基于机器学习分析 VoLTE 视频通话质量的研究及应用

钟其柱

（中国移动通信集团广东有限公司中山分公司，广东 中山 528400）

摘　要：针对目前评估 VoLTE 视频通话质量的方法的缺点，提出了一种基于机器学习和网络指标参数的 VoLTE 视频通话质量评估方法。首先，采集解码核心网的网络参数指标数据，并进行预处理；然后，选取用于 VoLTE 视频通话质量评估的关键特征，并通过对比选取合适的机器学习算法，构建 VoLTE 视频质量评估的评估模型，从而实现不依赖于测试环境和原始视频的 VoLTE 视频通话质量实时评估。通过对从 XDR（用户话单）数据中提取的特征指标数据进行预处理研究，解决了特征指标的标准化问题，便于指标特征输入评估模型；通过特征工程选出评估 VoLTE 视频通话的关键特征，减少特征维数，从而降低了算法的复杂度；同时采用先进的机器学习技术保证和提升算法的评估准确性。

关键词：机器学习；梯度提升树；VoLTE；视频通话质量

中图分类号：TN919.8

文献标识码：A

doi: 10.11959/j.issn.1000−0801.2020021

Research and application of VoLTE video call quality based on machine learning

ZHONG Qizhu

Zhongshan Branch of China Mobile Group Guangdong Corporation, Zhongshan 528400, China

Abstract: To overcome the shortcomings of current methods for evaluating VoLTE video call quality, a method for evaluating VoLTE video call quality without reference based on machine learning and network index parameters was proposed. Firstly, the network parameters of the decoding core network were collected and preprocessed; then, the key features for VoLTE video call quality assessment were selected, and a reference-free evaluation model for VoLTE video quality assessment was constructed by comparing and selecting appropriate machine learning algorithms, so as to achieve real-time VoLTE video call quality assessment independent of the test environment and the original video. By researching the preprocessing of feature index data extracted from XDR data, the standardization of feature index was solved, and the evaluation model of feature input was convenient; the key features of VoLTE video call were selected and evaluated by feature engineering, which reduced the feature dimension and the complexity of the algorithm; at the same time, advanced machine learning technology was adopted to ensure and enhance the algorithm assessment accuracy.

Key words: machine learning, gradient boosting decison tree, voice over long-term evolution, video call quality

收稿日期：2019−04−28；修回日期：2020−01−08

1　引言

VoLTE（voice over LTE）是在日益成熟的长期演进（long term evolution，LTE）网络基础上所架构的、基于 IP 多媒体子系统（IP multimedia subsystem，IMS）的 TD-LTE（time division-long term evolution）语音的目标解决方案[1]。VoLTE 不仅能提供更好的语音质量，也能提供更好的视频质量[2]。但由于移动互联网的出现，微信和 QQ 走在了视频通话应用的前沿，凭借 OTT（over the top）的技术优势很快占据了视频通话市场，VoLTE 面临前所未有的压力和挑战。随着移动 4G 用户数量与日俱增，广大用户对 VoLTE 业务质量的要求和需求也越来越高。因此，VoLTE 视频用户感知的分析和优化，成为关注的热点，而 VoLTE 视频通话质量的评估则是分析和优化的基础。

现有的视频质量评估方法基本都是一些特定环境下的评估方法。这些方法，要么只关注视频本身，要么只关注网络指标，要么评估指标参数获取困难，某些方法还对评估条件进行了一些假设，给评估方法设置了假设限制。参考文献[3]提供了一种通过在 VoLTE 终端抓取分组进行视频通话用户体验质量评估方法。参考文献[4]提供了一种通过获取视频帧时长间隔和清晰度进行视频通话质量评估方法。参考文献[5]公开了一种基于神经网络的 VoIP（voice over internet protocol）无参考视频通信质量客观评价方法，主要通过升级 VoIP 软件收集网络指标参数并用于实时通信场景下的 VoIP 视频通信质量监测，能够对 VoIP 视频质量做出准确且实时的评价。

从上述分析可知，目前的视频质量评估方法基本都是在特定环境下的评估方法。为了突破现有移动多媒体通信中的 VoLTE 视频通话质量评估方法存在的局限性，本文尝试运用机器学习的方法，通过分析网络传输数据中的网络指标和视频控制参数，对 VoLTE 视频通话质量进行评估。

2　方案概述

VoLTE 视频通话时，手机终端通过麦克风采集音频、摄像头采集图像，分别经过编码器编码成音频和视频流，再通过调制解调器（Modem）发往网络，经过无线网、核心网，再传送到被叫手机终端，分别经过解码器还原成音频和视频。基本流程如图 1 所示。

