非平稳环境中的样本高效经验回放

Tianyang Duan*, Zongyuan Zhang*, Songxiao Guo*, Yuanye Zhao†, Zheng Lin‡, Zihan Fang§,
Yi Liu§, Dianxin Luan¶, Dong Huang‖, Heming Cui*, Yong Cui**
*Department of Computer Science, The University of Hong Kong, Hong Kong, China
†College of International Education, Hebei University of Economics and Business, China
‡Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China
§Department of Computer Science, City University of Hong Kong, Hong Kong, China
¶Institute for Imaging, Data and Communications, University of Edinburgh, UK
‖School of Computing, National University of Singapore, Singapore
**Department of Computer Science and Technology, Tsinghua University, China

摘要—强化学习（RL）在非平稳环境中的挑战在于，变化的动力学和奖励迅速使过去的经历过时。传统的经验回放（ER）方法，特别是那些使用 TD 误差优先级的方法，在区分由于智能体策略变化引起的变化与环境引起的变 化方面存在困难，导致在动态条件下学习效率低下。为了解决这一挑战，我们提出了环境动力学差异（DoE），这是一个能够隔离环境变化对价值函数影响的指标。在此基础上，我们引入了环境优先级经验回放（DEER），这是一种自适应的 ER 框架，根据策略更新和环境变化来优先处理转换。DEER 使用一个二元分类器来检测环境的变化，并在每次转变前和之后应用不同的优先级策略，从而实现更高效的样本学习。在四个非平稳基准测试上的实验表明，与最先进的 ER 方法相比，DEER 使离线算法的性能提高了 11.54%。

***Index Terms*—强化学习，非平稳环境，经验回放，离策略算法**

I. 介绍

强化学习（RL）[1], [2] 是一种强大的动态序列决策规划方法，广泛应用于实际场景中 [3]。然而，实际环境往往是非平稳的，随着时间的推移，环境动力学和奖励信号会发生变化 [4]。这使得适应性规划对于应对现实世界的复杂性和不可预测性至关重要。

离策略强化学习（RL）利用历史经验来解决高维、连续动作空间中稀疏和昂贵采样的挑战 [5]。这种效率主要通过经验回放（ER）实现，经验回放存储并重用过去的转换以打破时间相关性并提高数据效率。最近的研究通过引入非均匀采样改进了 ER，基于时序差分误差（TD-error）的优先级在平稳环境中显著提高了样本效率 [6], [7]。TD-error 量化了估计回报与实际回报之间的差异，使代理能够识别和优先处理提供最有信息量经验的转换。因此，具有较高 TD-error 的转换被更频繁地采样，从而加速收敛并提高性能。

然而，在非平稳环境中，历史经验很快就会过时，这可能会破坏有效样本选择并误导学习 [8]。大多数现有方法忽视了过时转移的负面影响，尤其是在 TD 误差受到环境变化和策略更新双重影响的情况下。当价值函数适应新环境后，之前收集的过渡通常表现出更高的 TD 误差，并优先考虑这些过渡会放大不相关经验，降低训练效率和性能。类似的问题也出现在基于奖励 [9] 或基于频率采样 [6] 的情况中。最终，仅专注于策略改进的优先策略无法应对动态环境并且不能准确评估存储过渡的相关性。

为了解决非平稳环境的挑战，我们提出了环境差异（DoE），这是一个量化环境变化对状态转换影响的原则性指标。通过测量动态变化前后动作值函数之间的偏差——同时排除策略改进的影响——DoE 精确地将价值变化归因于潜在的环境动力学。在此基础上，我们提出了一种样本高效的回放缓冲框架 Discrepancy of Environment Prioritized Experience Replay（DEER），该框架自适应地优先考虑经验样本以进行策略优化和环境适应。DEER 利用一个二元分类器通过估计相邻时间窗口中的奖励序列分布来检测动态变化，并对检测到的变化前后的转换应用不同的优先级策略。在变化前具有较低 DoE 的转换被认为更相关并被优先处理，而变化后的转换则使用 TD 误差与实时 DoE 基于密度差异

