



专题：通信与 AI 融合

智能语音业务质量评估保障方案

李光宇, 梁燕萍, 余立, 王昆, 陈扬铭, 刘思佳
(中国移动通信有限公司研究院, 北京 100053)

摘要: 语音通话是电信网络下的主流业务之一, 业界主流的语音业务评估方案通过对比收/发端语音音频, 计算得到 POLQA MOS 评估业务质量。对于电信运营商, 在不侵犯用户隐私的前提下感知业务质量, 进而进行针对性的业务质量保障是重点工作之一。提出了通过数理统计及神经网络算法, 学习网络侧数据包交互信息与终端侧语音业务质量的映射关系, 形成高精度评估模型; 在此基础上, 结合无线网用户及小区级多维数据信息, 分析输出业务质差的原因及解决方案, 形成高效准确保障业务质量的系统。实际应用效果证明, 该系统可通过网络侧数据准实时完成语音业务质量评估, 并输出质差解决方案, 针对网络中的 VoLTE 语音业务进行质量保障工作。

关键词: VoLTE 业务; 质量保障; 网络智能化

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021110

AI based scheme on voice service quality evaluation and assurance

LI Guangyu, LIANG Yanping, YU Li, WANG Kun, CHEN Yangming, LIU Sijia
China Mobile Research Institute, Beijing 100053, China

Abstract: The voice service is one of the mainstream services under the telecommunications network. The mainstream voice service quality evaluation method in the industry is to compare the voice and audio of the transceiver and calculate the POLQA MOS, which is the standard benchmark for voice quality analysis. For telecom operators, perceive voice service quality efficiently without any infringement of user privacy, and carry out precise service quality assurance, is one of the key tasks. A system using mathematical statistics and neural network to learn the mapping between voice service feature data from network side and voice quality from user side to generate evaluation model with high accuracy and precision was proposed. On this basis, the system made use of data information from multiple fields including wireless network user level and cell level, to analyze reasons and work out solutions for poor service quality, so as to assure service quality efficiently and accurately. The result shows that the scheme has achieved good practical application results.

Key words: VoLTE service, quality assurance, network intelligence

1 引言

长期以来,作为运营商的自营业务,语音业务一直是运营商需要重点保障的核心业务。随着 LTE 网络、5G NR 网络的建设和发展,运营商用户数的持续增加,VoLTE 业务量也在逐年增长,业务质量保障需要投入的人力也随之增加。传统的 VoLTE 业务质量保障工作通常由用户的语音业务质差投诉触发,优化人员需要在业务质差位置附近区域进行道路测试,使用收发终端拨打语音业务并播放通用语料库,通过测试工具对比分析收发端的语料质量差异计算得到 POLQA(感知客观听力质量评估) MOS(mean opinion score)作为语音质量的评分,针对 MOS 值较低的情况进行信令分析、问题定位及解决。传统的问题排查及优化手段存在覆盖面窄、投入资源高、解决周期长的问题,且需要同时拿到收发端语料内容,无法满足运营商的日常优化需求。为了在不获取收发端语音内容的前提下进行质量评估,ITU 提出了多种通过业务性能指标分析运算得出业务质量结果的无参考方案,以 E-model 为代表,通过业务丢包率、相对时延等业务传输同步、时延引起的质量损伤作为计算因子,估算得出 MOS,但估算精度不够理想。

本文介绍了一套通过网络侧数据包传输特征数据、用户测量数据进行 VoLTE 业务质量评估及保障的技术方案,该方案基于网络中全量语音业务通信特征数据,通过人工智能算法建模并利用模型进行业务质量评估,可针对运营商全量语音业务进行高精度质量分析及质差定位解决,以满足运营商实际生产环节中的优化工作需求。该方案由网络数据采集、数据特征分析、质量评估、业务质差定位及解决 4 部分组成,目前已取得较好的保障效果。

2 语音业务质量保障方案

相对于传统的语音业务质量评估工作,本文提出的方案通过在业务特征的设计与生成、业务

特征的分析与选择、业务评估模型的搭建与优化 3 方面的重点优化与创新,大幅提升了质量评估的准确性,保证了业务质差的发现成功率。

2.1 业务特征的设计与生成

VoLTE 业务发起之前,用户会进行网络附着及 IMS 注册流程。完成注册后主被叫用户建立路由,开始业务数据传输。VoLTE 业务基于实时传输协议(real-time transport protocol, RTP)进行数据包的双向实时传输,并通过实时传输控制协议(real-time transport control protocol, RTCP)进行 QoS(quality of service,服务质量)控制。VoLTE 语音业务在数据包传输过程中,通话期每 20 ms 发送一个语音包,包括 RTP/UDP/RLC-security 压缩头;每 160 ms 生成一个 SID 静默帧,通过噪音提高用户的业务感受^[1-2]。

