

一种基于Wkmeans聚类的LTE外部干扰智能定位方法

An Intelligent Orientation Method of LTE External Interference Based on Wkmeans Clustering

赵伟¹,王勇²,陈乐¹(1. 中国联通浙江省分公司,浙江杭州325800;2. 中国联合网络通信集团有限公司,北京100033)
Zhao Wei¹,Wang Yong²,Chen Le¹(1. China Unicom Zhejiang Branch, Hangzhou 325800, China; 2. China United Network Communications Group Co.,Ltd., Beijing 100033, China)

摘要:

提出了一种基于上下行路损相似度来预测干扰源位置的方法,通过提取基站24 h的干扰时域特征,识别出干扰小区,将受相同干扰源影响的小区分组后生成干扰字典。一方面,计算每个干扰源的上行路损向量;另一方面,利用栅格化后的MDT数据,计算出每个栅格的下行路损向量。通过引入权重聚类算法,将上下行路损向量的相似度作为权重参数,进而定位可能的干扰源栅格位置。该方法已在某省LTE网络试点推行,不同场景干扰源定位的误差在50 m、150 m及200 m范围内不等。

Abstract:

It proposes a method based on the similarity of the uplink and downlink path loss to predict the location of the interference source. The 24-hour interference time domain characteristics of the base station are extracted to identify interfered cells, and the cells affected by the same interference source are grouped to generate an interference dictionary. On the one hand, the uplink path loss vector of each interference source is calculated, on the other hand, the downlink path loss vector of each grid is calculated by using the rasterized MDT data. By introducing a weighted clustering algorithm, the similarity of the uplink and downlink path loss vectors is used as the weight parameter to orientate the possible interference source grid location. This method has been implemented on a pilot program in one province, and the distance error of interference source location is within the range of 50 m, 150 m or 200 m depending on different scenarios.

Keywords:

External interference; Interference orientation; Weighted clustering

关键词:

外部干扰; 干扰定位; 权重聚类

doi:10.12045/j.issn.1007-3043.2021.05.016

文章编号:1007-3043(2021)05-0072-06

中图分类号:TN929.5

文献标识码:A

开放科学(资源服务)标识码(OSID):



引用格式:赵伟,王勇,陈乐.一种基于Wkmeans聚类的LTE外部干扰智能定位方法[J].邮电设计技术,2021(5):72-77.

1 LTE干扰排查现状及难点

在大流量背景下,移动用户不断增高的网络感知要求与网络覆盖不平衡的矛盾日益突出,进而因私装干扰器导致LTE外部干扰现象明显增多。现有干扰器主要以直放站、信号放大器、部分不达标的民用设备等等为主。干扰信号经由扩频后将引入上行干扰。

目前,针对LTE网络上行系统外部干扰定位的相关研究较少。现行排查方法以工程师现场扫频排查为主,存在以下4个难点。

a) 人工成本高。每排查一个外部干扰源,至少需2名经验丰富的一线工程师前往干扰现场,手持扫频仪器在受干扰基站覆盖范围内(大于500 m)进行地毯式排查。通常至少需要2次以上现场排查方能确定干扰源真实位置,且对扫频工程师的经验和专业水平要求高。

收稿日期:2021-04-12

b) 扫频仪器精度有限。当前常用的扫频仪器往往至多能对周围 50 m 范围进行扫频,且随着使用年限的增长会存在一定程度的精度误差。

c) 现场环境复杂。许多干扰器位于城区,障碍物多、建筑层数高、同网及异网基站密集,且往往发现同一片区域中存在多个干扰源,大大增加了定位的难度。

d) 干扰器工作原理复杂且工作时间不定,部分干扰器仅在白天或特定时间段开启,若依照传统模式仅选取业务闲时(即 02:00—04:00)的干扰电平数据来进行人工排查,容易忽略该类干扰器。

2 干扰定位建模基础理论

在通信系统中,发送端发出的无线信号经由信道传递至接收端。受发射器本身特性及传播环境的影响,接收端所收到的无线信号较原始发送信号均有一定程度的衰减。该衰减即为路径损耗,单位为 dB。若在自由空间内对无线信号的传播进行建模,接收端与发送端传播距离为 d ,发送端天线发射频率为 f ,则传播损耗 L_{bs} 可表示为:

$$L_{bs}(\text{dB}) = 32.45 + 20\lg d(\text{km}) + 20\lg f(\text{MHz}) \quad (1)$$

