

生成式 QoE 建模：电信网络的轻量级方法

文帝尼·奈尔¹，卡尼卡·萨切德¹，布雷杰什·拉尔¹，

1 印度理工学院德里分校

* 布雷杰什·拉尔@ee.iitd.ac.in

摘要

体验质量 (QoE) 预测在优化资源管理和提升用户满意度方面发挥着关键作用，这适用于电信和 OTT 服务。尽管最近的进展主要依赖于深度学习模型，本研究引入了一个轻量级生成建模框架，该框架平衡了计算效率、可解释性和预测准确性。通过验证向量量化 (VQ) 作为预处理技术的有效性，连续网络特征被有效地转换为离散分类符号，从而可以与隐马尔科夫模型 (HMM) 结合进行时间序列建模。这个 VQ-HMM 管道增强了模型捕捉动态 QoE 模式的能力，并支持对新数据和未见数据的概率推理。在公开可用的时间序列数据集上的实验结果——包括客观指标和主观的 QoE 评分——证明了这种方法在实时环境和资源受限环境中是可行的，其中推断延迟也是一个关键因素。该框架为复杂的深度学习方法提供了一种可扩展的替代方案，特别是在计算资源有限或延迟约束至关重要的场景中。

1 介绍

用户对服务质量的期望随着新应用和服务的出现而稳步增长。未来的电信网络将整合多种技术，应用于各种不同的部署环境，确保服务质量 (QoS) 仍然是网络提供商的核心关注点。作为更先进且以用户为中心的概念，QoE 在 QoS 的基础上进一步扩展，捕捉客户对服务的感知满意度或不满。它反映了用户体验的端到端评估，涉及许多仍在被发现的主观元素，正如 [7] 中所指出的。

关键的是，QoE 随着时间的推移而展开——这一点在评估中至关重要，正如在 [1]、[2]、[15]、[13] 中所强调的。[18] 中的研究反对通用建模策略，并倡导以用户为中心、上下文感知的机器学习方法，以更有效地预测 QoE。随着沉浸式服务的预期增长，实时且贴近用户地评估 QoE 的能力将变得越来越重要 [21]。各种关于 QoE 评估的考量，包括使用具有高预测准确性的复杂时间序列模型 [17]、[18]、[10]，在多种应用场景中继续被探索。

QoE 是由客观指标和主观维度之间的复杂互动形成的。图 1 说明了 QoE 堆栈。该图展示了影响 QoE 的因素，包括系统级、情境和人为因素。尽管本研究认识到人类、网络、应用和情境变量的作用，但由于包含感知维度 (PDs) 的数据集有限，它主要关注于系统级别的建模。

