Real-world bandit applications: Bridging the gap between theory and practice

Audrey Durand

McGill

EWRL 2018
The bandit setting
(Thompson, 1933; Robbins, 1952; Lai and Robbins, 1985)

Flow

1. (Receive context)
2. Select action
3. Observe reward
4. Go back to step 1
Structure in bandit problems

Contextual bandits
Reward = function of context

Structured bandits
Reward = function of action

Audrey Durand
Real-world bandit applications
Capturing structure with kernel regression

Non-linear $f(x)$

$$f(x) = \varphi(x)^\top \theta$$

Kernel $k(x, x') = \varphi(x)^\top \varphi(x') \quad \forall x, x' \in \mathcal{X}$
Posterior model

Gathering observations $y_1, y_2, \ldots, y_N$ at locations $x_1, x_2, \ldots, x_N$

Assuming that
\[ y_i = f(x_i) + \xi_i \text{ with } \xi_i \sim \mathcal{N}(0, \sigma^2) \]
\[ \theta \sim \mathcal{N}_d(0, \Sigma) \text{ with } \Sigma = \frac{\sigma^2}{\lambda} I_d \text{ for } \lambda > 0 \]

$\Rightarrow$ Gaussian posterior on $f$

\[
\hat{f}_{\lambda, N | x_1, \ldots, x_N, y_1, \ldots, y_N} \sim \mathcal{N}\left(\left[f_{\lambda, N}(x)\right]_{x \in \mathcal{X}}, \frac{\sigma^2}{\lambda} \left[k_{\lambda, N}(x, x')\right]_{x, x' \in \mathcal{X}}\right)
\]

with

\[ f_{\lambda, N}(x) = k_N(x)\top(K_N + \lambda I_N)^{-1}y_N, \]
\[ k_{\lambda, N}(x, x') = k(x, x') - k_N(x)\top(K_N + \lambda I_N)^{-1}k_N(x') \]
Gaussian Process (GP)
(Rasmussen and Williams, 2006)

- For $\lambda = \sigma^2$, i.e. assuming $\theta \sim \mathcal{N}_d(0, I_d)$

(a) $N = 1$
(b) $N = 10$
(c) $N = 50$

- Theoretical guarantees in the streaming setting
- Many applications in bandits, e.g. Kernel UCB, Kernel TS, CGP/GP-UCB, GP-TS
Adaptive treatment allocation for mice trials as contextual bandits

Joint work with Joelle Pineau, Georgios D. Mitsis, Katerina Strati, Charis Achilleos, and Demetris Iacovides
@ Machine Learning for Healthcare (MLHC) 2018
Collecting data for personalized medicine

Treatment adapted to the disease

Which treatment should be allocated to patients with cancer given the stage of their disease?

\[
v = \frac{\pi}{6} (lw)^{3/2} \quad \text{(Tomayko and Reynolds, 1989)}
\]

Data acquisition setup

- Mice with induced cancer tumours
- Options: none, 5-FU, imiquimod, and imiquimod + 5-FU
- Natural death or sacrifice when tumour volume is critical
Initial data acquisition

RCT (Randomized Clinical Trial)

- 6 mice, 163 triplets \((v_i, a_i, v'_i)\)
- Quick degradation of subjects
- Limited covering of the tumour volume space

(a) Tumour evolution

![Graph showing tumour evolution over time](image)

(b) Tumour volume distribution

![Graph showing tumour volume distribution](image)
Adaptive allocation

ACT (Adaptive Clinical Trial) (Thompson, 1933)

- Favor selection of better treatments
- Reduce the exposition of less effective treatments

Contextual bandits

- Context: tumour volume $x_t$
- Actions: 4 possible treatment options
- Treatment effect: post-treatment tumour volume $x'_t$
- Reward: $y_t = -x'_t$
Select next action given the context

BESA (Best Empirical Sampled Average) (Baransi et al., 2014)

- Fair comparison of empirical estimators
- Opportunities for action to show how good they are

GP BESA for two actions
Parameters: context \(x_t\), two actions \(a\) and \(b\)

