 RSRP(右)对比图

## 参考文献

- [1] 赵元福. 对移动通信的应用及其发展的分析[J]. 无线互联科技, 2012, (9): 46-46.
- [2] 江湧. 移动心痛的分类和发展史[J]. 无线电技术与信息, 2004, (3):41-43.
- [3] 徐伟. 基于 LTE 系统无线传播模型校正的研究及应用[D]. 2016.
- [4] 梁金. 浅谈 3G 技术[J]. 黑龙江科技信息, 2010, (27):59-59.
- [5] 崔春风, 王晓云. 3GPP LTE——3G 向 4G 演进的关键一步[J]. 电信网技术, 2005(9):19-22.
- [6] 翟冠楠, 李昭勇. 5G 无线通信技术概念及相关应用[J]. 电信网技术, 2013(9):1-6.
- [7] Li J, Tian M, He Y, et al. Effect Evaluation System of Massage Chair Based on EEG[C]//2019 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, 2019: 1565-1569.
- [8] Zhang J, Fogelman-Soulié F, Largeron C. Towards Automatic Complex Feature Engineering[C]//International Conference on Web Information Systems Engineering. Springer, Cham, 2018: 312-322.
- [9] Calin D, Kaya A Ö, Abouliatim A, et al. On the feasibility of outdoor-to-indoor LTE small cell deployments: Field trial experiments and performance prediction[C]//2013 IEEE Globecom Workshops (GC Wkshps). IEEE, 2013: 714-719.
- [10] Weitzen J, Wakim R, Webster E. Comparing RSRP, CQI, and SINR measurements with predictions for coordinated and uncoordinated LTE small cell networks[C]//2015 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS). IEEE, 2015: 1-5.
- [11] 溜哒兔 , LightGBM的算法介绍: LightGBM算法的特别之处,  
<https://www.cnblogs.com/ldt-/p/10206356.html>, 2019 年 9 月 22 日