



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	同济大学
参赛队号	19102470040
1.	朱宇斌
2.	高 盼
队员姓名 3.	刘甜甜

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：

随着 5G NR 技术在全球范围内的应用，运营商在部署 5G 网络时如何合理选择覆盖区域内的基站站址，对于满足用户的通信需求十分必要。在整个无线网络规划流程中，高效的网络估算对于精确的 5G 网络部署有着非常重要的意义。无线 LTE 网络已在全球普及，全球几十亿用户，每时每刻都会产生大量数据。如何合理地运用这些数据来辅助无线网络建设就成为了一个重要的课题。

本文以华为提供的真实环境中采集的多个小区历史工程参数数据、地图格式数据和平均信号接收功率（RSRP）标签数据为基础，结合无线传播经验模型 Cost 231-Hata，3GPP Uma NLOS 模型，从特征转换、几何位置计算、经验传播模型影响因子组合提取等方式，进行特征设计。通过对所设计特征的多个角度筛选，获得合理的特征子集。我们考虑机器学习模型与深度模型的结合对该问题进行建模求解，对无线传播历史数据训练以获得可泛化的预测模型。通过华为 Model Arts 平台的测试，更新调整机器模型的参数，提高其泛化性能。

针对问题一，查看历史数据的分布情况，首先进行数据清洗，结合通信理论专家知识，剔除异常数据。在特征设计部分，我们根据数据的统计分析结果，对原始特征进行特征转换，减少冗余特征。考虑物理角度与几何位置，计算潜在的物理特征，进行特征提取。利用无线传播经验模型 Cost 231-Hata 和 3GPP Uma NLOS 模型的影响因子，设计新的特征属性。挖掘出可以使用的特征。

针对问题二，从三个角度对设计的特征进行特征筛选。首先剔除取值变化较小的特征。然后应用 Pearson 相关系数和 Spearman 秩相关数对特征与预测目标的相关性进行计算。为了弥补单一相关系数对于特征相关性计算的片面性，最后引入机器学习模型，使用 Lasso 回归、Ridge 回归以及决策树回归模型进行特征重要性排序，以获得输入最终预测模型的有效特征。

针对问题三，我们采用了简单线性模型、决策树模型、神经网络模型预测 RSRP 值。首先，我们分析了数据集的结构特点并进行了合理的分割使其满足模型训练要求。然后我们分别运用了简单线性模型、决策树模型和神经网络模型进行训练，并做了相应的调参处理。最后我们分析了各个模型的综合表现能力，并提出了对该课题进一步的研究思路。

关键词：5G、无线传播模型、机器学习、神经网络

目 录

目 录	2
1 问题重述	3
1.1 问题背景	3
1.2 问题描述	3
2 合理假设及符号系统	4
2.1 问题假设	4
2.2 项目符号	4
3 问题一的建模与求解	6
3.1 问题分析	6
3.2 模型建立	6
3.2.1 异常值处理	6
3.2.2 基于经典传统经验信道模型 Cost 231-Hata 的特征设计	9
3.2.3 基于角度的特征设计	11
3.2.4 基于几何位置计算的特征设计	12
3.2.5 基于 3GPP Uma NLOS 模型的特征设计	14
3.3 特征设计结果	14
4 问题二的建模与求解	16
4.1 问题分析	16
4.2 模型建立	16
4.2.1 去除取值变化小的特征	16
4.2.2 单变量特征选择	16
4.2.3 基于模型的特征排序和正则化	18
4.3 特征选择结果	20
5 问题三的建模与求解	22
5.1 问题分析	22
5.1.1 特征集分析	22
5.1.2 评价指标	22
5.1.3 求解目标	23
5.2 模型建立	23
5.2.1 数据集处理	23
5.2.2 模型选择	24
5.3 模型解算及结果分析	25
5.3.1 本地预测结果分析	25
5.3.2 平台测试结果展示	26
5.3.3 结论与展望	27
参考文献	28

1 问题重述

1.1 问题背景

随着 5G NR 技术的发展，5G 在全球范围内的应用也在不断地扩大。运营商在部署 5G 网络的过程中，需要合理地选择覆盖区域内的基站站址，进而通过部署基站来满足用户的通信需求。在整个无线网络规划流程中，高效的网络估算对于精确的 5G 网络部署有着非常重要的意义。无线传播模型正是通过对目标通信覆盖区域内的无线电波传播特性进行预测，使得小区覆盖范围、小区间网络干扰以及通信速率等指标的估算成为可能。由于无线电波传播环境复杂，会受到传播路径上各种因素的影响，如平原、山体、建筑物、湖泊、海洋、森林、大气、地球自身曲率等，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，所以建立一个准确的模型是一项非常艰巨的任务。

现有的无线传播模型可以按照研究方法进行区分，一般分为：经验模型、理论模型和改进型经验模型。经验模型的获得是从经验数据中获取固定的拟合公式，典型的模型有 Cost 231-Hata、Okumura 等。理论模型是根据电磁波传播理论，考虑电磁波在空间中的反射、绕射、折射等来进行损耗计算，比较有代表性的是 Volcano 模型。改进型经验模型是通过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，典型的有 Standard Propagation Model (SPM)。

在实际传播模型建模中，为了获得符合目标地区实际环境的传播模型，需要收集大量额外的实测数据、工程参数以及电子地图用来对传播模型进行校正。此外无线 LTE 网络已在全球普及，全球几十亿用户，每时每刻都会产生大量数据。如何合理地运用这些数据来辅助无线网络建设就成为了一个重要的课题。

1.2 问题描述

近年来，大数据驱动的 AI 机器学习技术获得了长足的进步，并且在语言、图像处理领域获得了非常成功的运用。伴随着并行计算架构的发展，机器学习技术也具备了在线运算的能力，其高实时性以及低复杂度使得其与无线通信的紧密结合成为了可能。

如何运用已有的数据，建立 AI 模型，学习数据中的隐含关系，不仅能够高效解决现实世界中的实际问题，而且节省人力和财力资源，对于辅助现实世界中的工程问题有极大的帮助。

基于华为提供的多个小区的工程参数数据、地图数据和 RSRP 标签数据，通过运用机器学习模型在历史数据上建立无线传播模型，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。

问题一：根据 Cost 231-Hata 模型以及小区的工程参数数据、地图数据和 RSRP 标签数据集信息设计合适的特征，并阐述原因。

问题二：基于提供的各小区数据集，设计多个合适的特征，计算这些特征与目标的相关性，并将结果量化、排序，形成表格，并阐明设计这些特征的原因和用于排序的量化数值的计算方法。

问题三：通过建立预测模型来进行 RSRP 的预测。根据建立的特征集以及赛题提供的训练数据集，建立基于 AI 的无线传播模型来对不同地理位置的 RSRP 进行预测。

2 合理假设及符号系统

2.1 问题假设

1. 小区内任意测试点测出的 RSRP 结果不受临近小区基站信号干扰，多个小区的 RSRP 既不会标量相加也不会矢量叠加；
2. 水平方向角具体指的是正北方向与发射机发射方向形成的夹角，垂直下倾角具体指的是水平线与信号线射线形成的锐角。
3. 测量 RSRP 时用户均靠近本单元格建筑物，即将处于建筑物内部和外部视为同一种情况；
4. 参照 SPM 模型，默认用户天线高度为 1.5m；
5. 根据理想情况下单天线波瓣图可以认为（图 2-1），信号线是信号最强的方向；

Single Antenna (N=1): Space Response

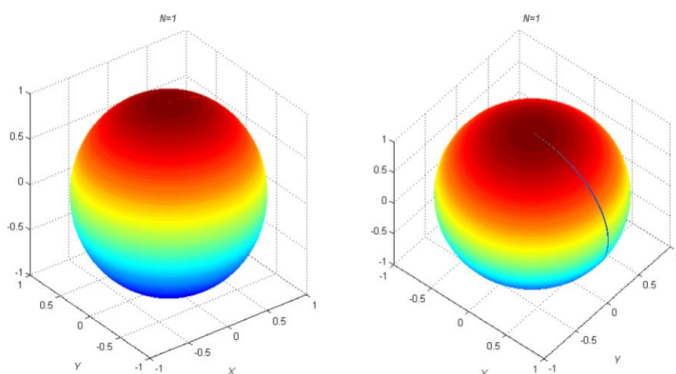


图 2-1 理想情况下单天线波瓣图（右图为旋转 90° 的情况）

6. 每个小区发射频率不固定，相同频率的信号在传播时不互相干扰。

2.2 项目符号

表 2-1 项目符号及其意义

项目 符号	意义
L_b	传播路径损耗 (dB)
f	载波频率 (MHz)
h_b	发射机有效高度 (m)
d	链路距离 (m)
α	用户天线高度纠正项 (dB)
C_m	场景纠正常数 (dB)
h_R	用户天线有效高度 (m)
θ_{md}	发射机垂直电下倾角 (degree)
θ_{ed}	发射机垂直机械下倾角 (degree)
θ_d	发射机下倾角 (degree)
θ_a	发射机水平方向角 (degree)
θ_h	测量点水平方向角 (degree)
$\Delta\theta$	发射机至测量点水平方向差角 (degree)
d_d	测量点至信号线直接距离 (m)
d_e	测量点至信号线有效距离 (m)
d_p	信号线传播至测量点传播距离 (m)
Δh_v	信号线与测量点垂直落差 (m)

Δh_d	信号线与测量点直接落差 (m)
Δh_a	发射塔与测量点海拔高度差 (m)

3 问题一的建模与求解

3.1 问题分析

无线信号在自由空间传播，其衰落主要是由于山体、树木、房屋等建筑物的遮挡造成了信号传播的损耗，尤其是 5G 网络，信号遮挡对其影响作用更大。

目前，无线信号传播模型可以按照研究方法大体分为两类：一类是基于无线电传播理论的理论分析方法，根据电磁波的传播理论，由于电磁波在空间中传播会出现反射、折射、绕射等产生损耗，因此无线信号的损失可以据此来计算，例如 Volcano 模型、WaveSight 模型、WinProp 模型等；另一类是建立在大量测试数据和经验公式基础上的实测统计方法，例如 OkumuraHata 模型、COST231-Hata 模型、Keenan-Motley 模型、SPM 模型等。

