以用户 QoE 预测值为奖励的视频自适应比特率算法

叶 进*，肖庆宇，陈梓晗，陈贵豪，李陶深

(广西大学计算机与电子信息学院 南宁 530004)

【摘要】该文提出了一种基于深度学习的用户体验质量预测网络 (UQPN)，通过当前视频播放状态预测当前用户的 QoE 并进行建模，旨在采用 UQPN 替代以往方法的奖励函数，使得生成的自适应比特率算法做出更符合用户需求的比特率决策。实验证明与已有的奖励函数相比，UQPN 的预测与真实 QoE 的相关系数更高，以该网络作为强化学习奖励得到的算法能够将用户体验质量提高 20%。

关 键 词 自适应比特率算法； 用户体验质量； 强化学习； 奖励

中图分类号 TP393 **文献标志码** A doi:10.12178/1001-0548.2020325

A Video Adaptive Bitrate Algorithm with User QoE Prediction as Reward

YE Jin*, XIAO Qing-yu, CHEN Zi-han, CHEN Gui-hao, and LI Tao-shen

(School of Computer and Electronic Information, Guangxi University Nanning 530004)

Abstract This paper proposes a deep learning-based user QoE prediction network (UQPN)). In this work, the current user's QoE is predicted and modeled based on the current video playback states, and UQPN is used to replace the existing reward functions, in this way the generated ABR algorithm can make bitrate decisions more in line with user requirements. Experiments and the comparison with the existing reward functions show that he correlation coefficient of UQPN prediction and user QoE is higher, and the algorithm using UQPN as reinforcement learning reward can improve user QoE by at least 20%.

Key words adaptive bitrate algorithm; quality of experience; reinforcement learning; reward

近年来，基于 HTTP 的视频流观看需求迅速增长。为了在各种网络条件下实现流畅的视频播放，客户端视频播放器采用自适应比特率 (adaptive bitrate, ABR) 算法来动态确定每个视频块的比特率以优化视频质量。这样做的目标是使视频比特率适应潜在的网络条件来最大化用户的体验质量 (quality of experience, QoE)。但是由于网络流量的高突发性，为每一个视频块选择一个合适的比特率是具有挑战性的。

国际电信联盟 (international telecommunication union, ITU) 对 QoE 进行了明确的定义[1]，即一个应用或一项服务的整体可接受性，它由终端用户的主观感知决定。当 QoE 较差时，用户可能会更早关闭视频页面，这导致视频内容提供方的大量经济损失。而在视频传输场景下，QoE 是指用户在某一次观看视频后对这次观看体验的接受性。一些现有研究以评分的形式直接从用户处获取 QoE，文献 [2] 则采用一些应用层或网络层的指标来定义 QoE。

现有的 ABR 算法采用固定的控制规则来选择未来的视频比特率。但这类方法具有很强的假设性，难以适应不同的网络环境。因此利用强化学习 (reinforcement learning, RL) 生成 ABR 的方法被提出，能从零开始学习并生成算法而无需任何网络假设，这类方法通过提高训练时的奖励值来优化神经网络，而奖励定义为 QoE 函数。但奖励函数往往被预先设置且设置时缺乏现实依据，因此该类基于 RL 的方法具有获得相对良好的奖励值的能力，但它们也可能为用户提供与用户期望不匹配的观看体验。

收稿日期：2020－08－08；修回日期：2020－11－15
基金项目：国家自然科学基金 (61762030, 61872387)；广西自然科学基金 (2018JJA70209)
作者简介：叶进 (1970－)，女，博士，教授，主要从事网络协议设计、数据中心网络等方面的研究. E-mail：yejin@gxu.edu.cn

播放视频时用户 QoE 受到多种因素影响，以准确的 QoE 值作为 RL 训练时的奖励，能让 ABR 朝着最大化 QoE 的方向做出比特率决策。QoE 与视频播放时的指标密切相关，其中包括视频播放时的卡顿持续时间、平均播放比特率和比特率的变化值等。恰当的奖励函数设计能使奖励值的变化更贴近真实用户的 QoE。但如何确定用于 ABR 的奖励，目前缺乏统一的标准，而现有方法中的奖励函数在训练之前就被预先设置，且设置过程缺乏描述和依据，无法得知是否与用户真实意图相匹配。

