文章编号： 1671-7449(2024)02-0194-09

无人机辅助智能交通系统中面向视频多播的资源优化

薛　斌，张志才*，付　芳

（山西大学 物理电子工程学院，山西 太原 030006）

摘　要： 在智能交通系统(Intelligent Transportation System，ITS)中，实时交通视频可以辅助网联车做出更理智的决策。然而，由于车载传感器位置和数量的限制，车辆无法全面掌握交通环境，不利于行车安全；此外，当多个车辆用户请求相同的视频内容时，传统的单播传输模式存在效率低下的问题。为了解决这些问题，基于非正交多址接入(Non-Orthogonal Multiple Access，NOMA)和可伸缩视频编码(Scalable Video Coding，SVC)技术，提出了一种无人机(Unmanned Aerial Vehicle，UAV)辅助ITS的交通视频多播方案。通过联合优化车辆分组和功率分配策略，最大化车辆接收的长期视频质量。将该优化问题建模为一个马尔可夫决策过程(Markov Decision Process，MDP)，并采用Soft Actor-Critic(SAC)算法来求解。大量仿真结果表明，该算法具有很强的探索能力，而且收益性能优于传统Actor-Critic(AC)算法。

关键词： 交通视频多播；可伸缩视频编码；非正交多址接入；Soft Actor-Critic

中图分类号： TP393　　**文献标识码：** A　　**doi：** 10.3969/j.issn.1671-7449.2024.02.013

引用格式： 薛斌，张志才，付芳. 无人机辅助智能交通系统中面向视频多播的资源优化[J]. 测试技术学报，2024，38(2)：194-202.

XUE Bin，ZHANG Zhicai，FU Fang. Resource optimization for video multicast in unmanned aerial vehicle assisted intelligent transportation system[J]. Journal of Test and Measurement Technology，2024，38(2)：194-202.

Resource Optimization for Video Multicast in Unmanned Aerial Vehicle Assisted Intelligent Transportation System

XUE Bin，ZHANG Zhicai*，FU Fang

（School of Physics and Electronic Engineering，Shanxi University，Taiyuan 030006，China）

Abstract： In Intelligent Transportation System(ITS)，real-time traffic video can be used to assist Connected Autonomous Vehicles(CAVs) to make more sensible decisions. However，due to the limitation of locations and qualities of the onboard sensors，CAVs cannot observe the full traffic situations. In addition，when vehicle users subscribe to the same video content simultaneously，traditional unicast transmission mode is spectrum inefficient，especially for the traffic congestion scenario. In order to solve these problems，in this paper，we propose a traffic video multicasting scheme based on Non-Orthogonal Multiple Access (NOMA) and Scalable Video Coding(SVC) for Unmanned Aerial Vehicle(UAV) enabled

收稿日期： 2023-01-21

基金项目： 山西省基础研究计划自然科学研究面上资助项目(202103021224024)；山西省基础研究计划青年科学研究资助项目(20210302123021)；山西省重点研发计划资助项目(202202020101004)

作者简介： 薛　斌(1998—)，男，硕士生，主要从事5G资源分配、强化学习等研究。E-mail：1339517493@qq.com。

***通信作者：** 张志才(1982—)，男，讲师，硕士生导师，主要从事移动边缘计算、车联网和机器学习。E-mail：zzcai@sxu.edu.cn。

ITS. We aim to maximize the long-term video quality received by vehicles by jointly optimizing the vehicle grouping and power allocation strategies. We model the problem as a Markov Decision Process (MDP) and leverage Soft Actor-Critic (SAC) algorithm to solve the MDP. Extensive simulation results indicate that the proposed algorithm has a strong exploration ability, and outperforms the traditional Actor-Critic (AC) algorithm in terms of the reward performance.

Key words: traffic video multicasting; scalable video coding; non-orthogonal multiple access; soft actor-critic

0 引 言

实时、准确和全面的交通信息对于网联车辆（Connected Autonomous Vehicles, CAVs）的驾驶安全问题至关重要[1]。作为未来智能交通系统（Intelligent Transport Systems, ITS）的关键组成部分，CAVs采用传统方法如摄像头、激光雷达、GPS等收集交通数据[2,3]。尽管这些方法可以及时感知有用的数据，但受到传感设备的位置、数量和质量的限制，CAVs无法完全观察到交通状况，这极大地增加了CAVs行驶的危险性。而无人机上的高清摄像头捕捉到的交通全局视角可以传递给CAVs，以帮助其做出更明智的决策[4]。此外当同一区域内的车辆请求相同的视频内容时，传统的单播传输模式存在效率低下的问题。多播作为一种有效的解决方案[5]，可以利用有限的频谱资源为更多车辆提供数据/视频应用。然而，为保证在多播组中车辆都能正确接收/解码视频内容，信道质量最差的车辆（例如，区域边缘车辆）会导致多播速率的瓶颈[6]。

可伸缩视频编码（Scalable Video Coding, SVC）集成到无线视频多播中，可以减少信道质量最差的车辆带来的瓶颈效应[7]。使用SVC，将视频流编码为一个基础层和多个增强层，接收端必须准确解码基础层信息才可以重建视频[8]。然而，现有的SVC多播方案多采用正交频分多址接入（Orthogonal Frequency Division Multiple Access, OFDMA）技术，需占用大量频谱资源，不适用交通繁忙的密集场景[9,10]。

非正交多址接入（Non-Orthogonal Multiple Access, NOMA）通过功率复用可以在同一信道上为多个用户提供服务，进而提高频谱利用率[11]。DING等[12]验证了由于用户信道不对称性，NOMA比OMA显著提高频谱效率并缓解传输延迟。DING等[13]提出当用户经历不同的信道质量时，NOMA实现的小区边缘吞吐量和频谱效率优于OMA。CHINGOSKA等[14]提出用于无线供电通信网络上行链路的NOMA，并与OMA相比扩大了上行链路的可实现速率区域。