华为提供了真实环境中多个小区的工程参数数据、地图数据以及 RSRP 标签。该问题可以看做有标签的机器学习问题，依据这些历史数据，提取关键特征，利用机器学习模型对平均信号接受率（RSRP）进行预测是本次比赛的核心内容。这一题的关键是利用已有的历史数据，结合无线信号传播模型，挖掘对于预测平均信号接收率（RSRP）有贡献的特征，并进行相应的分析与解释。3.2 节将重点介绍如何建立模型和选取特征。

3.2 模型建立

无线信号传播模型在建立时，需要根据不同的传播场景建立对应的传播经验模型，并通过收集大量的工程参数与实际信号接收功率之间的差异对经验模型公式进行修正。已知城市中经典的传统经验信道模型 Cost 231-Hata 为：

$$L_b = 46.3 + 33.9 \log_{10} f - 13.82 \log_{10} h_b - \alpha + (44.9 - 6.55 \log_{10} h_b) \log_{10} d + C_m \quad (3.1)$$

其中 PL 表示传播路径损耗（dB）， f 为载波频率（MHz）， h_b 为基站天线有效高度（m）， d 为链路距离（m）， α 为用户天线高度纠正项（dB）， C_m 为场景纠正常数（dB）。

在对数据进行特征设计前，首先要进行数据清洗，剔除异常数据，减少噪声的影响。

3.2.1 异常值处理

受实际测量误差和数据质量影响，我们应当在设计特征前合理地清洗离群数据、异常数据、缺失属性数据，图 3-1 表示了数据可视化的结果。

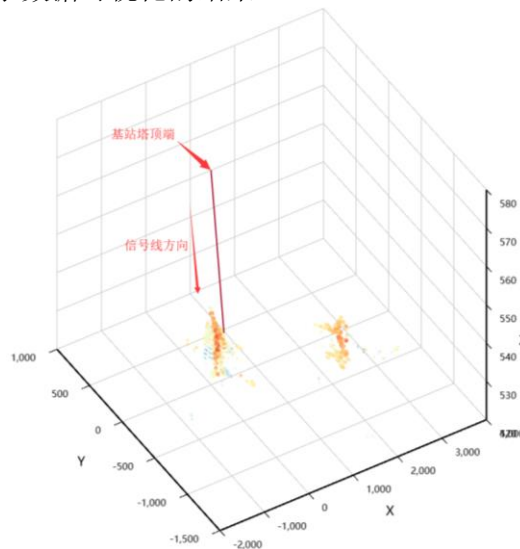


图 3-1 数据可视化结果（train_108401.csv）

1. 剔除距离异常数据

有效的数据应该分布在发射机周围，不应过远或过近。由于天线采用定向天线，只有在主波束方向上的测量点才是有效的，因此，异常数据应该包括：

- （1）距离过远：通常将与基站之间的距离为 2 倍站间距的数据认为是距离过远的异常数

据，需要过滤掉。在本次数据处理中，过滤掉测量点与基站距离大于 800 m 的数据。

(2) 距离过近：通常将与基站之间的距离过近的数据也认为是需要过滤掉的异常数据，可选择设定在 1/10 ~ 1/30 站间距之间，在本次比赛中，过滤掉与基站距离小于 10 m 的数据。

(3) 天线背侧：现有移动通信网络的发射机多使用定向天线，我们假设本次比赛提供的数据中，基站的天线也是定向天线。因此某一数据只有在当前的当前施主小区对应天线的主波束方向上的测量点才是有效的，所以需将非主波束平面方向的数据过滤掉。

图 3-2 表示数据集中存在离群数据的情况。

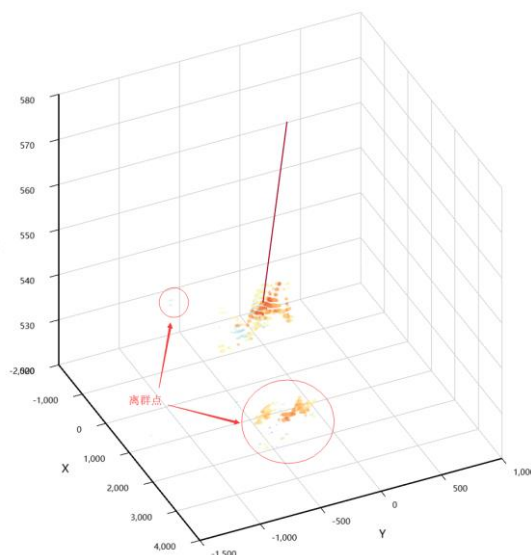


图 3-2 数据集中存在离群数据的情况 (train_108401.csv)

2. 剔除建筑物高度与地物索引不符的异常数据

栅格 (X, Y) 上建筑物的高度分布如图 3-3 所示。从图中可以看出，大量栅格没有建筑物存在即 Building Height 为 0。直接对 Building Height 大于 0 的测量数据进行分布统计分析，无法看到明显的特征。

建筑物的高度应该与观测点的地物类型相关，图 3-4 表示了对每个地物类型下的建筑物高度进行统计的结果。不难观察到在某些标定地物类型的区域内出现了与该地物类型不符的建筑，对于这些异常数据应该进行剔除。

