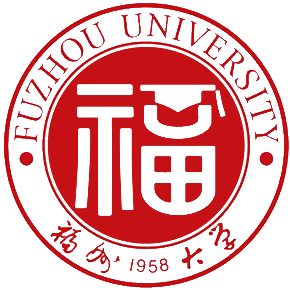
****  **** 

**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** |  |
| **参赛队号** |  |
| **队员姓名** | **1.** |
| **2.** |
| **3.** |

**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：

本文旨在

针对问题一

针对问题二

目 录

[目 录 2](#_Toc19956309)

[1 问题重述 4](#_Toc19956310)

[1.1 问题背景 4](#_Toc19956311)

[1.2 问题描述 4](#_Toc19956312)

[2 合理假设及符号系统 5](#_Toc19956313)

[3 问题一的建模与求解 6](#_Toc19956314)

[3.1 问题分析 6](#_Toc19956315)

[3.2 模型建立 6](#_Toc19956316)

[3.3 模型解算及结果分析 6](#_Toc19956317)

[4 问题二的建模与求解 7](#_Toc19956318)

[4.1 问题分析 7](#_Toc19956319)

[4.2 模型建立 7](#_Toc19956320)

[4.3 模型解算及结果分析 7](#_Toc19956321)

[5 问题三的建模与求解 8](#_Toc19956322)

[5.1 问题分析 8](#_Toc19956323)

[5.2 模型建立 8](#_Toc19956324)

[5.3 模型解算及结果分析 8](#_Toc19956325)

[参考文献 9](#_Toc19956326)

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。根据自己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。

1. 合理假设及符号系统

## 问题假设

1. 假设相邻基站测量信号相互正交
2. 水平方向角是发射机发射方向和正北方形成的夹角，对小区覆盖方向有着重要
3. 根据领域专家知识设计特征，请结合给出的信道模型作答，可以通过公式代入数值进行计算，也可以将公式扩展/截取，以建立合适的特征为准
4. 不考虑天线增益
5. 问题一是使用信道模型设计特征并阐述原因，问题二是设计（除第一问之外的）多个特征并阐述，同时对特征的重要性进行排序。
6. 假设用户天线高度都相同
7. 移动通信网络的发射机使用定向天线

## 2.2 项目符号

1. 问题一的建模与求解

## 问题分析

### 无线信号传播特性

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。

### 无线信号传播模型

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（标准传播模型）等。

## 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：

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

### 异常值处理

绘制原始数据的三维空间分布情况，我们可以看到。。。

### （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，所以需将非主波束方向的数据过滤掉。~~

### （2）剔除建筑物高度与地物索引不符的异常数据

统计栅格（X, Y）上建筑物的高度分布，大量栅格没有建筑物存在，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。

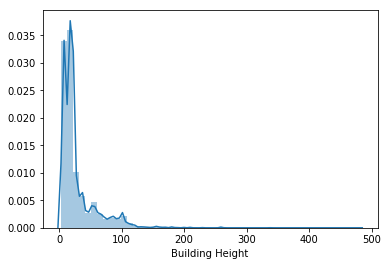
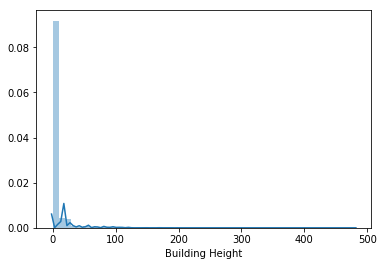


图 3 - 1 Building Height 分布

建筑物的高度应该与观测点的地物类型相关，对每个地物类型下的建筑物高度进行统计，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。

