



基于可解释机器学习模型的5G下行空口关键性能研究

林劫

(中国电信股份有限公司福州分公司, 福建 福州 350001)

摘要: 小区下行有效吞吐率和用户面时延是5G空口的重要性能指标。为研究并解释其背后的影响因素, 使用局部线性森林结合Shapley值法进行预测和解释。实验结果显示, 相比传统方法, 该方法具有较高的预测准确性。在此基础上通过计算各影响因素对结果的贡献并分析其影响变化趋势, 找出问题小区对应的问题因素。最后, 通过贡献重要性排序, 推演重要因素改变对结果的影响, 实现更加全面理解和分析性能指标的目的。

关键词: 有效吞吐率; 用户面时延; 局部线性森林; Shapley值

中图分类号: TN915

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025122

Study on the key performance of 5G downlink air interface based on an explainable machine learning model

LIN Jie

Fuzhou Branch of China Telecom Co., Ltd., Fuzhou 350001, China

Abstract: The cell downlink effective throughput and user plane latency are important performance indicators of the 5G system's air interface. To study the impact factors behind them, the integration of Shapley value with local linear forest model was used to make predictions with explanations. The experiment results show that compared to traditional method, this method achieves a better predict precision. Based on this, by calculating each factor's contribution towards the results and analyzing the corresponding trends, a problematic factor for each problematic cell was found. Finally, by sorting the importance of each factor's contribution, changes in results were inferred by changes in important factors, thus more comprehensive understanding and analysis of the performance indicators were achieved.

Key words: effective throughput, user plane latency, local linear forest, Shapley value

0 引言

近年来, 5G网络迎来了用户数量和流量的双重爆发式增长。据统计, 截至2024年11月

月底, 全国5G用户数量已突破10亿大关^[1]。这一规模庞大的用户群体既展示了5G的广泛应用, 也为网络运营维护带来了全新的挑战与要求。常见数据业务性能需求见表1。在端到端分析中,

这些要求可以粗略分解为无线侧和非无线侧，而无线空口侧的“变数”往往最大。因此，深入分析这些要求背后的无线空口侧相关因素及其影响大小可以为网络规划与优化提供重要依据。

表 1 常见数据业务性能需求

业务	带宽/(Mbit·s ⁻¹)	时延/ms
网页浏览	> 5	< 40
4K 视频	> 50	< 50
在线游戏	> 10	< 30
文件下载	> 100	< 100

在各类无线问题分析方法中，性能指标一直是重要抓手。对于速率、时延这类结果变量，传统方法可以将其与相关的性能指标进行关联，并分析其时序变化或找到 TopN 问题小区进行优化。然而该方法也存在如下不足：(1) 性能指标之间可能存在相关性，不易分离和研究单个指标对结果影响的变化趋势；(2) 多个性能指标对结果都有影响，不易按照影响大小排序；(3) 如果需要将结果提升 X 个百分点，不易明确需要将哪些指标提升 Y 个百分点来完成目标。

针对上述问题，本文展示一种可解释性机器学习方法并将其应用于 5G 小区下行有效吞吐率和下行用户面时延的研究中。具体而言，将小区有效吞吐率和时延分别作为结果变量，通过理论分析找出关键影响因素的性能指标作为特征变量并使用局部线性森林 (local linear forest, LLF)^[2] 进行建模，再用 Shapley 值法^[3]对模型结果进行解释，分离这些性能指标各自对结果的贡献，评估贡献大小和变化趋势并据此推算重要性能指标变化对结果的综合影响。

1 相关工作

现有的速率、时延研究首先是基于其产生机制进行的，例如，文献[4]研究了 5G 容量计算的 4 种方式，文献[5]研究了增强移动宽带 (enhanced mobile broadband, eMBB) 空口时延影响因素。机制研究往往更注重理论推导，在现实场景中，多用户时用户行为的随机性、基站的调度策略等也会对结果产生影响。文献[6]研究了实际双频网络中多层次的 5G 速率影响因素并基于专家经验给出了优化方法。专家经验在网络优化方面直接而有效，但无法提供所有信息，如较难定量描述各因素对结果影响的变化趋势，而趋势或相关性研究正是机器学习所擅长的。本文使用的森林模型结合 Shapley 值的研究方法在多个领域均有应用，如医学^[7]、气象学^[8]。在移动网络方面，文献[9]使用基于 Shap 的降维聚类技术对问题进行分类。