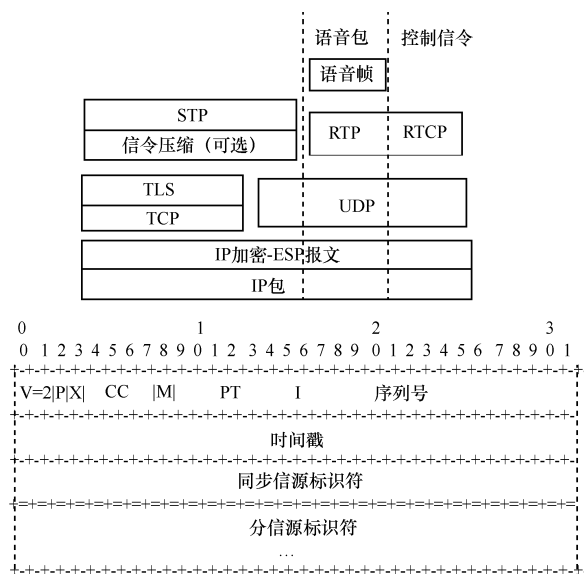


图1 VoLTE 协议栈及 RTP 包头示意图

VoLTE 协议栈及 RTP 包头结构如图 1 所示,在业务质量监控场景下,业务特征分析不能涉及用户隐私,因此针对 RTP 之下的包头信息进行分析。包头中序列号(sequence number, SN)用于标识发送的 RTP 数据包编号,在业务质量分析中可用于计算丢包情况;时间戳(timestamp)标识 RTP 数据分组中第一个字节的采样时间,可用于



计算时延及抖动信息^[3]。

RTP 本身没有提供任何机制来确保实时的传输或其他的 QoS 保证, RTCP 通过不同报文类型计算端到端丢包数、丢包率、包间抖动等指标, 可以更准确地评估语音业务质量信息^[4]。

目前, VoLTE 语音质量主要通过丢包、时延及抖动 3 类特征进行分析, 当使用业务级特征数据进行网络中全量语音质量分析时, 无法获取端到端信息, 因此无法获取业务的时延特征。使用现有方案进行业务质量评估时仍存在特征粒度较粗, 无法精细化评估秒级业务质量变化情况的弊端。以 RTP 丢包率及抖动分析为例, 目前丢包对业务质量的影响主要通过丢包率是否超过一定门限进行判断, 但针对连续丢包及间歇性丢包对业务质量的不同影响程度未做详细分析。间歇性丢包与连续丢包对比如图 2 所示, 当语音业务中出现连续丢包时, 虽然丢包率相同, 但业务质量会有较明显恶化, 现有方案无法发现问题。

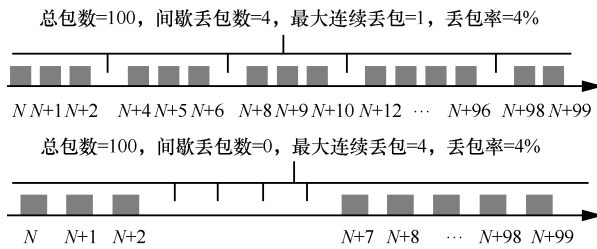


图2 间歇性丢包与连续丢包对比

同样, 现有方案针对业务抖动检测采用滑动叠加机制, 将当前抖动绝对值与前一时刻的抖动加权相加, 计算式如下:

$$J(i) = 15/16 \cdot J(i-1) + |D(i-1; i)|/16$$

$$D(i; j) = (R_j - R_i) - (S_j - S_i) = (R_j - S_j) - (R_i - S_i)$$

(1)