本文提出用户 QoE 预测网络 (user QoE prediction network, UQPN)，以真实用户数据进行监督学习并预测用户 QoE 的方法。UQPN 将视频流状态作为输入，输出为现在用户的 QoE 预测分数，并以 UQPN 作为“奖励函数”。本文提出了一种基于 RL 的 ABR 算法，引入 UQPN 加入 ABR 训练过程，避免了奖励函数建模的盲目性，从而使 ABR 算法可以在满足用户要求的方向上进行训练。

1　相关工作

基于客户端的 ABR 算法主要分为两种类型：基于模型的方法和基于学习的方法[3]。

第一类方法考虑了吞吐量的预测值和视频缓冲区大小等因素来选择比特率。文献 [4] 通过过去视频块大小和下载时间预测网络吞吐量，并以此作为未来吞吐量的估计值，估计值大时选择高视频比特率。另一些方法通过观察缓冲区大小来避免卡顿事件，并以此作为标准为下一个视频块选择尽可能高的比特率。文献 [5] 提出了一个线性标准阈值来控制可用的播放缓冲区大小。以 model predictive control (MPC)[6]为代表的混合策略综合考虑了吞吐量预测值和缓冲区大小，进行下一个视频块的比特率决策。此外，文献 [7] 研究了电池电量与移动流媒体 QoE 的关系。文献 [8] 提出的 Oboe 对现有 ABR 策略参数进行自动调整，使现有算法能够找到更佳的参数配置。此类方法往往针对某些网络条件，并在具有较强假设的前提下进行设计，严重依赖于微调的参数，难以适用于不同的网络环境。

基于学习的方法针对上述不足进行了改进，在获取到不同网络条件下的经验后，该类算法能够显著提高 ABR 的性能。D-DASH (a deep Q-learning frame work for dynamic adaptive streaming over HTTP)[9] 结合了深度学习和强化学习技术，利用深度 Q-learning 这种基于价值的强化学习方法优化视频的 QoE。在相同的网络条件下，Tiyuntsong[10] 用生成对抗网络，通过两个智能体的竞争来朝着规则或特定的奖励进行自我优化。Pensieve[11] 采用最新的 A3C[12] 算法生成 ABR算法模型，其中包含两个神经网络模型，一个用于比特率决策，另一个用于评估当前状态并给出状态价值，实验结果显示其性能优于基于模型的方法。HOT Dash[13] 将视频中的帧区分为热点和非热点，并将热点部分在带宽允许时优先传输，这样的做法使用户能够高质量的观看特定视频块。Comyco[14]针对该类方法采样效率低的缺陷，通过模仿即时求解器给出的专家轨迹来训练策略，这不仅可以避免多余的探索，还可以更好地利用收集的样本。

上述方法在训练时拥有相同的目标：最大化累计奖励值。基于学习的方法多采用线性 QoE 公式作为奖励函数，应用层的网络或播放器参数作为其输入，每一项参数给与固定的权重以表示对其的重视程度，但是权重的设置过程缺乏描述和依据。因此出现了一些采用机器学习的方法对用户的 QoE 进行预测。Video ATLAS[15] 是一种机器学习框架，其中结合了许多与 QoE 相关的特征，包括客观视频质量、卡顿以及记忆特征进行 QoE 预测。在此基础上，文献 [16] 采用非线性自回归外生模型来在连续时间上对 QoE 进行预测，在帧级别的粒度上测量 QoE，并利用了多模型联合预测来提升准确率。文献 [17] 选择长短期记忆网络 (long short-term memory, LSTM) 来捕捉 QoE 在时序上的依赖关系，并在真实的用户数据上展现了良好的性能。

综上，现有强化学习的训练目标都可以被描述成使预期的累计奖励值达到最大化，而基于 RL 的 ABR 算法输出比特率决策，视频播放器以该比特率请求下载下一个视频块。下载完成后状态发生转移，奖励函数以这些状态指标作为输入，计算得到下一步的奖励值，从而使算法模型沿着奖励值的梯度方向进行更新，因此奖励函数的设置对于算法性能具有重要影响。如果奖励函数设计未经充分考虑，一般会导致网络不收敛，结果不优或者使模型无法按照希望的方法做出决策。