的混合方法进行排名。这种方法保持了回放缓冲区的多样性，并动态分配采样优先级以满足代理需求。在四个非平稳 Mujoco 标准测试中进行了广泛的实验表明，相较于最先进的经验回放方法，DEER 进一步提高了离线策略算法的性能，提升了 11.54%。此外，在极端非平稳设置（200% 变化）下，相对于最佳 ER 方法，DEER 为离线策略算法额外带来了 22.53% 的性能提升。

II. 相关工作

经验回放缓冲通过存储和重采样过去的经历来提高样本效率和学习稳定性 [10]。虽然它有可能加速在非平稳环境中的适应，但这一优势仍需进一步探索。主要的方法是 PER（优先体验回放），根据 TD 误差对转换进行抽样，在各种 RL 基准测试中取得了显著的收益 [6], [7]。除了基于 TD 误差的采样，还开发了几种替代策略：PSER 增加了前导于重要事件的转换的优先级 [11]；ReF-ER 将采样限制在由政策相似性定义的“近似政策”转换上 [12]；而 AER 强调与代理当前状态类似的转换 [6]。此外，HER 及其变体 [13]–[15] 通过重新标记目标来解决稀疏奖励环境的问题，从而提供更丰富的奖励信号。RB-PER [16] 将更高的优先级分配给较少采样的转换，促进采样多样性并增强对非平稳性的适应能力。

III. 方法论

A. 问题形式化

非平稳环境中的强化学习被公式化为一组马尔可夫决策过程（MDPs）[17]，表示为 $\{\langle \mathcal{S}, \mathcal{A}, P_i, R_i, T_i, \gamma \rangle\}_{i=0}^{\infty}$。这里，$\mathcal{S}$ 是状态空间，$\mathcal{A}$ 是动作空间。在每个时间步骤 t，代理使用策略 $\pi(a_t \mid s_t)$ 选择一个动作 $a_t \in \mathcal{A}$，根据状态转移概率 $P(s_{t+1} \mid s_t, a_t)$ 转移到下一个状态 $st+1 \in \mathcal{S}$，并收到一个奖励 $r_t = R(s_t, a_t)$。非平稳环境中的状态转移函数定义为：

$$P\left(s_{t+1} \mid s_t, a_t\right) = \begin{cases} P_0\left(s_{t+1} \mid s_t, a_t\right), 0 \le t \le T_0 \\ \dots \\ P_i\left(s_{t+1} \mid s_t, a_t\right), T_{i-1} < t \le T_i \\ \dots \end{cases} \tag{1}$$

其中 T_i 表示环境动态发生变化的时间步。非平稳环境的奖励函数定义为：

$$R\left(s_t, a_t\right) = \begin{cases} R_0\left(s_t, a_t\right), 0 \le t \le T_0 \\ \dots \\ R_i\left(s_t, a_t\right), T_{i-1} < t \le T_i \\ \dots \end{cases} \tag{2}$$

代理的目标是学习一个策略 π 以最大化期望折损回报 $\mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k}\right]$，其中 γ 表示折扣因子。

B. 环境差异（DoE）

一个有效的经验回放机制对于非平稳环境应该优先考虑有助于智能体快速适应动态变化的转换。为了实现这一点，我们通过分析当前策略 $\pi(a \mid s)$ 下生成的智能体-环境轨迹 $\tau = [s_0, a_0, s_1, a_1, \dots]$ 来跟踪环境动力学的变化。由于状态转移函数捕获了环境的动力学特性，这些轨迹中的变化提供了可靠的动态转变信号：

$$P\left(\tau\right) = P\left(s_0\right) \prod_{t=0}^{\infty} \pi\left(a_t \mid s_t\right) P\left(s_{t+1} \mid s_t, a_t\right), \tag{3}$$