其中, J 代表当前抖动结果, D 表示两数据包间抖动, S 与 R 分别表示数据包的发送与接收时间。

现有方案没有考虑较长时间范围内抖动的累加效应, 且没有考虑正向抖动和负向抖动的抵消

情况。当数据包间隔在标准间隔上下波动时, 终端会通过缓冲机制进行调整, 保证业务质量不受影响; 而当数据包间隔不断增长时, 会导致业务时延逐渐增大, 影响业务体验。这种情况下, 现有方案无法区分前者与后者的差别。随机抖动与正向抖动的对比如图 3 所示, 图 3 (a) 的情况时延正常波动, 业务质量较好, 图 3 (b) 的情况时延由于抖动的正向累加不断增加, 业务质量劣化明显, 但两种情况下现有方案计算出的抖动值相同, 无法发现图 3 (b) 的劣化问题。

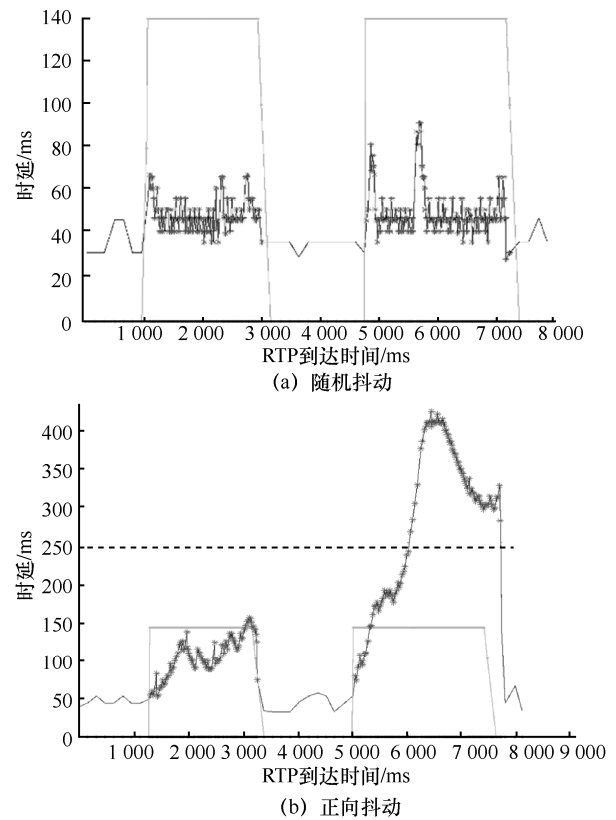


图3 随机抖动与正向抖动的对比

为了解决现有业务质量评估方案的问题, 本文介绍的方案将 RTP 包间抖动的正、负方向, 丢包连续性等问题考虑在内, 提出了 3 类新指标: 相对时延指标、连续丢包数指标以及最大包间指标。通过新指标的补充运用, 可以解决现有方案问题, 优化业务质量评估效果。

连续丢包数可以通过最大连续丢包、连续丢

包较多的问题出现次数等维度,在整体丢包率的基础上额外发现业务的局部性或突发性质量差问题。连续丢包的计算方式为:将所有收到的 RTP 数据包根据序列号从小到大排序,分别计算相邻序列号的差值,若差值大于 1,则存在连续丢包。计算式如下:

$$\begin{aligned} \text{CPL}(1) &= 0 \\ \text{CPL}(i) &= \text{SN}(i) - \text{SN}(i-1) - 1, i \geq 2 \end{aligned} \quad (2)$$

其中, CPL 代表连续丢包数, SN 代表排序后的 RTP 包序列。

相对时延代表每个数据包相对于业务过程中第一个数据包时延的变化情况,计算方式为首包时延加上当前包与首包间所有抖动的累加结果,抖动累加时考虑正向负向。计算式如下:

$$\begin{aligned} \text{RD}(1) &= 0 \\ \text{RD}(i) &= \text{RD}(i-1) + J(i), i \geq 2 \\ \text{RD}'(i) &= \text{RD}(i) - \min\{\text{RD}(i), i \in \{1, 2, \dots, N\}\}, i \geq 1 \end{aligned} \quad (3)$$

其中, RD 代表相对时延, J 代表抖动, RD' 代表校准后的相对时延,目的是避免第一个收到的包并非切片内第一个发送包,从而影响相对时延序列的情况。

在 VoLTE 业务中,通过相对时延,可以准确分辨随机抖动与连续正向抖动对通话质量的不同影响结果。以图 4 为例,图 4(a)和图 4(b)分别是平滑滤波前后的抖动情况,可以看到 1 500~3 000 ms 时间范围内是随机抖动,由图 4(c)可见相对时延计算结果在 130 ms 以内,通过图 4(d)波形图看到收发端语音无明显差别,此时通话过程未感受到质量劣化;而 5 000~6 500 ms 时间范围内发生了连续正向抖动,相对时延在 6 200 ms 后达到 200 ms 以上且持续增加,通过波形图可以看到方框标注部分有明显的收端延后情况,此时通话质量下降明显。

