

Learn to Compress CSI and Allocate Resources in Vehicular Networks 复现研究报告

摘要

资源分配对于车联网 (V2X) 网络的性能具有直接且深远的影响。在该论文中,精心构建了一种融合集中式决策与分布式资源共享的混合架构 (即 C-决策方案),旨在最大程度地提升所有车辆的长期总速率。为有效降低网络信令开销,每辆车均借助深度神经网络对其观测信息进行压缩,随后将压缩后的信息反馈至集中式决策单元。该集中式决策单元运用深度 Q 网络实施资源分配,并将所得决策结果传输给每一辆车。同时,为每辆车增设一个量化层,该量化层能够学习对连续反馈进行量化操作。此外,该论文精心设计了一种平衡车对车 (V2V) 链路和车对基础设施 (V2I) 链路传输的机制。为进一步推动分布式频谱共享,针对每一条 V2V 链路,该论文创新性地提出了分布式决策与频谱共享架构 (即 D-决策方案)。经由大量的仿真实验,研究结果表明,所提出的 C-决策方案方案和 D-决策方案均能实现近乎最优的性能表现,并且在反馈间隔变动、输入噪声以及反馈噪声等复杂情况下,展现出良好的鲁棒性。

关键词: 车载网络; 深度强化学习; 频谱共享;

1 引言

将道路上的车辆连接成一个动态通信网络,通常称为车联网 (V2X),正逐渐成为现实,使我们的日常驾车出行更加安全便捷。借助 V2X 实现车辆、行人和其他道路实体之间的协调,除了提供无处不在的信息娱乐服务外,还可以缓解交通拥堵、提升道路安全状况。在该论文中,我们利用学习方法的数据驱动能力,在高度动态的 V2X 网络中优化频谱共享性能和减少网络信令开销。在高移动性的环境中,基站要获取准确的全局信道状态信息 (CSI) 是具有挑战性的,因为 CSI 变化迅速,同时又决定着频谱共享性能。更糟糕的是,所有车辆将各自的本地 CSI 传输到基站的信令开销也给稀缺的频谱资源带来了沉重负担。反馈开销也随着可用信道数量和车辆数量线性增加。为解决这个问题,我们想要在动态 V2X 网络场景中自动做出明智的频谱共享决策。该论文的贡献如下:

(1) 我们利用 DNN 和 RL 的能力,设计了一种用于车载频谱共享的集中式决策和分布式实现架构,该架构最大化所有车辆的长期总速率。我们使用加权总速率奖励来动态平衡 V2I 和 V2V 性能。

(2) 我们利用每个车辆的 DNN 压缩本地观测信息,并通过量化层进一步增强,以减少网络信令开销,同时实现理想性能。我们还开发了一种分布式决策架构,允许每个车辆在本地做出频谱共享决策,并设计了二进制反馈以减少信令开销。

(3) 基于大量的计算机模拟，我们证明了所提出的两种架构都可以实现接近最优的性能，并且对反馈间隔变化、输入噪声和反馈噪声具有鲁棒性。此外，还给出了每个 V2V 链路的连续反馈和反馈比特的最优数量，以在信令开销和性能损失之间取得平衡。

2 相关工作

由于车辆的高移动性和复杂的时变通信环境，在车载网络中保证多样化的服务质量(QoS)要求极具挑战性，例如超大容量、超高可靠性和低延迟等。为解决这些问题，在 V2X 场景中进行有效的频谱共享资源分配十分必要。现有车载网络频谱共享的相关工作主要可分为两类：集中式方案和分布式方案。

2.1 集中式方案

对于集中式方案，决策通常在给定节点集中做出，在集中式方案中，决策节点需要获取准确的 CSI、所有 V2V 链路的干扰信息以及每个 V2V 链路的发射功率，以做出频谱共享决策。然而，将每个 V2V 链路的所有此类信息报告给决策节点会给反馈链路带来沉重负担，甚至在实际中变得不可行。

2.2 分布式方案

对于分布式方案，每个 V2V 链路在对其他 V2V 链路只有部分或很少了解的情况下做出自己的决策。在分布式方案中，做出决策之前，V2V 链路可能与邻居交换部分或不交换信道信息。然而，在 V2X 场景中，由于每个 V2V 链路在地理位置上与其他 V2V 链路相隔，它只能观察到其周围环境的部分信息。这可能会导致一些信道过度拥塞，而其他信道未得到充分利用，从而导致性能大幅下降。

