

脑启发视频用户体验 评测关键技术

Key Techniques of Brain Inspired Video QoE Prediction



陶晓明 /TAO Xiaoming¹, 杜冰 /DU Bing², 段一平 /DUAN Yiping¹

(1. 清华大学, 中国 北京 100084;

2. 北京科技大学, 中国 北京 100083)

(1. Tsinghua University, Beijing 100084, China;

2. University of Science and Technology Beijing, Beijing 100083, China)

摘要: 基于脑电图 (EEG) 响应趋同性的生理学机理, 研究了基于脑电图响应特征的体验质量 (QoE) 度量方法, 实现小样本稳定度量; 建立网络关键性能指标 (KPI)、服务关键质量指标 (KQI) 与 QoE 之间的映射关系模型。该模型可有效地提升多媒体服务和网络资源协同的优化空间, 为显著提升多媒体业务支持能力提供新途径。

关键词: 用户体验; 深度学习; 脑电图; 评测

Abstract: Based on the physiological mechanism of electroencephalogram (EEG) response convergence, the quality of experience (QoE) measurement method based on EEG response characteristics is studied to realize small sample stability measurement. Furthermore, the key performance indicator (KPI) and key quality indicator (KQI) are established based on this model. By this way, the optimization space of multimedia services and network resources collaboration can be effectively improved, and a new way to significantly enhance multimedia business support capabilities is provided.

Keywords: quality of experience; deep learning; electroencephalogram; evaluation

DOI: 10.12142/ZTETJ.202101008

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210204.0943.002.html>

网络出版日期: 2021-02-04

收稿日期: 2020-12-22

1 视频体验质量评测概述

无线多媒体业务的爆炸式增长已成为无线通信行业快速、持续发展的主要推动力。多媒体信息在为用户带来多类型、全方位的视听体验的同时, 也决定了用户对多媒体业务的主观评价必将受多维因素的综合影响; 因此用户的体验质量 (QoE)^[1] 才是衡量服务质量的最优指标。然而, 传统

通信技术的发展是基于通信系统中各层面客观技术指标来提升用户的服务质量 (QoS)。该发展思路只能单纯地提升一个或多个客观通信指标, 同时可能会偏离用户满意度提升这一根本目标, 从而造成通信和计算资源的浪费。可以看到, 与传统通信业务相比, 无线视频通信的功能定位已从准确、快速、及时地完成信息传输, 扩展为满足用户多样化的通信业务体验需求。未来, 无线视频通信质量的评价方法应当具备更强的主观属性和整体属性。也就是说, 无线视频通信质量评价的

结果应当能全面反映一项技术或者业务是否能够满足其用户在特定应用场景下的各方面体验需求。

面向 QoE 的视频通信成为全球学术界的研究热点。现有视频通信系统大部分采用基于 QoS 的度量方法, 即以微观精确性为前提的客观评价指标, 具体包括峰值信噪比 (PSNR)、结构相似性 (SSIM) 等^[2-3]。这个度量方法使得网络服务能力的提高单纯依赖于传输带宽的增加。A. V. MOORSEL 等学者于 2001 年率先提出 QoE 的概念, 用于度量用户在使用网络基础设施、

基金项目: 国家重点研发计划 (2019YFB1803404);
国家自然科学基金 (61925105、61801260)

享受网络服务过程中的主观感受^[4]。研究表明,人们在获取信息时,往往只关注主观上的视听体验,并不关注客观上的 QoS 通信指标,例如带宽、排队时延、时延抖动、吞吐量、丢包率、峰值信噪比等^[5]。也就是说,用户体验与网络传输能力并不构成比例关系,单纯通过增加传输带宽来提高用户体验具有局限性;而面向 QoE 的视频通信有可能从根本上减少业务对无线网络带宽需求的压力。如果以体验质量 QoE 为准则,比如关注度、清晰度、流畅度等,那么可放松对精确性的要求,获得新的优化空间,提高视频通信的效率。

