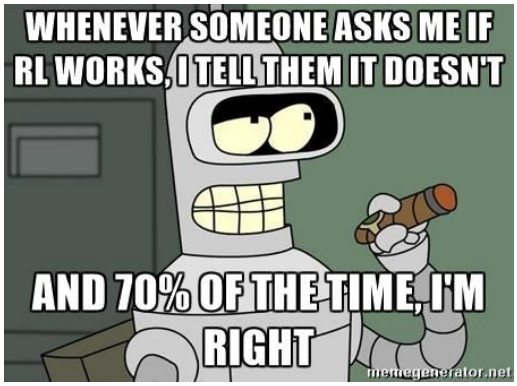
Do Androids Dream of Great Success?

Ni Lao

Machines are still bad at learning from a few examples

There has been a surge of recent interest in applying reinforcement learning (RL) to various domains including program synthesis, dialogue generation, deep architecture search, Atari games and continuous control. However, current training procedures rely on huge number of try and errors, and the performance is still hit and miss. In contrast, human can reliably learn from only a few examples.

“Whenever someone asks me if reinforcement learning can solve their problem, I tell them it can’t. I think this is right at least 70% of the time.” said Alex Irpan from Google Brain Robotics in his blog post. “AlphaGo... was an unambiguous win for deep RL, and that doesn’t happen very often.” “it’s disappointing that deep RL is still orders of magnitude above a practical level of sample efficiency.”. For example, the best learning algorithm (DeepMind RainbowDQN) “passes median human performance on 57 Atari games at about 18 million frames of game play, while most humans can pick up a game within a few minutes.”



There is also this “classic exploration-exploitation problem that has dogged reinforcement learning since time immemorial. Your data comes from your current policy. If your current policy explores too much you get junk data and learn nothing. Exploit too much and you burn-in behaviors that aren’t optimal.”

Mammals learn from past experiences by dreaming

In early 2000s scientists from MIT already discovered that animals have complex dreams and are able to retain and recall long sequences of events while they are asleep. While the Rapid Eye Movement (REM) replay lasts several minutes and playback experiences in approximately real time, the slow wave sleep (SWS) replay intermittently in brief bursts, each compressing the behavioral sequence in time by approximately 20-fold. Related work in humans also suggests

that the amount of REM and SWS sleep is correlated with subsequent enhancement of performance on learned tasks.

One fundamental question is that since animals accumulate vast amount of experiences during their days, which piece of experience should they dream about to achieve best learning?

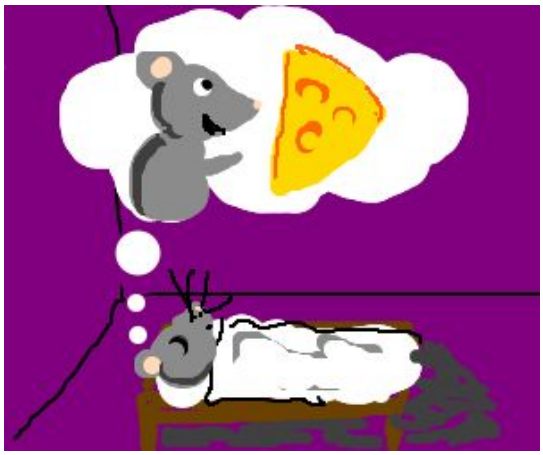


Image courtesy of Kote on Drawception.com, 2012

Recent studies in sleep and dreaming have indicated that by consolidating memory traces with high emotional/motivational value "sleep and dreaming may offer a neurobehavioral substrate for the offline reprocessing of emotions, associative learning, and exploratory behaviors, resulting in improved memory organization, waking emotion regulation, social skills, and creativity." For example a recent experiment at University College London found that when rats rest, their brains simulate journeys to a desired outcome such as a tasty treat in its maze environment. The scientists concluded that “such goal-biased preplay may support preparation for future experiences in novel environments.”

Reinforcement learning with a memory of past experiences

The question of how human and other animals can learn so efficiently and reliably give rise to our framework of reinforcement learning with a memory of past experiences or ‘experience replay’. The general idea is that learning (specifically the optimization of a policy-encoding deep neural model) is a lot more effective if a memory of interesting experiences can be incorporated. Because high-reward experiences remain in the memory, they will not be forgotten, and maybe repeatedly revisit during training if necessary.

Our previous work linearly interpolates the maximum likelihood (ML) training objective and reinforcement learning (RL) objective. While the RL objective represents the expected performance of an agent, the ML objective (measuring how similar the agents behavior is to an oracle behavior) both speedups and stabilizes training. When applied to the task of question answering from a large knowledge graph (WebQuestionsSP), it achieves the state-of-the-art

performance with weak supervision. Despite of its effectiveness, this interpolation strategy introduces bias due to the ML objective, and also depends on the assumption of a way to identify the best performing experience for each task (a question to be answered).

In our new paper, we further develop the experience replay technology by removing the ML bias, and the assumption of a single most effective experience. Here we still consider the task of weakly supervised program synthesis from natural language, but on a more challenging dataset WikiTableQuestions. This dataset involves more complex programs, which compute answers from Wikipedia tables, and therefore need to deal with a much larger program space with falsely rewarded program hypothesis. Our proposed new approach optimizes the agent's policy by separately considering the experiences inside the memory and those freshly generated according to the agent's current policy. When the memory is big, one experience is sampled from it for training each time. We showed that as long as this sampling is done according to both the agent's current model probability and the experiences' reward values, then the overall training objective is still RL and avoids introducing bias from the ML objective.

On the challenging WikiTableQuestions benchmark we achieve an accuracy of 46.2% on the test set, significantly outperforming the previous state-of-the-art of 43.7%. Interestingly, on the Salesforce WikiSQL benchmark, we also achieve an accuracy of 70.9% without the supervision of gold programs, outperforming several strong fully supervised baselines.

Image courtesy of Robot Dreams (The Robot Series), Byron Preiss Visual Publications, 2012

References

- Do Androids Dream of Electric Sheep? by Philip K. Dick. 1968, Doubleday
- Temporally Structured Replay of Awake Hippocampal Ensemble Activity during Rapid Eye Movement Sleep, Kenway Louie, Matthew A.Wilson, Neuron, 2001
- Memory of Sequential Experience in the Hippocampus during Slow Wave Sleep. Albert K.Lee, Matthew A.Wilson, Neuron, 2002
- Rats dream about their tasks during slow wave sleep, MIT News, 2002

- Sleep and dreaming are for important matters, L. Perogamvros, T. T. Dang-Vu, M. Desseilles, and S. Schwartz, Front Psychol. 2013
- Do Rats Dream of a Journey to a Brighter Future?, Neuroscience News, June 26, 2015
- “Hippocampal place cells construct reward related sequences through unexplored space” by H Freyja Ólafsdóttir, Caswell Barry, Aman B Saleem, Demis Hassabis, and Hugo J Spiers, in eLife, June 26 2015
- Prefrontal cortex as a meta-reinforcement learning system, Jane X Wang, Zeb Kurth-Nelson, Dharshan Kumaran, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Demis Hassabis, Matthew Botvinick, 2018, Nature Neuroscience
- “Deep RL is popular because it’s the only area in ML where it’s socially acceptable to train on the test set.”, Jacob Andreas from Berkeley