Thinking Far Ahead: Long-horizon planning using deep model-based reinforcement learning

Piotr Kozakowski

Reinforcement learning (RL) is a branch of Artificial Intelligence concerned with training independent agents via interaction. Such agents operate in a given environment, gaining new experience and learning from it to improve their behavior. A simpler term might be "learning by trial and error". Importantly, reinforcement learning agents automatically discover useful representations to reason about the world.

A wide range of problems can be framed in terms of agents interacting with environments, promising wide-scale applicability of reinforcement learning methods in real-world problems. New research in this area constantly pushes the boundary between what problems computers excel in, and in which problems humans are still better. Spectacular examples include robotic control, complex strategic games, or modern computer games. Recent research has even studied the application of RL systems to controlling economy, or reducing impact of pandemies.

However, RL struggles with tasks spanning long time horizons, which are very common in the real world. Solving such problems with RL is possible, but requires tremendous amounts of human effort and computational resources. Examples include OpenAI Five, a system that has defeated the world champions in Dota 2, or AlphaStar, a system besting world-class players in StarCraft II.

Another way to solve such is *planning.* Planning methods do not learn, but instead use clever algorithms to find suitable strategies, possibly spanning long time horizons. However, planning systems require good representations - we need to clearly define what is a state of the environment and what are the possible choices for the agent. Planning with a bad representation can be very computationally expensive, and finding good representations requires a lot of work.

So on one hand we have RL methods, which learn their own representations but struggle with long horizons, and on the other hand we have planning algorithms, which handle long horizons well, but need good representations to be specified from the outside. Fortunately, there are ways to combine the two approaches, combining the benefits of both.

Model-based RL assumes access to a *model* of the environment, that can be used to predict the outcomes of actions and formulate plans. Model-based RL systems can combine planning and learning, achieving the best of both worlds. AlphaZero is an example of a model-based RL system, achieving superhuman performance in a variety of complex strategical games. AlphaZero trains neural networks to predict the promising moves, and later uses them to guide the planning algorithm, so it allocates its computational resources much more effectively.

In this project, we are going to investigate various methods of combining planning and learning to obtain more effective intelligent agents. We will explore the synergies between the two approaches, and design systems that (a) plan in such a way to make learning more efficient, (b) learn in such a way to make planning more efficient, and (c) learn their own models of the environment to achieve full autonomy.

We expect that the methods we are going to develop in this project will allow us to build better AI agents, reasoning over longer time periods, learning faster and solving more complex tasks.