

基于GBDT算法的

TA偏离因子结构识别模型研究

Research on TA Deviation Factor
Structure Identification Model
Based on GBDT Algorithm

李翔明¹,尹以雁¹,臧玉²,何璐璐¹,胡坚¹,刘璐³,张倩¹,阙莹淑¹(1. 中国移动通信集团云南有限公司,云南昆明 650228;2. 上海大唐移动通信设备有限公司,上海 200233;3. 中国移动通信集团设计院有限公司重庆分公司,重庆 401121)
Li Xiangming¹,Yin Yiyan¹,Zang Yu²,He Lulu¹,Hu Jian¹,Liu Lu³,Zhang Qian¹,Que Yunshu¹(1. China Mobile Communications Group Yunnan Co., Ltd., Kunming 650228, China; 2. Shanghai Datang Mobile Communication Equipment Co., Ltd., Shanghai 200233, China; 3. China Mobile Communications Group Design Institute Co., Ltd. Chongqing Branch, Chongqing 401121, China)

摘要:

随着4G/5G网络结构问题日益复杂,传统的网络结构识别模型面临着样本数据量不足和鲁棒性差的问题。提出一个基于TA识别覆盖特征判断网络结构问题的模型,它利用时间提前(Timing Advance, TA)偏离因子模型,结合梯度提升决策树(Gradient Boosting Decision Tree, GBDT)算法对训练样本数据进行特征提取,迭代优化损失函数以增强模型算法的预测精确度,为网络结构问题识别提供了一种新方法。

关键词:

GBDT算法;TA偏离因子;机器学习;结构识别
doi:10.12045/j.issn.1007-3043.2026.02.008
文章编号:1007-3043(2026)02-0039-06
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):



Abstract:

With the increasing complexity of 4G/5G network structure problems, traditional network structure recognition models face the problems of insufficient sample data and poor robustness. It proposes a model for determining network structure problems based on TA recognition coverage feature, which utilizes the Time Advance (TA) deviation sub model and combines the Gradient Boosting Decision Tree (GBDT) algorithm to extract features from training sample data. The loss function is iteratively optimized to enhance the prediction accuracy of the model algorithm, providing a new method for network structure problem recognition.

Keywords:

GBDT algorithm; TA deviation factor; Machine learning; Structure identification

引用格式:李翔明,尹以雁,臧玉,等. 基于GBDT算法的TA偏离因子结构识别模型研究[J]. 邮电设计技术,2026(2):39-44.

1 概述

基站数量的急剧增长在不断提升网络信号覆盖的同时,使网络结构问题日益凸显,如超高站带来的越区覆盖、塔下黑等问题,超近站引起的重叠覆盖、干扰等问题,都直接影响用户的网络感知体验。因此,研发高效、精确识别并解决这些网络结构问题的方法显得尤为迫切和重要。在此背景下,本文提出了一个基于TA识别覆盖特征判断网络结构问题的模型。

2 TA技术原理

TA表征的是UE与天线端口之间的距离,用户在随机接入过程中,当基站接收到RA Preamble信令后,会向用户设备发送接入响应(RAR),在接收到RAR并成功发送第1条CRC校验正确的上行数据后,eNodeB会根据RAR中的TA值来统计相关指标,不同UE到达基站的距离是不同的,传播时延决定了TA的大小,一般 $TA=2\times$ 传播时延^[1],具体如图1所示。