对建筑物高度与地物类型不符的异常数据进行处理后，第 10-14 号地物类型下建筑物高度分布情况如图 3-5 所示。

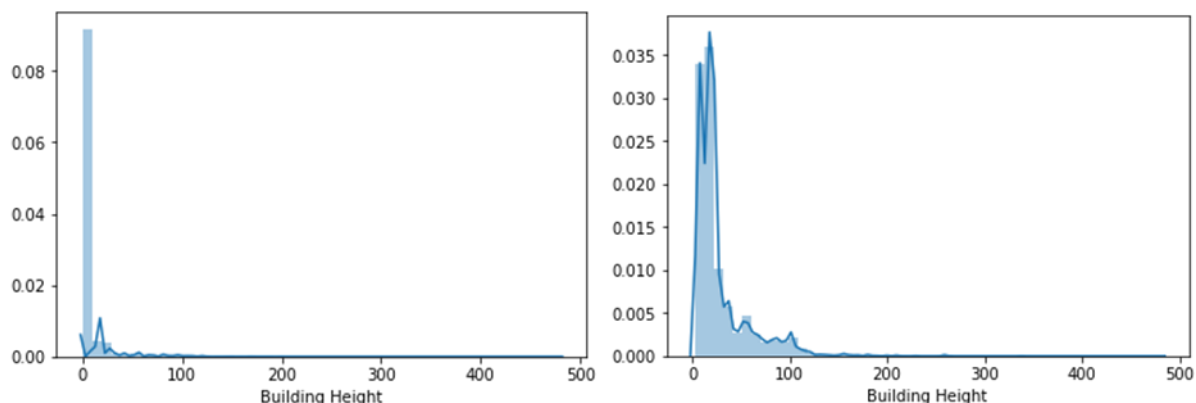


图 3-3 建筑物高度分布

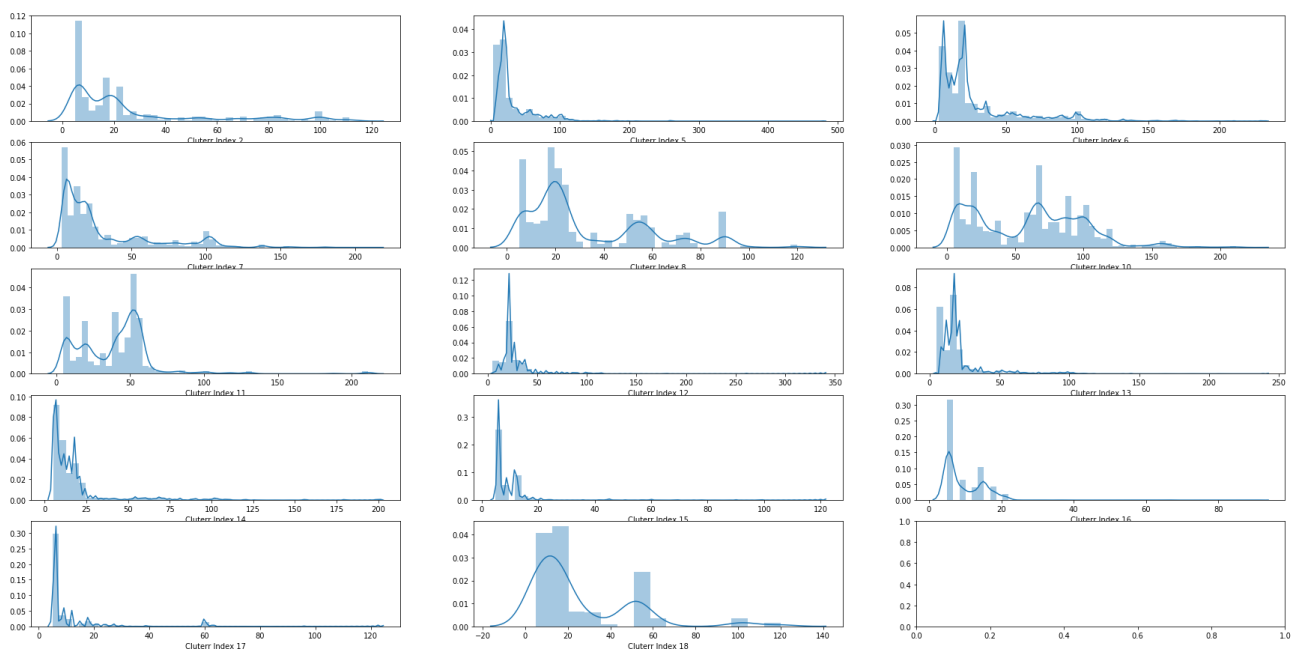


图 3-4 不同地物类型下建筑物高度分布

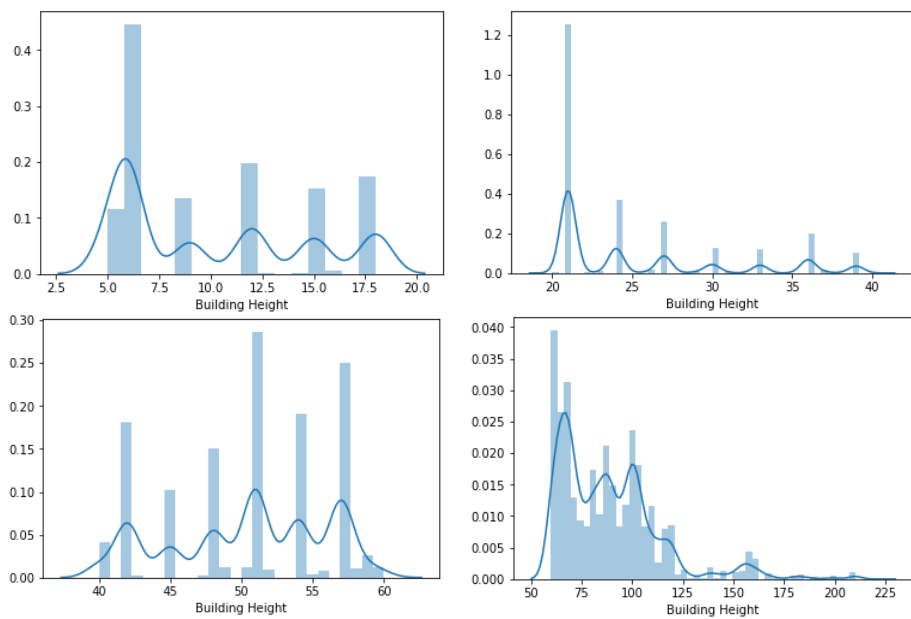


图 3-5 剔除异常值后地物类型与 Building Height 分布

3. 剔除基站高度缺失的异常数据

图 3-6 简单地表示了数据集中存在基站高度缺失地情况。

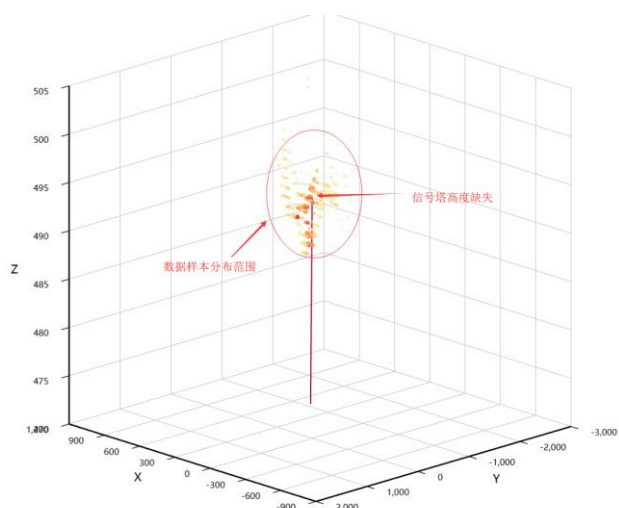


图 3-6 数据集中存在基站高度缺失的情况 (train_133001.csv)

4. 剔除与实际情况不符的异常数据

根据信号传播的规律不难得出，RSRP 在距信号线近的地方强度高，在距信号线远的地方强度低。图 3-7 中记录了数据集中存在与这一规律不符数据的情况。

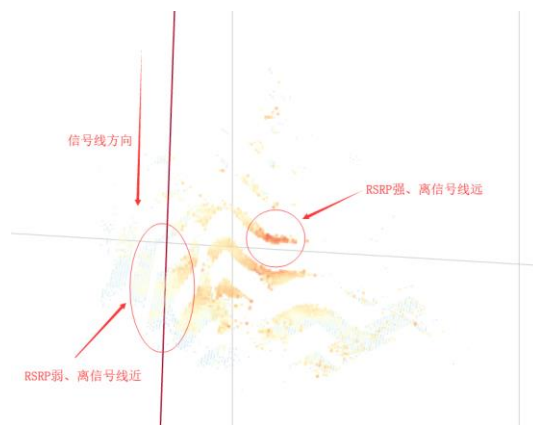


图 3-7 数据集中存在与实际情况不符数据的情况 (train_2068601.csv)

3.2.2 基于经典传统经验信道模型 Cost 231-Hata 的特征设计

1. 基于小区发射机中心频率 f 的特征

根据城市经典模型 Cost 231-Hata，对于路径传播损耗影响因子 f （载波频率），数据统计结果如表 3-1 所示。

表 3-1 小区发射机中心频率

频率 (MHz)	2585	2624.6	2604.8
数据量	3,172,693	65,810	24,765

表 3-2 基于中心频率的特征构建

频率值 (MHz)	特征列	值	解释说明
2585	F1	0 或 1	0 表示非此频率 1 表示为此频率
2624.6	F2	0 或 1	
2604.8	F3	0 或 1	

对于小区中心发射机的中心频率，数据中出现三个值，每一个值对应的数据量大于 10,000

条，因此基于发射机频率建立新的类别特征，结果如表 3-2 所示。

2. 基于小区发射机有效高度 h_b 的特征

基站天线的有效高度是影响传播路径损耗的一个因子，对其数据分布进行统计，结果如表 3-3 所示。

表 3-3 小区发射机相对地面高度统计

特征	均值	方差	25%	50%	75%	0 值数量
Height	22.8578	9.5797	17	24	29	103,585

数据中包含一万多条 Height 值为 0 的数据，考虑到基站高度对最终信号损失的影响，我们使用数据参数 Height 的均值替换 0 值。

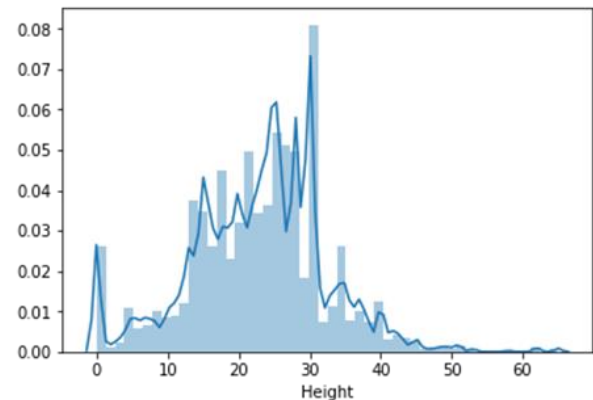


图 3-8 基站高度分布

3. 基于链路距离 d 的特征

原始数据提供测量点以及小区基站在栅格中的坐标，根据坐标计算测量点与基站在欧式空间的距离。测量点与基站的欧式空间距离分布如图 3-9 所示，其分别展示了所有测量点与基站的距离分布，筛选距离小于 5000 m 与 3000 m 测量点后的距离分布，大部分数据都是在小于 3000 m 范围内的.考虑信号对于空间距离的有效性，选择距离 10-3000m 范围内的测量点作为有效数据。

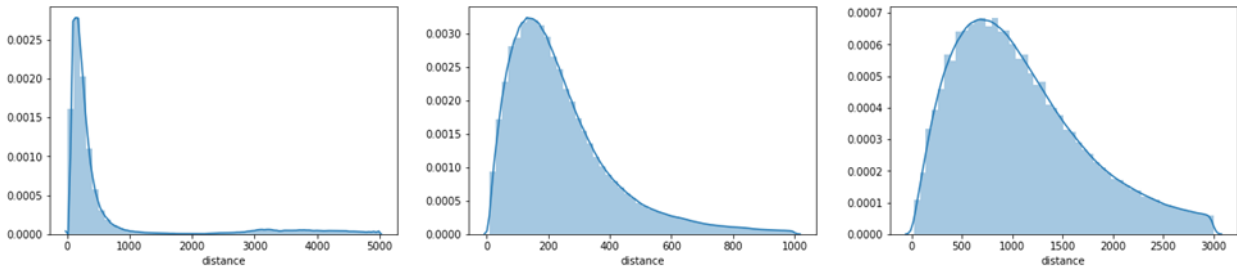


图 3-9 测量点与基站距离分布（左图：<5000m；中图：<1000m；右图：<3000m）

4. 基于 Cost 231-Hata 模型影响因子组合特征

根据城市经典模型 Cost 231-Hata 公式，对其中的各个影响因子项进行组合，并计算与目标“RSRP”的相关性，组合方式如表 3-4 所示，其相关性矩阵如图 3-10 所示。

表 3-4 特征组合方式

特征列	组合方式	相关性
H_D	Height * Distance	-0.26
F_sub_H	Frequency Band – Height	0.02
F_add_D	Frequency Band + Distance	0.02
F_sub_hd	Frequency Band – Height * Distance	0.26

F_sub_H_add_D	Frequency Band – Height + Distance	-0.32
F_sub_H_sub_HD	Frequency Band – Height – Height * Distance	0.22
F_add_D_sub_HD	Frequency Band + Distance – Height * Distance	0.10
F_sub_H_add_D_sub_HD	Frequency Band – Height + Distance – Height * Distance	0.24

