Review-17

Mastering Atari with Discrete World Models

World models enable learning using imagined sequences via planning in latent space which directly results in efficient generalization and sample-efficient learning. However, scaling world models to high-dimensional task spaces such as Atari has been an open challenge. To this end, the work introduced DreamerV2 which is an improved version of DreamerV1. DreamerV2 is a Reinforcement Learning (RL) agent learns behaviors purely from latent predictions and builds a discrete model of the environment. The world model trains the policy to yield optimal behaviors in real environment. DreamerV2 is competitive to model-free RL algorithms such as Rainbow on 55 games from the Atari benchmark.

The original framework of DreamerV2 is adopted from DreamerV1 which consists of a world model and behavioral learning. The world model is learned separately from the RL policy which, on the other hand, learns behaviors using imagined sequences in the latent of discrete world model. In contrast to DreamerV1, the improved version makes use of categorical latents which utilize straight-through gradients in the world model. Furthermore, DreamerV2 combines the Reinforce estimator and dynamics gradients for training the actor policy using backpropagation. This leads to efficient, yet noisy behaviors in agents. Lastly, the agent is trained using an improved loss function which comprises of KL balancing constituting of separate scaling factors in prior and posterior cross-entropy terms in order to motivate accurate estimation of temporal prior. DreamerV2 is evaluated on a suite of 55 games from the Atari benchmark and compared with state-of-the-art model-free agents. Scores are normalized with respect to a professional games as well as the world record corresponding to each game. The work argues that clipped version of latter is a better method for comparison since it provides reasonable basis for evaluation in scenarios wherein the agent or player are extremely dominant.

DreamerV2 demonstrates improved performance across all comparison metrics and highlights the efficacy of learning world models. Suitability of categorical latents is further validated in comparison to prior gaussian latents in DreamerV1 using a short ablation study. Additionally, DreamerV2 utilizes equivalent compute as the model-free Rainbow. However, the experimental setup presents several limitations. Due to the high number of improvements, a comprehensive abaltion study may not be possible but DreamerV2 could be compared to DreamerV1 or its baselines from the original work. Additionally, DreamerV2 being a model-based approach, is not compared to state-of-the-art methods such as SimPLe which are oterhwise found to be sample efficient and competitive. Lastly, the summary of modifications provides various changes which were tried but did not work. However, the summary does not provide any insights into why do these changes fail to improve performance.