

# 基于数据挖掘的 RPMA 低功耗广域网网络规划方法

朱晓荣, 沈瑶

(南京邮电大学江苏省无线通信重点实验室, 江苏 南京 210003)

**摘要:** 针对 RPMA 低功耗广域网基站密度大、业务分布不均匀等特点, 提出了一种基于数据挖掘的网络规划方法。首先, 利用提升回归树算法建立了信号质量预测模型, 用于提取网络的覆盖分布空间模式; 然后, 针对覆盖分布空间模式, 采用加权 k-centroids 分簇算法得到适应当前模式的最优基站部署; 最后, 根据总目标函数判定得到最终的基站拓扑。通过真实数据集的仿真实验结果表明, 与传统的网络规划方法相比, 所提的方法很好地提升低功耗广域网网络的覆盖质量。

**关键词:** 低功耗广域网; 提升回归树; 加权 k-centroids; 基站部署

**中图分类号:** TN915.81

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-436x.2019050

## RPMA low-power wide-area network planning method based on data mining

ZHU Xiaorong, SHEN Yao

Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

**Abstract:** A network planning method based on data mining was proposed for RPMA low-power wide-area network with large density of base stations and uneven traffic distribution. First, a signal quality prediction model was established by using the boosting regression trees algorithm, which was used to extract the coverage distribution spacial pattern of the network. Then, the weighted k-centroids clustering algorithm was utilized to obtain the optimal base station deployment for the current spacial pattern. Finally, according to the total objective function, the best base station topology was determined. Experiment results with the real data sets show that compared with the traditional network planning method, the proposed method can improve the coverage of low-power wide-area networks.

**Key words:** low power wide area network, boosting regression trees, weighted k-centroids, base station deployment

### 1 引言

随着物联网的快速发展, 联网设备的数量将有望增长到 500 亿, 并且业务量将增加一千倍以上<sup>[1]</sup>, 传统的短距离无线技术和蜂窝网技术已经无法满足多样化的物联网业务需求, 因此, 新的通信模式——低功耗广域网 (LPWAN, low power wide area

network)<sup>[2]</sup>应运而生。LPWAN 主要满足大连接、低速率的物联网业务, 具有覆盖范围大、连接成本低、功耗低等特点, 主要包括 NB-IoT (narrow band internet of things)、LORA (long range)、RPMA (random phase multiple access) 等无线通信技术, 支持超大规模数量的设备接入网络。

然而, 在 RPMA 等 LPWAN 中基站密度大, 覆

收稿日期: 2018-04-23; 修回日期: 2019-01-16

基金项目: 江苏省研究生科研实践创新计划基金资助项目 (No.KYCX17\_0766); 国家自然科学基金资助项目 (No.61871237); 江苏省高校自然科学基金研究重大项目基金资助项目 (No.16KJA510005)

**Foundation Items:** The Post Graduate Research & Practice Innovation Program of Jiangsu Province (No.KYCX17\_0766), The National Natural Science Foundation of China (No.61871237), The Natural Science Foundation of the Higher Education Institutions of Jiangsu (No.16KJA510005)

盖距离达到 2~3 km，业务分布不均匀<sup>[3]</sup>，导致基站部署难度大，因此，面对 LPWAN 网络规划带来的挑战，需要针对其自身的特点，对网络进行合理的部署与优化，从而提高网络服务质量。在 LPWAN 的网络规划中，基站部署决定了网络的整体性能，过密的部署会给基站带来很大的干扰；过疏的部署会影响边缘区域的覆盖质量或者造成覆盖盲区<sup>[4]</sup>。此外，基站的选址也是影响网络质量的关键因素之一，不合理的选址会导致部分区域的信号覆盖质量差、容量不足等问题，造成网络运营困难，增加网络建设的成本<sup>[5]</sup>。由此可见，合适的基站数量和站址规划在网络部署中起到了重要作用。然而，基站站址的确定属于 NP-hard 问题<sup>[6]</sup>，如果采用传统的选址模型来分析站址问题的各种因素，就会导致所建模型中的变量与约束条件的维度灾难，此选址方法并不科学。网络规划不仅需要考虑覆盖，还要考虑业务分布，需要处理和整合时空特性<sup>[7]</sup>，这使网络规划问题更加复杂，需要设计出合理的网络规划方案。

