

基于机器学习算法的 5G重点业务KQI和KPI关系研究

Research on Relationship
Between KQI and KPI for
Key 5G Services Based on
Machine Learning Algorithms

任小强¹,罗志文²,陈苗苗¹,王春佳¹(1.西南交通大学希望学院,四川成都610400;2.中国移动通信集团甘肃有限公司,甘肃兰州730000)

Ren Xiaoqiang¹, Luo Zhiwen², Chen Miaomiao¹, Wang Chunjia¹(1. Southwest Jiaotong University Hope College, Chengdu 610400, China; 2. China Mobile Group Gansu Co., Ltd., Lanzhou 730000, China)

摘要:

近年来,多媒体内容成为移动网络用户最重要的业务之一,用户更加关注端到端的5G网络用户业务感知质量。应用机器学习算法,构建5G网络中端到端视频业务和文件下载业务的感知质量模型,探索关键质量指标(KQI)与关键性能指标(KPI)的关系。仿真结果显示,该评估方法对提升5G网络运维的质量、实时性和准确性具有指导意义。

关键词:

关键质量指标;5G;关键性能指标;机器学习
doi:10.12045/j.issn.1007-3043.2022.10.006
文章编号:1007-3043(2022)10-0028-04
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:

In recent years, multimedia content has become one of the most important services for mobile network users, and users pay more attention to end-to-end user service perception quality. It applies machine learning algorithms to evaluate the perceived quality model of end-to-end video services and file download services in 5G networks, and explores the relationship between key quality indicators (KQI) and the underlying KPI indicators. According to the simulation results, this evaluation method can improve the quality, real-time and accuracy of 5G network operation and maintenance.

Keywords:

KQI; 5G; KPI; Machine learning

引用格式:任小强,罗志文,陈苗苗,等.基于机器学习算法的5G重点业务KQI和KPI关系研究[J].邮电设计技术,2022(10):28-31.

0 引言

服务质量(QoS)根据关键性能指标(KPI)来定义,可反映网络质量,如丢包、延迟和抖动的情况。体验质量(QoE)则表示用户主观感知的质量,用关键质量指标(KQI)表示,例如“非常好”“好”“差”。KPI不是面向应用服务的,没有考虑用户业务感知,只能反映部分网络性能。即使KPI正常时,也存在用户业务感知

差的情况。5G网络采用网络切片技术,旨在提供具有不同业务需求的端到端应用。预计未来几年,5G网络业务感知问题主要依靠KQI指标优化定界。由于KQI指标涉及网元多,数据流程长,端到端影响因素众多等问题,运维人员主要靠专家经验数据判断,准确定位问题原因较难。随着计算机处理性能的不不断提升,机器学习算法得到了较好发展,例如线性回归(LR)、逐步线性回归(SWLR)、支持向量回归(SVR)、随机森林回归(RFR)和决策树回归(DTR)等机器学习算法在行业都得到了较好应用^[1-2]。为了改善用户业务感知,提升网络运行性能,本文提出基于机器学习算法的KQI与KPI指标关系分析模型,重点评估5G网络中端到端视频业务和文件下载业务的业务感知质量模型,探索关键质量指标(KQI)与底层KPI指标的关系。

基金项目:教育部2021年第2批产学研合作协同育人项目《Python3入门机器学习》课程改革与实践(202102041003);中国电子劳动学会“产教融合、校企合作”教育教学改革课题研究项目(Ciel2021271)

收稿日期:2022-08-26

1 算法总体框架

QoS通过KPI表示,QoE通过KQI表示,最终用户所体验的KQI可以由KPI通过复杂的映射关系来确定^[3]。本文选择机器学习算法完成KPI到KQI指标的映射分析。总体建模框架如图1所示,包含3个阶段:数据收集阶段、训练阶段和应用阶段。这种模型框架允许使用直接测量的KPI指标(如DT测试、UE信令跟踪)或基于OTT收集的数据集来训练模型,运营商可以通过调整配置参数来优化网络,该模型甚至可以预测出未知的业务感知指标。