在LTE网络中,最小时间单位 $T_s=1/(15\ 000\times 2\ 048)$,单位是s,1个 T_s 对应的时间提前量代表的距离约为4.89

收稿日期:2026-02-16

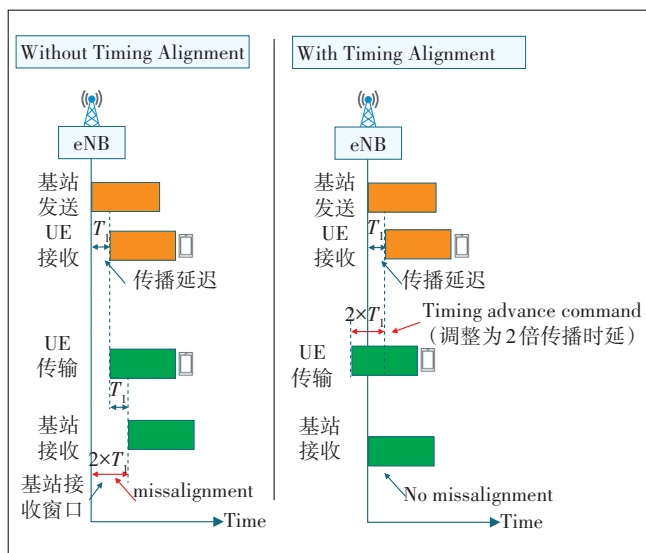


图1 TA定时提前技术原理

m。UE调整上行发射时间的公式为：

$$N_{\text{ta}} = \text{TA} \times 16T_s \quad (1)$$

当定时提前量(TA)被设定为1时,实际距离即 $16 \times 4.89 \text{ m} = 78.24 \text{ m}$ 。

在NR网络中,最小时间单位 $T_c = 1/(380\,000 \times 4\,096)$,单位是s,UE调整上行发射时间的公式为：

$$N_{\text{ta}} = \text{TA} \times 16 \times \frac{64}{2^\mu} \times T_c \quad (2)$$

不同SCS取值下,每个时间符号对应的距离如表1所示。

表1 不同SCS取值对应的TA映射关系

μ	子载波间隔	TA精度/s	TA距离/m
0	15	$16 \times 64 \times T_c$	78.13
1	30	$16 \times 32 \times T_c$	39.06
2	60	$16 \times 16 \times T_c$	19.53
3	120	$16 \times 8 \times T_c$	9.77

3 研究策略及技术方案

3.1 总体实现流程

通过对某省网络进行覆盖与业务的特征聚类,根据不同场景提取覆盖特征,同时构建TA偏离因子模型,与基站覆盖进行关联性分析。同时,通过GBDT算法对覆盖、容量、性能、工程参数等多个决策树参数进行迭代,利用加法模型逐步构建多个回归树,累加上述回归树的输出值,最终实现基于TA偏离因子模型的网络结构问题的定量预测,总体实现流程如图2所示。

