基于强化学习的移动视频流业务码率自适应算法研究进展

杜丽娜[1,2]，卓力[1,2]，杨硕[1,2]，李嘉锋[1,2]，张菁[1,2]

（1. 北京工业大学计算智能与智能系统北京市重点实验室，北京 100124；2. 北京工业大学信息学部，北京 100124）

摘 要：近几年来，随着 HTTP 自适应流媒体（HAS）视频数据集和网络轨迹数据集的不断推出，强化学习、深度学习等机器学习方法被不断应用到码率自适应（ABR）算法中，通过交互学习来确定码率控制的最优策略，取得了远超过传统启发式方法的性能。在分析 ABR 算法研究难点的基础上，重点阐述了基于强化学习（包括深度强化学习）的 ABR 算法研究进展。此外，总结了代表性的 HAS 视频数据集和网络轨迹数据集，介绍了算法性能的评价准则，最后探讨了 ABR 研究目前存在的问题和未来的方向。

关键词：强化学习；码率自适应算法；用户质量体验；深度学习；深度强化学习

中图分类号：TP391

文献标识码：A

DOI: 10.11959/j.issn.1000−436x.2021178

Survey on reinforcement learning based adaptive bit rate algorithm for mobile video streaming services

DU Li'na[1,2], ZHUO Li[1,2], YANG Shuo[1,2], LI Jiafeng[1,2], ZHANG Jing[1,2]

1. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China
2. Information Department, Beijing University of Technology, Beijing 100124, China

Abstract: In recent years, with the continuous release of HTTP adaptive streaming (HAS) video datasets and network trace datasets, the machine learning methods, such as deep learning and reinforcement learning, have been continuously applied to adaptive bit rate (ABR) algorithms, which obtain the optimal strategy of rate control through interactive learning, and achieve superior performance that surpasses the traditional heuristic methods. Based on the analysis of the research difficulties of ABR algorithms, the research advances of ABR algorithms based on reinforcement learning (including deep reinforcement learning) was investigated. Furthermore, several representative HAS video datasets and network trace datasets were summarized, the evaluation metrics of the performance were depicted. Finally, the existing problems and the future tendency of ABR research were discussed.

Keywords: reinforcement learning, ABR algorithm, QoE, deep learning, deep reinforcement learning

1 引言

随着移动通信技术和移动智能终端的迅猛发展，移动互联网用户数量以及移动视频业务流量呈指数级增长趋势。思科公司的报告预测，到 2022 年，视频业务流量将占据全球互联网数据流量的 82%[1]。与传统互联网在线视频业务相比，移动环境下的视频业务呈现出空间移动性、时间碎片性以及社交关联性等新的复杂特性。面向复杂的移动环境以及移动视频业务所呈现出的新特性，如何在有限的网络资源条件下保障移动视频业务的用户体验质量（QoE, quality of experience），成为当下学术界和工业界共同关注的热点课题。

QoE 是指用户对设备、网络、系统、应用或业

收稿日期：2021−03−15；**修回日期**：2021−06−10

通信作者：卓力，zhuoli@bjut.edu.cn

基金项目：国家自然科学基金资助项目（No.61531006）；北京市教委−市基金联合资助项目（No.KZ201910005007）

Foundation Items: The National Natural Science Foundation of China (No.61531006), Beijing Municipal Education Commission Cooperation Beijing Natural Science Foundation (No.KZ201910005007)

务的质量和性能的整体主观感受[2]，反映了用户在接受服务时的满意或舒适程度。国际电信联盟[3]对 QoE 的定义为“用户使用一项应用或服务时感到的快乐或烦恼程度”。

目前的移动视频流媒体业务普遍采用基于 HTTP 的自适应流媒体（HAS, HTTP adaptive streaming）技术进行传输。HAS 技术可以根据网络的时变、波动特性以及客户端的播放状态自适应调整视频码率，有效提高用户的质量体验。2012 年，MPEG 和 3GPP 联合推出了基于 HTTP 的动态自适应流媒体（DASH, dynamic adaptive streaming over HTTP）国际标准[4]，以满足移动流媒体业务日益发展的需求。除此之外，工业界也推出了各种 HAS 协议，如 Adobe 公司推出了 HTTP 动态流媒体（HDS, HTTP dynamic streaming）[5]，苹果公司推出了 HLS（HTTP live streaming）[6]，微软公司推出了 MSS（Microsoft smooth streaming）[7]等。由于 HAS 技术具有兼容性好、扩展性强、易部署等优点，目前已被广泛应用于各种移动流媒体业务系统中。

HAS 流程如图 1 所示。对原始视频以不同的码率进行编码，将每个视频码流分割成不同的片段存储于服务器端。客户端根据当前的网络状况、自身硬件的处理能力、缓存状态等，动态调整视频流的码率，以提升用户的 QoE。可以看出，在客户端部署的码率自适应（ABR, adaptive bit rate）算法，是 HAS 技术的核心所在。

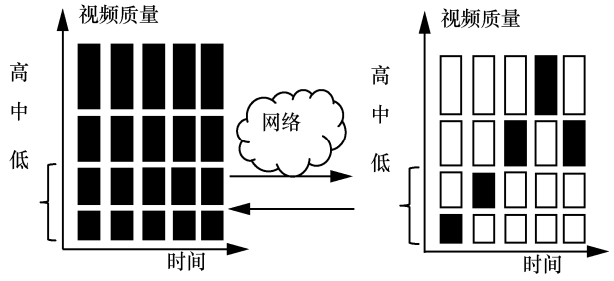

图 1 HAS 流程

在实际应用过程中，ABR 算法仍然面临许多的困难与挑战。这些挑战主要来自以下几个方面。

1) 网络状态的多变性以及不可预测性。

2) 决策的累积效应。即前面的决策会影响未来的码率选择，这就需要 ABR 算法决策具有一定的前瞻性和预测性。

3) 实时性。ABR 算法的决策时间不能过长。

4) 公平性和稳定性。多个客户端同时通过瓶颈链路竞争共享带宽时，面临公平性和稳定性问题。

5) QoE 影响因素众多且难以量化。QoE 的影响因素既包括主观因素，又包括客观因素，这些因素相互影响，尤其是主观因素难以量化。如何精确地预测用户 QoE 是设计 ABR 算法不可回避的问题。

近年来，随着强化学习（RL, reinforcement learning）和深度学习在各个领域的广泛应用，学者将其应用于 ABR 算法中，取得了诸多有意义的成果。这些方法以直接优化用户的 QoE 为目的，从大量数据中学习码率自适应的最优策略，可以获得比传统启发式方法更好的性能，因此逐步成为目前 ABR 研究的主流方向。

目前国内外已经有关于 ABR 算法的研究综述，比如，2017 年，Kua 等[8]从客户端、服务器端和网络内部 3 个角度出发，总结了 ABR 算法的研究进展。此外，作者还阐释了直播和点播之间的区别以及具体实施过程中媒体呈现描述（MPD, media presentation description）文件的差异。2018 年，Ayad 等[9]从代码层面分析了 DASH 标准开源播放器 DASH-IF、谷歌的 DASH Media Source 以及 Bitmovin 播放器的实现细节，并评估了上述 3 个播放器以及 Netflix、YouTube、Vimeo 共 6 个播放器的性能。本文则对近几年来出现的基于 RL（包括深度强化学习）的 ABR 算法进行了综述，对代表性 ABR 算法的思路、策略、性能等进行了深入的分析总结。

2 基于强化学习的 ABR 算法研究进展

2.1 强化学习

RL 的基本架构如图 2 所示。RL 包含 3 个基本要素：智能体、环境和奖励函数。其基本原理为：智能体通过从环境中观察到的状态选择一个动作用于环境，环境接受该动作更新状态并产生一个强化信号（奖或惩）反馈给智能体。如果智能体的某个行为策略导致环境正的奖赏，那么智能体产生这个行为策略的趋势便会加强。智能体的目标是在每个离散状态下发现最优策略，以使期望的折扣奖赏最大，并根据强化信号和当前状态再选择下一个动作。选择的动作不仅影响当前的强化值，还影响环境下一时刻的状态及最终的强化值。

不同于监督学习和非监督学习，RL 把学习看作试探评价过程，侧重在交互中学习，由智能体在与环境的交互过程中根据获得的奖励或惩罚不断地学习，最终学习到更加适应环境的策略。