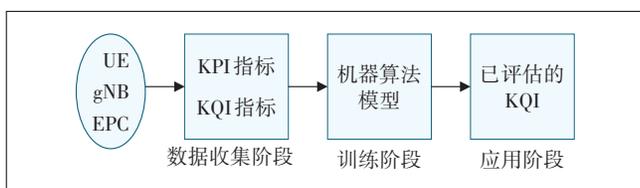


图1 算法总体框架

测量数据集作为模型的输入,来源于无线侧KPI参数、无线基站配置参数、DT测试数据和采集的KQI指标数据。此外,网络和配置参数以及其他数据集,也包括用户设备(UE)和应用程序层服务所在的服务器之间路径上的中间网络的信息,如核心网、传输网、Internet骨干网和应用程序提供者的网络。

训练阶段将前面收集的数据集合作为模型的输入,将KQI与其他指标关联起来。这里选择监督回归技术对KQI与KPI的关系进行建模分析。为了防止过拟合问题,这里使用经典的K-Fold交叉验证算法,它将原始数据分成K组(K-Fold),将每个子集数据分别做一次验证集,其余的K-1组子集数据作为训练集,这样会得到K个模型。这K个模型分别在验证集中评估结果,最后对误差MSE(Mean Squared Error)加权平均就得到交叉验证误差。为了评价模型质量,引入决定系数 R^2 来判断回归方程拟合的程度,该系数表示目标变量的预测值和实际值之间的相关程度平方的百分比,是评价模型质量的常用系数之一。 R^2 为1表示该模型可以完美的预测数据, R^2 为0表示完全无法预测数据。同时采用RMSE(均方根误差)衡量预测值与真实值之间的偏差,RMSE对数据中的异常值较为敏感,如果得到的RMSE较小,则可以用它建立预测模型,否则重新建模。

在应用阶段,模型将网络的数据配置和性能作为输入,并获得这些条件下预测的业务感知KQI指标。

在此阶段,该模型还考虑了额外获取数据集的可能性,以重新训练模型,处理之前无法获得的数据。

2 算法模型设计

2.1 测试网络环境

为了收集测试数据,利用中国移动某省的5G实验网络,包括10个5G宏站,5个5G室分站点,1套包含AMF、SMF、UPF和计费网元的网元结构。所有参数设置可以通过iManager U2000客户端进行自助配置。数据收集部分包括1个测试终端UE、pioneer测试软件、GPS和1个应用服务器组成。

在数据收集阶段,pioneer测试软件收集UE在不同参数配置下的训练数据集,测试开始时,UE收集网络无线KPI指标,启动业务服务时,pioneer测试软件收集业务感知指标KQI。这里主要分析了文件下载业务和视频业务。文件下载业务的KQI为文件传输延迟(IFTD)、文件传输平均吞吐量(FATH)、总文件传输延迟(TFTD)^[4]。视频业务的KQI指标包括视频流启动延迟(VSSD),即终端对视频流的初始请求到视频可以开始播放的瞬时的延迟;视频流吞吐量(VSTH),即视频流数据在完整播放期间的平均下载速度;视频流失帧频率、失帧次数和失帧时间(VSS)。

在培训阶段,使用Matlab编写ML建模模块,对结果进行处理并建立模型。该模块在远程客户端执行,通过SSH(Secure Shell)接口从终端接收特定无线网络配置参数和该网络配置下的KQI指标。

