

# 基于 LSTM 无线智能传播模型

## 摘要

本文通过对无线传播实测数据进行挖掘分析，基于 Cost 231-Hata 模型、结合实测数据字段构造出 34 个特征，对特征进行相关性分析，并根据相关性的大小挑选出 25 个有效特征构建平均信号接收功率(RSRP)构建 LSTM(长短期记忆网络)预测模型。该模型训练集线下 RMSE 为 0.67，线上测试集为 15.6。训练集线下弱覆盖识别率 (PCRR : Poor coverage recognition rate)为 0.46。

关键字：无线传播，Cost 231-Hata，LSTM

## 一、研究背景及意义

### 1.1 研究背景

随着 5G 技术的成功，5G 技术在近年开始进行商业应用落地。运营商在部署 5G 网络的过程中，需要合理规划在不同的区域进行建设 5G 基站，以满足居民的使用需求。在整个无线网络规划流程中，需要考虑到不同区域的环境因素，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，受到各种环境因素影响，比如山川、平原、城市之间的大气密度、地球自身曲率等，使得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。

在传统的无线传播模型的建立过程中，往往首先需要对传播场景进行划分，每一个场景对应一个传播经验模型。然而，经验模型在实际使用中往往不够精确，所以仍然需要通过采集大量的工程参数以及实际平均信号接收功率测量值进行经验模型公式的修正。从所述过程中可以看到，传播模型建立本质上是一个函数拟合的过程，即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。

### 1.2 研究意义

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。

随着大数据 AI 技术的发展，使得在庞大且复杂的数据处理上有了实际上的挖掘可能。通过对实测数据进行数据挖掘传播特征，基于信号传播原理进行数学建模，建立一个无限传播 AI 系统，精确评估出基站选址与各个用户的传播效率关系，为运营商在实际建设基站选址中提供可靠性的数据参考，从而降低网络建设成本和提高网络建设效率。

二、文献综述

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)。

三、字段特征符号定义

字段名称	含义	单位
$X_c$	小区所属站点的栅格位置，X 坐标	-
$Y_c$	小区所属站点的栅格位置，Y 坐标	-
$h_b$	基站天线有效高度	m
$H_c$	发射机相对地面的高度	m
$\theta_{ED}$	发射机垂直电下倾角	Deg
$\theta_{MD}$	发射机垂直机械下倾角	MHz
$f$	射机中心频率	MHz
$P$	射机发射功率	dBm
$x$	栅格位置，X 坐标	-
$y$	栅格位置，Y 坐标	-
$A_{(x,y)}$	目标栅格 $(x,y)$ 上的海拔	m
$A_{c(x,y)}$	发射机栅格 $(x,y)$ 上的海拔	
$H_{b(x,y)}$	栅格 $(x,y)$ 上的建筑高度	m

$I_c$	小区所属站点的栅格位置地物类型索引	-
$I_{(x,y)}$	目标栅格 $(x,y)$ 上地物类型索引	-
$d$	目标栅格与发射机水平距离	km
$\Delta h_v$	目标栅格与信号线相对高度	m
$S$	目标栅格与天线的距离	m
$L$	信号线到目标栅格竖直方向上的距离	m
$PL$	传播路径损耗	dB
$N_c$	小区所属栅格发射机的数量	-
$N_t$	发射机服务目标数量	-

