Internship Proposal (Research): "Learning Restless Bandits, with Applications to Stochastic Scheduling"

Key Words Restless bandits. Reinforcement learning. Index policies; Maintenance problems. Stochastic scheduling.

Lab Laboratoire d'Informatique de Grenoble (LIG), Grenoble, France (head: Noël De Palma)

Team: POLARIS (head: Arnaud Legrand)

Background

Multi-armed restless bandit problems (MARBPs) are used to model optimal resource allocation problems. In a MARBP, there is a collection of stochastic binary-action (active/passive) projects that evolve over time, and the goal is to allocate resource to these projects in real time. A typical example of such a problem is given by stochastic scheduling, in which a machine needs to share its time between various resources. Other applications include asset management [6] or maintenance problems [3].

Until recently, MARBPs have been mostly studied from the computational point of view: for a given problem, how to compute a policy that gives a closte-to-optimal solution. While thse problems have been proved to be computationally hard [7], there exist computationally efficient policies, like Whittle index policies [8], which performs extremely well in practice while being computationally efficient [4].

Goal of the internship

The goal of the internship is to explore what we can do when the model information (transition probabilities and sometimes rewards) is unknown. To do that, we address this problem by deploying reinforcement learning (RL) based agents. Although RL algorithms have been recently explored for some restless bandits models [5, 2, 1], their performance guarantee and applicability in stochastic scheduling remains limited. We want to contribute to the following questions:

1. Starting from stochastic scheduling, propose a data-driven approach to this problem by modeling it as a restless bandit;

2. Develop reinforcement learning algorithms for the general multi-armed restless bandit problem where each bandit has Markovian transitions;

3. Envision application to other domains such as dynamic asset allocation, outsourcing warranty repairs, perishable inventory routing, etc.

The objective is to provide a framework for optimal control of stochastic distributed agents. This project is mostly of theoretical nature: it aims at contributing to the field of optimal control and reinforcement learning by developing new control mechanisms for distributed systems.

Additional information The project will benefic drom a bilateral partnership between France and India via the "équipe-projet associée" AIRBEA.

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

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] K. E. Avrachenkov and V. S. Borkar. "Whittle index based Q-learning for restless bandits with average reward". In: *Automatica* 139 (2022), p. 110186.

[2] A. Biswas, G. Aggarwal, P. Varakantham, and M. Tambe. "Learn to intervene: An adaptive learning policy for restless bandits in application to preventive healthcare". In: *arXiv preprint arXiv:2105.07965* (2021).

[3] S. Ford, M. P. Atkinson, K. Glazebrook, and P. Jacko. "On the dynamic allocation of assets subject to failure". In: *European Journal of Operational Research* 284.1 (2020), pp. 227–239.

[4] N. Gast, B. Gaujal, and K. Khun. "Computing whittle (and gittins) index in subcubic time". In: *arXiv preprint arXiv:2203.05207* (2022).

[5] N. Gast, B. Gaujal, and C. Yan. "(Close to) Optimal Policies for Finite Horizon Restless Bandits". In: *arXiv preprint arXiv:2106.10067* (2021).

[6] K. D. Glazebrook, H. Mitchell, and P. Ansell. "Index policies for the maintenance of a collection of machines by a set of repairmen". In: *European Journal of Operational Research* 165.1 (2005), pp. 267–284.

[7] C. H. Papadimitriou and J. N. Tsitsiklis. "The complexity of optimal queueing network control". In: *Proceedings of IEEE 9th Annual Conference on Structure in Complexity Theory*. IEEE. 1994, pp. 318–322.

[8] P. Whittle. "Restless bandits: Activity allocation in a changing world". In: *Journal of applied probability* 25.A (1988), pp. 287–298.