如图 1 所示，本方案主要通过研究 eNode B 与核心网之间传输的数据分组（包括 SIP（session initiation protocol）、RTP（real-time transport protocol）和 RTCP（RTP control protocol）数据分组）中的网络参数和指标，借助机器学习建模，最终得到一种评估 VoLTE 视频通话质量的方法。

该方法中通过对比多种模型算法，支持向量、随机森林、梯度提升树等最终确定梯度提升树作为 VoLTE 视频通话质量的评估模型。梯度提升树模型详细算法在下文中将具体介绍，而整个模型的建立流程如图 2 所示。

本方案主要从以下几点来说明该方法的创新性、实用性以及准确性。

（1）创新性

方案创新性地使用机器学习的方法，结合数据分组、缺失值填充、特征转换及数据转换多种算法，对 VoLTE 视频的质量进行评估，保证了评估结果更准确，同时很大程度上降低了人工的参与，提高效率。

（2）实用性

方案针对全网进行数据采集，编解码获取用户的参数指标，实现了对全网、区域、用户多维的质量评估，能够准确反映网络状况和辅助发现问题，对网络维护运营有很大的帮助。

（3）准确性

通过对用户真实感知的学习预测全网用户感知，从用户自身出发；同时充分考虑了视频通话的全流程，包含了多数的视频失真类型，极大地提高了质量评估准确性。

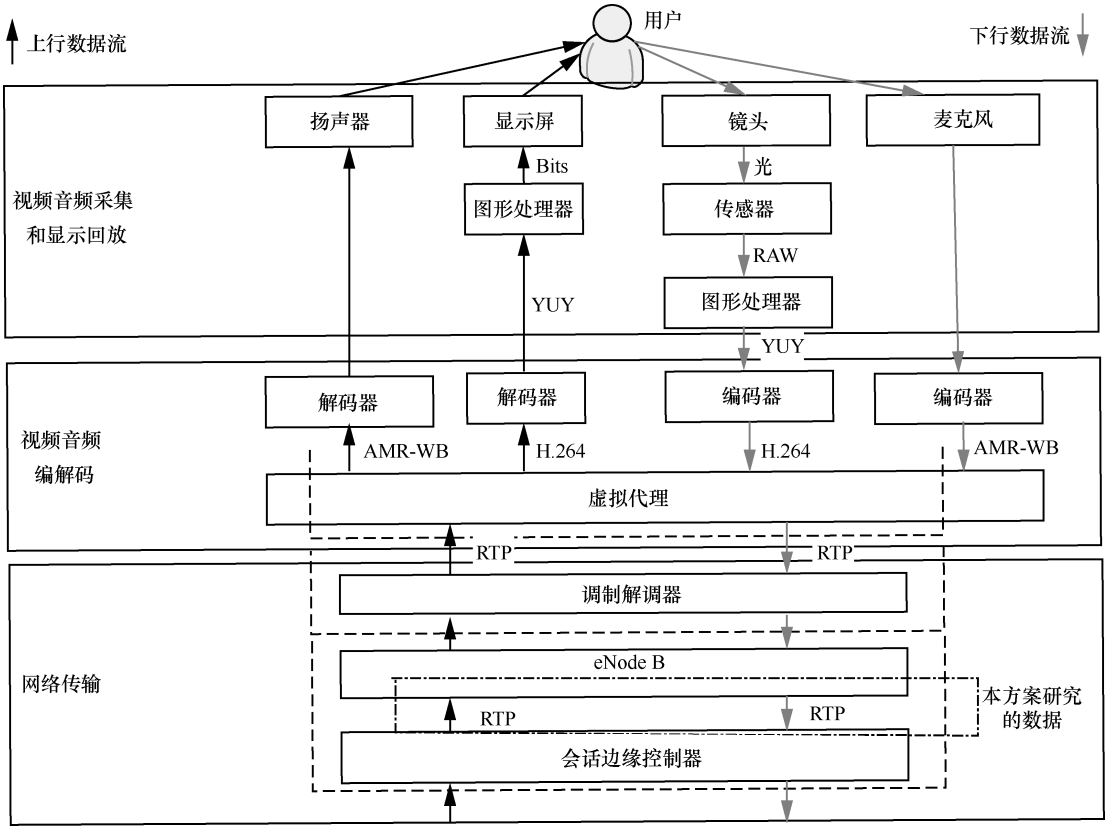

图 1　VoLTE 视频通话数据流走向

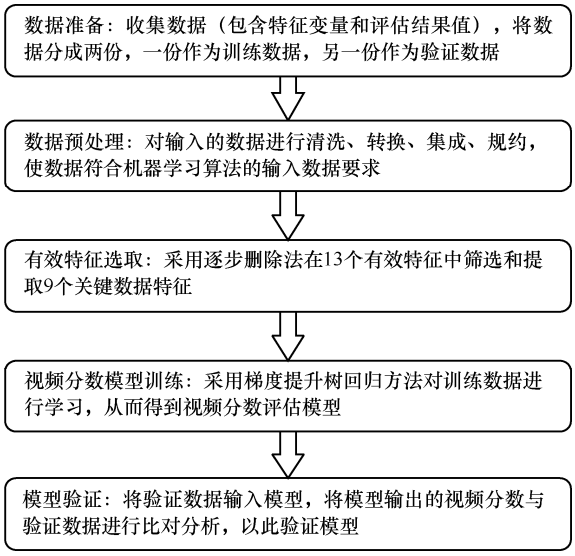

图 2　评估模型建立流程

3　模型建立

3.1　数据收集

原始评估数据是核心网 S1u 口（LTE 网络中基站和用户之间的数据传输接口）分光采集并解码所得，训练集数据则通过拨测产生。

此处将原始数据集存储为 csv 文件格式，作为训练数据集，包含了特征变量和预测值。数据格式范例见表 1。

最后一列是需要预测的分数。经过试验测试发现，vtMOS 值在 3.0 以上是可以接受的视频，没有卡顿等问题；低于 3.0 则认为视频质量较差，会引起用户体验差、反感，甚至引起投诉问题。

3.2　数据预处理

