Probabilistic Prediction Models for Data-Efficient RL

Marc Deisenroth

Imperial College London
PROWLER.io

@mpd37
m.deisenroth@imperial.ac.uk
marc@prowler.io

AAMAS/ICML/IJCAI Workshop on Prediction and Generative Modeling in Reinforcement Learning

July 15, 2018

Model-based Reinforcement Learning

- Models of the transition function
- Learned model serves as a proxy of real environment
- Learn policy using the model instead of the environment

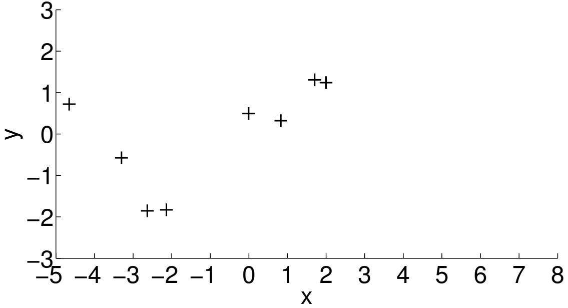
  ⏩ Reduce interactions with the system

Model-based Reinforcement Learning

- Models of the transition function
- Learned model serves as a proxy of real environment
- Learn policy using the model instead of the environment
  ⏩ Reduce interactions with the system
- **Model bias/errors**
- Probabilistic prediction models in RL
  - Account for uncertainty ⏩ Mitigate effect of model errors
  - Exploration ("natural" and "safe")
  - Meta learning
  - Incorporation of engineering priors

Model Errors/Bias

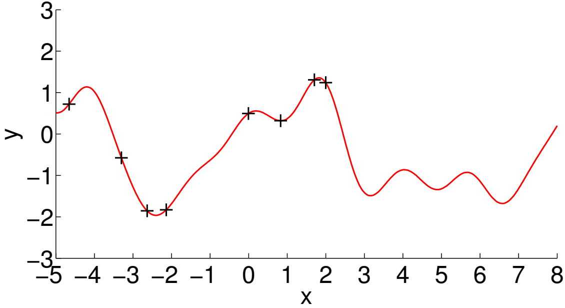
Model learning problem: Find a function $f : x \mapsto f(x) = y$



Observed function values

Model Errors/Bias

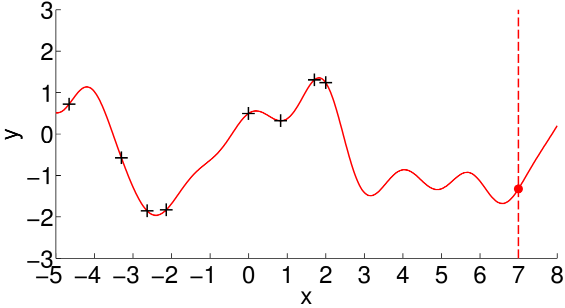
Model learning problem: Find a function $f : x \mapsto f(x) = y$



Plausible model

Model Errors/Bias

Model learning problem: Find a function $f : x \mapsto f(x) = y$

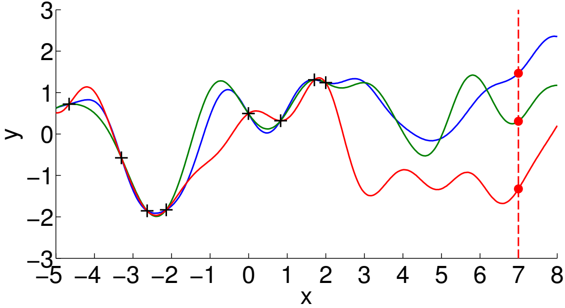


Plausible model

Predictions? Decision Making?

Model Errors/Bias

Model learning problem: Find a function $f : x \mapsto f(x) = y$

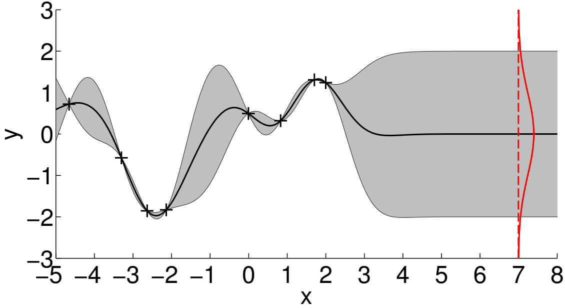


More plausible models

Predictions? Decision Making? Model Errors!

