Automatic Control III

Lecture 10 – A snapshot of some ongoing research

Thomas Schön
Division of Systems and Control
Department of Information Technology
Uppsala University.

Email: thomas.schon@it.uu.se,
www: user.it.uu.se/~thosc112
Data-efficient control in continuous state-action spaces using very high-dimensional observations remains a key challenge in developing fully autonomous systems.

Recent paper by Google Deepmind


The so-called Pixels-to-torques problem.

**General goal**: Control the system to a state where a certain target frame $x_{\text{ref}}$ without any prior knowledge of the system at hand.

Key ingredient: a **deep dynamical model** for learning a low-dimensional feature embedding of images jointly with a predictive model in this low-dimensional feature space.
Problem set-up and objective

Objective: Find a closed-loop control law $\pi^*$, that minimize

$$V^\pi = \sum_{t=0}^{N-1} c(s_t, u_t),$$

where $c$ denotes a cost, $s_t \in \mathbb{R}^{N_s}$ is the continuous-valued state, and $u_t \in \mathbb{R}^{N_u}$ are continuous control signals.

Two additional challenges:

1. The agent has no access to the true state $s_t$, but perceives the environment only through high-dimensional pixel information $s_t \in \mathbb{R}^{N_x}$ (images).
2. A good control policy is required in only a few trials.
We will present these results in Beijing in two weeks time:

Niklas Wahlström, Thomas B. Schön and Marc Peter Deisenroth. Learning deep dynamical models from image pixels. In Proceedings of the 17th IFAC Symposium on System Identification (SYSID), Beijing, China, October 2015.
Here we show the learned feature space $z \in \mathbb{R}^2$ for different pendulum angles between $0^\circ$ and $360^\circ$. The DDM has learned to generate features that represent the angle of the pendulum, as they are mapped to a circle-like shape accounting for the wrap-around property of an angle.
Nonlinear controller

Using the deep dynamical model, nonlinear MPC is used to find an optimal (open-loop) control sequence $u^*_0, \ldots, u^*_{K-1}$, such that the predicted features $\hat{z}_1, \ldots, \hat{z}_K$ gets as close to the target feature $z_{\text{ref}}$ as possible, which results in the objective

$$u^*_0, \ldots, u^*_{K-1} \in \arg \min_{u_0:K-1} \sum_{t=0}^{K-1} \| \hat{z}_t - z_{\text{ref}} \|^2 + \lambda \| u_t \|^2,$$

where $\| \hat{z}_t - z_{\text{ref}} \|^2$ is a cost associated with the deviation of the predicted features $\hat{z}_0:K-1$ from the reference feature $z_{\text{ref}}$, and $\| u_t \|^2$ penalizes the amplitude of the control signals.
Examples – planar pendulum

True video frames

Predicted video frames

(a) Planar pendulum
Example – planar double pendulum

True video frames

\[ x_{t+0}, x_{t+1}, x_{t+2}, x_{t+3}, x_{t+4}, x_{t+5}, x_{t+6}, x_{t+7}, x_{t+8} \]

Predicted video frames

\[ \hat{x}_{t+0}, \hat{x}_{t+1}, \hat{x}_{t+2}, \hat{x}_{t+3}, \hat{x}_{t+4}, \hat{x}_{t+5}, \hat{x}_{t+6}, \hat{x}_{t+7}, \hat{x}_{t+8} \]

(b) Planar double pendulum
If you want to read more about this, you can find the following papers from my web site.

Niklas Wahlström, Thomas B. Schön and Marc Peter Deisenroth. Learning deep dynamical models from image pixels. In Proceedings of the 17th IFAC Symposium on System Identification (SYSID), Beijing, China, October 2015.