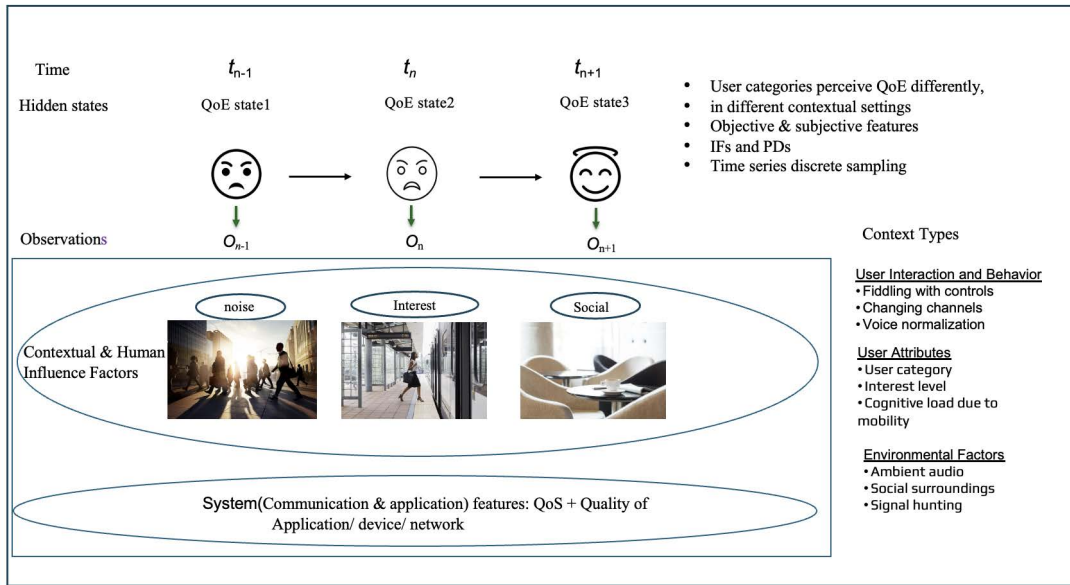


图 1. QoE 堆栈

尽管深度学习模型显示出潜力，但它们的计算复杂性和数据需求往往限制了在实时或资源受限环境中的应用。实验评估了不同建模技术在公开可用的时间序列数据集上的效果，这些数据集结合了客观和主观指标。通过与其他模型的比较分析，我们强调了预测准确性和计算延迟之间的权衡，突出了电信环境中实用的优势。

本工作的关键贡献在于验证向量量化（VQ）作为一种有效的预处理技术，将连续值网络特征转换为离散分类符号，使其能够集成到序列建模框架中进行 QoE 估计。这种 VQ-HMM 管道增强了模型捕捉时间 QoE 动态的能力，显著提高了预测准确性。此外，该方法在无参考（NR）和近参考（NRR）场景下具有应用潜力——这些情况下用户 QoE 的直接真实值不可用。HMM 的概率生成结构能够在未见用户数据上进行概率推理、半监督学习及合成 QoE 序列生成，使得这种方法非常适合部署到轻量级边缘平台和可穿戴设备上。据我们所知，这是首次将 VQ 驱动的 HMM 框架应用于公开可用的数据集上的低延迟 QoE 估计，并且还能够在资源受限环境中支持数据增强。

2 相关工作

QoE 估计依赖于客观特征的结合——这些特征作为可量化、可重复的质量指标，以及主观 PDs，这对于捕捉以用户为中心的经验细微差别 [1], [2], [4] 是至关重要的。许多现有的研究使用符合 ITU-T P.10/G.100 建议的 [20] 数据集，通常采用绝对类别评分（ACR）量表来标注 QoE。这种标准化标注有助于在不同的模型和研究之间进行一致的基准测试。尽管最近的 QoE 建模工作主

要集中在深度学习框架上——包括判别式和生成式——基础的时间序列生成模型如 HMMs 仍然被显著低估。这尤其体现在可获得主观采集的真实世界数据集的情境中。早期将 HMM 应用于 QoE 的一项尝试在 [16] 中提出，其中该模型用于预测异构接入网络（HAN）中模拟切换过程中的 QoE，依赖于由 ITU-T 估计框架指导的一小组网络级特征。类似地，[11] 提出了一个利用 HMM 的自适应资源分配策略，尽管是在合成数据集上进行评估的。虽然这些研究证实了基于 HMM 建模的理论潜力，但它们并未在现实世界中用户感知的 QoE 数据上验证这种方法，据我们所知，也没有提供公开访问的数据集用于进一步的基准测试。这限制了其实际应用性和可重复性。

一项由 [14] 进行的全面调查回顾了 QoE 建模的最新进展，并强调了实时 PDs 在估计用户体验中的重要性。尽管 [19] 中的数据集提供了一组丰富的主观维度，但它仅限于会话结束时收集的回顾性 QoE 标注。同样，[1] 和 [2] 研究了具有显著 PD 覆盖率的不同数据集，但仍然是仅为会话后评估。在 [5] 中探讨了将以人为本的特点整合到 QoE 框架中的更广泛挑战，并倡导一种包含主观量化和综合建模的全方位统一视角。此外，主观测试的限制——包括成本、可扩展性和预测泛化能力——在 [8] 中得到了强调。