目前,通信领域对于 QoE 的研究,主要集中在度量和建模^[6]。主观评价的方法是从用户的感知出发,让用户直接对所使用的业务做出评价,因此它是能够最直接、准确地反映用户体验的方法,具体包括平均意见评分(MOS)、差分平均意见评分(DMOS)、恰可识别失真(JND)等。视频视觉主观质量评价的研究者们提出了最小可觉失真(JND)^[7]的重要概念,并将其作为评判视频视觉质量是否达到最优的标准值,以对 QoE 进行度量。文献[8]将场景因素与人类因素作为模型的调节因素,提出了真实环境下基于加权函数的 QoE 估计框架。华为 mlab^[9]提出了移动视频平均意见评分(MV-MOS),对视频的体验质量进行度量。上述方法主要在于建立关键指标和用户 QoE 之间的映射关系,从而反映用户体验质量。这类度量方法主要依赖于用户行为,在一定程度上反映用户的主观感受,为视频通信提供了新的思路。通信过程中的编解码、传输等都会影响视频质量。因此,面向 QoE 的视频质量评价、面向 QoE 的视频编码以及面向 QoE 的视频质量增强是本文研究的主要内容。文章中,我们将

从 QoE 评价、QoE 编码、QoE 提升 3 个方面提升视频通信系统的效率。

面向 QoE 的视频通信面临着新的机遇和挑战。当前的视频通信 QoS 缺少面向人类感知的评价准则。现有视频通信系统的服务质量多使用 PSNR、SSIM 等 QoS 评价指标,它们以微观精确性为前提,不能反映用户感知质量。视频通信系统的最终目标是终端用户,衡量通信服务品质的根本标准是用户的 QoE^[10-11],即用户的宏观满意度。当前,基于用户的评价方法依赖于用户的行为反馈,这易受人们认知偏差的影响;基于业务的评价方法尚难以实现在实际业务场景下对业务的综合体验质量的准确评价,且难以推广应用。视频感知质量是视频 QoE 的一个客观评价指标,常用的评价方法有特征相似性、相对熵^[12]、用 Fréchet Inception 距离^[13]、Inception Score^[14]等,但它们难以直接准确地反应主观感受;因此,如何建立一套视频感知质量评价准则,为面向 QoE 的视频通信提供通用的评价标准,是必须要解决的问题。

QoE 的模型本质上是可测参数到视频 QoE 的映射,包括 QoE 的度量与评测。QoE 的度量是指 QoE 的标尺,目前不同场景、不同业务存在 QoE 度量标准不统一的问题,难以真实反映用户的 QoE。传统方法往往采用主观评分,因人而异、差异大,难以找出度量关系。这样一来,传统方法会需要大量样本,如利用 10 万左右的用户打分,才能归纳出视频质量退化与 QoE 相对稳定的度量关系。近年来的研究表明,通过测量人类视听刺激的早中期神经信号,脑电图(EEG)可以反映主观感受,并可有效排除个体的影响。QoE 评测本质上是从可测参数到用户体验的映射,包括数学模型和机器学习模型两大类。数学模型主要包括对数模型、线性回归、信息增

益、相关性分析、E 函数模型等,也是目前使用较多的方法。数学模型的主要步骤是通过采集大量的数据,包括可测参数和用户打分,建立可测参数和 QoE 的数学关系,从而建立图表、分段函数等数学映射公式,其本质是数据拟合的思想。

影响 QoE 的因素众多,因此通过数据拟合的方法,是很难获得显式的数学表达的。近几年,机器学习模型取得了突破性进展,被大量应用于计算机视觉任务中。机器学习模型主要是采用数据驱动的方法,从大量数据中挖掘有意义的规律或模式,这在一定程度上克服了线性回归和相关性分析方法的缺点。因此,将机器学习的方法用于 QoE 模型被开始大量地研究。该方法的核心思想是从大量数据中学习可测参数和视频 QoE 间的复杂映射关系。一方面,视频 QoE 可以建模为一个预测问题,即使用神经网络建立影响 QoE 的因素与 QoE 分数之间的非线性复杂映射关系,用于视频 QoE 评测,例如,基于逻辑斯蒂回归的预测模型。另一方面,视频 QoE 可以建模为一个分类问题(通常为五分类问题),可以支持向量机(SVM)和决策树等分类算法。该类算法通常难以有效处理高维数据,无法为大规模业务提供稳定评测。