Model Errors/Bias

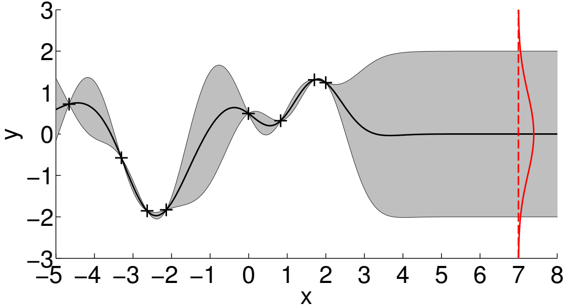
Model learning problem: Find a function $f : x \mapsto f(x) = y$



Distribution over plausible functions

Model Errors/Bias

Model learning problem: Find a function $f : x \mapsto f(x) = y$



Distribution over plausible functions

- Express uncertainty about the underlying function to be robust to model errors
- Gaussian process for model learning (Rasmussen & Williams, 2006)

Fast Reinforcement Learning

PILCO Framework: High-Level Steps

1. Learn probabilistic model of transition function
2. Compute long-term predictions and expected cost/reward using the model
3. Policy improvement
4. Apply controller to system

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Probabilistic Model Essential?

DEMO

- Probabilistic model: GP
- Deterministic model: Mean function of GP (still nonparametric)

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Probabilistic Model Essential?

DEMO

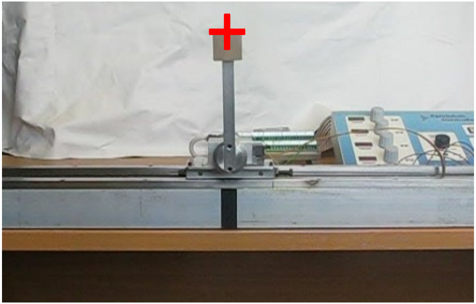
- Probabilistic model: GP
- Deterministic model: Mean function of GP (still nonparametric)

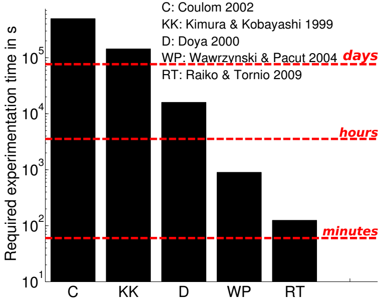
Table: Average learning success with nonparametric (NP) transition models

	GP	"Deterministic" GP
Learning success	**94.52%**	**0%**

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Standard Benchmark Problem: Cart-Pole Swing-up

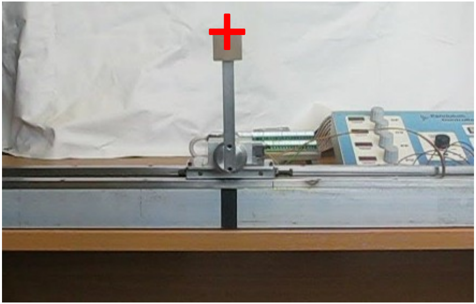


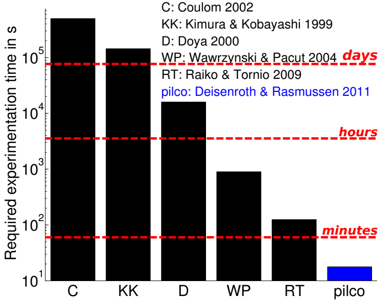


- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch
- Saturating cost function $1 - \exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$
- Code available at `https://github.com/icl-sml/pilco-matlab`

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

Standard Benchmark Problem: Cart-Pole Swing-up

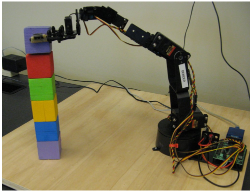


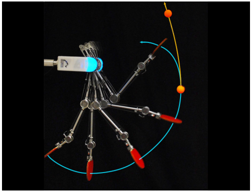


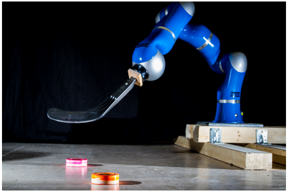
- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch
- Saturating cost function $1 - \exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$
- Code available at `https://github.com/icl-sml/pilco-matlab`
- **Unprecedented learning speed** compared to state-of-the-art

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

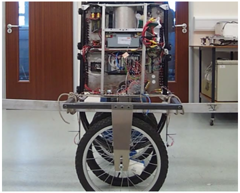
Other Real-World Applications

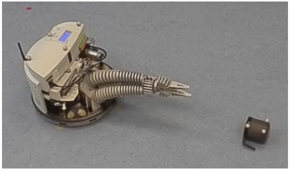

with D Fox


with P Englert et al.


with A Kupcsik et al.


