A Sequence Prediction Perspective for Intuitive Physics Reasoning and Interaction

Du Jiang
School of Intelligence Science and Technology
Peking University
dujiang@pku.edu.cn

Abstract

In this essay, we discuss a sequence prediction-based approach to empower deep-learning systems with intuitive physics reasoning capabilities. Our approach involves pre-training a physics-informed model to predict future states based on real-world observations. This model is then fine-tuned using reinforcement learning (RL) techniques, aligning it with RL's success in training large language models. Thus it is possible for our systems to better understand complex, dynamic environments. This approach holds potential across diverse domains, including robotics, simulations, and autonomous systems.

1 Introduction

Intuitive physics reasoning has long been a fundamental challenge in the pursuit of creating machines that can understand and interact with the physical world as adeptly as humans[1]. Developing machine learning systems capable of predicting and comprehending the behavior of objects and the environment based on underlying physical principles is essential for a wide range of applications, including robotics, autonomous vehicles, simulations, and decision-making. In this context, this essay discusses a sequence prediction-based approach that leverages the power of pre-training with physics-informed models followed by fine-tuning using reinforcement learning (RL) techniques.

One of the most important controversial topics in intuitive physics is nature versus nurture. Actually, evolution itself, which has inspired many deep learning works[4], can be deemed as a powerful algorithm that provides a good base for our lifelong learning. From this perspective, when training a machine agent, we may consider some two-stage approaches: a "nature" stage provides an algorithm and a "nurture" stage applies the algorithm based on real-world observations, or the former provides a well-initialized model for following learning or fine-tuning. Inspired by GPT-4.0[2] and RLHF[3], we consider creating machine learning systems that can reason about and interact with the physical world in a manner that combines the robustness of physics knowledge with the adaptability of RL, by pre-training a model to predict future states in real-world observations, and then fine-tuning it using RL. This essay briefly discusses this process, aiming to provide some insights for enhancing safety, efficiency, and adaptability in real-world applications.

2 Methodology

Pre-training a physics-informed model to predict the next time state based on observations of the real world and then using it as an intuitive physics engine for RL tasks, is a reasonable and promising approach. It combines elements of predictive modeling, physics simulation, and reinforcement learning (RL) to build a robust machine learning system for understanding and interacting with the physical world.

2.1 Pretraining with Physics-Informed Models

Pre-training a foundation model with knowledge of physics principles can provide a solid foundation for understanding how objects and the physical world behave, which can be especially helpful in RL tasks where the agent must interact with the environment effectively. One possible implementation is training the foundation model on sequence prediction or recovery tasks for discretized physical dynamics. We can use a neural network-based perception model to encode the observations into a tensor sequence. To obtain an effective intuitive physics engine, we may include the states and actions of the agent as a part of the sequence.

2.1.1 RL-Based Fine-Tuning

By including agents' information in the sequence and treating interaction with the environment as a sequence prediction task, we can frame the problem in a way that aligns with RL-based GPT training paradigms. This allows us to leverage RL techniques for learning and optimizing the agent's behavior over time. Especially for tasks in which real-world observations or simulation results are hard to collect, the framework allows us to utilize a reward model that enables the usage of some algorithms with data efficiency, e.g., Proximal Policy Optimization (PPO)[5]. In addition, by incorporating the violation of expectation paradigm into the reward model, we can encourage agents to learn and predict physical interactions in a manner consistent with human intuition.

3 Benefits

Interaction Efficiency Since we construct an intuitive physics engine that approximates real-world dynamics using a parametric neural network, the efficiency of interaction is directly ensured by the efficiency of model inference and meets our expectations.

Transfer Learning Benefits Leveraging pre-trained models for physics understanding can offer transfer learning benefits. One can apply the knowledge learned from one scenario to new, related scenarios, making the system more adaptable and efficient in learning.

Sample Efficiency RL training paradigms, such as PPO and Reinforcement Learning from Human Feedback (RLHF)[3], have shown promise in training large language models with limited interaction data. This can be advantageous in scenarios where collecting real-world interaction data can be expensive or dangerous.

Complex Environment and Various Signals The approach allows the agent to learn from interactions and adapt its behavior over time and thus is well-suited for complex, dynamic environments where the outcomes of actions are not always straightforward to predict. Also, RL-based optimization allows non-differentiable signals and thus meets the requirement of diverse tasks.

4 Potential Problems

4.1 Optimization Problems

Challenges in Training Be aware that training RL models, especially large ones, can be computationally intensive and may require substantial resources. Proper reward design and exploration strategies are critical to successful RL training.

Evaluation and Fine-Tuning After pre-training and RL training, it's important to evaluate the performance of our model and fine-tune it as needed. This may involve addressing biases, refining the reward structure, and ensuring safety and robustness.

4.2 Other Considerations for Real-World Applications

Be mindful of ethical considerations, particularly in real-world applications. Ensuring that the AI system's behavior aligns with ethical principles and adheres to safety standards is essential. Consider how to make the system's reasoning and predictions interpretable, as understanding the model's behavior is crucial for trust, debugging, and accountability.

References

[1] James R. Kubricht, Keith J. Holyoak, and Hongjing Lu. Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, 21(10):749–759, 2017. ISSN 1364-6613. doi: https://doi.org/10.1016/j.tics.2017.06.002. URL https://www.sciencedirect.com/science/article/pii/S1364661317301262. 1

[2] OpenAI. Gpt-4 technical report, 2023. 1

[3] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022. 1, 2

[4] Evgenia Papavasileiou, Jan Cornelis, and Bart Jansen. A Systematic Literature Review of the Successors of "NeuroEvolution of Augmenting Topologies". *Evolutionary Computation*, 29(1): 1–73, 03 2021. ISSN 1063-6560. doi: 10.1162/evco_a_00282. URL https://doi.org/10.1162/evco_a_00282. 1

[5] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 2017. URL http://arxiv.org/abs/1707.06347. 2