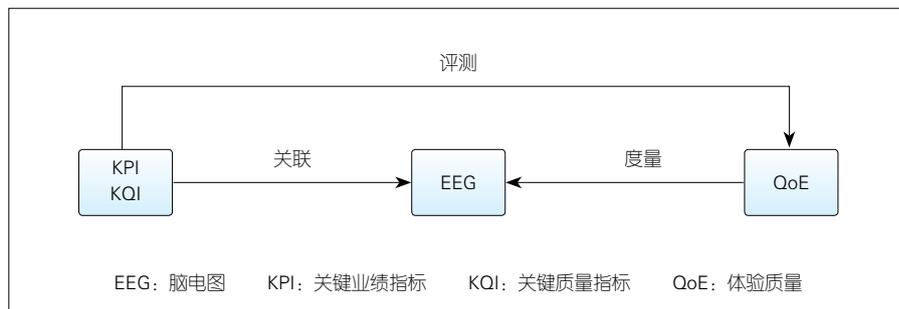
2 脑启发 QoE 评测关键技术

借助 EEG 实验手段,我们测量专业被试者观看不同播放质量视频时相对稳定的脑电响应,剖析无线网络参数、多媒体业务特性对用户体验质量的影响作用。另外,还利用多参数特征选择方法提取关键特征,量化其与用户的感知体验的相关性,从而确定影响用户体验质量的关键特征参数。利用机器学习方法,挖掘这些关键特征指标与用户主观体验之间的映射关

系，建立从可测参数到用户主观体验的预测模型。脑启发 QoE 评测主要包括 3 个方面的关键技术，如图 1 所示。

2.1 基于脑电响应特征的 QoE 度量方法

对体验质量评价方法的研究除了需要准确理解体验质量的定义外，还需要全面考虑体验质量的各项影响因素，清晰梳理体验质量的形成过程并科学地提出体验质量的预测方法。同时，体验质量评价方法的建立本质上是一项多学科交叉的研究内容。在传统信息与通信科学的基础上，生理学、心理学和社会科学等有关学科也是本领域研究的基础。EEG 系统是一种能够连续测量并记录人类头皮不同部位电位信号的设备，它通过在头皮不同部位放置弱贴合的电极，实现非侵入式的测量。对头皮电位的有效记录与分析为人类探索自身思维活动提供了一种科学、客观、可行的研究手段。脑电信号的空间分布、波形等特征也根据具体感官刺激事件的类型有着显著的模式。借助脑电图实验手段，对用户观看不同播放质量视频过程中的脑电信号进行实时测量，并通过基于事件相关电位（ERPs）和时频特征的分类算法，对用户关于视频关键性能指标的感知及认知行为进行量化表征，进而测定其感知极限。此外，从大规模用户日常移动视频业务实测数入手，将智能计算引入用户质量体验预测模型，并利用机器学习的计算工具，挖掘无线视频业务参数与用户体验中间分值之间复杂的映射关系，建立符合用户体验的评价标准。利用时域 P300 成分检测与小波变换时频分析，确定了用户对于卡顿、清晰度、启动时延等关键性能指标的感知极限；根据实测数据，计算视频业务参数和用户体验之间的斯皮尔曼相关系数，确定对用户体验有显著影响的关键网络及业



▲图 1 脑启发 QoE 评测关键技术

务指标；为建立无线视频业务参数与用户体验分值之间模型，引入深度学习方法，对关键的视频业务参数进行归一化计算，并利用深度感知器模型，训练无线视频业务参数与用户体验分值的非线性变换关系，从而获得基于客观业务参数的用户体验映射模型，实现对广大区域内多用户体验质量的在线实时预测。