与模拟数据方法不同，诸如 [3], [9] 和 [4] 的数据集提供了跨多种观看场景的实时 QoE 标注。这些数据集已经被广泛建模，但 HMM 在这些数据上的应用仍然有限。尽管在特定于 QoE 的文献中存在这一差距，HMM 仍在其他领域得到了广泛应用——尤其是在自然语言处理 (NLP) 和基因组学中，它们对序列和时间依赖性的建模能力已经在众多研究 [12] 等中得到验证。

表.1 提供了早期段落中提到的所选研究的关键方法和策略的简要概述。表.2 概述了相关工作的比较。

3 方法论

本节概述了用于 QoE 预测的建模框架，包括其理论基础、设计理由、架构组件和评估设置。所提出的方法将 VQ 步骤与 HMM 集成在一起，以轻量级且可解释的方式捕捉时间上的 QoE 动态。以下子部分详细介绍了生成模型的基础、使用 HMM 的理由、系统架构以及所采用的数据集和评估策略。

3.1 理论基础

确定性建模和生成式建模代表了统计建模和机器学习领域中的两种不同方法。确定性模型，如线性回归和支持向量机，旨在学习条件概率 $P(Y | X)$ ，其中 X 表示观测特征而 Y 表示目标变量。这些模型设计用于根据固定规则或判别式学习将输入映射到输出。

相比之下，生成模型旨在学习联合概率分布。

$$P(X, Y) = P(Y) \cdot P(X | Y)$$

表 1. 相关工作

论文标题/作者	可用数据集	使用的方法论	重要发现	实际应用	限制
QoE Estimation and Prediction using Hidden Markov Models in Heterogeneous Access Networks / K. Mitra	No	HMM	HMM can accurately estimate QoE in handoffs in HANs (using the ITU-T E-model)	Application in Handoff protocols in HAN	1. Limited in capturing nuances of user perception, as it assumes fixed mappings between network metrics with Objectively estimated QoE. 2. Simulates two observation variables (One Way Delay, Round Trip Time) with Gaussian distribution
Context aware QoE modelling, measurement, and prediction in mobile computing systems / K. Mitra	No	Naïve Bayes, Gaussian Naïve Bayes, Hybrid Bayesian Networks	CaQoEM efficiently learns and predicts QoE using BNs	Context aware approach called CaQoEM for QoE modelling. Novel introduction to QoE understanding	Non temporal
Study of temporal effects on subjective video quality of experience / Bampis	Yes	Correlation	Significant dataset for accuracy studies on temporal subjective QoE	Streaming	Incorporating additional perceptual dimensions into observations can significantly enhance accuracy
A hierarchical classification model of QoE Influence Factors/ Amour	Yes	Deterministic	Subjective perceptual dimensions are of significant importance	Importance of relevant IFs	Subjective QoE but only in retrospect
Streaming video qoe modeling and prediction: A long short term memory approach/ Eswara	Yes	LSTM	Enhanced Temporal accuracy. Significant dataset for accuracy studies on temporal subjective QoE	Streaming video in telecom networks	1. Challenge in resource limited deployments. 2. Requires incorporating additional perceptual dimension observations to increase accuracy in QoE.
This proposed work	Yes	HMM	Transformation of multivariate continuous data enhancing temporal accuracy	Effective data transformation and validation of light weight generative model under NR and NRR conditions for subjective QoE	QoE is the sole subjective feature (PD). Requires incorporating additional perceptual dimension observations can significantly enhance accuracy

表 2. 相关研究的总结比较

研究	数据集类型	HMM 使用	现实世界验证	公共访问
Mira et al [16]	Simulated	QoE prediction	X	X
Gandotra et al [11]	Synthetic	QoE prediction	X	X
Kougiountzidis et al [14]	Survey	X	-	-
Porcu et al [19]	Retrospective Subjective	X	✓	✓
Amour et al [1] [2]	Retrospective PD	X	✓	✓
Bampis et al [4], Eswara et al [9] [10]	Real Time subjective	X	✓	✓

这允许进行概率推理，生成合成样本，并支持半监督学习，正如文献如 [6] 中所讨论的。与确定性模型不同，生成方法可以捕捉数据中的隐藏结构并在存在缺失信息的情况下建模关系。