3 本文方法

3.1 本文方法概述

在 V2X 网络环境中，一种较为简易的分布式策略能够许可各个 V2V 链路自主地选取信道，旨在实现自身数据速率的最大化。然而，鉴于不同 V2V 链路之间存在的干扰情况，单纯追求局部速率最大化常常致使全局性能无法达到最优状态。从另一方面来看，V2X 场景下的基站具备充足的计算与存储资源，足以达成高效的资源分配任务。凭借机器学习技术，我们设计了一种集中式决策方案，该方案依赖于每个 V2V 链路分布式学习所获取的压缩信息。为达成上述目标，每个 V2V 链路率先开展对本地观测信息的压缩学习，这些信息涵盖了信道增益、源自其他 V2V 链路以及 V2I 链路的观测干扰、发射功率等等。随后，将经过压缩处理后的信息反馈至基站。依据来自所有 V2V 链路的反馈信息，基站运用强化学习(RL)算法为所有 V2V 链路制定出最优决策。最后，基站把决策结果以广播形式传递给所有 V2V 链路。如图 1所示：

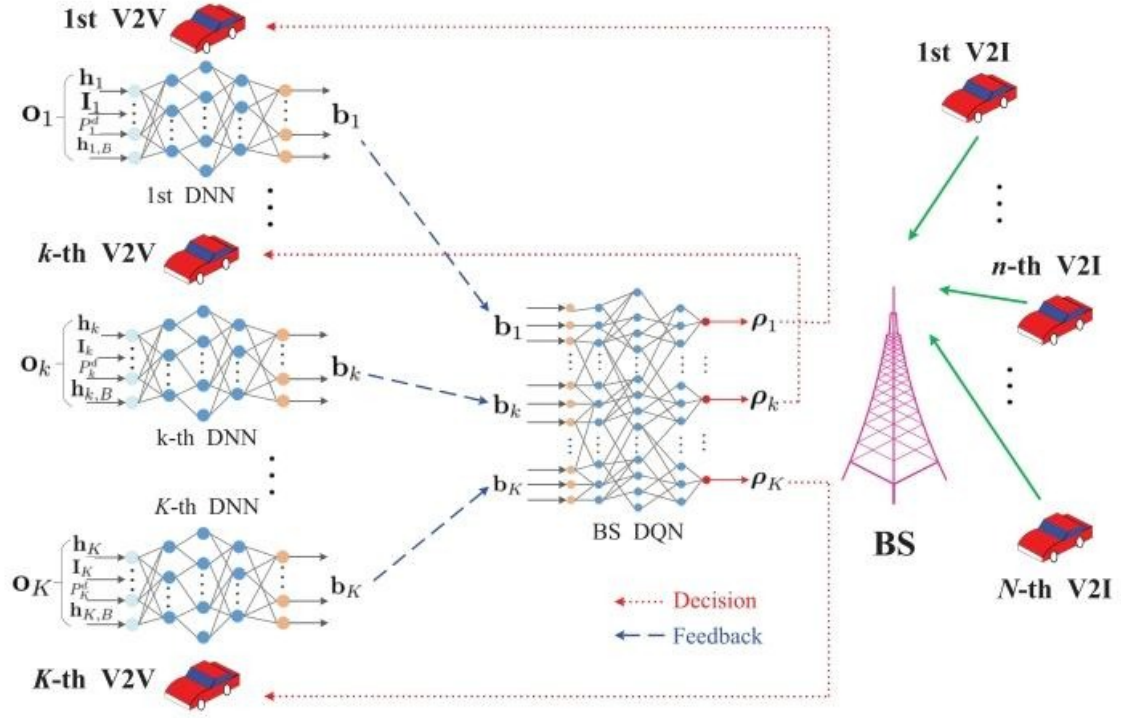


图 1. C 决策方案中 V2V 链路和基站的神经网络架构 [1]