图 2　强化学习基本架构

2.2　基于强化学习的 ABR 算法

RL 为 ABR 算法的设计提供了一种新的技术手段[10]，其本质可以看成高维状态空间的特征选择[11]或者序列决策。在不同的前提条件下，RL 也可以转化成动态规划（DP, dynamic programming）、马尔可夫决策过程（MDP, Markov decision process）、隐马尔可夫模型（HMM, hidden Markov model）等。

目前，大多数 ABR 算法的研究工作集中在 QoE 建模和训练最佳策略来提高用户 QoE 等方面。将 RL 应用于码率选择的整体架构如图 3 所示。智能体的输入是前几个视频片段的码率以及下载时间、下一个视频片段的可选码率等信息，输出是拟选择的码率。环境包括网络状态和客户端播放器的状态等，奖励函数普遍用 QoE 模型表达。人们通过设计不同的算法来定义智能体如何做出决策。

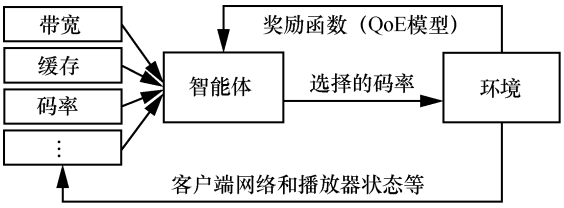

图 3　将强化学习应用于码率选择的整体架构

表 1 是目前比较具有代表性的几种基于强化学习的 ABR 算法。从表 1 可以看出，这些 ABR 算法采用的 RL 方法主要包括 Q-learning、HMM、MDP 和 DP 等，其中 Q-learning 是最常用的算法。Q-learning 的经典实现方式是列举出所有的状态和动作，构建 Q 函数表，然后迭代计算各种状态下执行每个动作的预期回报的最大值。例如，Claeys 等[12]采用 Q-learning 方法来解决 ABR 问题，将环境模型划分成 2 500 万个状态，由于状态空间太大，导致该模型难以收敛。作者之后进行了改进，减少了环境状态的组成元素，并重新设计了奖励函数，提升了性能[13]。由于 Q-learning 方法在计算和存储 Q 函数表时需要耗费大量的时间与空间，因此基于 Q-learning 的 ABR 算法需要在 Q 函数表空间−时间复杂度和 QoE 性能之间进行折中。

RobustMPC[19]算法是一种基于控制理论的 ABR 算法，该算法综合考虑带宽预测器预测的带宽、缓冲区状态以及用户的 QoE 来选择码率，所采用的 QoE 模型被广泛应用于各种基于 RL 的 ABR 算法中，如式(1)所示。

$$\mathrm{QoE}=\alpha\sum_{n=1}^{N}q(R_n)-\beta\sum_{n=1}^{N}T_n-\gamma\sum_{n=1}^{N-1}\left|q(R_{n+1})-q(R_n)\right| \quad (1)$$

其中，R_n 表示第 n 个视频片段的码率，$q(R_n)$表示视频质量，T_n 表示卡顿时长，最后一项表示平滑度，α、β、γ表示每一项的权重。

RobustMPC 算法对带宽预测的准确性要求较高，当带宽预测不够准确时，算法性能会出现明显下降。为此，Sun 等[14]利用隐马尔可夫模型进行带宽预测，该算法在带宽剧烈抖动的情况下依然有较好的性能。Chiariotti 等[15]采用并行学习技术，可提高学习率并限制次优选择，从而实现快速而准确的

表 1　　**基于强化学习的代表性 ABR 算法**

ABR 算法	强化学习方法	奖励函数
文献[12]	Q-learning	$R=C_1R_{\text{quality}}+C_2R_{\text{oscillation}}+C_3R_{\text{bufferfilling}}+C_4R_{\text{bufferchange}}$，其中，$R_{\text{quality}}$ 表示视频质量，$R_{\text{oscillation}}$ 表示质量切换，$R_{\text{bufferilling}}$ 表示缓冲区占有率，$R_{\text{bufferchage}}$ 表示缓冲区的变化，C_1、C_2、C_3、C_4 表示每一项的权重
文献[13]	Q-learning	$R=R_{\text{quality}}+R_{\text{switches}}+R_{\text{bufferfilling}}$，其中，$R_{\text{quality}}$ 表示视频质量，R_{switch} 表示质量切换，$R_{\text{bufferilling}}$ 表示缓冲区状态
文献[14]	HMM	$\mathrm{QoE}=\alpha\sum_{n=1}^{N}q(R_n)-\beta\sum_{n=1}^{N}T_n-\gamma\sum_{n=1}^{N-1}\left\|q(R_{n+1})-q(R_n)\right\|$，其中，$R_n$ 表示第 n 个视频片段的码率，$q(R_n)$表示视频质量，T_n 表示卡顿时长，最后一项表示平滑度，α、β、γ 表示每一项的权重
文献[15]	MDP	$R=q_t-\beta\left\|q_t-q_{t-1}\right\|$，其中，$q_t$ 表示视频质量，β表示权重
文献[16]	DP	$y(q,\lambda)=\left(\prod_{i=0}^{N-1}q_i\right)^{\frac{1}{N}}-\frac{\lambda}{N-1}\sum_{i=0}^{N-2}(q_i-q_{i+1})^2$，其中，$q_i$ 表示第 i 个视频片段的质量，N 表示视频片段数量，λ表示权重，第一项表示视频质量，第二项表示视频质量的波动
文献[17]	DP	$\max\sum_n d(n-1,n)U_a(Q(n))$，其中，$U_a(Q(n))$ 表示 α 的公平性[18]，$Q(n)$ 表示第 n 个片段的码率。若发生质量切换，$d(n-1,n)=1$；否则，$d(n-1,n)=0.9$

学习过程，并迅速收敛于稳定的奖励。通过选择最佳码率，使长期预期奖励最大化。

还有一部分研究者采用 DP 来解决 ABR 问题。例如 Andelin 等[16]采用 DP 来解决码率的最优选择问题，设计奖励函数时考虑了视频质量以及视频质量切换等因素。仿真网络下的实验结果表明，该算法可以有效提升用户的 QoE。

PANDA/CQ[17]是另一种比较具有代表性的使用 DP 的 ABR 算法，该算法的重点在于奖励函数的设计，目的是减少不必要的质量切换对 QoE 产生的影响。García 等[20]提出基于随机 DP 的 ABR 算法，目的是学习到保障用户 QoE 的最佳请求策略。该算法的奖励函数是视频质量、卡顿频率和时长以及质量切换的线性表达式，通过对卡顿、质量切换等不同指标的实验结果分析，证明该算法可以实现视频质量和卡顿之间的均衡。

2.3 小结

总体来看，基于 RL 的 ABR 算法以最优化 QoE 为目标，通过数据驱动的方式训练 ABR 算法。算法不依赖预先设计的模型或对环境的假设，通过观察和实验逐渐学习到码率自适应的最佳策略，取得了比启发式方法更好的性能。然而，采用传统的 RL 算法设计的 ABR 算法存在维度灾难以及收敛缓慢的问题。受限于算法对于高维数据的表达能力，基于 RL 的 ABR 算法的动作空间和样本空间都很小，难以应对具有很大状态空间和动作空间的情况。

3 基于深度强化学习的 ABR 算法研究进展

近年来，深度学习在计算机视觉、自然语言处理、语音识别等领域取得了极大的成功。研究者将深度学习与 RL 相结合，提出各种深度强化学习（DRL, deep reinforcement learning）算法，如 2013 年 Mnih[21]提出的深度 Q 网络（DQN, deep q-network），可利用深度学习自动提取大规模输入数据的抽象表征，使深度强化学习能够解决各种复杂的决策任务。

2017 年以来，研究者开始将 DRL 应用到 ABR 算法中，Pensieve 算法[22]是使用 DRL 解决 ABR 问题最有代表性的算法之一，已成为目前基于 DRL 的 ABR 算法的基准。Pensieve 算法采用 Actor-Critic[23]算法实现 ABR，其算法结构如图 4 所示。智能体的输入是上一个视频片段的下载时间、带宽估计值、下一个视频片段的可选大小、当前缓冲区占有率、剩余视频片段数量以及前几个视频片段的码率，输出是所选择的码率，奖励函数是 RobustMPC 算法采用的 QoE 模型，如式(1)所示。深度神经网络（DNN, deep neural network）由一维卷积神经网络（CNN, convolutional neural network）和全连接层组成，输出为下一个视频片段的码率。DNN 作用是提高算法对高维数据的表达能力，进而缓解“维度灾难”问题。