B Bischoff (Bosch), ECML 2013

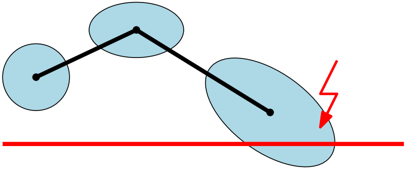

A McHutchon (U Cambridge)


with B Bischoff et al.

▶▶ Application to a wide range of robotic systems

Deisenroth et al. (RSS, 2011): *Learning to Control a Low-Cost Manipulator using Data-efficient Reinforcement Learning*
Englert et al. (ICRA, 2013): *Model-based Imitation Learning by Probabilistic Trajectory Matching*
Deisenroth et al. (ICRA, 2014): *Multi-Task Policy Search for Robotics*
Kupcsik et al. (AAAI, 2013): *Data-Efficient Generalization of Robot Skills with Contextual Policy Search*

Safe Exploration



- Deal with real-world safety constraints
- Use probabilistic model to predict whether constraints are violated (e.g., Sui et al., 2015; Berkenkamp et al., 2017)
- Adjust policy if necessary (during policy learning)

➡ Safe exploration within an MPC-based RL setting

Probabilistic MPC in RL

- GP model for transition dynamics
- Repeat (while executing the policy):
  1. In current state $\boldsymbol{x}_t$, determine optimal control sequence $\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_H^*$
  2. Apply first control $\boldsymbol{u}_1^*$ in state $\boldsymbol{x}_t$
  3. Transition to next state $\boldsymbol{x}_{t+1}$
  4. Update transition model

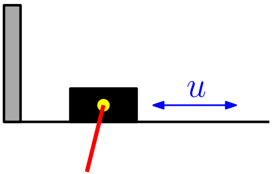
Kamthe & Deisenroth (AISTATS, 2018): *Data-Efficient Reinforcement Learning with Probabilistic Model Predictive Control*

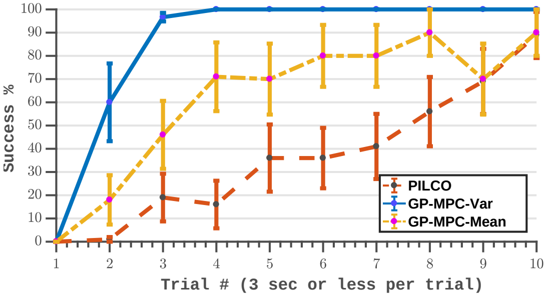
Probabilistic MPC in RL

- GP model for transition dynamics
- Repeat (while executing the policy):
  1. In current state $\boldsymbol{x}_t$, determine optimal control sequence $\boldsymbol{u}_1^*, \dots, \boldsymbol{u}_H^*$
  2. Apply first control $\boldsymbol{u}_1^*$ in state $\boldsymbol{x}_t$
  3. Transition to next state $\boldsymbol{x}_{t+1}$
  4. Update transition model
- Theoretical guarantees with GP dynamics models via Pontryagin's Maximum Principle
- Principled treatment of state/control constraints
- Including the most recent state transition in the model significantly improves robustness to model errors

Kamthe & Deisenroth (AISTATS, 2018): *Data-Efficient Reinforcement Learning with Probabilistic Model Predictive Control*

Experimental Results: Constraints





PILCO	16/100	constraint violations
GP-MPC-Mean	21/100	constraint violations
GP-MPC-Var	3/100	constraint violations

▸▸ **Propagating model uncertainty important for safety**

Kamthe & Deisenroth (AISTATS, 2018): *Data-Efficient Reinforcement Learning with Probabilistic Model Predictive Control*

Meta Learning

- Different robot configurations (link lengths, weights, ...)
- Re-use experience gathered so far generalize to new dynamics that are similar
- Accelerated learning