同时学术界和工业界对多播业务资源分配的方法也展开广泛研究。TAN等[15]提出了一种DASH（Dynamic Adaptive Streaming over HTTP）多播方案，通过分组算法和自适应码率算法以提高用户体验质量。UL ZUHRA等[16]针对LTE（Long Term Evolution）组播传输的分组和资源分配问题，设计了一个模拟退火算法。然而，以上算法主要是针对静态结构问题，对于动态结构问题效果不佳。深度强化学习通过智能体与环境不断交互学习到最优策略，在求解动态结构问题时速度更快，效率更高[17,18]。

因此，本文提出了一种UAV辅助ITS中基于SAC（Soft Actor-Critic）算法的NOMA-SVC多播业务的资源优化研究方案。通过联合优化车辆分组和功率分配策略，最大限度地提高了车辆接收的长期视频质量。SAC是基于熵最大化的算法，具有比传统AC（Actor-Critic）算法更强的探索能力。

1 系统架构

图1考虑了UAV辅助ITS场景。其中配备高清摄像头和通信模块的UAV在道路正上方悬停，为覆盖范围内车辆提供实时交通多播服务。路边单元（Road Side Unit, RSU）沿道路部署，可为UAV提供计算服务。

1.1 SVC模型

UAV拍摄的交通视频由SVC编码器编码，其中视频流由一系列视频序列组成，每个序列的持续时间为Δ_t s，让$\mathcal{T}=\{1,\cdots,t,\cdots,T\}$表示时隙集。在SVC标准中，将视频序列编码成一组图片（Group of Pictures, GOP），每个GOP包含I层：

一个基本层(BL)和($I-1$)个增强层(EL)。BL层能够自我解码，并且可以将多个EL数据附加到BL层上以提高视频质量，即具有更多EL的视频意味着更高的视频质量。相应地，规定SVC视频流的视频质量划分为I级。

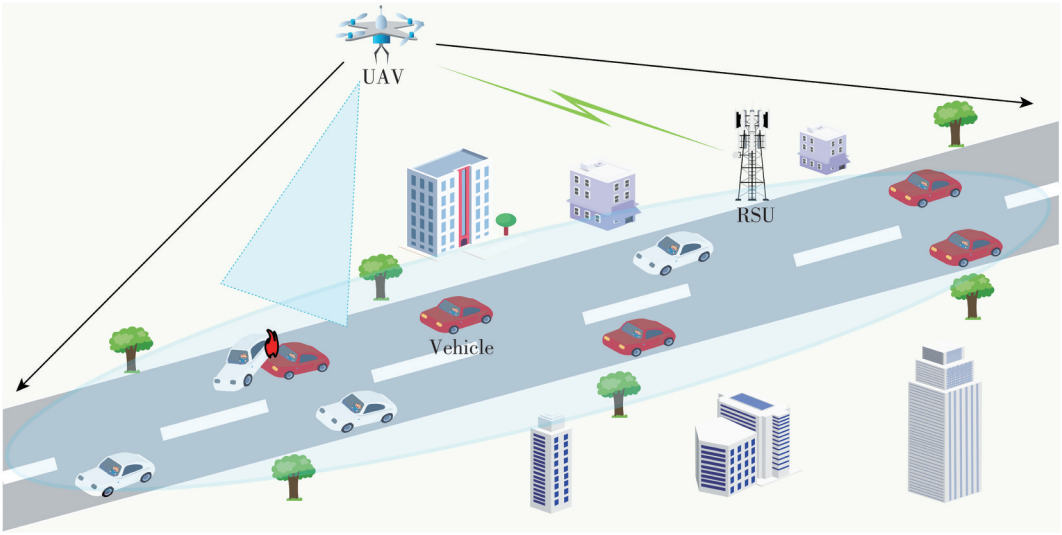

图 1 系统模型

Fig. 1 System model

1.2 用户分组模型

假设$V=\{1,2,\cdots,v\}$车辆在时隙t请求交通流视频服务，$q_v^t=(x_v^t,y_v^t,0)$表示车辆v在时隙t的道路坐标，$q_u=(x_u,y_u,H)$表示UAV的悬停坐标。因此，可以计算UAV和车辆v之间的距离

$$d_{v,u}^t=\sqrt{(x_v^t-x_u)^2+(y_v^t-y_u)^2+H^2},\forall v\in V。\tag{1}$$

UAV-to-Vehicle(U2V)无线信道遵循自由空间路径损耗数学模型，因此，在时隙t信道增益表示为

$$h_{v,u}^t=\zeta_0\left(d_v^t(q_u)\right)^{-2},\forall v\in V,\tag{2}$$

式中：ζ_0是参考距离1 m处的信道增益。设$\mathcal{G}=\{G_i|i\in\mathcal{I}\}$表示分组的集合，$\mathcal{I}=\{1,2,\cdots,I\}$。并且$G_i$组中的车辆可以接收和解码相同质量的视频。$C=\{c_{i,v}^t\in\{0,1\}|1\leqslant v\leqslant V,1\leqslant i\leqslant I\}$表示在时隙$t$车辆关联策略，其中$c_{i,v}^t=1$表示车辆$v$分配到组$G_i$，否则$c_{i,v}^t=0$。由于车辆可能会接收多层数据，因此，车辆将相应地分到多个组。图2是车辆分组图的示例，其中具有高信道质量的车辆v_6，v_7与G_1，G_2和G_3相关联，可以分别在同一时间段接收/解码BL，EL1和EL2数据，并且这些不同的层组合在一起，以生成车辆的最终视频流。另一方面，信道质量较恶劣的车辆v_2，v_3与G_1进行唯一关联，车辆仅接收/解码BL数据。而信道质量极其恶劣的车辆v_1需要抛弃，以保证其他车辆接收视频的质量。因此$c_{i,v}^t$应满足以下条件

$$c_{i,v}^t\in\{0,1\},\forall i\in\mathcal{I},v\in V,\tag{3}$$

$$0\leqslant\sum_{i=1}^{I}c_{i,v}^t\leqslant I,\forall i\in\mathcal{I},v\in V。\tag{4}$$