具体而言，DNN 用来拟合 DRL 中的各种函数，如价值函数和策略函数等。研究结果表明，在强化学习中引入 DNN，有助于学习到更好的策略，从而获得更优的性能。

2018 年，Paul 等[24]对 Pensieve 算法进行了扩展，在低带宽环境下测试了 Pensieve 算法的性能。尽管 Pensieve 取得了远超启发式算法的性能，但是依然存在很大的改进空间。为此，研究者从奖励函数、深度神经网络结构、实时性等角度出发，提出了各种基于 DRL 的 ABR 算法。

3.1 奖励函数设计

QoE 的影响因素众多，不同的因素对 QoE 有着

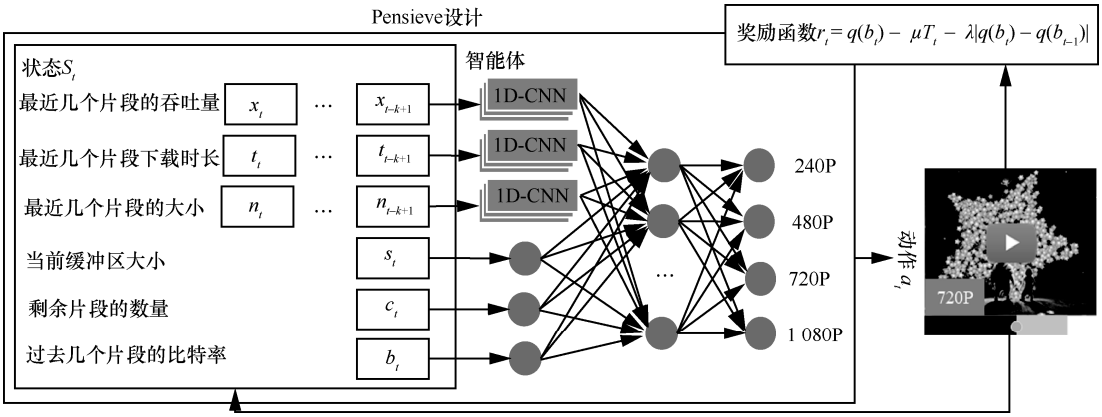

图 4 Pensieve 算法结构

不同的影响，而现有的用作奖励函数的 QoE 模型忽略了视频的内容特性、用户的行为偏好等影响因素。为此，研究者从不同的角度对奖励函数进行了改进。

1) 考虑用户的行为偏好、兴趣度等主观因素对 QoE 的影响

Sengupta 等[25]提出 HotDASH 算法，考虑了用户的行为偏好，将用户对视频片段的偏好程度写入 MPD 文件，以便环境将该状态信息传递给智能体，预提取满足用户偏好的视频片段。其奖励函数在式(1)的基础上考虑了用户的偏好，采用 Actor-Critic 算法训练策略。实验结果表明，该算法的性能优于现有的启发式方法以及 Pensieve 等基于 DRL 的 ABR 算法。

Gao 等[26]认为，大多数 ABR 算法都没有考虑到视频的语义信息，而语义信息在很大程度上决定了视频内容的信息量和趣味性，从而影响用户的 QoE。为此，作者提出了一种基于内容兴趣度（CoI, content of interest）的 ABR 算法。首先设计一种用于识别视频 CoI 的深度学习方法，然后将视频的 CoI 信息作为智能体的一个输入，采用 DQN 方法实现 ABR 算法。

文献[27]提出了一种情感内容感知的 ABR 算法。作者考虑了用户对情感内容（AC, affective content）的偏好，将视频帧的 AC 分为 6 类，采用 ResNet 网络对视频帧的 AC 进行分类，并输出其属于每一类别的置信度。之后对视频片段中所有帧的类别置信度求平均，作为整个视频片段的置信度，并将用户的情感偏好加入奖励函数中，根据用户的情感偏好来优化 QoE。

这些方法在进行 QoE 建模时，充分考虑了用户的主观因素，建立的模型可以更好地表征人的感知，因此获得了更好的 QoE 性能。

2) 考虑视频内容特性对 QoE 的影响

文献[28]观察到用户在视频的某些部分对低质量的敏感度高于其他部分。为此，该算法对每个视频进行众包实验，以得出用户在视频不同部分的质量敏感性，将敏感性作为权重对 KSQI 模型[29]进行修正，并将该模型应用到 ABR 算法中。实验结果表明，这些改进可以有效提高 QoE。

文献[30]建立了一种连续性 QoE 模型，该模型考虑了前一段时间出现的卡顿、平均的视频质量对当前 QoE 的影响。研究结果表明，将该模型作为奖励函数，其性能要优于 Pensieve 算法。

上述的研究结果表明，通过改进奖励函数，可以在一定程度上提高 ABR 算法的性能。

3.2　深度强化学习方法

学者将各种先进的深度学习方法与强化学习相结合，提升 ABR 算法的有效性。

Tiyuntsong 方法[31]将生成对抗网络与 Actor-Critic 相结合，生成对抗网络用来从过去的状态中提取隐藏的特征，采用 Self-Play 深度强化学习方法，通过训练 2 个相互竞争的智能体来自动学习码率控制策略。

Zwei 方法[32]将 DRL 与蒙特卡罗算法相结合，在蒙特卡罗搜索过程中，根据当前策略从起始状态中采样多个轨迹，通过平均每个轨迹对的竞争结果来估计预期的长期获胜率，其中结果表示哪个轨迹更接近 2 条轨迹之间的实际需求。估计获胜率后，通过增加获胜样本的概率和减少失败样本的可能性，采用近端策略优化[33]来优化 DNN，以获得更优的策略。

针对现有的 ABR 算法主要侧重于优化所有客户端的整体 QoE，而忽略了不同用户 QoE 多样性的问题，Huo 等[34]提出了多用户感知的元学习方法，该方法将元学习和多任务 DRL 相结合，通过主观实验分析验证了不同用户在 QoE 影响因素上的差异，并且量化这个差异。在此基础上，作者修正了奖励函数，借鉴 Frans 等[35]提出的方法建立了一种多任务深度强化学习框架，实现了多用户偏好的 QoE 优化。

深度学习与强化学习相结合，虽然可以利用深度学习强大的表征能力提升性能，但同时也面临样本效率低、难以收敛、计算复杂度高等问题。为此，很多学者针对性地提出了各种解决方案。

针对无模型深度强化学习方法难以收敛等问题，Liu 等[36]采用支持向量回归（SVR, support vector regression）建立了一个 chunk-wise 感知的奖励函数，将 Pensieve 算法中的 Actor-Critic 算法替换为深度 Q-learning 算法，并加入 double Q-learning、Dueling network 以及 multi-step 等机制对深度 Q-learning 网络进行改进，使网络训练能更快收敛，并获得更高的平均 QoE。

Saleem 等[37]提出了一种基于 double deep Q-learning 的 ABR 算法，并对算法的性能进行了较全面的评估，主要评价指标有峰值信噪比（PSNR,

peak signal to noise ratio）、结构相似性（SSIM, structural similarity index measure）、卡顿频率和质量切换等客观指标以及用户的 QoE。实验结果表明，与 Actor-Critic 方法相比，采用 double deep Q-learning 具有更快的收敛速度。

针对现有方法样本效率低和对视频质量信息感知不足的缺点，Huang 等[38]提出了 Comyco 算法，该算法主要包括及时求解器、仿真播放器、经验池以及神经网络 4 个模块。其中，及时求解器负责提供专家策略，经验池负责存储专家策略，神经网络负责训练策略、做出决策。

Comyco 算法利用模仿学习解决强化学习样本效率低的问题，以视频的感知质量作为奖励函数。此外，作者认为视频质量上升和下降对 QoE 的影响不同，因此对奖励函数进行了改进，采用视频多方法评价融合（VMAF, video multimethod assessment fusion）准则来度量视频质量 $q(R_n)$。

DAVS[39]是一种基于学徒学习的 ABR 算法，该算法不需要奖励函数作为指导，而是从专家示例中学习更好的策略。考虑到视频内容上的差异性，将视频分为动态片段（纹理丰富度高且运动复杂度高）和静态片段，并为动态片段分配更高的质量。