3.2 参数名称及其含义

参数名称	含义
$g_n[n]$	第 n 个 CU 到基站在第 n 个信道上的信道功率增益
$g_{n,k}[n]$	第 n 个 CU 到第 k 个 D2D 对的接收机在第 n 个信道上的交叉信道功率增益
$h_{k,B}[n]$	第 k 个 V2V 链路的发射机到基站在第 n 个信道上的交叉信道功率增益
$h_{l,k}[n]$	第 l 个 D2D 对的发射机到第 k 个 D2D 对的接收机在第 n 个信道上的交叉信道功率增益
$h_k[n]$	第 k 个 V2V 链路在第 n 个信道上的信道功率增益
$\gamma_n^c[n]$	第 n 个 V2I 链路的接收信号与干扰加噪声比
$\gamma_d^k[n]$	第 k 个 V2V 链路在第 n 个信道上的接收信号与干扰加噪声比
P_n^c	第 n 个 V2I 链路的发射功率
P_k^d	第 k 个 D2D 对的发射功率
σ^2	噪声功率
$\rho_k[n]$	信道分配指示符, 若第 k 个 D2D 用户选择第 n 个信道, $\rho_k[n]=1$, 否则 $\rho_k[n]=0$
$C_n^c[n]$	第 n 个 V2I 链路在第 n 个信道上的容量
$C_k^d[n]$	第 k 个 V2V 链路在第 n 个信道上的容量
B	信道带宽
$I_k[n]$	第 k 个 V2V 链路的干扰功率
$h_{k,B}[n]$	第 k 个 V2V 链路到所有 V2I 链路的交叉信道增益

3.3 系统模型

我们考虑一个车载通信网络，这个通信网络中有 N 个蜂窝用户（CUs）和 K 对共存的设备对设备（D2D）用户。其中所有的设备都配备单天线。其中令集合 $N = \{1, 2, \dots, N\}$ 代表着 CUs 集合， $K = \{1, 2, \dots, K\}$ 代表着 D2D 对集合。在车载通信网络中，每一对 D2D 用户借助建立的 V2V 链路，进行重要且简短消息的交互，此类消息涵盖与安全相关的信息等。而各个 CU 则凭借 V2I 链路，为带宽密集型应用提供支持，像社交网络以及视频流等应用。为保障 CUs 的服务质量（QoS），我们假定所有的 V2I 链路均被分配正交无线电资源。在不影响一般性的前提下，进一步假设每个 CU 占用一个信道用于上行传输。为提升频谱利用效率，所有的 V2V 链路均与 V2I 链路共同分享频谱资源。由此， N 被称作信道集。第 n 个 V2I 链路的接收信号与干扰加噪声比（SINR）可表示为：

$$\gamma_n^c[n] = \frac{P_n^c g_n[n]}{\sum_{k=1}^K \rho_k[n] P_k^d h_{k,B}[n] + \sigma^2}$$

假设每个 D2D 对仅占用一个信道，即 $\sum_{n=1}^N \rho_k[n] \leq 1$ 。那么第 n 个 V2I 链路在第 n 个信道上的容量可写为：

$$C_n^c[n] = B \log_2(1 + \gamma_n^c[n])$$

类似的，第 k 个 V2V 链路在第 n 个信道上的 SINR 可写为：

$$\gamma_k^d[n] = \frac{\rho_k[n] P_k^d h_k[n]}{I_k[n] + \sigma^2}$$

其中第 k 个 V2V 链路的干扰功率 $I_k[n]$ 为：

$$I_k[n] = \sum_{l \neq k}^K \rho_l[n] P_l^d h_{l,k}[n] + P_n^c g_{n,k}[n]$$

那么第 k 个 V2V 链路在第 n 个信道上的容量可写为：

$$C_k^d[n] = B \log_2(1 + \gamma_k^d[n])$$

3.4 C 决策方案

如图 1 所示，对于资源分配，我们采取深度强化学习的方式来解决，设计每个 V2V 链路的深度神经网络（DNN）和用于集中控制决策的基站的深度 Q 网络（DQN），这是一种集中式决策和分布式频谱共享架构，称之为 C 决策方案。