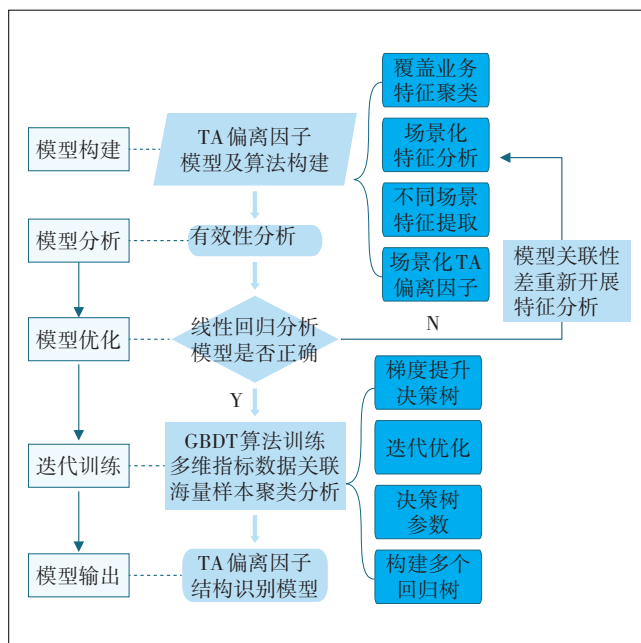


图2 总体实现流程

示。

3.2 业务特征分析

根据某省的采样数据,将4G小区划分为4个象限,分别表示不同的覆盖距离和业务量组合,具体如图3所示。

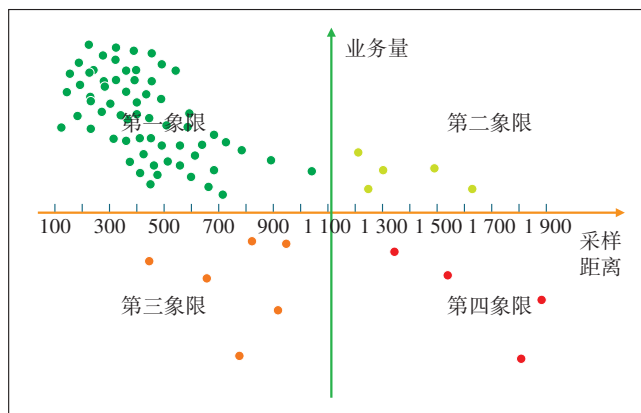


图3 覆盖距离和业务量组合象限图

通过聚类分析,大部分小区位于第一象限,即覆盖距离较近,业务量较高,符合理论预期。然而,也有部分小区位于第二、三、四象限,表现出不合理的覆盖距离和业务承载组合,需进一步探索覆盖与业务的具体关联性。

在城区场景,通过分析大量样本数据发现,主力业务承载频段的覆盖距离不超过700 m,城区场景下TA覆盖范围与业务分布如图4所示。

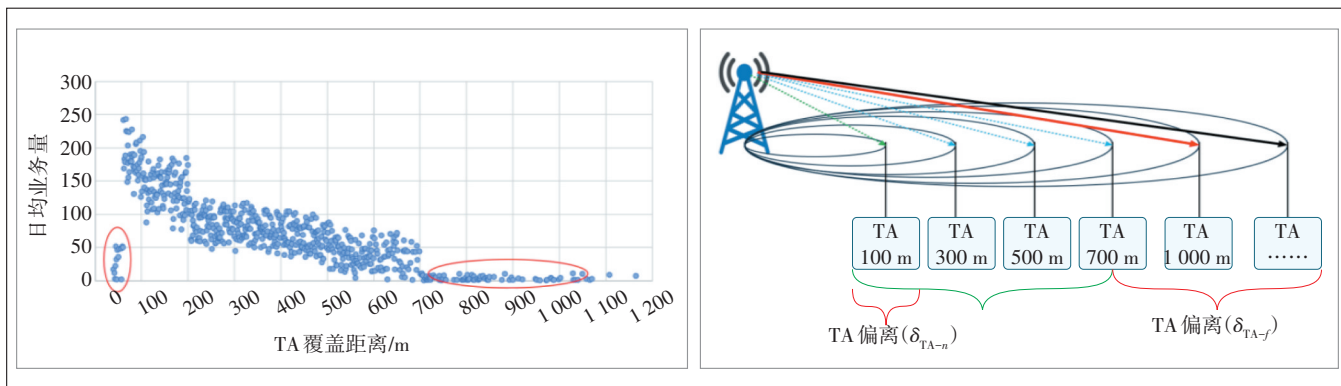


图4 城区室外小区TA覆盖范围与业务分布

在农村场景,通过分析大量样本数据发现,主力业务承载频段的覆盖距离不超过1 200 m,农村场景下TA覆盖范围与业务分布如图5所示。

3.3 算法模型构建

通过对城市农村基站小区覆盖距离与业务的综合分析,可以得到以下结论,结论A:城区覆盖业务集中区不大于TA9(≈700 m);结论B:农村覆盖业务集中区不大于TA15(≈1 200 m)。

实际画像TA分布结果值和理想规划预期的偏差可以用TA偏离因子(δ_{TA})来表示,其中过近偏离因子用 $\delta_{(TA-n)}$ (不分场景)表示,城区过远偏离因子用 $\delta_{(TA-f)}$ 表示,农村过远偏离因子用 $\delta_{(TA-l)}$ 表示。设 N 为采样总数, X_i 是第 i 个采样的长度(以m为单位),首先确定超出理论规划模型(≈700 m)的采样数量 $N_{>700}$ 。

$$N_{>700} = \sum_{i=1}^N [X_i > 700] \quad (3)$$

$[X_i > 700]$ 是一个逻辑表达式,如果 X_i 大于700,则其值为1,否则为0。因此, $\sum_{i=1}^N [X_i > 700]$ 实际上是对所有满足条件的采样进行计数。然后,计算 X_i 大于