3.3 深度神经网络结构设计

随着深度学习的不断发展，人们提出了各种 DNN 结构，众多研究者对 DRL 算法中的 DNN 进行了改进。

2017 年，Gadaleta 等[40]提出了 D-DASH 方法。该方法采用 deep Q-learning 方法实现码率自适应，其中的深度神经网络采用的是多层感知机以及长短期记忆（LSTM, long short-term memory）等网络结构。实验结果表明，与基于强化学习的 ABR 算法相比，D-DASH 方法表现出更快的收敛速度以及更好的性能。

LASH 是 Lekharu 等[41]提出的一种 ABR 算法，该算法采用 LSTM 代替部分 1D-CNN，将输入部分分为两组：一组输入 LSTM 网络，另一组输入全连接层，将两部分的输出合并后，输入下一层全连接层。实验结果表明，加入 LSTM 使模型能够更准确地学习到时序特征。与 Pensieve 算法相比，平均 QoE 提升了约 8.84%。

总体来说，深度神经网络可以有效解决传统强化学习算法中“维度灾难”的问题，提高算法对于高维数据的表达能力，通过对深度强化学习中的价值函数、策略函数等进行有效拟合，使网络可以学习到更好的策略。

3.4 面向网络传输需求的改进

移动视频流在网络传输过程中，要综合考虑网络状况、视频编码方式、实时传输需求等多种因素。很多学者针对这些因素开展了研究工作。

单个方案往往无法对复杂多变的网络情况进行充分的建模。为此，有学者通过集成多个 ABR 算法来提高用户的 QoE。比如，Zhang 等[42]提出了集成自适应流媒体（EAS, ensemble adaptive streaming）方法，该方法首先对网络环境进行分类，之后为每种网络环境分别设计了相应的 ABR 算法以适应其特性，依据网络状态动态匹配合适的算法。实验结果表明，EAS 可以应对更具挑战性的网络环境。

Zhao 等[43]设计了时延控制模型来控制跳帧，并通过集合 2 个不同网络轨迹下训练出来的 ABR 算法来提高用户的 QoE。Akhtar 等[44]提出了一种 Oboe 方法。该方法首先对不同的网络状态进行建模，并训练出不同网络环境的最佳参数，之后在线检测网络状态改变点，并自适应调整参数。实验结果表明，Oboe 方法显著提升了 RobustMPC、Pensieve 等算法的性能。

众所周知，视频传输质量在很大程度上取决于可用带宽资源。为此，Yeo 等[45]提出 NAS 方法，应用图像超分辨率 DNN 来提升视频的增强质量，最大化用户的 QoE。在进行码率控制时，利用深度强化学习来选择下载的视频片段的码率和 DNN 块，降低了视频传输质量对带宽资源的依赖程度。

针对视频流的实时传输需求，研究者开展了 ABR 算法的研究。QARC 是 Huang 等[46]提出的一种面向实时视频流的码率自适应算法，该算法包括视频质量预测网络（VQPN, video quality prediction network）和视频质量增强学习（VQRL, video quality reinforcement learning）2 个子网络，VQPN 根据历史视频帧的质量预测当前视频帧的质量，其输入是前几个视频帧，输出是下一时刻视频帧的质量；VQRL 则依据网络状态和 VQPN 的预测情况选择合适的码率。VQRL 采用 Actor-Critic 算法，输入是 VQPN 预测的视频帧的质量、最近 k 个发送和接收到的视频帧的码率、时延以及数据分组丢失率，奖励函数考虑了视频质量、时延、码率和质量切换，输出是所选择的码率。

虽然 QARC 取得了很好的性能，但是作者认为其缺乏可解释性，因此于 2019 年进一步提出了升级版 EQARC（explainable QARC）[47]。EQARC 包括 EVQPN（explainable VQPN）和 EVQRL（explainable VQRL）2 个子网络：EVQPN 由 VQPN 中加入注意力机制形成，用来更准确地预测当前视频帧的质量；EVQRL 采用 2D-CNN 代替 VQRL 中的 1D-CNN，获得了比 QARC 算法更优的性能。

Deeplive[48]是另一个针对实时视频流的 ABR 算法，其奖励函数综合考虑了码率、卡顿时长、端到端的时延、丢帧以及质量切换等因素，输入包括过去 k 个视频帧的码率、卡顿时长、缓冲区大小和端到端时延，动作包括所选择的码率、目标缓冲区以及时延限制，输出是所选择的码率。实验结果表明，在 4 种不同的网络环境下，Deeplive 方法获得的 QoE 比 Pensieve 算法平均提高了 15%以上。

有学者将深度强化学习与可扩展视频编码（SVC, scalable video coding）相结合，来提高用户的 QoE。一般的 ABR 算法中，视频普遍以恒定码率进行编码，而 SVC 可以提供细粒度的码率切换，为 HAS 客户端提供更大的灵活性，在可变网络条件下，可以减少卡顿的发生。但是，SVC 会引入额外开销，并增加获取视频块的 HTTP 请求数量。

LAAVS[49]方法采用 SVC 和非 SVC 相结合的方式，可以在不影响视频质量的情况下显著减少视频的卡顿。实验结果表明，该方法可以取得比 Pensieve 更高的平均 QoE。

Grad-HYBJ[50]是一种基于深度强化学习的 ABR 算法，该算法将 SVC 的质量控制机制加入算法中，考虑了编码开销对 QoE 的影响。同时，提出了一种混合编码（HYBJ, jump-enabled hybrid coding）方法来减轻这方面的影响。与 Pensieve 算法相比，平均 QoE 提升了约 17%。

3.5　工业界提出的 ABR 算法

基于深度强化学习的 ABR 算法取得了远超启发式方法的性能，逐渐成为研究热点，同时也引起了工业界的重视。针对现有算法存在的仅针对单用户、未考虑运营商带宽成本等问题，工业界开展了深入的研究。

RESA[51]是爱奇艺提出的一种实时的 ABR 算法，该算法综合考虑流畅度、清晰度和平滑度（视频质量等级切换次数占总视频片段数量的比例）3 个方面的因素，加权组合建立 QoE 模型，QoE 模型的参数依据用户的观看行为偏好进行设置。真实用户的在线评估结果表明，该算法在为优质用户和普通用户提供不同 QoE 的同时，还可以控制服务提供商的带宽成本。

快手联合清华大学从训练效率和模型复杂度 2 个方面对 Comyco 算法进行了改进，提出了 Lifelong-Comyco 算法[52]。该算法由外循环系统和内循环系统两部分组成：内循环系统是 Comyco，外循环系统的目标是进一步减少训练时所需的训练集。首先对客户端估计的带宽数据进行整理，找出线下最优解。然后查看当前线上策略与线下最优解所取得的 QoE 的差距，当差距超过某个值时，将当前带宽数据放入要训练的数据集中。最后采用终身学习的方法训练神经网络，使网络可以在不忘记过去表现良好的带宽数据的情况下，记住表现不好的带宽数据。与 Pensieve 算法相比，训练效率提高了 1 700 倍，训练速度提高了 16 倍，平均 QoE 提高了 4.57%～9.93%。而且该模型的浮点计算量仅为轻量级神经网络 ShuffleNetV2[53]的 0.15%，可以成功部署在笔记本电脑等移动设备上。

Stick[54]是一种低复杂度的 ABR 算法，该算法将传统的启发式方法 BBA（buffer-based algorithm）[55]与基于学习的方法相结合，通过深度学习方法增强 BBA 的性能。与此同时，BBA 又能给基于深度学习的算法带来更多的领域知识，从而降低模型的浮点运算量（FLOPS, floating point operation per second）。Stick 算法主要包含 Stick 和 Trigger 这 2 个模块，Stick 模块利用线下训练好的神经网络，根据当前客户端接收的状态输出连续值，用于控制 BBA 算法的阈值；Trigger 模块用于决定是否开启 Stick 模块的轻量级神经网络，从而进一步降低 Stick 神经网络的整体浮点计算量。在多个数据集上的实验结果表明，该算法获得的平均 QoE 要高于 Pensieve 算法，且模型的 FLOP 降低了 88%。

