Contextual bandits to help patient follow-up

Emilie Kaufmann, Odalric-Ambrym Maillard, Timothée Mathieu, Philippe Preux

Univ. Lille, Inria Scool

Where and When? 4 to 6 months internship (spring-summer 2024) at Scool, Inria Lille, Villeneuve d'Ascq, France.

Expected Background: Master in statistics/data science/machine learning. Prior knowledge of bandit algorithms is a plus but not mandatory. The candidate should have some experience working with data (e.g. using Python).

Keywords: sequential decision making, contextual bandits, precision medecine.

1 Context

This internship offers is within the ANR project BIP-UP (Bandits Improve Patients follow-UP) between Inria Lille and the Lille hospital, in which we seek to develop new machine learning tools to help patient follow-up, with a focus on the particular use case of patients that have undergone bariatric surgery. Bariatric surgery is a medical procedure to help individuals with obesity lose weight by making changes to their digestive system, and is known to require a long follow-up period to avoid relapse or complications. Using a "large" ($n = 1300$) database available at the hospital, we developed of a first model to predict the weight loss after surgery, based on decision trees (Saux et al., 2023). We now want to take a step further and consider the construction of decision support tools, possibly relying on this prediction model. We hope to leverage the rich literature on *contextual bandits* for this purpose.

Contextual bandits Bandit models (Lattimore and Szepesvari, 2019) are powerful tools for sequential decision making. In particular in contextual bandits (Li et al., 2010), a context x_t is revealed at time t (e.g. a patient descriptor), an action a_t is chosen at time t and a *reward* r_t, depending on both the context and the action is received. The goal is to design an action selection mechanism that maximizes the total reward received, or to find a good policy, that is a mapping from context to action that yield large reward (or small cost), on average. Contextual bandits have been extensively studied for application to recommender systems or the display of advertising (Chapelle and Li, 2011). More recently, they started to be considered for applications to mobile health, in which the goal is to adaptively propose interventions to patients using a digital application in order to maintain a desired healthy behavior (e.g. exercising or stop smoking), see, e.g., (Yom-Tov et al., 2017; Tomkins et al., 2021). In our context, the kind of interventions we seek to propose is different: adaptively decide when to schedule the next visit of a patient, in order to both minimize his or her well-being and to save doctor's time.

2 Goals

Patient follow-up for of bariatric surgery Together with the medical team, our goal is to propose a new adaptive follow-up protocol based on a system in which patients having had bariatric surgery would be asked to input (e.g. on a weekly basis) information about their weight and their well-being (e.g. declare some unpleasant digestive symptoms, or report their glycemy in case of diabetes). After receiving the information the system could produce some suggestions about when a medical visit (to a general practicioner, or to the hospital) is needed. The goal of the learning phase is twofolds: treat the patients well during the study (maximize some appropriately defined notion of reward) and find a good policy, that is a good decision rule that maps the current patient states to when its ideal next visit should happen. This policy could be in a subsequent randomized trial compared to the standard of care (a pre-defined timing of the visits fixed after the surgery) in order to access its benefit for the patient.

Objectives of the internship Besides the practical challenges (defining the appropriate context, actions, and rewards in collaboration with the medical team), the goal of this internship is to tackle different problematic related to the general use of contextual bandits:

- The objective of maximizing rewards and finding a good policy (i.e. a good arms) are known to be antagonistic for classical (i.e. non-contextual) bandit problems (Bubeck et al., 2011). How can we achieve a good trade-off between both objectives in the contextual case?
- Our weight predictor has the nice feature of being interpretable as it consists of a decision tree. Can we find a good policy for the more complex visit prediction task among decision trees? How to learn it online?
- In the learning phase, either doctors (double checking whether visits should indeed occur) or patients could be non-compliant. How to we take that into account to still learn a good policy?
- Can we transfer a good policy learnt for a certain population of patients (context distribution) to another?

3 Practical information

The internship will take place in the Scool team at Inria Lille in spring-summer 2024. Scool is a growing team with seven permanent researchers and around 15 PhD students and post-docs, working on sequential decision making (adaptive testing, bandits and reinforcement learning). There is a possibly to continue as a PhD student within the BIP-UP project. To apply, please send your resume to `emilie.kaufmann@inria.fr` and `philippe.preux@inria.fr`.

References

Bertsimas, D., Klasnja, P. V., Murphy, S. A., and Na, L. (2022). Data-driven interpretable policy construction for personalized mobile health. In *ICDH*, pages 13–22. IEEE.

Bubeck, S., Munos, R., and Stoltz, G. (2011). Pure Exploration in Finitely Armed and Continuous Armed Bandits. *Theoretical Computer Science 412, 1832-1852*, 412:1832–1852.

Chapelle, O. and Li, L. (2011). An empirical evaluation of Thompson Sampling. In *Advances in Neural Information Processing Systems*.

Kaufmann, E., Ménard, P., Domingues, O. D., Jonsson, A., Leurent, E., and Valko, M. (2021). Adaptive reward-free exploration. In *Algorithmic Learning Theory (ALT)*.

Krishnamurthy, S. K., Zhan, R., Athey, S., and Brunskill, E. (2023). Proportional response: Contextual bandits for simple and cumulative regret minimization. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Lattimore, T. and Szepesvari, C. (2019). *Bandit Algorithms*. Cambridge University Press.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *WWW*.

Pace, A., Chan, A. J., and van der Schaar, M. (2022). POETREE: interpretable policy learning with adaptive decision trees. In *ICLR*.

Saux, P., Bauvin, P., Raverdy, V., Teigny, J., Verkindt, H., Soumphonphakdy, T., Debert, M., Jacobs, A., Jacobs, D., Monpellier, V., Lee, P. C., Lim, C. H., Andersson-Assarsson, J. C., Carlsson, L. M. S., Svensson, P., Galtier, F., Dezfoulian, G., Moldovanu, M., Andrieux, S., Couster, J., Lepage, M., Lembo, E., Verrastro, O., Robert, M., Salminen, P., Mingrone, G., Peterli, R., Cohen, R. V., Zerrweck, C., Nocca, D., Roux, C. W. L., Caiazzo, R., Preux, P., and Pattou, F. (2023). Development and validation of an interpretable machine learning-based calculator for predicting 5-year weight trajectories after bariatric surgery: a multinational retrospective cohort SOPHIA study. *The Lancet Digital Health*, 5 (10).

Tomkins, S., Liao, P., Klasnja, P. V., and Murphy, S. A. (2021). IntelligentPooling: Practical Thompson sampling for mHealth. *Machine Learning*, 110(9):2685–2727.

Yom-Tov, E., Feraru, G., Kozdoba, M., Mannor, S., Tennenholtz, M., and Hochberg, I. (2017). Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. *Journal of medical Internet research*, 19(10).