其中 $P(s_0)$ 表示初始状态分布。给定固定初始状态分布和策略的情况下，任何环境动态的变化都会直接影响到轨迹的分布。由于奖励仅由奖励函数决定，因此奖励序列的 概率密度 ，$\boldsymbol{r}_\tau = [r_0, r_1, \dots]$，是由轨迹分布引起的。因此，可以通过分析奖励序列的变化来近似环境动态的变化。

为此，我们提出了一种基于密度比估计（DRE）[18] 的环境动态监测方法，利用在奖励序列上训练的二元分类器。具体来说，在每个时间步 t，我们构建两个相邻的滑动窗口：一个标记为 $l = 0$ 的参考窗口 $\widetilde{\boldsymbol{r}}_{rf} = \{\boldsymbol{r}_{i:j} \mid t - 2m + 1 \le i < j \le t - m\}$ 和一个标记为 $l = 1$ 的测试窗口 $\widetilde{\boldsymbol{r}}_{te} = \{\boldsymbol{r}_{i:j} \mid t - m < i < j \le t\}$，其中 m 表示窗口大小。二元分类器被训练来区分来自参考窗口和测试窗口的样本，从而估计它们分布之间的密度比。形式上，分类器将条件分布建模如下：

$$P\left(\boldsymbol{r} \mid f\left(\boldsymbol{r}\right) = l\right) = \begin{cases} P_{rf}\left(\boldsymbol{r}\right) & \text{if } l = 0 \\ P_{te}\left(\boldsymbol{r}\right) & \text{if } l = 1 \end{cases} \tag{4}$$

其中 $P_{rf}\left(\cdot\right)$ 和 $P_{te}\left(\cdot\right)$ 分别表示参考窗口和测试窗口的密度函数，而 $f\left(\cdot\right)$ 代表二元分类器。我们采用一个多层感知器（MLP）作为二分类器，并通过最小化以下交叉熵损失函数来优化其参数：

$$\mathcal{L}\left(f\right) = -\frac{1}{|\widetilde{\boldsymbol{r}}_{rf}|} \sum_{\boldsymbol{r} \in \widetilde{\boldsymbol{r}}_{rf}} \log\left(1 - f\left(\boldsymbol{r}\right)\right) - \frac{1}{|\widetilde{\boldsymbol{r}}_{te}|} \sum_{\boldsymbol{r} \in \widetilde{\boldsymbol{r}}_{te}} \log f\left(\boldsymbol{r}\right). \tag{5}$$

如果时间步长 $t - m$ 处的密度比分数 $S\left(\widetilde{\boldsymbol{r}}_{te}, \widetilde{\boldsymbol{r}}_{rf}\right)$ 超过预定义阈值 μ（即 $S \ge \mu$），则识别出一个变化点。具体来

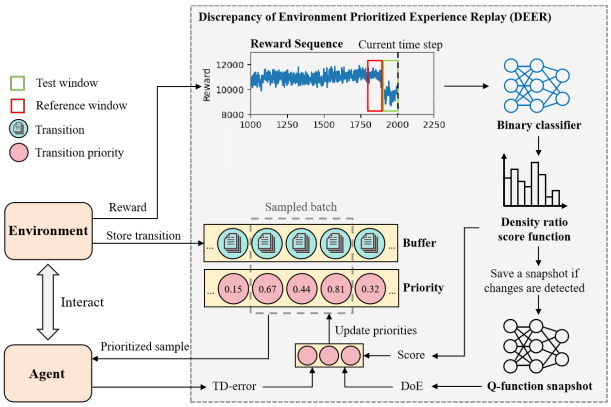


图 1: DEER 的工作流程。

说，我们采用 Jensen-Shannon 散度作为密度比分数函数，其定义为：

$$S(\widetilde{\boldsymbol{r}}_{te}, \widetilde{\boldsymbol{r}}_{rf}) = \log 2 + \frac{1}{2n_{te}} \sum_{\boldsymbol{r} \in \widetilde{\boldsymbol{r}}_{te}} \log P(\boldsymbol{r} \mid f(\boldsymbol{r}) = 1) + \frac{1}{2n_{rf}} \sum_{\boldsymbol{r} \in \widetilde{\boldsymbol{r}}_{rf}} \log P(\boldsymbol{r} \mid f(\boldsymbol{r}) = 0). \quad (6)$$