3.5 C 决策方案 V2V 的 DNN 设计

首先对于每个 V2V 链路需要设计一个 DNN，用于对观测信息的压缩，从而减少传输观测信息所需要的开销。对于每个 V2V 链路 k 需要观察周围环境并收集信息，需要收集的信息包括其传输功率、所有信道的当前信道功率增益和干扰功率。则第 k 个 V2V 链路观测信息可以表示为：

$$o_k = \{h_k, I_k, P_k^d, h_{k,B}\}$$

然后每个 V2V 链路使用 DNN 将观测到的信息 o_k 压缩，压缩后的信息为 b_k (DNN 的输出)，被反馈给基站的 DQN。为了减少开销每个 V2V 链路反馈的是压缩后的信息 b_k ，而不是观测信息 o_k ，所有 V2V 链路旨在从长远来看最大化其全局总速率，同时最小化反馈信息 b_k 。因此，所有 V2V 链路的 DNN 参数和 DQN 参数将共同确定，以最大化整个 V2X 网络的总速率。

3.6 基站的深度 Q 网络

为了做出适当的资源共享决策，我们在基站引入深度强化学习架构，如图 1 所示。为了最大化所有链路的长期总速率，我们通过将基站视为智能体来运用强化学习技术。在强化学习中，智能体通过采取行动与其周围环境（称为环境）交互，然后从环境中观察到相应的数值奖励。智能体的目标是找到最优行动，以使预期奖励总和最大化。在每个时间步 t 中，智能体从状态空间 S 中获取状态 S_t ，然后在动作空间 A 中选取一个动作 A_t ，然后得到一个奖励 R_{t+1} ，然后状态由 S_t 转变成 S_{t+1} 。定义将所有 V2V 链路反馈回来的信息作为当前状态 s ，可以表示为：

$$s = \{b_1, b_2, \dots, b_K\}, \forall k \in \mathcal{K}$$

然后，基站的行动是确定每个 V2V 链路的信道指示符 $\rho_k[n]$ 的值，定义基站的行动 a 为：

$$a = \{\rho_1, \rho_2, \dots, \rho_K\}, \forall k \in \mathcal{K}$$

其中 $\rho_K = \{\rho_K[n]\}, \forall n \in \mathcal{N}$ 。最后，对于奖励的设置，这是强化学习中重要的一环，我们的目标是在 V2X 场景中，最大化 V2V 链路的长期总速率，同时还需要保证 V2I 的服务质量。所以我们在设计奖励函数是需要将这两个因素都考虑。但在大多是情况 V2V 链路都携带着重要信息，所以我们还是以 V2V 链路的速率为首要目标。所以将奖励函数表示为：

$$R = \lambda_c \sum_{n=1}^N C_n^c[n] + \lambda_d \sum_{k=1}^K C_k^d$$

其中 λ_c 和 λ_d 为非负权重，用于权衡 V2I 和 V2V 链路的性能。

然而，在大多情况下，状态和动作的空间非常大，将所有的动作价值函数存储下来是不现实的。所以通常使用函数逼近来估计这些动作价值函数。采用由 θ 参数化的深度神经网络 (DNN) 来表示动作价值函数，因此称为 DQN。在每个时间步将 $(S_t, A_t, R_{t+1}, S_{t+1})$ 转移元组存储在重放记忆中。累积智能体在 MDP 中的经验。在每个时间步，从重放记忆中均匀采样一小批经验 D ，称为经验重放，使用随机梯度下降方法的变体更新网络参数，以最小化如下所示的平方误差：

$$\sum_{t \in D} \left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a; \theta^-) - Q(S_t, A_t; \theta) \right]^2$$

其中 θ^- 参数来自与参数 θ 的复制，用于稳定学习的过程。

3.7 D 决策方案

为了进一步促进分布式频谱共享并降低计算复杂度，我们提出了如图 2 所示所示的分布式决策与频谱共享架构（称为 D 决策方案），以使每个 V2V 链路能够在本地做出频谱共享决策。