目前，国内外对网络规划已经做了大量的研究。文献[5]针对异构网络环境下，研究已安排预算的基站规划问题，其目标是在给定预算的条件下，最大化业务需求点的数量，同时也要满足业务需求点的速率需求。文献[8]认为 LTE (long term evolution) 基站规划的优化任务是确定基站的数量和位置，基于这 2 个目标，结合规划过程的 2 个重要约束——区域覆盖约束和基站容量约束，提出了一种最优的 LTE 无线规划方法。文献[9]认为基站规划策略不应该只关注如何减少基站数目，能效也是一个重要的指标，所以考虑以最小的能量消耗为目标来获取最优的基站数目和基站站址。文献[10]认为蜂窝网络的核心目标是能够保证用户的服务质量 (QoS, quality of service) 和无缝覆盖的吞吐量，基于这 2 个目标，采用区域划分技术解决异构网络中的基站规划问题，提出以负载均衡的方式部署基站，所提方案不仅能实现部署总成本达到最低（需要部署的基站数目最少），还能获得更好的网络性能。文献[11]联合所提出的基站定位算法和无线资源控制算法自动地规划蜂窝网基站位置，实现以最少的基站数目提供必要的覆盖和容量。

上述文献主要针对蜂窝网络进行规划，对 LPWAN 的技术进行综述，并未提出合理的规划方案。另外，很多研究者都把基站规划问题当作优化

问题进行处理，针对不同的研究场景，提出优化目标和约束，再采用合适的算法进行解决。然而，这些文献所提出的网络规划方法是以大量的假设为前提，算法模型有一定的局限性，没有从本质上提出有效的规划方法来快速地规划和部署大量的基站。

针对上述问题，本文将大数据的分析方法和网络规划相结合，以低功耗广域网在通信系统中的应用为背景，利用获得的网络数据进行性能分析，将基站选址问题由传统的模型驱动转变为数据驱动，以海量的数据为分析主线，克服传统网络规划模型的求解缺点，并结合聚类算法探索以数据驱动的基站选址方法，从而提升站点选择的合理化水平。

## 2 系统模型

本文研究场景如图 1 所示。RPMA 网络属于典型的星形拓扑结构，多个终端以无线的方式连接到邻近的 RPMA 网关，由网关负责接收来自终端的上行链路数据，并将数据聚集到各自的回程连接，实现多路数据的收集和转发。网络服务器和网关之间通过 4G/5G/以太网回传建立通信链路，其中，网络服务器主要负责介质访问 (MAC, media access control) 层处理，包括网关的管理和选择、重复数据分组的消除、进程的确认等。

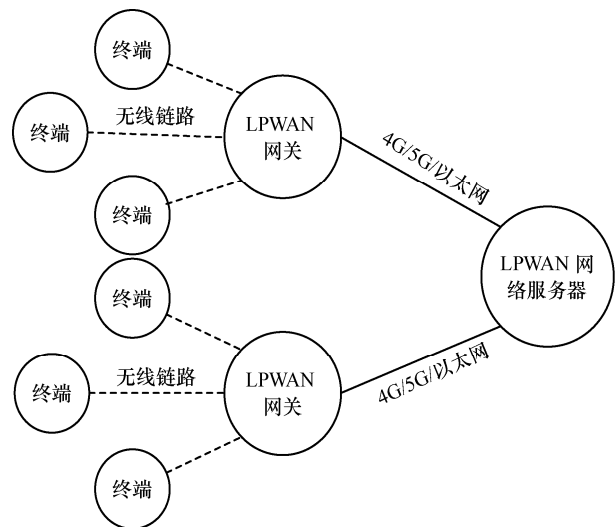


图 1 RPMA 星形网络拓扑结构

针对 RPMA 网络特点，本文提出了一种基于数据挖掘的网络规划方法，系统框图如图 2 所示。首先，考虑网络规划的覆盖目标，采集 RPMA 网络的实测数据，主要包括基站基本信息数据、终端测试

点数据、地理位置数据等。基于网络规划知识数据库，对实测数据进行初步清洗与分析，去除具有大量重复和缺省值的属性，并且通过分析提取影响信号覆盖质量的相关特征，将其输入到学习模型中进行训练，从而获得最终的基站部署，其中，学习模型分为预测模型和规划模型，如图 2 所示。本文中，基站站址的选定是根据每一次基站部署好后网络的空间覆盖情况进行调整，由于每一次基站调整，会使整个网络的信号覆盖情况发生变化，因此，需要针对网络信号覆盖的变化，利用相应的数据学习得到预测模型，对每一次的网络拓扑下的覆盖情况进行预测。而规划模型则是根据预测所得的覆盖情况，确定合适的基站部署。整个规划的总目标是通过每一次基站调整逐渐缩小信号覆盖差的区域，并且使区域中信号覆盖质量接近所要求的标准。