从式(1)可以看出, $L_{bs}(\text{dB})$ 是与 d, f 正相关的函数 $L(d, f)$ 。实际传播过程中,任一地形下的某发送端实际路径损耗是在 $L_{bs}(\text{dB})$ 的基础上,加上衍射损耗、地形损耗等损耗因子综合矫正后的结果,传播规律服从 $L(d, f)$ 。因为干扰器和基站均为信号发送端的一种形式,其所受干扰强度和终端距离及发射器本身的增益呈正相关。

图 1 给出了 LTE 外部干扰传播模式及其定位原理示意图。在图 1(a)中,基站 C_1, C_2 和 C_3 所在区域受干扰器 I 影响,可知距离干扰源最近的基站 C_1 受到强干扰,基站 C_2 、基站 C_3 则存在不同程度的弱干扰。相似地,干扰范围内的用户终端 U_1, U_2 和 U_3 因所处位置不同,其干扰信号强度依次递减。目前常见的 LTE 干扰定位模型采用计算用户上下行路损差值的方式来进行干扰源定位。假定正常 LTE 小区在任意位置上下行路损的差值应小于阈值(如 11 dB),已知干扰信号强度和终端所处距离呈正相关,因而上下行路损差值最大的位置即为干扰源所在位置。由于该方法涉及到上下行路损绝对值的计算,受环境的复杂性影响,难以对现网基站进行精准建模。此外,该方法需在基站侧开启额外的非标上行电平信息测量,受制于设备

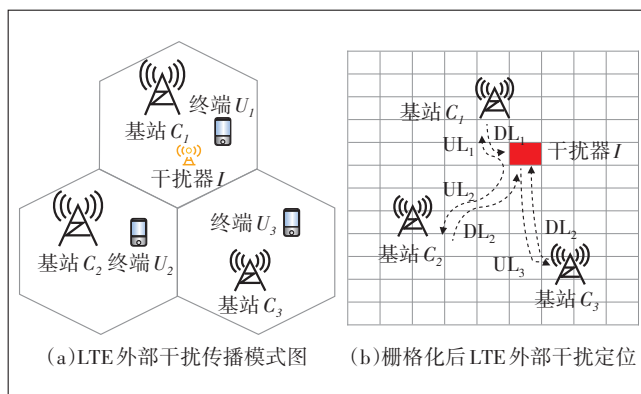


图 1 LTE 外部干扰传播模式及其定位原理示意图

厂商系统的局限性及准确性,较难推广至实际生产中。

因而,本文提出了一种通过计算上下行路损相似度的方法来进行干扰源定位,基于栅格化后的 MDT 数据,通过寻找上下行路损变化最为相似处来识别干扰源的位置。如图 1(b)所示,干扰源收到基站 C_1, C_2, C_3 的信号强度(即 DL_1, DL_2, DL_3 的路损大小)与其到基站 C_1, C_2, C_3 的距离 d_1, d_2, d_3 正相关。相似地,干扰器发送至基站 C_1, C_2, C_3 的干扰信号强度(即 UL_1, UL_2, UL_3)与其传播距离亦和距离 d_1, d_2, d_3 呈正相关。因此,下行路损向量(DL_1, DL_2, DL_3)与上行路损向量(UL_1, UL_2, UL_3)虽绝对值大小不同,但符合相同的波动规律,且距离干扰源越近的位置其相似性越高。通过计算上下行路损的相似度,本文采用了权重聚类算法对候选栅格进一步聚类,可以获取干扰源的预测位置。

3 干扰定位建模方法

根据小区上行 PRB 时域特征来识别干扰小区,生成干扰字典,结合栅格化后的 MDT 数据得到上下行路损向量的相似度,基于权重聚类算法定位可能的干扰源位置。模型整体流程如图 2 所示。