这里， $P(Y)$ 模型描述了输出类别或状态的先验分布，而 $P(X | Y)$ 模型描述了给定特定输出时观察到输入的可能性。一旦联合分布被学习，模型就可以通过从 $P(X, Y)$ 中采样来生成新的、合理的示例 (X', Y') 。

这种生成能力在序列建模任务中尤为强大，例如时间上的 *QoE* 估计，其中可以通过模型有效捕捉时间依赖性和潜在状态转换。在 *QoE* 估计的背景下，该估计会随时间变化，并且通常受到潜在因素的影响，生成式建模提供了一个令人信服的框架。特别是，生成模型非常适合需要进行时序推理、结构化序列预测以及对不完整或部分观测数据进行推断的任务。它们生成合成样本的能力和概率解释的支持与在动态和上下文相关的环境中建模用户体验所面临的挑战非常契合。

3.2 模型合理性说明：为何选择 HMMs ?

HMMs 提供了一种结构化且高效的方式来建模存在隐藏状态的序列数据。当观察到的数据是由随时间演变的潜在过程生成时，它们特别有效，使得 HMM 在 *QoE* 估计中高度相关。*QoE* 不仅与时间有关，还受到不可直接观测的潜在感知状态的影响。HMM 可以通过建模隐藏状态之间的概率转移以及每个状态下可观察输出的发射来捕捉这一点。这种结构允许 HMM 基于嘈杂或不完整的观测推断出最有可能的基本用户体验状态序列——在现实部署场景中是一项有价值的的能力。

与通常需要大量训练数据和计算资源的深度学习模型相比，HMM 更轻量级、可解释且更容易实时或边缘环境中部署。它们的透明性使得更好地理解状态转换以及输入特征对感知 *QoE* 的影响成为可能。此外，使用 HMM 可促进概率推理、时间对齐和合成数据生成，所有这些对于 NR 和 NRR 条件下的 *QoE* 建模都是必不可少的。

3.3 系统架构和工作流

所提出的 *QoE* 预测管道从预处理阶段开始，在该阶段中，多变量网络观测值被映射到潜在的用户体验状态。这是通过 VQ 实现的，它将连续值输入特征转换为离散分类令牌，采用 K 均值聚

类算法。每个令牌代表原始观察的量化版本，使能够使用离散状态序列框架进行有效建模。这些量化向量然后传递给一阶 HMM，该模型被训练根据观测到的令牌序列推断出潜在隐藏状态。训练过程涉及从标记序列中估计状态转移矩阵、观测发射矩阵和初始状态概率。对新未标记序列的推理使用维特比算法进行，计算与观察输入相对应的最可能隐藏状态序列。

该架构支持监督训练和无监督推理，使其适应各种部署条件。在具有有限或嘈杂地面真实数据的情况下尤其有效，因为模型的概率结构允许稳健的学习和解释。VQ 和 HMM 的集成导致了一个紧凑且可解释的管道，非常适合实时和资源受限环境。

总结部署的两个阶段为训练阶段和推理阶段：

训练阶段

- 输入：带标签的序列（观测值+对应的状态）
- 估计转换概率 A
- 估计发射概率 B
- 估计初始分布 π

推理阶段

- 输入：未见的多变量观察序列
- 应用维特比算法以找到最可能的状态序列

3.4 数据集和实验设置

实验在多种公开可用的数据集 [1, 4, 10, 19] 上进行。本工作中提及的数据集的选择依据是：

1. 时间序列
2. 多变量观测
3. 用户评估的主观 QoE

本节中展示的样本结果源自通过控制实验生成的数据集 [9]，该数据集包含六个基于多变量网络的预测因子。输入特征由从原始遥测数据派生出的多变量序列组成，包括诸如网络吞吐量、延迟和数据包丢失等参数。这些连续值特征通过矢量量化（VQ）进行离散化处理，将每个数据点映射到一组预定义的离散标记之一。量化步骤不仅降低了计算复杂度，还有助于通过隐马尔可夫模型（HMMs）应用离散状态序列建模。