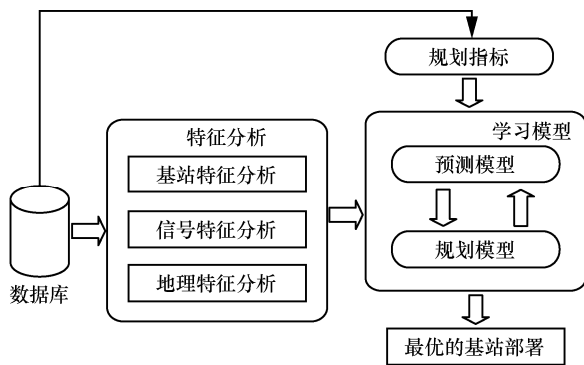


图 2 基于数据挖掘的网络规划系统框架

### 3 信号质量预测模型

如图 2 所示，本文首先从覆盖目标着手，针对无线网络中弱覆盖问题进行分析，重点对覆盖盲区和弱覆盖区域进行优化，并且根据网络覆盖情况进行基站位置的调整，使调整后的基站能满足所要求的覆盖效果。一般来说，区域弱覆盖主要是因为接收到的信号强度不足造成的，具体影响因素涉及以下三方面：1) 基站侧影响覆盖的因素，比如发射功率、天线方位角、天线挂高、天线增益等；2) 信号传输路径影响覆盖的因素，比如由于障碍物遮挡造成的路径损耗、阴影衰落等；3) 干扰对覆盖的影响，比如多个相邻基站重叠覆盖区内产生的同频干扰，以及建筑、山峦等地表物对电波反射造成的多径干扰等。

综合上述分析，可以知道网络中某个位置的终

端接收信号质量，基本和上述三方面的因素相关，是揉合了这些因素之后的结果，因此，本文考虑先得到信号质量和这些因素之间的映射关系，即进行信号质量预测，用于辅助基站的站址确定。

信号质量预测问题在机器学习任务是属于回归问题，即运用机器学习算法可以训练得到特定的某个函数将所输入的一系列变量映射为一个连续输出值，因此，可以先通过现有的路测数据来构建出适应当前无线网络环境的数据模型。当需要对新规划方案的覆盖效果进行预测时，只要给出对应的无线网络特征，就可以预测出符合新规划方案的信号覆盖情况，然后根据预测结果，进行基站站址的进一步调整。

#### 3.1 数据特征选取

信号质量预测是基于当前部署网络所获取的基站侧数据和测试点侧数据，如表 1 和表 2 所示。

表 1 基站侧属性	
属性名	说明
B_lat	基站经度
B_lon	基站纬度
B_alt	基站高度
B_power	发射功率
A_height	天线挂高
B_name	基站名
deploy region	部署地区号
last connect time	最后连接时间
last connect address	最后连接地址
UL PER(UL PDU error rate)	上行链路 PDU 误码率
UL SNR	上行链路信噪比
UL SF	上行链路扩频因子
UL signaled excess SNR	上行信号的信噪比盈余
UL target SNR error	上行信号信噪比与预定值的误差
network state	网络状态

表 2 测试点侧属性	
属性名	说明
DL RSSI	下行链路接收信号强度指示器
DL SF	下行链路传输一个 SDU 时的平均扩频因子
P_lat	测试点经度
P_lon	测试点纬度

首先，需要对数据进行初步清洗与分析，去除具有大量重复和缺省值的属性，如 UL PER、network

state 等。对于 deploy region 属性，本文认为经纬度已经代表了基站的位置差异特性，该属性可以剔除。另外，结合上述所分析的三方面区域覆盖影响因素，剔除 last connect time、last connect address 等无关属性，最终选定基站位置 B\_loc（包括经纬度）、基站高度 B\_alt、基站功率 B\_power、天线挂高 A\_height 和终端位置 P\_loc（包括经纬度）作为输入特征，将这些输入特征整合为一条记录，如式(1)所示。

$$x_k = \{B\_loc_k, B\_alt_k, B\_power_k, A\_height_k, P\_loc_k\} \quad (1)$$

这些记录的集合作为信号质量预测模型的训练数据集。由于 RPMA 网络采用功率控制，收到的上行信号强度总是在接收灵敏度附近，所以选定终端接收的下行 RSSI (received signal strength indication) 作为衡量信号质量的指标，即输出变量。建立无线网络数据模型的过程，就是通过训练已有的数据集找到两者之间的映射函数  $f$  的过程，如式(2)所示。

$$y_k = f(x_k) \quad (2)$$

其中， $y_k$  的值就是给定输入特征变量  $x_k$  下的预测信号质量值。

### 3.2 信号预测模型建立

本文采用提升回归树算法<sup>[12]</sup>来构建式(2)所示的函数映射关系。提升回归树 (BRT, boosting regression tree) 算法属于集成学习方法中的一种，通过集成多个基学习器共同完成学习任务。相比于单一的回归算法，如线性回归、对数几率回归算法等，BRT 算法以组合多个决策树的方式，能够获取更加优越的泛化能力，从而提升了模型的预测精度。BRT 模型可以用  $M$  棵决策树的加法模型，如式(3)所示。

$$f_M(x) = \sum_{m=1}^M T(x; \gamma_m) \quad (3)$$

其中，每棵树表示为

$$T(x; \gamma) = \sum_{j=1}^J c_j I(x \in R_j) \quad (4)$$

其中，参数  $\gamma = \{(R_1, c_1), (R_2, c_2), \dots, (R_J, c_J)\}$  代表每棵树在输入变量集上的区域划分  $R_1, R_2, \dots, R_J$  和在相应区域上的常数  $c_1, c_2, \dots, c_J$ ， $J$  是决策树的叶子节点数。