2.2 KQI指标建模

为了实现KQI指标评估,初始值参数为参考信号接收功率(RSRP),参考信号接收质量(RSRQ),接收信号强度指示器(RSSI),可用BW和网络负载这些指标当做自变量。测试UE捕获的KQI数据被认为是要构建模型的因变量。为了评估测试终端UE在文件传输业务下的KQI,把每个文件的大小也作为自变量。使用9000个测试数据集,包含9种文件大小(1KB、10KB、100KB、500KB、1MB、5MB、10MB、20MB、100MB)、4种可用带宽(5、10、15、20MHz)、3种网络负载(空负载、低负载和满负载),为这些参数的所有可能组合收集了KPI和KQI。另外,在不同网络条件下收集了2400个视频业务样本数据,每个视频长度为60min,形成360P、720P和1080P3种不同分辨率的视频。

评估FTP业务的KQI与前面描述的KQI相同。而对于视频业务,评估的KQI为VSSD和VSTH。由于在

实验过程中产生的失帧数量并不显著, VSSD 指标没有被考虑。这里选择缓存视频流达到规定分辨率所需的最大时间(VTMaxQ)作为一个KQI指标^[5-6]。

在训练阶段评估了5种主要的机器学习算法: 线性回归(LR)、逐步线性回归(SW-LR)、支持向量回归(SVR)、随机森林回归(RFR)和决策树回归(DTR)。LR是最经典的回归技术, 它定义了一个因变量和几个自变量之间的线性关系。SW-LR是LR的改进形式, 它利用逐步回归算法筛选并剔除引起多重共线性的变量, 得到最优变量集合。支持向量机是一种用于回归的自适应向量机。DTR通过跟踪决策树结构中从初始节点到包含预测的最终节点的分支来预测输入数据。RFR是由多棵回归树加权组合成的预测模型。

3 算法结果分析

图2是5种回归技术按照K-Fold交叉验证算法进行分析的结果, 对于文件传输业务, RFR和DTR回归算法的相关性最好, TFTD和FATH的 R^2 值接近1, 可以根据TFTD和FATH指标优化客户感知。对于IFTD, 所有回归技术的 R^2 值都接近于零, 这个指标不能用于衡量用户的业务感知, 对于视频业务, RFR、DTR、SVR分析的VSTH和VTMaxQ指标相关性较高, 相关性系数总体大于0.8, 但是VSSD的相关性系数小于0.6, 不能作为无线网络优化重点参数。

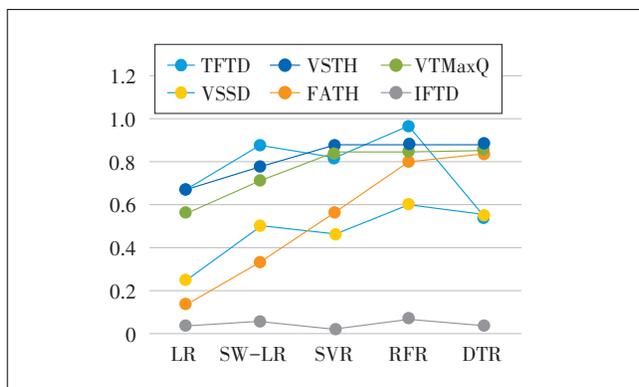


图2 算法分析结果

为了评估每种算法的处理时间, 表1比较了每个机器学习算法和经过训练得到KQI指标所需的执行时间。

对于FTP服务, 在回归过程中, 变量的重要性是决定哪个预测因子对结果有更大意义的关键因素。决策树模型给出了量化每个变量重要性的简单方法, 图3显示了使用决策树模型分析得到的KPI指标权重。权

表1 训练得到KQI指标所需的执行时间

回归算法	TFTD/s	FATH/s	VTMaxQ/s	VSTH/s
LR	0.002 0	0.002	0.001	0.001
SW-LR	0.028 0	0.028	0.012	0.012
SVR	0.027 8	0.815	0.009	0.014
RFR	0.009 0	0.009	0.006	0.006
DTR	0.005 0	0.005	0.005	0.004