这些预测因子基于已建立的视频质量评估（VQA）指标，如多尺度结构相似性指数（MSSSIM）和峰值信噪比（PSNR）。感知到的 QoE 是通过从 1 到 100 的连续用户报告评分量表捕获的。在所有预测因子中，MSSSIM 和 PSNR 对预测准确性贡献最大。虽然其他系统参数的单独影响相对较低，但它们在提升整体 QoE 估计方面的贡献仍然显著。

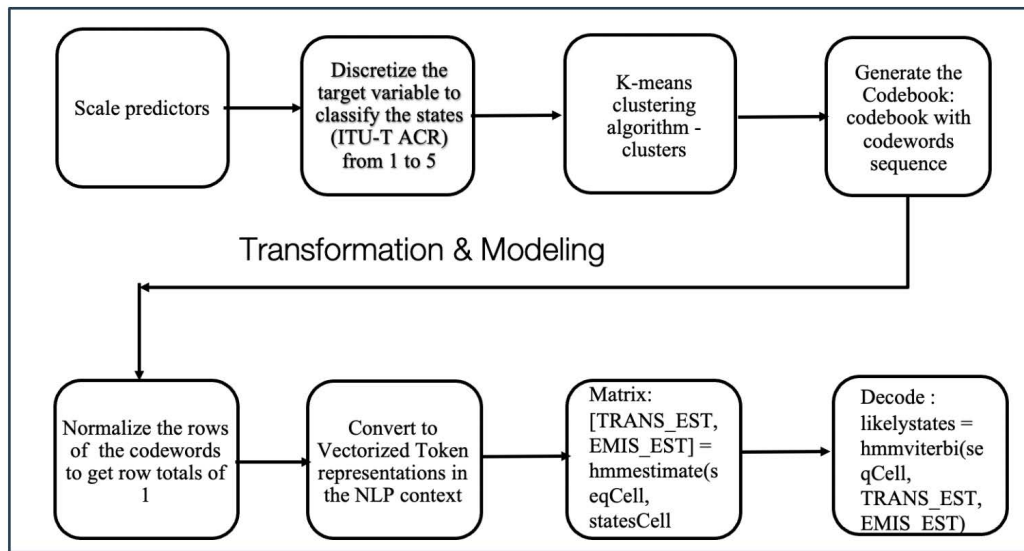


图 2. 变换和建模管道

变换和建模方法总结如图 2 所示。这说明了一个工作流程的概览，其中多元网络数据使用 VQ 进行量化，并使用 HMM 进行建模以预测 QoE。

3.5 限制

在这个领域中的一个关键限制是缺乏包含实时环境中上下文感知或以人为中心特性的公开数据集。因此，本研究依赖于强调系统级（网络）指标的数据集。虽然系统特征不是影响用户体验质量的唯一决定因素，但它们具有显著的影响，并且能够实现相对准确的预测建模。先前的研究表明，几种高级建模技术仅基于这些系统参数已经成功实现了高预测准确性。然而，缺乏全面、多维度的数据集，特别是那些整合用户上下文、内容特征和感知反馈的数据集，继续对构建整体的以用户为中心的用户体验质量模型构成挑战。

4 结果与讨论

4.1 准确性和延迟权衡

图 3 展示了多个模型的比较评估，重点关注准确性和预测推理延迟。评估的模型包括 HMMs、长短时记忆网络 (LSTMs)、生成对抗网络 (GANs) 和变分自编码器 (VAEs)。值得注意的是，GANs 和 VAEs 表现出更高的延迟，但这些未在图中显示。