700 m业务的占比。

$$\text{占比} = \frac{N_{>700}}{N} \times 100\% \quad (4)$$

即大于700 m的业务数量除以总业务数量,再乘以100%,得到的就是该部分业务占总业务的比例,以百分比形式表示。综上,TA偏离因子 $\delta_{(TA-f)}$ 可表示为:

$$\delta_{(TA-f)} = \left[\frac{\sum_{i=1}^N [X_i > 700]}{N} \right] \times 100\% \quad (5)$$

通过以上数据模型算法,可以得到TA偏离因子 $\delta_{(TA-l)}$ 及 $\delta_{(TA-n)}$:

$$\delta_{(TA-l)} = \left[\frac{\sum_{i=1}^N [X_i > 1200]}{N} \right] \times 100\% \quad (6)$$

$$\delta_{(TA-n)} = \left[\frac{\sum_{i=1}^N [X_i < 100]}{N} \right] \times 100\% \quad (7)$$

综上所述,构建3类TA偏离因子 δ_{TA} 的模型及算

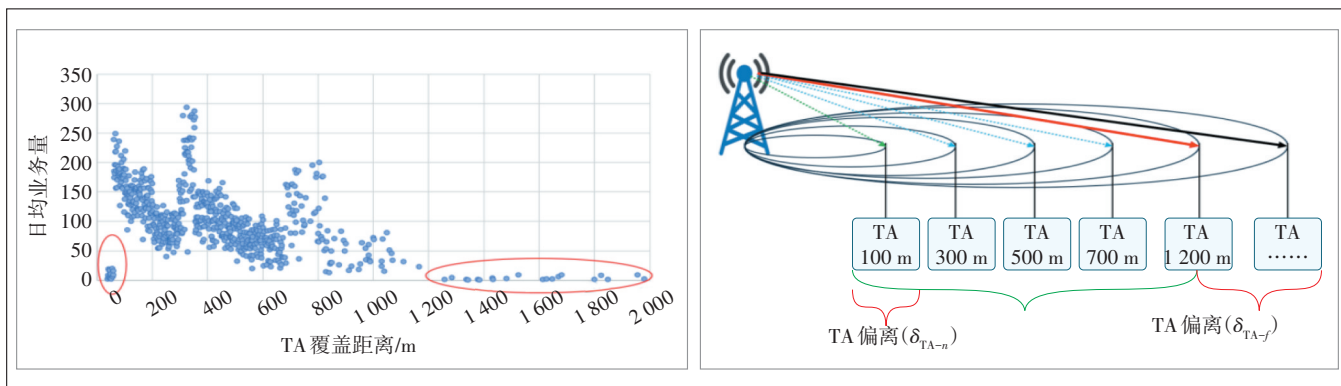


图5 农村室外小区TA覆盖范围与业务分布

法,即TA 700 m之外偏离因子 $\delta_{(TA-f)}$ 、TA 1 200 m之外偏离因子 $\delta_{(TA-l)}$ 和TA 100 m以内偏离因子 $\delta_{(TA-a)}$ 。

3.4 模型关联性验证

通过线性回归(Linear Regression)模型探究因变量与自变量之间存在的关联性。

将现网采样数据建模,探索TA偏离因子与覆盖距离的关系,尝试在站址结构问题分析中引入TA偏离因子,定位站址结构问题,结果如图6所示。

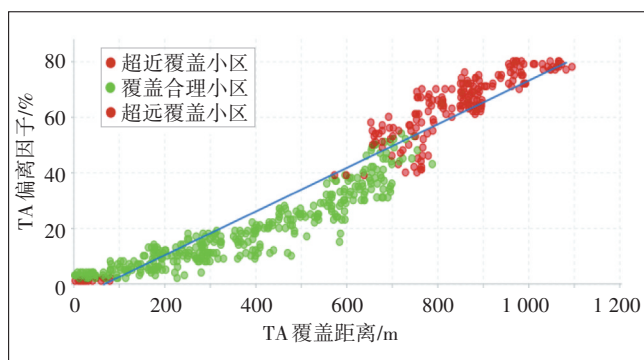


图6 TA偏离因子与覆盖的线性回归拟合

线性回归方程的 R^2 用来衡量线性回归模型的拟合度,回归模型展现出了其对因变量变化性的阐释与说明能力^[2],计算公式为:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (8)$$