BRT 采用前向分步算法，按照从前向后的顺序学习每一棵决策树，即通过优化如式(5)所示的损失

函数学习每棵树的参数。

$$\hat{\gamma} = \arg \min_{\gamma_m} \sum_{i=1}^N L(y_i, f_{m-1}(x_i) + T(x_i; \gamma_m)) \quad (5)$$

式(5)中的损失函数采用平方误差，即样本的预测值和实际值差的平方和，如式(6)所示。

$$L(y_i, f_{m-1}(x_i) + T(x_i; \gamma_m)) = (y_i - f_{m-1}(x_i) - T(x_i; \gamma_m))^2 \quad (6)$$

而  $e_m = y_i - f_{m-1}(x_i)$  是利用当前模型  $f_{m-1}(x_i)$  拟合数据时导致的残差，所以利用 BRT 算法解决回归问题时，只需要对每次模型拟合的残差进行拟合。具体算法过程如算法 1 所示。

**算法 1** 基于最小平方误差代价函数的提升回归树算法

输入 训练数据集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\},$$

$$x_i \in X \subseteq R^n, y_i \in Y \subseteq R$$

输出 提升树  $f_M(x)$

1) 初始化  $f_0(x) = 0$

2) for  $m = 1$  to  $M$  do

①  $e_m = y_i - f_{m-1}(x_i)$ ,  $i = 1, \dots, N$

②  $\gamma_m = \arg \min_{\gamma_m} \sum_{i=1}^N L(y_i, f_{m-1}(x_i) + T(x_i; \gamma_m))$

③ update  $f_m(x) = f_{m-1}(x) + T(x; \gamma_m)$

end for

3) 得到回归问题提升树

$$f_M(x) = \sum_{m=1}^M T(x; \gamma_m)$$

### 3.3 预测变量的重要性度量

本文所要研究的信号预测模型，不仅要求精确地预测出信号质量，还希望能够通过模型了解哪些变量是影响信号覆盖质量的主要因素。

在训练单棵决策树时，输入的变量对响应变量的影响程度不同，用  $J_n(T)$  表示第  $n$  个输入变量  $X_n$  对于响应变量的相关性度量<sup>[12]</sup>，如式(7)所示。

$$J_n(T) = \sum_{i=1}^{L-1} \hat{j}_n^2 I(v_i = n) \quad (7)$$

其中， $T$  表示具有  $L$  个叶节点的决策树； $\hat{j}_n^2 I(v_i = n)$  表示当前决策树在第  $i$  个非叶节点上以  $X_n$  作为分裂变量时，对应的平方误差风险改进。输入变量  $X_n$  对于整个 BRT 模型的相关性的平方  $J_n^2$ ，即该变量在上述 BRT 模型上使所有决策树预测损失改进的

平均, 如式(8)所示。

$$J_n^2 = \frac{1}{M} \sum_{m=1}^M J_n^2(T_m) \quad (8)$$

当比较各个输入变量对预测结果的影响力时, 一般先将式(8)结果进行归一化处理, 即令所有输入变量对结果的相对影响力之和为 1, 以百分数的形式来表示每个变量的重要性。

#### 4 基站位置确定

典型的  $K$  均值聚类算法是将数据集  $X=\{x_1, \dots, x_n\}$  中的  $n$  点进行划分, 把原来独立的  $n$  个点通过设定距离相似度划分进  $K$  个簇当中, 簇集合  $C=\{c_1, \dots, c_k\}$ 。一般会以两点之间的欧式距离作为相似性度量, 把数据点划分进距离最近的簇中心所在的簇中。算法一般是以最小化簇内位置误差平方和 (SSE, sum of the squared error) 为目标函数, 如式(9)所示。

$$SSE = \sum_{j=1}^k \sum_{x_i \in c_j} \text{dist}(c_j, x_i)^2 \quad (9)$$

在经典的 K-means 算法中, 每个数据点对定位簇中心的位置有着同样的重要性。然而, 本文把基站位置的选择当作基于覆盖分布空间模式的加权问题进行处理, 即认为空间中的每一个点不再对簇中心有等价的影响, 引入权重来衡量数据点对基站位置的影响度, 从而提出加权 K-centroids 算法。