HMM-VQ 展示了最高的预测准确性为 0.77，并且具有最低的推理延迟 0.0015 秒，使其非常适合

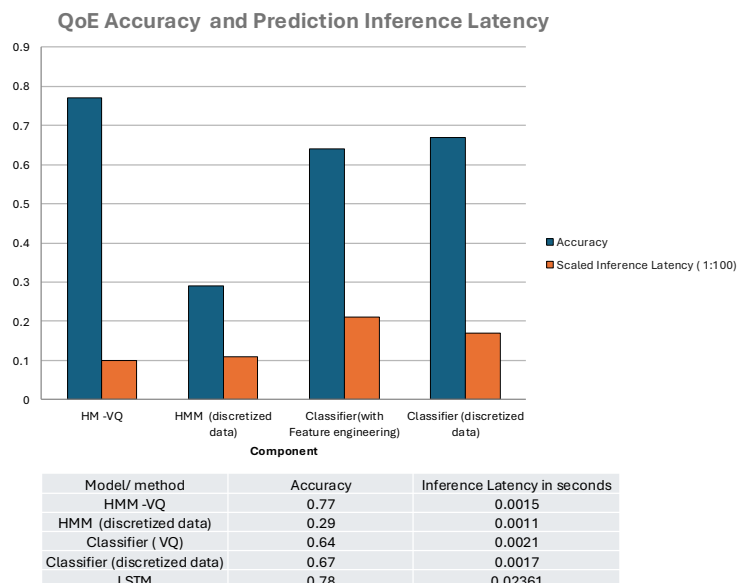


图 3. 模型在准确性和延迟方面的比较分析

实时 QoE 预测。特定领域的特征工程整合显著提升了模型捕捉时间序列 QoE 变化的能力，提供了预测准确性和计算效率之间的最优平衡。相比之下，使用分箱数据的 HMM 表现出次优性能，其准确性为 0.29，尽管它的推理延迟仅为 0.0011 秒。这种准确性上的显著下降突显了标量离散化的潜在缺点，这可能导致关键的时间和上下文信息丢失，从而影响模型的预测能力。分类器模型表现出不同的性能结果。具有特征工程的分类器达到 0.64 的准确性和 0.0021 秒的延迟，而使用离散化数据的分类器实现了稍高的准确性 0.67，并且伴随较低的延迟 0.0017 秒。这些结果显示，在这种特定情况下，离散化可能会简化特征空间并通过降低复杂性来提升分类器性能。

LSTM 模型虽然可以通过微调实现更高的精度，但在推理延迟方面却有大约 0.023 秒的显著增加，这比基线模型高出约 10 倍。诸如 LSTMs、GANs 和 VAEs 等高级模型，尽管展示了强大的表示能力，但也引入了显著的推理延迟。因此，这些模型更适合用于离线处理或基于云的大批量推理，因为它们较高的延迟使它们不适合延时敏感的应用程序，如边缘计算或可穿戴设备。总而言之，HMM-VQ 模型在实时 QoE 预测中表现出最有效的解决方案，提供了高精度和极低的延迟，因此非常适合部署在资源受限、实时环境中。

4.2 精度召回分析

图 4 比较了基于分类器和基于 HMM 的模型在多个视频序列中使用 MATLAB 工具包的表现。对于序列 1，分类器达到了 0.41 的精确度，而基于 HMM 推导出的“可能”序列将这一数值提高到了 0.68，表明假阳性显著减少。召回率也从 0.64（分类器）增加到 0.74（HMM），F1 分数从 0.50 增加到 0.69，反映了基于 HMM 的方法在精确度和召回率之间更为有效的平衡。