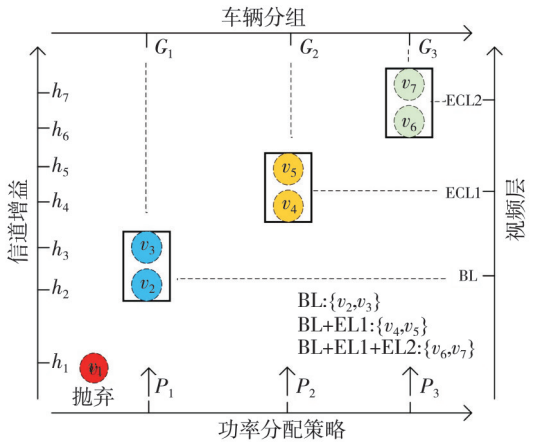

图 2 用户分组模型

Fig. 2 User grouping model

1.3 NOMA-SVC发送模型

设$\mathcal{N}=\{N_i|i\in\mathcal{I}\}$表示NOMA层集合，$\mathcal{I}=\{1,2,\cdots,I\}$。图3是NOMA-SVC发送器框图示例。规定BL通过第一NOMA层即N_1中的信号X_1传输；同时EL1通过N_2中的信号X_2传输；EL2通过N_3中的信号X_3传输，以此类推，且这些信号是同时传输的，X_i的传输功率在时隙t将分配

$P_i^t=P_u\cdot\beta_i^t$，其中 $0\leqslant\beta_i^t\leqslant1$ 表示功率比，P_u 为UAV传输功率。满足以下条件

$$\beta_1^t>\beta_2^t>\beta_3^t,\ \sum_{i=1}^{I}\beta_i^t=1。\tag{5}$$

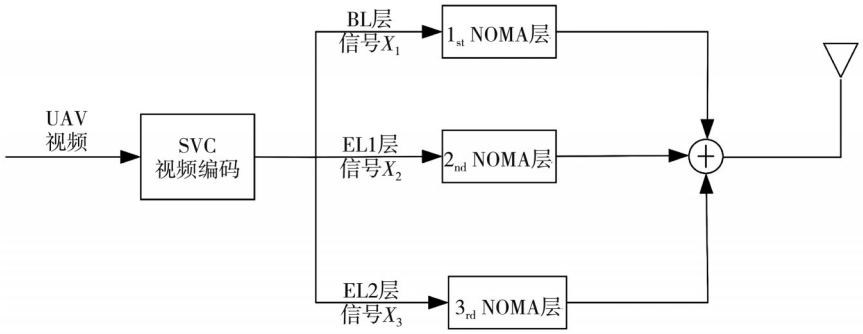

图 3　NOMA-SVC发送器框图

Fig. 3　The block diagram of NOMA-SVC transmitter

值得注意的是，组 G_i 与第 i NOMA层即 N_i 匹配，这意味着组 G_i 中车辆的最差信道质量将用作 N_i 的信道增益，即 $g_i^t=\min\{h_{v,u}^t,v\in G_i\}$。根据NOMA原理，在时隙 t 由UAV多播的叠加传输信号表示为

$$X^t=\sum_{i=1}^{I}X_i\sqrt{P_u\beta_i^t}。\tag{6}$$

1.4　NOMA-SVC接收模型

组 G_i 中车辆在时隙 t 的接收信号可以表示为

$$Y_i^t=g_i^t\cdot\left(\sum_{i=1}^{I}X_i\sqrt{P_u\beta_i^t}\right)+\varpi,\tag{7}$$

式中：ϖ 是信道自然噪声。在接收器处，进行连续干扰消除。与具有较恶劣信道质量的组（例如 G_1）相比，具有高信道质量的组（例如 G_3）可以接收/解码更多视频层，并实现高视频质量。由于同一组中的车辆需要解码相同的信息，因此，只存在组间干扰，而不考虑组内干扰。

图 4 是NOMA-SVC接收器框图示例。连续干扰消除解码过程可以描述为 G_1 中车辆解码高功率信号 X_1 时，将信号 X_2 和 X_3 视为噪声/干扰。G_2 中车辆在解码信号 X_2 之前，首先用连续干扰消除解码信号 X_1，并从接收到的叠加信号中删除 X_1。G_3 中车辆在解码信号 X_3 之前删除信号 X_1 和 X_2。

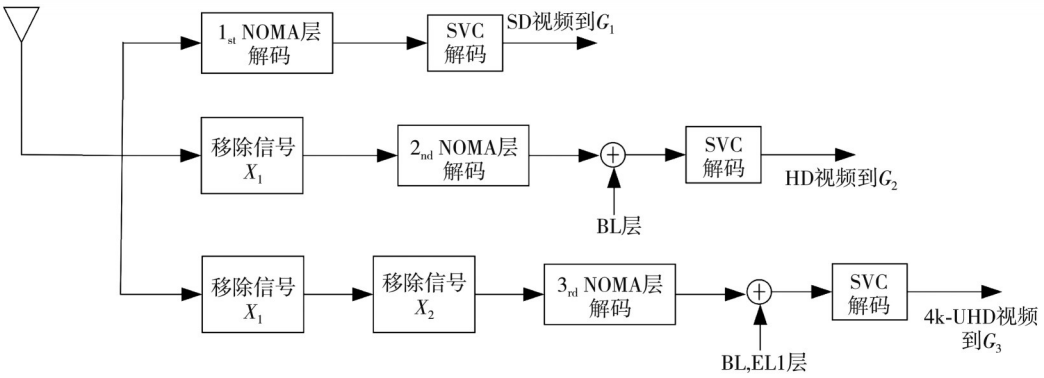

图 4　NOMA-SVC接收器框图

Fig. 4　The block diagram of NOMA-SVC receiver

在不丧失通用性的情况下，$G_j(j>i)$ 可以连续地解码信号 X_i，并将其从接收信号中移除。连续干扰消除解码过程按从基本层到增强层的升序排列。通常，接近UAV的车辆具有更好的信干噪比（$SINR$）。在这种情况下，他们可以接收/解码更多的视频层并获得更高的视频质量。