基于以上分析可见机制研究、专家经验和机器学习方法各有优点，本文结合三者，首先通过理论研究提取最具相关性的性能指标作为特征变量，既能避免虚假导致的伪回归问题，又能提高分析效率和可靠性，之后再专家经验对建模结果展示的趋势进行分析并据此提出优化意见。

2 建模流程

2.1 特征选择

本文研究的 5G 小区为开启多用户多入多出 (multi-user multiple-input multiple-output, MU-MIMO) 能力的宏基站，基础信息见表 2。

小区下行有效吞吐率由式 (1) 给出：

表 2 基础信息

制式	频段	带宽	基站类别	时隙周期	时隙配置	子载波间隔	通道数	MU-MIMO 能力	载波聚合
TDD	3.5 GHz 或 3.6 GHz	100 MHz (273 RB)	宏基站	2.5 ms 双周期	DDDSU DDSUU	30 kHz	32TR 或 64TR	使能	无



$$\text{Thr} = \frac{\text{TB}}{t} \quad (1)$$

其中, TB是下行混合自动重传请求 (hybrid automatic repeat request, HARQ) 反馈确认字符 (acknowledgement, ACK) 即正确发送的传输块 (transport block, TB) 总和, t 为小区内媒体访问控制 (media access control, MAC) 层下行有数据发送的总时长。式 (1) 计算的是多用户时的平均结果, 而单用户的吞吐量由式 (2) 近似给出^[4]:

$$T = N \times (1 - \text{Oh}) \times R \times O \times S \quad (2)$$

其中, N 是小区分配的时频资源, 计算1s内用于下行的资源粒子 (resource element, RE) 总数, 由制式、子载波间隔及基站调度决定, Oh 是用于开销的资源比例, $1 - \text{Oh}$ 表示用于数据传输的资源比例, R 和 O 分别是每个RE对应的目标码率和调制阶数, 由调制编码方案 (modulation and coding scheme, MCS) 决定, 而MCS主要由信道条件决定, S 是空分复用流数 (RANK)。

由式 (1) 可见重传率会对结果有影响, 再结合式 (2), 共梳理出8个性能指标作为模型的特征变量, 特征变量见表3。PDCCH (physical downlink control channel) 是物理下行控制信道, CCE (control channel element) 是PDCCH的基本组成单位, DCI (downlink control information) 是下行控制信息, PRB (physical resource block)

是物理资源块。

小区空口下行用户面时延由式 (3) 计算:

$$t_{\text{avg}} = \frac{t_{\text{airtotal}}}{p_{\text{airtotal}}} \quad (3)$$

其中, t_{airtotal} 为小区下行业务数据包空口成功发送总时延, p_{airtotal} 为小区下行业务数据包空口成功发送总包数。一个下行数据包的传输流程主要包括: 基站处理、下行数据发送、终端接收并对齐时隙边界、终端处理、终端等待上行时隙到来并反馈ACK/NACK、上行数据发送、基站接收并对齐时隙边界。其中每一步均会引入时延, 若存在误码需要重传, 则上述一些步骤会再多一次, 引入额外时延。下行数据包处理流程如图1所示。



图1 下行数据包处理流程

表3中的特征变量包括负荷、信道条件、重传等因素, 预计这些对时延都会有影响, 因此可复用表3中的特征变量作为研究时延模型的输入变量。另外基站配置的终端非连续接收 (discontinuous reception, DRX) 可能引入额外的调度时延, 这部分配置全网统一, 其影响预计可通过用户数和负荷间接反映。