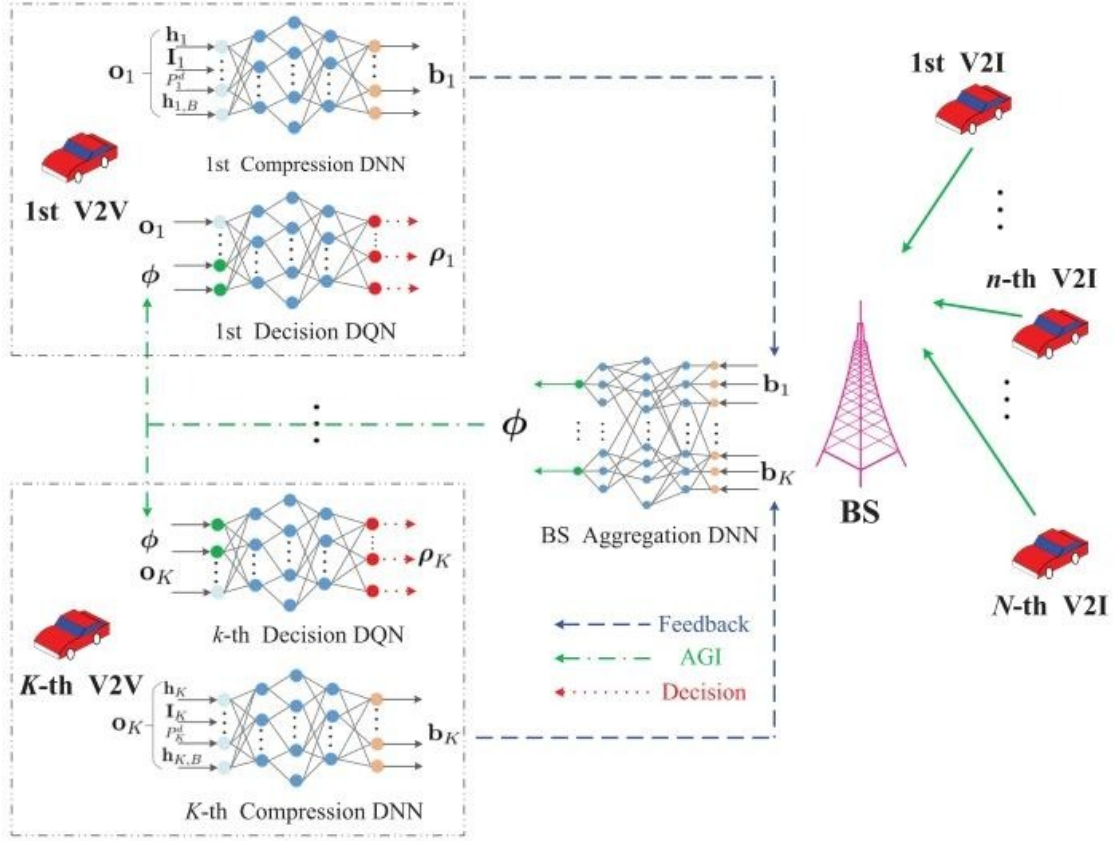


图 2. D 决策方案中车对车 (V2V) 链路和基站的神经网络架构 [1]

3.8 D 决策方案 V2V 和基站的 DNN 设计

D 决策方案需要每个 V2V 链路通过自身的 DNN 压缩其观测信息，这与 3.5 中的 DNN 一致。然后基站收到反馈信息后，需要将所有的反馈信息，通过基站自身的 DNN 聚合成一个全局消息 (AGI) 广播给所有 V2V 链路。最后，每个 V2V 链路将其本地观测和 AGI 组合作为其决策 DQN 的输入，以决定在哪个信道上进行传输。

4 复现细节

4.1 与已有开源代码对比

论文中设置了许多消融实验,复现过程中参考了源代码来源:<https://github.com/CooperLWang/LearnCompressCSI-RA-V2X-Code>。然后同时对代码理解并进行整理。

4.2 实验环境搭建

使用工具: pycharm(python 3.6)
主要库及其版本

库名称	版本
Keras	2.2.4
TensorFlow	1.12.0
numpy	1.16.5

4.3 创新点

(1) 添加正则化项：在 DQN 和 DNN 模型的损失函数中添加正则化项，如 L1 或 L2 正则化，防止模型过拟合。过拟合可能导致模型在实际 V2X 网络环境中的泛化能力下降，而正则化可以约束模型的复杂度，提高模型对不同场景和数据的适应性，增强模型的稳定性。

(2) 在数据预处理阶段，检测并处理可能存在的异常值。异常值可能对模型训练产生较大影响，导致模型参数更新不稳定。可以通过统计方法（如计算均值和标准差，设定阈值来判断异常值）或基于领域知识的方法来识别和处理异常值，确保模型输入数据的质量，从而提高模型的稳定性和准确性。