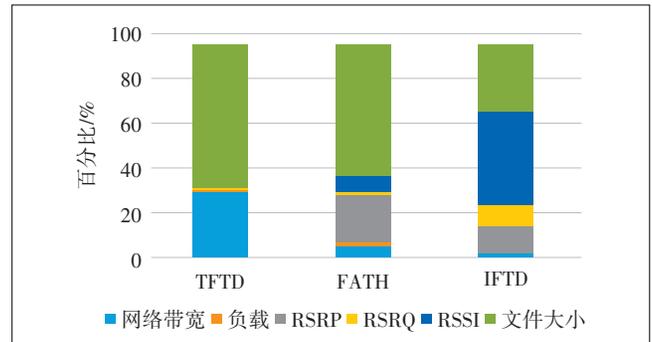


图3 KQI与对应KPI指标权重

重值越低, 表示它的重要性就越低。结果表明, 可用的BW、负载是决定TFTD的关键指标, 而RSRP和RSSI对FATH有显著影响。对于IFTD, 所有的KQI指标基本没有相关性。

在视频流业务中, SINR反映了UE和5G基站之间信道内的干扰和噪声。SINR、RSRP参数与VSTH之间应该有一个映射模型, 可以通过重点优化SINR、RSRP指标来提升视频业务的KQI指标。根据仿真数据, 预测了FATH和SINR之间的映射模型如下:

$$FATH_{PSNR} = a \log_2(SINR + b) + c \quad (1)$$

其中PSNR为视频流FATH指标的峰值信噪比。令 $I(i, j)$ 表示原始视频流图像帧的强度值, $K(i, j)$ 表示失真视频流图像帧的强度, 然后均方误差(MSE)定义如式(2)所示, 其中 m, n 表示视频图像的行数和列数。

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (2)$$

视频帧的PSNR指标定义如式(3)所示, $\text{Max}[I(i, j)]$ 为视频帧的最大强度值; 对于8位图像, 通常 $\text{Max}[I(i, j)] = 255$ 。

$$FATH_{PSNR} = 10 \lg \left\{ \frac{\text{Max}[I(i, j)]^2}{MSE} \right\} \quad (3)$$

采用曲线拟合的方法, 得到 $FATH_{PSNR}$ 与SINR的关系如图4所示, 在这种情况下, 参数 a 和 b 分别为-15.39和79.52, 可以看到拟合程度接近96%, 说明此模型是可行的, 预测和模拟结果之间的匹配验证了本文的映

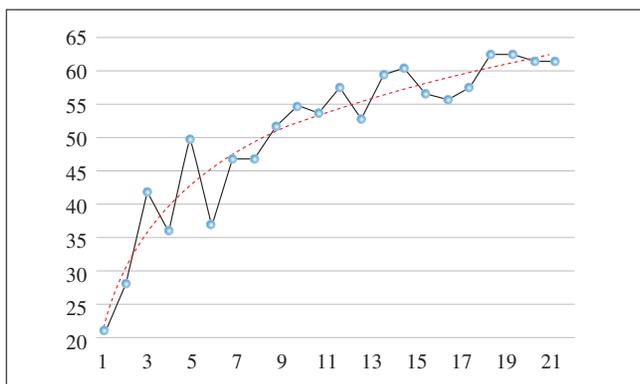


图4 FATH_{PSNR}与SINR关系

射模型。

R^2 决定系数表示回归方程的拟合程度,用于度量回归算法模型的性能。结果表明,使用完整数据集训练的模型性能优于使用部分数据集训练的模型。对于完整的数据集, R^2 的值与图4中分析结果类似,差异不大,如图5所示。然而,对于DTR和RFR回归算法来说,TFTD中使用部分数据集与完整数据集分析的差异较大,如图6所示,这是因为文件数量增加时,DTR和RFR算法的性能大大降低,从而对TFTD指标具有明显的影响。