其中,SSR为回归平方和,表示因变量 Y 的变化可以被自变量 X 解释的部分;SST为总平方和,表示因变量 Y 的总变化;SSE为残差平方和,表示因变量 Y 的变化中不能被自变量 X 解释的部分。式(8)可进一步表示为:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - Y)^2} \quad (9)$$

其中, Y_i 表示实际观测到的数据点, \hat{Y}_i 表示根据回归模型预测得出的相应值, Y 表示所有观测数据的算术平均值。当数值趋近于1时,意味着变量间存在着较强的正相关关系;相反,若数值接近0,则表明变量间缺乏显著的相关性。

通过线性回归分析,TA偏离因子与覆盖的 $R^2=0.91$,表示TA偏离因子模型与小区覆盖呈现强相关性,可在评估覆盖距离、判断结构问题方面使用。

3.5 基于GBDT算法训练

GBDT是一种先进的机器学习算法,其核心机理

为循环地构建决策树模型,每一轮迭代均依据前一轮的输出结果进行调整与优化,并通过逐步减少预测残差的过程实现数据的精确分类或回归分析。

GBDT算法通过多个迭代周期来逐步构建模型,每个周期都会生成一个弱学习器,在GBDT框架中,弱分类器通常选取分类与回归树(Classification And Regression Tree, CART)。鉴于对模型复杂度(要求简单、高偏差)的控制,每个CART树的深度通常不会被限制得太深。最终,通过将所有迭代轮次中训练得到的CART弱分类器按照一定权重进行累加,形成强大的总分类器^[3]。模型最终可以描述为:

$$f_M(x) = \sum_{i=1}^M T(x; \theta_i) \quad (10)$$

其中, M 表示模型所拥有的CART树的数量, $T(x; \theta_i)$ 表示第 i 棵回归树对样本的预测结果,其中 θ 表示每一棵回归树中的参数,GBDT模型最终的结果是 M 个回归树预测的加和,具体如图7所示。

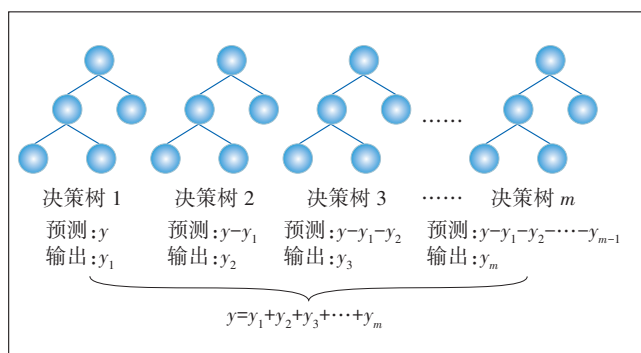


图7 GBDT算法原理

3.5.1 GBDT训练图

GBDT算法的关键是每一棵树在学习前面所有树预测和的残差,最终的预测是每一棵树预测的叠加值^[4]。基模型采用决策树,为有效控制模型复杂度并确保其保持弱模型特性,各决策树的深度被有意限制在一个较浅的层次。在解决回归问题时,标准做法是使用平方误差损失函数,其数学表达式为:

$$L(y, \hat{y}) = (y - \hat{y})^2 \quad (11)$$

其中, L 代表损失函数, y 是数据的真实值(或目标值), \hat{y} 是模型的预测值。

GBDT基模型训练如图8所示,其中参数 w 就是每个基模型的决策树的参数,也就是每个基模型的每个分裂特征和每个分裂阈值。

3.5.2 初始化过程

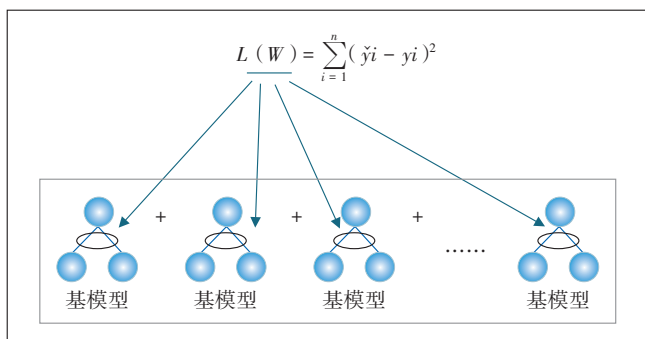


图8 GBDT基模型训练

创造第1棵回归树 $f_1(x)$,它在回归问题中用回归树拟合目标值的结果:

$$f_1(x) = \underset{c}{\operatorname{arg\,min}} \sum_{i=1}^N L(y_i, c) \quad (12)$$

3.5.3 迭代优化

在构建GBDT模型的过程中,对于从第2棵到第 m 棵的每一棵回归树,需确定其训练目标,这一目标具体表现为前一棵或前几棵树预测结果与实际观测值之间的残差:

$$r_{mi} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right] f(x) = f_{m-1}(x) \quad (13)$$

在构建当前第 m 棵子树时,需系统地探索所有可行的分割点及其对应的阈值,以寻找一个最优的预测值 c (通常称为叶子节点的输出值),该值应能最大程度地拟合前序树预测后的残差,具体公式为:

$$c_{mj} = \underset{c}{\operatorname{arg\,min}} \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i) + c) \quad (14)$$

在描述第 m 棵子树的构建过程中,引入符号 R_{mj} 来表示该树所有可能的叶子节点区域,这些区域对应不同的预测值集合,即第 m 棵回归树可能产生的所有预测值,其中 j 的取值范围是从1到 J 。接下来更新模型的预测函数 $f_m(x)$,其表达式为 $f_m(x) = f_{m-1}(x_i) + \sum_{i=1}^J c_{mj} I(x \in R_{mj})$,其中 I 是一个指示函数,当样本 x 落在特定的叶子节点区域 R_{mj} 内时, I 的值为1,否则为0。这样的更新机制旨在通过在新一轮模型预测的基础上,加上当前树基于叶子节点预测值的调整,从而逐步逼近真实值。最后得到回归树:

$$F(x) = f_M(x) = \sum_{m=1}^M f_m(x) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (15)$$

GBDT的核心在于利用加法模型逐步构建多个回

归树,最终预测结果由这些回归树的输出值累加。在每一轮迭代中,新构建的回归树主要聚焦于拟合前一轮模型预测与实际观测值之间的残差,以此作为训练目标,旨在逐步减少预测误差,提升整体模型的预测性能。

3.5.4 模型训练结果

通过对某省50万样本数据开展梯度提升决策树训练,研究TA偏离因子对结构问题识别预测。

在城区场景,当TA偏离因子 $\delta_{(TA-f)}$ 大于等于40%,且弱覆盖关键KPI劣化时,则判定为过远覆盖;当TA偏离因子 $\delta_{(TA-n)}$ 大于等于60%,且容量关键KPI劣化时,则判定为过近覆盖;当TA偏离因子 $\delta_{(TA-f)}$ 大于等于30%,下倾角不在2~12°范围内,且弱覆盖关键KPI劣化时,则判定为疑似超高覆盖;当TA偏离因子 $\delta_{(TA-n)}$ 大于等于60%,下倾角不在2~12°范围内,且容量关键KPI劣化时,则判定为疑似超低覆盖,具体如图9所示。

4 应用效果

传统的结构问题识别方法主要通过TA平均覆盖距离按照固定的值进行判断,存在诸多局限性,在面对不同覆盖场景时,一刀切式模型的鲁棒性较差,需要人工进行现场勘察确认,不仅耗时长且成本高^[5]。