3.1 干扰小区特征提取

为了进行外部干扰定位,首先需提取小区的干扰特征,具体步骤如下。

步骤 1,通过 LTE 的统计测量报告 MRS 中 MR_RIPPRB 来获取各小区小时级 0~99 每个 PRB 的干扰噪声电平值,为每个小区构建如式(2)所示的 24×100 维干扰特征矩阵 C 来保留完整的干扰时域特征。

$$C = \begin{bmatrix} C_{0,0} & \cdots & C_{0,99} \\ \vdots & \vdots & \vdots \\ C_{24,0} & \cdots & C_{24,99} \end{bmatrix} \quad (2)$$

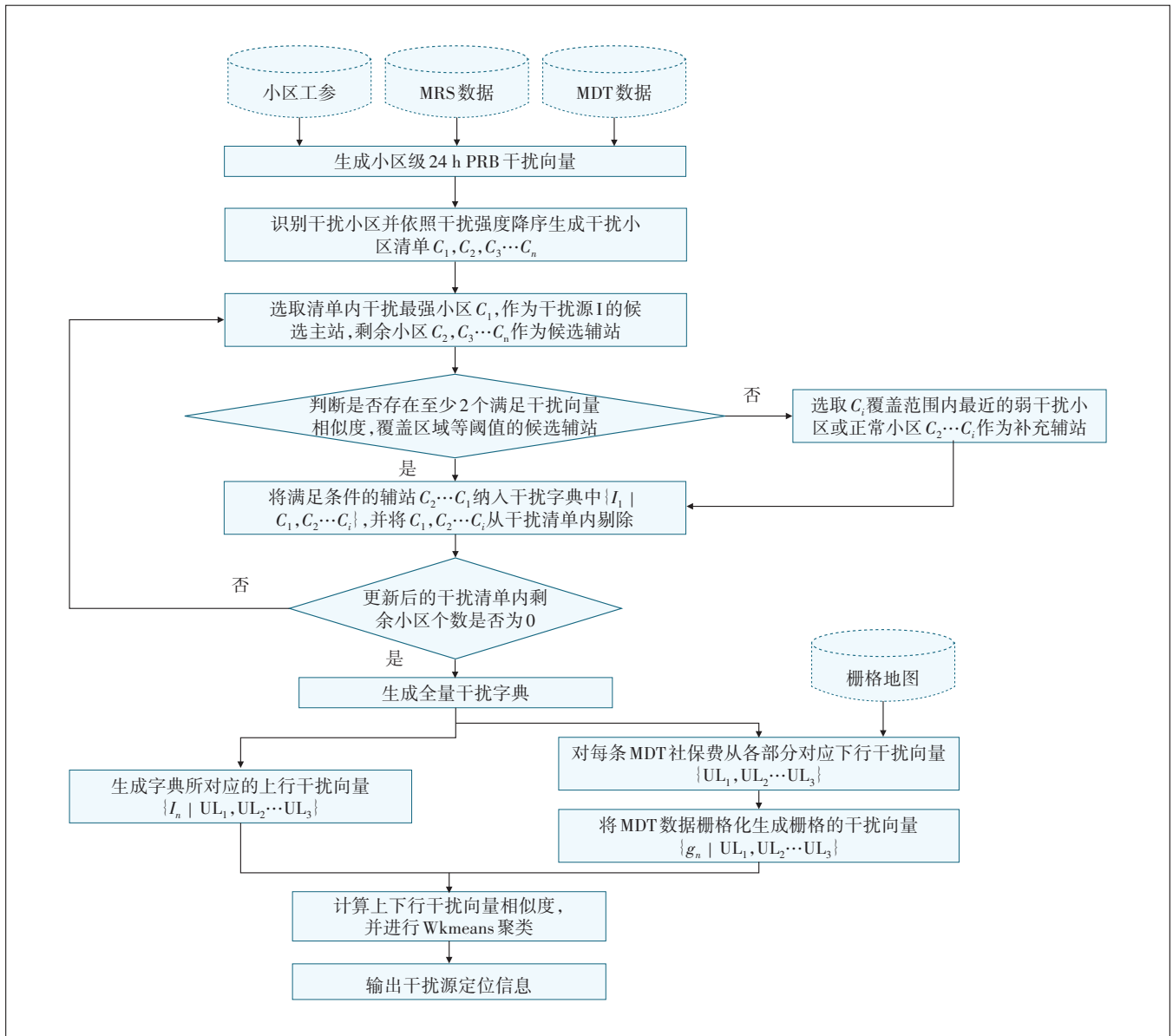


图2 LTE外部干扰定位流程图

步骤2,根据生成的特征矩阵针对任意小区 C_i 依照式(2)构建干扰强度评估函数 $IS(C_i)$:

$$IS(C_i) = f(\overline{PRB}_{\text{全天}}, PRB_{\text{忙时}} - PRB_{\text{闲时}}, \max(PRB), \max(N_{\text{连续干扰PRB个数}})) \quad (3)$$

式中:

- f ——干扰强度评估函数
- $\overline{PRB}_{\text{全天}}$ ——当日小区平均干扰电平值
- $PRB_{\text{忙时}} - PRB_{\text{闲时}}$ ——小区忙闲时干扰差值
- $\max(PRB)$ ——小区当日内的干扰最强小时的干

扰PRB电平值

$\max(N_{\text{连续干扰PRB个数}})$ ——当日小区小时级最大连续干扰个数的

通过 $IS(C_i)$ 对小区的干扰强度进行综合评分,对选定小区按干扰强度进行降序排序,生成最终干扰小区清单 $C = \{C_i | i \in [1, n]\}$ 。

步骤3,利用皮尔逊相似系数两两评估干扰小区清单内小区间的干扰相似度。已知 C 的干扰特征矩阵为 $C_{a,b}$, C' 干扰特征矩阵为 $C'_{c,d}$, 则0:00时2个小区的干扰特征向量可分别表示为: $\overline{C_{0,b}}$ 与 $\overline{C'_{0,d}}$ 。任意2个向量的皮尔逊相似度计算公式可表示为:

$$P_{X,Y} = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2\sum(Y-\bar{Y})^2}} \quad (4)$$

将 $\overline{C_{0,b}}$ 与 $\overline{C_{0,d}}$ 代入, 可得小区在 0:00 时的干扰相似度 P_0 。遍历 24 h 的干扰特征数据后求平均, 可获得 2 小区的平均皮尔逊干扰相似度 \bar{P} 。 $\bar{P} \in [-1, 1]$ 。当 $\bar{P} = 0$ 时, 2 矩阵不相关; $\bar{P} \in (0, 1]$ 时, 相关性随着 \bar{P} 的增大而增大, 呈正相关性。 $\bar{P} \in [-1, 0)$ 时, 相关性随着 \bar{P} 的减小而增大, 呈负相关性。在现网实际建模中, 往往发现 $C_{a,b}$ 和 $C_{c,d}$ 可能不是 24×100 的矩阵, 即 $a \neq b$ 或 $c \neq d$, 或 $a \neq b$ 且 $c \neq d$, 此现象通常为以下 2 种情况导致。

a) 小区因临时断站或节电等原因导致缺失部分小时或个别 PRB 的数据。

b) 现网不同小区的容量不同, 主流 20 MHz 带宽的小区可占用 100 个 PRB, 但部分 5~15 MHz 的小区仅分别占用 25、50、75 个 PRB, 进而导致剩余 PRB 的干扰电平值为空。

考虑到此处的数据缺失具有实际意义, 若统一填充异常值(如“-1”)进行维度占用将产生误差, 因而需注意在两两向量相似度计算前将矩阵裁剪至维度一致, 即令 $a = b = \min(a, b)$, $c = d = \min(c, d)$ 。

3.2 干扰字典生成

已知受相同干扰源影响的小区在时域 PRB 特征上存在高度相关性, 根据这一特性将干扰清单内的小区按照不同干扰源进行分组, 进而生成干扰字典。具体步骤如下。

步骤 1, 选取清单 C 内干扰最强小区 C_1 , 作为干扰源 I 的候选干扰主站 I_1 , 令剩余小区为 $C_{a1}, C_{a2}, C_{a3}, \dots, C_{an}$, 并作为候选辅站。

