CIS 700-003 FINAL PROJECT PROPOSAL

KUN HUANG [HUANGKUN@SEAS]
ZHIHAO RUAN [RUANZH@SEAS]

DevDAgger: Data Efficient Visual Imitation Learning Through Variational Autoencoders and Gaussian Process

1. BACKGROUND

Imitation learning is commonly used in planning and control in high-dimensional state space with non-linear dynamical model in robotics. It offers robots the ability to manipulate in complex environment without explicitly modelling the environment by imitating the control actions from an expert (i.e., human). Dataset Aggregation (DAgger) [1], in particular, is a popular online imitation learning algorithm that trains the controller while aggregating new data queried from the expert simultaneously. Traditional DAgger uses a neural network to train the controller, requiring lots of expert data which is often costly to obtain in many situations. On the contrary, Gaussian Process (GP) is a commonly used model-free data efficient technique for regression tasks. Recent literature also shows a particular interest in vision-based robotic planning and control problems (visuomotor control) [2, 3, 4]. As a consequence, we propose a *Data-Efficient Vision-based DAgger (DevDAgger)* algorithm for imitation learning tasks, combining the advantage of Gaussian Process and latent representation learning with Variational Autoencoders (VAEs).

2. RELATED WORK

DAgger was proposed in 2011 as an online supervised learning algorithm for imitation learning by constantly aggregating data from the expert and imitating expert's behaviors [1]. Recent works have addressed some drawbacks of DAgger by improving the expert data querying efficiency [5]. Uncertainty estimation was also introduced to DAgger in [6] to improve the training-time safety.

Traditional exact GP suffers from the limitation of dataset in order to achieve reasonable amount of training time and thus cannot be put into direct use in large scale learning tasks. However, as the computational efficiency of exact GP has been greatly improved recently [7, 8], we decided to utilize the power of exact GP in our DAgger solution. At the same time, although the GP suffers from the curse of dimensionality, recent works have shown that using Variational Autoencoders to map the structured high-dimensional inputs to latent space [9] dramatically expands the capability of GPs.

3. Proposed Methods

Our proposed approach is composed of two parts: a novel probabilistic visuomotor controller and a uncertainty-based decision rule for DAgger to optimize expert query frequency.

In a visual imitation learning setting, the visuomotor controller takes an image O_t as the input, and outputs the control action a_t that's learned from the expert's demonstrations. Traditionally such visuomotor is realized by end-to-end deep convolutional neural networks [10]. However, in order to exploit the structural information embedded in the high-dimensional visual inputs, we propose to use a Variational Autoencoder to explicitly map the images to low-dimensional latent representations z_t in an unsupervised manner. To learn an uncertainty-aware controller based on the latent representations z_t, we propose to use Gaussian process due to its capability of fitting functions with accurate estimations of the uncertainties while at the same time using fewer data.

The trade-off between exploration using the learned novice policy π_{nov} and exploitation using the expert policy π_{exp} can be automatically benefited from the knowledge of uncertainties. We propose to use the uncertainty estimation coming from the Gaussian process to derive a sampling-based decision rule for DAgger. The underlying intuition is that when the uncertainty is high, we need to query the expert for their demonstration; otherwise, we can simply use our learned novice policy to collect trajectories. As a result, we should be able to significantly reduce the amount of expert demonstrations that is required for training a novice policy.

References

[1] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.

[2] Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen, Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke, and Sergey Levine. Scalable deep reinforcement learning for vision-based robotic manipulation. In Aude Billard, Anca Dragan, Jan Peters, and Jun Morimoto, editors, *Proceedings of The 2nd Conference on Robot Learning*, volume 87 of *Proceedings of Machine Learning Research*, pages 651–673. PMLR, 29–31 Oct 2018.

[3] Frederik Ebert, Chelsea Finn, Sudeep Dasari, Annie Xie, Alex Lee, and Sergey Levine. Visual foresight: Model-based deep reinforcement learning for vision-based robotic control, 2018.

[4] Annie Xie, Avi Singh, Sergey Levine, and Chelsea Finn. Few-shot goal inference for visuomotor learning and planning. In *Conference on Robot Learning*, pages 40–52. PMLR, 2018.

[5] Jiakai Zhang and Kyunghyun Cho. Query-efficient imitation learning for end-to-end simulated driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[6] Kunal Menda, Katherine Driggs-Campbell, and Mykel J Kochenderfer. Ensembledagger: A bayesian approach to safe imitation learning. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5041–5048. IEEE, 2019.

[7] Ke Wang, Geoff Pleiss, Jacob Gardner, Stephen Tyree, Kilian Q Weinberger, and Andrew Gordon Wilson. Exact gaussian processes on a million data points. *Advances in Neural Information Processing Systems*, 32:14648–14659, 2019.

[8] Jacob R. Gardner, Geoff Pleiss, David S. Bindel, Kilian Q. Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *NeurIPS*, 2018.

[9] Antoine Grosnit, Rasul Tutunov, Alexandre Max Maraval, Ryan-Rhys Griffiths, Alexander I Cowen-Rivers, Lin Yang, Lin Zhu, Wenlong Lyu, Zhitang Chen, Jun Wang, et al. High-dimensional bayesian optimisation with variational autoencoders and deep metric learning. *arXiv preprint arXiv:2106.03609*, 2021.

[10] Chelsea Finn, Tianhe Yu, Tianhao Zhang, Pieter Abbeel, and Sergey Levine. One-shot visual imitation learning via meta-learning. In Sergey Levine, Vincent Vanhoucke, and Ken Goldberg, editors, *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78 of *Proceedings of Machine Learning Research*, pages 357–368. PMLR, 13–15 Nov 2017.

APPENDIX A. OTHER PROJECT IDEAS

(1) Probabilistic methods for imitation learning:
- Learning from demonstration with model-based Gaussian process. *CoRL '19*. [Paper]
- Gaussian process based model predictive controller for imitation learning. *IEEE RAS on Humanoid Robots '17.* [Paper]
- Keep it Simple: Data-efficient Learning for Controlling Complex Systems with Simple Models. *IEEE Robotics and Automation Letters (RA-L) 2021*. [Paper]
- EnsembleDAgger: A Bayesian Approach to Safe Imitation Learning. *IROS '19.* [Paper]
- PILCO: A Model-Based and Data-Efficient Approach to Policy Search. *ICML '11.* [Paper]
- Deep Reinforcement Learning in a Handful of Trials using Probabilistic Dynamics Models. *NeurIPS '18.* [Paper]
- Model-based imitation learning by probabilistic trajectory matching. *ICRA '13.* [Paper]
- Generative predecessor models for sample-efficient imitation learning. *ICLR '19.* [Paper]
- Generative Adversarial Imitation Learning (GAIL). *NeurIPS '16.* [Paper]

(2) BayesOpt for robotics simulator calibration for Sim2Real transfer learning:
- A User's Guide to Calibrating Robotics Simulators. *CoRL '20*. [Paper]
- BayesSim: adaptive domain randomization via probabilistic inference for robotics simulators. *RSS '19*. [Paper]

(3) GP for control:
- Learning to Control an Unstable System with One Minute of Data: Leveraging Gaussian Process Differentiation in Predictive Control. *IROS '21*. [Paper] [Yisong Yue @Caltech]

(4) GP for planning under uncertainty:
- Active Policy Learning for Robot Planning and Exploration under Uncertainty. *RSS '07.* [Paper]