任何机器学习过程都无法省略对数据的预处理，数据预处理直接决定了算法模型精准度能达到的“天花板”，而算法模型的优化只能决定是否抵达“天花板”[6]。数据预处理是对数据的清洗、转换、集成、规约，从而形成机器学习算法能训练的数据。数据预处理流程如图 3 所示。

表 1　数据输入格式范例

音频编码	音频码率/(kbit·s^{-1})	音频 IP 分组丢失率	音频时延/ms	视频编码	视频分辨率	视频 IP 码率/(bit·s^{-1})	视频帧率/(f·s^{-1})	视频 IP 分组丢失率	视频时延/ms	音频 RTP 分组丢失率	视频 RTP 分组丢失率	视频 RTP 码率/(bit·s^{-1})	vtMOS
AMR-WB	23.85	0	23.35	H.264	VGA	521 257	30.04	0	52.35	0	0	521 214	2.76
AMR-WB	23.85	0	23.23	H.264	VGA	521 225	30.04	0	49.23	0	0	521 413	2.77
AMR-WB	23.85	0	25.67	H.264	VGA	521 633	30.03	0	68.67	0	0	521 014	2.72
AMR-WB	23.85	0.01	45.56	H.264	VGA	217 681	12.10	0.63	341.56	0.01	0.67	684 073	0.99
AMR-WB	23.85	0	20.89	H.264	VGA	815 485	21.24	0	30.24	0.01	0.03	776 154	3.32
AMR-WB	23.85	0	18.88	H.264	VGA	810 223	21.27	0	20.11	0	0.01	771 365	4.02

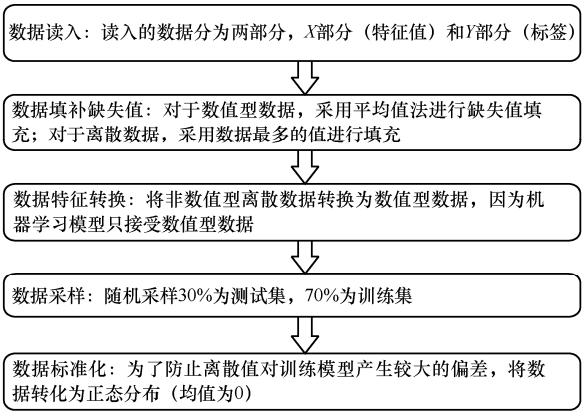

图 3　数据预处理流程

（1）数据读入

将收集的数据读入，并将数据中的参数指标部分作为特征值，放入特征训练集，将结果值放入结果训练集。

（2）数据填补缺失值

读入数据后会发现，有些指标值存在缺失值的情况，具体见表 2。

表 2 中方框内单元格字段存在值缺失，采用“平均值”法进行填充，也就是计算这一列的平均值，填充缺失值。填充后结果见表 3。

从表 3 可以看到，原来缺失的音频 RTP 分组丢失率指标值填充成了该列的平均值 0.000 2；原来缺失的视频 RTP 分组丢失率指标值填充为 0.000 7；视频 RTP 码率指标值填充为 772 290.4。