步骤 2, 按照以下规则构建干扰字典判决函数:

a) 候选辅站与候选干扰主站之间干扰特征向量的相似度是否大于给定阈值(如 0.8)。

b) 候选辅站与候选干扰主站之间是否存在共覆盖区域。

步骤 3, 判断满足要求的候选辅站个数大于等于 2 个。若大于等于 2 个, 至多选取干扰强度最强的前 5 个站点作为干扰辅助站点; 若小于 2 个, 选取 I_1 覆盖范围内最近的弱干扰小区或正常小区 $C_{a1}, C_{a2}, C_{a3}, \dots, C_{ai}$, 作为补充辅站。

步骤 4, 将满足条件的辅站 $C_{a1}, C_{a2}, C_{a3}, \dots, C_{ai}$ 依照干扰强度降序纳入干扰字典中 $\{I_1|C_{a1}, C_{a2}, \dots, C_{ai}\}$, 并将

I_1 , 对应的站点从初始干扰清单 C 内剔除。

步骤 5, 判断更新后的干扰清单 C 内的站点个数是否为 0。若不为 0, 返回步骤 1 继续运算, 直到遍历完清单内所有站点。

步骤 6, 输出干扰字典 $\{I_j|C_{a1}, C_{a2}, \dots, C_{ai}\}$ 。其中, j 代表区域内可预测干扰源。对于字典内任意干扰源 I_j , 有 i 个小区受其干扰影响, i 的取值范围在 $[2, 6]$ 内。以图 1 (b) 为例, 干扰源 I 对应的干扰字典可表示 $\{C_1|C_2, C_3\}$, 其中干扰强度 $IS(C_1) > IS(C_2) > IS(C_3)$ 。

3.3 上行干扰向量生成

根据传播模型, 任意干扰源的上行路损可表示为:

$$UL_i = P_{\text{干扰源发射功率}} - PRB_i \quad (5)$$

针对干扰字典中某特定干扰源, 其周边受影响的干扰站点可表示为 $(C_i, C_{a1}, C_{a2}, C_{a3}, \dots, C_{ai})$, 对应的上行路损向量可表示为:

$$\overline{UL} = (UL_i, UL_{a1}, UL_{a2}, UL_{a3}, UL_{a4}, UL_{a5}) \quad (6)$$

对该向量按式(7)进行差值化处理,

$$\Delta \overline{UL} = \overline{UL} - UL_i = (0, \Delta UL_{a1}, \Delta UL_{a2}, \Delta UL_{a3}, \Delta UL_{a4}, \Delta UL_{a5}) \quad (7)$$

其中 $\Delta UL_{ai} = UL_{ai} - UL_i = PRB_i - PRB_{ai}$

3.4 下行干扰向量生成

根据传播模型, 任意干扰源的下行路损可表示为:

$$DL_i = P_{i\text{基站发射功率}} - G_{i\text{天线增益}} - RSRP_i \quad (8)$$

从 MDT 数据中可获取干扰主站每个采样点主服务器小区 RSRP 电平值以及邻区 RSRP 电平值, 则每条 MDT 的下行路损向量可表示为:

$$\overline{DL} = (DL_i, DL_{a1}, DL_{a2}, DL_{a3}, DL_{a4}, DL_{a5}) \quad (9)$$

对该向量按式(10)进行差值化处理:

$$\Delta \overline{DL} = \overline{DL} - DL_i = (0, \Delta DL_{a1}, \Delta DL_{a2}, \Delta DL_{a3}, \Delta DL_{a4}, \Delta DL_{a5}) \quad (10)$$

其中 $\Delta DL_{ai} = DL_{ai} - DL_i = (P_{ai\text{基站发射功率}} - G_{ai\text{天线增益}} - RSRP_{ai}) - (P_{i\text{基站发射功率}} - G_{i\text{天线增益}} - RSRP_i)$ 。将计算得到的向量落入 50×50 的栅格地图中, 可获得每个栅格所对应的下行干扰路损向量。

3.5 权重聚类定位干扰源

已知下行路损与上行路损向量波动规律越相似的栅格即为干扰源所在的栅格。因而依照式(11), 将上下行路损向量代入余弦相似度计算公式进行相似度 W_i 的计算。

$$W_i = \cos(\overline{\Delta UL}, \overline{\Delta DL}) = \frac{\sum_{i=1}^n (\Delta UL_i \times \Delta DL_i)}{\sqrt{\sum_{i=1}^n \Delta UL_i^2} \times \sqrt{\sum_{i=1}^n \Delta DL_i^2}} \quad (11)$$