在检测到变化点后，我们提出使用环境差异（DoE）指标来严格量化环境动态变化对状态-动作价值估计的影响。具体来说，DoE 测量给定状态-动作对 (s_k, a_k) 在环境变化前后的 Q 函数差异：

$$\begin{aligned} DoE(s_k, a_k) &= \mathbb{E}_{s \sim P_i, a \sim \pi} \left[\sum_{j=k}^{\infty} \gamma^{k-t} R_i(s_j, a_j) \right] \\ &\quad - \mathbb{E}_{s \sim P_{i-1}, a \sim \pi} \left[\sum_{j=k}^{\infty} \gamma^{j-t} R_{i-1}(s_j, a_j) \right] \\ &= Q_i(s_k, a_k) - Q_{i-1}(s_k, a_k), \end{aligned} \quad (7)$$

其中 Q_{i-1} 表示来自先前环境的 Q 函数 $\langle P_{i-1}, R_{i-1} \rangle$，而 Q_i 表示在新动态 $\langle P_i, R_i \rangle$ 下的当前 Q 函数。

C. 环境优先经验回放（DEER）的不一致问题

我们进一步提出了 DEER，如图 1所示。DEER 优先考虑在变化前具有低环境变化程度（DoE）的转换，因为这些转换受环境变化的影响较小，因此在变化后仍保持其相关性。特别地，时间步 $k (k \leq T_{i-1})$ 收集的转换分配的优先级定义为:

$$p_k = 2\sigma(-|DoE(s_k, a_k)|), \quad (8)$$

其中 p_k 表示在时间步 $k (k \leq T_{i-1})$ 收集的转换的优先级，而 σ 代表用于归一化优先级的 sigmoid 函数。对于变化后的转换，DEER 使用一种由实时密度比分数指导的混合优先采样策略，该策略结合了 TD 错误和 DoE。

表 I: 变化后 SAC+HalfCheetah 在不同非平稳水平下 10^5 步的平均剧集奖励。

Offset	0%	50%	200%
DEER	11894.85 ± 182.36	**11789.66 ± 283.81**	**9856.27 ± 477.04**
PER	11834.68 ± 244.31	10266.77 ± 327.79	5204.22 ± 763.55
RB-PER	11754.29 ± 202.61	10506.37 ± 364.37	7332.36 ± 318.96
CER	**11975.13 ± 256.47**	9428.88 ± 199.17	8043.76 ± 467.35
LA3P	11895.44 ± 267.72	11024.15 ± 787.15	6614.79 ± 285.27

密度比率分数（公式 6）量化了奖励序列的变化，其中高分表示代理正在适应改变的环境。在这种情况下，具有较高 DoE 的转换被优先考虑以加速适应。相反，当密度比分数低时，采样强调具有更高 TD 错误的转换以促进策略细化。这种自适应优先级采样机制被形式化定义如下：

$$\begin{aligned} p_k &= (1 - S(\widetilde{\boldsymbol{r}}_{te}, \widetilde{\boldsymbol{r}}_{rf}))(2\sigma(|TD(s_k, a_k, r_k, s_{k+1})|) - 1) \\ &\quad + S(\widetilde{\boldsymbol{r}}_{te}, \widetilde{\boldsymbol{r}}_{rf})(2\sigma(|DoE(s_k, a_k)|) - 1), \end{aligned} \quad (9)$$

其中，TD 表示 TD 错误 [19]，p_k 表示在时间步 $k (k > T_{i-1})$ 时收集的转换的优先级，而 $S(\widetilde{\boldsymbol{r}}_{te}, \widetilde{\boldsymbol{r}}_{rf})$ 表示当前时间步 t 的密度比分数。为了缓解由于频繁重播高优先级过渡而引起的过拟合，我们将前置和后置转换的采样概率标准化如下：

$$P(k) = \frac{p_k^{\alpha}}{\sum_i p_i^{\alpha}}, \quad (10)$$

其中 $\alpha \in (0, 1]$ 控制优先程度。由于优先级抽样改变了用于 Q 函数估计的分布，我们应用重要性采样来纠正由此产生的偏差。具体来说，每次转换的更新权重计算为：