3.6　小结

综上所述，近几年基于 DRL 的 ABR 算法研究工作主要集中在以下几个方面：1) 建立更精确的用户 QoE 模型来指导 ABR 算法；2) 设计合适的 DNN，优化网络训练策略，使网络可以更好地拟合深度强化学习中的价值函数和策略函数，进而学习到更好的策略；3) 应对网络状态的多变性以及用户对于高质量视频的需求；4) 针对深度强化学习方法样本效率低、难以收敛等问题设计更合适的算法。

表 2 列举了目前几种代表性的基于深度强化学习的 ABR 算法，从算法面向的业务类型、奖励函数、强化学习算法、网络轨迹数据集、性能评价指标（包括平均 QoE 提升和节约带宽）等方面进行了总结。在目前的研究工作中，对算法性能进行测试时，普遍采用的网络轨迹数据集是 HSDPA[56]和 FCC[57]，也有少量研究工作采用 4G LTE[58]和 Oboe[44]等网络数据集。测试视频则普遍采用的是 Envivio 序列。文献[38-39]则在不同类型的视频（如游戏、电影、新闻以及运动等）上对 Comyco 和 DAVS 算法进行了测试。测试结果表明，算法的性能与视频的内容密切相关。具体来说，对于电影、新闻和运动这几类场景切换较频繁的视频，Comyco 算法可以获得更高的 QoE[38]。对 DAVS 算法的测试结果表明，对纹理丰富度高且运动复杂度高的视频，应该分配更高的质量[39]。

表 2　　基于深度强化学习的代表性 ABR 算法

算法	业务类型	奖励函数	强化学习算法	网络轨迹数据集	性能评价指标	
					平均 QoE 提升	节约带宽
文献[22]	点播	$\mathrm{QoE}=\alpha\sum_{n=1}^{N}q(R_n)-\beta\sum_{n=1}^{N}T_n-\gamma\sum_{n=1}^{N-1}\lvert q(R_{n+1})-q(R_n)\rvert$，其中，$N$ 表示视频片段的数量，$q(R_n)$表示视频质量，T_n 表示卡顿时长，α、β、γ表示每一项的权重，最后一项表示平滑度	Actor -Critic	HSDPA FCC	Baseline	Baseline
文献[25]	点播	$\mathrm{QoE}_{\mathrm{hotDASH}}=\sum_{h\in H}q(R_h)+\sum_{v\in V-H}q(R_v)-\beta\sum_{n=1}^{N}T_n-\sum_{n=1}^{N-1}\lvert q(R_{n+1})-q(R_n)\rvert$，其中，$n$ 个视频片段记为 $V=\{v_1, v_2, v_3, \cdots, v_n\}$，$m$ 个用户更加偏好的片段记为 $H=\{h_1, h_2, h_3, \cdots, h_m\}$，$H\subset V$，$m\leqslant n$	Actor -Critic	HSDPA FCC	30%	—
文献[28]	点播	$Q_t=P_t+S_t+A_t$，其中，P_t、S_t、A_t 分别表示视频质量、卡顿时长、平滑度，$P_t=\sum_{i=1}^{N}\omega_t q_i$，$\omega_t$ 表示用户对第 t 个视频片段的敏感性	—	HSDPA FCC	5.7%	27.9%
文献[27]	点播	$\mathrm{QoE}=v_n\sum_{n=1}^{N}q(R_n)-\beta\sum_{n=1}^{N}T_n-\gamma\sum_{n=1}^{N-1}\lvert q(R_{n+1})-q(R_n)\rvert$，其中，$v_n$ 表示用户的情感偏好，其他变量与 Pensieve 算法一致	DP	FCC	—	—
文献[36]	点播	$\mathrm{QoE}=v_n\sum_{n=1}^{N}q(R_n)-\beta\sum_{n=1}^{N}T_n-\gamma\sum_{n=1}^{N-1}\lvert q(R_{n+1})-q(R_n)\rvert$，其中，各项权重$\alpha$、$\beta$、$\gamma$通过 SVR 获得，其他变量与 Pensieve 算法一致	double Q-learning& Dueling network	4G LTE	—	—
文献[38]	点播	$\mathrm{QoE}=\alpha\sum_{n=1}^{N}q(R_n)-\beta\sum_{n=1}^{N}T_n+\gamma\sum_{n=1}^{N-1}\lvert q(R_{n+1})-q(R_n)\rvert_+-\delta\sum_{n=1}^{N-1}\lvert q(R_{n+1})-q(R_n)\rvert_-$，其中，$q(R_n)$表示第 n 个视频片段的 VMAF 值，α、β、γ、δ表示每一项的权重，其他变量与 Pensieve 算法一致	模仿学习	HSDPA FCC Oboe	7.37%	—
文献[41]	点播	同 Pensieve 算法一致	Actor -Critic	HSDPA	8.84%	—
文献[45]	点播	$\mathrm{QoE}=\alpha\sum_{n=1}^{N}q(R_n)-\beta\sum_{n=1}^{N}T_n-\gamma\sum_{n=1}^{N-1}\lvert q(R_{n+1})-q(R_n)\rvert$，其中，$q(R_n)$表示第 n 个视频片段的 SSIM 值，其他与 Pensieve 算法一致	Actor -Critic	HSDPA FCC	43.08%	17.13%
文献[48]	直播	$\mathrm{QoE}=\alpha\sum_{k=1}^{K}Q(k)+\beta\sum_{k=1}^{K}R(k)+\gamma\sum_{k=1}^{K}L(k)-\delta\sum_{k=1}^{K}F(k)+\varphi S(i)$，其中，$Q(k)$、$R(k)$、$L(k)$、$F(k)$分别表示第 k 帧的质量、卡顿时长、延时、跳帧数，$S(i)$表示质量切换，K 表示视频帧的数量，α、β、γ、δ、φ表示权重	Double -DQN	—	15%	—
文献[50]	点播	$\mathrm{QoE}=\sum_{n=1}^{N}\log\left(\frac{R_n}{R_{\min}}\right)-\log\left(\frac{R_n}{R_{\min}}\right)\sum_{n=1}^{N}T_n-\sum_{n=1}^{N-1}\lvert\log(R_{n+1})-\log(R_n)\rvert\cdot\frac{\max(R_{n+1},R_n)}{\min(R_{n+1},R_n)}$，其中，$N$ 表示视频段的总数，R_n 表示视频片段的比特率，T_n 表示卡顿时长，$R_{\min}$ 和 $R_{\max}$ 分别表示最低和最高比特率	Actor -Critic	HSDPA FCC Belgium	17%	—
文献[51]	点播	$\mathrm{QoE}=\alpha\mathrm{Resolution\ Score}-\beta\mathrm{Fluency\ Score}-\gamma\mathrm{Smoothness\ Score}+C$，其中，Resolution Score 表示分辨率得分，Fluency Score 表示流畅度得分，Smoothness Score 表示平滑度，α、β、γ 表示权重，C 表示常数	Actor -Critic	真实环境	—	—

虽然基于深度强化学习的 ABR 算法有着远超启发式方法的性能，但是其复杂度较高，往往难以直接在客户端实现，尤其是针对计算和存储资源非常有限的移动设备。为此，Meng 等[60]提出 Pitree 框架，该框架将 ABR 转化为决策问题，具有通用性强、高性能和可扩展等优点，可以在牺牲少量性能的情况下，大大提升处理速度。在多个数据集上的实验结果表明，将 Pensieve 以及 HotDASH 等复杂度较高的算法转为决策树问题，可以有效节约计算和存储资源，进而将算法部署于视频播放器。

4　常用公共数据集和性能评价标准

对算法进行性能评价时，往往需要在同一个数据集上进行，本节首先介绍各种 HAS 视频数据集以及网络轨迹数据集，然后介绍评价标准。

4.1　常用公共数据集

数据是基于机器学习的 ABR 算法研究中最重要的工具之一。ABR 算法的常用公共数据集包括 HAS 视频数据集以及网络轨迹数据集，近年来越来越多数据集的公开为 QoE 模型和 ABR 算法的设计提供了良好的基础。

4.1.1　HAS 视频数据集

采用 HAS 传输的视频可能存在卡顿（包括初始延时）和质量切换 2 种失真。早期的数据集大部分为手工生成且包含失真类型单一，数据规模有限。近几年，数据集规模不断扩大，包含的失真类型、采用的 ABR 算法和网络环境更加丰富。