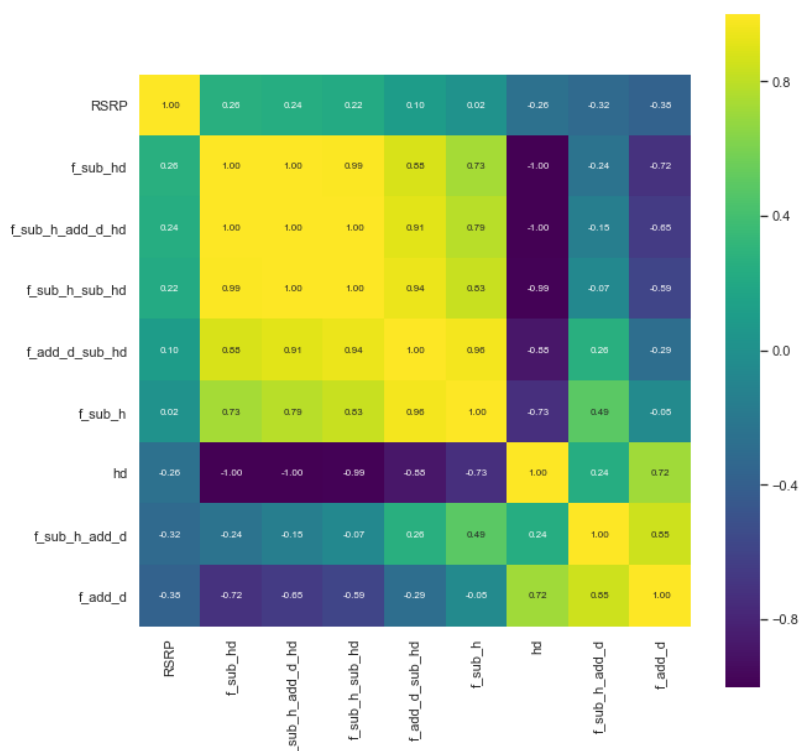


图 3-10 组合特征与目标的相关性矩阵

3.2.3 基于角度的特征设计

机械下倾角 θ_{md} 是天线面板后面的支架控制的，若该特征对于信号的传播损耗有影响，可以通过人工调整角度的方式帮助减少信号传播的损耗。为了直观理解该特征，我们对该数据进行统计分析，结果如表 3-5 所示，其频数分布图如图 3-11 所示。

表 3-5 机械下倾角频数统计

序号	机械下倾角	数据量
1	3	3547306
2	0	1474321
3	4	1393921
4	5	1138256
5	2	890579
6	6	600736
7	1	291082
8	8	279904
9	10	267628
10	7	211828
11	9	105965
12	12	37327
13	11	19966
14	13	15620
15	26	4855
16	15	4082
17	16	3612

18	-1	2688
19	-2	2680
20	14	2585

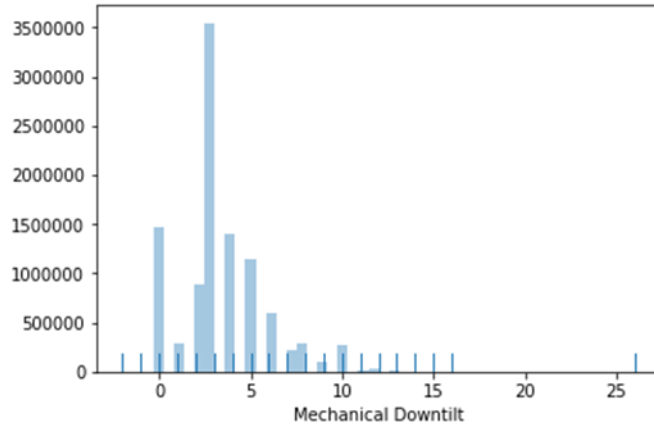


图 3-11 机械下倾角频数分布图

垂直电下倾角 θ_{ed} 是通过调整天线内部的线圈实现的，其频数分布图如图 3-12 所示。

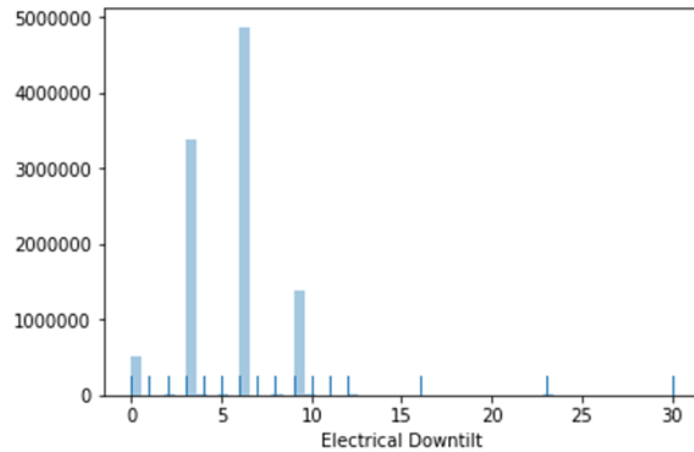


图 3-12 垂直电下倾角频数分布图

根据两个下倾角的分布情况，我们可以看出大多数的小区发射器的下倾角都是类似的，实际信号线的下倾角是机械下倾角与电下倾角的和，因此我们合并“机械下倾角”与“电下倾角”为“下倾角”特征列，以减轻特征的冗余。

$$\theta_d = \theta_{md} + \theta_{ed} \quad (3.2)$$

3.2.4 基于几何位置计算的特征设计

信号线的传播路径极大地影响了 RSRP 在空间中的分布，本节拟推算出直接距离 d_d 、有效距离 d_e 和传播距离 d_p 的一般公式。其中，直接距离 d_d 指测量点距过测量点作与海拔垂线与信号线所成平面的垂直平面形成的交线交信号线的点的距离，有效距离 d_e 指测量点距离其作垂线交信号线垂点的距离，传播距离 d_p 指发射机发射点距的垂点的距离。三维坐标系下的模型如图 3-13 所示。

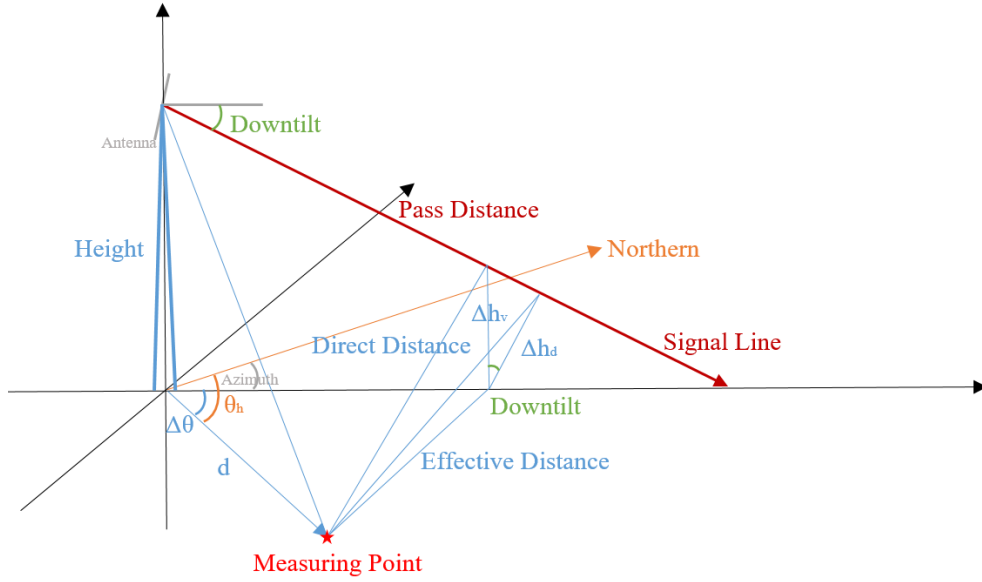


图 3-13 三维坐标系下建模示意图

根据信号塔坐标 (x_{cell}, y_{cell}) 和测量点坐标 (x, y) 易得

$$d = \sqrt{(x_{cell} - x)^2 + (y_{cell} - y)^2} \quad (3.3)$$

$$\theta_h = \arctan\left(\frac{x_{cell} - x}{y_{cell} - y}\right) \quad (3.4)$$

进而得到发射机至测量点水平方向差角 $\Delta\theta$

$$\Delta\theta = \theta_h - \theta_a \quad (3.5)$$

于是可以计算 d_d 、 d_e 和 d_p

$$\Delta h_v = h - d * \cos \Delta\theta * \tan \theta_a \quad (3.6)$$

$$\Delta h_d = \Delta h_v * \cos \theta_d \quad (3.7)$$

$$d_d = \sqrt{(d * \sin \Delta\theta)^2 + \Delta h_v^2} \quad (3.8)$$

$$d_e = \sqrt{(d * \sin \Delta\theta)^2 + \Delta h_d^2} \quad (3.9)$$

$$d_p = \sqrt{h^2 + d^2 - d_e^2} \quad (3.10)$$

最终把有效距离 d_e 和传播距离 d_p 作为候选特征。

另外，地图数据提供了小区发射机所在区域的海拔高度，以及测量点栅格位置的海拔高度，因此可以将两列特征合并为“相对海拔高度”。

$$\Delta h_a = \text{altitude}(\text{measuring point}) - \text{altitude}(\text{signal}) \quad (3.11)$$

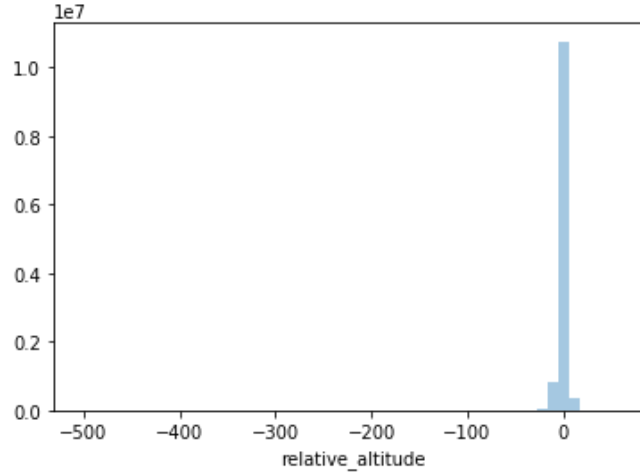


图 3-14 相对海拔高度统计分布情况

3.2.5 基于 3GPP Uma NLOS 模型的特征设计

由于 Cost231-Hata 模型只适用于 1.5-2GHz 的频率范围，从对数据的统计中可以看到，华为所提供的数据小区发射机中心频率在 2GHz 以上，因此考虑选用 3GPP 模型。城区 5G 宏蜂窝传播模型 3GPP Uma NLOS 为：

$$L_b = 161.04 - 7.1 \log_{10} W + 7.5 \log_{10} h - \left(24.37 - 3.7 \left(\frac{h}{h_b} \right)^2 \right) \log_{10} h_b \\ + (43.42 - 3.1 \log_{10} h_b)(\log_{10} d - 3) + 20 \log_{10} f - (3.2(\log_{10}(17.625h_R))^2 \\ - 4.97 - 0.6(h_R - 1.5)) \quad (3.12)$$