G_i 在时隙 t 的 $SINR$ 由组内最差信道质量的车辆决定

$$SINR_i^t=\begin{cases}\dfrac{\beta_i^t}{\sum\limits_{i=i+1}^{I}\beta_i^t+\dfrac{\sigma^2}{P_u g_i^t}} & (i<I),\\ \dfrac{\beta_i^t P_u g_i^t}{\sigma^2} & (i=I)。\end{cases}\tag{8}$$

根据 $SINR_i^t<SINR_j^t(\forall j>i)$，由于 $g_i^t<g_j^t$[19]，使用香农容量来计算时隙 t 解码信号 X_i 时 G_i 实现的最小下载速率

$$R_{i,i}^{t}\left(\beta_i^t, c_{i,v}^t\right)=B\log_2\left(1+SINR_i^t\right) \tag{9}$$

式中：B是所有组的可用带宽。定义了指标e_i^t，从物理层的角度显示时隙t时G_i中车辆接收视频层的有效性

$$e_i^t\left(\beta_i^t, c_{i,v}^t\right)=\begin{cases}1, & \text{if } R_{i,i}^{t}\left(\beta_i^t, c_{i,v}^t\right)\geqslant R_{\min}^{i},\\ 0, & \text{otherwise}。\end{cases} \tag{10}$$

当$R_{i,i}^t$大于用$R_{\min}^i$表示的GOP视频层的总和速率时，$e_i^t=1$，这意味着G_i内车辆可以正确接收/解码信号X_i，否则$e_i^t=0$。注意$R_{j,j}^t>R_{i,i}^t\geqslant R_{\min}^i$有$SINR_i^t<SINR_j^t(\forall j>i)$，这意味着$G_j(j>i)$可以正确解码和去除信号$X_i$。

SVC视频解码过程需要连续的视频层，即当第I层视频可以正确解码仅是因为基本层到I-1层已成功接收/解码。G_i在时隙t已正确解码的视频层的数量可以表示为

$$\mathcal{L}_i^t=\sum_{i=1}^{i}\prod_{i=1}^{i}e_i^t。 \tag{11}$$

1.5 QoE模型和效用函数

在本文中，QoE衡量车辆用户对视频质量的满意度。由于同一组中的车辆具有类似的信道质量，他们可以实现相同的视频质量，并具有类似的QoE要求。用户体验遵循对数定律，QoE函数可以以对数形式建模，用于网页浏览和视频下载应用。因此，在每个时隙t，G_i中车辆QoE可以定义为$\mathcal{L}_i^t$的对数函数

$$\mathcal{Q}_i^t(\mathcal{L}_i^t)=\log_2\left(1+\frac{\mathcal{L}_i^t}{\mathcal{L}}\right)\cdot\prod_{i=1}^{i}e_i^t。 \tag{12}$$

如果G_i对所有可用视频层进行解码，则其最大值为1，即$\mathcal{L}_i^t=\mathcal{L}$。用$\Omega^t$表示第$t$个时隙所有组内车辆的总体效用

$$\Omega^t(\beta_i^t, c_{i,v}^t)=\sum_{i=1}^{I}\sum_{v=1}^{V}c_{i,v}^t\omega_i\mathcal{Q}_i^t \tag{13}$$

式中：ω_i是第i个NOMA层即N_i数据的价格系数，可以在视频层选择方面平衡公平性和效率。

1.6 问题表述

通过联合优化车辆关联策略$c_{i,v}^t$和功率分配策略β_i^t，使所有车辆接收的长期视频质量最大化。本文的问题表述为

$$\max_{\beta_i^t, c_{i,v}^t}:\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{v=1}^{V}c_{i,v}^t\omega_i\mathcal{Q}_i^t, \tag{14}$$

$$\text{subject to. } c_{i,v}^t\in\{0,1\},\forall i\in\mathcal{I},v\in V, \tag{14(a)}$$

$$0\leqslant\sum_{i=1}^{I}c_{i,v}^t\leqslant I,\forall i\in\mathcal{I},v\in V, \tag{14(b)}$$

$$\beta_1^t>\beta_2^t>\beta_3^t,\ \sum_{i=1}^{I}\beta_i^t=1, \tag{14(c)}$$

式中：$\omega_i(\omega_i\geqslant 0)$(\$/vehicle)是第i个NOMA层即N_i数据的价格系数。式(14(a))表示时隙t时车辆v是否分配到G_i，式(14(b))表示时隙t时车辆v允许分配到多个组中但最大不超过I。式(14(c))表示时隙t时功率分配策略。

问题式(14)是一个长期优化问题，需要随着时间的推移依次做出一系列决策。传统的方法，如静态优化和博弈论无力处理该问题，它们试图根据当前状态搜索最优/次优解，以最大限度地提高即时回报[20]。基于上述观察，将问题式(14)建模为马尔可夫决策过程(MDP)。将MDP用一个元组表示：$\langle S,A,P,r\rangle$，每个元素具体描述如下：

S是环境的状态空间。包括：第t个时隙开始时的车辆坐标$q_v^t=(x_v^t,y_v^t,0)$。

A是环境的动作空间。包括：车辆关联策略$c_{i,v}^t$和功率分配策略β_i^t。

P是状态转移概率函数。具体而言，车辆在下一个时隙的坐标由车辆的当前位置、速度和加速度决定。

r是即时奖励函数。定义为：$r=\sum_{i=1}^{I}\sum_{v=1}^{V}c_{i,v}^t\omega_i\mathcal{Q}_i^t$。

之后，采用深度强化学习算法，即Soft Actor-Critic(SAC)来求解上述MDP问题。

2 基于Soft Actor-Critic的算法设计

2.1 Soft值函数

与现有的强化学习算法的不同之处在于，算法中加入了熵项$H(\pi(a|s))=-\log\pi(a|s)$，增加了策略的探索性，从而加快了学习速度，并且避免了策略过早收敛到局部最优值。其中策略$\pi(a|s)$称为期望熵，表示为

$$J(\pi)=E\left\{\sum_{t=0}^{\infty}\gamma^t[r_t-\lambda\log\pi(a_t|s_t)]\,|\,\pi\right\}, \tag{15}$$