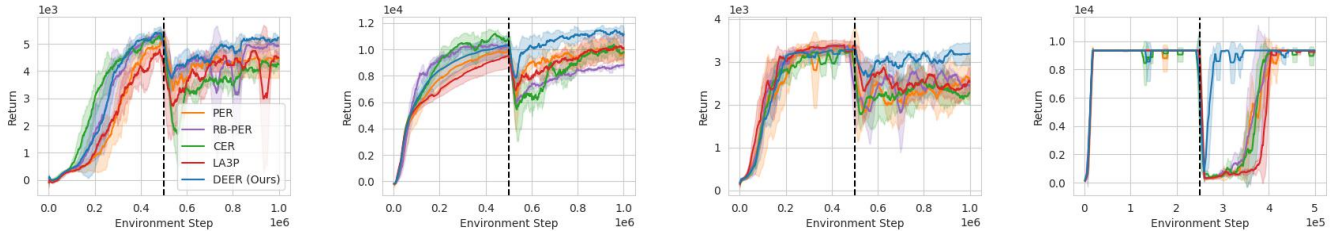
$$w_k = \frac{1}{(N \cdot P(k))^{\beta} \cdot \max_i w_i}, \quad (11)$$

其中 N 表示回放缓冲区大小，$\beta \in (0, 1]$ 控制偏差校正的程度。

IV. 评估

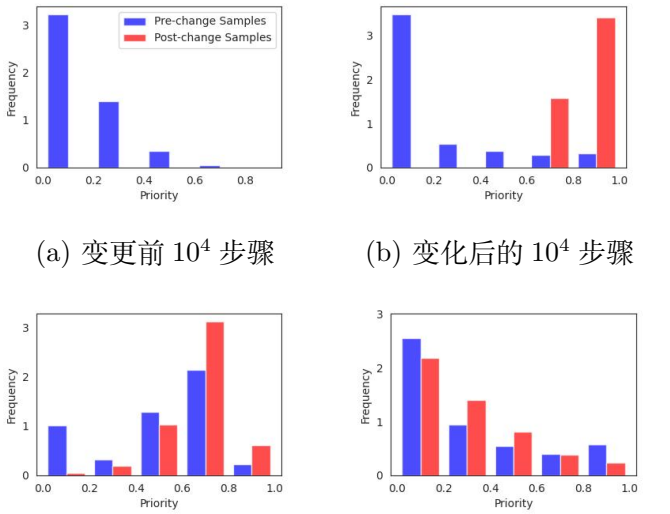
A. 实验设置

我们评估了所提出的与 SAC 算法 [20] 在四个标准 MuJoCo Gymnasium 连续控制任务上的集成：Ant-v4、HalfCheetah-v4、Hopper-v4 以及 Inverted Double Pendulum-v4 (ID-Pendulum-v4)。为了模拟非平稳环境，我们向摩擦和关节阻尼系数引入了偏移量——设置为仍允许 SAC 使用均匀采样收敛的最大值。我们比较



