Multi-armed Bandit Algorithms (MAB) for Price Optimization in Retail

Jeffery Tang, Aman Sidhu, Sebastian Pilarski, Prof. Daniel Varro, Prof. Maxime Cohen

Electrical and Computer Engineering, Systems Engineering Lab, McGill University + Bensadoun School of Retail Management, Retail Innovation Lab, McGill University

Motivation

Improve pricing in retail to provide consumers with more attractive prices, improve business profits, and reduce food waste

MAB Algorithm

- The multi-armed bandits (MAB) is a type of reinforcement learning
- Based on its actions, the rewards it obtains can teach it to choose actions that maximize reward

- Types of MAB:
  - **Explore** – randomly choose your actions
  - **Exploit** – always choose what you think is the best option
  - **Greedy epsilon** – exploit but every so often, you explore
  - **Contextual** – choose actions based on the context. E.g. pricing strawberries based on season, inventory, demand, etc.

Slot machines used to be called one-arm bandits.

Turbodega

We thank Turbodega for offering us their 2021-2022 product and store datasets. They also agreed to be a case study for applying contextual information such as geographical insights into a bandits setting.

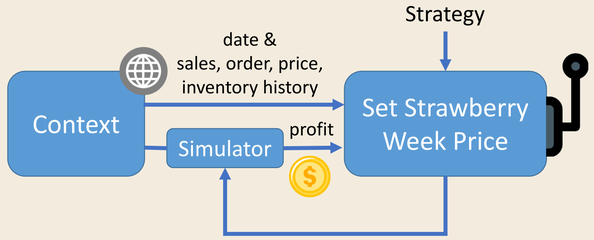
Research Focus

1) Explore MAB algorithms for dynamic pricing in food retail

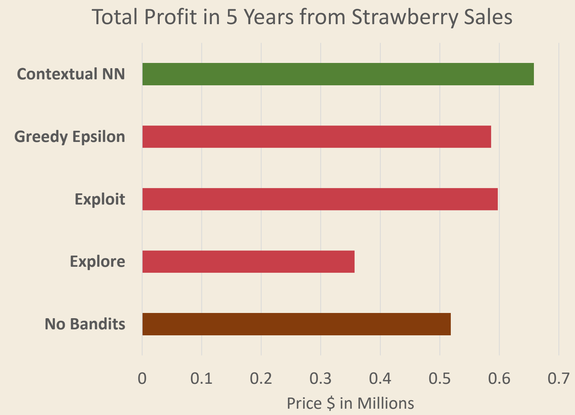
2) Extract insights from Turbodega data for case study

MAB Simulation

- Pilarski's demand simulator lets you set the price and order of certain fruits and vegetables weekly/daily and returns the profit
- Diagram of my MAB implementation hooked up to the simulator:

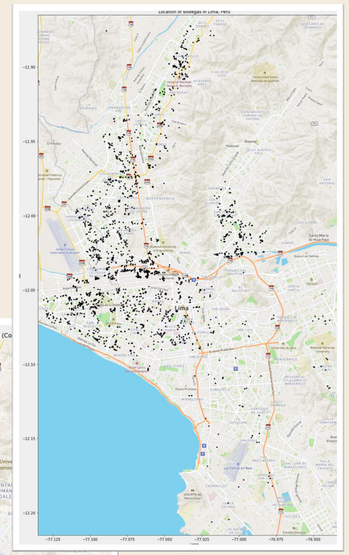


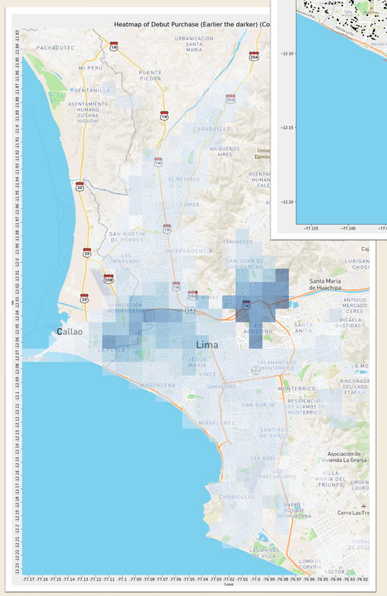
Profits averaged over a 100 runs for each strategy:



Geographical Insights

- Context can help MABs make more informed decisions
- For us, context includes the location, age, and spending of stores, some of it shown in these maps:





The top map shows the locations of stores that use Turbodega. For the left heatmap, the darker the tile, the older those stores are.

TURBODEGA

Conclusions

- Found anomalies, design flaws, common product categories, inactive stores, and important insights in Turbodega's dataset
- Improved upon baseline profits using MABs in Pilarski's demand simulator

Future Work

- Improve upon my MAB implementation, so that its approach is less greedy or exploratory and more contextual and consistent
- Apply my MAB to Turbodega's dataset
- Incorporate Sidhu's demand prediction in MAB

References

S. Pilarski, S. Pilarski and D. Varró, "Optimal Policy for Bernoulli Bandits: Computation and Algorithm Gauge," in IEEE Transactions on Artificial Intelligence, vol. 2, no. 1, pp. 2-17, Feb. 2021, doi: 10.1109/TAI.2021.3074122.

Cohen, Maxime C., et al. Demand Prediction in Retail a Practical Guide to Leverage Data and Predictive Analytics. Springer, 2022.

Sponsors

This research is financially supported by NSERC-USRA and Professor Varro through NSERC-RGPIN.