其中, $W_i \in [-1, 1]$ 。 W_i 越大, 向量间相似度越高, 反之则相似度越低, 可认为 $\min(W_i)$ 所在的栅格为本文前述方法所认定的干扰源位置。

但是, 同一小区往往会受到多个干扰源的共同影响, 处于覆盖边缘或阻挡位置的栅格有可能因为弱覆盖等原因反而与上行干扰向量存在高度相关性。最终筛选得到的高相似性栅格集合可能在空间上呈现离散的点状分布态势。图3(a)即为一典型场景。相似度最高的栅格(相似度0.99)与真实干扰源存在一个栅格距离的偏差, 而覆盖范围内右下角(图3(b))存在大量高相似度的栅格。

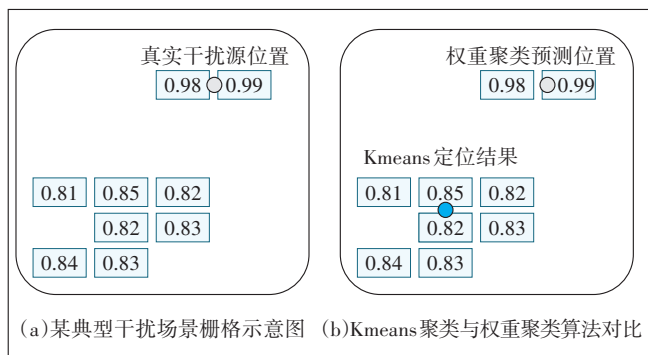


图3 权重聚类算法应用示意图

如图3(b)中所示, 常规Kmeans聚类算法以数据点到初始聚类中心的欧式距离作为优化的目标函数, 聚类结果会因部分密集低权重的栅格产生偏差。而本文算法将上下行路损相似度作为栅格的权重参数 W_i 纳入目标函数中, 将迭代目标更改为计算每个栅格到质心的加权距离和。该方法强化了高相似性栅格在聚类时的重要性。相较于直接选取相似度最高的栅格, 最终预测位置和真实干扰源位置的误差进一步缩小, 有效提升了定位精度。具体实现步骤如下:

步骤1, 初始化聚类个数 $k = 1$ 。

步骤2, 假设 m_c 代表质心, g_i 代表候选栅格样本, π_c 代表生成的聚类簇, 依照公式(12)迭代计算每个栅格到质心的加权距离和 D , 生成初始聚类结果如下:

$$D\left(\left\{\pi_c\right\}_{c=1}^k\right)_{\text{WKmeans}} = \sum_{c=1}^k \sum_{g_i \in \pi_c} W_i \|g_i - m_c\|^2, m_c = \frac{\sum_{g_i \in \pi_c} W_i g_i}{\sum_{g_i \in \pi_c} W_i} \quad (12)$$

步骤3, 按式(13)计算当前轮廓系数, 其中 a 表示某个样本与其所在簇内其他样本的平均距离, b 表示某个样本与其他簇样本的平均距离。 S_k 越接近1则代表聚类性能越好。

$$S_k = \frac{b - a}{\max(a, b)}, S \in [-1, 1] \quad (13)$$

步骤4, 令聚类个数 $k=k+1$, 重复步骤2和步骤3。

步骤5, 比较 S_k 和 S_{k+1} 。若 $1 > S_{k+1} > S_k$, 重复步骤2~步骤5; 否则循环结束, 输出聚类结果。

3.6 指标评价

为评估本文算法在实际生产环境的预测能力, 通过人工现场实地排查, 假设预测位置和其真实干扰源 i 的距离误差为 d_i , 共计有 n 个样本, 则平均误差为

$$\frac{\sum_{i=0}^n d_i}{n}$$

4 干扰定位模型效果验证

本文使用的数据集为某省所有在网小区某天小时级每PRB的干扰噪声电平值来进行分析处理。结合人工扫频现场排查, 使用本文所提出的算法, 在某省试点某频段小区, 共排查出172个干扰小区, 干扰小区识别准确率为97.18%, 输出63个干扰源可能位置。对63个干扰源定位误差分析如表1所示。