通过使用填充算法对空值指标项进行非 0 填充，避免了空值无法被训练模型学习的问题，同

表 2　缺失部分数据的数据

音频编码	音频码率/(kbit·s^{-1})	音频时延/ms	视频编码	视频分辨率	视频 IP 码率/(bit·s^{-1})	视频帧率/(f·s^{-1})	视频 IP 分组丢失率	视频时延/ms	音频 RTP 分组丢失率	视频 RTP 分组丢失率	视频 RTP 码率/(bit·s^{-1})	vtMOS
AMR-WB	23.85	0.000 1	H.264	VGA	321 979	27.37	0.000 1	38.6				2.32
AMR-WB	23.85	0.000 5	H.264	VGA	322 158	22.22	0.003 2	46.46				2.25
AMR-WB	23.85	0	H.264	VGA	637 498	26.24	0.000 3	47.9				3
AMR-WB	23.85	0	H.264	VGA	318 787	24.63	0.000 4	47.21				2.32
AMR-WB	23.85	0.000 1	H.264	VGA	365 991	26.79	0.000 6	36.06				2.44

表 3　缺失数据填充后的数据

音频编码	音频码率/(kbit·s^{-1})	音频时延/ms	视频编码	视频分辨率	视频 IP 码率/(bit·s^{-1})	视频帧率/(f·s^{-1})	视频 IP 分组丢失率	视频时延/ms	音频 RTP 分组丢失率	视频 RTP 分组丢失率	视频 RTP 码率/(bit·s^{-1})	vtMOS
AMR-WB	23.85	0.000 1	H.264	VGA	321 979	27.37	0.000 1	38.6	0.000 2	0.000 7	772 290.4	2.32
AMR-WB	23.85	0.000 5	H.264	VGA	322 158	22.22	0.003 2	46.46	0.000 2	0.000 7	772 290.4	2.25
AMR-WB	23.85	0	H.264	VGA	637 498	26.24	0.000 3	47.9	0.000 2	0.000 7	772 290.4	3
AMR-WB	23.85	0	H.264	VGA	318 787	24.63	0.000 4	47.21	0.000 2	0.000 7	772 290.4	2.32
AMR-WB	23.85	0.000 1	H.264	VGA	365 991	26.79	0.000 6	36.06	0.000 2	0.000 7	772 290.4	2.44

表 4　音频编码、视频编码、视频分辨率值为字符值的原始数据

音频编码	音频码率/(kbit·s^{-1})	音频 IP 分组丢失率	音频时延/ms	视频编码	视频分辨率	视频 IP 码率/(bit·s^{-1})	视频 IP 帧率/（f·s^{-1}）	视频 IP 分组丢失率
AMR-WB	23.9	0.000 05	32	H.265	VGA	460 235	13.38	0.000 5
AMR-WB	23.9	0	31.91	H.265	VGA	463 141	14.8	0.000 8
AMR-WB	23.9	0.000 09	27.45	H.264	VGA	321 979	27.37	0.000 1
AMR-WB	23.9	0.000 5	31.37	H.264	VGA	322 158	22.22	0.003 2
AMR-WB	23.9	0.000 02	29.53	H.264	VGA	637 498	26.24	0.000 3
AMR-WB	23.9	0	28.82	H.264	VGA	318 787	24.63	0.000 4

时又比 0 填充更接近网络实际情况。

（3）数据特征转换

音频编码、视频编码、视频分辨率值为字符值的原始数据见表 4，音频编码、视频编码、视频分辨率这 3 个指标为字符型值，非数值，不能被模型所接受，因此，需要进行特征转换。

采用 one-hot 编码的方式对离散数据特征进行转换，转换格式见表 5。例如视频编码有 H.264、H.265 两种数据值，需要转换为两个字段。

表 5　视频编码的字符型值转换为数字编码

IS_H.264	IS_H.265
1	0
0	1

同样地，对视频分辨率、音频编码等非数值型数据进行数据特征转换，转换成机器学习可以识别的数字型数据。

（4）数据采样

为了能更好地验证训练模型，将采集的原始学习数据拆分为训练集、测试集两部分，其中训练集数据只用于训练模型，而测试集只用于验证模型训练的好坏。随机抽取 75%的原始数据作为训练集，剩下的 25%作为测试集，于是数据被划分为两份：训练集、测试集，而每一份数据里面的内容没有任何变化。

（5）数据标准化

经过处理的数据符合标准正态分布，即均值为 0、标准差为 1，其转化函数为：$x^* = \frac{x-\mu}{\sigma}$，其中：$x^*$是标准化后的值，$x$ 是原始值，μ 是平均值，σ 是标准差。