其中 W 表示街道宽度， h 表示建筑物高度。

由于 5G 信号的传播较容易收到建筑物的遮挡，因此根据栅格处建筑物的高度，应提取“建筑物是否存在”的特征，通过计算该特征的相关性。

3.3 特征设计结果

综合 3.1 节、3.2 节所述，本章设计的特征如表 3-6 所示。

表 3-6 特征设计结果汇总

特征	解释
$\log_{10} f$	发射机中心频率 f 的对数值
F_1	发射机中心频率类型 1
F_2	发射机中心频率类型 2
F_3	发射机中心频率类型 3
$\log_{10} h_b$	发射机有效高度 h_b 的对数值
$\log_{10} d$	链路距离 d 的对数值
$L_{b-cost231}$	Cost231-Hata 经验估计结果
L_{b-nlos}	3GPP Uma NLOS 经验估计结果
L_{b-spm}	SPM 经验估计结果
$\log_{10} d_e$	有效距离 d_e 的对数值
$\log_{10} d_p$	传播距离 d_p 的对数值
H_D	Height * Distance
F_sub_H	$f - h_b$
F_add_D	$f + d$

$F_{sub\ hd}$	$f - h_b * d$
$F_{sub\ H\ add\ D}$	$f - h_b + d$
$F_{sub\ H\ sub\ HD}$	$f - h_b - h_b * d$
$F_{add\ D\ sub\ HD}$	$f + d - h_b * d$
$F_{sub\ H\ add\ D\ sub\ HD}$	$f - h_b + d - h_b * d$

4 问题二的建模与求解

4.1 问题分析

对于机器学习模型，选择有意义的特征作为模型的输入是十分必要的，特征选择的质量一定程度上决定了模型预测的结果。基于问题一，通过基础特征工程、高阶特征设计（基于几何位置特征、基于传统信道模型参数）的特征构造结果，问题二将针对特征进行筛选。

筛选特征时依据：

- 特征的发散性
- 特征与目标的相关性

基于提供的各小区数据集，设计除去基于经验公式计算的特征外的多个合适的特征，计算这些特征与目标的相关性，并将结果量化、排序，形成表格，并阐明设计这些特征的原因和用于排序的量化数值的计算方法。

4.2 模型建立

在机器学习中，为了使模型得到较好的结果，我们通常采用增加特征的方法，但事实是，随着特征的增加，当特征数量超过一定值的时候，机器学习的效果反而下降，这就是所谓的“维度灾难”。因此选择合适且对模型有较大贡献的特征是十分必要的。

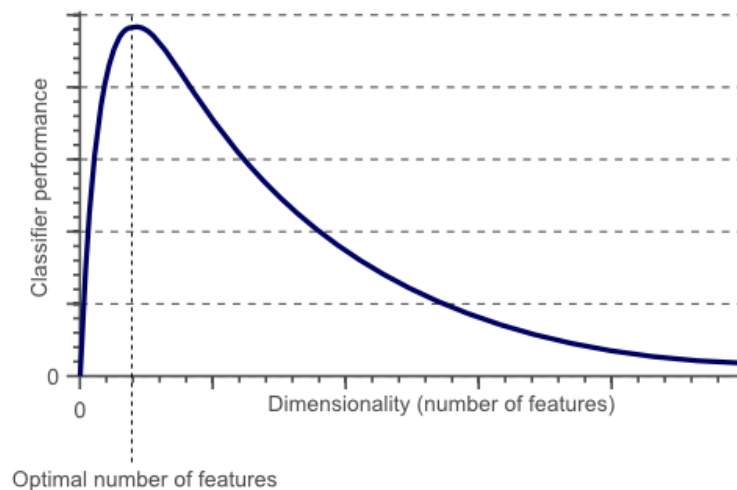


图 4-1 维度与分类器性能之间的关系

本次比赛中采用的特征选择方法有：去除取值变化小的特征、单变量特征选择、线性模型和正则化方法。

4.2.1 去除取值变化小的特征

若某一特征大部分都取相同值，即其方差接近于 0，说明样本在该特征上的差异不大，此特征对模型的最终贡献并不大，可以删除。我们对于数据中所有特征进行方差统计。

方差变化不大的特征有：

表 4-1 特征的方差

特征	Height	Frequency Band	RS Power
方差	0.1927	0.0009	2.5216

4.2.2 单变量特征选择

单变量特征选择指对每一个特征进行测试该特征相关变量之间的变化关系。

1. Pearson 相关系数

Pearson 相关系数考察两个变量之间的相关程度，对于两个变量 X 和 Y ，当相关系数为 0

时，两变量无关系；当 X 值增大（减小），Y 值增大（减小），两变量为正相关，相关系数在 0-1 之间；当 X 值增大（减小），Y 值减小（增大），两变量为负相关，相关系数在-1 与 0 之间。相关系数的绝对值越大，相关性越强，通常情况下相关系数与相关强度的关系如表 4-2 所示

表 4-2 Pearson 相关系数与相关强度关系

相关系数的绝对值	相关强度
0.8 - 1.0	极强相关
0.6 - 0.8	强相关
0.4 - 0.6	中等强度相关
0.2 - 0.4	弱相关
0 - 0.2	极弱相关或无相关

Pearson 相关系数的计算公式为

$$p_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (4.1)$$

利用 Pearson 相关系数计算特征与目标相关性的量化排序结果如表 4-3 所示。

表 4-3 特征名称及其与目标的相关性排序

排序	特征名称	相关性（绝对值）
1	F_add_D	0.38
2	d	0.38
3	$F_sub_h_add_D$	0.32
4	Δh_v	0.29
5	H_D	0.26
6	F_sub_HD	0.26
7	d_e	0.23
8	$F_sub_H_sub_HD$	0.22
9	$F_add_D_sub_HD$	0.10
10	$\Delta\theta$	0.10

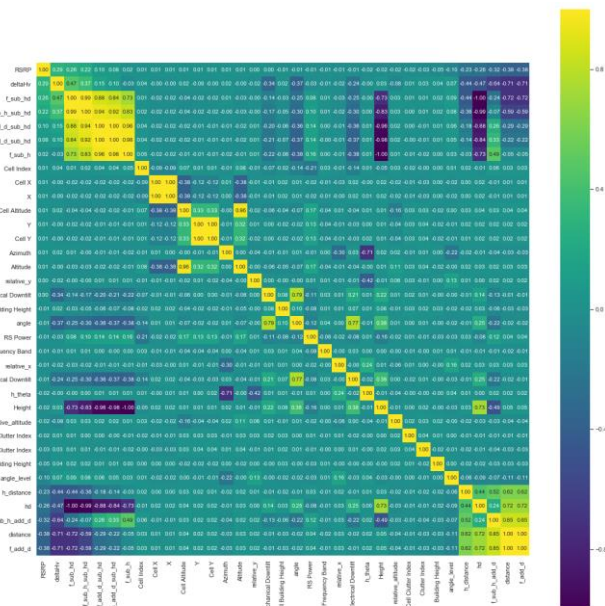


图 4-2 Pearson 相关系数热力图

根据 Pearson 相关系数与相关强度的关系，我们剔除相关系数的绝对值小于 0.2 的特征列。

Pearson 相关系数只能衡量特征与目标之间的线性相关性，因此作为特征排序机制，它只对线性关系敏感，如果关系是非线性的，Pearson 相关性可能会接近 0。

2. Spearman 秩相关系数

由于 Pearson 相关系数描述的是线性关系，为了弥补这一相关性度量值的不足，我们同时引入 Spearman 秩相关系数，以查看不服从正态分布以及分类变量与目标值的相关性。其定义为

ρ = 1 - (6 ∑ d_i^2) / (n(n^2 - 1)) (4.2)

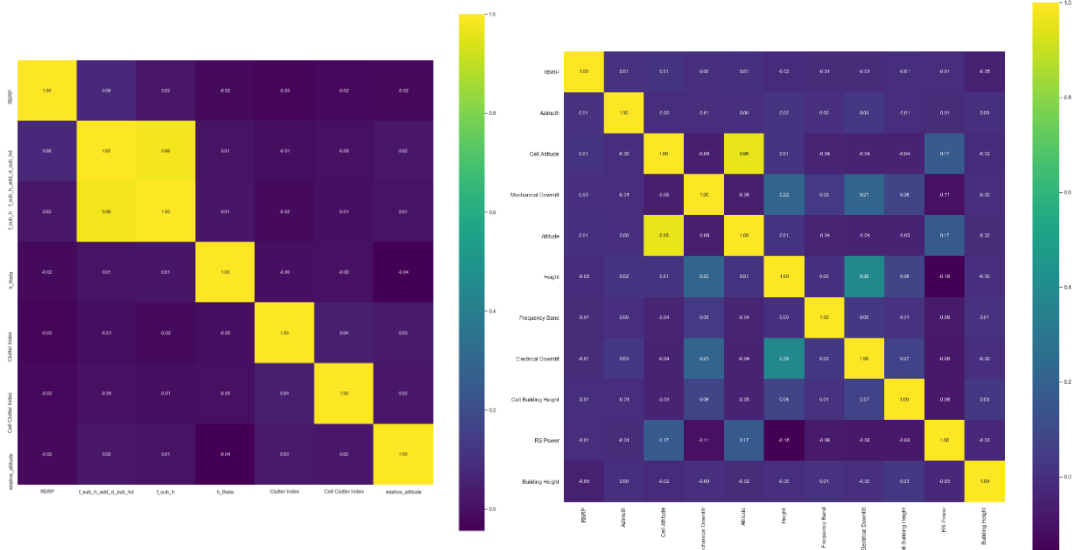


图 4-3 Spearman 相关性热力图

图 4-3 展示了不同特征之间的 Spearman 秩相关系数，结果显示，我们数据中基本都是连续型的数值变量，因此 Spearman 相关系数基本都是 0。其 Pearson 相关系数是更适合相关性的描述指标。