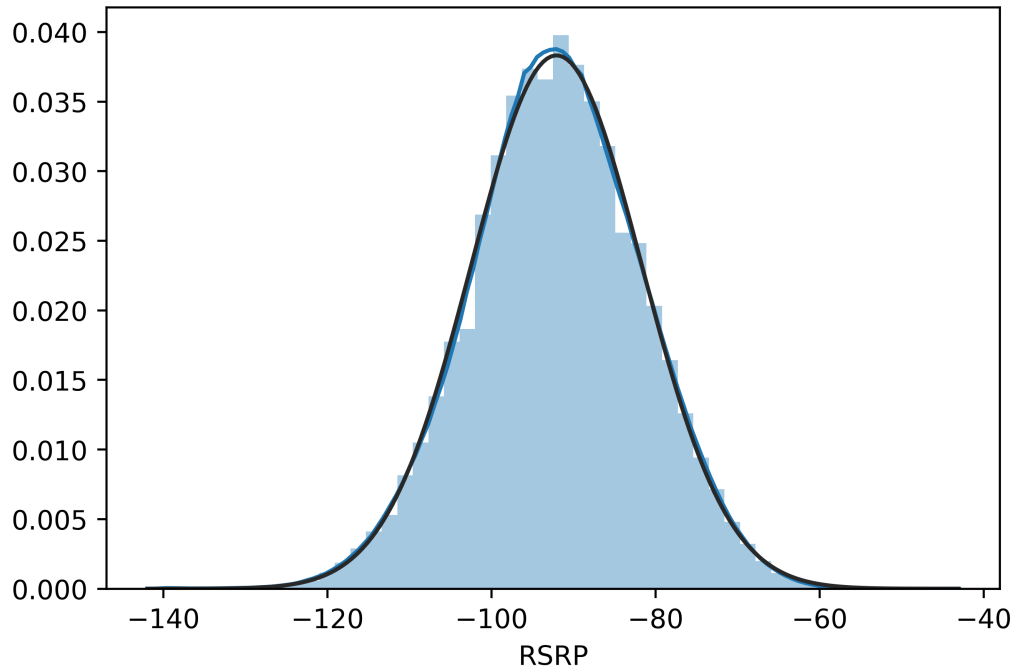
#### 四、数据汇总及清洗

##### 4.1 剔除与数据定义相违数据

对 4000 个训练集汇总在一起，共得到 12011833 条样本数据（不含表头）。由于题目地物类型名称的编号对各种建筑有高度规定，比如地物类型编号为 10 的建筑高度定义为大于 60 米，因此在数据中可以把该类地物类型建筑高度小于 60 米的作为异常数据并进行剔除。综合对比所有栅格建筑高度和对应的地物建筑规定定义，共发现 1175100 个异常数据，占总体数量 9.78%。

##### 4.2 中心极限定律剔除极端数据

先对数据 RSRP 数据进行描述直方图，如图下。可见其分布服从正态分布，应用中心极限定律，或者叫 $3\sigma$ 准则（又叫拉依达准则），把 RSRP 数据取值不在 $(\mu - 3\sigma, \mu + 3\sigma)$ 内的数据剔除，其中 $\mu$ 为 RSRP 的均值， $\sigma$ 为标准差。目的在于剔除出现概率不到 0.3%的极端数据。此操作共剔除 2732 个样本，占比 0.23%。



## 五、特征工程

### 5.1 离散化 RSRP

根据题目的标准，弱覆盖判决门限 $P_{th}$ 的值定为-103 dBm，可以先对小于 $P_{th}$ 则为弱覆盖并标记为 1，若大于等于 $P_{th}$ 则为非弱覆盖并标记为 0。在后面对于 RSRP 预测前，可以先对 RSRP 是否属于弱覆盖进行预测，然后把预测的弱覆盖标签作为一个特征再对 RSRP 进行预测。

### 5.2 One-Hot 地物类型

由于地物类型为离散化的数值，但这些数值并不直接与 RSRP 形成相关。在回归，分类，聚类等机器学习算法中，特征之间距离的计算或相似度的计算是非常重要的，而我们常用的距离或相似度的计算都是在欧式空间的相似度计算，计算余弦相似性，基于的就是欧式空间。One-Hot 处理后也方便这次回归任务中对比不同地物类型对 RSRP 的影响。

### 5.3 目标与发射塔海拔高度差

一般情况下，如果目标与发射机海拔高度差越低，说明目标和发射机的地形类似，环境越接近。举个例子，如果海拔高度差很大，比如发射机一般在居民聚集地方，多数为平原，如果目标在山上，则目标接收到的信号较弱。计算公式为：

