Paper Review: Learning Latent Dynamics for Planning from Pixels

Jordan Coblin
University of Alberta
coblin@ualberta.ca

1 Answers to the 5C's Questions

Category: New technique and prototype of this technique.

Context: Model-based reinforcement learning, variational inference, continuous control problem with partial observability, recurrent state space models (Karl et al., 2016; Doerr et al., 2018).

Correctness: Assumptions around latent space being preferable to image generation, dynamics model benefiting from stochastic and deterministic components

Contributions: Performing planning using a learned dynamics model with a compact latent space. Implementing a dynamics model with both deterministic and stochastic components. Generalization of the variational objective to support multi-step predictions.

Clarity: Abstract and introduction are very clear about the purpose of the work. Sufficient background to motivate the approach, and clearly reported experimental results.

2 Summary of the Paper

This paper introduces a novel model-based reinforcement learning agent called PlaNet that learns a latent dynamics model which can be used for planning in high-dimensional environments. A latent dynamics model generates trajectories in latent space instead of the input space (i.e. in this case, pixel-based images), allowing for fast computation during planning. The authors propose a dynamics model that has both deterministic and stochastic components, in order to more accurately predict transitions multiple time-steps into the future. This hybrid model, which they name a ”recurrent state space model” (RSSM) is shown to exceed performance of purely stochastic or purely deterministic models across several control tasks.

This dynamics model learned over real experience is then used to train an agent solely via planning, using the ”cross entropy method” to infer a distribution over optimal action sequences. Basically, this involves sampling action sequences from a Gaussian distribution, evaluating each sequence using the dynamics model, and then adjusting the Gaussian distribution parameters based on the top K sequences. In order to enhance the dynamics model (i.e. provide data for unexplored parts of the environment), the agent also injects trajectories that are sampled from its dynamics model into the dataset used to train the model.

Lastly, the paper introduces a regularization technique for the dynamics model objective function that they call ”latent overshooting”. This technique is meant to extend the standard variational lower bound to support better multi-step predictions. Latent overshooting was found to improve performance for other dynamics models, but not for RSSM.

3 Follow-up Questions

- It seems strange to refine the dynamics model by training on data sampled from itself. Couldn't this result in compounding drift from the real environment?

- Why use MPC over some other planning algorithm? What are other alternatives?

- How does the RSSM model compare to other approaches at dynamics model learning/system identification? The paper compares their method with a purely deterministic and purely stochastic model, but not with SOTA dynamics models.

- What are some potential explanations for why latent overshooting didn't end up improving RSSM?

4 Terms I Did Not Understand

- Model predictive control (MPC)
- Non-linear Kalman filter
- The Cross-Entropy Method
- Bayesian filtering

5 Relevant References

Lars Buesing, Theophane Weber, Sebastien Racaniere, S. M. Ali Eslami, Danilo Rezende, David P. Reichert, Fabio Viola, Frederic Besse, Karol Gregor, Demis Hassabis, and Daan Wierstra. 2018. Learning and querying fast generative models for reinforcement learning.

Andreas Doerr, Christian Daniel, Martin Schiegg, Duy Nguyen-Tuong, Stefan Schaal, Marc Toussaint, and Sebastian Trimpe. 2018. Probabilistic recurrent state-space models.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. 2019. Dream to control: Learning behaviors by latent imagination.

Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. 2020. Mastering atari with discrete world models.

Lukasz Kaiser, Mohammad Babaeizadeh, Piotr Milos, Blazej Osinski, Roy H Campbell, Konrad Czechowski, Dumitru Erhan, Chelsea Finn, Piotr Kozakowski, Sergey Levine, Afroz Mohiuddin, Ryan Sepassi, George Tucker, and Henryk Michalewski. 2019. Model-based reinforcement learning for atari.

Maximilian Karl, Maximilian Soelch, Justin Bayer, and Patrick van der Smagt. 2016. Deep variational bayes filters: Unsupervised learning of state space models from raw data.

Alex X. Lee, Anusha Nagabandi, Pieter Abbeel, and Sergey Levine. 2020. Stochastic latent actor-critic: Deep reinforcement learning with a latent variable model. 33:741–752.