表3 特征变量

序号	变量名	计算方式	说明
1	下行平均MCS	MCS表 ^[10] 各索引的平均值	对应式 (2) 中 R 和 O
2	接入距离	上行提前量 (timing advance, TA) 值的平均值并根据频段换算	与大尺度衰落相关, 影响式 (2) 中的 R 和 O
3	下行平均激活用户数	小区内下行缓存中有数据的用户数的平均值	式 (1) 为多用户场景
4	平均使用的PDCCH CCE个数	小区内各个类别DCI使用的PDCCH CCE个数之和的平均值	式 (2) 中 Oh 的主要来源
5	下行平均RANK	下行传输不同RANK值的平均值	对应式 (2) 中的 S
6	MU-MIMO 平均配对层数	下行PRB不同配对层数的平均值	配对用户通过共享相同频谱资源提高整体吞吐量
7	下行PRB平均利用率	下行平均使用的PRB个数/下行平均可用PRB个数	表示小区整体带宽利用情况
8	下行HARQ重传率	下行重传的TB数/(下行重传+新传的TB数)	式 (1) 分子考虑成功传输, 而分母考虑所有传输

2.2 数据集选择和处理

本文选择5 985个5G小区并连续取7日15 min级别的性能数据,即每个小区包含 $7 \times 24 \times 4 = 672$ 条15 min数据。之后每个小区再从672条数据中找出平均激活用户数最大的那一条,重新组合成5 985条最大激活用户数数据集输入模型。取最大激活用户数原因在于其能减少随机性差异,使结果更符合统计特征。同时,用户数是唯一“外部”特征,基本不受其他特征影响。另外,5G的设计初衷之一就是服务更多用户,这样做也更能体现5G空口的“真实能力”。

2.3 模型输入

本文使用LLF进行建模和预测。LLF结合了随机森林(random forest, RF)^[11]的灵活性和局部线性回归的平滑特性,不仅能够处理数据中的非线性关系,在处理具有平滑趋势特征和稀疏数据时,LLF相比RF还能表现出独特的优势,特别适用于本文数据集的情况。

使用R包grf^[12]配置LLF,所有超参数均采用默认设置。数据集的80%作为训练集、20%作为测试集。模型的预测效果通过均方根误差(root mean square error, RMSE)和中位数绝对百分比误差(median absolute percentage error, MdAPE)两个指标进行评估,如式(4)和式(5)所示:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$\text{MdAPE} = \text{Median} \left(\left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \right) \quad (5)$$

这两个指标分别从绝对误差和相对误差的角度评估模型性能。选择MdAPE而非传统的平均绝对百分比误差(mean absolute percentage error, MAPE)原因为当实际值接近于0时(如数据集中的超低吞吐率小区),MAPE会产生异常大的误差值(因分母接近0),从而对整体评估结果产生过度影响。

LLF内部决策过程难以直观理解和解释,这

种“黑盒”特性限制了其在一些需要高度可解释性的领域的应用。为了克服这一局限,本文引入Shapley值法对预测结果进行解释。Shapley值最初用于合作博弈中的利益分配问题^[13],将其用于机器学习的核心思想是通过计算特征在所有可能的组合顺序下对模型预测的边际贡献来量化单个特征的贡献。通过计算,模型预测可以表示为一个加性计算式,即模型预测等于基准值和所有特征的Shapley值相加,如式(6)所示^[3]:

$$f(x^*) = \phi_0 + \sum_{j=1}^M \phi_j \quad (6)$$

其中, $f(x^*)$ 是模型对待解释样本 x^* 的预测值。 ϕ_0 是预测集中所有样本结果变量的均值(基准值), ϕ_j 是第 j 个特征的Shapley值,表示该变量对预测结果的贡献, M 是特征总数。式(6)使得模型预测具有类似线性回归中回归系数的解释性。