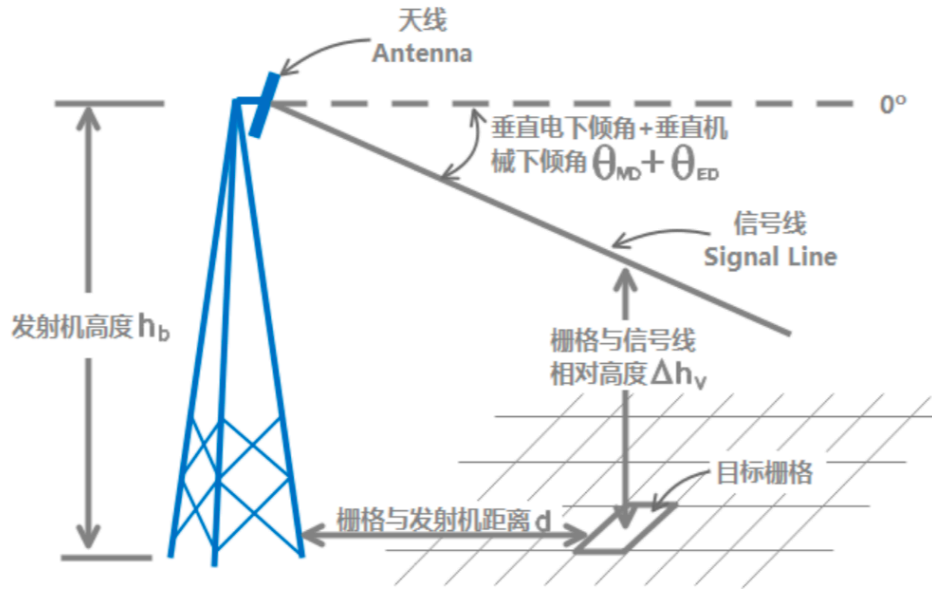
$$\Delta A = A_c - A$$

其中 $A_c$ 为发射机的海拔；

$A$ 为目标的海拔

#### 5.4 天线离目标有效高度

当目标与发射机的海拔一致的时候，天线的有效高度为发射塔高度，如下图所示：



但实际情况中，特别是在西部山区，发射机和目标的海拔不一样。天线离，目标有效高度修正为：

$$h_b = H_c + A_c - A$$

其中 $H_c$ 为发射机高度； $A_c$ 为发射机的海拔； $A$ 为目标的海拔。

天线离目标越高，信号越弱。

#### 5.5 目标与发射机水平距离

衡量信号接收效率指标之一，一般称负相关关系。计算公式如下：

$$d = \frac{5}{1000} \sqrt{(X_c - x)^2 + (Y_c - y)^2}$$

其中 $(X_c, Y_c)$ 为发射机栅格位置， $(x, y)$ 为目标栅格位置， $d$ 的单位为km。

特别注意的时，有时候目标与发射机的栅格位置相同，即 $d$ 的值为0。这种情况会干扰到后面特征的计算，对于这种情况，假设目标的期望位置为栅格的中心，而发射机为栅格的左上方，此时可以重新计算出为 $5\sqrt{2}/2$ 米。