数据标准化的意义在于，将所有数据都缩放在一个特定的区间，以防少量太大、太小的样本对整体训练产生极大的影响，大大影响模型的泛化能力。

视频 IP 码率原始值见表 6，对于视频 IP 码率指标，其最大值和最小值之间相差较大，不利于模型稳定训练，需要进行标准化处理。

首先计算视频IP码率指标值的平均值为435 682，标准差为 126 501.89。然后对视频 IP 码率指标值进行标准化处理，标准化后结果见表 7。

从表 7 可以看到，原始数据被缩放到了一个固定的区间范围（即标准化处理后列数据均值为 0、标准差为 1）。

3.3　有效特征选取

特征工程就是从原始数据中找出模型所需要的特征数据的过程，它的目的就是获取最少的训练数据特征，使得机器学习模型的输入维度最少，同时结果又逼近实际。

本步骤是对训练集的特征进行选择，筛选出对预测结果有用的数据特征，采用“相关系数”“均方差”“平均绝对误差”3 种指标对特征的筛选结果进行评估。

VoLTE 视频通话的原始特征有音频编码、视

表 6　视频 IP 码率原始值

音频编码	音频码率/(kbit·s^{-1})	音频 IP 分组丢失率	音频时延/ms	视频编码	视频分辨率	视频 IP 码率/(bit·s^{-1})	视频帧率/(f·s^{-1})	视频 IP 分组丢失率	视频时延/ms	音频 RTP 分组丢失率	视频 RTP 分组丢失率	视频 RTP 码率/(bit·s^{-1})
AMR-WB	23.90	0.000 1	32.00	H.265	VGA	460 235	13.38	0.001	30.90	0.035 8	0.656 7	204 311.0
AMR-WB	23.90	0	31.91	H.265	VGA	463 141	14.80	0.001	30.73	0.020 5	0.586 4	242 380.0
AMR-WB	23.90	0.000 1	27.45	H.264	VGA	321 979	27.37	0	38.60	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0.000 5	31.37	H.264	VGA	322 158	22.22	0.003	46.46	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0	29.53	H.264	VGA	637 498	26.24	0	47.90	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0	28.82	H.264	VGA	318 787	24.63	0	47.21	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0.000 1	29.58	H.264	VGA	365 991	26.79	0.001	36.06	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0.000 1	28.16	H.264	VGA	595 667	22.28	0.002	44.00	0.000 2	0.000 7	772 290.4

表 7　视频 IP 码率数值标准化处理后数值

音频编码	音频码率/(kbit·s^{-1})	音频 IP 分组丢失率	音频时延/ms	视频编码	视频分辨率	视频 IP 码率/(bit·s^{-1})	视频帧率/(f·s^{-1})	视频 IP 分组丢失率	视频时延	音频 RTP 分组丢失率	视频 RTP 分组丢失率	视频 RTP 码率/(bit·s^{-1})
AMR-WB	23.90	0.000 1	32.00	H.265	VGA	0.19	13.38	0.001	30.90	0.035 8	0.656 7	204 311.0
AMR-WB	23.90	0	31.91	H.265	VGA	0.22	14.80	0.001	30.73	0.020 5	0.586 4	242 380.0
AMR-WB	23.90	0.000 1	27.45	H.264	VGA	−0.90	27.37	0	38.60	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0.000 5	31.37	H.264	VGA	−0.90	22.22	0.003	46.46	0.0002	0.000 7	772 290.4
AMR-WB	23.90	0	29.53	H.264	VGA	1.60	26.24	0	47.90	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0	28.82	H.264	VGA	−0.92	24.63	0	47.21	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0.000 1	29.58	H.264	VGA	−0.55	26.79	0.001	36.06	0.000 2	0.000 7	772 290.4
AMR-WB	23.90	0.000 1	28.16	H.264	VGA	1.26	22.28	0.002	44.00	0.000 2	0.000 7	772 290.4

频编码等。下面采用固定模型为“梯度提升树”的情况，使用逐步删除法进行有效特征筛选，具体见表 8。

一开始将全部的特征作为训练数据集的输入，得到一个预测结果，然后逐个删除特征，观察模型预测结果变化。通过实验发现，当删除音频 IP 分组丢失率、音频时延、视频 IP 码率、视频帧率、视频 IP 分组丢失率这几个特征中的一个时，相关系数大幅降低，均方差和平均绝对误差大幅升高，模型的效果明显降低；而当删除视频时延、音频 RTP 分组丢失率、视频 RTP 分组丢失率、视频 RTP 码率中的一个时，相关系数降低微小，均方差和平均绝对误差增幅微小，模型效果稍有降低，但不明显。所以由此推断：音频 IP 分组丢失率、音频时延、视频 IP 码率、视频帧率、视频 IP 分组丢失率这几个特征对模型效果是最重要的，视频时延、音频 RTP 分组丢失率、视频 RTP 分组丢失率、视频 RTP 码率对模型效果是次重要的。