应用基于GBDT算法的TA偏离因子后,在结构问题定界时,通过梯度迭代学习,在各维度数据中提取结构特征,提高对结构问题的识别能力,在问题识别方面,识别准确率由39.55%提升至87.30%,在感知提升方面,整治站点平均感知速率由115 Mbit/s提升至159 Mbit/s,具体如图10和图11所示。

5 结束语

网络结构问题是影响用户感知的关键,在建设阶段,强有力的设计会审管控可有效规避结构问题,然而,在实际网络运营过程中,城市化推进以及用户需求的调整都会带来网络结构的变化。

本文提出的基于TA偏离因子的结构问题定界方法,通过构建TA偏离因子模型,实现了对结构问题的高效精准识别,实施区域结构识别定界准确率和整治站点平均感知速率均显著提升。未来在网络规划、网络建设中将进一步探索该方法的应用优化,最大化规避网络结构问题流转至后续优化阶段处理。

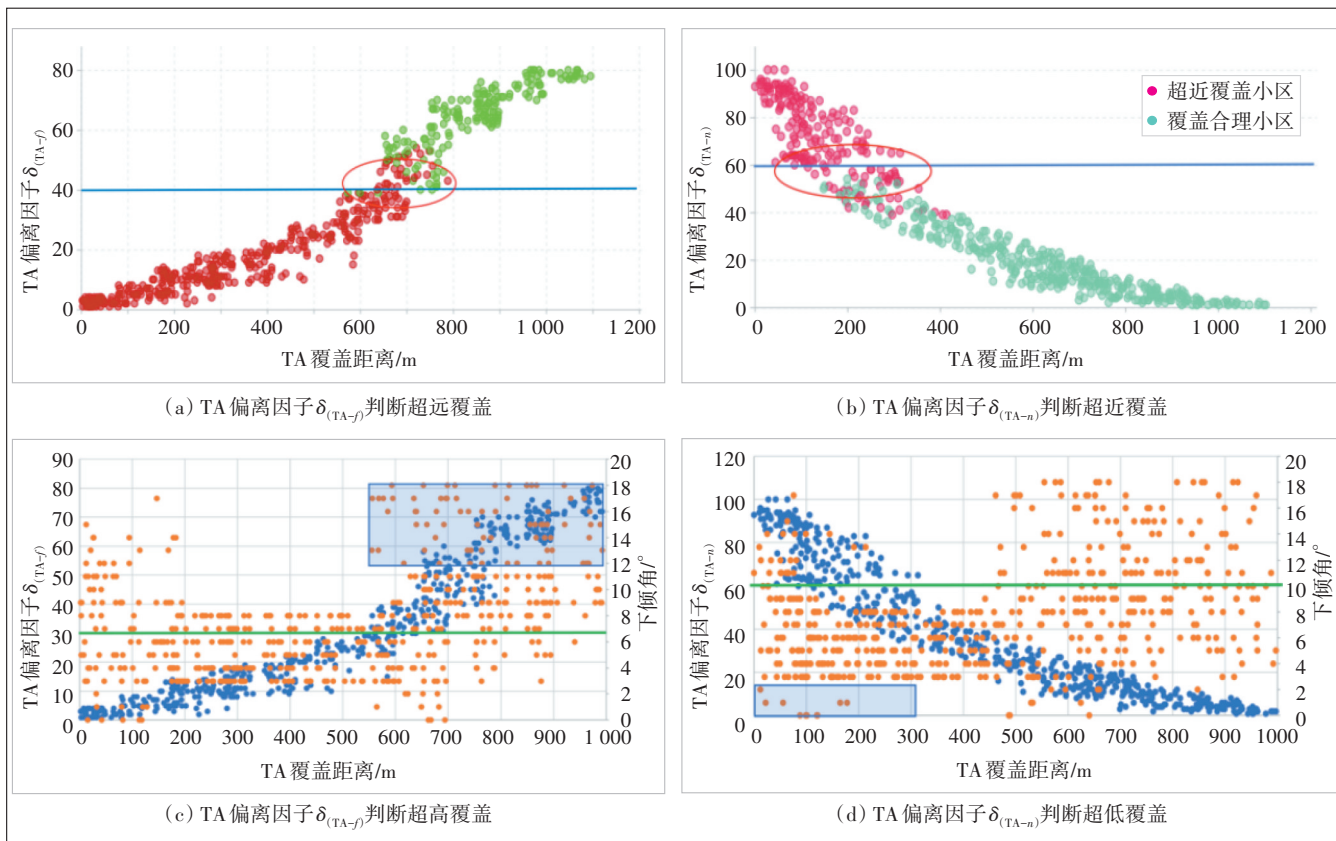


图9 训练效果

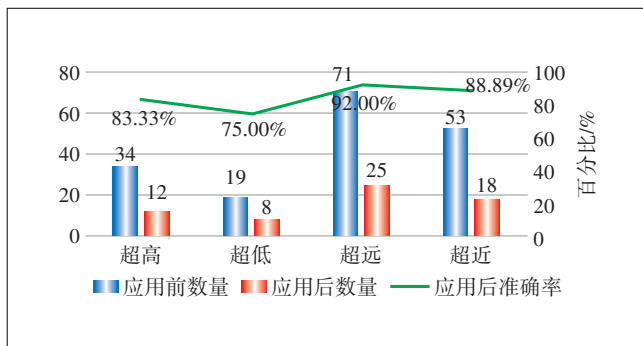


图10 TA 偏离因子应用效果

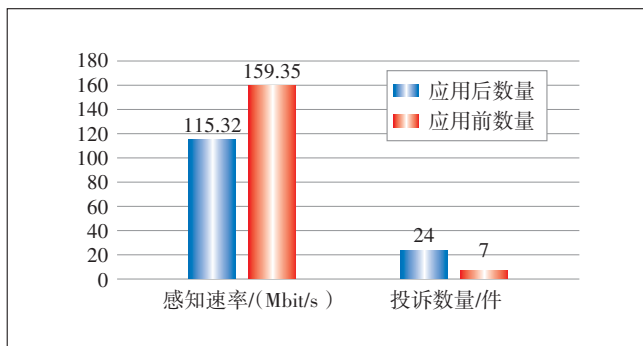


图11 TA 偏离因子感知效果

参考文献:

- [1] 赵海舜. 基于SoC的LTE物理层设计与实现[D]. 西安: 西安电子科技大学, 2015.