#### 5.6 目标栅格与信号线相对高度

衡量信号接收效率指标之一，目标栅格与信号线相对高度越高，说明信号越弱。计算公式如下：

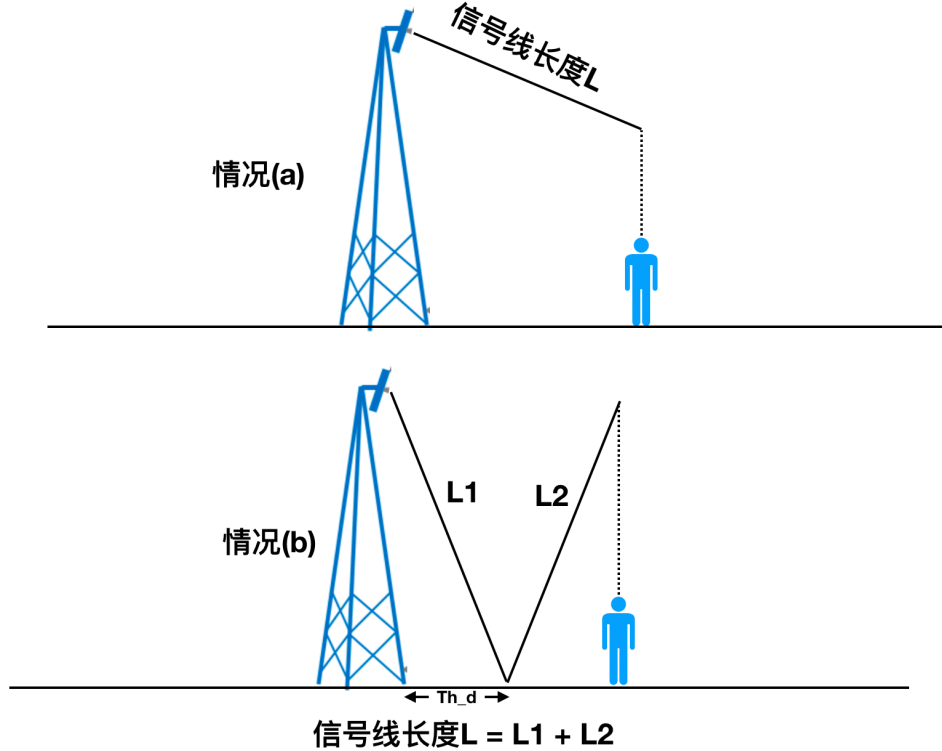
$$\Delta h_v = h_b - 1000d \cdot \tan(\theta_{ED} + \theta_{MD})$$

其中 $h_b$ 为天线离目标有效高度； $d$ 为目标与发射机水平距离； $\theta_{ED}$ 为发射机垂直电下倾角； $\theta_{MD}$ 为发射机垂直机械下倾角。

### 5.7 信号线长度

信号线越长，受到环境干扰的概率越大，传播途中遇到被高层建筑挡住的情况越多，最后在到用户身上的信号越弱。信号线需要考虑到两张情况，第一种情况下图情况(a)，信号线直接穿过目标的上方。其计算公式如下：

$$L = 1000d\sqrt{1 + (\tan(\theta_{ED} + \theta_{MD}))^2}$$



另外一种情况则需要通过反射才能穿过目标上方。由于反射具有对称性，此时信号线长度为：

$$Th_d = h_b / \tan(\theta_{ED} + \theta_{MD})$$

$$L_1 = \sqrt{h_b^2 + Th_d^2}$$

$$L_2 = \sqrt{\Delta h_v^2 + (d - Th_d)^2}$$

$$L = L_1 + L_2$$

### 5.8 信号线是否未到达目标前反射地面

根据 5.7 可知，可以根据  $Th_d$  是否小于  $d$  判断信号线是否需要反射才能判断目标上方。一般而言，信号通过反射后其传播效率会很降低。实际上也可以通过目标栅格与信号线相对高度  $\Delta h_v$  的正负来判断。

### 5.9 目标栅格离天线距离

与信号线长度类似，多维度反映目标与发射机的空间距离。

$$S = \sqrt{(1000d)^2 + h_b^2}$$

### 5.10 传播路径损耗

根据城市中的经典模型 Cost 231-Hata，传播路径损耗 PL 的计算公式如下：

$$PL = 46.3 + 33.9 \ln f - 13.82 \ln h_b - \alpha + (44.9 - 6.55 \ln h_b) \ln d + C_m$$

其中  $f$  为射机中心频率(MHz)；

$h_b$  为天线离目标有效高度(m)；

$\alpha$  为用户天线高度纠正项(dB)；

$d$  为目标与发射机水平距离(km)；

$C_m$  为场景纠正常数(dB)。

由于  $\alpha$  与  $h_b$  的关系式和  $C_m$  参数未知，因为把这两项忽视掉，得到一个有“偏差”的传播路径损耗模型：

$$PL = 46.3 + 33.9 \ln f - 13.82 \ln h_b + (44.9 - 6.55 \ln h_b) \ln d$$

$PL$  作为一个新的特征，可以继而该简化模型下的 RSRP。

### 5.11 理论下的 RSRP

根据题目知道，RSRP 与 PL 的关系式为：

$$RSRP = P - PL$$

其中  $P$  为射机发射功率。简化模型下的 RSRP 可以作为实测 RSRP 的一个参考。

### 5.12 发射机密度

仔细观察，可以看到有些栅格不止一台发射机，如果从供需关系来看，发射机就是供应。供应越充足，则用户平均分配到的资源越多。另外，不同发射机的信号线过于密集也有互相影响。定义以目标为中心方圆 5 公里范围内发射机数量为发射机密度。