使用R包shapr^[14]计算Shapley值,Shapley值可正可负,正值表示该特征对预测结果有正向贡献,负值则表示有负向贡献。Shapley值计算结果样例如图2所示,展示了一个例子,其中LLF预测的下行吞吐率为185.2 Mbit/s,横坐标表示Shapley值,每个特征的Shapley值所表示的贡献如图2所示,贡献值总和=预测值-基准值。

通过这种方式,可以明确每个特征对模型预测结果的影响程度和方向。因此,Shapley值提供了一种系统化的方法来解释模型的预测。

在计算每个样本每个特征的Shapley值后,按照式(7)计算每个特征对结果的贡献重要性评分:

$$\text{SI}_j = \frac{1}{n} \sum_{i=1}^n |\phi_{ji}| \quad (7)$$

式(7)表示特征 j 的评分等于所有样本的特征 j 的Shapley值的绝对值之和的平均值。这样无论是正贡献还是负贡献,只要绝对值大,最终重要性就会大。

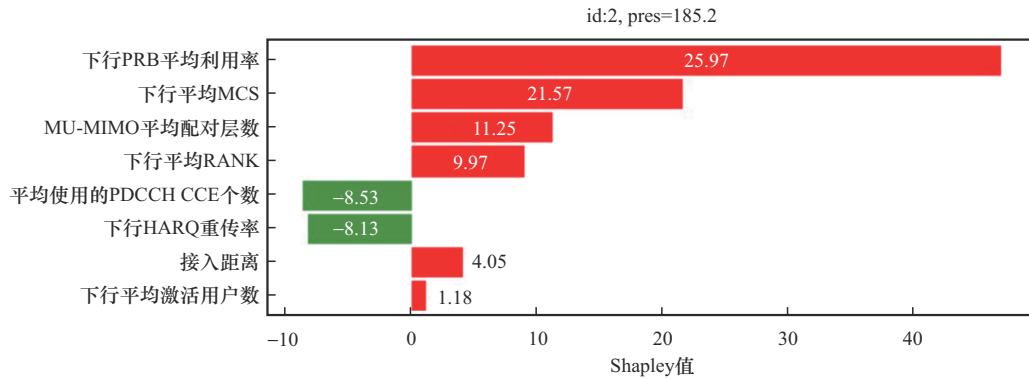


图2 Shapley值计算结果样例

3 结果分析

3.1 建模结果

不同数据集的LLF和RF建模结果见表4。

LLF和RF模型的绝对误差和相对误差均较低且相对均值预测均有巨大提升。LLF在有效吞吐率的预测表现明显优于RF，而在空口时延的预测表现与RF基本持平。因此后续Shapley值的计算基于LLF的预测结果。

3.2 Shapley值分析

计算测试集所有样本的Shapley值并按特征汇总，再按照式(7)计算每个特征的重要性得分，最后按照总和占比进行归一化，不同特征的重要性占比见表5。对于有效吞吐率，PRB利用率和MCS重要性得分占比最高，RANK占比最低。可能原因为虽然理论上RANK1和RANK4相差4倍的吞吐率，但实际的室外宏基站RANK多集中在2附近，发散程度低，例如，与MCS比较可见高平均MCS未必会对应高平均RANK，MCS和RANK变动关系如图3所示。因此导致RANK变动产生的平均贡献要低。

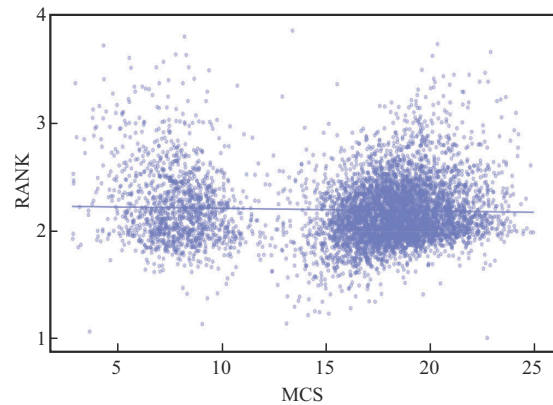


图3 MCS和RANK变动关系

对于时延，HARQ重传率的重要性得分最高且远大于其他特征，而接入距离的重要性得分最低。接入距离的变化可以通过TA补偿，因此理论上接入距离对时延无影响。

影响有效吞吐率前5特征贡献的变化趋势如图4所示，影响用户面时延前5特征的贡献变化趋势如图5所示，X轴表示每个特征的实际取值，Y轴表示该特征的Shapley值。为观察下行平均激活用户数变化对结果的影响，每张图使用激活用户数着色，颜色越深表示该样本用户数越多。

