Contextual bandits to help patient follow-up

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

Inria Scool, Lille

Context These internship offers are 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.

A simplified model Each patient is described by a context vector $x \in \mathbb{R}^d$ ($d = 7$ in our current predictive model) obtained before the surgery ($t = 0$). This model can be used to output a "nominal weight," describing the patient's ideal weight evolution after a surgery over a 10-year period. We denote $t \mapsto f_x(t)$ as the ideal weight curve associated with patient x. The objective is threefold: to keep the patient's weight curve close to the ideal curve, avoid complications (often correlated with either too low or too high weight), and prevent "unnecessary" interventions. Different types of interventions can be considered:

- Proposal of the next hospital follow-up visit date.
- Asking the patient to send their weight value at a specific time.

We focus on a simplified decision problem at $t = 1$ year, where a doctor sees the patient and observes potential additional variables, including the weight at $t = 1$ year. Let $\overline{x}$ be this "extended context" available at the 1 year visit. Given $\overline{x}$, we can choose one of the following actions:

- $a = 0$: Ask the patient to come back at $t = 5$ years.
- $a = 1$: Ask the patient to come back at $t = 5$ years and to send their weight at $t = 2$ years.
- $a = 2$: Ask the patient to come back at $t = 2$ years and $t = 5$ years.

In all cases, the cost of the action for a patient $\overline{x}$ will be observed at $t = 5$ years and will be a weighting (to be defined) of three things:

- The difference in weight at 5 years from the nominal weight $f_x(5)$
  (or an indicator that we are "too far" from this weight).
- A binary variable indicating whether complications occurred between 1 and 5 years
  (or a continuous variable taking into account the "cost" of complications).
- The cost of the visit at 2 years if $a = 2$ or the cost of the weight collection intervention if $a = 1$.

Let $c(\overline{x}, a)$ be the cost of action a for patient $\overline{x}$ (which is actually a random variable and depends on the patient's evolution given intervention a). Two types of objectives could be considered:

1. Learn a policy $\pi : \overline{\mathcal{X}} \to \{0, 1, 2\}$ that proposes an action to each patient, minimizing the average cost $C_\pi = \mathbb{E}[c(\overline{x}, \pi(\overline{x}))]$. The expectation is over the distribution of $\overline{x}$ (which can be estimated or simulated using our data) and the distribution of $c(\overline{x}, a)$ given $\overline{x}$ (as these interventions were not performed, estimating this will be challenging, but we can try to observe what happens on "realistic" simulators, which is the topic of the first internship described below). We could also take into account some more risk-averse criterion.
2. Simulate a sequence of patients $n = 1, 2, ...$ When patient $\overline{x}_n$ reaches 1 year, make a decision a_n to minimize the sum of costs $\sum_{n=1}^{N} c(\overline{x}_n, a_n)$. This is closer to the classic contextual bandit objective, but here we need to take into account the quite long delay between the choice of the action and the observation of the cost. Safety criteria may also be considered.

First internship topic (advisors: Odalric-Ambrym Maillard and Timothée Mathieu) We will develop a realistic simulator for $c(\overline{x}, a)$ based on the weight predictor (possibly adding certain features, such as updating the prediction with intermediate observations) to estimate the optimal policy defined in 1. The cost function, as envisaged, essentially depends on two things: weight and the presence or absence of complications. We propose simulating them as follows:

- For $a = 2$, simulate a possible "natural" trajectory using the weight predictor, which is built from patients who have had all visits (mostly).
- For $a = 0$, with probability $\alpha_{\overline{x}}$, simulate a "natural" trajectory, with probability $1 - \alpha_{\overline{x}}$, add a drift (positive or negative, depending on the patient type?).
- For $a = 1$, with probability $\beta_{\overline{x}}$, simulate a "natural" trajectory, with probability $1 - \beta_{\overline{x}}$, add a drift.
- State that a complication occurs with a certain probability if the weight falls outside a certain window or state that the probability of complication increases with the distance of the weight at 5 years from the ideal weight.

The goal of the internship is to construct and to use this simulator to find an optimal policy. Finding a good policy means sequentially simulating a patient, choosing an intervention, simulating the associated cost, and stopping when we think we have found a good policy. One difficulty here is that the context space ($\mathcal{X}$) is continuous. Given the difficulty to design a good cost function, we may also be interesting to solve a "cost-free exploration" problem, i.e., perform enough trials to find the optimal policy for *any* cost function given after exploration (i.e., corresponding to different weighting of the different criteria). This relates to the problem of reward-free exploration in reinforcement learning (Kaufmann et al., 2021).

Contact: `odalric.maillard@inria.fr`, `timothee.mathieu@inria.fr`

Second internship topic (advisors: Emilie Kaufmann and Philippe Preux) The second project is to review the existing literature on mobile health applications and to see whether it can be useful to tackle the challenges of our applications. In particular, we will investigate the following things:

- taking into account potentially large delays between the intervention and the observation of the cost/reward
- leveraging intermediate observations (enrich the model with patients who send their weight value every year, for example, to refine the estimation of the reward we have not yet received).
- considering alternative formulation with explicit safety constraints (minimize cost under constraint that there is no complications)
- can you simultaneously minimize the cumulative cost and learn a good policy?

We will try to keep a focus on building interpretable decision rules, following the decision tree approach that was taken for our weight predictor. Some recent papers have taken a similar approach for different health applications (Bertsimas et al., 2022; Pace et al., 2022).

Contact: `emilie.kaufmann@univ-lille.fr`, `philippe.preux@inria.fr`

Practical information The internship will take place in the Scool team at Inria Lille. 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 are some opportunities to start a PhD in the team in 2024, including one on the BIP-UP project.

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.

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)*.

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).