5.13 发射机服务目标数量

同一个发射机可能同时给多个目标传递信号，目标越多，平均接受到的资源越少。举个例子，在节假日特别火爆的景区会可能因为旅客太多而把信号“挤”没了。发射机服务目标数量可以通过以小区唯一标示分组统计各个小区的样本数量的出。

5.14 目标竞争强度

发射机同时服务的目标数量与目标接受到的信号效率成负相关。同时，一般地目标离信号线越近，首先分配到的资源越好，即 RSRP 与目标栅格与信号线相对高度 $\Delta h_v$ 负相关。定义目标竞争强度公式如下：

$$S_t = \frac{1}{N_t \cdot \Delta h_v}$$

其中， $N_t$ 为发射机同时服务的目标数量；  
 $\Delta h_v$ 为目标栅格与信号线相对高度。

六、特征与目标的相关性及标准化

6.1 相关性分析

除了上面列举出新生成的特征除外，还以原数据中部分字段作为输入特征，对 RSRP 进行相关分析，相关系数以皮尔逊相关系数( Pearson correlation coefficient)作为指标，其计算公式如下：

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

其中 $cov(X,Y)$ 为特征(X)与 RSRP(Y)的协方差， $\sigma_X$ 、 $\sigma_Y$ 分别为特征和 RSRP 的标准差。对各个特征的相关系数的绝对值进行排序，得到如下结果：

排序	特征名称	与目标相关性
1	是否弱覆盖	-0.644904192
2	目标与发射机水平距离	-0.173958419
3	目标栅格离天线距离	-0.173958419
4	信号线长度	-0.168816465
5	传播路径损耗	-0.162879332



6	理论下的 RSRP	0.161400928
7	目标与信号线相对高度	-0.143847816
8	目标上方信号线离天线水平距离	-0.143825851
9	修正目标与信号线相对高度	-0.143220892
10	发射机密度	0.117068974
11	发射机同时服务目标数量	0.073693587
12	地物类型编号是否为 6	0.05790112
13	小区发射机中心频率	-0.039814888
14	目标位置建筑物高度	-0.037029989
15	地物类型编号是否为 5	-0.035060859
16	小区发射机发射功率	-0.03467534
17	地物类型编号是否为 15	-0.034245691
18	小区站点所在栅格建筑高度	-0.030252188
19	目标所在栅格海拔	0.029392925
20	地物类型编号是否为 14	-0.024639135
21	小区所在栅格海拔	0.022527131
22	小区发射机相对地面的高度	0.018499911
23	地物类型编号是否为 10	0.018048143
24	地物类型编号是否为 13	-0.013888373
25	天线离目标有效高度	0.011798498
26	地物类型编号是否为 11	0.009837749
27	地物类型编号是否为 7	-0.009710903
28	地物类型编号是否为 17	-0.008709736
29	地物类型编号是否为 2	0.008337622
30	地物类型编号是否为 16	-0.006946319
31	地物类型编号是否为 18	0.005506292
32	地物类型编号是否为 8	0.004465046
33	目标竞争强度	0.003655378
34	地物类型编号是否为 12	-0.002152588

由于后面的特征相关系数过低，为了节省计算资源和防止过拟合，只前 25 个特征作为输入，即选择相关系数的绝对值大于 0.01 的特征。在特征分析时，可以知道特征是否弱覆盖与 RSRP 最相关，在训练集可以直接通过 RSRP 进行

离散化获得，但是在测试集中，并不知道 RSRP。因此先要构建一个能够分类无线传播是否弱覆盖的模型。在测试集的数据中，首先预测是否弱覆盖，让预测值也作为一个新特征再进行预测 RSRP 的值。