已有基于 RL 的 ABR 算法均以量化的 QoE 作为奖励值。QoE 由播放中的指标如视频平均比特率、卡顿时间、比特率切换值等构成，每项指标赋予固定的权重表达对其重视程度。但由于用户的主观因素 (如期望、体验经历) 和环境因素，QoE 的量化十分复杂。奖励函数中权重的设置体现了用户

对不同指标的倾向，而定量描述用户对这样的事件的倾向，用以确定奖励函数的设置是一项难以实施的工作。本文认为，ABR 算法奖励值应该体现对播放质量变化事件的相应惩罚或奖励，应该针对用户 QoE 进行大量采样和训练建模，用以研究用户 QoE 预测的方法。因此本文提出了 UQPN，从用户数据出发训练得到 QoE 预测模型代替以往的函数，以此网络用于训练，能够获得更加符合用户需求的 ABR 算法模型。

2 详细设计

本文提出 UQPN 并让其加入 RL 训练过程，因为 RL 训练的目标为最大化累计奖励值，所以有了 UQPN 输出更准确的 QoE 预测值作为奖励，可以使 ABR 学会做出令用户 QoE 更佳的比特率决策。据此设计的 ABR 算法其整体系统结构如图 1 所示。

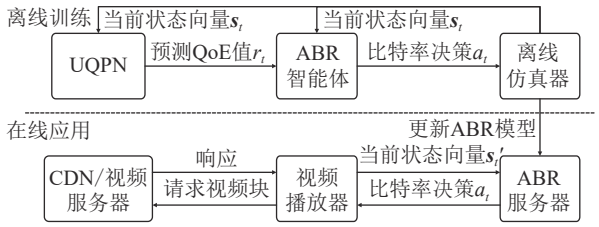

图 1 ABR 算法整体结构

该方法采用离线训练模式，在离线仿真器上读取收集好的数据进行训练。仿真训练时，设某一视频块下载完成的时间为 t，离线仿真器将不同的状态向量 $\boldsymbol{s}_t$ 和 $\boldsymbol{s}'_t$ 输出至 UQPN 和 ABR 智能体。UQPN 接受当前状态后将用户 QoE 预测值 r_t 输出给 ABR 智能体作为奖励用于训练。智能体接收该值并基于该值采用策略梯度法更新神经网络，随后发送下一视频块的码率决策 a_t 至仿真器，开始下一块的模拟下载。待训练过程结束后，将最新的 ABR 模型发送至在线使用的 ABR 服务器进行替换，为在线视频播放器提供自适应比特率服务。与离线训练阶段相比，在线应用阶段则无需奖励机制，由 ABR 获取播放器状态并做出码率决策即可。

2.1 UQPN 设计

使用 RL 生成 ABR 模型时，受到奖励函数建模困难和权重难以确定的困扰。因此，本节给出 UQPN 的设计细节，该网络能够接收当前视频流状态并输出当前 QoE 预测值。训练 UQPN 让其“学会”捕捉用户数据中潜藏的信息，以 UQPN 作为奖励能够使 ABR 模型做出更迎合用户需求的码率决策。

UQPN 的神经网络结构采用具有双隐藏层的多层感知器 (multi-layer perceptron, MLP) 结构。MLP 是一种前馈人工神经网络，层与层之间的节点进行全连接。每当下载完成一个视频块，设此时时刻为 t，UQPN 收到状态输入 $\boldsymbol{s}_t=\{x,n,m\}$，$x$ 为已下载完成的上个视频块的比特率；n 为上个视频块的比特率切换值；如果某个块的比特率和上一个块的比特率不相等，则有第 i 个块的比特率切换值 $n_i=|x_i-x_{i-1}|$。比特率切换值越大，代表用户观看视频时的质量波动越大；m 则表示上个视频块播放过程中的卡顿持续时间。在特征选择时，使用 Multi-RELIEF[18] 方法来筛选特征，它计算每个特征对于 QoE 贡献的权重。其中以现有数据集 LIVE-NFLX-II[19] 中记录的用户评分作为其观看的 QoE，各个视频流中记录的参数作为特征来计算权重。取权重值最大的前 3 位特征作为状态输入 $\boldsymbol{s}_t$。接收输入 $\boldsymbol{s}_t$ 后，UQPN 给出当前状态的 QoE 预测值，即奖励值 r_t。

