Localisation-Aware Informative Path Planning via Meta Bayesian Reinforcement Learning

Christopher Iliffe Sprague, Petter Ögren
Robotics, Perception, and Learning (RPL)
KTH Royal Institute of Technology
Swedish Maritime Robotics Centre (SMaRC)
{sprague, petter}@kth.se

In this paper, we investigate using *reinforcement learning* (RL) to learn an optimal control, resulting in a continuous trajectory that minimise the uncertainty of a bathymetric model, whilst respecting resource constraints (e.g. energy or time) and taking into account localisation-uncertainty.

Autonomous underwater vehicles (AUV) are rapidly becoming crucial to furthering scientists' understanding of oceanographic phenomena — due to their unique capability to gather in-situ data in inaccessible environments. This is especially true in polar regions, beneath ice-shelves and sea-ice, which are among the least accessible environments on Earth. In these settings, AUV autonomy is a key enabler for current and forthcoming operations [1].

Despite the plethora of sensors that AUVs often wield, their path-planning autonomy typically extends only to coverage-based navigation [2] based on dead-reckoning (e.g. with lawnmower patterns), where the waypoints are precoordinated by the operator. Although this approach works well in practice and is effective at mitigating risks [1], it is limited to the expectations of the operator [3] — and it is likely that more deliberative data-driven exploration methods could provide more efficiency and fidelity [4], as well as more fruitful scientific findings.

This is especially pertinent in frontier environments, where their very nature is not well understood [4]. In this case, it is difficult to preconceive the distribution and scale of environmental features [3]. Hence, it is paramount to avoid dependencies on operator expectations and the shortcommings of coverage-based methods by employing *adaptive*-sampling strategies to ensure both efficiency (e.g. minimising expended energy or time) and fidelity (i.e. maximising data certainty).

The core problem surrounding this is *informative path planning* (IPP) — where the objective is to plan how to navigate a given environment in order to maximise the certainty of gathered data whilst simultaneously considering resource constraints. This problem has been addressed by a growing body of research, particularly in the aquatic [5], [6] and aerial [7] domains, where environmental field reconstruction is intended. Optimising this problem leads to paths that outperform conventional coverge-based methods [7], leading to savings in both time and money.

In most contemporary approaches to IPP, *Gaussian processes* (GP) [8] are used to incrementally model environmental fields, as they conveniently provide uncertainty quantification, allowing map-entropy minimisation to be used as the primary objective. Unfortunately, however, most of such approaches incorrectly assume perfect state information despite the significant rôle that localisation uncertainty plays in the certainty of observations [9]–[11]. Additionally, a common issue with this approach is the increasing cost of GP regression (krigging) as dense sensory information is accumulated. To circumvent these issues, we employ stochastic variational GPs [12], which naturally enable us to incorporate localisation uncertainty (stochastic inputs) and use stochastic optimisation to quickly regress large amounts a sensory data.

Among the most attractive adaptive approaches to IPP are those that plan *continuous* trajectories, offering greater scalability compared to *discrete* approaches — our work proceeds in this fashion. In this regime, a variety of path-planning techniques have emerged, ranging from rapidly-exploring random trees [13] to evolutionary methods [7], [9] and *Bayesian optimisation* (BO) [14], [15], to name a few. However, the use of RL has been virtually unexplored in this setting, despite its favourable ability to generalise and extract implicit information to improve data-efficiency [16].

Recently, in the conventional BO setting, it was shown in [16] that *acquisition functions* (AF) could be learnt using RL algorithms, resulting in adaptive neural AFs that learn to use the structure of the underlying objective function to accelerate the optimisation process. It was also shown that the structure learnt in one context could be used to accelerate optimisation in another — this is known as *meta-learning*. As shown in [10], [15], [17], BO is particularly well suited to IPP, as AFs provide a principled way to make decisions — balancing exploration and exploitation — based on the incrementally built posterior arising from the GP-modelled environmental field. However, these approaches only consider analytical AFs, which do not adapt their policies to the given context in the way that neural AFs can. In this paper, we employ RL for BO in IPP, but, in contrast to existing IPP approaches that use BO, we learn a policy to compute the optimal control rather than the optimal waypoint. Unlike continuous IPP approaches that plan geometric trajectories (e.g. splines), computing the optimal control automatically guarantees the dynamic feasibility of trajectories. Our approach will first be evaluated in simulation, and later with real bathymetric data.