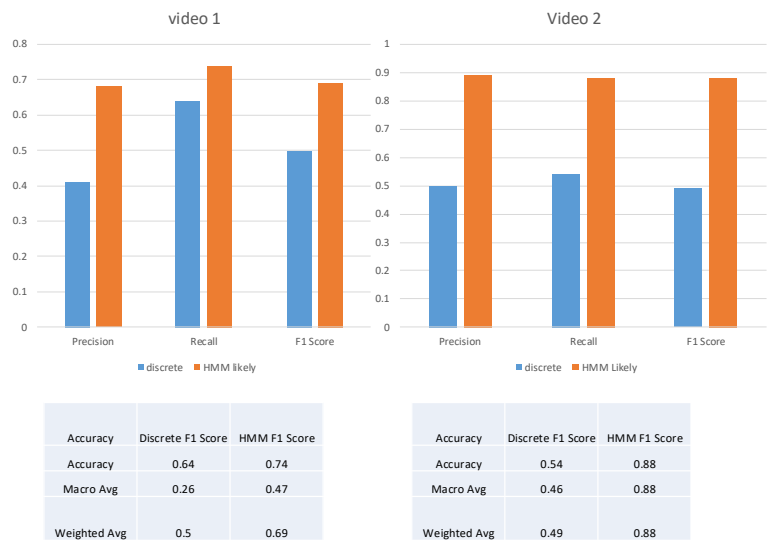


图 4. 精确率-召回率分析

4.3 捕获时间质量-of-体验变化

图 5 描述了时间变化分析的代表性结果。尽管使用了一阶 HMM，该模型有效地捕捉到了与网络级观察相匹配的 QoE 状态序列的时间动态。除了在状态转换期间（仅持续几毫秒）出现瞬时偏差外，推断出的“可能”状态序列紧密匹配地面真实情况，如相应的时间序列图所示。HMM 的生成结构不仅有助于准确推断隐藏状态，还可以用于生成合成序列以供下游训练和评估。利用 VQ 显著降低了输入维度，确保了适用于对延迟敏感的边缘部署的计算效率。至关重要的是，模型的可解释架构——通过学习到的状态转移和发射概率——为理解 QoE 动态提供了透明度，与黑盒深度学习模型形成对比。这种低延迟推理、降低复杂性和可解释性的结合突显了 VQ-HMM 框架在未来的电信网络中进行可扩展 QoE 估计的实际应用价值。

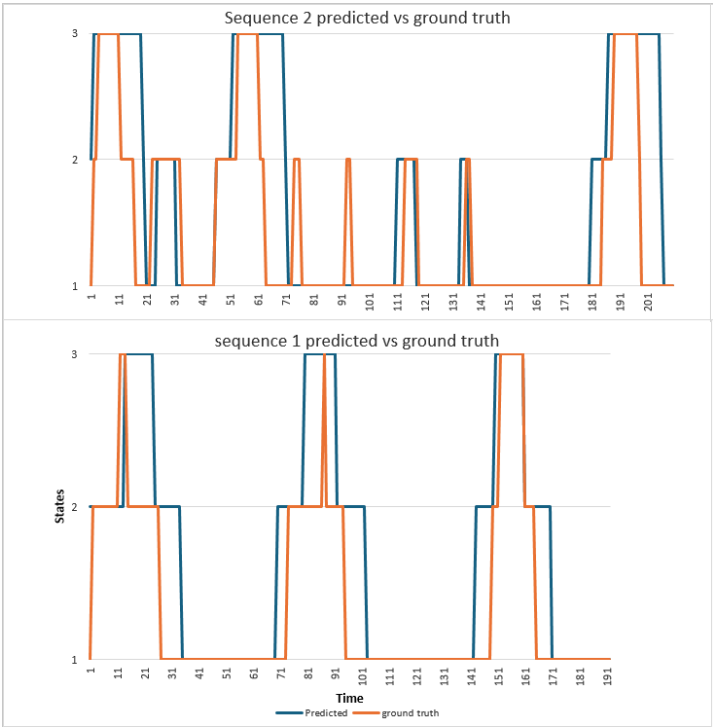


图 5. 时域实验结果

5 结论与未来工作

本文提出了一种轻量级的生成框架用于 QoE 预测，结合了矢量量化和隐马尔可夫模型 (VQ-HMM)，实现了高效、可解释的序列建模。该框架能够以高精度和低计算复杂度有效捕捉时序 QoE 动态，展示了其在实时资源受限应用中的适用性。

实验结果验证了该模型在 NRR 和 NR 场景中的潜在性能，尤其是在主观反馈有限的情况下。除了预测准确性外，该模型通过可解释的状态转换提供了透明度，解决了基于深度学习方法的关键局限性。

未来的工作将纳入感知维度，以更好地使预测与用户为中心的 QoE 因素保持一致，探索混合生成-判别模型以增强表示能力，并扩展实时标注数据集以提高在各种部署环境中的泛化能力。