5 实验结果分析

5.1 C 决策方案结果

模型训练 Loss 图如图 3：

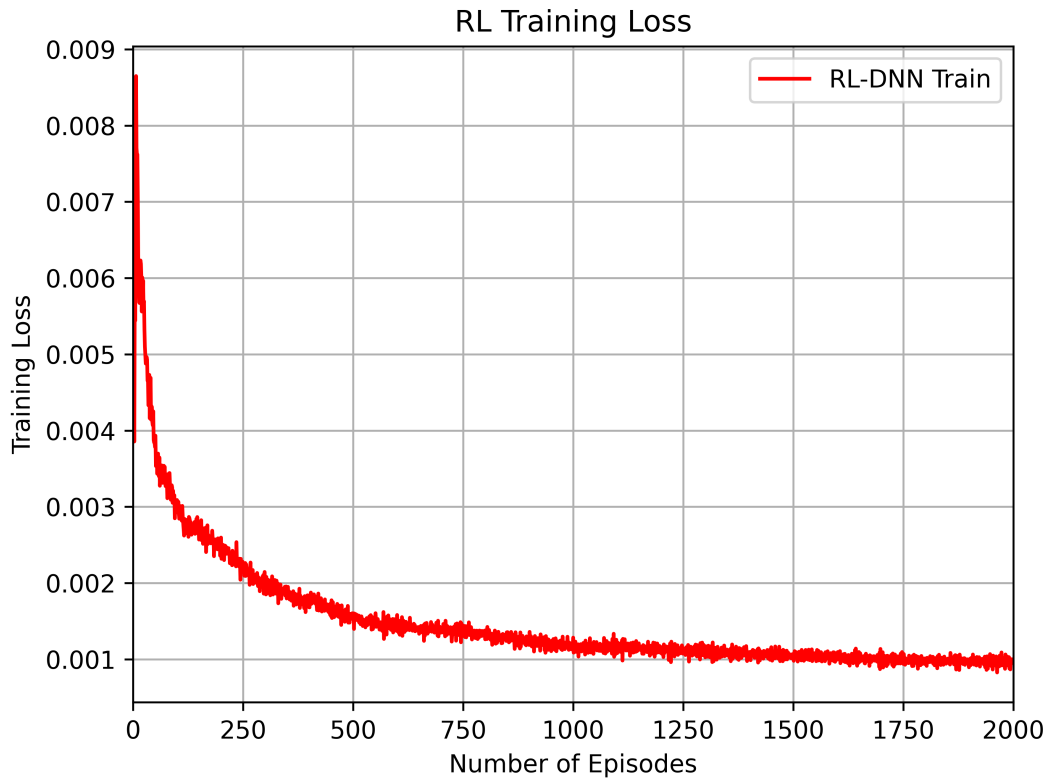


图 3. Train loss

如图 3 所示，模型在训练至 2000 轮时 loss 趋于稳定，其稳定在 0.001 附近。模型评估平均回报图如图 4

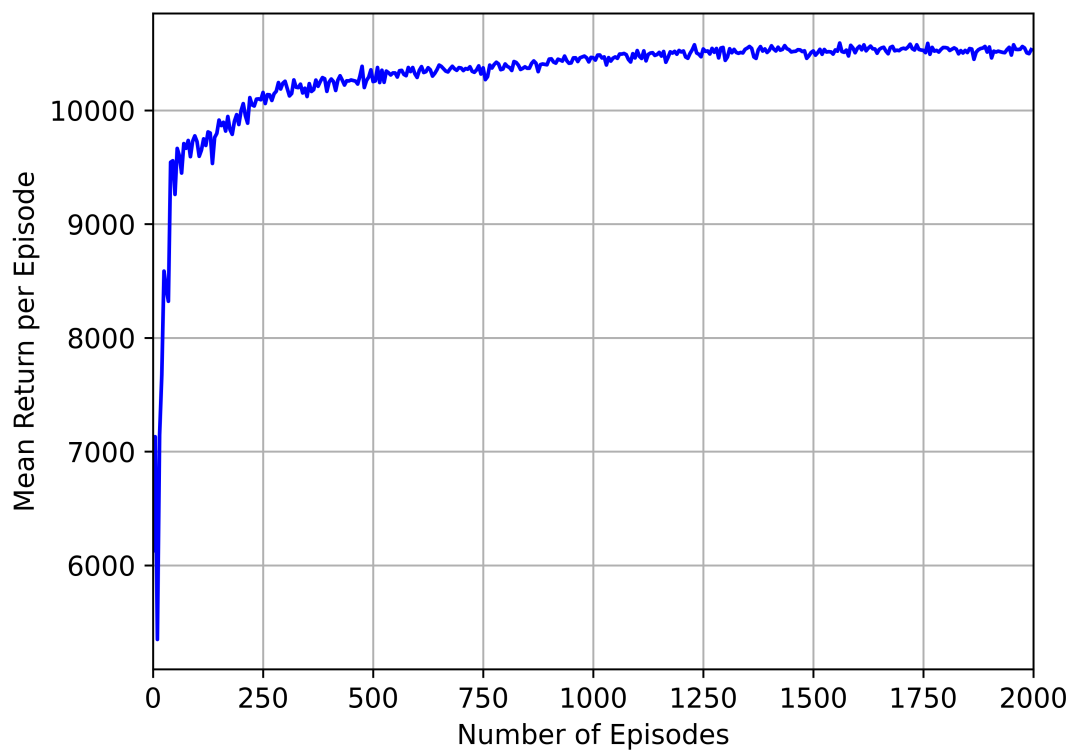