UQPN 以梯度下降法训练，采用反向传播的方式更新网络节点，使训练集上的累计误差不断减小。训练开始之前，以输入层的下一层为第一层，为所有节点随机初始化权重和偏置值。其中设第 i 层的权重矩阵为 $\boldsymbol{W}^i$、偏置为 b^i。节点的激活函数采用 sigmoid，令其为 f。于是，第 i 隐藏层的节点激活值为：

$$\text{Output}^i = f((\boldsymbol{W}^i)^{\mathrm{T}}\boldsymbol{x}^{i-1}+b^i) \tag{1}$$

式中，$\boldsymbol{x}^{i-1}$ 为第 i−1 层所有节点的输出组成的矩阵。

最终输出层的激活值为：

$$r = f((\boldsymbol{W}^L)^{\mathrm{T}}\text{Output}^{L-1}+b^L) \tag{2}$$

式中，L 为神经网络层数。

训练时的损失函数定义为训练数据集之中用户观看视频给出的 QoE 得分 y 与 UQPN 输出值 r 之间差值的平方，并加入正则化项以防止过拟合。训练过程中通过反向传播算法逐一计算误差的偏导数，并以此更新神经网络参数来达到最小化累计误差的目标。

2.2 基于 UQPN 的 ABR 算法

在此基础上，本文提出一种强化学习的 ABR 算法。UQPN 训练完成后，令其加入 RL 训练，替代以往的奖励函数给出奖励值。该方法的基本训练算法使用 A3C，这是一种高效的 actor-critic 算法，

其中包括用于做出决策的 actor 网络和预测状态价值的 critic 网络。训练时采用策略梯度算法更新网络参数，梯度方向则是能使 UQPN 输出值增加的方向。

在每一个视频块完成下载的时刻 t，actor 网络接收状态观察向量 $\boldsymbol{s}'_t$ 并输出比特率决策 a_t。同样的，使用 Multi-RELIEF 方法筛选特征后，考虑到该方法网络结构的复杂性和客户端获取特征的可行性，定义状态观察向量$\boldsymbol{s}'_t=\{\boldsymbol{o}_t,\boldsymbol{n}_t,\boldsymbol{a}_t,\boldsymbol{e}_t,\boldsymbol{\tau}_t,B\}$，其中 $\boldsymbol{o}_t$为 t 之前的 k 个视频块的下载时的吞吐量测量值；$\boldsymbol{n}_t$为 t 之前的 k 个视频块的大小；$\boldsymbol{a}_t$表示 t 之前的 k 个视频块的比特率；$\boldsymbol{e}_t$储存了 t 之前的 k 个视频块各自下载时播放视频的卡顿时间；$\boldsymbol{\tau}_t$为 t 之前的 k 个视频块的下载时间；B 为当前播放器缓冲区的大小。

需要注意的是，actor 网络的实际输出并非某一确定值，而是一个概率分布。即在某一状态下特定比特率被选择的概率，将其标识为 $\pi(\boldsymbol{s}',a)$，输入状态和动作后输出概率。而具有可管理的、可调整的神经网络权重集 θ 的网络，标识为 $\pi_\theta(\boldsymbol{s}',a)$。因此，训练目标，即累积奖励相对于 θ 的梯度可表示为：

$$\nabla \mathrm{E}\left[\sum_{t=0}^{\infty}\gamma^t r_t\right]=\mathrm{E}_{\pi_\theta}\left[\nabla_\theta \log_{\pi_\theta}(\boldsymbol{s}',a)A^{\pi_\theta}(\boldsymbol{s}',a)\right] \tag{3}$$

因此，actor 网络的神经网络权重集 θ 更新公式为：

$$\theta \leftarrow \theta+\alpha\sum_{\mathrm{t}}\nabla_\theta \log_{\pi_\theta}(\boldsymbol{s}'_t,a_t)A_{\pi_\theta}(\boldsymbol{s}'_t,a_t)+\beta\nabla_\theta H(\pi_\theta(\cdot|\boldsymbol{s}'_t)) \tag{4}$$

