CSE6243: Advanced Machine Learning **Fall 2023**

Lecture 15: Foundation Models for Decision Making

Lecturer: Bo Dai *Scribes: Rishi Banerjee*

Note: *LaTeX template courtesy of UC Berkeley EECS Department.*

15.1 New Content

Machine learning has made significant advances in text-to-image/video and language generation tasks, which have been underlied by foundation models.

Foundation models are large pre-trained model used in machine learning and natural language processing. They form the foundation for various tasks and are trained on extensive internet search data, which enables them to be useful for a large amount of knowledge and language patterns. Prominent examples are the GPT series by OpenAI and Google's BERT model. (This explanation was even produced by ChatGPT).

The creation and use of foundation models is not sufficient for all tasks due to the following criteria. The first problem is that there may not exist enough data for build a sufficient model. This is a problem for scenarios like:

- Scientific discoveries
- Rare events for safety

Foundation models also run into issues where you would want a model to produce outcomes that are better than data they are trained on. This is a problem for scenarios like:

- Building optimal robot policies from failed robot executions
- Generating faster programs from available code data

The solutions of the problems are as follows: in order to solve problems where there are not enough data, the solution is to collect more data, while for the scenarios where the model must produce better results than the data, the solution is to optimize the generated action. This optimization can be done through leveraging past work in reinforcement learning, planning, search, control and optimization.

15.1.1 Challenges of Sequential Decision-Making

If there was not enough data, an important aspect of collecting data is sample efficiency. For example, in typical RL systems in order to solve a system like Brick Breaker, it would take 38 hours for the system to learn how to play the game, while it would take a human only minutes.

Along with this, in order to understand how to optimize actions, one other problem is how to generalize the results of a foundation models for a specific game to work on another game. Another issue is that for general decision-making problems, these models currently lack the broad knowledge like that of physics, vision, and language.

15.1.2 How Foundations Models Acquire Broad Knowledge

How can we leverage these capabilities in order to make decisions for problems?

- Representation learning can extract information even if it is containing multiple failed executions
- Reasoning would leverage some planning and search algorithms in order to take actions
- Modeling generating intermediate steps would build reasoning capability
- How can we optimize the internet data read in order to extract knowledge

15.1.3 Representation Learning

Representation learning is learning representations for images or language.

Examples:

- Contrastive Learning - SimCLR, CLIP
- Denoising AutoEncoders - BERT, MAE

In Learning from Expert Demonstrations, the goal is to find state to action mappings that allow for the agent to operate successfully. When performing this RL task, it is better to perform representation over the states being learned through pretraining. During pretraining, we would learn the representation that maps the image or observation space to the latent representation, then would learn the policy mapping from latent states to actions. The benefit of this is that the representations have lower dimensionality, which leads to fewer expert demonstrations necessary as the hypothesis space is smaller. Another benefit of this method is that it is provably bounded performance difference between expert policy and the learned policies, as shown in slide 16 of the presentation.

This method has shown benefits for continuous control problems. These tasks consisted of suboptimal behavior for continuous control tasks. Pretraining with representation learning outperformed the tasks for imitation learning, offline and online RL tasks. As a result, it is possible to use suboptimal data for representation learning. As a result, contrastive learning and denoising autoencoding can be used for approximate dynamics models.

15.1.4 Reasoning

Case Study: How do you teach an LLM to do math?
Simply passing in input-output pairs to the model will not teach the model the mathematical rules since it would only be able to perform rote memorization. In order to establish understanding, not memorization, it is necessary to pass in intermediate reasoning steps in order to understand the procedure.

Input-output pairs for actions is the current method of doing most deep RL tasks. As a result, RL can utilize search algorithms in order to improve upon current methods. As a result, these procedural cloning methods outperform the typical behavioral cloning. In this method, it is necessary to predict the intermediate steps before generating the final actions. Intermediate steps are in machine-native language of the agent. By teaching intermediate computations in order to have agent able to emulate planning and search procedures. Note that it is not necessary to train the data in natural language, can repeat in machine-native language. This is analogous to the reasoning series found in large language models.

15.1.5 Using Internet Data

Using internet web crawl protocols and the entire Wikipedia corpus, it was possible to generate a human-like chatbot. Along with this data-source, there exists many other data sources that are not fully labelled. For example, there are panarama scans, videos with corresponding subtitles, simulated robotic manipulation, and static images, which can have action-based natural language descriptions generated for them. As a result, we can ask the question: Can we create a world-like simulator from the multimodal data of the Internet?

As a result, the speaker presents UniSim, which was a universal simulator for text-to-video generation. This leverages the fact that simulated data has a lot of controlled manipulation and actions, with low dimensional action space. Please see the link for UniSim for more information. The possibilities are to increase sample efficiency and generalization and UniSim has been shown to help perform large-scale online RL tasks. It has also be shown that real-world agents could learn from a zero-shot transfer of simulations. As a result, the takeaway is that rich interactive data could exist to improve decision making. VLMs can also be repurposed in order to derive actions by parameterizing different components of MDPs.

15.1.6 Future Steps

- Foundations models can be used for control and embodiment
- Process can also be used in order for other processes such as materials discovery
- We can apply different models in order to train single models on multi-task environment to take advantage of synergies between representations between tasks
- Models can be improved by algorithms like RLHF and interactive data
- New data will be produced and generated by deploying models and improving algorithms