Sæmundsson et al. (UAI, 2018): *Meta Reinforcement Learning with Latent Variable Gaussian Processes*

Meta Learning

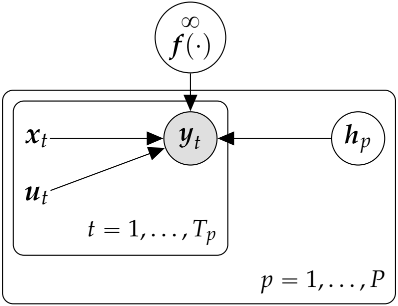
- Different robot configurations (link lengths, weights, ...)
- Re-use experience gathered so far generalize to new dynamics that are similar
- Accelerated learning

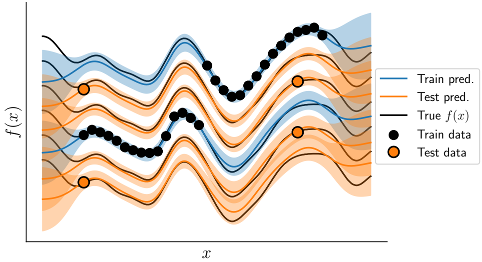
Approach:

- Model (unknown) configurations with latent variable
- Disentangle global and task specific properties
- Online inference of models of unseen configurations
- Few-shot model-based RL

Sæmundsson et al. (UAI, 2018): *Meta Reinforcement Learning with Latent Variable Gaussian Processes*

Meta Model Learning with Latent Variables

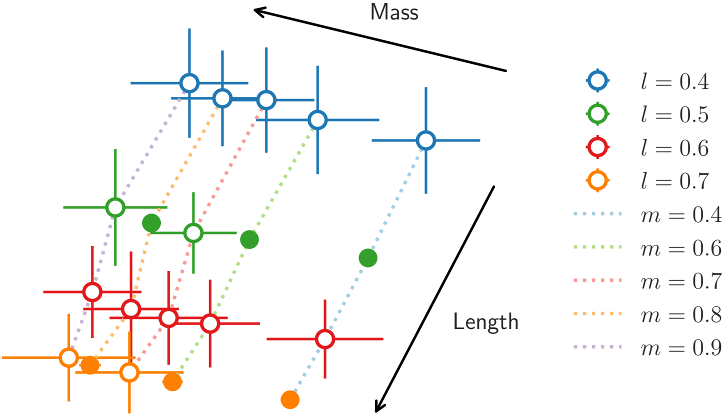




- GP captures global properties of the dynamics
- Latent variable $\boldsymbol{h}$ describes local configuration
  - ▶▶ Variational inference to learn latent configuration
- Fast online inference of new configurations (no model re-training required)

Sæmundsson et al. (UAI, 2018): *Meta Reinforcement Learning with Latent Variable Gaussian Processes*

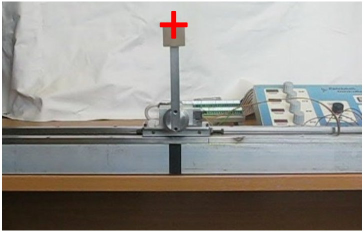
Latent Embeddings



- Latent variable h encodes length l and mass m of the cart pole
- 6 training tasks, 14 held-out test tasks

Sæmundsson et al. (UAI, 2018): *Meta Reinforcement Learning with Latent Variable Gaussian Processes*

Meta-RL (Cart Pole): Training



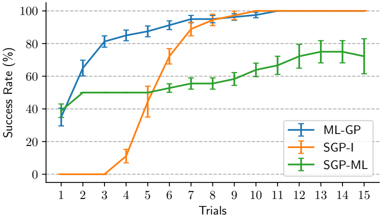
- Pre-trained on 6 training configurations until solved

Model	Training (s)	Description
Independent	16.1 ± 0.4	Independent GP-MPC
Aggregated	23.7 ± 1.4	Aggregated experience (no latents)
Meta learning	$\mathbf{15.1 \pm 0.5}$	Aggregated experience (with latents)

▸▸ **Meta learning can help speeding up RL**

Sæmundsson et al. (UAI, 2018): *Meta Reinforcement Learning with Latent Variable Gaussian Processes*