图4显示PRB利用率、MCS、激活用户数、

表4 不同数据集的LLF和RF建模结果

序号	结果	LLF-RMSE	LLF-MdAPE	RF-RMSE	RF-MdAPE	Avg-RMSE	Avg-MdAPE
1	下行有效吞吐率	16.23 Mbit/s	7.09%	22.56 Mbit/s	9.44%	48.17 Mbit/s	23.87%
2	下行空口时延	0.22 ms	7.55%	0.22 ms	7.5%	0.5ms	19.38%

表5 不同特征的重要性占比

重要性	下行 PRB 平均利用率	下行平均 MCS	下行 HARQ 重传率	下行平均激活用户数	平均使用的 PDCCH CCE 个数	MU-MIMO 平均配对层数	接入距离	下行平均 RANK
吞吐率	27.89%	19.61%	14.72%	12.56%	8.00%	6.75%	5.83%	4.64%
时延	6.85%	8.35%	48.76%	4.62%	11.30%	5.77%	4.32%	10.03%

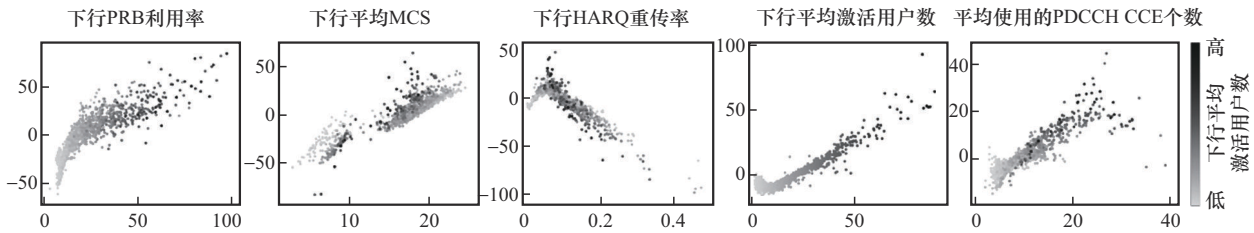


图4 影响有效吞吐率前5特征贡献的变化趋势

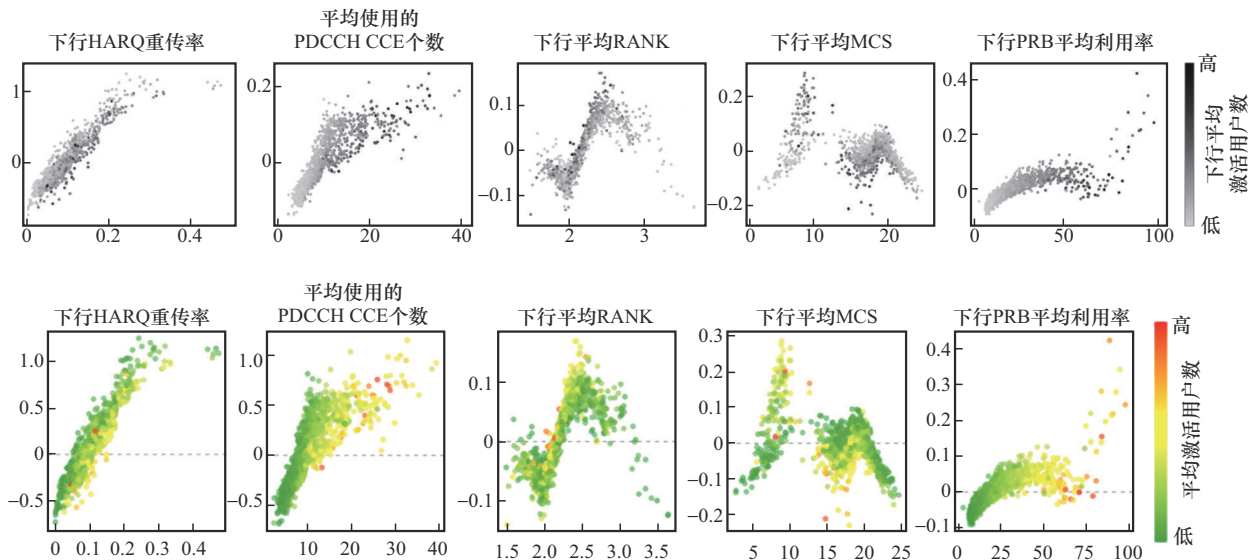


图5 影响用户面时延前5特征贡献的变化趋势

使用的PDCCH CCE个数的取值越大对有效吞吐率的贡献越大,相反HARQ重传率取值越大,负面影响越大。另外在高用户数时,特征贡献的斜率要大于低用户数时的情况。例如,MCS的计算结果表明在无线环境良好时,追求更高的用户数能带来更大的小区吞吐量,而在无线环境较差时,需要改善无线环境,此时更高的用户数更容易导致小区吞吐量变低。由于目前5G网络整体PRB利用率依然偏低,从重要性排序看,在当前阶段更应该有针对性地对低PRB利用率小区进行排查,解决除负荷因素外,导致低

PRB利用率的隐患和问题,如小区接入或覆盖问题、隐性故障等。在此基础上还应同步关注网络优化,尤其是与MCS强相关的MR、CQI等指标优化,为未来网络能承载更高的流量和用户数提供有利条件。