(a) 蚂蚁　(b) 半猎豹　(c) 霍珀　(d) ID-摆

图 2: 样本效率比较在四个任务中的 SAC。虚线标记环境动态变化的时间点。



(a) 变更前 10^4 步骤　(b) 变化后的 10^4 步骤

(c) 变化后的 10^5 步骤　(d) 变化后的 5×10^5 步骤

图 3: SAC+DEER 在 HalfCheetah 中不同训练阶段前后变化样本的优先分布。

了几种经验回放（ER）方法，包括 PER [19]、RB-PER [16]、CER [21]、LA3P [22] 以及我们提出的 DEER。训练在 ID-Pendulum 上运行了 5×10^5 步，而在其他任务上则运行了 1×10^6 步，并且在中途引入了非平稳性。结果基于五次运行取平均值，标准差以阴影区域表示。所有网络均使用两个隐藏层，每层 256 个单元，学习率为 1×10^{-3}，折扣因子为 0.99，回放缓冲区大小为 1×10^6，批量大小为 256。对于 DEER，我们设置 $\alpha = 0.6$，$\beta = 0.4$，并使用一个二元分类器（带有两个隐藏层的 MLP，每层 100 个单元）进行最多 50 次迭代训练。检测窗口大小为 500，每个窗口包含 10 个样本（每个样本包含 50 个数据点），检测阈值为 0.5。

B. 实验结果与分析

图 2a–2d展示了将各种 ER 方法与 SAC 算法集成时的回放曲线。值得注意的是，在两种情况下，DEER（用红线表示）的整体回报均高于基线方法。在各种环境变化中，DEER 的奖励减少较少且恢复速度更快，这表明其适应动态环境的能力更强。

为了进一步分析 DEER 的采样行为，图 3展示了在不同训练阶段回放缓冲区中样本优先级分布情况。总体而言，变化后的过渡优先级高于变化前的过渡优先级，在环境动态变化初期这种差异尤为显著。随着密度比分数随训练进程减少，整体样本分布倾向于恢复到变化前的状态。这反映了 DEER 平衡使用变化前后经验的策略：早期利用新经验进行快速政策调整，随后逐渐并有选择地重复使用旧经验以维持稳定性和长期记忆。

为了分析非平稳程度的影响，表 I 展示了环境变化后在 SAC+HalfCheetah 任务中 100 个时段的平均奖励，评估了不同程度的非平稳性：极端（200%偏移）、轻微（50%偏移）和稳定（0%偏移）。值得注意的是，标准（100%偏移）对应于图 2b 所示的结果。结果显示，在高度非平稳设置中，DEER 显著优于其他方法，比最佳基准在 200%偏移时高出约 22.53%的奖励，这突显了 DEER 对急剧环境变化的强大适应能力。在轻微非平稳条件下，DEER 仍保持微弱性能优势；而在稳定环境中，所有方法表现相当，表明当环境不变时，DEER 不会引入负面影响。

V. 结论

我们提出了一种名为 DEER 的混合度量优先级方法，该方法能够动态调整采样权重以解决非平稳环境中的样本效率问题。通过根据当前环境变化优先考虑有价值的样本，DEER 提高了样本效率。作为一种潜在的未

来方向，我们期待将我们的方法扩展到改进各种应用性能，如大型语言模型 [23]–[25]、多模态训练 [26], [27]、分布式机器学习 [28]–[35] 和自动驾驶 [36]–[38]。

参考文献

[1] Zongyuan Zhang, Tianyang Duan, Zheng Lin, Dong Huang, Zihan Fang, Zekai Sun, Ling Xiong, Hongbin Liang, Heming Cui, Yong Cui, et al., "Robust deep reinforcement learning in robotics via adaptive gradient-masked adversarial attacks," *arXiv preprint arXiv:2503.20844*, 2025.

[2] Tianyang Duan, Zongyuan Zhang, Zheng Lin, Yue Gao, Ling Xiong, Yong Cui, Hongbin Liang, Xianhao Chen, Heming Cui, and Dong Huang, "Rethinking adversarial attacks in reinforcement learning from policy distribution perspective," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[3] Chen Tang, Ben Abbatematteo, Jiaheng Hu, Rohan Chandra, Roberto Martín-Martín, and Peter Stone, "Deep reinforcement learning for robotics: A survey of real-world successes," in *Proc. AAAI Conf. Artif. Intell.*, 2025, vol. 39, pp. 28694–28698.

[4] Khimya Khetarpal, Matthew Riemer, Irina Rish, and Doina Precup, "Towards continual reinforcement learning: A review and perspectives," *J. Artif. Intell. Res.*, vol. 75, pp. 1401–1476, 2022.