3.4　视频分数模型训练

视频分数预测属于机器学习里面的回归模型。在机器学习算法中，常用的回归模型算法有：线性回归、决策树回归、支持向量回归、K 最近邻回归、随机森林、极端随机森林、梯度提升树等模型[7]。以下实验是固定特征组合为音频编码、音频码率、音频 IP 分组丢失率、音频时延、视频编码、视频分辨率、视频 IP 码率、视频帧率、视频 IP 分组丢失率、视频时延、音频 RTP 分组丢失

表 8　特征筛选对比结果

特征组合	相关系数	均方差	平均绝对误差
音频编码、音频码率、音频 IP 分组丢失率、音频时延、视频编码、视频分辨率、视频 IP 码率、视频帧率、视频 IP 分组丢失率、视频时延、音频 RTP 分组丢失率、视频 RTP 分组丢失率、视频 RTP 码率	0.94	0.05	0.138
音频 IP 分组丢失率、音频时延、视频编码、视频分辨率、视频 IP 码率、视频帧率、视频 IP 分组丢失率、视频时延、音频 RTP 分组丢失率、视频 RTP 分组丢失率、视频 RTP 码率	0.93	0.06	0.144
音频时延、视频编码、视频分辨率、视频 IP 码率、视频帧率、视频 IP 分组丢失率、视频时延、音频 RTP 分组丢失率、视频 RTP 分组丢失率、视频 RTP 码率	0.64	1.99	1.45
音频编码、音频码率音频 IP 分组丢失率、视频编码、视频分辨率、视频 IP 码率、视频帧率、视频 IP 分组丢失率、视频时延、音频 RTP 分组丢失率、视频 RTP 分组丢失率、视频 RTP 码率	0.57	2.33	1.57
音频编码、音频码率、音频 IP 分组丢失率、音频时延、视频 IP 码率、视频帧率、视频 IP 分组丢失率、视频时延、音频 RTP 分组丢失率、视频 RTP 分组丢失率、视频 RTP 码率	0.56	2.36	1.67
音频 IP 分组丢失率、音频时延、视频 IP 码率、视频帧率、视频 IP 分组丢失率	0.89	0.07	0.157

率、视频 RTP 分组丢失率、码率视频 RTP 码率的情况，实验结果见表 9。

表 9　常用回归模型实验结果

算法模型	相关系数	均方差	平均绝对误差
线性回归	0.67	0.82	0.9
决策树回归	0.71	0.76	0.84
支持向量回归	0.76	0.68	0.76
K 最近邻回归	0.69	0.79	0.88
随机森林	0.89	0.11	0.17
极端随机森林	0.92	0.12	0.15
梯度提升树	0.94	0.05	0.138

经过综合评判各项指标，发现梯度提升树在本数据集的训练效果最好，因此，采用梯度提升树学习模型进行训练学习。下面介绍梯度提升树的训练过程。

（1）梯度提升树模型的输入、输出数据

输入数据：训练数据集（此处是视频特征与质量的数据集），数据集抽象表示见表 10。

将（$X_1, X_2, X_3, \cdots, X_{13}$）表示为 x，Y 表示为 y。于是数据集表示为 $T=\{(x_1,y_1),(x_2,y_2),\cdots\}$，其中 (x_1,y_1)表示输入数据的第一行。

输出结果：梯度提升树模型公式 $f_m(x)$。

表 10　数据集抽象表示

特征指标	抽象表示
音频编码	X_1
音频码率	X_2
音频 IP 分组丢失率	X_3
音频时延	X_4
视频编码	X_5
视频分辨率	X_6
视频 IP 码率	X_7
视频帧率	X_8
视频 IP 分组丢失率	X_9
视频时延	X_{10}
音频 RTP 分组丢失率	X_{11}
视频 RTP 分组丢失率	X_{12}
视频 RTP 码率	X_{13}
vtMOS	Y

