Multi-world Model in Continual Reinforcement Learning

Kevin Shen
The University of British Columbia
kevins00@student.ubc.ca

Abstract

World Models are made of generative networks that can predict future states of a single environment which it was trained on. This research proposes a Multi-world Model, a foundational model built from World Models for the field of continual reinforcement learning that is trained on many different environments, enabling it to generalize state sequence predictions even for unseen settings.

Introduction

Reinforcement learning (RL) is one of the major fields in machine learning where an agent receives varying rewards based on the actions it takes in a given environment with the objective of learning the optimal policy that would maximize its cumulative reward. Unlike traditional RL, where an agent typically learns a single task in a stationary environment, continual RL deals with non-stationary environments where the agent must adapt to new tasks over time while maintaining its performance on older ones. This is particularly challenging due to the issue of catastrophic forgetting, where neural networks tend to drastically overwrite old knowledge with new information.

One way to tackle this problem is using World Models (Ha and Schmidhuber 2018), which are generative networks that can predict future states of a given environment after learning its compressed spatial and temporal representations. World Models have shown much success in both RL and continual RL settings (Kessler et al. 2023). However, World Models are environment specific and very limited research has been done on exploring their generalizability to unseen environments. Given the uprising of foundational models, and by leveraging their functionality and architecture in conjunction with World Models, this research proposes a "Multi-world Model", which is a foundational model that can predict image sequences of both familiar and novel environments.

Background

There has been significant progress for foundational models' application in RL. Recently, DeepMind released a foundational model with promising results that lay the grounds for increasingly general and versatile RL agents which can perform well across big and complex domains (Bauer et al. 2023). The prospect of extending these advancements to continual RL offers a promising avenue for future exploration.

While the research done on foundational models in continual RL is limited, there are several research that use generative neural network models to train continual reinforcement learning agents with desirable results such as DreamerV3 (Hafner et al. 2023). DreamerV3 is a general and scalable algorithm based on World Models which trains agents using latent imagination. It demonstrated remarkable predictive prowess by accurately generating 45 frames into the future given only the starting 5 frames and the agent's action sequence. DreamerV3 is the first algorithm to collect diamonds in Minecraft from scratch without human data or curricula, a long-standing challenge in artificial intelligence.

This serves as an important basis that foundational models are also likely to do well, if not better with their state-of-the-art scalable and generative capabilities.

Prior Work by the Applicant

In my 2023 summer internship at Tencent, I co-authored a research paper accepted by Transactions on Machine Learning Research (TMLR) titled "Replay-enhanced Continual Reinforcement Learning" which used adaptive normalization and policy distillation to enhance the generality and stability of a continual RL model (Zhang et al. 2023). Through understanding more of this area of work and conducting multiple experiments, I found that replay-based methods are effective for averting catastrophic forgetting for continual RL problems albeit the challenges it faces such as poor generalizability when the tasks increase in dissimilarity. Hence,

it could be apt to leverage foundational models known for their expansive pre-training and generalization capabilities, propelling the field of continual RL towards more robust solutions.

Approach

Multi-world Model builds on World Models' architecture, which is made up of a vision component that uses a Variational Encoder (VAE) model and a memory component that uses a Recurrent Neural Network (RNN) model. Since the Multi-world Model will be trained on multiple different environments, encompassing a much larger and more diverse set of data as compared to World Models which are trained on a single environment per model, its model architecture will need to be expanded and altered accordingly. Besides adding more layers to the VAE and RNN models, architectures commonly used in foundational models can also be employed.

In particular, a potential upgrade would be to replace the RNN model with Transformers (Vaswani et al. 2017) to leverage its parallelizability, scalability and effectiveness in handling long-range dependencies. It could be critical to investigate how the attention mechanism (Vaswani et al. 2017) can be applied for such long image sequence prediction tasks since it is common for RL agents to only receive a reward many steps after a key action was taken. Inspiration can be drawn from a similar field of work which has achieved such success as seen in Video Prediction by Efficient Transformers (Ye and Bilodeau 2023). Nonetheless, the task at hand presents an additional layer of complexity compared to video prediction since the following states of an environment changes based on the different actions taken by the agent. As an additional measure, the robustness of the model can be enhanced by aligning the image representation with text representations of environment states (Schwartz et al. 2019) generated from large language models such as GPT-4 and vision language models such as DALL-E.