2.2 QoE 关键特征选择

影响视频 QoE 的参数种类各异，数量繁多，例如，比特率、误码率（BER）、信号强度、缓冲速率、缓冲时延、重缓冲比率、重缓冲次数、视频播放时长等，可综合反映视频的清晰度、流畅度、关注度等。从大规模实测数据中，可以采集可测参数和相应的用户打分。高维可测参数对用户 QoE 的影响是有差别的。因此，通过计算可测参数和用户打分的相关系数，进行特征选择。我们把每 100 条用户对视频的打分为一组，每一条用户行为有多个可测参数和用户关于清晰度、启动延时、卡顿和视频总印象的主观打分。我们计算这 100 条用户行为的每一个特征和 4 个打分的斯皮尔曼相关性系数，找到影响 QoE 的关键因素。通过计算相关系数，得到可测参数对 QoE 的一个“加权”，之后就可以产生和用户主观打分相关性较强的特征子集。我们可以采用阈值法、双向搜索方法等提取关键特征。

2.3 基于深度学习的用户体验质量预测模型

在进行 QoE 度量之后，如何建立可测参数 KPI 和 KQI 到 QoE 的映射，仍然是需要解决的问题。综合考虑客观与主观因素，定量分析各参数与 QoE 的关系后，可以采用深度学习的方法，定义适用于无线环境的视频质量评价模型。通过采集大量实测数据，获得网络参数、业务参数和用户打分，构建视频 QoE 评测数据集，并根据采集到的实际数据，采用 One-hot 进行编码。数据集可以分为训练数据集和测试数据集。首先，通过分析，确定对用户体验有显著影响的关键网络及业务指标，在数据输入深度学习模型之前，需要先对关键指标进行归一化。该深度学习模型由两部分组成：全连接层和分类预测层（使用了分类器）。根据深度学习的原理，整个模型分为两个过程：训练和预测。在训练过程中，采集的大量真实数据经过 One-hot 编码后，输入深度神经网络，学习网络业务客观指标和视频平均意见分数（VMOS）之间的非线性映射关系，并实现其在线模式。预测过程是将实时采集的数据输入到训练好的网络中，从而得到用户体验质量的预测值。

2.4 视频 QoE 评测系统

随着 4G 网络的快速发展及 5G 业务层面多样化，传统的投诉问卷调查无法满足用户需求，因此构建视频

QoE 评测系统以实时预测视频 QoE 是必须要解决的问题。该评测系统主要包括：数据采集系统、脑电感知系统、智能模型。该系统主要的输入为：初始缓冲时延、卡顿数组、卡顿量、卡顿总时长、播放时长、视频总时长、分辨率、帧率等参数；输出为：用户感知分值。通过采集用户观看视频过程中采集的参数，调用视频 EEG 标准模型接口，可反馈当前观看视频的用户感知量化分值。

数据采集是指通过数据接口采集网络数据，并在云端建立采集数据库，基于不同业务类型，提取相应字段输出给脑电感知系统和智能模型两个单元。

脑电感知系统是系统的第一核心单元。脑电感知系统具体是指基于生物特征识别技术建立脑电标准数据库，再使用输入的网络采集数据，提取相应业务的脑电数据库进行模型映射，并将评分准则校正结果输出给智能模型进行在线质量评测。

智能模型是系统的第二核心单元。智能模型具体是指基于已有算法库，根据用户应用选择相应的人工智能模型，并使用输入的网络采集数据，依托脑电感知系统输入的评分准则校正结果，进行质量评测以及其他的可扩展应用结果输出。智能模型是基于 Keras + Tensorflow 框架的，采用模块化和参数化设计，主要包括数据载入和参数输入、数据清洗、模型训练单元、模型精度评估单元，以及评测结果应用程序编程接口（API）等主要模块。