表1 各场景下算法评估对比表

场景	测试样本个数	平均误差/m	误差中位数/m	误差范围/m
开阔区域(农村、城郊)	15	36	38	[0, 46]
中等密集区域(城镇或相对开阔的楼宇)	20	76	93	[0, 127]
复杂区域(居民区、密集楼宇)	28	115	134	[0, 204]

对比发现, 本模型在无线环境相对简单的场景误差精度可达46 m, 而随着环境复杂度的提升, 预测误差也随之增大, 平均误差最高不超过115 m, 误差范围控制在204 m之内, 显著提升干扰排查效率。

图4为模型验证的3种典型案例, 分别为厂区、办公楼宇、居民区3类场景。其中黑色方块代表模型预测位置, 橙色三角形则代表真实干扰源所在位置。

a) 某厂区位于远郊, 处于空旷开阔地, 周围无线环境相对简单。该厂区附近基站24 h内均出现不同程度的全频段干扰。经现场排查后发现干扰器实际为厂区内私装的信号放大器。对比预测结果, 可以看



图4 3类典型场景干扰预测对比及现场实物图

到模型对该类场景适应性较好,预测误差不超过 50 m。

b) 某办公楼宇处于中等密集城区,周边楼宇相对较多,且周边存在多个不同频段的基站。该楼宇附近基站在凌晨呈现弱干扰,而在白天工作时间 08:00—20:00 存在强干扰。人工排查后发现为楼宇梯控系统所致,对比预测结果,可以看到模型对该类场景适应性较好,预测误差不超过 50 m。

c) 某居民区为别墅区,其地下室信号覆盖不佳。多幢住户私装放大器,片区内干扰器密集,影响周边多个基站。模型迭代预测出 2 个可能的干扰源位置。预测误差较前面 2 类场景稍高,但未超过 150 m。

5 结束语

本文描述了一种基于权重聚类的 LTE 外部干扰定位和识别的方法。相较于传统通过计算上下行链路不平衡预测的方式,本文所提出的算法一方面引入相似度概念,减少对传播空间建模的误差;另一方面通过引入多个站点进行辅助定位,结合人工智能中的权重聚类算法,进一步提高了定位的精准度。本文所述方法已在东南某沿海 X 省成功试点,实时性高、误差精度小,可有效估计干扰源的位置,降低人工成本,具

备实际生产指导意义,具有良好地推广性与复制性。

参考文献:

- [1] 谢华. 大数据在移动通信中的应用探讨[J]. 科技创业家, 2014 (2): 125-126.
- [2] 江华,孙圆圆. 电力无线专网多业务隔离技术研究[J]. 宁夏电力, 2015(3): 51-56.
- [3] 李宝磊,周俊,任晓华. TD-LTE 网络优化关键问题的研究[J]. 电信工程技术与标准化, 2015, (1): 57-61.
- [4] 黄庆秋. TD-LTE 网络优化关键问题的研究[J]. 网络安全技术与应用, 2018, 211(7): 74-75.
- [5] 孙士保,秦克云. 改进的 k-平均聚类算法研究[J]. 计算机工程, 2007, 33(13): 200-201.
- [6] 李桂林,陈晓云. 关于聚类分析中相似度的讨论[J]. 计算机工程与应用, 2004, 40(31): 64-65.
- [7] 屈晓光. TD-LTE 干扰分析[J]. 科技致富向导, 2015(5): 276-276.
- [8] 钟逸秋,冯赞,曾纪宽. LTE 干扰分类和优化方法探究[J]. 通讯世界, 2016(9): 5-5.

作者简介:

赵伟,毕业于香港理工大学,工程师,硕士,主要研究方向为网络优化、大数据及人工智能在移动通信中运用;王勇,毕业于北京大学,高级工程师,硕士,主要研究方向为网络优化;陈乐,毕业于英国曼彻斯特大学,工程师,硕士,主要从事大数据及人工智能在移动通信中运用方面的研究工作。