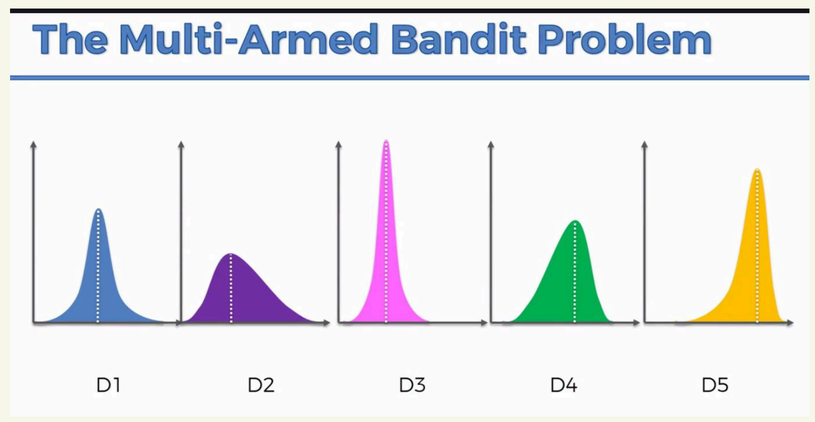
Multi-Armed bandits

Multi-armed bandits are a class of decision-making problems with a wide variety of applications. But the efficiency of these algorithms varies based on the environment. In this project we look at a few algorithms and compare their performance in non-linear kernelized environments with noisy observations.

Author
M.K. Herrebout

Affiliations
A thesis submitted to EEMCS faculty
Delft University of Technology
Supervisor: J. Olkhovskaia



Introduction

Multi-Armed Bandit problems have the goal of optimizing the trade-off between exploration and exploitation of all choices. Each decision yields some reward, and the goal is to minimize the regret that follows from a combination of decisions, that is to say to minimize the difference between the set of decisions made, and the set of optimal decisions. In particular, these algorithms deal with the trade-off between choosing the best-known option and exploring new, possibly better options. These algorithms are widely used in reinforcement learning, optimization and economics, where decisions need to be made without all the information and with some uncertainty.

Objective

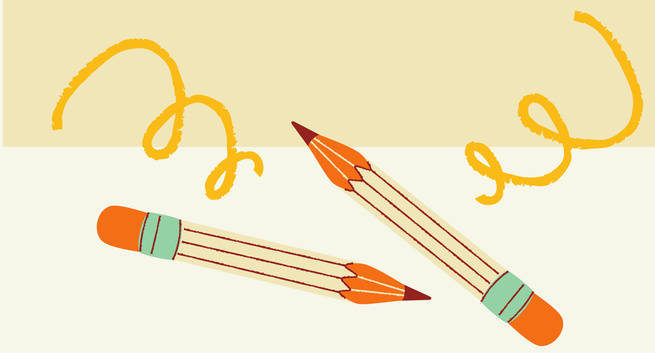
This project aims to compare between and draw conclusions about different algorithms for a specific environment in MAB problems.

Methodology

We used the SMPyBandits library for the implementation of multi-armed bandits and extend the framework for a tailored kernelized algorithm, KernelUCB. We compare the results by plotting the regret of each algorithm over a fixed timespan for several different reward functions.

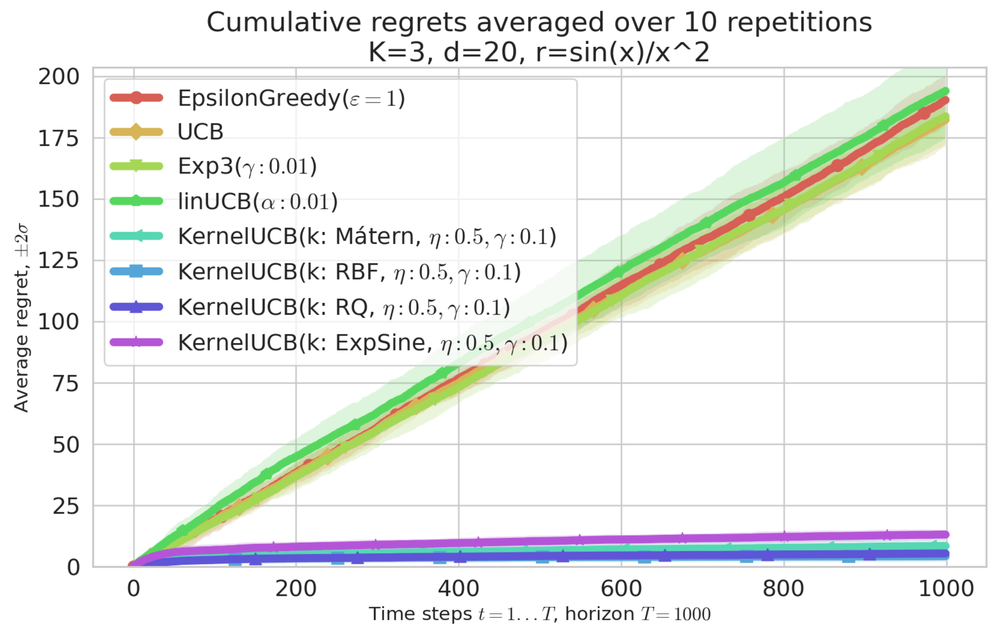
Results

As expected, UCB, Exp3 and random choice never converge in non-linear environments, but KernelUCB almost always does. Some kernel functions are better in some environments than others though. Furthermore, LinUCB sometimes performs well, and even better than KernelUCB in non-linear non-periodic environments.



Analysis

UCB, Exp3 and random choice perform poorly in all of the non-linear environments that were tested. Surprisingly, LinUCB sometimes still does well and even better than KernelUCB. KernelUCB does well in almost all non-linear settings though, and especially in periodic environments with a sine or cosine term. KernelUCB also does not always converge though, and it is not always the best choice. This really depends on the environment, and needs to be judged on a case-by-case basis; there is no one clear winner that performs best for all environments.



Conclusion

KernelUCB is a powerful algorithm that works well for a lot of non-linear environments, but it is not perfect. There are more kernel functions to try, and there are more MAB algorithms to try depending on the setting of the problem. Nonetheless, this research has proven that KernelUCB is a good contender in non-linear scenarios.

Related literature

1. Tor Lattimore and Csaba Szepesvári. Bandit algorithms. Cambridge University Press, 2020
2. P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. SIAM journal on computing, 32(1):48–77, 2002
3. W. Chu, L. Li, L. Reyzin, and R. Schapire. Contextual bandits with linear payoff functions. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 208–214. JMLR Workshop and Conference Proceedings, 2011
4. B. Sch¨olkopf and A. J. Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2002

Acknowledgements

- Image credits top header: https://stats.stackexchange.com/questions/326449/multi-armed-bandit-problem?ref=mlq.ai

Created by Gan Khoon Lay from Noun Project