式中，α 为学习速率。此外，在更新时加入一个正则化项，其中 β 为超参数，预先设置以表达对探索的重视程度。$A^{\pi_\theta}(\boldsymbol{s}',a)$ 为优势函数，代表当确定在状态 $\boldsymbol{s}'$下选择动作 a 时，预期总报酬与从策略中获取的预期报酬相比的差异。在式 (4) 中用 $A(\boldsymbol{s}',a)$ 作为$A^{\pi_\theta}(\boldsymbol{s}',a)$ 的无偏估计，而优势函数 $A(\boldsymbol{s}',a)$ 定义如下：

$$A(\boldsymbol{s}'_t,a_t)=r_t+\gamma V(\boldsymbol{s}'_{t+1})-V(\boldsymbol{s}'_t) \tag{5}$$

环境部署应用 a_t 后状态由 $\boldsymbol{s}'_t$ 转移至 $\boldsymbol{s}'_{t+1}$，$\boldsymbol{s}'_{t+1}$ 的预期奖励估计为 $V(\boldsymbol{s}'_{t+1})$，γ 为未来折扣系数，$\gamma\in[0,1]$。γ=1 时表示未来状态和当前状态同等权重。而 critic 网络接收 $\boldsymbol{s}'_t$ 后输出状态价值 $V(\boldsymbol{s}'_t)$，以评价当前状态好坏。对于 critic 网络的更新，使用时序差分法更新所有 critic 网络的神经网络权重集 θ_v，对于每次 t，critic 估计值和真实值之间的误差可以表示为：

$$\mathrm{Error}_t=(r_t+\gamma V(\boldsymbol{s}'_{t+1}|\theta_v)-V(\boldsymbol{s}'_t|\theta_v))^2 \tag{6}$$

θ_v 的更新公式为：

$$\theta_{\mathrm{v}} \leftarrow \theta_v-\mu\sum_t\nabla_{\theta_v}\mathrm{Error}_t \tag{7}$$

式中，μ为 critic 网络的学习速率。为提升训练速度，算法使用多个 ABR 智能体并行训练，每个智能体的输入不同。默认情况下，本文工作按照 Pensieve 建议，使用 16 个并行智能体。这样互不干预的独立训练可获得不同的经验。每个智能体将其获得的数据发送给中央智能体，该智能体会对其进行汇总以生成一个 ABR 算法模型。对于中央智能体接收到的每组数据，它都使用 actor-critic 算法来计算梯度并进行更新。最后，中央智能体更新 actor 网络，并将新模型返回给其余智能体使用。

2.3　模型更新

当前的客户端视频播放器网络条件多变，且流量行为变化复杂，为了保证 ABR 模型决策的有效性和对环境变化的适应性，算法中设置了触发更换新模型的触发机制。当客户端播放器播放视频时记录其吞吐量，即每次视频播放完成时都可以获得该次播放的网络吞吐量追踪。其次，播放完成或页面关闭时，向用户询问该次播放的 QoE 评分并记录。考虑到用户评分收集难度较大，当吞吐量追踪获取到一定数量时，离线 ABR 智能体在现有模型基础上进一步训练，结束后则进行模型更新，即将刚经过训练的 ABR 模型部署至在线的 ABR 服务器。具体算法如下。

算法 1：ABR 模型更新算法

输入：吞吐量追踪向量 $\boldsymbol{o}$，用户 QoE 评分向量 $\boldsymbol{N}$，现有 ABR 模型 π，UQPN 模型 r

输出：模型更新结果

初始化更新所需阈值 S;

if len(o)≥S

　if len(N)≥0

　for i=1: len(N)

　　以 N_i 作为训练数据更新 r;

　end for;

　end if;

　for i=1: len($\boldsymbol{o}$)

o_i 作为网络数据、r 输出奖励来更新 π；

end for;

将 π 部署至在线 ABR 服务器进行替换;

end if;

其中 S 为模型更新的吞吐量追踪数量阈值，该值应随客户端具体需求变化。当客户端网络条件变化较为频繁时，可以适当减小 S 以更多地更新模型。网络条件较为稳定时，可以适当增大 S 以减少更新次数。