本算法的输入包括  $n$  个终端数据点位置集合  $P=\{p_1, \dots, p_n\}$  和初始的基站位置集合  $B=\{b_1, \dots, b_k\}$ 。由于是基于现有的基站站址进行规划, 所以可以把当前的基站位置和数量作为聚类的初始化参数, 即簇的初始化中心和聚簇数目。本文算法以归一化距离判定数据点归属于哪一个簇中, 即定义从属函数  $f(b_j|p_i)$ <sup>[13]</sup>, 如式(10)所示。

$$f(b_j|p_i) = \frac{\|p_i - b_j\|^2}{\sum_{j=1}^k \|p_i - b_j\|^2} \quad (10)$$

根据  $f(b_j|p_i)$  可以将数据点  $p_i$  分配进离其最近的基站位置  $b_j$  所在的簇中。

分配好数据点以后, 本算法需要对簇中心的位置即站址的位置进行迭代调节, 主要从距离影响度和覆盖权重两方面考虑。对于距离影响度来说, 由于基站要尽可能覆盖远处的终端, 对于离基站近的终端自然信号会相对好一些, 所以它离基站的距离

影响度要小, 而离基站远的终端可能会因为楼层建筑物的遮挡、信号传播的特点等因素导致终端接收到的信号衰减, 所以远处的终端离基站的距离影响度要大一些, 本文应用上述定义的从属函数  $f(b_j|p_i)$  来表示距离影响度。对于覆盖权重来说, 优化基站位置是希望在基站覆盖内终端的接收信号能尽可能好, 所以需要重点关注覆盖质量差的终端, 赋予它对基站位置调整更大的影响权重。根据上一阶段中得到的覆盖分布空间模式, 对于每个数据点产生相应的权重值  $w(p_i)$ 。结合  $f(b_j|p_i)$  和  $w(p_i)$ , 可得到簇中心位置的迭代计算式, 如式(11)所示。

$$b_j = \frac{\sum_{i=1}^n f(b_j|p_i)w(p_i)p_i}{\sum_{i=1}^n f(b_j|p_i)w(p_i)} \quad (11)$$

当  $b_j$  不再发生变化时候, 结束迭代, 得到最佳的簇中心位置即基站站址。加权 K-centroids 算法的步骤如算法 2 所示。

**算法 2** 加权 K-centroids 分簇算法决定基站位置

**输入**  $n$  个测试点集合  $P$  和候选基站站址  $L_{b_j}$  的数量

**输出**  $k$  个簇的中心位置

1) 以现有的基站位置和数量作为初始的簇中心位置和簇数量

2) repeat

① for  $i=1, 2, \dots, N$  do

根据从属函数将点  $p_i$  划分进离它最近的基站位置  $b_j$  所在的簇中, 有

$$f(b_j|p_i) = \frac{\|p_i - b_j\|^2}{\sum_{j=1}^k \|p_i - b_j\|^2}$$

end for

② for  $j=1, 2, \dots, k$  do

$$b_j = \frac{\sum_{i=1}^n f(b_j|p_i)w(p_i)p_i}{\sum_{i=1}^n f(b_j|p_i)w(p_i)}$$

其中,  $w(p_i)$  代表与覆盖空间模式相关的权重值

end for

3) until  $b_j$  不再发生变化

由加权 K-centroids 算法得到的网络拓扑, 已经针对当前网络覆盖情况进行了优化, 但是并不一定是最终的最优结果, 仍然需要对其进行覆盖预测分析, 根据分析结果再次进行基站位置优化, 直到满足下述的总目标函数, 得到最优的网络拓扑。

整个规划过程的总目标函数, 如式(12)所示。

$$\min \sum_{i \in \{i | y_i \leq \bar{y}\}} (y_i - \bar{y})^2 \quad (12)$$

其中,  $y_i$  值代表在某点利用 BRT 算法预测出的覆盖强弱值 RSSI;  $\bar{y}$  是理论上满足覆盖要求所要达到的信号良好覆盖阈值;  $i \in \{i | y_i \leq \bar{y}\}$  表示区域中信号质量值低于阈值的测试点, 即覆盖信号质量差的测试点。以两者的最小平方误差和作为整个规划过程迭代终止的目标函数。

5 仿真分析与性能评估

本实验数据来源于 37 个 RPMA 基站和经过数据清洗后的 131 454 条测试点的路测数据, 包含基站基本信息数据、终端测试点数据以及相应的地理位置数据, 用于验证本文所提出的网络规划方法的可行性, 并且采用 Python Matplotlib 工具将实验结果可视化。