Acknowledgement

This work was supported by Stiftelsen for StrategiskForskning (SSF) through the Swedish Maritime Robotics Centre (SMaRC) (IRC15-0046).

References

[1] P. King, G. Williams, R. Coleman, K. Zürcher, I. Bowden-Floyd, A. S. Ronan, C. Kaminski, J.-M. Laframboise, S. D. McPhail, J. P. Wilkinson, A. D. Bowen, P. Dutrieux, N. Bose, A. K. Wåhlin, J. Andersson, P. Boxall, M. Sherlock, and T. Maki, “Deploying an auv beneath the sørsdal ice shelf: Recommendations from an expert-panel workshop,” *2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV)*, pp. 1–6, 2018.

[2] E. Galceran and M. Carreras, “A survey on coverage path planning for robotics,” *Robotics Auton. Syst.*, vol. 61, pp. 1258–1276, 2013.

[3] F. Thompson and D. Guihen, “Review of mission planning for autonomous marine vehicle fleets,” *J. Field Robotics*, vol. 36, pp. 333–354, 2019.

[4] J. Bellingham, P. E. Dupont, P. Fischer, L. Floridi, R. J. Full, N. Jacobstein, V. Kumar, M. McNutt, R. D. Merrifield, B. J. Nelson, B. Scassellati, M. Taddeo, R. Taylor, M. M. Veloso, Z. L. Wang, and R. J. Wood, “The grand challenges of science robotics,” *Science Robotics*, vol. 3, 2018.

[5] K.-C. Ma, L. Liu, H. K. Heidarsson, and G. S. Sukhatme, “Data-driven learning and planning for environmental sampling,” *J. Field Robotics*, vol. 35, pp. 643–661, 2018.

[6] H. Yetkin, J. McMahon, N. Topin, A. Wolek, Z. Waters, and D. J. Stilwell, “Online planning for autonomous underwater vehicles performing information gathering tasks in large subsea environments,” *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6354–6361, 2019.

[7] M. Popović, T. Vidal-Calleja, G. Hitz, J. J. Chung, I. Sa, R. Siegwart, and J. Nieto, “An informative path planning framework for uav-based terrain monitoring,” *Autonomous Robots*, pp. 1–23, 2020.

[8] C. E. Rasmussen and C. K. I. Williams, “Gaussian processes for machine learning,” in *Adaptive computation and machine learning*, 2005.

[9] M. Popovic, T. Vidal-Calleja, J. J. Chung, J. Nieto, and R. Siegwart, “Informative path planning and mapping for active sensing under localization uncertainty,” *arXiv preprint arXiv:1902.09660*, 2019.

[10] R. Oliveira, L. Ott, V. Guizilini, and F. Ramos, “Bayesian optimisation for safe navigation under localisation uncertainty,” in *Robotics Research*. Springer, 2020, pp. 489–504.

[11] M. G. Jadidi, J. V. Miro, and G. Dissanayake, “Sampling-based incremental information gathering with applications to robotic exploration and environmental monitoring,” *The International Journal of Robotics Research*, vol. 38, pp. 658 – 685, 2019.

[12] J. Hensman, A. Matthews, and Z. Ghahramani, “Scalable variational gaussian process classification,” 2015.

[13] L. Schmid, M. Pantic, R. Khanna, L. Ott, R. Siegwart, and J. Nieto, “An efficient sampling-based method for online informative path planning in unknown environments,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1500–1507, 2020.

[14] R. Marchant and F. T. Ramos, “Bayesian optimisation for informative continuous path planning,” *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6136–6143, 2014.

[15] P. Morere, R. Marchant, and F. Ramos, “Continuous state-action-observation pomdps for trajectory planning with bayesian optimisation,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 8779–8786.

[16] M. Volpp, L. Fröhlich, K. Fischer, A. Doerr, S. Falkner, F. Hutter, and C. Daniel, “Meta-learning acquisition functions for transfer learning in bayesian optimization,” in *ICLR 2020*, 2019.

[17] R. Marchant, F. Ramos, S. Sanner, *et al.*, “Sequential bayesian optimisation for spatial-temporal monitoring.” in *UAI*, 2014, pp. 553–562.