Internship Proposal: Mean Field Optimal Control for Multi-Armed Bandit Problems

Supervisors:
Nicolas Gast (Inria) – http://mescal.imag.fr/membres/nicolas.gast/
Patrick Loiseau (Inria) – http://lig-membres.imag.fr/loiseapa/

Grenoble. Spring/Summer 2019.

Keywords: Online Learning, Mean Field Approximation, Stochastic Optimal Control

Lab: Laboratoire d'Informatique de Grenoble (LIG), Grenoble, France (head: Eric Gaussier)
Team: POLARIS (head: Arnaud Legrand)

1 Background

Mutli-armed bandits are classical models of sequential decision making problems in which a controller (or learner) needs to decide at each time step how to allocate its resources to a finite set of alternatives (called arms). They are widely used in online learning today as they provide theoretical tools to solve many practical problems: choosing which ads you see when you visit a web page, optimally routing packets in communication networks, designing efficient strategies in complex games such as the game of go, etc. When some information is available about the different arms, this falls into the class of Markovian bandits and restless multi-armed bandits. If the number of arms is small, optimal decision rules can be computed by dynamic programming. The complexity of such an approach, however, grows quickly with the number of arms which calls for *approximations*.

Mean field methods have become a common tool to provide approximations for systems of interacting agents or entities. They are used in many domains (statistical physics, game theory [5], performance and design of distributed systems [4] and more recently multi-agent reinforcement learning [8]). The idea of the mean-field approximation is that when the size of the population n is large, a single individual has a minor influence on the rest of the individuals. The mean field approximation consists in neglecting the effect that one individual has on the others. This approximation is known to be exact as n goes to infinity, but it can very poor for finite systems. Our recent progresses [3], however, show that it is possible to correct these methods to study systems with a relatively small number of entities ($n \approx 10$), which makes it a promising direction to derive efficient approximations for multi-armed bandit problems.

2 Goal of the internship

The main objective on the internship is to design (approximately) optimal allocation strategies in multi-armed bandit problems by using tools from *mean field optimal control*. Mean field optimal

control problems arise naturally as the limit of centralized [2, 1] or decentralized (games [5]) control problems and, as described above, the main idea is that mean field allows one to simplify control problems by *relaxing* a system with dependent agents into a system of independent agents.

The intern will first formalize a theoretical framework and find the proper relaxation to use corrected mean field approximations [3]. He/she will then work on solving theoretically the relaxed and corrected model (characterization of the solutions, etc.) in order to design efficient and approximately optimal learning policies. Finally, he/she will implement the obtained algorithms and perform numerical evaluations. As a benchmark, we will compare the obtained policies with the classical index policies that are known to be asymptotically optimal for restless bandits [6, 7].

3 Contact

For more information, please contact `nicolas.gast@inria.fr`.

4 Location

The intern will be hosted in the POLARIS team. The POLARIS team is a joint team between Inria and LIG (Grenoble Computer Science Laboratory) and is located on Grenoble University main campus (`https://batiment.imag.fr/`).

References

[1] Massimo Fornasier and Francesco Solombrino. "Mean-field optimal control". In: *ESAIM: Control, Optimisation and Calculus of Variations* 20.4 (2014), pp. 1123–1152.

[2] Nicolas Gast, Bruno Gaujal, and Jean-Yves Le Boudec. "Mean field for Markov decision processes: from discrete to continuous optimization". In: *IEEE Transactions on Automatic Control* 57.9 (2012), pp. 2266–2280.

[3] Nicolas Gast and Benny Van Houdt. "A Refined Mean Field Approximation". In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems - SIGMETRICS* 1.2 (Dec. 2017), 33:1–33:28. ISSN: 2476-1249. DOI: 10.1145/3154491. URL: http://doi.acm.org/10.1145/3154491.

[4] Nicolas Gast and Benny Van Houdt. "Transient and steady-state regime of a family of list-based cache replacement algorithms". In: *ACM SIGMETRICS Performance Evaluation Review* 43.1 (2015), pp. 123–136.

[5] Jean-Michel Lasry and Pierre-Louis Lions. "Mean field games". In: *Japanese journal of mathematics* 2.1 (2007), pp. 229–260.

[6] I. M. Verloop. "Asymptotically optimal priority policies for indexable and nonindexable restless bandits". In: *Ann. Appl. Probab.* 26.4 (Aug. 2016), pp. 1947–1995. DOI: 10.1214/15-AAP1137. URL: https://doi.org/10.1214/15-AAP1137.

[7] Richard R Weber and Gideon Weiss. "On an index policy for restless bandits". In: *Journal of Applied Probability* 27.3 (1990), pp. 637–648.

[8] Yaodong Yang et al. "Mean Field Multi-Agent Reinforcement Learning". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. Stockholmsmässan, Stockholm Sweden: PMLR, 2018, pp. 5571–5580. URL: http://proceedings.mlr.press/v80/yang18d.html.