UNIVERSITY OF GOTHENBURG

CHALMERS UNIVERSITY OF TECHNOLOGY

Bayesian Analysis of Combinatorial Gaussian Process Bandits

With Applications to Energy-Efficient Navigation

Jack Sandberg[1] Niklas Åkerblom[12] Morteza Haghir Chehreghani[1]

[1]Chalmers University of Technology and University of Gothenburg [2]Volvo Car Corporation

ICLR FFI WASP WALLENBERG AI AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

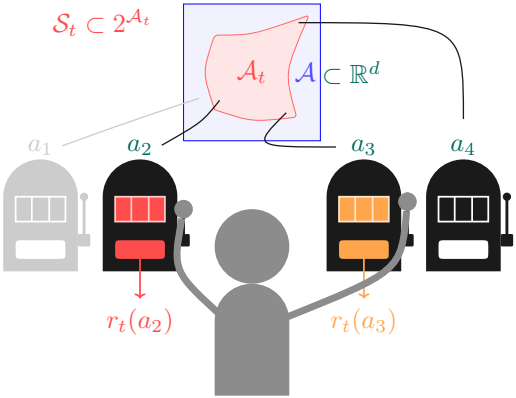
Online energy-efficient navigation

Map data from OpenStreetMap: www.openstreetmap.org/copyright

- Drive from A to B using minimal energy.
- *Online*: learn whilst collecting data.
- Minimize total energy used over T steps.
- *Goal:* Speed up learning and use less energy $\implies$ Increased effective range.
- *How:* Add contextual information using Gaussian processes.

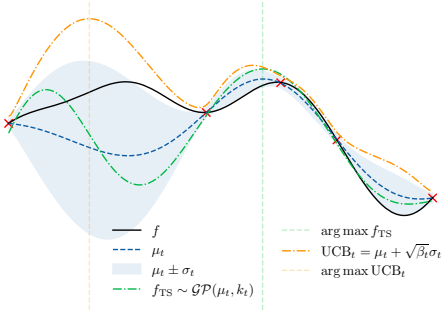
Combinatorial volatile Gaussian process semi-bandit

- Sequential decision making under uncertainty.
- Maximize total reward.
- Combinatorial: Select combination of arms (e.g. a path in the road network).
- Volatile: Time-varying arm availability.
- Contextual $\subset$ Volatile.
- Gaussian process: Reward depends smoothly on the arm features.



$$\mathbb{E}[R_t] = f(a_2) + f(a_3), f \sim \mathcal{GP}(\mu, k)$$

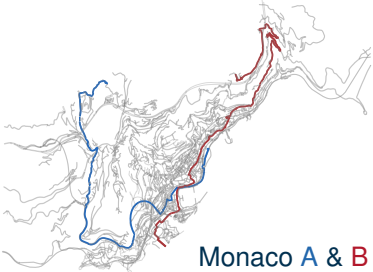
Contributions



- Extend previous Bayesian regret bounds for GP-UCB and GP-TS to the *infinite, volatile* and *combinatorial setting*.
- First regret bound for GP-BayesUCB (Kaufmann et al., 2012; Nuara et al., 2018).
- Develop a combinatorial and contextual bandit framework for the online energy-efficient navigation problem.

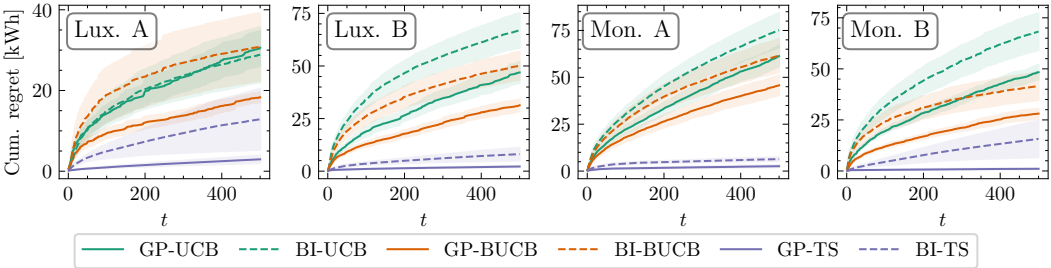
Experiments: Networks and baselines

- Real-world road networks of two European cities.
- 5169 and 3474 edges.
- Baselines: Bayesian inference (BI) with UCB, BayesUCB and TS (Åkerblom et al., 2023)



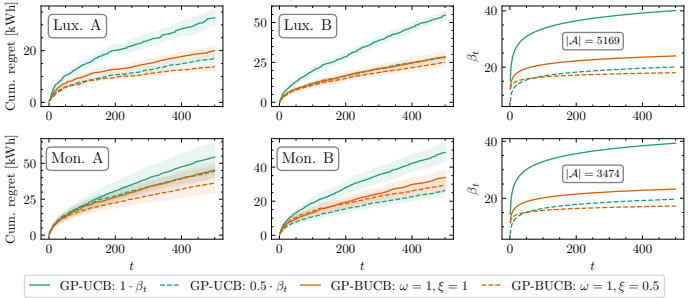
Map data from OpenStreetMap: www.openstreetmap.org/copyright

Experiments: Algorithms



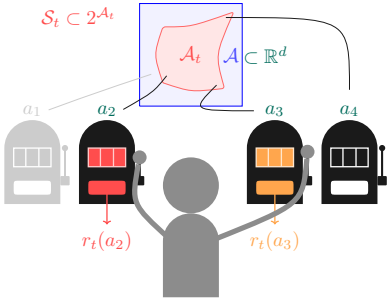
- Contextual features reduce regret and lowers the energy consumption.
- GP-TS has lowest regret.
- Theoretical confidence parameter β_t large $\implies$ Excessive exploration for GP-UCB and GP-BUCB.

Experiments: β_t parametrization

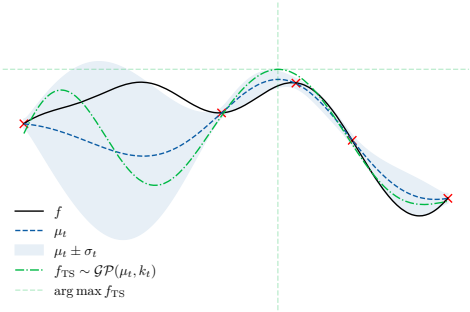


- GP-UCB and GP-BUCB differ in parametrization of confidence parameter β_t.
- Theoretical results for GP-BUCB holds for wider choices of β_t.
- Experiment with tuning β_t.
- GP-BayesUCB is more flexible than GP-UCB.

Conclusion



$\mathbb{E}[R_t] = f(a_2) + f(a_3), f \sim \mathcal{GP}(\mu, k)$



References

Niklas Åkerblom, Yuxin Chen, and Morteza Haghir Chehreghani. Online learning of energy consumption for navigation of electric vehicles. *Artificial Intelligence*, 317:103879, April 2023.

Emilie Kaufmann, Olivier Cappe, and Aurelien Garivier. On Bayesian Upper Confidence Bounds for Bandit Problems. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, pp. 592–600. PMLR, March 2012.

Alessandro Nuara, Francesco Trovò, Nicola Gatti, and Marcello Restelli. A Combinatorial-Bandit Algorithm for the Online Joint Bid/Budget Optimization of Pay-per-Click Advertising Campaigns. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), April 2018.