Paper summary: Dyna, ax integrated architecture for learning, planning and reacting

Marko Guberina 19970821-1256

March 20, 2022

1 Idea in few sentances

So you have you're model-free learner and you're happy with how it works. Collecting samples is hard. How about you use those samples to learn the dynamics and generate your own samples? Then you'd have more samples and thus a better agent which is also able to learn safely by imagining stuff in its head.

This paper concretizes this idea, outlines it in pseudocode and discusses it's benefits and drawbacks.

2 Explanation of the central concept

3 Methodology

4 Initial rambly notes

This papper is old, early 90s old. Here R. Sutton lays down the theoretical foundation for learning a model and a policy in tandem.

4.1 Abstract

Just says he introduces Dyna without going into any detail. Everything is completely theoretical here.

4.2 Introduction

Dyna architecture attempts to integrate (I'm guessing the correct notation as Sutton doesn't provide any):

- trial-and-error learning of an optimal *reactive policy* : $\pi : \boldsymbol{s}_t \rightarrow \boldsymbol{a}_t$
- learning domain knowledge in the form of an *action model*: $p : (\boldsymbol{s}_t, \boldsymbol{a}_t) \rightarrow \boldsymbol{s}_{t+1}$

- planning: finding the optimal reactive policy given domain knowledge π_p
- reactive execution: there is no planning, just reactions to percieved states

Algorithm pseudocode Repeat forever:

1. observe the world's state and reactively choose an action based on it
2. observe resulting reward and state
3. apply RL to this experience
4. update action model based on this experience
5. repeat K times:
   (a) choose a hypothetical world state and action
   (b) predict resultant reward and new state using the action model
   (c) apply RL to this hypothetical experience

4.3 Method

There is no method, only the theoretical architecture is introduced. However, it's potential implementations, advantages and shortcomings are disscused.

4.4 Other stuff