[5] Denis Yarats, Amy Zhang, Ilya Kostrikov, Brandon Amos, Joelle Pineau, and Rob Fergus, "Improving sample efficiency in model-free reinforcement learning from images," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, pp. 10674–10681.

[6] Peiquan Sun, Wengang Zhou, and Houqiang Li, "Attentive experience replay," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, pp. 5900–5907.

[7] Hu Li, Xuezhong Qian, and Wei Song, "Prioritized experience replay based on dynamics priority," *Sci. Rep.*, vol. 14, no. 1, pp. 6014, 2024.

[8] Ali Rahimi-Kalahroudi, Janarthanan Rajendran, Ida Momennejad, Harm van Seijen, and Sarath Chandar, "Replay Buffer with Local Forgetting for Adapting to Local Environment Changes in Deep Model-Based Reinforcement Learning," in *Proc. Conf. Lifelong Learn. Agents*, 2023, pp. 21–42.

[9] Xi Cao, Huaiyu Wan, Youfang Lin, and Sheng Han, "High-value prioritized experience replay for off-policy reinforcement learning," in *Proc. IEEE Int. Conf. Tools Artif. Intell.*, 2019, pp. 1510–1514.

[10] Long-Ji Lin, "Self-improving reactive agents based on reinforcement learning, planning and teaching," *Mach. Learn.*, vol. 8, pp. 293–321, 1992.

[11] Marc Brittain, Josh Bertram, Xuxi Yang, and Peng Wei, "Prioritized sequence experience replay," *arXiv preprint arXiv:1905.12726*, 2019.

[12] Guido Novati and Petros Koumoutsakos, "Remember and forget for experience replay," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 4851–4860.

[13] Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, OpenAI Pieter Abbeel, and Wojciech Zaremba, "Hindsight experience replay," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[14] Yongle Luo, Yuxin Wang, Kun Dong, Qiang Zhang, Erkang Cheng, Zhiyong Sun, and Bo Song, "Relay Hindsight Experience Replay: Self-guided continual reinforcement learning for sequential object manipulation tasks with sparse rewards," *Neurocomputing*, vol. 557, pp. 126620, 2023.

[15] Erdi Sayar, Vladislav Vintaykin, Giovanni Iacca, and Alois Knoll, "Hindsight Experience Replay with Evolutionary Decision Trees for Curriculum Goal Generation," in *Proc. Int. Conf. Appl. Evol. Comput.*, 2024, pp. 3–18.

[16] Derek Li, Andrew Jacobsen, and Adam White, "Revisiting experience replay in non-stationary environments," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2021.

[17] Sindhu Padakandla, "A survey of reinforcement learning algorithms for dynamically varying environments," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–25, 2021.

[18] Jonathan Hillman and Toby Dylan Hocking, "Optimizing roc curves with a sort-based surrogate loss function for binary classification and changepoint detection," *arXiv preprint arXiv:2107.01285*, 2021.

[19] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver, "Prioritized Experience Replay," *arXiv preprint arXiv:1511.05952*, 2015.

[20] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1861–1870.

[21] Shangtong Zhang and Richard S Sutton, "A deeper look at experience replay," *arXiv preprint arXiv:1712.01275*, 2017.

[22] Baturay Saglam, Furkan B Mutlu, Dogan C Cicek, and Suleyman S Kozat, "Actor prioritized experience replay," *J. Artif. Intell. Res.*, vol. 78, pp. 639–672, 2023.

[23] Zheng Lin, Yuxin Zhang, Zhe Chen, Zihan Fang, Xianhao Chen, Praneeth Vepakomma, Wei Ni, Jun Luo, and Yue Gao, "HSplitLoRA: A Heterogeneous Split Parameter-Efficient Fine-Tuning Framework for Large Language Models," *arXiv preprint arXiv:2505.02795*, 2025.

[24] Zihan Fang, Zheng Lin, Zhe Chen, Xianhao Chen, Yue Gao, and Yuguang Fang, "Automated Federated Pipeline for Parameter-Efficient Fine-Tuning of Large Language Models," *arXiv preprint arXiv:2404.06448*, 2024.