图5显示HARQ重传率、使用的PDCCH CCE个数取值越大对时延的贡献越大,其中HARQ重传率贡献的线性关系最明显,影响绝对值最大。RANK和MCS的贡献在两端为负贡献而在中间大部分为正贡献,可能的因为他们所代表的在无线环境较好时能降低时延,而在无线环



境较差时系统通过降流、降阶等保证可靠传输从而时延也不会增大太多。PRB利用率在达到75%之前基本对时延无影响,超过75%之后对时延有明显影响。

基于以上分析,可确认问题小区及对应的问题特征。根据第三代合作伙伴计划(3rd Generation Partnership Project, 3GPP) 37910自评估报告^[15],在30 kHz、10%初传错误概率时时延的典型值为1.46 ms,取此作为时延问题小区的门限,吞吐率依据经验取低于100 Mbit/s作为问题小区门限。每个小区对于时延取贡献最大的特征作为问题特征,对于吞吐率取贡献最小的特征作为问题特征。问题特征在问题小区中的分布见表6。表6中的分布与表5中的重要性大致匹配,因此在优化吞吐率或时延问题小区时可有两种思路:(1)直接针对问题特征进行优化;(2)若问题特征与结果相关性较低或实施较困难,如RANK之于时延,则依然可按照表5的特征重要性排序优化。

3.3 推演

根据以上结果做一个保守推演,假设小区的关键影响特征取值从所有取值中的0.25分位提升至中位数,同时假设其他特征保持不变推演结果的变化。对于有效吞吐率,选择PRB利用率和MCS两个特征,选取分位数附近10个点的均值作为分位数的贡献,计算显示两个特征的提升能引起有效吞吐率合计提升24.4 Mbit/s。对于空口用户面时延,选取HARQ重传率一个特征,计算显示其分位数由0.25提升至中位数时,时延劣化0.11 ms。

4 结束语

本文使用局部线性森林对小区下行有效吞吐率及下行空口用户面时延进行建模,并结合Shapley值法详细讨论了各相关性能指标对结果的贡献变化趋势和重要性排序。本文研究也可看作对传统网络运营中基于专家经验确定性能指标优化门限的一种补充。未来可纳入更多适合端到端分析的特征进行建模,将本文方法扩展至端到端分析,从而更全面地评估各项指标。