图 4. Average return per episode

如图 4 所示，随着轮数的提高，平均回报也在升高，在 2000 轮时趋于不变，逐渐收敛。

5.2 D 决策方案结果

模型训练 Loss 图如图 5：

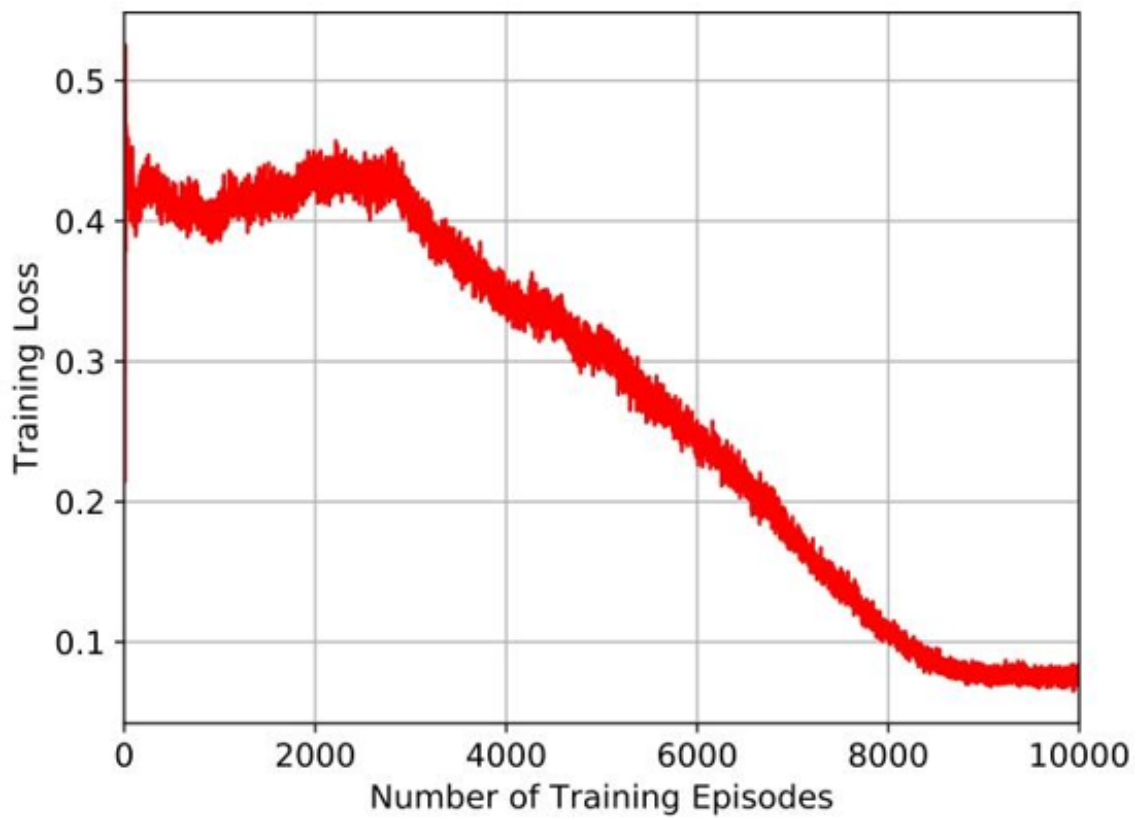


图 5. Train loss [1]