[25] Zheng Lin, Xuanjie Hu, Yuxin Zhang, Zhe Chen, Zihan Fang, Xianhao Chen, Ang Li, Praneeth Vepakomma, and Yue Gao, "SplitLoRA: A Split Parameter-Efficient Fine-Tuning Framework for Large Language Models," *arXiv preprint arXiv:2407.00952*, 2024.

[26] Zihan Fang, Zheng Lin, Senkang Hu, Yihang Tao, Yiqin Deng, Xianhao Chen, and Yuguang Fang, "Dynamic uncertainty-aware multimodal fusion for outdoor health monitoring," *arXiv preprint arXiv:2508.09085*, 2025.

[27] Yongyang Tang, Zhe Chen, Ang Li, Tianyue Zheng, Zheng Lin, Jia Xu, Pin Lv, Zhe Sun, and Yue Gao, "MERIT: Multimodal Wearable Vital Sign Waveform Monitoring," *arXiv preprint arXiv:2410.00392*, 2024.

[28] Yuxin Zhang, Haoyu Chen, Zheng Lin, Zhe Chen, and Jin Zhao, "Fedac: An adaptive clustered federated learning framework for heterogeneous data," *arXiv preprint arXiv:2403.16460*, 2024.

[29] Zheng Lin, Yuxin Zhang, Zhe Chen, Zihan Fang, Cong Wu, Xianhao Chen, Yue Gao, and Jun Luo, "Leo-split: A semi-supervised split learning framework over leo satellite networks," *arXiv preprint arXiv:2501.01293*, 2025.

[30] Mingda Hu, Jingjing Zhang, Xiong Wang, Shengyun Liu, and Zheng Lin, "Accelerating Federated Learning with Model Segmentation for Edge Networks," *IEEE Trans. Green Commun. Netw.*, 2024.

[31] Yuxin Zhang, Haoyu Chen, Zheng Lin, Zhe Chen, and Jin Zhao, "LCFed: An Efficient Clustered Federated Learning Framework for Heterogeneous Data," *arXiv preprint arXiv:2501.01850*, 2025.

[32] Zehang Lin, Zheng Lin, Miao Yang, Jianhao Huang, Yuxin Zhang, Zihan Fang, Xia Du, Zhe Chen, Shunzhi Zhu, and Wei Ni, "Sl-acc: A communication-efficient split learning framework with adaptive channel-wise compression," *arXiv preprint arXiv:2508.12984*, 2025.

[33] Zheng Lin, Guanqiao Qu, Wei Wei, Xianhao Chen, and Kin K Leung, "Adaptsfl: Adaptive Split Federated Learning in Resource-Constrained Edge Networks," *IEEE Trans. Netw.*, 2024.

[34] Song Lyu, Zheng Lin, Guanqiao Qu, Xianhao Chen, Xiaoxia Huang, and Pan Li, "Optimal resource allocation for u-shaped parallel split learning," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 197–202.

[35] Zheng Lin, Guangyu Zhu, Yiqin Deng, Xianhao Chen, Yue Gao, Kaibin Huang, and Yuguang Fang, "Efficient Parallel Split Learning over Resource-Constrained Wireless Edge Networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 10, pp. 9224–9239, 2024.

[36] Zheng Lin, Lifeng Wang, Jie Ding, Yuedong Xu, and Bo Tan, "Tracking and transmission design in terahertz v2i networks," *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 3586–3598, 2022.

[37] Zihan Fang, Zheng Lin, Senkang Hu, Hangcheng Cao, Yiqin Deng, Xianhao Chen, and Yuguang Fang, "IC3M: In-Car Multimodal Multi-Object Monitoring for Abnormal Status of Both Driver and Passengers," *arXiv preprint arXiv:2410.02592*, 2024.

[38] Zheng Lin, Lifeng Wang, Jie Ding, Bo Tan, and Shi Jin, "Channel Power Gain Estimation for Terahertz Vehicle-to-Infrastructure Networks," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 155–159, 2022.