式中：γ为折扣因子；λ是温度参数。熵期望表示为

$$Q^{\pi}(s,a)=E\left\{\sum_{t=0}^{\infty}\gamma^t[r_t-\lambda\log\pi(a_t|s_t)]\,|\,s_0=s,a_0=a,\pi\right\}。 \tag{16}$$

定理 1　根据 Q 值函数可以求出状态值函数

$$V^{\pi}(s)=\lambda\log\int_{A}\exp\left(\frac{1}{\lambda}Q^{\pi}(s,a)\right)da。\tag{17}$$

对应的最优策略为

$$\pi'(\cdot|s)=\exp\left(\frac{1}{\lambda}Q^{\pi}(s,a)-V^{\pi}(s)\right),\tag{18}$$

式中：Q 值函数和状态值函数在 SAC 算法中分别对应的是 soft Q 值函数和 soft 状态值函数。

2.2　Critic 部分

Q 值函数 Q 和状态值函数 V 由深度神经网络（Deep Neural Networks，DNN）来拟合。通过采用一组权重为 $\theta=\{\theta_1,\theta_2,\cdots\theta_N\}$ 的 DNN 来使得 Soft Q 值参数化，即 $Q(s,a)\approx Q_{\theta}(s,a)$。DNN 通过从经验池中随机采样 (s,a,r,s')，然后更新参数 θ。软 Q 值函数由均方误差（Mean Square Error，MSE）表示为

$$L(\theta)=E\left[\frac{1}{2}(Q_{\theta}(s,a)-\hat{Q}(s,a))^2\right],\tag{19}$$

式中，$\hat{Q}(s,a)$ 满足 $\hat{Q}(s,a)=r+\gamma\hat{V}_{\hat{\theta}}(s')$，而 $\hat{V}_{\hat{\theta}}(s')$ 是目标状态值。参数 θ 表示为

$$\theta(t+1)\leftarrow\theta(t)-\alpha_{c,t}\nabla_{\theta}L(\theta),\tag{20}$$

式中：$\alpha_{c,t}>0$ 是 critic 部分的学习率；$\nabla_{\theta}L(\theta)$ 是 $L(\theta)$ 的梯度，表示为

$$\nabla_{\theta}L(\theta)=\nabla_{\theta}Q_{\theta}(s,a)[Q_{\theta}(s,a)-r-\gamma\hat{V}_{\hat{\theta}}(s')]。\tag{21}$$

SAC 算法采用一组权重为 $\vartheta=\{\vartheta_1,\vartheta_2,\cdots,\vartheta_N\}$ 的 DNN 来近似状态值函数，即 $V(s)=V_{\vartheta}(s)$。软 V 值函数由均方误差 MSE 表示为

$$L(\vartheta)=E\left[\frac{1}{2}V_{\vartheta}(s)-E[Q_{\theta}(s,a)-\lambda\log\pi_{\varphi}(a|s)]\right]^2。\tag{22}$$

对应的 $L(\vartheta)$ 梯度和参数 ϑ 的更新为

$$\nabla_{\vartheta}L(\vartheta)=\nabla_{\vartheta}V_{\vartheta}(s)[V_{\vartheta}(s)-[Q_{\theta}(s,a)-\lambda\log\pi_{\varphi}(a|s)]],\tag{23}$$

$$\vartheta(t+1)\leftarrow\vartheta(t)-\alpha_{c,t}\nabla_{\vartheta}L(\vartheta)。\tag{24}$$

相应的目标状态值函数 $\hat{V}_{\hat{\vartheta}}(s')$ 的参数是 $\vartheta=\{\vartheta_1,\vartheta_2,\cdots,\vartheta_N\}$。参数 $\hat{\vartheta}$ 的更新为

$$\hat{\vartheta}(t+1)\leftarrow k\vartheta(t)+(1-k)\hat{\vartheta}(t),\tag{25}$$

式中：k 是平滑因子，满足 $0<k<1$。

2.3　Actor 部分

Actor 网络采用参数为 φ 的 DNN 深度神经网络来表示参数化策略 $\pi_{\varphi}(\cdot|s)$。不同于其他算法的是，SAC 算法是通过 Actor 网络来输出含有平均值和标准差的高斯分布，使用 KL 散度来最小化下面的期望值，从而获得最优的策略

$$L(\varphi)=E(D_{KL}\pi_{\varphi}(\cdot|s)\,\|\,\pi'(\cdot|s))。\tag{26}$$

本文通过随机神经网络变换来重新参数化策略，即 $a=g_{\varphi}(\tau',s)$，其中 τ 是从标准正态分布采样的动作噪声向量. 则

$$L(\varphi)=E[\lambda\log\pi_{\varphi}(g_{\varphi}(\tau;s)|s)-Q^{\pi}(s,g_{\varphi}(\tau;s))+V^{\pi}(s)]。\tag{27}$$

梯度 φ 表示为

$$\nabla_{\varphi}L(\varphi)=\nabla_{\varphi}\lambda\log\pi_{\varphi}(a|s)+\nabla_{a}\lambda\log\pi_{\varphi}(a|s)-\nabla_{a}Q_{\theta}(s,a))\nabla_{\varphi}g_{\varphi}(\tau;s)。\tag{28}$$

在梯度下降的方向上更新策略参数 φ，表示为

$$\varphi(t+1)\leftarrow\varphi(t)-\alpha_{a,t}\nabla_{\pi}L(\varphi),\tag{29}$$

式中：$\alpha_{a,t}>0$，表示 actor 的学习率。

定理 2　SAC 算法通过交替进行策略评估（Critic）和策略改进（Actor），从而达到收敛。即对于任意的 $\forall\pi,\pi^{*}\in\Pi(\pi^{*}\neq\pi)$，和 $\forall(s,a)\in S\times A$，满足 $Q^{\pi^{*}}(s,a)>Q^{\pi}(s,a)$，$\forall\pi\in\Pi$ 收敛到策略 $\pi^{*}(\cdot|s)$。