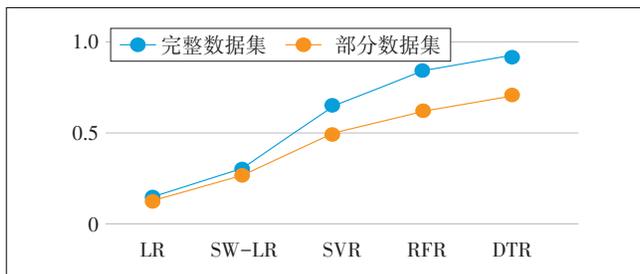


图5 数据集对VSTH的影响分析

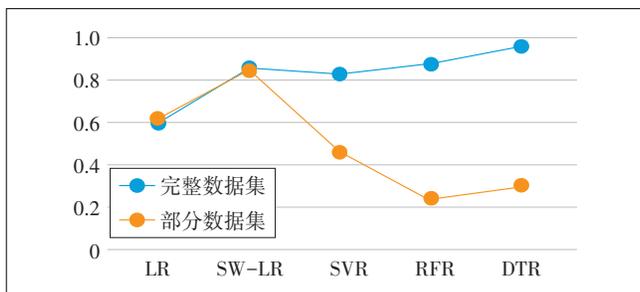


图6 数据集对TFTD的影响分析

4 应用效果分析

4.1 低速率基站优化

真实场景的测试表明,造成文件下载业务感知差

的主要因素有MR覆盖率(32%)、RRC连接平均数高(25%)和MR重叠覆盖度高(21%)等。该模型关联分析的定位准确性为83.3%,能有效分析现网小区KQI劣化的主要原因,按本算法给出的优化建议实施后,5日平均下载速率环比提升13%~20%。

4.2 视频业务优化

选取视频业务质差强相关的VSTH指标作为KQI,分析其与无线KPI的关联关系。应用本模型对498个基站进行了关联性分析。造成VSTH指标差的几个主要因素有SINR>20 dB的采样点占比(29%)、UE高发射功率占比(24%)和平均TA值(9%)等。算法自动定位VSTH指标差的主要原因是SINR<-3 dB的采样点占比过高,优化后,5日平均下行丢包率从优化前3.39%下降到3.06%。

5 结束语

本文针对5G的视频流业务和文件下载,通过机器学习算法构建2种业务的KQI指标评估模型,分析KQI指标评估模型与KPI指标的关系。网络运营商可以使用该模型来预测最终用户的QoE水平,提高运维效率。

参考文献:

- [1] AGUAYO L, FORTES S, BAENA C, et al. A multivariate time-series based approach for quality modeling in wireless networks [J]. Sensors, 2021, 21(6): 2017.
- [2] BAENA E, FORTES S, BARCO R. KQI performance evaluation of 3GPP LBT priorities for indoor unLicensed coexistence scenarios [J]. Electronics, 2020, 9(10): 1701.
- [3] 刘子可. 移动通信基站故障类型的识别[D]. 北京: 北京交通大学, 2020.
- [4] HERRERA-GARCIA A, FORTES S, BAENA E, et al. Modeling of key quality indicators for end-to-end network management: preparing for 5G [J]. IEEE Vehicular Technology Magazine, 2019, 14(4): 76-84.
- [5] 任小强, 敬嘉亮. 无线侧提升VoLTE用户语音感知的优化策略研究[J]. 邮电设计技术, 2020(4): 46-49.
- [6] 蒋仕宝, 杜翠凤, 聂丹彤. 融合CNN与LSTM的网络质量KQI数据特征提取与投诉预警[J]. 移动通信, 2020, 44(2): 69-75.

作者简介:

任小强, 毕业于成都理工大学, 讲师, 硕士, 主要从事无线网优化和数据挖掘工作; 罗志文, 毕业于电子科技大学, 工程师, 学士, 主要从事无线网优化工作; 陈苗苗, 毕业于四川大学, 助教, 硕士, 主要从事机器学习研究工作; 王春佳, 毕业于四川师范大学, 工程师, 硕士, 主要从事数据挖掘工作。