表 3 是最常用的几个公开 HAS 视频数据集，给出了各个数据集的发布时间、QoE 影响因素、原始视频数量、网络轨迹数量、失真视频数量、观看设备以及视频质量感知度量准则等信息。

得克萨斯大学公布的 LIVE-NFLX-II[65]是目前最全面的数据集之一。该数据集在 7 种网络环境下，采用 4 种 ABR 算法由 15 个原始视频生成了 420 个失真视频。数据集提供了每个视频的连续性和回顾性主观 MOS 得分。

滑铁卢大学建立的 SQoE-III数据集[66]包含 450 个视频。将 20 个不同内容类型的源视频序列编码为 11 个码率等级（235～7 000 kbit/s），并存储于服务端。客户端选择了 6 个具有代表性的 ABR 算法，在 13 个网络环境下进行仿真，采用 ITU-R 绝对类别标度进行主观测试，并进行打分。数据集中给出了 MOS、VMAF 等多种视频质量打分结果。

2020 年，SQoE-III的作者公布了 SQoE-IV[67]。SQoE-IV 是目前为止最大的公开数据集，共包含 1 350 个失真视频，5 个时长 30 s 的不同内容的原始视频经过 H.264 和 HEVC 编码器编码为 13 个码率等级，在 9 个网络环境条件下进行仿真。客户端选择了 5 个具有代表性的 ABR 算法，测试者在 3 种设备上观看打分。在此基础上，作者对现有的 11 种 QoE 模型的性能进行了评估。实验结果表明，最新的 QoE 模型与主观评分之间的相关性还有待提高，QoE 模型和 ABR 算法都有改进的余地。

4.1.2　网络轨迹数据集

近年来，研究者公布了多个网络轨迹数据集。表 4 是最常用的几个公开网络轨迹数据集，给出了各个数据集的发布时间、移动模式、持续时间、轨迹数目、带宽范围等信息。

已经公开的数据集中包括 3G[56]、4G[58]、5G[69]等多种网络通信环境下的轨迹数据，其中 FCC 数据集是美国联邦通信委员会依据美国测量宽带计划

表 3　　**代表性公开 HAS 视频数据集**

数据库名称	发布时间	QoE 影响因素	原始视频数量	网络轨迹种类	失真视频数量	观看设备	视频感知质量度量准则
LIVAMVQA[61]	2012 年	卡顿	10	—	200	Phone, Tablet	MOS, MS-SSIM, SSIM, PSNR
LIVE QHVS[62]	2014 年	质量切换	3	—	15	HDTV	MOS, MS-SSIM, SSIM, PSNR
LIVE Mobile Stall Video Database-II[63]	2014 年	卡顿	24	—	176	Apple iPhone 5	MOS
LIVE Stall Study[64]	2017 年	质量切换、卡顿	26	—	174	PC	MOS
LIVE-NFLX-II[65]	2018 年	质量切换、卡顿	15	7	420	Computer monitor	MOS, VMAF, SSIM, PSNR
Waterloo SQoE-III[66]	2018 年	质量切换、卡顿、初始缓冲时间	20	13	450	HDTV	MOS, VMAF, SSIM, PSNR
Waterloo SQoE-IV[67]	2020 年	质量切换、卡顿	5	9	1350	Phone, HDTV, UHDTV	MOS, VMAF, SSIM

表 4 代表性公开网络轨迹数据集

数据库名称	发布时间	移动模式	时间间隔/持续时间	轨迹数目	带宽范围
HSDPA[56]	2013 年	地铁、电车、火车、公共汽车、轮渡和小轿车	1 s/30 min	86	0～3 Mbit/s
FCC[57]	2016 年	—	1 s/约 3.7 天	1 000	—
Belgium[59]	2016 年	步行、自行车、公共汽车、电车、火车和小轿车	1 s/5 h	40	0~111 Mbit/s
3G&4G[68]	2016 年	车辆行驶环境	4G-10 s，3G-15 s/约 15 h，约 38 h	56 754	0~3 Mbit/s
Oboe[44]	2018 年	—	—	571	0~6 Mbit/s
4G LTE[58]	2018 年	静态、行人、汽车、公共汽车和火车	1 s/15 min	135	0~173 Mbit/s
5G[69]	2020 年	应用程序（文件下载等）和移动性模式（静态、驾驶）	1 s/约 3 142 min	83	0~1 Gbit/s

定期公布的带宽数据集。目前常用的 FCC 数据集于 2016 年公布。

4.2 评价指标

早期的 ABR 算法常用平均比特率、卡顿次数、卡顿频率、平滑性等指标来度量 ABR 算法性能的优劣，无法准确反映用户 QoE。目前的 ABR 算法普遍采用 QoE 提升程度、视频感知质量、卡顿时长、平滑性以及 QoE 相同情况下节约带宽的比例等指标来度量算法的性能。由于目前没有公认的 QoE 评价标准，各算法中所采用的 QoE 评价模型往往是研究者自行建立的。

考虑到平均比特率无法准确反映视频的主观感受质量，有部分算法采用 SSIM、VMAF 等指标来评价视频的感知质量。其中 VMAF 是 Netflix 公司提出的一种全参考视频质量客观评价指标，该指标采用视觉信息保真度、细节损失程度和运动信息等评估方法对视频质量进行度量，使用 SVR 将 3 个指标进行融合，得到最终的评估结果，这种方式使 VMAF 可以保留每种质量评估方法的优势。相比于 PSNR、SSIM 等视频质量客观评价准则，VMAF 指标更接近于用户的主观感受，可以与人类的主观评价保持一致，因此被越来越多地应用于 ABR 算法中，对视频感知质量进行度量。

图 5 和图 6 给出了 Comyco、Pitree_Comyco、Pensieve、RobustMPC、BOLA[70]和 Rate_based[71]共 6 种代表性 ABR 算法在 FCC 数据集上的测试结果。其中 Comyco、Pensieve 是基于 DRL 的 ABR 算法，Pitree_Comyco 是将 Comyco 算法通过 Pitree 框架转化为决策树的算法，RobustMPC、BOLA 和 Rate_based 是传统的启发式 ABR 算法。图 5 是 6 种算法的平均 QoE 累积分布函数（CDF, cumulative distribution function）对比结果，其中横坐标是平均 QoE，纵坐标是概率值。图 5 很好地反映了在不同区间内平均 QoE 的分布概率。图 6 进一步给出了平均 QoE、平均卡顿时长和平均比特率 3 个指标的对比实验结果。为便于对比，对每个指标进行了归一化处理，其计算方式为实际值与最大值的比值，纵坐标表示其归一化后的值。从图 6 可以看出，基于深度强化学习的 ABR 算法性能要远高于传统的启发式算法，其平均 QoE 可以提升 1.71%～4.54%，平均卡顿时长可以减少 7.04%～34.6%，平均比特率可以提升 2.91%～3.98%。

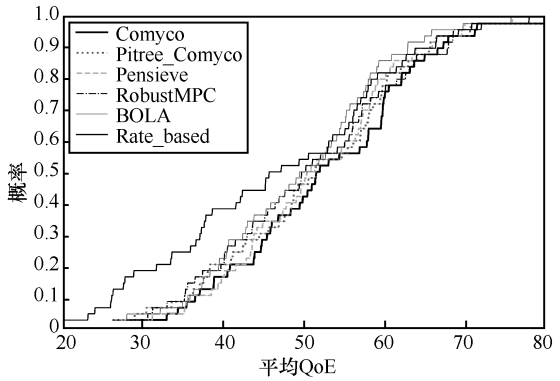

图 5 FCC 数据集上多种算法的 CDF

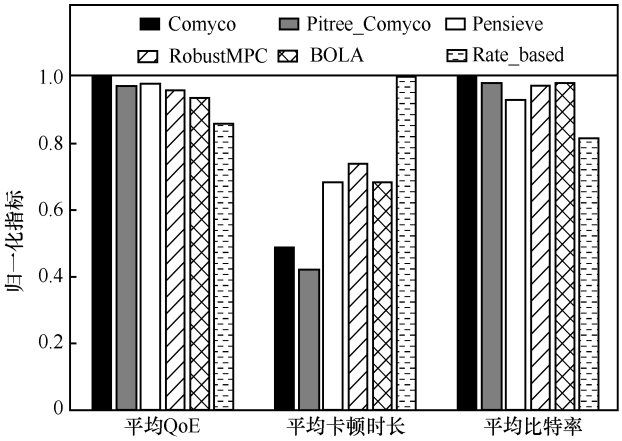

图 6 FCC 数据集上多种算法性能对比

图 7 给出了 Pensieve、RobustMPC 以及基于缓存的 Buffer_based[55]算法在 Oboe 数据集上的测试结果，其中 Buffer_based 是传统的启发式 ABR 算法。从图 7 可以看出，Pensieve 算法的性能明显优于其他 2 种启发式的算法，其平均 QoE 可以提升 1.10%～10.16%，平均卡顿时长可以减少 3.73%～7.65%，平均比特率可以提升 1.93%～6.45%。