2.4　Soft actor-critic 算法

图 5 为 soft actor-critic 算法的体系结构，其中包含了 3 个部分：actor 网络、critic 网络和记忆存储器（replay memory）。actor 网络将状态映射到动作，critic 网络则负责估计状态和状态-动作对的值，而记忆存储器则用于存储经验。actor 使用参数化的深度神经网络近似策略 $\pi_{\varphi}(\cdot|s)$，并根据当前环境的状态 s 选择并执行动作 a。当接收到奖励 r 和观察到下一个状态 s' 之后，将元组 (s,a,r,s') 存储在记忆存储器中。为了避免不同样本之间的相关性，critic 会从记忆存储器中随机采样经验，并分别用来训练状态值网络和 Q 值网络。通过随机梯度下降法，critic 产生的损失函数 $L(\theta)$ 和 $L(\varphi)$ 反向传播，用来更新深度神经网络的参数，同时用 $V_{\vartheta}(s)$ 和 $Q_{\theta}(s,a)$ 更新 actor 的参数。学习过程交替进行在以下两方面：一是使用当前策略从环境中收集经验，二是根据从记忆存储器中采样的批次更新函数逼近器的参数，直到时间结束或达到最终状态[21]。该算法的学习率满足

$$\sum_{t=0}^{\infty}\alpha_{c,t}=\infty,\ \sum_{t=0}^{\infty}\alpha_{c,t}^{2}<\infty,\tag{30}$$

$$\sum_{t=0}^{\infty}\alpha_{a,t}=\infty,\ \sum_{t=0}^{\infty}\alpha_{a,t}^{2}<\infty,\ \lim_{t\to\infty}\frac{\alpha_{a,t}}{\alpha_{c,t}}=0。\tag{31}$$

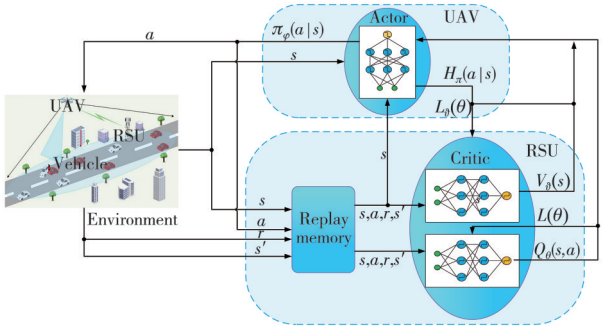

图 5 Soft actor-critic 框架

Fig. 5 The framework of soft actor-critic

3 仿真结果分析

本节给出了所提算法的仿真结果。仿真在基于 Python 的模拟器上实现，其中软件环境为 TensorFlow1. 15. 0，Python3. 6. 5。硬件环境为基于 CPU 的服务器，具有 8 GB 3 200 MHz DDR4，3. 0 GHz AMD Core R5 和 512 G 内存。

仿真中在 200 m 的城市道路上随机生成车辆，并生成车辆轨迹。无人机部署在道路正上方，用于捕获交通视频并将其多播给车辆，悬停高度 H 从 50～125 m 不等，无人机的传输功率 P_u 设置为 0. 01～0. 1 W[22]，带宽 B=0. 9 MHZ。每个时隙的长度 Δ_t=1 s。视频层的数量 I=3。道路中央部署了一个 RSU，用于协助无人机训练策略。载波频率为 5. 9 GHz，总带 20 MHZ[23]。学习参数设置为折扣因子 γ=0. 99，温度参数初始值 λ=0. 05，经验池大小 200 000，一次训练所选取样本数为 256。学习率讨论如下：

图 6 显示了所提出算法不同 Actor 的学习率收敛性能。

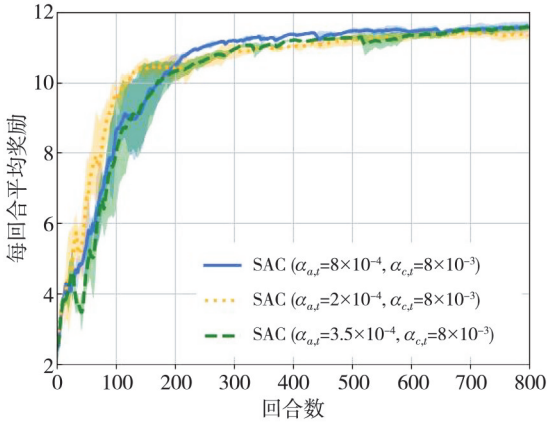

图 6 不同 Actor 学习率

Fig. 6 Different actor learning rates

其满足式(30)、式(31)并且通过反复试验来设置。在仿真中，Critic 的学习率固定为 $\alpha_{c,t}=8\times10^{-3}$。在图中较深的线表示平均值，阴影区域表示平均值±标准误差，反映了曲线的方差。从图 6 中可以观察到，当 $\alpha_{a,t}=2\times10^{-3}$ 时所提算法有高方差和低奖励，并且大约在 300 回合(每回合包含 200 步)后达到收敛。然而，如果学习率下降到 $\alpha_{a,t}=3.5\times10^{-4}$ 时学习速度变慢。相比于 $\alpha_{a,t}=2\times10^{-3}$ 和 $\alpha_{a,t}=3.5\times10^{-4}$ 情况，$\alpha_{a,t}=8\times10^{-4}$ 是最佳学习率，在平均收益和方差方面表现优异。

图 7 显示了所提出算法当 $\alpha_{a,t}=8\times10^{-4}$ 的不同 Critic 的学习率收敛性能。可知 Critic 的最佳学习率是 $\alpha_{c,t}=8\times10^{-3}$。当学习率下降到 $\alpha_{c,t}=1\times10^{-3}$ 时学习速度变慢，学习率上升到 $\alpha_{c,t}=3\times10^{-2}$ 时有高方差和低奖励。因此，在下面的仿真中，学习率设置为 $\alpha_{a,t}=8\times10^{-4}$ 和 $\alpha_{c,t}=8\times10^{-3}$。