4.2.3 基于模型的特征排序和正则化

针对每个单独的特征和响应变量建立预测模型。对于线性关系可以使用线性模型，若特征和目标变量之间的关系是非线性的，可以使用基于树的方法（决策树、随机森林），或者扩展的线性模型。

正则化就是把额外的约束或者惩罚项加到已有模型（损失函数）上，以防止过拟合并提高泛化能力。损失函数由原来的E(X,Y)变为E(X,Y) + α||w||，其中w是模型参数组成的向量，||·||一般是L₁或者L₂范数，α是一个可调的参数，控制着正则化的强度。当用在线性模型上时，L₁正则化和L₂正则化也称为 Lasso 和 Ridge。

对于特征选择的部分，我们分别应用 Lasso、Ridge、决策树模型进行特征排序。

1. L₁正则化（Lasso）

L₁正则化将系数w的L₁范数作为惩罚项加到损失函数上，由于正则项非零，这就迫使那些弱的特征所对应的系数变成 0。因此L₁正则化往往会使得学到的模型很稀疏（系数w经常为 0），这个特性使得L₁正则化成为一种很好的特征选择方法。

我们针对数据应用 Lasso 模型进行训练，最终获得特征重要性的排序，如表 4-4 所示。

表 4-4 基于 Lasso 模型特征重要性排序

排序	特征	特征重要性
1	d	-10.376293
2	F_add D	-1.565335

3	<i>Clutter Index</i>	-0.157417
4	Δh_a	-0.094187
5	<i>Cell Clutter Index</i>	-0.069195
6	<i>Building Height</i>	-0.051014
7	$\Delta \theta$	-0.024382
8	<i>RS Power</i>	-0.008247
9	d_e	-0.005963
10	θ_h	-0.000430
11	Δh_v	0.020409
12	$F_{sub_H_sub_HD}$	0.028422

从特征重要性的排序结果来看，测量点栅格与小区发射机之间的距离特征是对机器学习模型贡献最大的特征，基于该距离特征设计的其他组合特征也对模型有一定的贡献度。Lasso 模型发现了“地物索引”与预测目标的相关性，因此“地物索引”特征最终应该保留。

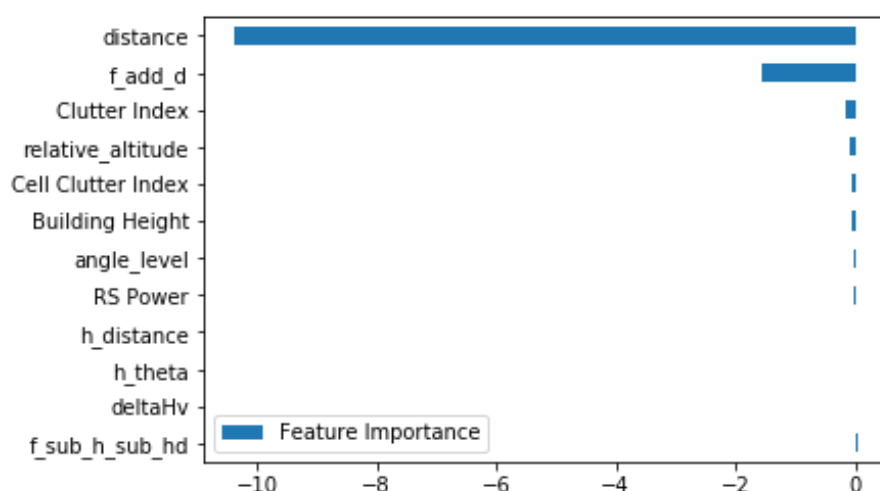


图 4-4 基于 Lasso 模型的特征重要性

2. L_2 正则化 (Ridge regression)

L_2 正则化将系数向量的 L_2 范数添加到了损失函数中。由于 L_2 惩罚项中系数是二次方的，这使得 L_2 和 L_1 有着诸多差异，最明显的一点就是， L_2 正则化会让系数的取值变得平均。对于关联特征，这意味着他们能够获得更相近的对应系数。 L_2 正则化会让各个系数趋于相同的特点。

L_2 正则化对于特征选择来说一种稳定的模型，不像 L_1 正则化那样，系数会因为细微的数据变化而波动。所以 L_2 正则化和 L_1 正则化提供的价值是不同的， L_2 正则化对于特征理解来说更加有用——表示能力强的特征对应的系数是非零。

表 4-5 基于 Ridge 模型特征重要性排序

序号	特征	特征重要性
1	f	-65.363017
2	F_{add_D}	-43.069413
3	F_{sub_HD}	-38.166066
4	h	-36.292658
5	F_{sub_H}	-29.070466
6	d_a	-27.197108
7	d	22.293753
8	$F_{sub_H_add_D_sub_HD}$	20.420179
9	$F_{add_D_sub_HD}$	-15.872271
10	$F_{sub_H_add_D}$	-6.776769
11	$F_{sub_H_sub_HD}$	-1.873345

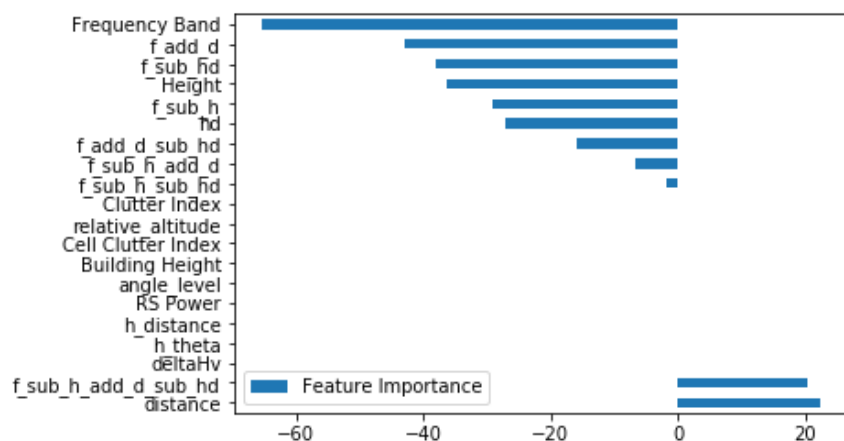


图 4-5 基于 Ridge 模型特征重要性

3. DecisionTreeRegressor

单独对每一个特征进行建模，并进行交叉验证，得到每一个特征对于决策树模型的贡献度，汇总于表 4-3，由于篇幅限制，仅展示前 12 个特征的重要性。

表 4-6 特征对于决策树模型的重要性排序

序号	特征	特征重要性
1	d	0.147
2	F_add_D	0.147
3	$F_sub_H_add_D$	0.116
4	Δh_v	0.106
5	$H_distance$	0.078
6	H_D	0.076
7	F_sub_HD	0.076
8	$F_sub_H_sub_HD$	0.055
9	$\Delta \theta$	0.021
10	θ_h	0.020
11	$F_add_D_sub_HD$	0.019
12	$F_sub_H_add_D_sub_HD$	0.012

4.3 特征选择结果

根据 4.1 和 4.2 小节，最终用于训练模型的特征如表 4-7 所示。

表 4-7

序号	特征
1	h
2	d
3	$\Delta \theta$
4	f
5	Δh_v
6	$Clutter\ Index$
7	Δh_a
8	$Building\ Height$
9	$Cell\ Clutter\ Index$
10	d_e
11	θ_h
12	H_D

13	F_add_D
14	F_sub_HD
15	$F_sub_H_add_D$
16	$F_sub_H_sub_HD$
17	$F_add_D_sub_HD$
18	$F_sub_H_add_D_sub_HD$

5 问题三的建模与求解

5.1 问题分析

5.1.1 特征集分析

进行了特征工程之后，可以得到一个整体的特征数据集，这个数据集一共有 1200 万条数据，其中每条数据有 9 项特征数据，其中有一项为真实的 RSRP 数据。

特征数据包括多种数据类型，有 0-1 分布的布尔型数据、整形分布的类型数据、以及其他浮点数据。数据集中各个维度的特征都已经进行了初始化，对数值过大的数据进行了对数处理，对异常值数据进行了合适的转换，使得整体的数值分布集中于统一的数量级，为之后的模型数据输入提供了方便。清理了异常数据后数据量从 1200 万条变成了 1042 万条。

5.1.2 评价指标

模型的评价指标包括 PCRR 和 RMSE，下面分别介绍两个评价指标的计算方法和意义。

1. 弱覆盖识别率

弱覆盖识别率 PCRR (poor coverage recognition rate) 反映的是某一个区域被准确的划分为弱覆盖区域的 F1 值。其中某一个区域定义为弱覆盖区域的指标是其消耗强度小于 -103dBm。这一个指标可以更好的帮助运营商规划和优化网络的分布，从而提高用户体验。PCRR 的具体计算过程如下。