### 6.2 数据标准化

由于各特征之间的量纲不同，为了防止计算精度偏差问题，一般先进行标准化处理。而且在后面用到神经网络模型，标准化后的数据能保证网络良好的收敛。标准化公式为：

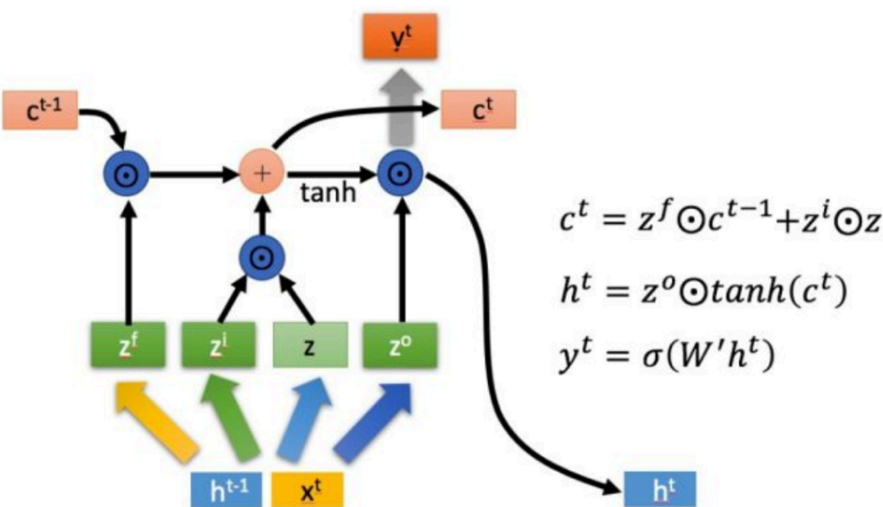
$$\hat{X} = \frac{X - \mu}{\sigma_X}$$

上式中， $X$ 为原特征数据， $\mu$ 为特征数据均值， $\sigma_X$ 为标准差； $\hat{X}$ 为标准化后的数据。

## 七、基于 LSTM 的无线传播是否弱覆盖分类模型

### 7.1 模型说明

LSTM(长短期记忆网络)本质是是 RNN(递归神经网络)，在传统的 RNN 中在梯度反向传播阶段，与神经元之间的权重矩阵可能由于时步数而相乘了很多次，如果权重矩阵特征比较小时，那么会出现梯度消失；相反，如果权重过大，则会出现梯度爆炸。LSTM 引入一种具有四个主要元素的存储单元结构，四个元素分别为：输入门、自循环连、忘记门、输出门，其基本结构如下：

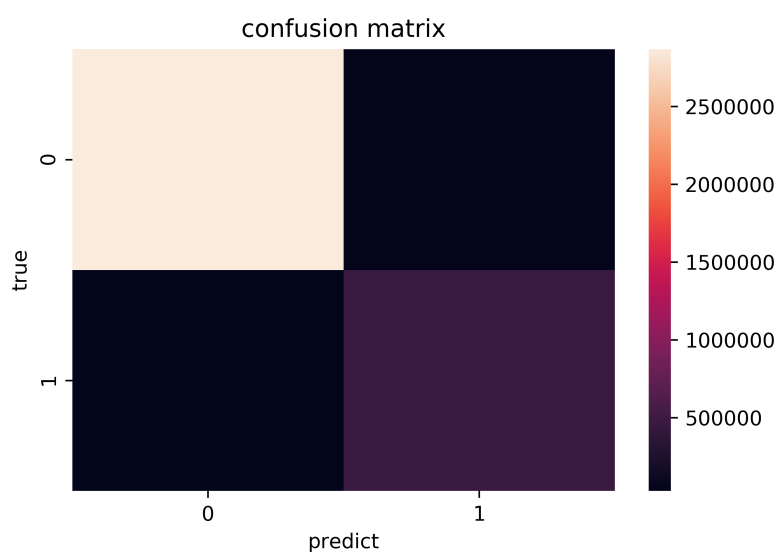


LSTM 由于其设计特点，在预测时间序列上有着显著的优势。除此外，LSTM 早已应用到各个方面上，如情感分析、图像识别、文本分析、回归预测等都有着不错的效果。本文先基于 tensorflow 架构构造一个经典的 LSTM 对无线传播是否弱覆盖进行分类预测，输入特征分别为本位第六章特征相关分析表中序号为第 2 至 25 这 24 个特征。LSTM 模型参数如下：