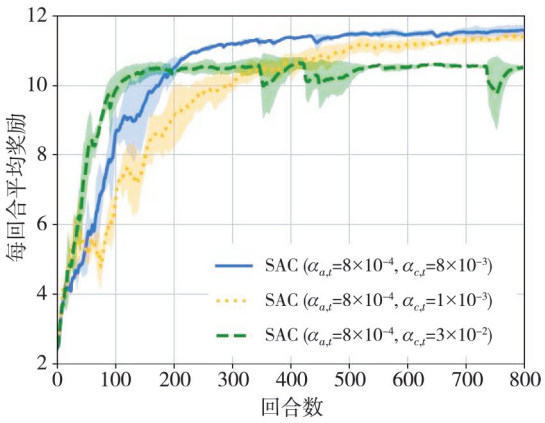

图 7 不同 Critic 学习率

Fig. 7 Different Critic learning rates

图 8 显示了车辆占不同视频层的百分比。图 8(a)作为基线，此时所有的价格系数相等。从图中可以看出，随着无人机传输功率的变大，视频质量稳步增长。图 8(b)～图 8(e)显示了当飞行高度 H=100 m 时，价格系数对视频质量的影响。对比图 8(a)和图 8(b)可知，对于固定的无人机传输功率，当由 0. 5 变成 5 时，具有 BL 数据的车辆百分比明显增加，而具有 BL+EL1 和 BL+EL1+EL2 数据的车辆的百分比明显降低。类似的现象也可以从图 8(c)、图 8(d)和图 8(e)中观察到，N_i 价格系数 ω_i 增加将导致相应视频层的百分比增加。因此，通过调整 N_i 价格系数 ω_i，该算法可以在公平性和效率之间取得平衡，这对于无人机辅助 ITS 中的交通多播服务尤为重要。

图 8（f）显示了当无人机发射功率为 0.06 W，N_i 价格系数 ω_i 全为 0.5 时不同无人机高度下的车辆占不同视频层的百分比，从图中可以看出，当无人机高度从 50 m 上升到 125 m 时，总体视频质量会稍微下降。这是因为较高的海拔导致较差的信道质量和较低的传输速率。

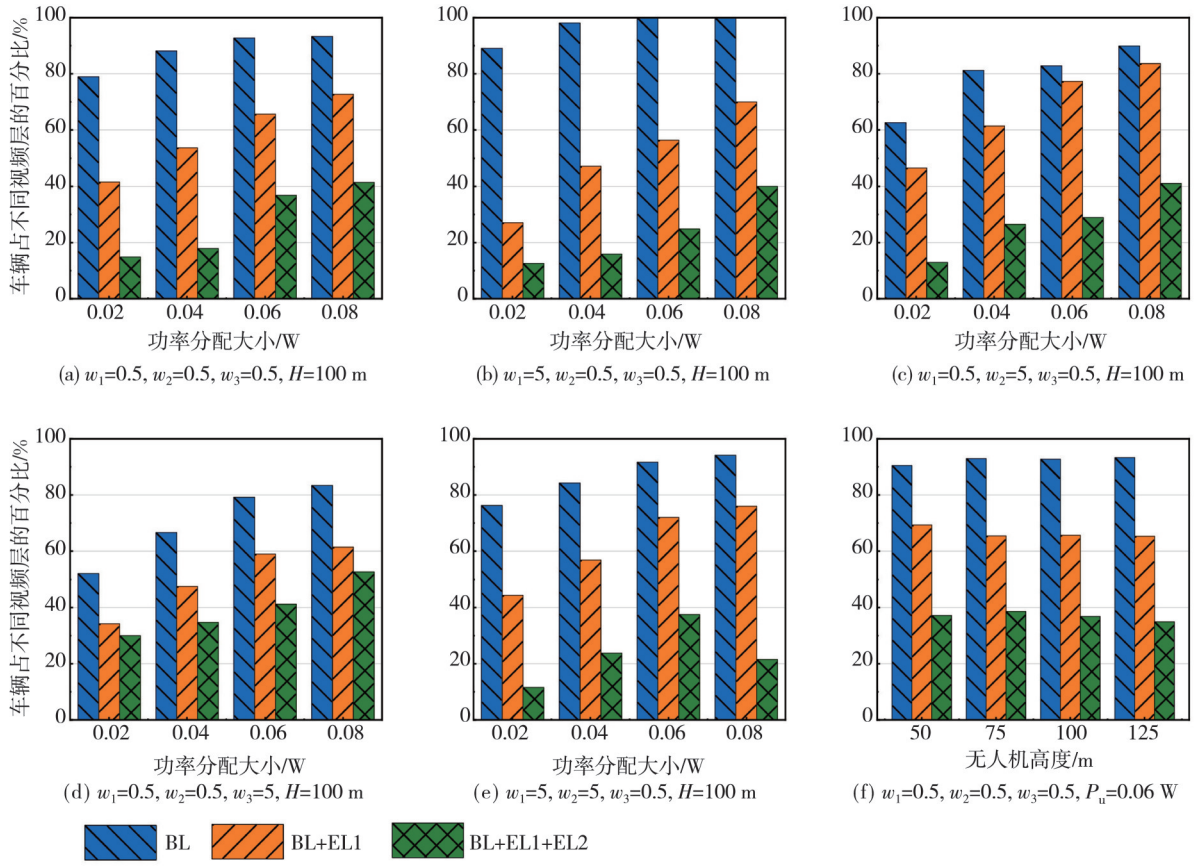

图 8　SAC 算法不同条件下车辆占不同视频层的百分比

Fig. 8　The percentages of vehicles with different video layers of SAC algorithm

图 9 为本文算法与其他基准算法的性能对比。通过比较传统 AC 算法和无学习（Without learning）的方案，可知本文算法平均回报值明显高于其他基准算法。主要是因为本文算法是基于熵最大化的，具有比其他深度强化学习算法更好的探索能力，能够快速稳定地获得更多的探索收益。