Meta-RL (Cart Pole): Few-Shot Generalization

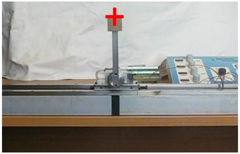


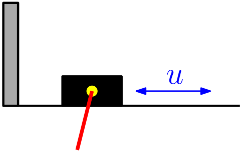
- Few-shot generalization on 4 unseen configurations
- Success: solve all 10 (6 training + 4 test) tasks
- Meta learning: blue
- Independent (GP-MPC): orange
- Aggregated experience model (no latents): green

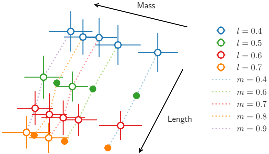
▶▶ **Meta RL generalizes well to unseen tasks**

Sæmundsson et al. (UAI, 2018): *Meta Reinforcement Learning with Latent Variable Gaussian Processes*

Wrap-up







- Probabilistic models in RL
  - Reduce model bias for data-efficient RL
  - Safe exploration
  - Meta learning with latent variables for few-shot learning
- Key to success: Probabilistic modeling and Bayesian inference

`m.deisenroth@imperial.ac.uk`
`marc@prowler.io`

Thank you for your attention

References I

[1] F. Berkenkamp, M. Turchetta, A. P. Schoellig, and A. Krause. Safe Model-based Reinforcement Learning with Stability Guarantees. In *Advances in Neural Information Processing Systems*, 2017.

[2] B. Bischoff, D. Nguyen-Tuong, T. Koller, H. Markert, and A. Knoll. Learning Throttle Valve Control Using Policy Search. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, 2013.

[3] B. Bischoff, D. Nguyen-Tuong, H. van Hoof, A. McHutchon, C. E. Rasmussen, A. Knoll, J. Peters, and M. P. Deisenroth. Policy Search For Learning Robot Control Using Sparse Data. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2014.

[4] R. Coulom. *Reinforcement Learning Using Neural Networks, with Applications to Motor Control*. PhD thesis, Institut National Polytechnique de Grenoble, 2002.

[5] M. P. Deisenroth, P. Englert, J. Peters, and D. Fox. Multi-Task Policy Search for Robotics. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2014.

[6] M. P. Deisenroth, D. Fox, and C. E. Rasmussen. Gaussian Processes for Data-Efficient Learning in Robotics and Control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):408–423, Feb. 2015.

[7] M. P. Deisenroth and C. E. Rasmussen. PILCO: A Model-Based and Data-Efficient Approach to Policy Search. In *Proceedings of the International Conference on Machine Learning*, pages 465–472. ACM, June 2011.

[8] M. P. Deisenroth, C. E. Rasmussen, and D. Fox. Learning to Control a Low-Cost Manipulator using Data-Efficient Reinforcement Learning. In *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2011.

[9] P. Englert, A. Paraschos, J. Peters, and M. P. Deisenroth. Model-based Imitation Learning by Probabilistic Trajectory Matching. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2013.

[10] S. Kamthe and M. P. Deisenroth. Data-Efficient Reinforcement Learning with Probabilistic Model Predictive Control. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, April 2018.

[11] H. Kimura and S. Kobayashi. Efficient Non-Linear Control by Combining Q-learning with Local Linear Controllers. In *Proceedings of the* 16^{th} *International Conference on Machine Learning*, pages 210–219, 1999.

References II

[12] A. Kupcsik, M. P. Deisenroth, J. Peters, and G. Neumann. Data-Efficient Generalization of Robot Skills with Contextual Policy Search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2013.

[13] T. Raiko and M. Tornio. Variational Bayesian Learning of Nonlinear Hidden State-Space Models for Model Predictive Control. *Neurocomputing*, 72(16–18):3702–3712, 2009.

[14] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA, 2006.

[15] S. Sæmundsson, K. Hofmann, and M. P. Deisenroth. Meta Reinforcement Learning with Latent Variable Gaussian Processes. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2018.

[16] Y. Sui, A. Gotovos, J. W. Burdick, and A. Krause. Safe Exploration for Optimization with Gaussian Processes. In *Proceedings of the International Conference on Machine Learning*, 2015.

[17] P. Wawrzynski and A. Pacut. Model-free off-policy Reinforcement Learning in Continuous Environment. In *Proceedings of the INNS-IEEE International Joint Conference on Neural Networks*, pages 1091–1096, 2004.