- Let \(N_{t-1}^{(a)} < N_{t-1}^{(b)}\)
- Subset \(S_t^{(a)} \leftarrow \) all observations from \(a\) history
- Subset \(S_t^{(b)} \leftarrow \) subsample of \(N_t^{(a)}\) from \(b\) history
- Compute posterior mean \(\tilde{f}_t^{(a)}\) using \(S_t^{(a)}\)
- Compute posterior mean \(\tilde{f}_t^{(b)}\) using \(S_t^{(b)}\)
- Select \(a_t = \arg\max_{i \in \{a, b\}} \tilde{f}_t^{(i)}(x_t)\)
Regression on subsets of observations

(a) \( N = 5 \)

(b) \( N = 10 \)

(c) \( N = 20 \)

(d) \( N = 50 \)

Audrey Durand
Real-world bandit applications
Animal experiments

Delayed feedback
Update the history upon the completion of a group

Group A mice 1 and 2
Group B mice 3, 4 and 5
Group C mice 6 and 7
Group D mice 8, 9 and 10

Baseline strategies for comparison

- No treatment (3 mice)
- Random (5 mice)
- 5-FU (4 mice)
Recall: The algorithm is updated after each group.
Animals live longer

(a) Per approach

(b) Per group of GP BESA

Preliminary results

- GP BESA median > 50% 5-FU median
- Visible improvement after each group

To validate on a larger cohort
A better state space covering

(a) Random allocation

(b) Adaptive allocation

Using data in a next phase

- More information on the tumor growth process
- 40% more data points of volume $>70\text{mm}^3$
Contextual bandits for adaptive clinical trials work well!

Even though:

- Contexts were not independent
- The model was (over) simplistic
- Rewards were delayed
- We lost guarantees on the regression model
Theory vs Practice

In theory...
▶ The kernel is well adapted to the function
▶ The noise variance $\sigma^2$ is known

In practice...
▶ Sometimes the kernel is picked from expert knowledge
▶ Sometimes the kernel is tuned using previous data
▶ Sometimes $\sigma^2$ is learned from previous data
▶ Sometimes an upper bound $\sigma^2_+ \geq \sigma^2$ is used
Theory vs Practice

In theory...

- The kernel is well adapted to the function
- The noise variance $\sigma^2$ is known

In practice...

- Sometimes the kernel is picked from expert knowledge
- Sometimes the kernel is tuned using previous data
- Sometimes $\sigma^2$ is learned from previous data
- Sometimes an upper bound $\sigma_+^2 \geq \sigma^2$ is used

Theory holds with this!
Impact of noise upper bound on regression model

Using $\sigma^2$
Empirical noise variance estimate

- Confidence intervals on empirical variance (Maillard, 2016)
- Example with $\sigma = 0.1$:

![Confidence intervals graph]

- Use these intervals to adapt the regularization ($\lambda$)
Example: Adapting $\lambda_t$ vs fixed $\lambda$ (unknown $\sigma$)

$$f(\lambda) = \sigma^2 + \frac{1}{C^2}$$

$$\lambda_t = \sigma^2_{+,t-1} + \frac{1}{C^2}$$
Theory can hold with more realistic assumptions

- Requires initial lower/upper bounds $\sigma_- / \sigma_+$ instead of $\sigma$
- Preserves guarantees on the resulting regression model
- Tightest confidence intervals than using fixed $\sigma_+$

All details in Durand, Maillard, Pineau (JMLR 2018)
Optimizing super-resolution imaging parameters as multi-objective structured bandits

Joint work with Flavie Lavoie-Cardinal, Paul De Koninck, Christian Gagné, Theresa Wiesner, Marc-André Gardner, Louis-Émile Robitaille, and Anthony Bilodeau
Observing structures at the nanoscale
(Hell and Wichmann, 1994)
Optimizing imaging parameters

Biology: The optimal parameters are not *always* the same

Find good parameters *during* the imaging task

- Maximize the acquisition of images useful to researchers
- Minimize trials of *poor* parameters

Structured bandits

- Action space: Space of parameters that can be tuned
- Reward function defined over the action space
Trade-off between different objectives

Conflicting objectives

- Maximize image quality (structure visible)
- Minimize photobleaching (structure maintained after imaging)