（2）模型训练过程

找出每个属性的最佳分列点，使得预测值跟实际值之间的均方差最小。举例说明见表 11。

表 11　举例说明数据

x_i	1	2	3	4	5	6	7	8	9	10
y_i	5.56	5.70	5.91	6.40	6.80	7.05	8.90	8.70	9.00	9.05

对于 x_i 来说，存在切分点（1.5，2.5，3.5，4.5，5.5，6.5，7.5，8.5，9.5）。对于上述的每一个切分点（将数据切分为 2 份），计算出预测值，例如切分点为 2.5，那么对应的预测值分别为：

$$c_1=(5.56+5.70)/2=5.63$$

$$c_2=(5.91+6.40+\cdots+9.05)/8=7.73 \quad (1)$$

于是得到一个简易模型：如果 $x<2.5$，预测值为 5.63；如果 $x>2.5$，预测值为 7.73。

以上过程会得到一个残差 $m(s)=\sum(y_i-c_1)^2+\sum(y_i-c_2)^2=12.07$，也就是预测值跟实际值的误差平方和。显然，存在一个分列点，使得模型的预测值跟实际值的误差平方和最小，把这个点称为最佳分列点。找到了最佳分列点，也就找到了最佳预测模型。残差见表 12。

表 12　残差

s	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
$m(s)$	15.72	12.07	8.36	5.78	3.91	1.93	8.01	11.73	15.74

从表 12 可以看出，以上数据集的最佳分列点是 6.5，因为对应的残差 $m(s)$最小。则对应的模型公式如下：

$$T_1(x)=\begin{cases}6.24, x<6.5\\8.91, x\geqslant 6.5\end{cases}$$

$$f_1(x)=T_1(x) \quad (2)$$

计算梯度的公式如下：

$$r_{mi}=-\left[\frac{\partial L(y_i,f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-i}(x)}=f(x_i)-y_i \quad (3)$$

梯度见表 13。

表 13　梯度

x_i	1	2	3	4	5	6	7	8	9	10
r_{2i}	−0.68	−0.54	-0.33	0.16	0.56	0.81	−0.01	−0.21	0.09	0.14

根据表 13 对梯度进行拟合，得到如下模型公式：

$$T_2(x)=\begin{cases}-0.52, x<3.5\\0.22, x\geqslant 3.5\end{cases}$$

$$f_2(x)=f_1(x)+T_2(x)=\begin{cases}5.72, x<3.5\\6.46, 3.5\leqslant x<6.5\\9.13, x\geqslant 6.5\end{cases} \quad (4)$$

用 $f_2(x)$拟合训练数据的平方损失误差为：

$$L(y,f_2(x))=\sum_{i=1}^{10}(y_i-f_2(x))^2=0.79 \quad (5)$$

继续求得：

$$T_3(x)=\begin{cases}0.15, x<6.5\\-0.22, x\geqslant 6.5\end{cases} \quad L(y,f_3(x))=0.47$$

$$T_4(x)=\begin{cases}-0.16, x<4.5\\0.11, x\geqslant 4.5\end{cases} \quad L(y,f_4(x))=0.30$$

$$T_5(x)=\begin{cases}0.07, x<6.5\\-0.11, x\geqslant 6.5\end{cases} \quad L(y,f_5(x))=0.23$$

$$T_6(x)=\begin{cases}-0.15, x<2.5\\0.04, x\geqslant 2.5\end{cases}$$

$$f_6(x)=f_5(x)+T_6(x)=T_1(x)+\cdots+T_5(x)+T_6(x)=$$

$$\begin{cases}5.63, x<2.5\\5.82, 2.5\leqslant x<3.5\\6.56, 3.5\leqslant x<4.5\\6.83, 4.5\leqslant x<6.5\\8.95, x\geqslant 6.5\end{cases} \quad (6)$$