由此可见,新特征的设计和实现可以有效地弥补现有质量评估方案的不足,更精细地发现业

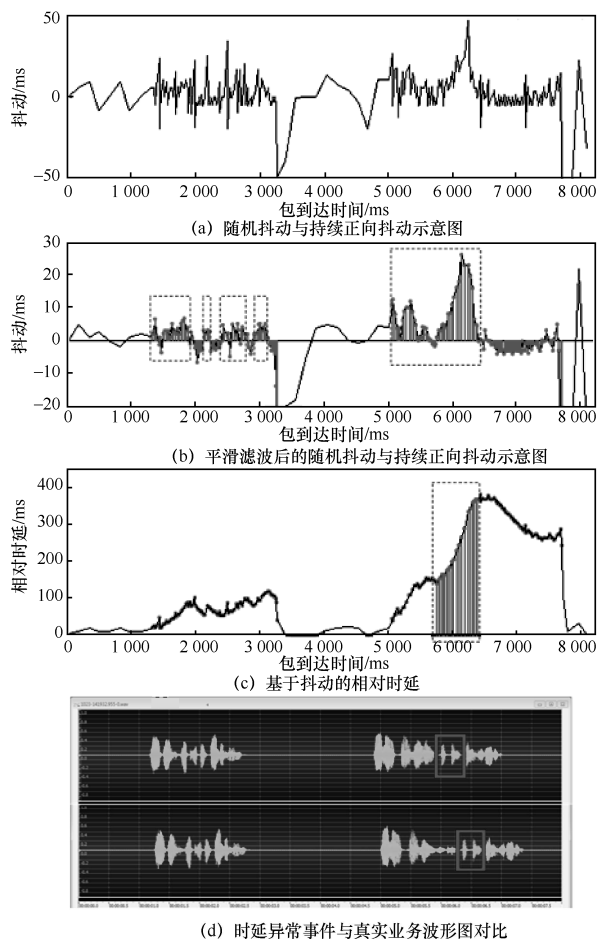


图 4 随机抖动与正向抖动的时延分析及波形图对比

务质量劣化的情况,帮助提升评估效果。基于连续丢包特征,可以设计最大连续丢包、连续丢包超过 n 个的次数等指标,而相对时延特征可以衍生得到最大相对时延、平均相对时延、非静默期相对时延等指标,这些指标均为语音业务质量评估的重要特征。

2.2 业务特征的分析与选择

当完成新特征的设计与实现后,特征分析模块针对基于业务知识设计、选出的重要特征进行重要性分析,目标是筛选其中对业务质量影响程度更高的指标并通过机器学习的手段完成特征处理,促进业务质量评估模型的精度提升。

本方案中使用的特征分析与选择方案融合了数理统计与机器学习领域常用的特征重要性分析工具。其中,数据统计方式通过绘制各个特征与



MOS 分的散点分布进行特征相关性分析^[5], 结果如图 5 所示。

可以看出, 最大相对时延及(非静默期)最大包间隔和 MOS 之间存在较明显的负相关性, 随着两类标签取值的增加, MOS 迅速下降, 符合业务逻辑。

机器学习的特征重要性分析通过决策树模型的构建及训练, 通过每个特征在质量评估过程中被选作划分点的次数来衡量特征指标的重要性^[6-7]。基于决策树模型得到的特征重要性排序如图 6 所示, 可以看出平均相对时延、最大包间隔、最大相对时延等均在预测 MOS 的模型中起到了比较重要的作用, 证明新增特征的设计具有必要性。

2.3 业务评估模型的搭建与优化

目前, 业界公认的 VoLTE 业务语音质量评估标

准是 POLQA MOS。该标准由 ITU 提出, 通过对比 VoLTE 业务过程中收/发端语音的声纹信息, 经过时间对齐、语音采样分析、感知分析等操作评估手段语音相对发端语音的劣化程度^[8]。在语音业务质量评估过程中, 本方案同样使用含有 POLQA MOS 的语音业务路测数据建立训练数据集, 针对训练数据进行特征工程及模型训练, 获取高精度模型。然后通过该模型评估每一通语音通话的业务质量。