Evaluation

Results will be evaluated across multiple continual RL benchmarks such as Continual World (Wołczyk et al. 2021). Performance curves on task sequences will be plotted to visualize the learning trajectory. If several different model architectures are proposed, they will first be compared internally before the best one is chosen to be compared with other state-of-the-art continual RL algorithms such as clonEx-SAC (Wołczyk et al. 2022) and PackNet (Mallya and Lazebnik 2018). Metrics for the performance, amount of forgetting and positive transfer will be defined so that more in-depth analysis can be conducted. Performance will also be tested against continual RL agents that are trained using World Models on single environments to check if the additional generalizability from the Multi-world Model produced better performing agents.

Discussion

Given the rapid advancement of foundational models as well as the emerging paradigm of continual RL, the integration of these two domains will likely be increasingly feasible and beneficial. The realization of the Multi-world Model could herald a breakthrough in the field of continual RL and make significant progress towards more generalizable AI systems which can understand and adapt to multifaceted real-world scenarios. It would mean the development of agents which are efficient hypothesis-driven learners capable of thinking before acting by simulating outcomes even in unseen settings, possibly matching or surpassing human cognition. This could yield numerous societal benefits such as robotics in healthcare. Another practical application lies in autonomous systems utilizing computer vision, like autonomous vehicles. Accurate image sequence predictions could substantially enhance the safety and robustness of these systems, allowing anticipation and reaction to unpredictable or hazardous situations.

Conclusion

This research outlined the potential for the Multi-world Model to significantly enhance the field of continual RL. By leveraging the generative and representational abilities of foundational models, this research aims to develop continual RL agents that are more resilient to catastrophic forgetting, exhibit improved generalizability across tasks, and showcase increased adaptability within non-stationary environments. Success in this endeavor could produce AI agents capable of lifelong learning and adaptation, with significant benefits across various sectors of society such as healthcare and autonomous systems.

References

Ha, D.; and Schmidhuber, J. 2018. World models. *arXiv preprint arXiv:1803.10122*.

Hafner, D.; Pasukonis, J.; Ba, J.; and Lillicrap, T. 2023. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*.

Kessler, S.; Miło´s, P.; Parker-Holder, J.; and Roberts, S. J. 2022. The surprising effectiveness of latent world models for continual reinforcement learning. *arXiv preprint arXiv:2211.15944*.

Mallya, A.; and Lazebnik, S. 2018. Packnet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 7765–7773.

Schwartz, E.; Tennenholtz, G.; Tessler, C.; and Mannor, S. 2019. Language is power: Representing states using natural language in reinforcement learning. *arXiv preprint arXiv:1910.02789*.

Team, A. A.; Bauer, J.; Baumli, K.; Baveja, S.; Behbahani, F.; Bhoopchand, A.; Bradley-Schmieg, N.; Chang, M.; Clay, N.; Collister, A.; et al. 2023. Human-timescale adaptation in an open-ended task space. *arXiv preprint arXiv:2301.07608*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wolczyk, M.; Zajac, M.; Pascanu, R.; Kuciński, Ł.; and Miłoś, P. 2022. Disentangling transfer in continual reinforcement learning. Advances in Neural Information Processing Systems, 35: 6304–6317.

Wołczyk, M.; Zajac, M.; Pascanu, R.; Kuciński, Ł.; and Miłoś, P. 2021. Continual world: A robotic benchmark for continual reinforcement learning. Advances in Neural Information Processing Systems, 34: 28496–28510.

Ye, X.; and Bilodeau, G.-A. 2023. Video prediction by efficient transformers. *Image and Vision Computing*, 130: 104612.

Zhang, T.; Shen, K. Z.; Lin, Z.; Yuan, B.; Wang, X.; Li, X.; and Ye, D. 2023. Replay-enhanced Continual Reinforcement Learning. *arXiv preprint arXiv:2311.11557*.