Aalto University

Investigating Bayesian Neural Network Dynamics Models for Model-Based Reinforcement Learning

Aidan Scannell Arno Solin Joni Pajarinen

FCAI Finnish Center for Artificial Intelligence

TL;DR

Reinforcement learning (RL) is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning. Instead of leveraging labelled/unlabelled corpora of data, RL's focus is on finding a balance between exploration (gathering new experience) and exploitation (of current knowledge). Intuition says that capturing uncertainty related to what a model knows about and does not know would play an important role in this, but so far this has been challenging in complicated deep models. We explore lightweight approximate inference methods in these models.

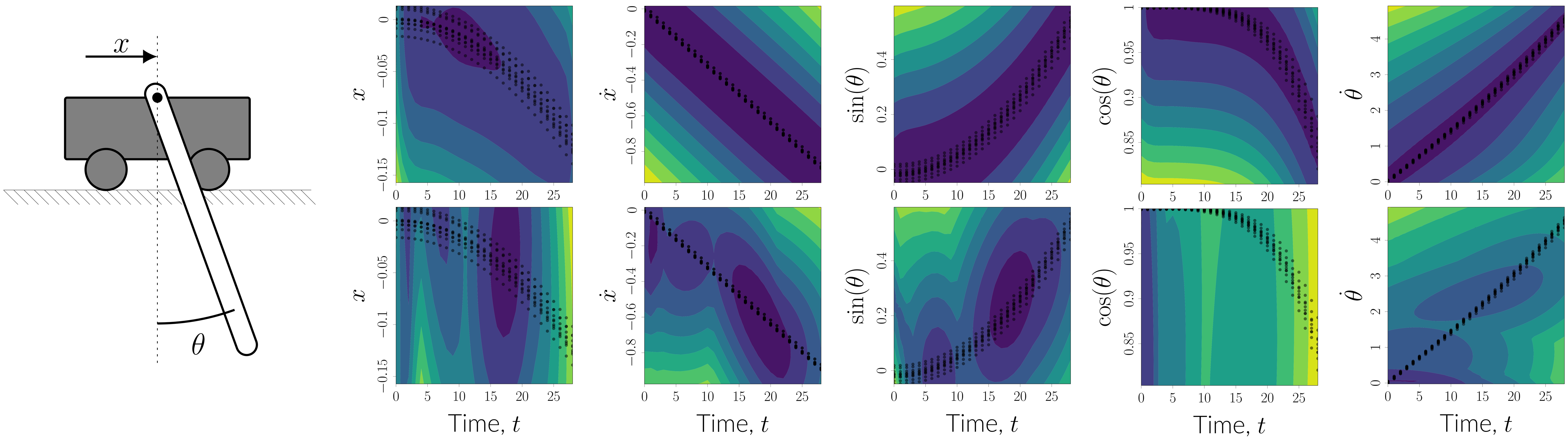

Figure 1: The inverted pendulum model (left) that is simple-enough to be captured by the baseline (top). We compare uncertainty estimates (un certain, scaled individually per state variable) for each state variable by a Gaussian process baseline (top) and a deep feedforward neural network (bottom) where we use the Laplace approximation for efficient inference. The dots visualize the progression of the trial trajectories through time; the the plots are cross sections of the 5D state-space moving along the median of the state variables.

Models & Methods

Model-based reinforcement learning

- Model-based reinforcement learning (MBRL) algorithms are more sample-efficient than their model-free counterparts
- This means that they are better fit to real-world use-cases where training is not necessarily done in a simulator
- However, they often perform poorly, due to their decision-making strategy exploiting inaccuracies in the learned dynamics model
- These inaccuracies arise because the dynamics are learned from state transitions which only represent a small subset of the environment
- As such, the dynamics model cannot be confident making predictions far away from these state transitions

For the model to know when it does not know

- This is known as epistemic uncertainty and in the limit of infinite data it is reduced (*i.e.* a data set containing all possible state transitions)
- If the dynamics model *knows what it does not know*, then this information can be used to prevent the decision-making strategy from exploiting the model's inaccuracies
- There exists principled methods to capture model uncertainty, but they do not necessarily scale well to complicated deep models in real-world RL
- We compare Gaussian process regression to a deep feedforward neural network where we use a Laplace approximation for approximate inference—a simple, yet efficient inference method for Bayesian neural networks

Experiments

- We study the standard toy problem where a cart has a pendulum attached to it and the goal is to swing up and keep the pendulum in balance (see Figure 1 left)
- The state space includes cart position x, speed $\dot{x}$, the sine and cosine of the pendulum angle, and the angular velocity $\dot{\theta}$
- As a baseline we use a Gaussian process (GP) model with a stationary prior covariance function, for which the resulting model becomes uncertain outside the observed trajectories
- We compare the GP model to a feedforward neural network with 4 hidden layers of width 10, and use a last-layer Laplace approximation around the MAP estimate of the network weights
- We let the model explore and try to visualize the quantified uncertainty in the state-space
- The resulting uncertainty plots show that even the non-stationary neural network model captures well the notion of being uncertain further away from explored regions of the state-space

Discussion

- We are interested in learning dynamics models using Bayesian neural networks and comparing them to other methods commonly used in model-based RL, such as ensembles
- We seek to compare different approximate inference techniques (*e.g.*, Laplace approximation, MC dropout), as well as ensemble methods, to understand why they either succeed or fail in different environments
- These methods have potential benefits in improving upon data-efficiency, practicality, and real-world use of model-based RL

aidan.scannell@aalto.fi