- 时步数 time\_step：3
- 学习率 lr：0.001
- 损失函数 loss：交叉熵( Cross Entropy Loss)
- 优化器：AdamOptimizer
- 最大迭代次数：500

## 7.2 模型评估

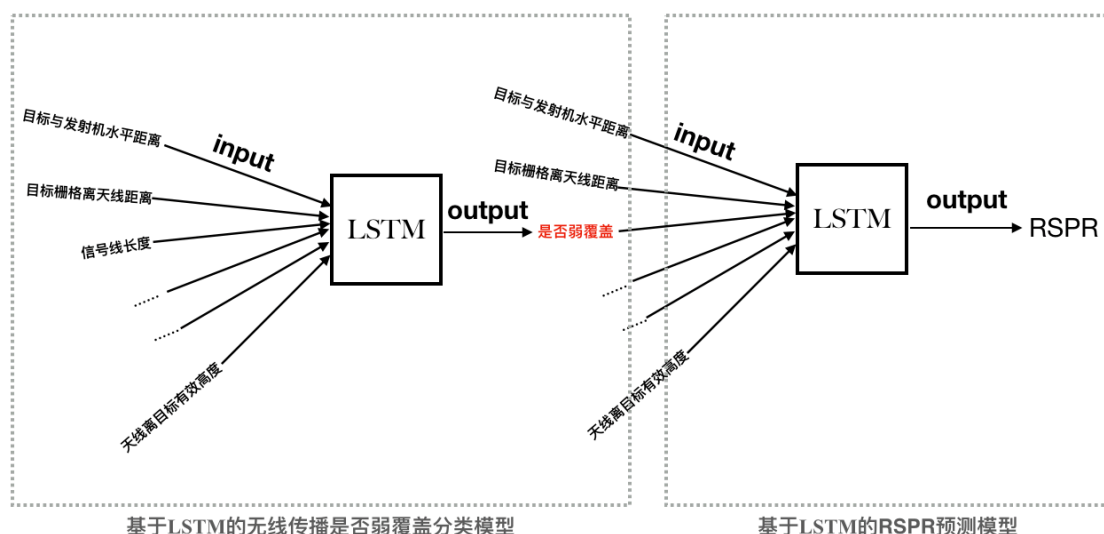
随机以 7:3 比例划分训练集和测试集，训练完 LSTM 后对测试集进行评估，当概率阈值取为 0.5 时，分类准确率为 94.43%。下图是为测试集的混淆矩阵：



## 八、基于 LSTM 的 RSPR 预测模型

### 8.1 模型说明

该模型是建立于本文第七章--基于 LSTM 的无线传播是否弱覆盖分类模型之上。先用一个分类 LSTM 模型预测出信号是否弱覆盖，把预测值也作为一个新特征再训练一个回归 LSTM 进行预测 RSPR 的值。具体结构流程如下所示：



模型参数跟之前的 LSTM 不太一样，因为之前的模型主要用于分类，而且输入特征的形状不一样，稍作调整部分参数和训练目标，参数如下：

- 时步数 time\_step：20
- 学习率 lr：0.001
- 损失函数 loss：均方根误差（RMSE：Root mean squared error）
- 优化器：AdamOptimizer
- 最大迭代次数：500