⇒ Include an expert in the learning loop

Multi-objective bandits

![Diagram of multi-objective bandits with nodes labeled Agent, Expert, Environment, Estimates, Action, and Rewards, and arrows indicating the flow of information.]
Producing estimates

Thompson Sampling (Thompson, 1933)
Decision based on sampling from the posterior distribution

Kernel TS in the multi-objective loop

- Model each objective function $f_{*,i}$ using kernel regression
- Sample $\tilde{f}_{t,i}$ from the posterior distribution on $f_{*,i}$
- Estimate at parameter configuration $x = (\tilde{f}_{t,i}(x))_{\forall i}$
  E.g. (sampled imaging quality, sampled photobleaching)
Presenting estimates to user

- Expert acts as an argmax on the preference function
Experiments on neuronal imaging: two objectives

Experimental conditions

- Three parameters (1000 configurations):
  - Excitation laser power
  - Depletion laser power
  - Duration of imaging per pixel

- Two objectives:
  - $f_{*,1}$: image quality (structure visible)
  - $f_{*,2}$: photobleaching (structure maintained after imaging)

- Image quality feedback provided by expert
Example: Image quality of actin rings

Poor images

Good images
Logarithmic regret, as expected

Figure 3: Multi-objective optimization of microscopy parameters for repeated STED imaging of LifeAct-GFP: a comparison across different cell types. a) Parameter configurations selected by Kernel TS during different imaging trials in neuronal (green), PC12 (blue), and HEK293 (orange) cells. b) Cumulative regret curve of (left) image quality alone (images with a quality score below 60%) and (right) image quality and photobleaching (images with a quality score below 60% or photobleaching above 75%). c) Example images obtained among the last 10 images of one optimization sequence for each cell type. The confocal image was taken before two consecutive STED images (labeled as STED-1 and STED-2). Note the differences in intensity scales across images to reflect differences in fluorescence intensity (confocal images) or photobleaching (STED1 vs STED2). Scale bar 1µm

- Neuron: Rat neuron
- PC12: Rat tumor cell line
- HEK293: Human embryonic kidney cells
Better images without manual tuning

Figure 3: Multi-objective optimization of microscopy parameters for repeated STED imaging of LifeAct-GFP: a comparison across different cell types. a) Parameter configurations selected by Kernel TS during different imaging trials in neuronal (green), PC12 (blue), and HEK293 (orange) cells. b) Cumulative regret curve of (left) image quality alone (images with a quality score below 60%) and (right) image quality and photobleaching (images with a quality score below 60% or photobleaching above 75%). c) Example images obtained among the last 10 images of one optimization sequence for each cell type. The confocal image was taken before two consecutive STED images (labeled as STED-1 and STED-2). Note the differences in intensity scales across images to reflect differences in fluorescence intensity (confocal images) or photobleaching (STED1 vs STED2). Scale bar 1 µm.
Fully automatized loop!

Kernel TS → Sampled options → Preference articulation → Action

Feedback

Optical microscope

Images

FCN + online analysis
Logarithmic regret again, without expert in the loop

- **Bassoon**: Protein from the nerve terminals active zone
- **Tubulin**: Protein of cytoskeleton
- **Actin**: Protein involved in cell motility and establishment/maintenance of cell junctions/shape
Fully automated imaging results

Actin-STAR635

Tubulin-STAR635P

Bassoon-STAR635P

STED - Image #1
Quality FCN : 0.57

STED - Image #1
Quality FCN : 0.19

STED - Image #1
Quality FCN : 0.22

Confocal

Audrey Durand  Real-world bandit applications  32
Structured bandits for online imaging parameters tuning with multiple objectives work well!

Even though:

- We may have experts in the loop
- We may even have neural networks in the loop
Take home

- Bandits have great applications
- Theory is important: we should make it applicable
Thanks!

Collaborators:

Joelle Pineau
Odalric-Ambrym Maillard
Christian Gagné

Georgios D. Mitsis
Katerina Strati
Demetris Iacovides
Charis Achilleos

Flavie-Lavoie Cardinal
Paul De Koninck
Theresa Wiesner
Marc-André Gardner
Louis-Émile Robitaille
Anthony Bilodeau