参考文献

1. L. Amour, S. Sami, S. Hoceini, and A. Mellouk. Building a large dataset for model-based qoe prediction in the mobile environment. In *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 313–317, 2015.
2. L. Amour, S. Souihi, S. Hoceini, and A. Mellouk. A hierarchical classification model of qoe influence factors. In *International Conference on Wired/Wireless Internet Communication*, pages 225–238. Springer, 2015.
3. C. G. Bampis and A. C. Bovik. Learning to predict streaming video qoe: Distortions, rebuffering and memory. *CoRR*, abs/1703.00633, 2017.
4. C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik. Study of temporal effects on subjective video quality of experience. *IEEE Transactions on Image Processing*, 26(11):5217–5231, 2017.
5. J. Baraković Husić, S. Baraković, E. Cero, N. Slamnik, M. Oćuz, A. Dedović, and O. Zupčić. Quality of experience for unified communications: A survey. *International Journal of Network Management*, 30(3):e2083, 2020.
6. C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
7. K. Brunnström, S. A. Beker, K. De Moor, A. Dooms, S. Egger, M.-N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, and M.-C. Larabi. Qualinet white paper on definitions of quality of experience, 2013. Non-published work.
8. Y. Chen, K. Wu, and Q. Zhang. From qos to qoe: A tutorial on video quality assessment. *IEEE Communications Surveys & Tutorials*, 17(2):1126–1165, 2014.

-
9. N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya. Streaming video qoe modeling and prediction: A long short-term memory approach, 2018.
 10. N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya. Streaming video qoe modeling and prediction: A long short-term memory approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(3):661–673, 2019.
 11. P. Gandotra, R. K. Jha, and S. Jain. Sector-based radio resource allocation (sbrra) algorithm for better quality of service and experience in device-to-device (d2d) communication. *IEEE Transactions on Vehicular Technology*, 67(7):5750–5765, 2017.
 12. A. S. Hsu and T. E. Griffiths. Effects of generative and discriminative learning on use of category variability. In *32nd annual conference of the cognitive science society*, 2010.
 13. J. B. Husić and S. Baraković. Multidimensional modelling of quality of experience for video streaming. *Computers in Human Behavior*, 129:107155, 2022.
 14. G. Kougiumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis. A survey on multimedia services qoe assessment and machine learning-based prediction. *IEEE Access*, 10:19507–19538, 2022.
 15. K. U. R. Laghari and K. Connelly. Toward total quality of experience: A qoe model in a communication ecosystem. *IEEE Communications Magazine*, 50(4):58–65, 2012.
 16. K. Mitra, C. Åhlund, and A. Zaslavsky. Qoe estimation and prediction using hidden markov models in heterogeneous access networks. In *Australasian Telecommunication Networks and Applications Conference (ATNAC) 2012*, pages 1–5. IEEE, 2012.
 17. S. Nádas, L. Ernström, L. Szilágyi, G. Patra, D. Krylov, and J. Lynam. To qoe or not to qoe. In *Proceedings of the 2024 Applied Networking Research Workshop, ANRW '24*, page 38 – 44, New York, NY, USA, 2024. Association for Computing Machinery.
 18. P. H. S. Panahi, A. H. Jalilvand, and A. Diyanat. Machine learning-driven open-source framework for assessing qoe in multimedia networks, 2024.
 19. S. Porcu, A. Floris, J.-N. Voigt-Antons, L. Atzori, and S. Möller. Estimation of the quality of experience during video streaming from facial expression and gaze direction. *IEEE Transactions on Network and Service Management*, 17:2702–2716, 2020.
 20. I. Rec. P. 10, vocabulary for performance, quality of service and quality of experience. *ITU-T, Geneva*, 2017.

-
21. R. Schatz, T. Hoßfeld, L. Janowski, and S. Egger. From packets to people: quality of experience as a new measurement challenge. In *Data traffic monitoring and analysis*, pages 219–263. Springer, 2013.