传统的 E-model 模型^[9]在进行语音业务质量分析时, 选取丢包率、时延等指标并人为制定计算式及参数, 得到 MOS 评估结果, 评估模型相对简单粗糙。本方案使用深度学习建立模型, 该模型参数丰富, 拟合能力强, 更有利于学习业务特征数据中 RTP 丢包、连续丢包、抖动、相对

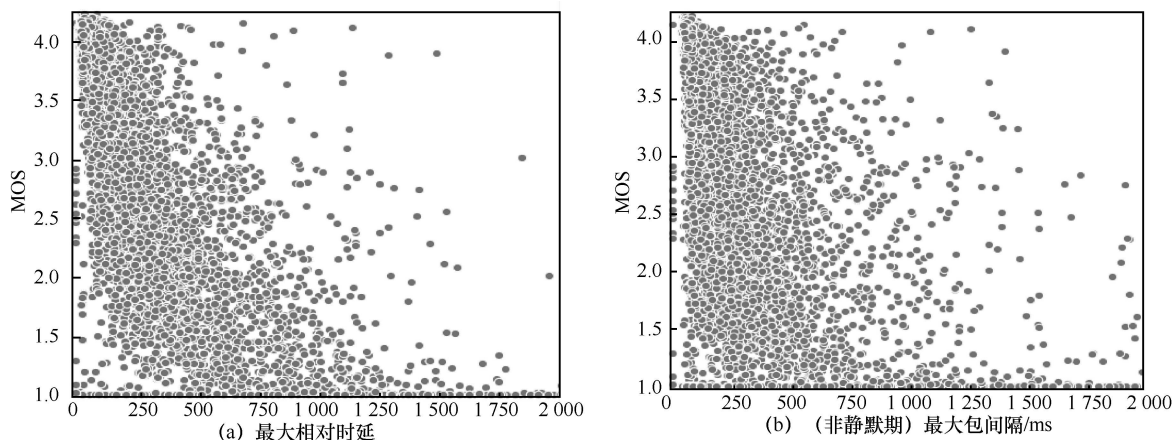


图 5 最大相对时延、(非静默期)最大包间隔与 MOS 的分布关系

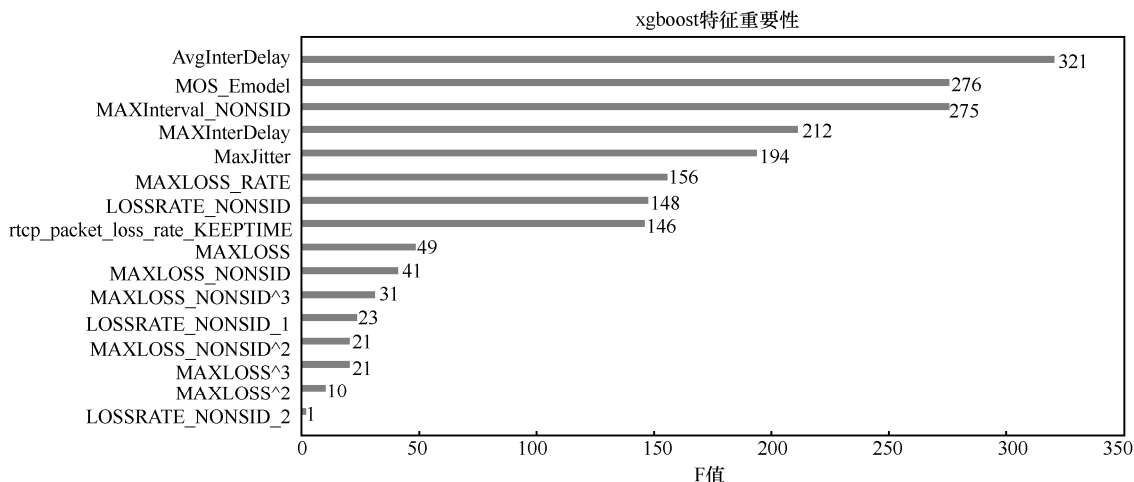


图 6 特征指标重要性排序

时延、包间隔等业务特征与 POLQA MOS 的复杂映射关系, 评估效果优于 E-model 模型。

本方案在进行质量分析时, 共有特征处理、建模优化、质量评估 3 个步骤。

(1) 特征处理指针对上文理论分析结合特征重要性分析确定的相对时延、丢包、包间隔等重要指标, 进行平方、立方、离散化等特征增强操作, 增加这些指标对最终 MOS 分析的影响程度。