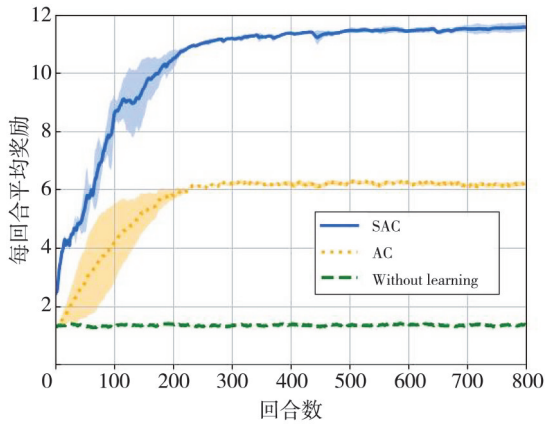

图 9　算法比较图

Fig. 9　Algorithm comparison diagram

4　结　语

提出了一种非正交多址的可伸缩视频多播资源优化研究方案。通过联合优化车辆分组和功率分配策略，最大限度地提高车辆接收的长期视频质量。其中无人机在空中悬停实时捕捉交通视频并多播传输至服务车辆。将该联合优化问题建模成马尔可夫决策过程。采用最先进的深度强化学习算法（SAC）来解决。仿真结果表明，该算法比传统 AC 算法有更强的探索能力。

参考文献：

[1] KUUTTI S, FALLAH S, KATSAROS K, et al. A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications[J]. IEEE Internet of Things Journal, 2018, 5(2): 829-846.

[2] WANG L, YANG H, QI X, et al. ICast: Fine-grained wireless video streaming over Internet of intelli-

gent vehicles[J]. IEEE Internet of Things Journal, 2018, 6(1): 111-123.
[3] KHAN M A, ECTORS W, BELLEMANS T, et al. Unmanned aerial vehicle-based traffic analysis: Methodological framework for automated multivehicle trajectory extraction[J]. Transportation Research Record, 2017(1): 25-33.
[4] MENOUAR H, GUVENC I, AKKAYA K, et al. UAV-enabled intelligent transportation systems for the smart city: applications and challenges[J]. IEEE Communications Magazine, 2017, 55(3): 22-28.
[5] ARANITI G, CONDOLUCI M, SCOPELLITI P, et al. Multicasting over emerging 5G networks: challenges and perspectives[J]. IEEE Network, 2017, 31(2): 80-89.
[6] AFOLABI R O, DADLANI A, KIM K. Multicast scheduling and resource allocation algorithms for OFDMA-based systems: a survey[J]. IEEE Communications Surveys & Tutorials, 2012, 15(1): 240-254.
[7] CHOU Z T, LIN Y H. Energy-efficient scalable video multicasting for overlapping groups in a mobile WiMAX network[J]. IEEE Transactions on Vehicular Technology, 2015, 65(8): 6403-6416.
[8] SHEN S H. Efficient SVC multicast streaming for video conferencing with SDN control[J]. IEEE Transactions on Network and Service Management, 2019, 16(2): 403-416.
[9] YOON J, ZHANG H, BANERJEE S, et al. Video multicast with joint resource allocation and adaptive modulation and coding in 4G networks[J]. IEEE/ACM Transactions on Networking, 2013, 22(5): 1531-1544.
[10] ZHOU H, JI Y, WANG X, et al. Joint resource allocation and user association for SVC multicast over heterogeneous cellular networks[J]. IEEE Transactions on Wireless Communications, 2015, 14(7): 3673-3684.
[11] LIU X, LIN B, ZHOU M, et al. NOMA-based cognitive spectrum access for 5G-enabled internet of things[J]. IEEE Network, 2021, 35(5): 290-297.
[12] DING Z, LEI X, KARAGIANNIDIS G K, et al. A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(10): 2181-2195.
[13] DING Z, YANG Z, FAN P, et al. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users[J]. IEEE Signal Processing Letters, 2014, 21(12): 1501-1505.
[14] CHINGOSKA H, HADZI-VELKOV Z, NIKOLOSKA I, et al. Resource allocation in wireless powered communication networks with non-orthogonal multiple access[J]. IEEE Wireless Communications Letters, 2016, 5(6): 684-687.
[15] TAN X, XU L, ZHENG Q, et al. QoE-driven DASH multicast scheme for 5G mobile edge network[J]. Journal of Communications and Information Networks, 2021, 6(2): 153-165.
[16] UL ZUHRA S, CHAPORKAR P, KARANDIKAR A. Toward optimal grouping and resource allocation for multicast streaming in LTE[J]. IEEE Transactions on Vehicular Technology, 2019, 68(12): 12239-12255.
[17] LUO J, CHEN Q, TANG L, et al. Adaptive resource allocation considering power-consumption outage: a deep reinforcement learning approach[J]. IEEE Transactions on Vehicular Technology, 2023, 72(6): 8111-8116.
[18] DU J, CHENG W, LU G, et al. Resource pricing and allocation in MEC enabled blockchain systems: an A3C deep reinforcement learning approach[J]. IEEE Transactions on Network Science and Engineering, 2021, 9(1): 33-44.
[19] LEI L, YUAN D, HO C K, et al. Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation[J]. IEEE Transactions on Wireless Communications, 2016, 15(12): 8580-8594.
[20] ZHANG Z, ZHANG Q, MIAO J, et al. Energy-efficient secure video streaming in UAV-enabled wireless networks: a safe-DQN approach[J]. IEEE Transactions on Green Communications and Networking, 2021, 5(4): 1892-1905.
[21] 康云鹏. 车联网中基于视频业务的资源分配研究[D]. 太原: 山西大学, 2020.
[22] 王钰宁, 刘晓霞, 胡云冰. 基于能效感知的无人机协助的视频数据传输[J]. 弹箭与制导学报, 2021, 41(6): 7-11.
WANG Yuning, LIU Xiaoxia, HU Yunbing. Energy efficiency awareness-based unmanned aerial vehicles-assisted video data transmission[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2021, 41(6): 7-11. (in Chinese)
[23] ZHANG L, ANSARI N. Latency-aware IoT service provisioning in UAV-aided mobile-edge computing networks[J]. IEEE Internet of Things Journal, 2020, 7(10): 10573-10580.