Explore the Context: Optimal Data Collection for Context-Conditional Dynamics Models

Jan Achterhold, Joerg Stueckler

Embodied Vision Group, Max Planck Institute for Intelligent Systems, Tübingen, Germany

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

CyberValley

Problem setting

- *Family of systems* (e.g., MountainCar with random terrain profiles), modeled by a parameterized dynamics model
- Given a test environment from the family of systems, what actions should an agent perform to *calibrate* the parameterized forward dynamics model?

Our approach

- Context-conditional dynamics model with context encoder for amortized inference
- Compute optimal calibration action sequence via Information-Gain maximization

→ Computational alternative to hand-crafted system identification signals!

Improvements compared to random actions for model calibration in terms of

- predictive accuracy of the calibrated model
- performance of control algorithms using the calibrated models for planning.

Model Recurrent context-conditional (multi-step) dynamics model $q(x_{1:N} \mid x_0, u_{0:N-1}, \beta) = \prod_{n=1}^{N} q(x_n \mid x_0, u_{0:n-1}, \beta)$ with context-encoder $q(\beta|C) = \mathcal{N}(\mu(C), \mathrm{diag}(\sigma^2(C))$ (combining ideas from [2, 3]).

Training procedure:

1. Training data: Apply random actions in environment samples (indexed by α)
2. Iteratively maximize $\mathbb{E}_{D^\alpha, C^\alpha}[\mathcal{J}(D^\alpha, C^\alpha)]$
   - $\mathcal{J}$ is a lower bound $\mathcal{J}(D^\alpha, C^\alpha) \leq \log p(D^\alpha|C^\alpha)$
   - D^α: target chunk of length H $[x_n, u_n, x_{n+1}, ..., u_{n+H-1}, x_{n+H}]$
   - C^α: set of transitions $\{(x, u, x^+)\}$ (context observations)

   C^α and D^α are sampled from pre-collected data (see (1)).

Lower bound: Similar to [2]:

$$\mathcal{J}(D^\alpha, C^\alpha) = \mathbb{E}_{\boldsymbol{\beta} \sim q(\boldsymbol{\beta}|D^\alpha \cup C^\alpha)} [\mathcal{J}_{\mathrm{logll}}(D^\alpha, \boldsymbol{\beta})] - \lambda_{\mathrm{KL}} \mathrm{KL}(q(\boldsymbol{\beta}|D^\alpha \cup C^\alpha) || q(\boldsymbol{\beta}|C^\alpha))$$

$\mathcal{J}_{\mathrm{logll}}$ consists of single-step and multi-step prediction log-likelihoods.

Calibration

Goal: Compute actions $u_{0:N-1}$ which are maximally *informative* for the belief β.

Approach: Maximize expected information gain [4] for actions $u_{0:N-1}$

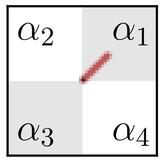
$$\mathrm{EIG}(u_{0:N-1}|x_0, T_0) = \mathbb{E}_{\substack{\beta_0 \sim q(\beta_0|T_0) \\ T \sim q(T|x_0, u_{0:N-1}, \beta_0)}} [H[q(\boldsymbol{\beta}|T_0)] - H[q(\boldsymbol{\beta}|T_0 \cup T)]]$$

given an initial state x_0 and already observed transitions T_0.

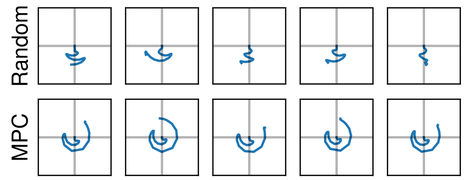
Open-Loop calibration: Computing the action sequence once at the beginning.

MPC calibration: Re-planning the action sequence after every applied action.

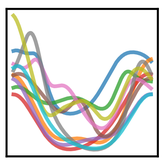
Baseline: Applying random actions to collect calibration transitions.



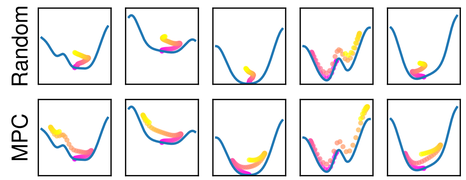
Pendulum dynamics vary per quadrant



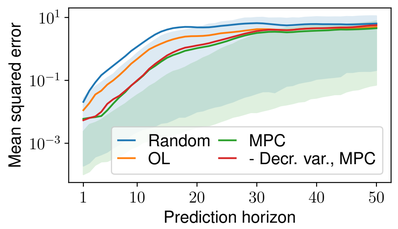
Pendulum angle extruded over time (starting at center) for random and MPC calibration rollouts



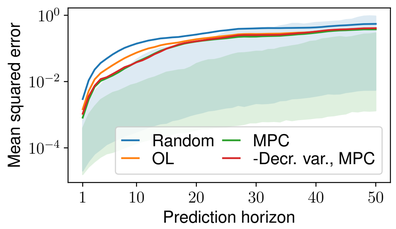
Sampled terrain profiles of the MountainCar environment



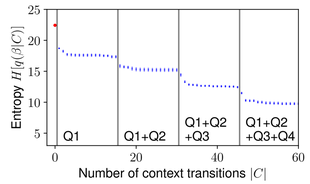
Random (top) vs MPC (bottom) calibration rollouts on 5 sampled MountainCar profiles (pink: t=1, yellow: t=50).



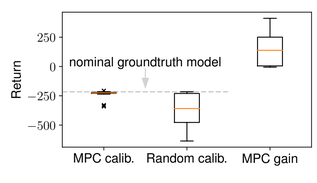
Prediction error of the calibrated Pendulum model



Prediction error of the calibrated MountainCar model



The entropy of β behaves reasonably for context sets containing transitions from a varying number of quadrants



MPC-calibrated models perform better in swing-up task than models calibrated with random actions

Paper and supplementary material available at:
`https://explorethecontext.is.tue.mpg.de`

Acknowledgements:
This work has been supported through the Max Planck Society and Cyber Valley. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Jan Achterhold.

References:

[1] Achterhold, J. and Stueckler, J. (2021). Explore the Context: Optimal Data Collection for Context-Conditional Dynamics Models. Accepted for publication at the *24th International Conference on Artificial Intelligence and Statistics (AISTATS)*.

[2] Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S. M. A., and Teh, Y. W. (2018). Neural processes. In *CoRR*, abs/1807.01622.

[3] Hafner, D., Lillicrap, T., Fischer, I., Villegas, R., Ha, D., Lee, H., and Davidson, J. (2019). Learning latent dynamics for planning from pixels. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 97 of Proceedings of Machine Learning Research, pages 2555–2565. PMLR.

[4] Lindley, D. V. (1956). On a measure of the information provided by an experiment. In *The Annals of Mathematical Statistics*, 27(4):986–1005.

jan.achterhold@tuebingen.mpg.de

joerg.stueckler@tuebingen.mpg.de

explorethecontext.is.tuebingen.mpg.de