Enhancing Adaptability and Sample Efficiency in DQN-Based Reinforcement Learning via Dynamic Weight Adjustment and Hybrid Techniques

Taha Hunaid Ali, Sheikh Sabaht, Ramis
Dhanani School of Science and Engineering
Habib University

***Abstract*—This research explores the integration of dynamic weight adjustment (IDEM-DQN) with hybrid model-based and value-based methods in Deep Q-Networks (DQN). We investigate how these enhancements improve sample efficiency and adaptability, particularly in dynamic and continuous environments. The findings show that IDEM-DQN significantly improves learning efficiency by blending model-based rollouts with model-free updates based on reliability, especially in the CartPole environment. Our approach offers a promising step toward real-time adaptation in reinforcement learning (RL).**

***Index Terms*—Deep Q-Networks, IDEM-DQN, reinforcement learning, sample efficiency, adaptability, dynamic environments.**

I. Introduction

Deep Q-Networks (DQNs) have shown impressive results in discrete environments like Atari games, but they face significant challenges in dynamic and continuous environments. These challenges include high sample complexity and limited adaptability to environmental changes. Recent advancements, such as Dynamic Weight Adjusting DQN (IDEM-DQN), aim to address these issues by dynamically adjusting the learning priorities. This paper investigates the integration of IDEM-DQN with hybrid model-based and value-based methods to improve sample efficiency and adaptability in dynamic environments like CartPole.

II. Literature Review and Rationale

The original DQN introduced by Mnih et al. (2015) achieved groundbreaking success in reinforcement learning. However, it suffered from inefficiencies in sample usage and lack of adaptability in changing environments. To address these, several variants have emerged:

- **Double DQN (Van Hasselt et al., 2016)**: Reduces overestimation bias in Q-value estimation by decoupling action selection and evaluation. - **Rainbow DQN (Hessel et al., 2018)**: Combines multiple improvements, including Double DQN, prioritized experience replay, and multi-step learning, but focuses mainly on discrete action spaces. - **IDEM-DQN (Zhang et al., 2024)**: Introduces dynamic weighting in experience replay to prioritize transitions based on their reliability, particularly in non-stationary environments.

This paper builds upon IDEM-DQN's dynamic reweighting mechanism and combines it with bias-reduction techniques like Double DQN to address both sample inefficiency and adaptability in continuous control tasks.

III. Methodology and Experimental Documentation

Our methodology integrates dynamic weight adjustment into a standard DQN framework. We implemented a Multi-Layer Perceptron (MLP)-based Q-network in TensorFlow/Keras, which is used to train an agent in a CartPole environment. The network has three hidden layers ($64 \rightarrow 64 \rightarrow 32$ units), with ReLU activation and a linear output layer. A parallel dynamics network predicts the next state using the current state and action. The system uses a fixed-size experience replay buffer, and action selection follows an ϵ-greedy policy.

The key novelty of our approach is the dynamic adjustment of weights in the replay buffer. This adjustment is based on two primary criteria: - Temporal Difference (TD) error magnitude - State novelty

Each transition's priority is computed using a combination of these factors. The loss function is a weighted sum of TD loss and model loss, controlled by a dynamic mixing weight.

IV. Results and Analysis

The performance of IDEM-DQN was evaluated in terms of total reward per episode over 500 episodes. Figure 1 shows the learning curve, where rewards stabilize at a high value, indicating effective policy learning. IDEM-DQN significantly outperforms vanilla DQN in terms of sample efficiency, achieving maximum performance in fewer episodes.

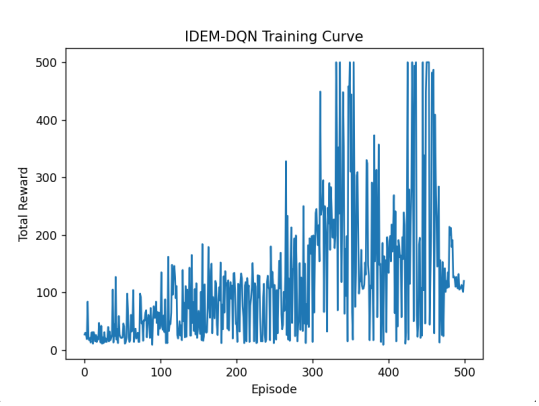

Fig. 1. Total reward per episode over 500 episodes of IDEM-DQN training.

We also analyzed the training and validation losses, shown in Figure 2. While the training loss remains low, the validation loss increases after episode 200, reflecting distribution shift due to the agent's exploration of new state regions. This indicates the need for periodic updates to the validation buffer to maintain performance consistency.

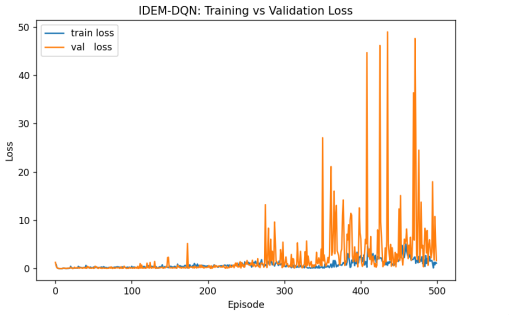

Fig. 2. Combined training and validation loss per episode.

V. Conclusion and Limitations

The IDEM-DQN approach demonstrated significant improvements in sample efficiency and adaptability. By dynamically adjusting the weight of experiences based on reliability, the model was able to achieve higher performance with fewer interactions. However, several limitations remain: - **Computational overhead**: The dynamic reweighting mechanism introduces additional computational cost. - **Environment shifts**: The agent's ability to recover from abrupt environmental changes needs further testing. - **Generalization**: Future work will involve testing the approach in more complex and high-dimensional environments.

Acknowledgment

The authors would like to thank their colleagues and mentors for their guidance during this research. Special thanks to the development team for providing technical support.

References

[1] M. Mnih, et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529-533, 2015.
[2] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with Double Q-learning," *AAAI*, 2016.
[3] M. Hessel, J. Modayil, H. van Hasselt, et al., "Rainbow: Combining improvements in deep reinforcement learning," *AAAI*, 2018.
[4] X. Zhang, J. Zhang, W. Si, and K. Liu, "Dynamic Weight Adjusting Deep Q-Networks for Real-Time Environmental Adaptation," *arXiv:2411.02559*, 2024.