3 仿真实验及结果

本节首先进行了相关性对比来验证 UPQN 的效果，然后对基于 UPQN 的 RL 奖励及其 ABR 算法进行了对比。其中相关性对比实验采用 LIVE-NFLX-II 数据集，包括训练所需的视频流信息和用户 QoE 信息。实验收集了由 15 个不同类型的视频、4 种不同的 ABR 算法、7 种不同的网络状态生成的视频流，以及由 65 个受试者给出的视频评分。对于每个视频流，在连续时间上生成了连续评分。数据集记录了视频卡顿状况和多种视频质量评价指标的变化。

对于 RL 奖励及其 ABR 算法，结合两个真实网络带宽数据集进行仿真：由 FCC 提供的宽带数据集[20] 和挪威收集的移动设备网络数据集[21]。仿真实验采用文献 [16] 提出的 QoE 预测方法作为评价标准，实验中包含多个测试视频流，每个视频流均需下载若干个视频块，因此实验采用每一视频块的平均 QoE 作为评价指标。

3.1 相关性对比

UPQN 为双隐藏层 MLP 结构，其中将第二隐藏层固定为 4 节点，进行第一层的节点实验性探索，发现第一隐藏层具有 12 节点时最优。因此论文使用上述 UQPN 网络结构在数据集上进行评估。表 1 使用了两个度量来对比 UQPN 模型与其他奖励函数的 QoE 预测的性能：线性相关系数 (linear correlation coefficient, LCC) 以及斯皮尔曼等级相关系数 (spearman rank order correlation coefficient, SROCC)。LCC 和 SROCC 度量的是两组数据之间的相关程度。对本实验来说，这两个指标值越大，预测的分值和 QoE 越接近。

表 1 可见经过真实用户数据训练得到的 UQPN 相比于现有方法的奖励函数平均提升了 12%～22.4% 的 LCC 和 11.6%～14.3% 的 SRCC。

表 1 UQPN 与其他奖励函数相关性对比

方法	LCC	SROCC
Pensive	0.6871	0.7244
MPC	0.7102	0.7324
Comyo	0.7507	0.7419
D-DASH	0.7043	0.7273
UQPN	**0.8413**	**0.8278**

3.2 不同 RL 奖励方法对比

本节主要考虑以下 3 种常用的 RL 算法：

Policy-Gradient：使用函数逼近器明确表示策略，并根据预期奖励相对于策略参数的梯度进行更新，并证明了具有任意可微函数逼近器的策略迭代之后可以收敛到局部最优策略。

A2C：A2C 是一种改进的 actor-critic 算法，使用优势函数代替 critic 网络中的原始奖励，可以作为衡量被选取动作值和所有动作平均值好坏的指标。

A3C：神经网络训练时，需要的数据是独立同分布的，因此 A3C 采用异步训练的方法，打破数据的相关性并加速了训练过程。

之后将每种 RL 算法中的奖励设定为由 3 种方法给出：Pensieve、Comyco 和 UQPN。其中前两者均为线性函数，由播放中比特率、卡顿时间等指标与固定权重的乘积组成。实验中 RL 智能体采用的输入与 Pensieve 中一致，并设置所有 ABR 模型训练次数为 10 000 次。实验中训练、测试数据均为离线仿真器读取网络带宽数据并模拟下载特定视频得出。

如图 2 所示，在两种用于测试的网络带宽数据下，与 Pensieve、Comyco 提出的奖励函数相比，UQPN 在 3 种不同的 RL 算法上的性能均更优，展现了良好的泛化能力。在 A3C 方法上 UQPN 的优势最为明显，相比另外两种方法的平均归一化 QoE 在挪威数据集上带来约 27.9% 提升并在 FCC 数据集上带来约 27.7% 的 QoE 提升。而在 A2C 方法上，UQPN 能够带来平均约 27.2% 和 18.3% 的 QoE 提升，在 Policy-Gradient 上则有约 15.4% 和 8.6% 的 QoE 性能上升。这是由算法的学习能力导致，A3C 算法的学习能力最强，同样的训练次数下更能够发挥 UQPN 的优势。而 Policy-Gradient 则相反，不同的奖励方法带来的差异并不明显。