$f_6(x)$就是最终得到的提升树，可以看到，在提升树的训练（学习）过程中，误差不断减少，预测值越来越准确。

以上就是梯度提升树模型的全过程。用上述训练好的模型公式对后面的未知数据集进行预测（评估）。

4　评估验证与结论

随机抽取 5 组数据进行验证，预测评估结果见表 14。

表 14　评估和结果验证

组别	组 1	组 2	组 3	组 4	组 5
音频编码	AMR-WB	AMR-WB	AMR-WB	AMR-WB	AMR-WB
音频码率/(kbit·s^{-1})	23.85	23.85	23.85	23.85	23.85
音频 IP 分组丢失率	0	0.002	0.001	0.001	0.002
音频时延/ms	22.54	22.54	22.54	22.54	22.54
视频编码	H264	H264	H264	H264	H264
视频分辨率	VGA	VGA	VGA	VGA	VGA
视频 IP 码率/(bit·s^{-1})	759 916	786 248	759 854	786 278	776 574
视频帧率/(f·s^{-1})	20.66	19.56	20.26	21.45	20.23
视频 IP 分组丢失率	0.000 7	0.002 2	0.001 2	0.000 9	0.001 6
视频时延/ms	30.57	41.58	30.57	41.58	41.58
音频 RTP 分组丢失率	0.000 0	0.000 7	0.000 4	0.000 0	0.000 5
视频 RTP 分组丢失率	0.000 7	0.001 0	0.000 5	0.000 0	0.000 7
视频 RTP 码率/(bit·s^{-1})	778 753	738 714	778 864	736 574	740 087
结果	3.27	3.05	3.18	3.09	3.22
vtMOS-梯度提升树	3.25	3.12	3.20	3.08	3.19
vtMOS-支持向量机	3.04	2.62	3.03	2.59	3.06
vtMOS-随机森林	3.89	3.36	3.78	3.33	3.29
误差-梯度提升树	0.61%	−2.30%	−0.63%	0.32%	0.93%
误差-支持向量机	7.03%	14.1%	4.72%	16.18%	4.97%
误差-随机森林	−18.96%	−10.16%	−18.87%	−7.77%	−2.17%

通过对比参考结果，首先在选择的 3 组算法模型中梯度提升树算法误差范围很小，误差均在±3%范围内，而其他两组算法误差较大。组 2 因为分组丢失率上升而出现了较大偏差，但梯度提升树的计算结果仍在可接受范围内。验证结果说明建立的新评估算法达到参考要求，可以对 VoLTE 视频通话进行评估。

通过对原始数据指标参数的预处理，使得数据特征能更好地被机器学习。同时对数据指标的不同数据类型筛选提取关键特征，减少了特征维度，降低了学习算法的复杂度。最后通过对比多种机器学习算法评估，最终采用梯度提升树模型建模进行视频质量的评估预测，解决了通过核心网参数指标实时评估 VoLTE 视频通话质量的问题。

5　结束语

本文首先简单介绍了机器学习技术和 VoLTE 视频通话流程，然后在基于相关研究之上，给出了一种 VoLTE 视频通话质量分析方法，应用于对 VoLTE 视频通话的特征学习过程，然后采用“相关系数”“均方差”“平均绝对误差”3 种指标选取最有效的特征参数及回归模型完成最优评估模型的选取，从而实现 VoLTE 视频通话质量评估。为了验证该评估模型的有效性，对现有的 VoLTE 视频通话原始数据，分别通过评估模型和其他全参考评估方法对所采集的数据进行实验分析，避免了人工评估导致的误差，提高了质量评估的效果。机器学习技术的创新使用，使得 VoLTE 视频通话

质量评估更快捷、更智能、更准确。同时，该方法的研发，打破了大厂商的软件评估算法的垄断，降低了运营商评估 VoLTE 视频通话质量的成本。

参考文献：

[1] 冯赟中. VoLTE 技术演进与改进对策研究[J]. 通讯世界: 下半月, 2016(7): 90-91.
FENG Y Z. Research on VoLTE technology evolution and improvement countermeasures[J]. Communication World: Second Half of the Month, 2016(7): 90-91.

[2] 温婧芙. 4G+高清语音 VoLTE[J]. 科技与创新, 2017(9): 15-16.
WEN J F. 4G+ VoLTE[J]. Technology and Innovation, 2017(9): 15-16.

[3] 北京理工大学. 一种 VoLTE 视频通话用户体验质量评估方法: 201710127004.X[P]. 2017-06-20.
Beijing Institute of Technology. VoLTE video call user experience quality evaluation method: 201710127004.X[P]. 2017-06-20.

[4] 北京理工大学. 基于神经网络的 VoIP 无参考视频通信质量客观评价方法: 201710268980.7[P]. 2017-08-08.
Beijing Institute of Technology. Objective evaluation method of VoIP unreferenced video communication quality based on neural network: 201710268980.7[P]. 2017-08-08.

[5] TOM M. Machine learning[M]. New York: McGraw-Hill, 1997.

[6] VALIANT L G. A theory of learnable[J]. Communications of the ACM, 1984, 27(11): 1134-1142.

[7] 陈凯, 朱钰. 机器学习及其相关算法综述[J]. 统计与信息论坛, 2007, 22(5): 105-112.
CHEN K, ZHU Y. Overview of machine learning and related algorithms[J]. Statistics and Information Forum, 2007, 22(5): 105-112.

[作者简介]

钟其柱（1985– ），男，中国移动通信集团广东有限公司中山分公司网络管理中心高级工程师，主要研究方向为移动通信网络的分析优化。

《电信科学》第五届编辑委员会