3 结束语

本文中，借助 EGG 实验手段，我们对用户观看不同播放质量的视频过程中的脑电信号进行实时测量。通过基于事件相关电位（ERPs）和时频特

征的分类算法，对用户关于视频关键性能指标的感知及认知行为进行量化表征，进而测定其感知极限。此外，我们从大规模用户日常移动视频业务实测数入手，将智能计算引入用户质量体验预测模型，并利用机器学习的计算工具，挖掘无线视频业务参数与用户体验中间分值之间复杂的映射关系，建立符合用户体验的评价标准，从而实现用户体验质量的实时评测。

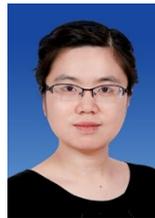
参考文献

- [1] BALACHANDRAN A, SEKAR V, AKELLA A, et al. Developing a predictive model of quality of experience for Internet video [C]//Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM. New York, NY, USA: ACM, 2013: 339–350. DOI:10.1145/2486001.2486025
- [2] MINNEN D, TODERICI G, COVELL M, et al. Spatially adaptive image compression using a tiled deep network [C]//2017 IEEE International Conference on Image Processing (ICIP). Beijing, China: IEEE, 2017: 2796–2800. DOI:10.1109/icip.2017.8296792
- [3] KLOPP J, WANG F Y, CHIEN S, et al. Learning a code-space predictor by exploiting intra-image-dependencies [EB/OL]. [2020–12–20]. <http://www.bmva.org/bmvc/2018/contents/papers/0491.pdf>
- [4] MOORSEL A. V. Metrics for the Internet age: quality of experience and quality of business [EB/OL]. [2020–12–20]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.24.3810>
- [5] TAO X M, DUAN Y P, XU M, et al. Learning QoE of mobile video transmission with deep neural network: a data-driven approach [J]. IEEE journal on selected areas in communications, 2019, 37(6): 1337–1348. DOI:10.1109/jsac.2019.2904359
- [6] PENG X, DUAN Y P, GENG B R, et al. A QoE-based alarm model for terminal video quality [C]//2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Ottawa, ON, Canada: IEEE, 2019: 1–5. DOI:10.1109/globalsip45357.2019.8969366
- [7] JAYANT N, JOHNSTON J, SAFRANEK R. Signal compression based on models of human perception [J]. Proceedings of the IEEE, 1993, 81(10): 1385–1422. DOI:10.1109/5.241504
- [8] MOLLER S, RAAKE A. Quality of experience: advanced concepts, applications and methods [M]. Germany: Springer, 2014
- [9] HUAWEU mLab [EB/OL]. [2020–12–20]. <http://mlab.huawei.com>
- [10] CHEN Y, WU K, ZHANG Q. From QoS to QoE: a survey and tutorial on state of art, evolution and future directions of video quality analysis [J]. IEEE communications surveys and tutori-

als, 2014, 99(1): 1

- [11] GREGPR K, BESSE F, REZENDE D J, et al. Towards conceptual compression [C]//NIPS' 16: Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2016: 3556–3564
- [12] MINNEN D, TODERICI G, SINGH S, et al. Image-dependent local entropy models for learned image compression [C]//2018 25th IEEE International Conference on Image Processing (ICIP). Athens, Greece: IEEE, 2018. DOI:10.1109/icip.2018.8451502
- [13] JOHNSTON N, VINCENT D, MINNEN D, et al. Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4385–4393. DOI:10.1109/cvpr.2018.00461
- [14] BALLE J, MINNEN D, SINGH S, et al. Variational image compression with a scale hyperprior [C]//6th International Conference on Learning Representations. Vancouver, BC, Canada: ICLR, 2018

作者简介



陶晓明，清华大学教授、博士生导师；主要从事无线多媒体通信理论及关键技术应用研究；曾获国家自然科学基金二等奖、国家技术发明二等奖、教育部科技进步一等奖等奖项，获得国家自然科学基金杰青项目资助，以及中国青年科技奖、中国青年女科学家奖等；发表论文 50 余篇，授权专利 40 余项。



杜冰，北京科技大学计算机与通信工程学院讲师；研究方向为无线多媒体通信。



段一平，清华大学电子工程系助理研究员；研究方向为无线多媒体通信。