对于某一个区域是否为弱覆盖区域以及是否被模型预测为弱覆盖区域，有下列四种方式划分该区域的特征。

True Positive(TP): 实际为弱覆盖区域，预测结果也为弱覆盖区域；

False Positive(FP): 实际为非弱覆盖区域，预测结果为弱覆盖区域；

False Negative(FN): 实际为弱覆盖区域，预测结果为非弱覆盖区域；

True Negative(TN): 实际为非弱覆盖区域，预测结果也为非弱覆盖区域。

统计了测试集所有的区域特征后，可以得到该模型准确率值 Precision 和召回率值 Recall，其定义为

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5.1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5.2)$$

其中 Precision 可以理解为模型预测某个区域为弱覆盖区域时，该区域实际也为弱覆盖区域的概率，Recall 可以理解为模型在所有的弱覆盖区域中准确定位到的弱覆盖区域的概率。在 Precision 和 Recall 的基础上可以计算 PCRR 的值，其定义为

$$\text{PCRR} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.3)$$

PCRR 值为 Precision 和 Recall 的调和平均值，可以理解为两个指标的综合，其反映的是该模型在强弱覆盖区域预测的综合评价指标。在赛题的要求中，PCRR 的值必须大于 20%才给与模型均方根误差的评价机会。

2. 均方根误差

均方根误差 RMSE (root mean squared error) 反映的预测值和真实值分布的整体偏差，以直观的形式反映了模型的准确性，这个评分也是赛题的最终排名指标，其具体的计算方法如下。

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (P^{(i)} - \hat{P}^{(i)})^2} \quad (5.4)$$

其中 $P^{(i)}$ 为模型预测的 RSRP 值， $\hat{P}^{(i)}$ 为真实的 RSRP 值。

5.1.3 求解目标

模型的最终目标是建立一个函数，函数的输入为所有提取的特征，函数的输出为一个预测的 RSRP 值，并且预测的 RSRP 值与真实的 RSRP 值尽量相近，因此这是一个典型的机器学习回归问题。在模型的训练过程中需要查看模型的 PCRR 值是否大于 20%，并且通过测试集中的均方根误差评判模型的优劣。

5.2 模型建立

5.2.1 数据集处理

依据模型训练的数据需要，对整体的数据集做合适的融合与拆分以及存储，以符合实际的模型训练需求。

1. 数据集分析

整个数据集包含 4000 个 CSV 文件，每一个文件代表一个发射塔对应的小区中所有测试区域的数据。依照发射塔之间的正交性，相邻发射塔之间无相互干扰关系，因此可以将所有的数据当成单独的单列数据处理，这样给数据的处理带来了优势。首先将所有的数据由 pandas 整合到一个 CSV 文件中，通过这个 CSV 文件对整个数据集做统一的处理，提取需要的特征，并剔除无关特征，形成可以直接用于机器学习模型的元数据。

2. 随机划分

根据上述的数据集分析，参考机器学习的通常做法，将总体数据集按照一定比例划分为训练集、交叉验证集和测试集，下面简单的介绍数据集划分原则以及相关的原理。

训练集（training set）的作用就是训练模型，形成模型的内部结构和参数估计，参数依赖于训练集，对于同一份数据的同一个模型，参数是统一的。但是在训练模型的过程中，需要引入部分人为规定的变量，比如神经网络隐层大小，决策树的深度等等，这些超参数会影响模型的结构，但是无法被训练集学习，因此需要引入验证集。

验证集（validation）的作用是对超参数的选择，对同一类模型，运用不同的超参数可能得到不同的训练效果，具体哪个超参数有利于模型的学习，需要验证集的验证结果确定。

测试集（test set）不参与模型的训练，也不参与模型的选择，与模型没有任何信息交互，可以用于拟合实际情况，要尽可能贴近实际中总体的分布，需要从总体中随机抽取。

对于一般规模的数据集一般采用 60%、20%、20% 的比例划分训练集、验证集和测试集，但是对于大规模数据，由于数据量庞大，数据具有同构性，小比例数据就能反映真实分布，因此一般采用 90%、5%、5% 的比例划分训练集、验证集和测试集。本次实验的数据规模在千万级，且数据具有相似的分布，因此按照大规模数据集随机划分训练集、验证集和测试集，其具体的数据量依次为 938 万条、52 万条、52 万条。

3. 结合地理信息

根据提供数据的特点，同一个发射塔编号的小区受到同一个信号源的影响，即该小区中的所有检测点收到的信号都具有相同的频率、相位等。因此我们可以合理的推测同一个小区检测点的训练可能影响到该小区其他测试点的测试结果，因为这两个检测点具有极其相似的部分特征。

为了保持测试数据集以及验证数据集的纯净性，我们改进了数据集的分割方法。我们首先按照小区编号将原始数据直接分割成训练集，验证集和测试集。这种情况下，同一个小区只能位于这三个数据集其中一个。这样就不会再出现同一个小区部分检测点位于训练集，部分检测点位于测试集而影响测试结果的情况。按照小区分割数据集后，后续的数据处理过程中每个数据集的处理过程完全一致。

运用新的数据集划分方法，我们得到了更加精确的模型度量，使得最终本地的 RMSE 测试结果同线上的实际 RMSE 分数接近。

5.2.2 模型选择

针对数据集整体的特征维度较小，特征的数值分布较为集中，数据的规模量庞大的特点，考虑在该数据集上运用如下三种模型：线性模型（Linear Regression）、神经网络（Neural Networks）、决策树模型（Decision Tree Model）。

1. 线性模型

参考传统的经验公式，其计算过程可以看成是各个计算的组成部分的线性组合，而线性组合的各个因子可以看成是特征数据，而每个因子的系数就是线性模型学习到的参数。这样，传统的模型如 Cost 231-Hata 可以看成是线性模型的经验拟合。在这个思想的驱动下，可以简单的将特征工程获取的特征提供给线性模型，在更丰富的特征以及更多的数据拟合的情况下，可以得到比传统模型更好的预测结果。

线性回归是指全部由线性变量组成的回归模型，其中最简单的单变量线性回归(Single Variable Linear Regression)是用来描述单个变量和对应输出结果的关系，其表达式为

$$Y = a * X + b \quad (5.5)$$

在实际的建模过程中遇到的问题往往更加复杂，用单个变量不能满足描述输出变量的关系，所以需要用到更多的变量来表示与输出之间的关系，也就是多变量线性回归(Multi Variable Linear Regression)。多变量线性回归模型。

$$Y = a_1 * X_1 + a_2 * X_2 + \dots + a_n * X_n + b \quad (5.6)$$

其中 a_i 为系数， X_i 是变量， b 为偏置。因为这个函数只有线性关系，所以只适用于建模线性可分数据，通过系数权重来区分每个特征变量的重要性。在本模型中，通过对 RMSE 损失函数的反向梯度下降确定这些权重 a_i 和偏置 b 。

线性模型在大规模数据的情况下训练速度快，并且依据权重系数给出每个变量的理解和解释，但是线性模型只能做线性的拟合，因此其对异常值很敏感，且无法拟合复杂形状的数据。在本模型中各个特征为结合模型提出，由于现实环境的复杂性，各个特征之间可能不止具有简单的线性关系，且实际的有效影响因素可能无法顺利挖掘，因此需要进一步引入更加复杂的，具有更强拟合能力的模型。

2. 决策树模型

决策树模型是运用于分类以及回归的一种树结构。决策树由节点和有向边组成，一般一棵决策树包含一个根节点、若干内部节点和若干叶节点。决策树的决策过程需要从决策树的根节点开始，待测数据与决策树中的特征节点进行比较，并按照比较结果选择选择下一比较分支，直到叶子节点作为最终的决策结果。决策树具有较强的鲁棒性，可以有效的处理缺失值和极端值。决策树模型另一个优点是可以同时处理连续值和离散值，这一点和特征输入是一致的。相比于普通神经网络的黑盒模型，决策树在逻辑上可以得到很好的解释性。

在本数据集中由于数据量过于庞大，而相应的单个数据维度较小，我们在原数据集以及特征集的基础上补充了大量的特征组合，以提高决策树的泛化能力。为了去除无关特征对模型的影响，我们同时对决策树进行了剪枝处理。在决策树的设计过程中，我们选取了最大深度，最小叶子结点数目作为决策树模型的超参数，我们对这两个参数进行了网格搜索，最终获取了最佳的决策树模型。

但是总体来说，单个决策树模型具有诸多限制，并且输入数据的特征维度依旧较小，这个因素从根本上限制了决策树的学习能力，因此决策树模型的拟合能力还有较大的提升空间。

3. 神经网络模型

由于线性模型拟合能力的缺失，因此考虑引入了神经网络模型，神经网络（Neural Networks）是由大量的、简单的处理单元（称为神经元）广泛地互相连接而形成的复杂网络系统，它反映了人脑功能的许多基本特征，是一个高度复杂的非线性动力学习系统

神经网络主要由输入层，若干隐层，和输出层构成，从输入层到隐层最后到输出层通过权重参数以及偏移参数连接。每一个计算结点获取数据后会通过一个激活函数对数据进行非线性处理，非线性的激活函数有效的提高了模型的拟合能力。整个过程可以通过一个迭代过程表示。

$$y_k = \sum_{i=0}^I W_{ik} * X_i + B_k \quad (5.7)$$

$$Y_k = f(y_k) \quad (5.8)$$

其中 X_i 代表上一层某个结点的输出向量， W_{ik} 为权重矩阵， B_k 为偏置值向量， y_k 为本层输入向量， Y_k 为本层通过激活函数 $f(\cdot)$ 之后的结果。

在本次模型训练中，神经网络部署为 8 个数据点的输入，接下来有两层隐层，隐层维度数分别为 16 和 8，最后汇聚到 1 的输出层，其中两个隐层采用了 \tanh 作为激活函数，最后输出采用了线性输出，不做额外的激活处理。模型的总体架构如表 5-1 和图 5-1 所示。