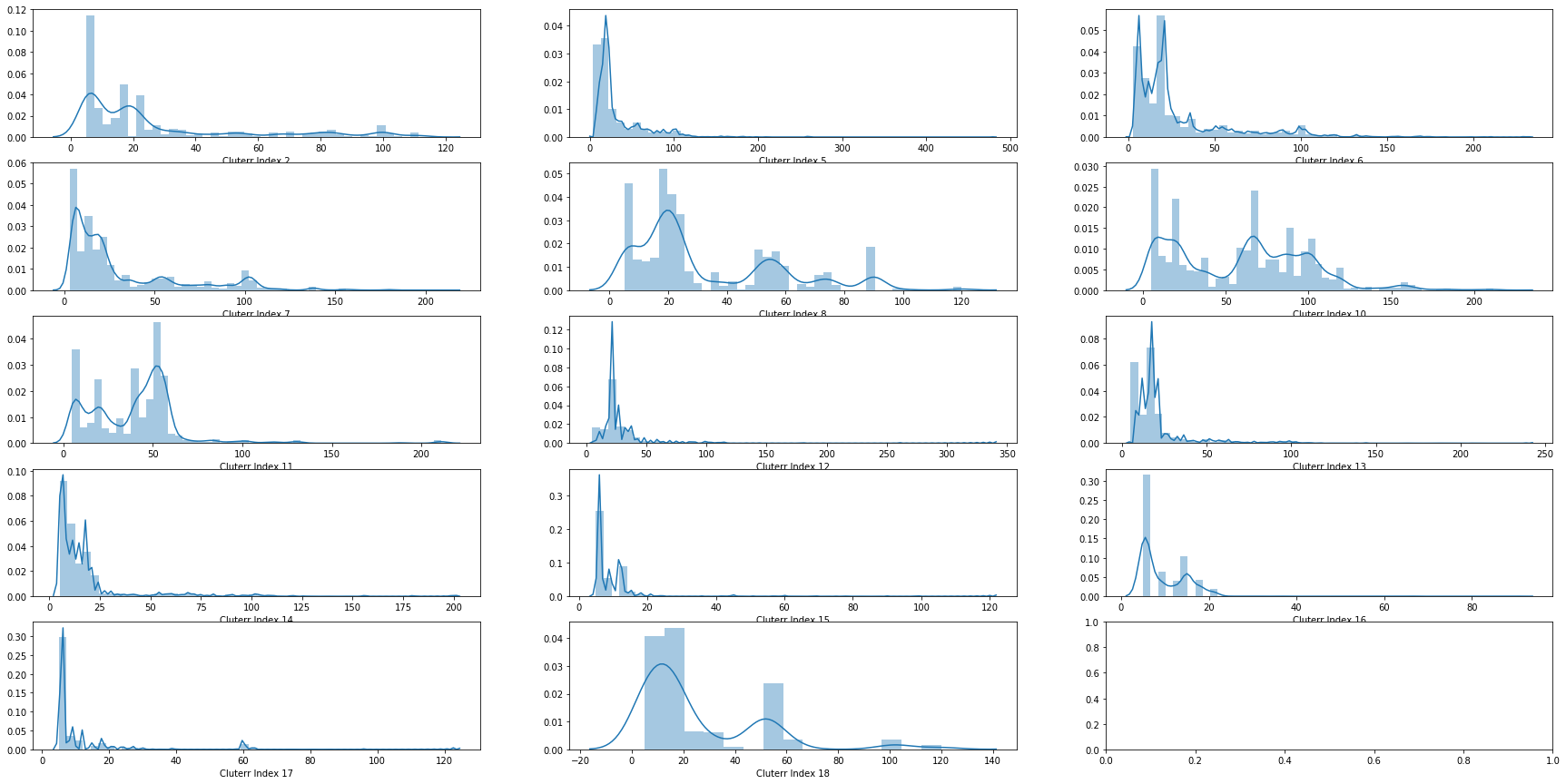


图 3 - 2 不同地物类型下建筑物高度分布

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下图：

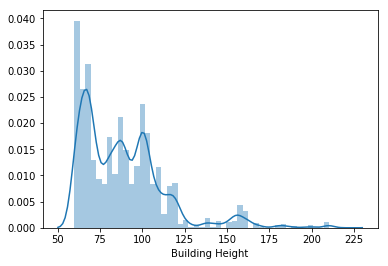
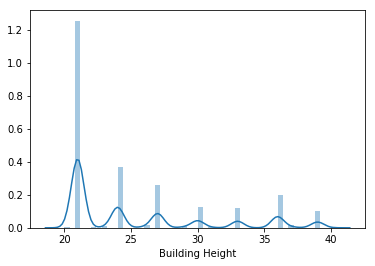
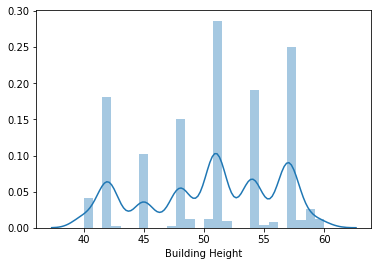
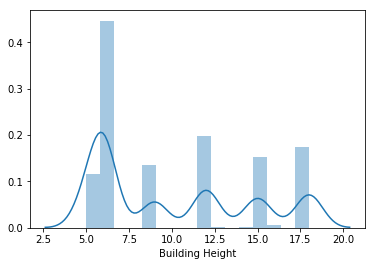


图 3- 3 剔除异常值后地物类型与Building Height 分布

### 基于经典的传统经验信道模型 Cost 231-Hata的特征工程

### （1）基于小区发射机中心频率的特征

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

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 |

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:

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 |

### （2）基于小区发射机有效高度 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如下表：

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。

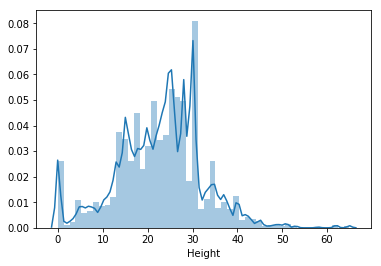


图 3- 4 基站高度分布

### （3）基于链路距离 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，对于测量点与基站的欧式空间距离，查看其分布如下：

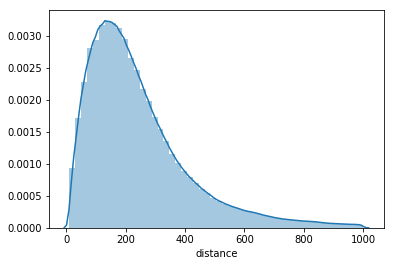
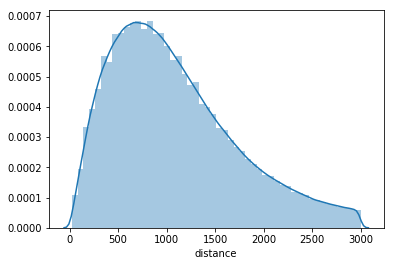
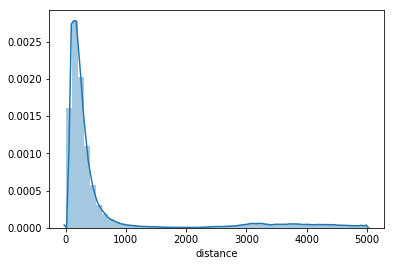


图 3- 5 测量点与基站距离分布

图3-2 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- 6 测量点与小区发射机距离分布

### （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，组合方式如下表：

表 3- 4 特征组合方式

|  |  |  |
| --- | --- | --- |
| 特征列 | 组合方式 | 与目标相关性 |
| 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 |

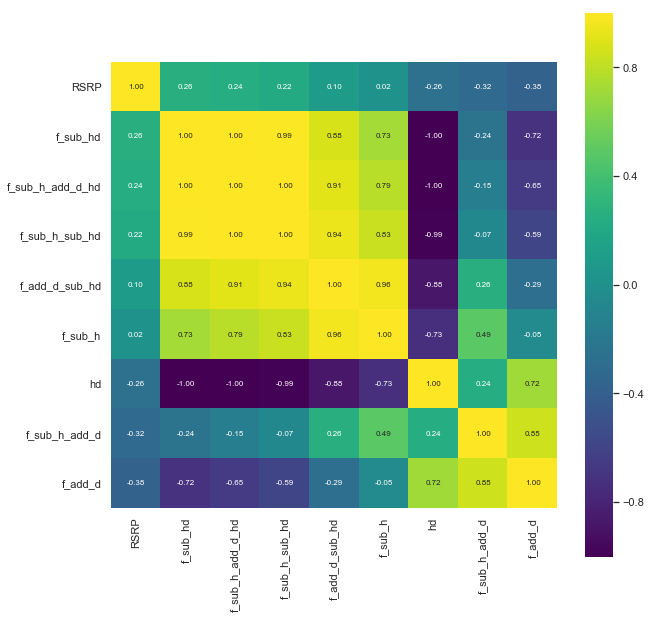


图 3- 7 组合特征与目标的相关性矩阵

### 基于3GPP Uma NLOS 模型的特征构建

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：

其中 表示频段， 表示基站天线的高度， 表示终端天线高度， 表示终端和基站间的距离，表示建筑物的高度，表示街道宽度。

一般取1.5m，对于城区调整范围为 , 农村范围调整范围为 ，据此，添加特征“Hut”

通过调整传播模型的系数， 使得利用传播模型计算得到的路径损耗值与实测路径损耗值误差最小

## 3.3 模型解算及结果分析

综合以上两小节所述，根据Cost 231-Hata模型设计的特征如下表：

|  |  |  |
| --- | --- | --- |
| 特征 | 解释 | 说明 |
| 发射机中心频率*f* |  |  |
| 发射机中心频率类型1 | 0或1 |  |
| 发射机中心频率类型2 | 0或1 |  |
| 发射机中心频率类型3 | 0或1 |  |
| 发射机有效高度 Height |  |  |
| 链路距离 *d* |  |  |
| H\_D | Height \* Distance |  |
| F\_sub\_H | Frequency Band – Height |  |
| F\_add\_D | Frequency Band + Distance |  |
| F\_sub\_hd | Frequency Band – Height \* Distance |  |
| F\_sub\_H\_add\_D | Frequency Band – Height + Distance |  |
| F\_sub\_H\_sub\_HD | Frequency Band – Height – Height \* Distance |  |
| F\_add\_D\_sub\_HD | Frequency Band + Distance – Height \* Distance |  |
| F\_sub\_H\_add\_D\_sub\_HD | Frequency Band – Height + Distance – Height \* Distance |  |

1. 问题二的建模与求解

## 问题分析

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。

除去基于经验公式计算的特征外，

## 4.2 模型建立

## 4.3 模型解算及结果分析

1. 问题三的建模与求解

## 5.1 问题分析

## 5.2 模型建立

## 5.3 模型解算及结果分析

参考文献