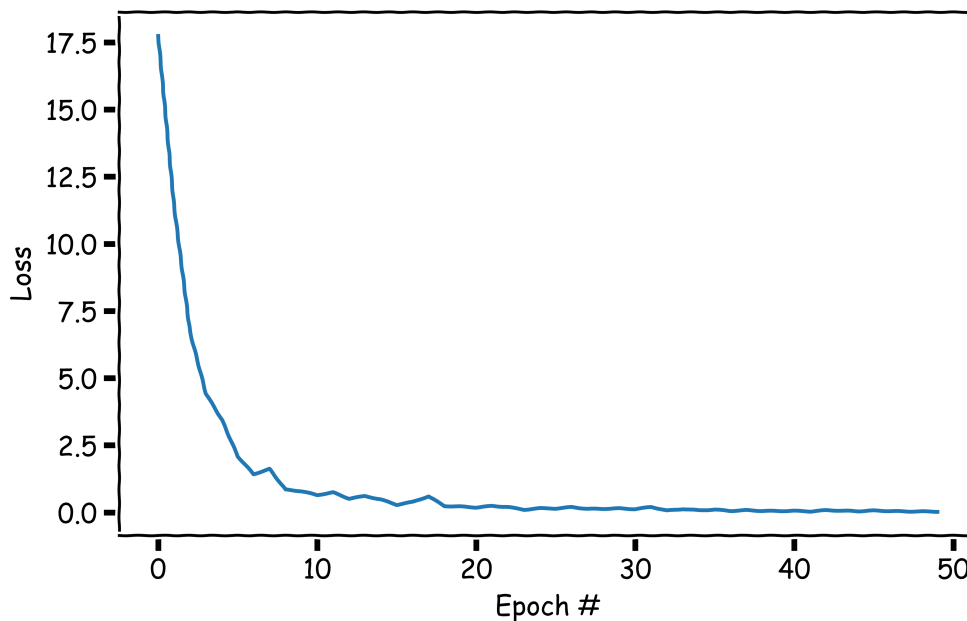
由于输入特征和目标都是标准化后的数据，因此模型预测的结果需要还原未标准化后的数据才能进行跟实测数据进行对比。逆标准化公式为：

$$Y = Y_{pred} \cdot \sigma + \mu$$

其中 $Y_{pred}$ 为模型预测值， $\sigma$ 为训练集 RSRP 的标准差， $\mu$ 为训练集 RSRP 的均值， $Y$ 逆标准化后的值。

## 8.2 模型评估

根据模型的损失函数记录来看，模型迭代 10 多次就开始收敛，未来节省计算资源和防止过拟合，模型迭代 15 次时停止，训练集线下 RMSE 的值最后在 0.67，线上测试集为 15.6。



除了均方根误差外，还需要计算弱覆盖识别率 (PCRR : Poor coverage recognition rate)，通过计算，训练集线下模型为 0.46，符合规定。

## 九、模型评价

### 9.1 模型优点

从数据定义上剔去了异常数据数据，借助 $3\sigma$ 准则剔除了 RSRP 极端数据，提高数据质量，为后面构建模型作铺垫；

通过无线传播伦理构造的特征有理论支撑，通过实测数据生成的特征也符合实际逻辑；

建立了两个 LSTM 模型，先对无线传播是否弱覆盖进行预测，能生成一个与 RSRP 相关性最高的一个特征，再进去预测 RSRP 的值。而且生成是否弱覆盖特征模型具有较高的准确率，在后面构建预测 RSRP 的模型具有参考意义。

### 9.2 缺点

与线上结果相比，训练模型出现过拟合情况。

## 参考文献

[1]Chhaya D. Tuning Of Cost-231 hata Model for Radio Wave Propagation Predictions[J]. 2013.

- [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. 2010.
- [3]Martinez Z O N, Rodriguez C, Arias O M. Propagation characteristics of Managua city based on Standard Propagation Model (SPM) at 850 MHz for 3G-WCDMA systems[C]// Central America & Panama Convention. 2015.
- [4]顾杰. 无线信道的确定性建模和参数分析[D].南京理工大学,2010.
- [5]魏畅. 无线信道的确定性建模与分析[D].南京邮电大学,2013.
- [6]程文强. 基于LTE网络传播模型优化及定位研究[D].武汉大学,2017.
- [7]李文娟. 大规模多天线信道测量及小尺度衰落特性研究[D].北京交通大学,2016.
- [8]唐越. 无线信道测量与建模方法研究[D].电子科技大学,2017.
- [9]胡金霞. 基于毫米波无线信道建模技术研究与应用[D].北京邮电大学,2019.
- [10]杨俊飞. 无线信道建模技术的研究与应用[D].浙江工业大学,2013.
- [11]黄俊然. 无线信道测量与建模中实测数据拟合的研究[D].天津大学,2010.
- [12]李忻. 新一代无线通信系统中的MIMO信道建模与多天线设计研究[D].电子科技大学,2005.