(2) 建模优化指通过模型训练和参数优化, 使得模型输出尽量接近 POLQA MOS, 同时尽量全面地发现质差业务。前者的优化方式通过模型训练尽可能缩小模型输出与 POLQA MOS 的均方差, 计算式如下:

$$MSE = \frac{1}{n} \sum_{i=0}^n (MOS_{\text{predict}} - MOS_{\text{true}})^2 \quad (4)$$

其中, n 代表语音业务数据样本量, MOS_{predict} 与 MOS_{true} 分别代表模型输出结果与 POLQA MOS。

(3) 质量评估: 模型发现质差业务的能力通过质差召回率衡量, 质差召回率即模型发现的质差数量与所有业务质差数量的比值。目前, 本文认为 MOS 低于 3 分即质差业务^[10]。

通过这样的训练目标, 评估模型在整体 MOS 评估准确性与质差业务检出能力之间达到均衡。模型示意图如图 7 所示。

完成建模优化后, 模型已经可以较准确地完成 MOS 评估, 并较全面地发现业务质差。此时使

用该模型评估网络中真实发生的 VoLTE 业务, 即可得到接近于路测工具得出的 MOS, 实现大规模的业务质量分析。

3 系统应用效果

本文提出的方案除了实验室验证, 也应用在真实网络环境并取得了较好的效果。本节将介绍方案目前的应用方式及效果。

3.1 质量保障方案结构及部署方式

本文介绍的语音业务质量保障方案由网络数据采集、数据特征分析、质量评估、业务质差定位及解决 4 个模块组成, 期间的工作流程及与外部系统的对接情况如图 8 所示。

数据采集模块对接网络大数据平台, 获取网络真实产生的语音业务特征数据用于质量评估, 获取用户测量 (MR) 数据用于质差原因定位。语音业务特征数据来源于各类业务过程中的数据包传输指标以及包头信息的解析运算, 涵盖的业务特征包括基本的数据包传输时延、抖动、传输包数、丢包数及丢包率等; 同时, 经过数据采集系统的迭代升级, 目前业务特征数据中已包含第 2 节介绍的最大相对时延、最大连续丢包数等新指标, 因此适合用于业务级质量分析的特征数据。实际优化工作中发现, VoLTE 业务的质差主要是由无线网络的覆盖、干扰、移动性参数配置等问题引起的, 而 MR 数据包含无线网络中用户测量