随着轮数的提高，模型在 10000 轮时 loss 趋于稳定，逐渐收敛。模型评估平均回报图如图 6

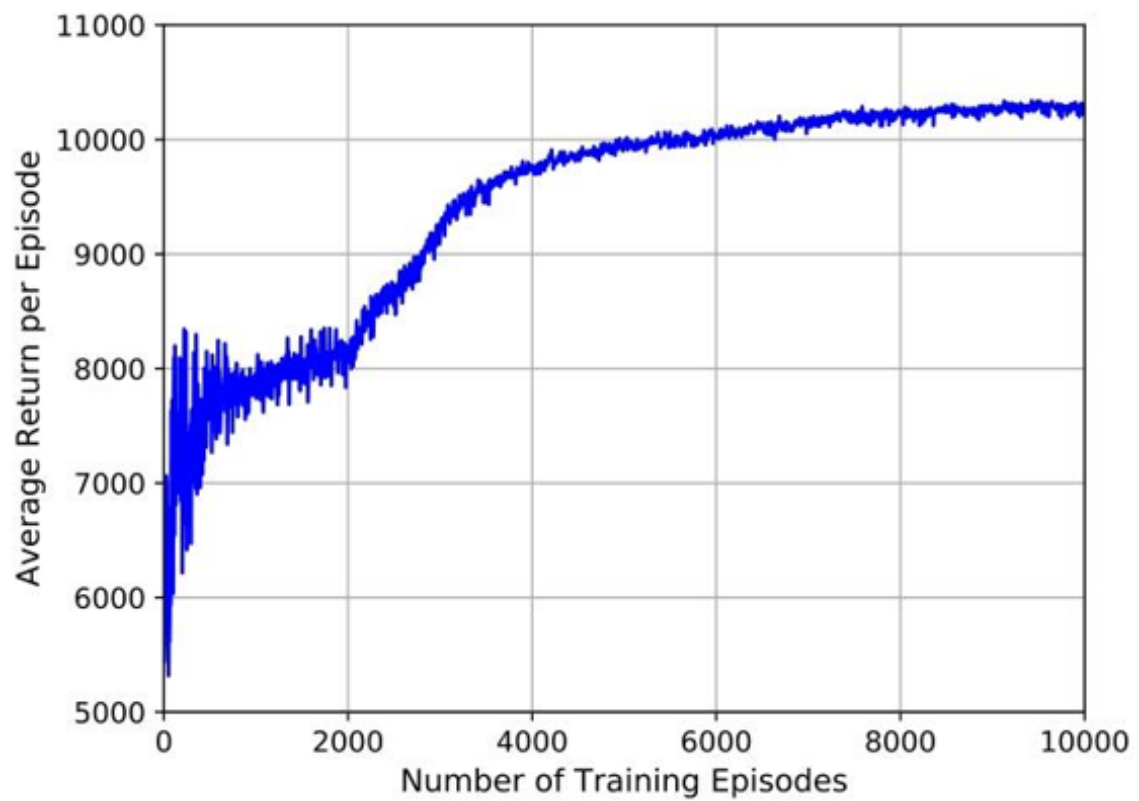


图 6. Average return per episode [1]

如图 D 决策方案的平均回报最终在 10000 轮时达到稳定。

6 总结与展望

本文对复现论文的内容进行了概括，主要是复现论文提出的两个方案，分别是 C 决策方案和 D 决策方案，并对 C 决策方案的方法和代码进行了复现，复现过程中，不足的就是没有将论文设置所有的消融实验全部复现出来，可能会导致实验结果比较少。对于提出的这两种方案对反馈间隔、输入噪声以及反馈噪声的变化分别具有较强的抗干扰能力，这验证了所提方案的稳健性。未来可以在联合功率控制和频谱共享问题展开研究

参考文献

- [1] Liang Wang, Hao Ye, Le Liang, and Geoffrey Ye Li. Learn to compress csi and allocate resources in vehicular networks. *IEEE Transactions on Communications*, 68(6):3640–3653, 2020.