参考文献:

- [1] 工业和信息化部. 2024年前11个月通信业经济运行情况[Z]. 2024.
MIIT. Economic operation of the telecommunications industry in the first 11 months of 2024[Z]. 2024
- [2] FRIEDBERG R, TIBSHIRANI J, ATHEY S, et al. Local linear forests[J]. Journal of Computational and Graphical Statistics, 2021, 30(2): 503-517.
- [3] AAS K, JULUM M, LØLAND A. Explaining individual predictions when features are dependent: more accurate approximations to Shapley values[J]. Artificial Intelligence, 2021, 298: 103502.
- [4] 汤建东,肖清华. 5G容量能力分析[J]. 邮电设计技术, 2020(3): 50-53.
TANG J D, XIAO Q H. Analysis of 5G system capacity ability[J]. Designing Techniques of Posts and Telecommunications, 2020(3): 50-53.
- [5] 魏垚,熊尚坤. 5G eMBB空口时延分析与增强技术研究[J]. 移动通信, 2019, 43(9): 42-46.
WEI Y, XIONG S K. Analysis on 5G eMBB uu interface latency and study on enhancement technique[J]. Mobile Communications, 2019, 43(9): 42-46.
- [6] 王瑞,乔辉,高超,等. 5G多频网用户速率优化研究[J]. 现代信息科技, 2025, 9(1): 8-12.
WANG R, QIAO H, GAO C, et al. Research on user rate opti-

表6 问题特征在问题小区中的分布

重要性	下行PRB平均利用率	下行平均MCS	下行HARQ重传率	下行平均激活用户数	平均使用的PDCCH CCE个数	MU-MIMO平均配对层数	接入距离	下行平均RANK
吞吐率<100 Mbit/s	51.12%	20.97%	15.36%	2.81%	0.56%	4.12%	1.31%	3.75%
时延>1.46 ms	2.12%	7.00%	59.45%	1.14%	12.54%	4.07%	1.63%	12.05%

- mization of 5G multi-frequency network[J]. *Modern Information Technology*, 2025, 9(1): 8-12.
- [7] CHOI J H, CHOI Y, LEE K S, et al. Explainable model using Shapley additive explanations approach on wound infection after wide soft tissue sarcoma resection: “big data” analysis based on health insurance review and assessment service hub[J]. *Medicina*, 2024, 60(2): 327.
- [8] HE Z Y, YANG Y J, FANG R Z, et al. Integration of Shapley additive explanations with random forest model for quantitative precipitation estimation of mesoscale convective systems[J]. *Frontiers in Environmental Science*, 2023, 10: 1057081.
- [9] CILÍNIO M, PEREIRA M, DUARTE D, et al. Explainable fault analysis in mobile networks: a SHAP-based supervised clustering approach[C]//*Proceedings of the 2023 16th International Conference on Signal Processing and Communication System (ICSPCS)*. Piscataway: IEEE Press, 2023: 1-9.
- [10] 3GPP. Physical layer procedures for data(Release 15): TS 38.214[S]. 2019. .
- [11] BREIMAN L. Random forests[J]. *Machine Learning*, 2019.
- [12] TIBSHIRANI J, ATHEY S, FRIEDBERG R, et al. Generalized random forests[EB]. 2024.
- [13] SHAPLEY LS. A value for n-person games[M]//*Contributions to the theory of games II*. Princeton: Princeton University Press, 1953: 307-317.
- [14] JULLUM M, OLSEN L H B, REDELMEIER A, et al. Prediction explanation with dependence-aware Shapley values[EB]. 2024.
- [15] 3GPP. Study on self evaluation towards IMT-2020 submission (Release 16): TS 37910[S]. 2019.

[作者简介]



林劫 (1998-), 男, 中国电信股份有限公司福州分公司工程师, 主要研究方向为无线网络优化。