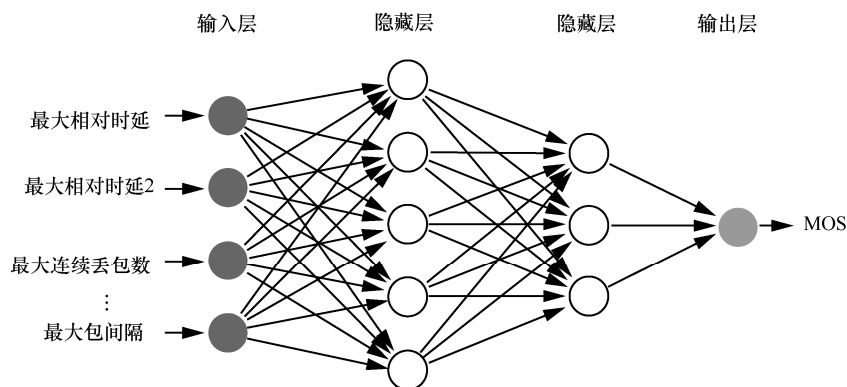


图 7 语音业务质量分析模型示意图

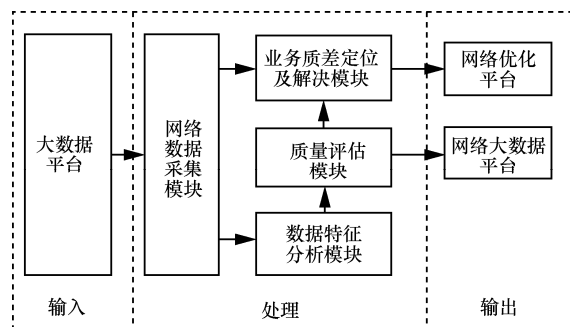


图8 保障方案功能模块结构及与外部系统的对接情况示意图

的服务小区及邻小区的信号质量，可用于覆盖及干扰分析，因此本方案通过网管系统的 MR 数据进行业务质差分析。

数据特征分析模块结合不同覆盖场景的数据特点及 VoLTE 业务流程特点，分析已有业务特征的重要性，并结合人工智能算法中的常用特征工程方案针对全量语音业务数据进行特征处理，输出质量评估模块用于语音业务 MOS 评估。

质量评估模块通过深度学习完成建模及优化后，在所有评估模型中挑选效果最优的模型对网络的语音业务单据进行推理，输出语音切片级 MOS、语音话单级 MOS，用于判断业务质量情况，并将推理结果返回至大数据平台。

业务质差定位及解决模块，通过关联分析质差业务对应的用户 MR 数据，将业务质差原因归纳为上下行弱覆盖、干扰、邻区漏配等 9 类问题，并针对每类问题制定解决方案，作为网络问题排查解决建议^[11]。目前，该模块与网络优化平台对接，在故障工单派发过程中将定位的质差原因及

解决方案以备注信息的形式随单下发，帮助一线网优人员迅速定位解决问题，保障用户业务体验。

3.2 质量保障方案应用效果

目前，本文提出的保障方案已在北京、浙江、湖南等多地应用，面向全省、市的全量 VoLTE 语音业务进行质量评估及质差保障。目前该方案每日评估的语音业务超过 3 亿条，经过实地测试验证，质量评估的效果相对于传统的 E-model 模型有明显提升，本文方案中语音业务质量评估效果与传统 E-model 评估结果对比见表 1。

表 1 本文方案中语音业务质量评估效果与传统 E-model 评估结果对比

统计项目	智能语音业务 质量分析	E-model (KQI 规范算法)
语音业务切片个数	3 068	3 068
MSE	0.09	0.20
MOS 评估准确率	84.7%	72.1%
质差业务召回率	92.6%	45.7%

业务质量较差部分的 MOS 评估结果对比如图 9 所示，可以看到本文介绍的方案评估的 MOS (MOS_AI) 相对于 E-model 模型的评估结果 (MOS_E-model) 更贴近于真实 MOS。表 1 中统计结果显示本文提出的保障方案对 MOS 的预测结果与 POLQA MOS 的误差相对 E-model 预测结果更小。目前，本方案将 MOS 预测结果与真实结果误差 ± 0.25 范围内认定为准确预测，表 1 结果显示业务保障方案评估准确率达到 84.7%，明显优

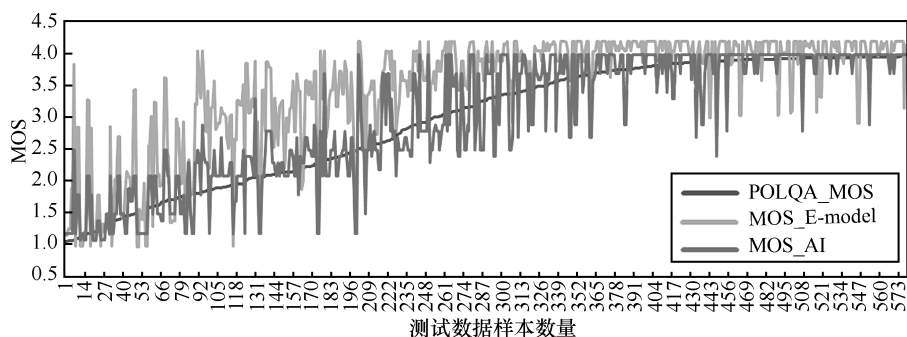


图9 MOS 评估结果对比（节选）

于 E-model 模型的 72.1%。

完成质量评估后, 质差定位及解决模块通过工单派发的形式已经辅助完成数千个质差小区的问题解决, 对应小区的业务质差数量明显降低。

4 结束语

本文介绍了一套基于网络侧获取的语音业务数据进行业务质量保障的方案, 该方案通过在业务特征的设计与生成、特征的重要性分析与选择、业务质量分析建模与优化工作上的创新与改进, 达到了比传统无参考评估方案有较大提升的业务质量评估效果, 同时针对业务质差进行原因定位及方案输出, 提升了实际生产中问题排查解决工作的效率, 改善了用户语音业务质差感知。

