Robust non-parametric bandit algorithms

Emilie Kaufmann (CNRS & CRIStAL, Inria Lille (Scool))

The objective of this post-doc is to investigate the use of non-parametric bandit algorithms, especially sub-sampling algorithms, beyond the standard setting in which they have been analyzed. The hope is to tackle in particular structured bandits problems (e.g. linear contextual bandits).

Context Several Sub-Sampling Dueling (SDA) algorithms have been proved to attain optimal (asymptotic) regret for any bandit problem in which the rewards belong to a one-dimensional exponential family [Baudry et al., 2020, Baudry et al., 2021a], yet the complete characterization of the distributions for which they can have logarithmic (if not optimal) regret remains elusive. Alternative resampling based approaches based on history perturbation have also been proposed [Kveton et al., 2019], but fail to attain optimal instance-dependent regret. Another avenue of research have consider variants of Dirichlet Sampling, first proposed as a Non-Parametric extension to Thompson Sampling to bounded reward distributions [Riou and Honda, 2020, Baudry et al., 2021b].

Objective SDA algorithms rely on pairwise comparisons between empirical means of (sub-samples) of the arms. To enhance their applicability, we will seek to develop novel comparison mechanisms, either using robust statistics or leveraging the structure (e.g. a linear regression model) to find how to equalize the quality of estimation of two arms. For example for Gaussian bandits with unknown variances, [Chan, 2020] suggests that the empirical variance has to be taken into account in the comparison between arms. In linear bandits, we hope to design efficient algorithms that have optimal instance-dependent regret when the set of arms is fixed (which optimistic approaches typically cannot achieve [Lattimore and Szepesvári, 2017]) and good worse-case regret in the contextual case, in which the arms' features can change in every round (which optimistic approaches can achieve [Abbasi-Yadkori et al., 2011]). So far, only the method of [Kveton et al., 2019] has been investigated for linear bandits [Kveton et al., 2020], and does not enjoy optimal-instance dependent guarantees. Other interesting forms of structured bandits are studied in the works of [Magureanu et al., 2014, Combes and Proutière, 2014, Degenne et al., 2020, Pesquerel et al., 2021].

Practical information. The post-doc will be working at Inria Lille in the Scool team-project, under the supervision of Emilie Kaufmann and Odalric Ambrym-Maillard. The Scool team is made of 7 permanent researchers and around 20 PhD students and post-docs, all working on sequential decision making. The monthly salary will be around 2500 € (prior taxes).

Contact: `emilie.kaufmann@univ-lille.fr`

References

[Abbasi-Yadkori et al., 2011] Abbasi-Yadkori, Y., D.Pál, and C.Szepesvári (2011). Improved Algorithms for Linear Stochastic Bandits. In *Advances in Neural Information Processing Systems*.

[Baudry et al., 2020] Baudry, D., Kaufmann, E., and Maillard, O.-A. (2020). Sub-sampling for Efficient Non-Parametric Bandit Exploration. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[Baudry et al., 2021a] Baudry, D., Russac, Y., and Cappé, O. (2021a). On limited-memory subsampling strategies for bandits. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*.

[Baudry et al., 2021b] Baudry, D., Saux, P., and Maillard, O. (2021b). From optimality to robustness: Dirichlet sampling strategies in stochastic bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[Chan, 2020] Chan, H. P. (2020). The multi-armed bandit problem: An efficient nonparametric solution. *The Annals of Statistics*, 48(1).

[Combes and Proutière, 2014] Combes, R. and Proutière, A. (2014). Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning (ICML)*.

[Degenne et al., 2020] Degenne, R., Shao, H., and Koolen, W. M. (2020). Structure adaptive algorithms for stochastic bandits. In *International Conference on Machine Learning (ICML)*.

[Kveton et al., 2019] Kveton, B., Szepesvári, C., Ghavamzadeh, M., and Boutilier, C. (2019). Perturbed-history exploration in stochastic multi-armed bandits. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*.

[Kveton et al., 2020] Kveton, B., Zaheer, M., Szepesvári, C., Li, L., Ghavamzadeh, M., and Boutilier, C. (2020). Randomized exploration in generalized linear bandits. In *AISTATS*.

[Lattimore and Szepesvári, 2017] Lattimore, T. and Szepesvári, C. (2017). The end of optimism? an asymptotic analysis of finite-armed linear bandits. In *AISTATS*.

[Magureanu et al., 2014] Magureanu, S., Combes, R., and Proutière, A. (2014). Lipschitz Bandits: Regret lower bounds and optimal algorithms. In *Proceedings on the 27th Conference On Learning Theory*.

[Pesquerel et al., 2021] Pesquerel, F., Saber, H., and Maillard, O. (2021). Stochastic bandits with groups of similar arms. In *Advances in Information Processing Systems (NeurIPS)*.

[Riou and Honda, 2020] Riou, C. and Honda, J. (2020). Bandit algorithms based on thompson sampling for bounded reward distributions. In *Algorithmic Learning Theory (ALT)*.