5.1 信号预测模型结果与分析

在应用提升回归树算法之前, 需要确定 3 个参数来调整 BRT 算法的学习过程。首先是基学习器数量。虽然其数量的增加, 会提升 BRT 算法对训练数据的拟合效果, 但是基学习器的数量超过一定的值, 很有可能会造成过拟合, 导致得到的模型对未知数据的预测效果差。其次是基学习器的大小, 它表示了被 BRT 模型捕捉到的多个特征之间相互作用的程度, 选用树的深度来控制基学习器的大小。对于这 2 个参数的选取, 本文采用了 sk-learn 中的 GridSearchCV 网格追踪法, 它能够根据给定的数据集, 遍历需要优化的参数的多种取值组合, 通过交叉验证获得最佳效果的参数取值, 基学习器数量为 530, 树的深度为 11。最后, 为了防止对训练数据的过度拟合, 会引入正则化因子, 即学习率 LR (learning rate), 来衡量每个基学习器对最终结果的影响程度, 一般 LR 设置为一个低于 0.1 的较小常数, 本文中设定为 0.1。

给定好参数之后, 选取数据集的 85%作为训练

数据集, 15%为测试数据集。图 3 横轴表示迭代次数 (即基学习器数量), 纵轴表示损失误差值, 图中线条分别表示了每一次迭代的测试误差和训练误差。从图中可以发现, 随着迭代次数的增加, 训练误差和测试误差都逐渐减少, 训练误差的减少说明模型在训练数据集上的拟合效果随着迭代次数的增加逐渐提高。图中显示测试误差要高于训练误差, 是由于测试集和训练集存在一定的差异性, 使模型在未知数据集上的学习能力要弱于原训练数据集, 属于正常现象。另外, 2 条曲线的趋势也说明 GridSearchCV 得到的参数适当。

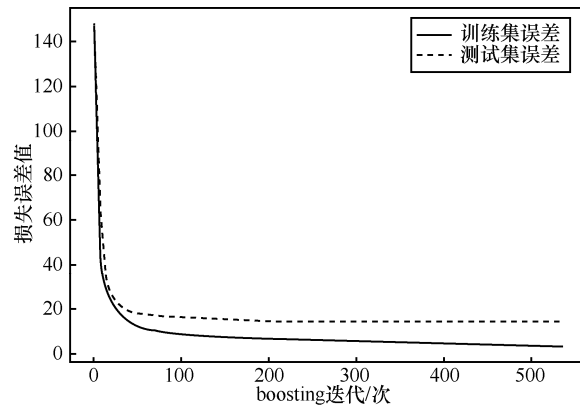


图 3 损失误差与基学习器数量的关系

图 4 表示了预测变量的相对重要性。分析认为测试点距离基站的距离 diff 与预测值 RSSI 的关系最为密切, 其相对重要性达到最高。其次是测试点位置和基站所处位置, 其中经度的相对重要性要小于纬度。天线挂高和基站高度的重要性稍低, 最低的是基站发射功率, 这是因为采集到的 RPMA 基站功率值只有 30 dBm、31 dBm 这 2 个值, 即其取值变化较小, 所以对预测值的影响作用不大。

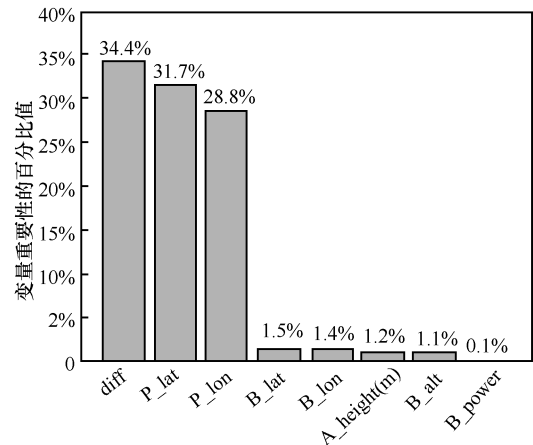


图 4 预测变量的相对重要性

### 5.2 最优基站部署确定

图 5 为所采集到的初始基站位置和测试点分布情况，以星形点标记基站的位置，圆点标记测试点的位置，其中圆点的颜色深浅代表 RSSI 值的大小，颜色越深代表 RSSI 值越低，信号覆盖情况越差，可见在初始的基站部署下，还是存在部分弱覆盖区域，图中 RSSI 单位为 dBm。