表 5-1 神经网络结构示意图

输入层	全连接层	激活层	全连接层	激活层	全连接层	全连接层	输出层
8	20	0	10	0	10	10	1

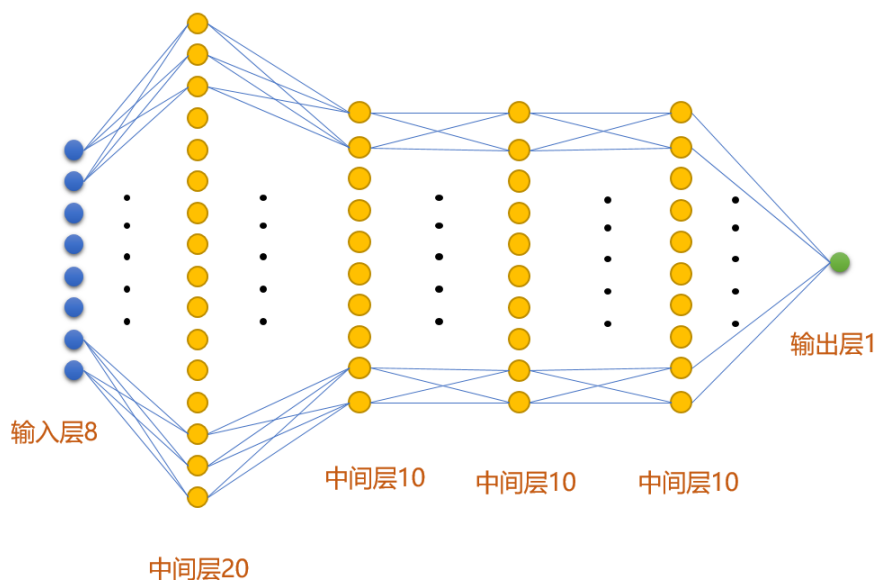


图 5-1 神经网络结构示意图

用于最终需要拟合的数据均为负值，因此选择了具有 $(-1, 1)$ 分布的激活函数 \tanh 。经过两轮的激活函数处理，模型具备了较好的提取非线性特征的能力。但是最后的输出的数据绝对值均小于 1，因此在第二层激活层之后补充了两层全连接层，用于将提取的非线性特征组合成预测的 RSRP 值。最后的模型输出就是两层全连接层的线性组合。

在模型训练过程中，模型使用了 Adam 作为模型的优化算法，学习率设置为了 0.001，损失函数为 RMES。综合考虑模型规模以及 tensorflow 训练时内存的加速算法，Batch 大小设置成了 128。

由于整体的数据量极大，虽然采用了简单的网络结构。但是如果直接将所有的数据读取进，由于读取大数据时的内存限制，模型的训练速度会急速下降，因此需要考虑更加快速的数据读取方式。在模型的训练过程中，我们将大数据集做了充分的分割，有效的提高了深度模型的学习速度。

5.3 模型解算及结果分析

5.3.1 本地预测结果分析

针对简单线性模型，决策树以及神经网络，我们分别在本地随机划分的数据集中进行训练和

测试，相应的训练结果如表 5-2 所示。

表 5-2 线性模型和决策树的本地训练结果

评价指标	线性模型	决策树
MSE	97.627	56.428
RMSE	9.880	7.511
MAE	7.845	5.109
R2	0.148	0.508

可以看出，作为 **baseline**，简单的线性模型的整体拟合能力较差，模型的相关性系数 **R2** 较低。但是线性模型给了我们启示，说明我们提取的特征和最终的结果之间的综合相关性依旧较低，我们需要设计更加复杂的模型进一步处理拟合的特征。

决策树模型对于由于该数据集的采样过程中可能产生一定量的异常数据，因此决策树模型的整体表现明显优于简单线性模型，模型的相关性系数 **R2** 也明显高于简单线性模型。

神经网络模型的 **RMSE** 为 10.320，神经网络具有数据的强拟合能力，但是在本地的结果可以看出神经网络的 **RMSE** 得分并没有明显的优于决策树模型。这可能和数据的特征提取有关，经过特征工程后，原始数据会经过大量的转换与提取，但同时由于训练所采用的特征为原始数据的特殊设计，因此可能丢失部分重要的细节，这些细节的丢失可能不利于神经网络的进一步拟合。

于此同时，我们设计的神经网络模型层数较浅，并且中间层的维度规模较低，这也可能限制了神经网络的拟合能力。后续我们可能考虑更加深层的网络以及更加泛化的特征提取，以提高神经网络的性能。

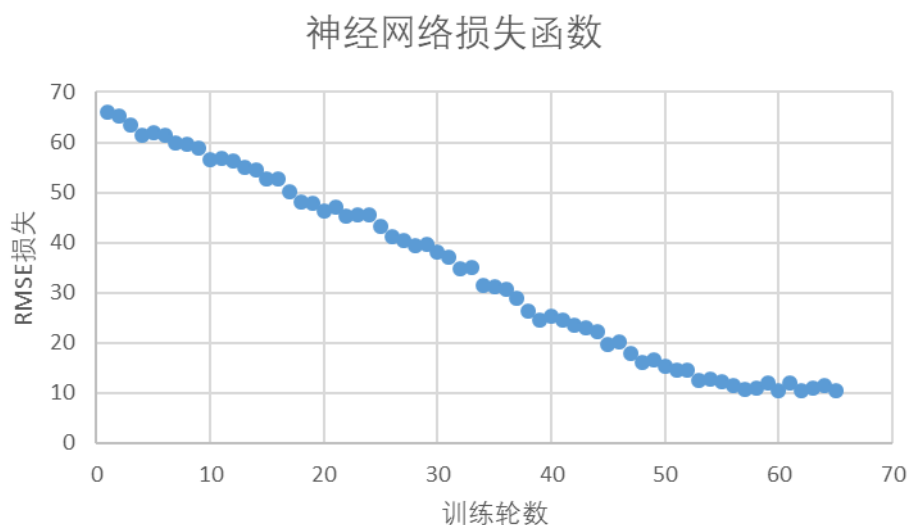


图 5-2 神经网络损失函数变化曲线

5.3.2 平台测试结果展示

结合本地的测试结果，我们最终确定使用决策树模型进行在线提交。通过在数据预处理阶段添加数据处理过程，并且提交训练好的决策树模型，这样使得本地的数据处理过程与云平台上的数据处理过程保持一致，我们最终在平台上部署了决策树模型。决策树模型的平台测试结果如图 5-3 所示。

作品名称	阶段	提交时间	得分	反馈信息
model-yubzhu-prod	初赛	2019-09-23 08:12:11	13.1483	success
model-yubzhu-prod	初赛	2019-09-22 21:11:44	13.7488	success
model-yubzhu-prod	初赛	2019-09-22 19:15:44	13.8268	success
model-yubzhu-prod	初赛	2019-09-22 17:12:22	无	scoring job failed
model-yubzhu-prod	初赛	2019-09-22 12:19:45	51.6903	success
model-yubzhu-prod	初赛	2019-09-22 10:44:17	无	scoring job failed
model-yubzhu-prod	初赛	2019-09-21 22:23:02	无	scoring job failed
model-yubzhu-test	初赛	2019-09-21 10:00:03	16.7258	success

图 5-3 华为 ModelArts 平台提交结果

平台测试结果为 13.1483，这个结果与本地的预测结果有较大差距，这可能和本地的测试环境和平台测试环境的差异有关。后续我们通过对本地数据集的合理划分，按照小区编号区分训练集和测试集，避免了测试集分布被训练过程感知的情况，最终的本地训练结果和平台测试结果保持了一致。

5.3.3 结论与展望

无论时简单的线性模型，还是决策树模型以及神经网络模型，我们发现部分的特征在三种模型中都产生了优于其他特征的性能，比如距离度量，环境度量。这表明整体的模型分布具有很强的空间相关性，而如何有效地挖掘出空间的相关性成为了模型拟合的关键。对于一个信号塔极其相应的信号分布，我们可以得出一个简单的结论，该信号分布是一个和空间环境严格相关的复杂函数，这个函数需要充分的考虑空间特征、地面材质特征等等。

这给我们进一步研究提供了可靠的思路。对于这种结构化的数据结构，类似于图片和网络结构，我们可以考虑采用结构化的处理方法。进一步研究方向可能是需要考虑在该数据集上运用 CNN、GCN 等结构化深度模型，在结构化莫的帮助下，可能得到更好的拟合效果。

参考文献

- [1] Martinez O N , Rodriguez C , Marvin A O . Propagation characteristics of Managua city based on Standard Propagation Model (SPM) at 850 MHz for 3G-WCDMA systems[C]// PROCEEDINGS OF THE 2014 IEEE CENTRAL AMERICA AND PANAMA CONVENTION (CONCAPAN XXXIV). IEEE, 2014.M awjoud S A , Mawjoud S A . Comparison of Propagation Model Accuracy for Long Term Evolution (LTE) Cellular Network[J]. International Journal of Computer Applications, 2013, 79(11):41-45.
- [2] Castro B S L , Gomes I R , Ribeiro F C J , et al. COST231-Hata and SUI Models performance using a LMS tuning algorithm on 5.8GHz in Amazon Region cities[C]// European Conference on Antennas & Propagation. IEEE, 2010.
- [3] Salehi H, Das S, Biswas S, et al. Data mining methodology employing artificial intelligence and a probabilistic approach for energy-efficient structural health monitoring with noisy and delayed signals[J]. Expert Systems With Applications, 2019: 259-272.
- [4] 于仰源, 孙宜军, 王磊, et al. 一种基于 MR 数据修正无线传播模型的方法[J]. 移动通信, 2019(3).
- [5] 蔡方萍. 基于相关向量机的无线通信系统信道检测[J]. 计算机工程与设计, 2017(10):60-63+255.
- [6] 臧运娟. 无线通信网络流量预测综述[J]. 电子设计工程, 2017, 25(4):150-153.
- [7] 薛玉洁, 周杰, 陈珍. 信道模型在无线通信环境中的适应性分析[J]. 黑龙江大学自然科学学报, 2019(1).