- [2] 任强. NB-IoT网络质差小区分析与优化方案[J]. 数字通信世界, 2020(6): 18-20.
- [3] 翁海兵, 杨阳, 黄颖. 基于GBDT算法的电力工程数据信息分析及预测方法研究[J]. 电子设计工程, 2023, 31(24): 154-158.
- [4] 张耀方, 陈坚. 基于GBDT算法的高速公路分车型交通流短时预测模型[J]. 公路, 2022, 67(1): 221-227.
- [5] 郑锦鹏, 邓志勇, 魏坚. 4G MR覆盖率优化思路研究[J]. 广东通信技术, 2019, 39(4): 15-18, 22.

作者简介:

李翔明, 毕业于电子科技大学, 高级工程师, 硕士, 主要从事无线网络规划优化工作; 尹以雁, 毕业于华南理工大学, 硕士, 主要研究方向为4G/5G无线网络优化工作; 臧玉, 毕业于南京信息职业技术学院, 主要从事无线网络优化工作; 何璐璐, 毕业于西北工业大学, 工程师, 硕士, 主要从事无线网络优化和智能化节能研究工作; 胡坚, 毕业于合肥工业大学, 高级工程师, 硕士, 主要研究方向为4G/5G无线网络优化; 刘璐, 毕业于贵州大学, 高级工程师, 学士, 主要从事网络智能优化研究工作; 张倩, 毕业于北京邮电大学, 工程师, 硕士, 主要从事网络质量管理; 阙黎淑, 毕业于重庆邮电大学, 工程师, 硕士, 主要从事网络质量管理。