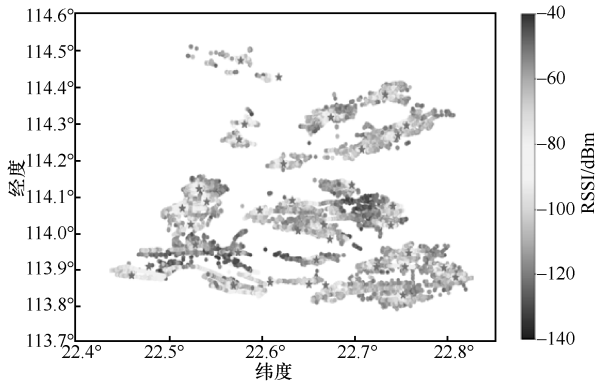


图 5 初始基站位置及周围测试点的 RSSI 值变化情况

本文中，采取基于覆盖分布空间模式来部署基站，以 RSSI 作为衡量覆盖强弱的权重值，结合加权 K-centroids 分簇算法，从而决定基站位置，并根据总目标函数值来判定当前基站部署是否达到最优。表 3 为每一轮规划迭代后，求得的总目标函数值，由表 3 可知，随着迭代次数的增多，总目标函数值逐渐减少，即信号覆盖情况逐渐得到改善，直到第 10 次迭代后终止，得到总目标函数的最小值为 535.41。图 6 为最终基站位置及周围测试点的 RSSI 值变化情况，可见深色区域相比于图 5 有所减少，即应用本文的规划方法后，信号覆盖情况得到改善。图 7 为对应的分簇结果，以黑色星形点标记基站的位置，圆点标记测试点的位置，同一颜色的点表示属于离其最近的基站簇中。

表 3 每轮迭代的总目标函数值

迭代次数/次	总目标函数值
1	640.93
2	581.79
3	554.43
4	548.89
5	548.66
6	547.08
7	546.87
8	539.41
9	535.54
10	535.41

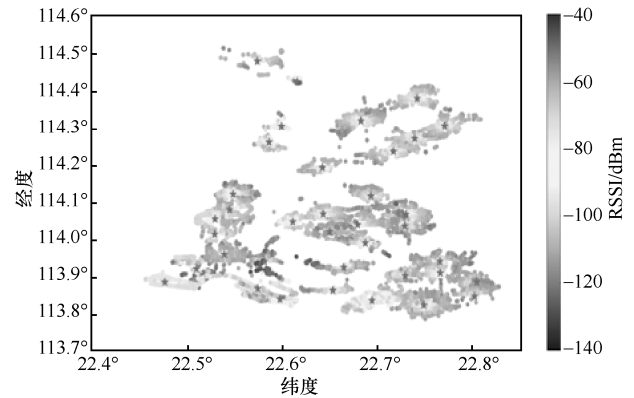


图 6 最终基站位置及周围测试点的 RSSI 值变化情况

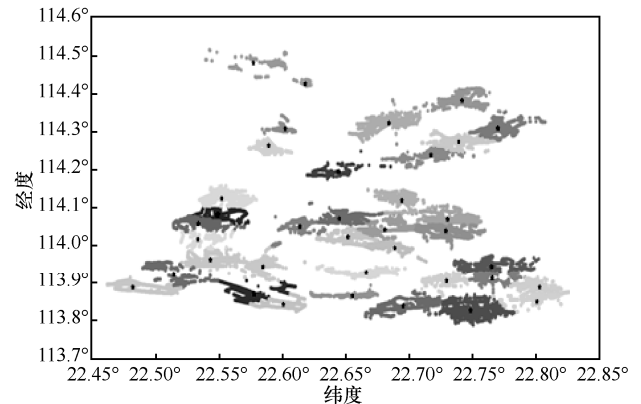


图 7 分簇结果

为了验证本文所提方法的优越性，将本文所提方法与基于 k-means 的优化方法<sup>[14]</sup>进行比较。利用同一组实测数据，使用基于 k-means 的优化方法进行基站位置调整，图 8 为应用该方法得到的基站位置及周围的测试点的 RSSI 值变化情况。相比于图 5，图 8 中深色区域减少，但与图 6 比较，深色区域依旧很多，可见应用 k-means 优化方法虽然能使信号覆盖情况有所改善，但与本文所提的方法相比，信号覆盖提升能力不足。同时，可以利用式(12)计算每一轮迭代后的总目标函数值，其迭代结果趋于 584.22，大于本文所提方法的目标函数最小值 535.41。因此，从信号覆盖率提升方面，本文所提的方法优于基于 k-means 的优化方法，能够更好地提升信号覆盖率。另外，从迭代次数上看，基于 k-means 的优化方法只考虑了位置距离来调节基站，需要迭代 16 次才能收敛到最小值，而本文方法在调节过程中引入了覆盖权重来协助调节，可以加速收敛，仅需要表 3 中所示的 10 次迭代，收敛速度上具有一定的优势。