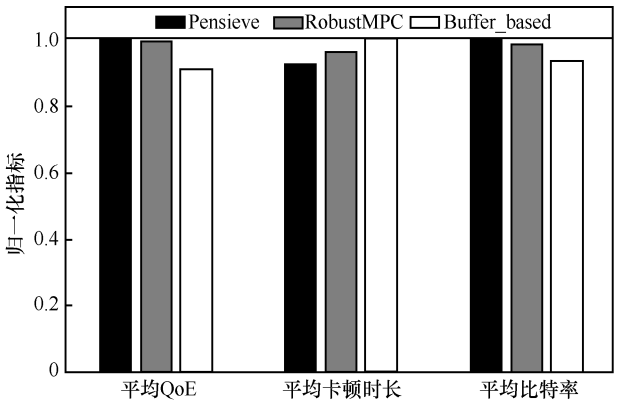

图 7　Oboe 数据集上多种算法性能对比

5　结束语

在客户端实施 ABR 算法是 HAS 保障用户 QoE 的关键所在。大量的研究结果表明，基于深度强化学习的 ABR 算法可以取得更优越的性能，是目前 ABR 的研究热点。但是，现阶段开展基于深度强化学习的 ABR 算法研究还面临以下问题。

1) 视频数据集规模。建立大规模视频主观质量评价数据集十分困难，一方面，主观实验费时费力，花费较高；另一方面，视频时长较长时，由于观测者的兴趣和精神能力有限，难以给出有效的主观评估。因此，建立大规模的主观视频质量评价数据集仍然是一个需要重点解决的问题。

2) 奖励函数的设计。基于深度强化学习的 ABR 算法采用 QoE 模型作为奖励函数，QoE 模型的准确性将直接影响算法的性能。因此如何建立更加准确的 QoE 模型是 ABR 算法需要解决的问题，还有很大的研究空间。

3) 深度神经网络的设计。智能体中的深度神经网络用来学习高维数据的有效表示，决定最优的码率控制策略。因此，针对应用场景的具体需求，利用最新的深度学习研究成果，设计适用于 ABR 算法的深度神经网络结构是一个重要的研究内容。

此外，将模仿学习、超分辨率重建、可扩展编码等方法引入 ABR 算法中，可以有效解决基于深度强化学习的 ABR 算法样本效率低、收敛缓慢等问题，也是未来比较有意义的研究方向。

参考文献：

[1] Cisco. Cisco visual networking index: Forecast and methodology[R]. 2019.

[2] XIAO A L, LIU J, LI Y Z, et al. Two-phase rate adaptation strategy for improving real-time video QoE in mobile networks[J]. China Communications, 2018, 15(10): 12-24.

[3] ITU-T. Definition of quality of experience, international telecommunication union, liaison statement, Ref: TD109rev2 (PLEN/12)[S]. 2007.

[4] STOCKHAMMER T. Dynamic adaptive streaming over HTTP: standards and design principles[C]//Proceedings of the Second Annual ACM Conference on Multimedia Systems. New York: ACM Press, 2011: 133-144.

[5] LEVKOV M. Video encoding and transcoding recommendations for HTTP dynamic streaming on the Adobe®Flash®Platform[R]. 2010.

[6] FECHEYR L A. A review of HTTP live streaming[R]. 2010.

[7] ZAMBELLI A. IIS smooth streaming technical overview[R]. 2009.

[8] KUA J, ARMITAGE G, BRANCH P. A survey of rate adaptation techniques for dynamic adaptive streaming over HTTP[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1842-1866.

[9] AYAD I, IM Y, KELLER E, et al. A practical evaluation of rate adaptation algorithms in HTTP-based adaptive streaming[J]. Computer Networks, 2018, 133: 90-103.

[10] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M]. Cambridge: MIT Press, 1998.

[11] FRANÇOIS L, HENDERSON P, ISLAM R, et al. An introduction to deep reinforcement learning[J]. Now Publishers, 2018, 11(3-4): 219-354.

[12] CLAEYS M, LATRÉ S, FAMAEY J, et al. Design of a Q-learning based client quality selection algorithm for HTTP adaptive video streaming[C]//Adaptive and Learning Agents Workshop (ALA). [S.n.:s.l.], 2013: 30-37.

[13] CLAEYS M, LATRÉ S, FAMAEY J, et al. Design and evaluation of a self-learning HTTP adaptive video streaming client[J]. IEEE Communications Letters, 2014, 18(4): 716-719.

[14] SUN Y, YIN X Q, JIANG J C, et al. CS_2P: improving video bitrate selection and adaptation with data-driven throughput prediction[C]//Proceedings of the 2016 ACM SIGCOMM Conference. Florianopolis Brazil. New York: ACM Press, 2016: 272-285.

[15] CHIARIOTTI F, D'ARONCO S, TONI L, et al. Online learning adaptation strategy for DASH clients[C]//Proceedings of the 7th International Conference on Multimedia Systems. New York: ACM Press, 2016: 1-12.

[16] ANDELIN T, CHETTY V, HARBAUGH D, et al. Quality selection for dynamic adaptive streaming over HTTP with scalable video coding[C]//Proceedings of the 3rd Multimedia Systems Conference on – MMSys'12. New York: ACM Press, 2012: 149-154.

[17] LI Z, BEGEN A C, GAHM J, et al. Streaming video over HTTP with consistent quality[C]//Proceedings of the 5th ACM Multimedia Systems Conference on – MMSys'14. New York: ACM Press, 2014:

248-258.

[18] KELLY F P, MAULLOO A K, TAN D K H. Rate control for communication networks: shadow prices, proportional fairness and stability[J]. Journal of the Operational Research Society, 1998, 49(3): 237-252.

[19] YIN X Q, 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.

[20] GARCÍA S, CABRERA J, GARCÍA N. Quality-optimization algorithm based on stochastic dynamic programming for MPEG DASH video streaming[C]//IEEE International Conference on Consumer Electronics (ICCE), Piscataway: IEEE Press, 2014: 574-575.

[21] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[J]. arXiv Preprint, arXiv:1312.5602, 2013.

[22] MAO H Z, NETRAVALI R, ALIZADEH M. Neural adaptive video streaming with Pensieve[C]//Proceedings of the Conference of the ACM Special Interest Group on Data Communication. New York: ACM Press, 2017: 197-210.

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

[24] PAUL C, HUDSON A. CS 244'18: recreating and extending pensieve[R]. 2018.

[25] SENGUPTA S, GANGULY N, CHAKRABORTY S, et al. HotDASH: hotspot aware adaptive video streaming using deep reinforcement learning[C]//2018 IEEE 26th International Conference on Network Protocols (ICNP). Piscataway: IEEE Press, 2018: 165-175.

[26] GAO G Y, DONG L S, ZHANG H Z, et al. Content-aware personalised rate adaptation for adaptive streaming via deep video analysis[C]//2019 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2019: 1-8.

[27] HU S H, XU M, ZHANG H M, et al. Affective content-aware adaptation scheme on QoE optimization of adaptive streaming over HTTP[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2020, 15(3s): 100.

[28] ZHANG X, OU Y Y, SEN S, et al. SENSEI: aligning video streaming quality with dynamic user sensitivity[C]//USENIX Symposium on Networked Systems Design and Implementation (NSDI). Berkeley: USENIX Association, 2021: 303-320.

[29] DUANMU Z F, LIU W T, CHEN D Q, et al. A knowledge-driven quality-of-experience model for adaptive streaming videos[J]. arXiv Preprint, arXiv:1911.07944, 2019.

[30] XIAO A L, HUANG X F, WU S, et al. Traffic-aware rate adaptation for improving time-varying QoE factors in mobile video streaming[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(4): 2392-2405.