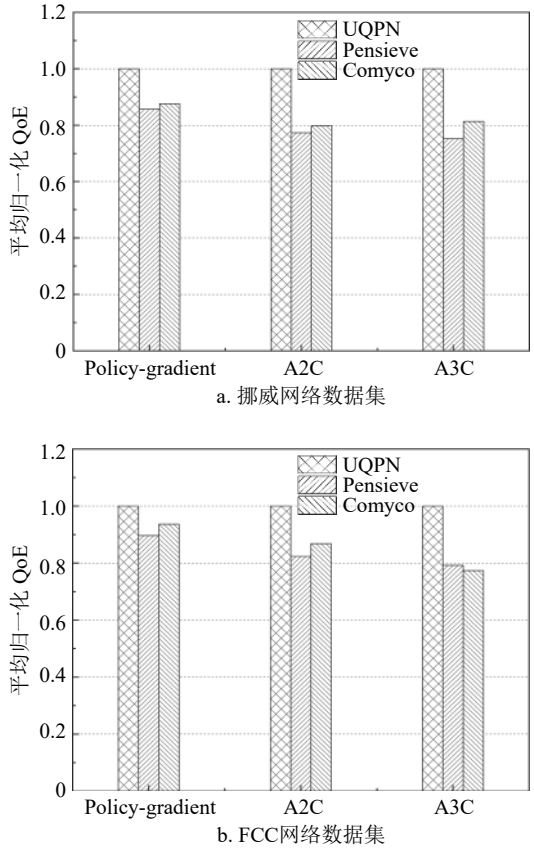

图 2　3 种奖励方法得到的 ABR 模型 QoE 对比

3.3　基于 RL 的 ABR 算法性能对比

图 3 给出了在两种用于测试的数据集下基于 UQPN 的强化学习 ABR 算法与 D-DASH 和 Pensieve 的对比。

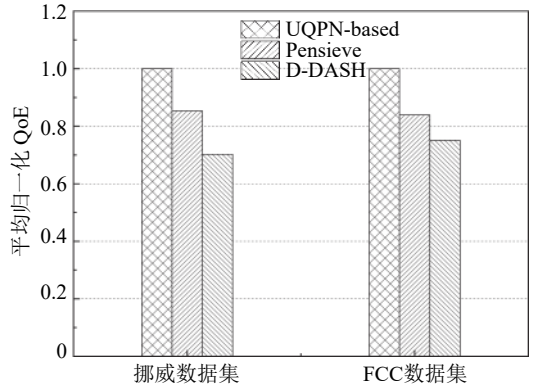

图 3　3 种基于 RL 的 ABR 方法性能对比

结果显示，该方法分别能够带来平均约 29.6% 和 26.1% 的归一化 QoE 提升。这意味着，将 UQPN 和本文提出的训练方法结合，以 UQPN 输出值作为奖励，使 ABR 策略模型在强化学习时有了更准确、合理的训练目标，因此该方法在宽带和移动两种真实网络条件下都能够比现有的基于 RL 的 ABR 方法给用户带来更好的 QoE。

4　结 束 语

本文提出了一种使用用户 QoE 预测值作为强化学习奖励的自适应比特率算法。有了更加准确的 QoE 预测值加入训练，该方法能够给用户带来更好的观看体验。该方法采用离线训练，仅基于收集的数据即可生成算法模型，其输入参数也易于获取，无需修改现有的流媒体视频传输框架，具备较好的可行性。未来的工作中，将考虑采用更细粒度、更准确的方法来探索用户在观看视频时的 QoE 变化，能够更准确把握用户在观看视频时的感受，为用户提供更好的观看体验。

参　考　文　献

[1] ITU-T Recommendation. Definition of quality of experience (QoE)[EB/OL]. [2007-07-16]. https://www.itu.int/rec/T-REC-P.10-201607-S!Amd5/en.

[2] MOK R K P, CHAN E W W, CHANG R K C. Measuring the quality of experience of HTTP video streaming[C]// Proceedings of the 12th IFIP/IEEE International Symposium on Integrated Network Management. Dublin, Ireland: IEEE, 2011: 485-492.

[3] BENTALEB A, TAANI B, BEGEN A C, et al. A survey on bitrate adaptation schemes for streaming media over HTTP[J]. IEEE Communications Surveys & Tutorials, 2019, 21(1): 562-585.