该方案的设计思路同样适用于 VoNR、VoIP 等语音业务, 随着 5G NR 用户数的快速增长及对应数据采集系统功能的逐渐完善, 后续也会将本文方案结合 VoNR 等业务特征进行适配优化, 使能更广泛的语音业务质量保障。

参考文献:

- [1] 3GPP. Multimedia subsystem (IMS): TS 23.228IP[S]. 2020
- [2] 杨黎明. 基于 RTP 协议的实时语音传输研究与实现[D]. 武汉: 华中科技大学, 2007.
YANG L M. Research and implementation of RTP-based real-time voice transmission[D]. Wuhan: Huazhong University of Science and Technology, 2007.
- [3] 田国庆, 唐红. 基于 RTP 协议的语音通信的 QoS 的设计与实现[C]//2004 计算机应用技术交流会议论文集. 重庆, 2004: 41-42, 49.
TIAN G Q, TANG H. Design and implementation of QoS for voice communication based on RTP protocol[C]//Proceedings of 2004 Computer Application Technology Exchange Conference. Chongqing, 2004: 41-42, 49.
- [4] RTP: A transport protocol for real-time applications; H. Schulzrinne et al; network working group request for comments: 3550[Z]. 2003.
- [5] 谢娟英, 吴肇中, 郑清泉. 基于信息增益与皮尔森相关系数的 2D 自适应特征选择算法[J]. 陕西师范大学学报(自然科学版), 2020, 48(6): 69-81.
XIE J Y, WU Z Z, ZHENG Q Q. An adaptive 2D feature selection algorithm based on information gain and Pearson correlation coefficient[J]. Journal of Shaanxi Normal University (Natural Science Edition), 2020, 48(6): 69-81.
- [6] 苏映雪. 特征选择算法研究[D]. 长沙: 国防科学技术大学, 2006.
SU Y X. Research on feature selection methods[D]. Changsha: National University of Defense Technology, 2006.
- [7] 叶菲, 罗景青. 基于信息增益和相关性的判定树特征选择算法[J]. 现代防御技术, 2009, 37(1): 87-89, 123.
YE F, LUO J Q. Decision tree feature selection algorithm based on information gain and correlation[J]. Modern Defence Technology, 2009, 37(1): 87-89, 123.
- [8] ITU-T. Perceptual objective listening quality assessment: P.863[S]. 2011.
- [9] 许可. VoLTE 语音质量评价方法及测试方案[J]. 电信网技术, 2014(5): 81-84.
XU K. VoLTE speech quality evaluation method and test scheme[J]. Telecommunications Network Technology, 2014(5): 81-84.
- [10] 曾俊. 路测 MOS3.0 占比提升研究[J]. 信息通信, 2017, 30(11): 250-251.
ZENG J. Study on road test MOS3.0 proportion improvement[J]. Information & Communications, 2017, 30(11): 250-251.
- [11] 刘韬, 吴涛, 王斌. XDR 与 MR 数据关联与应用[J]. 电信科学, 2019, 35(4): 120-130.
LIU T, WU T, WANG B. Association and application in XDR and MR data[J]. Telecommunications Science, 2019, 35(4): 120-130.

[作者简介]



李光宇 (1989-), 男, 中国移动通信有限公司研究院人工智能与智慧运营中心工程师, 主要研究方向为网络智能化技术及算法。

梁燕萍 (1986-), 女, 中国移动通信有限公司研究院人工智能与智慧运营中心工程师, 主要研究方向为网络智能化技术。

余立 (1981-), 男, 中国移动通信有限公司研究院人工智能与智慧运营中心副总经理、高级工程师, 主要研究方向为前沿移动通信技术、网络智能化、大数据和 IT 技术。

王昆 (1988-), 男, 现就职于中国移动通信有限公司研究院人工智能与智慧运营中心, 主要研究方向为网络智能化方案。

陈扬铭 (1994-), 女, 现就职于中国移动通信有限公司研究院人工智能与智慧运营中心, 主要研究方向为网络智能化产品方案设计、开发及管理。

刘思佳 (1994-), 女, 中国移动通信有限公司研究院人工智能与智慧运营中心工程师, 主要研究方向为网络智能化产品算法。