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

[32] HUANG T C, ZHANG R X, SUN L F. Self-play reinforcement learning for video transmission[C]//Proceedings of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video. New York: ACM Press, 2020: 1-10.

[33] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[J]. arXiv Preprint, arXiv:1707.06347, 2017.

[34] HUO L Y, WANG Z L, XU M, et al. A meta-learning framework for learning multi-user preferences in QoE optimization of DASH[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(9): 3210-3225.

[35] FRANS K, HO J, CHEN X, et al. Meta learning shared hierarchies[C]//International Conference on Learning Representations (ICLR). [S.n.:s.l.], 2018: 1-11.

[36] LIU J, TAO X M, LU J H. QoE-oriented rate adaptation for DASH with enhanced deep Q-learning[J]. IEEE Access, 2019, 7: 8454-8469.

[37] SALEEM M, SALEEM Y, ASIF H M S, et al. Quality enhanced multimedia content delivery for mobile cloud with deep reinforcement learning[J]. Wireless Communications and Mobile Computing, 2019, 2019: 1-15.

[38] HUANG T C, ZHOU C, ZHANG R X, et al. Comyco: quality-aware adaptive video streaming via imitation learning[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 429-437.

[39] LI W H, HUANG J W, WANG S Q, et al. DAVS: dynamic-chunk quality aware adaptive video streaming using apprenticeship learning[C]//GLOBECOM 2020 - 2020 IEEE Global Communications Conference. Piscataway: IEEE Press, 2020: 1-6.

[40] 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 and Networking, 2017, 3(4): 703-718.

[41] LEKHARU A, MOULII K Y, SUR A, et al. Deep learning based prediction model for adaptive video streaming[C]//2020 International Conference on Communication Systems & Networks (COMSNETS). Piscataway: IEEE Press, 2020: 152-159.

[42] ZHANG G H, LEE J Y B. Ensemble adaptive streaming – A new paradigm to generate streaming algorithms via specializations[J]. IEEE Transactions on Mobile Computing, 2020, 19(6): 1346-1358.

[43] ZHAO Y, SHEN Q W, LI W, et al. Latency aware adaptive video streaming using ensemble deep reinforcement learning[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 2647-2651.

[44] AKHTAR Z, NAM Y S, GOVINDAN R, et al. Oboe: auto-tuning video ABR algorithms to network conditions[C]//Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication. New York: ACM Press, 2018: 44-58.

[45] YEO H, JUNG Y, KIM J, et al. Neural adaptive content-aware internet video delivery[C]//Symposium on Operating Systems Design and Implementation (OSDI). [S.n.: s.l.], 2018: 645-661.

[46] HUANG T C, ZHANG R X, ZHOU C, et al. QARC: video quality aware rate control for real-time video streaming based on deep reinforcement learning[C]//Proceedings of the 26th ACM international conference on Multimedia. New York: ACM Press, 2018: 1208-1216.

[47] HUANG T C, ZHANG R X, WU C L, et al. Generalizing rate control strategies for realtime video streaming via learning from deep learning[C]//MMAsia '19: Proceedings of the ACM Multimedia Asia. New York: ACM Press, 2019: 1-6.

[48] TIAN Z, ZHAO L P, NIE L H, et al. Deeplive: QoE optimization for live video streaming through deep reinforcement learning[C]//2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS). Piscataway: IEEE Press, 2019: 827-831.

[49] NASRABADI A T, PRAKASH R. Layer-assisted adaptive video

streaming[C]//Proceedings of the 28th ACM SIGMM Workshop on Network and Operating Systems Support for Digital Audio and Video. Amsterdam Netherlands. New York: ACM Press, 2018: 31-36.

[50] LIU Y Z, JIANG B, GUO T, et al. Grad: learning for overhead-aware adaptive video streaming with scalable video coding[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM Press, 2020: 1-9.

[51] WANG Y N, WANG H L, SHANG J Y, et al. RESA: a real-time evaluation system for ABR[C]//2019 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2019: 1846-1851.

[52] HUANG T C, ZHOU C, YAO X, et al. Quality-aware neural adaptive video streaming with lifelong imitation learning[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(10): 2324-2342.

[53] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//Computer Vision - ECCV 2018. Berlin: Springer, 2018: 116-131.

[54] HUANG T C, ZHOU C, ZHANG R X, et al. Stick: a harmonious fusion of buffer-based and learning-based approach for adaptive streaming[C]//IEEE INFOCOM 2020 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2020: 1967-1976.

[55] HUANG T Y, JOHARI R, MCKEOWN N, et al. A buffer-based approach to rate adaptation: Evidence from a large video streaming service[C]//Special Interest Group on Data Communication (SIGCOMM). New York: ACM Press, 2014: 187-198.

[56] RIISER H, VIGMOSTAD P, GRIWODZ C, et al. Commute path bandwidth traces from 3G networks: analysis and applications[C]//Proceedings of the 4th ACM Multimedia Systems Conference on - MMSys '13. New York: ACM Press, 2013: 114-118.

[57] US Federal Communications Commission. Measuring Fixed Broadband Report[R]. 2016.

[58] RACA D, QUINLAN J J, ZAHRAN A H, et al. Beyond throughput: a 4G LTE dataset with channel and context metrics[C]//Proceedings of the 9th ACM Multimedia Systems Conference. New York: ACM Press, 2018: 460-465.

[59] HOOFT J V D, PETRANGELI S, WAUTERS T, et al. HTTP/2-based adaptive streaming of HEVC video over 4G/LTE networks[J]. IEEE Communications Letters, 2016, 20(11): 2177-2180.

[60] MENG Z L, CHEN J, GUO Y N, et al. PiTree: practical implementation of ABR algorithms using decision trees[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 2431-2439.

[61] MOORTHY A K, CHOI L K, BOVIK A C, et al. Video quality assessment on mobile devices: subjective, behavioral and objective studies[J]. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(6): 652-671.

[62] CHEN C, CHOI L K, VECIANA G D, et al. Modeling the time—varying subjective quality of HTTP video streams with rate adaptations[J]. IEEE Transactions on Image Processing, 2014, 23(5): 2206-2221.

[63] GHADIYARAM D, BOVIK A C, YEGANEH H, et al. Study of the effects of stalling events on the quality of experience of mobile streaming videos[C]//2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Piscataway: IEEE Press, 2014: 989-993.

[64] GHADIYARAM D, PAN J, BOVIK A C. A subjective and objective study of stalling events in mobile streaming videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(1): 183-197.

[65] BAMPIS C G, LI Z, KATSAVOUNIDIS I, et al. Towards perceptually optimized adaptive video streaming-a realistic quality of experience database[J]. IEEE Transactions on Image Processing, 2021, 30: 5182-5197.

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

[67] DUANMU Z F, CHEN D, LI Z W, et al. Assessing the quality-of-experience of adaptive bitrate video streaming[J]. arXiv Preprint, arXiv: 2008.08804, 2020.

[68] BOKANI A, HASSAN M, KANHERE S S, et al. Comprehensive mobile bandwidth traces from vehicular networks[C]//Proceedings of the 7th International Conference on Multimedia Systems. New York: ACM Press, 2016: 344-348.

[69] RACA D, LEAHY D, SREENAN C J, et al. Beyond throughput, the next generation: a 5G dataset with channel and context metrics[C]//Proceedings of the 11th ACM Multimedia Systems Conference. New York: ACM Press, 2020: 303-308.

[70] SPITERI K, URGAONKAR R, SITARAMAN R K. BOLA: near-optimal bitrate adaptation for online videos[J]. IEEE/ACM Transactions on Networking, 2020, 28(4): 1698-1711.

[71] JIANG J C, 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.

[作者简介]

杜丽娜（1995− ），女，山西忻州人，北京工业大学博士生，主要研究方向为视频质量评价、码率自适应算法。

卓力（1971− ），女，江苏徐州人，博士，北京工业大学教授、博士生导师，主要研究方向为图像/视频的编码与传输、多媒体大数据处理等。

杨硕（1993− ），男，河南商丘人，北京工业大学硕士生，主要研究方向为视频质量评价。

李嘉锋（1986− ），男，天津人，博士，北京工业大学讲师、硕士生导师，主要研究方向为计算机视觉、图像增强。

张菁（1975− ），女，广东梅县人，博士，北京工业大学教授、博士生导师，主要研究方向为图像/视频处理、图像识别和图像检索。