[4] JIANG J, SEKAR V, ZHANG H. Improving fairness, efficiency, and stability in HTTP-based adaptive video streaming with festive[J]. IEEE/ACM Transactions on Networking, 2014, 22(1): 326-340.

[5] HUANG T Y, JOHARI R, MCKEOWN N, et al. A buffer-based approach to rate adaptation: Evidence from a large video streaming service[J]. ACM Sigcomm Computer Communication Review, 2014, 44(4): 187-198.

[6] YIN X, JINDAL A, SEKAR V, et al. A control-theoretic approach for dynamic adaptive video streaming over HTTP[J]. ACM Sigcomm Computer Communication Review, 2015, 45(4): 325-338.

[7] 葛志辉, 张旭锋, 宋玲, 等. 基于电池电量感知的移动流媒体 QoE 优化策略[J]. 北京邮电大学学报, 2018, 41(6): 78-82.
GE Zhi-hui, ZHANG Xu-feng, SONG Ling, et al. Power-aware video QoE optimization strategy for mobile video streaming[J]. Journal of Beijing University of Posts and Telecommunications, 2018, 41(6): 78-82.

[8] AKHTAR Z, NAM Y S, GOVINDAN R, et al. Oboe: Auto-tuning video ABR algorithms to network conditions[C]// 2018 Conference of the ACM Special Interest Group. Budapest: ACM, 2018: 44-58.

[9] GADALETA M, CHIARIOTTI F, ROSSI M, et al. D-DASH: A deep Q-learning framework for DASH video streaming[J]. IEEE Transactions on Cognitive

Communications & Networking, 2017(4): 1.

[10] HUANG T, YAO X, WU C, et al. Tiyuntsong: A self-play reinforcement learning approach for ABR video streaming[C]//2019 IEEE International Conference on Multimedia and Expo (ICME). Shanghai: IEEE, 2019: 1678-1683.

[11] MAO H, NETRAVALI R, ALIZADEH M. Neural adaptive video streaming with pensieve[C]//Conference of the ACM Special Interest Group on Data Communication. Los Angeles: IEEE, 2017: 197-210.

[12] MNIH V, BADIA A P, MIRZA M, et al. Asynchronous methods for deep reinforcement learning[C]//The 33rd International Conference on International Conference on Machine Learning (ICML). New York: ACM, 2016: 1928-1937.

[13] SENGUPTA S, GANGULY N, CHAKRABORTY S, et al. HotDASH: Hotspot aware adaptive video streaming using deep reinforcement learning[C]//IEEE International Conference on Network Protocols. Cambridge: IEEE, 2018: 165-175.

[14] HUANG T, ZHOU C, ZHANG R, et al. Comyco: Quality-aware adaptive video streaming via imitation learning[C]//The 27th ACM International Conference on Multimedia. Nice: ACM, 2019: 429-437.

[15] BAMPIS C G, BOVIK A C. Feature-based prediction of streaming video QoE: Distortions, stalling and memory[J]. Signal Processing Image Communication, 2018(68): 218-228.

[16] BAMPIS C G, LI Z, KATSAVOUNIDIS I, et al. Recurrent and dynamic models for predicting streaming video quality of experience[J]. IEEE Transactions on Image Processing, 2018, 27(7): 3316-3331.

[17] ESWARA N, ASHIQUE S, PANCHBHAI A, et al. Streaming video QoE modeling and prediction: A long short-term memory approach[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(3): 661-673.

[18] YE K, ANTON F K, HERINGA J, et al. Multi-RELIEF: A method to recognize specificity determining residues from multiple sequence alignments using a machine-learning approach for feature weighting[J]. Bioinformatics, 2007, 24(1): 18-25.

[19] DUANMU Z, REHMAN A, WANG Z. A quality-of-experience database for adaptive video streaming[J]. IEEE Transactions on Broadcasting, 2018, 64(2): 474-487.

[20] Federal Communications Commission. Raw data-measuring broadband america[EB/OL]. [2020-6-8]. https://www.fcc.gov/reports-research/reports/measuring-broadband-america.

[21] HAAKON R, PAUL V, CARSTEN G, et al. Commute path bandwidth traces from 3G networks: Analysis and applications[C]//The 4th ACM Multimedia Systems Conference (MMSys). New York: ACM, 2013: 114-118.

编 辑 叶 芳