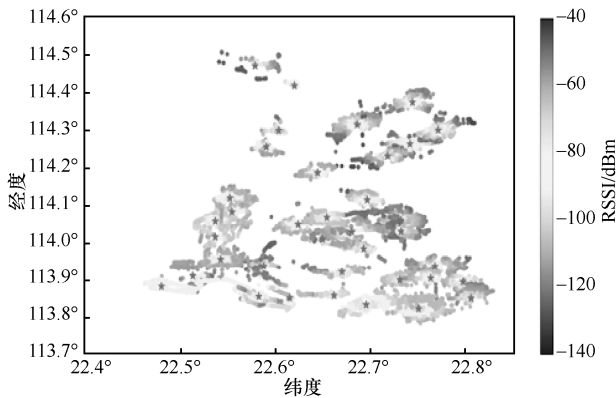


图8 应用基于 k-means 的优化方法得到的基站位置及周围测试点的 RSSI 值变化情况

## 6 结束语

针对 RPMA 网络特点, 本文提出了一种基于数据挖掘的网络规划方法。首先利用获取的实测数据对整体网络进行初步分析, 选取对覆盖质量的影响特征。然后, 采用 BRT 算法和 K-centroids 分簇算法对网络的覆盖分布空间模式进行提取, 并且获得最优的 RPMA 网络基站部署。最后, 利用实测数据验证了本文所提方法的可行性, 并与基于 k-means 的优化方法相比较。实验结果表明, 该方法能够很好地提升低功耗广域网网络的覆盖质量, 对网络规划具有一定的参考价值。

在实际网络规划中, 基站部署需要考虑多方面的因素, 而本文仅考虑了网络规划的覆盖目标, 因此, 下一步的工作将会引入容量目标, 结合两方面的目标进行基站的最优部署, 使得网络规划更加完善。

## 参考文献:

- [1] PATEL D, WON M. Experimental study on low power wide area networks (LPWAN) for mobile internet of things[C]// Vehicular Technology Conference. IEEE, 2017:1-5.
- [2] HERNANDEZ D M, PERALTA G, MANERO L, et al. Energy and coverage study of LPWAN schemes for Industry 4.0[C]// Electronics, Control, Measurement, Signals and their Application to Mechatronics. IEEE, 2017:1-6.
- [3] KRUPKA L, VOJTECH L, NERUDA M. The issue of LPWAN technology coexistence in IoT environment[C]//International Conference on Mechatronics-Mechatronika. IEEE, 2016:1-8.
- [4] YU G J, YEH K Y. A k-means based small cell deployment algorithm

for wireless access networks[C]//International Conference on Networking and Network Applications. IEEE, 2016:393-398.

- [5] WANG S, ZHAO W, WANG C. Budgeted cell planning for cellular networks with small cells[J]. IEEE Transactions on Vehicular Technology, 2015, 64(10):4797-4806.
- [6] AMALDI E, CAPONE A, MALUCELLI F. Planning UMTS base station location: optimization models with power control and algorithms[J]. IEEE Transactions on Wireless Communications, 2003, 2(5):939-952.
- [7] LEE C Y, KANG H G. Cell planning with capacity expansion in mobile communications: a tabu search approach[J]. IEEE Transactions on Vehicular Technology, 2000, 49(5):1678-1691.
- [8] GHAZZAI H, YAACOUB E, ALOUINI M S, et al. Optimized LTE cell planning with varying spatial and temporal user densities[J]. IEEE Transactions on Vehicular Technology, 2016, 65(3):1575-1589.
- [9] YANG Z H, CHEN M, WEN Y P, et al. Cell planning based on minimized power consumption for LTE networks[C]// IEEE Wireless Communications and Networking Conference. IEEE, 2016: 1-6.
- [10] WANG S, RAN C. Rethinking cellular network planning and optimization[J]. IEEE Wireless Communications, 2016, 23(2):118-125.
- [11] ISTV S, FAZEKAS P. An algorithm for automatic base station placement in cellular network deployment[C]// EUNICE/IFIP WG 6.6 Conference on Networked Services and Applications: Engineering, Control and Management. Springer-Verlag, 2010:21-30.
- [12] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. The Annals of Statistics, 2001, 29(5):1189-1232.
- [13] WEN R, YAN W, ZHANG A N. Weighted clustering of spatial pattern for optimal logistics hub deployment[C]//IEEE International Conference on Big Data. IEEE, 2016:3792-3797.
- [14] KANUNGO T, MOUNT D M, NETANYAHU N S, et al. An efficient k-means clustering algorithm: analysis and implementation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 40(7): 881-892.

## 作者简介



朱晓荣 (1977- ), 女, 山东临沂人, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为 5G 网络、异构网络、无线传感器网络等无线资源管理、跨层优化算法及协议设计、性能评估及建模分析等。



沈瑶 (1994- ), 女, 江苏常州人, 南京邮电大学硕士生, 主要研究方向为 5G 网络优化、无线大数据处理等。