Real-time Bidding Optimization with Reinforcement Learning

Thao Nguyen
CS668 - Analytics Capstone Project - Fall 2024
Professor: Dr. Christelle Scharff
Pace University, Seidenberg School of Computer Science & Information Systems

GitHub

Abstract

Real-Time Bidding (RTB) is a programmatic advertising technique that enables advertisers to bid on individual ad impressions in real time as they become available. This project presents a reinforcement learning framework using Q-learning and exploration-exploitation technique to optimize bids in RTB. The results demonstrate the ability of Q-learning to adaptively learn optimal bidding strategies that maximize reward based on clicks and conversions.

Research Questions

- How can Q-learning be applied to optimize RTB bidding strategies?
- What is the impact of user demographics and ad-specific features on bidding outcomes?
- How effective is the learned bidding strategy in maximizing rewards?
- How can the challenges of large-scale data and state spaces be addressed in RTB bid optimization?

Related Work

Real-time bidding (RTB) presents numerous research challenges, particularly in bid optimization with budget limitations. Existing research on bid optimization in RTB explores various methodologies, including smart pacing methods, linear bidding functions, and learning-based approaches. Most studies focus on maximizing Click-Through Rate (CTR) or conversion probabilities but lack the integration of both in a unified reward system. To evaluate the effectiveness of different bidding strategies, researchers utilize publicly available datasets such as iPinYou [1], which provides a valuable resource for benchmarking and analysis.

Dataset

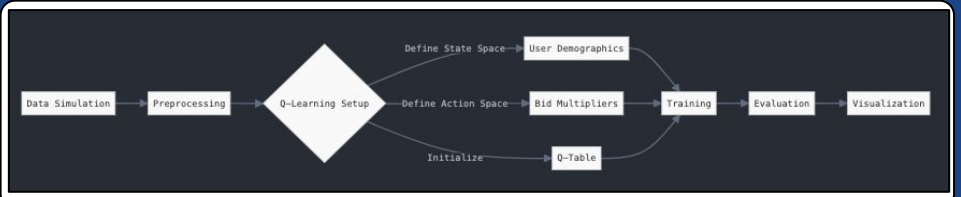
The 10,000 row-dataset "Online Advertisements Click-Through Rates" [by Mendeley Data, 2024] includes user demographics like age, gender, income, location; ad characteristics like types, topics, placements; and behavior metrics like clicks, click time, conversion rate. To simulate RTB and RL environment, I also added synthetic data columns like *Bid* and *Reward* (*Click* + 3×*Conversion*).

Methodology & Experiments

The preprocessing step involved simulating data with encoded categorical and scaled numerical variables for user and ad features. State spaces were defined by these features, while actions represented bid levels. Q-learning was employed to iteratively update *Q-values* for state-action pairs using *a reward function* integrating clicks and conversions to measure campaign effectiveness. The model was trained over 500 episodes using an ε-greedy policy to balance exploration and exploitation, and its performance was evaluated on unseen data to compute average rewards.

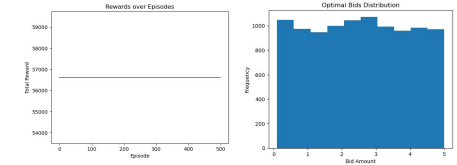
Q-values formula with α learning rate, γ discount factor, s state, a action:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma a' \max Q(s',a') - Q(s,a)]$$

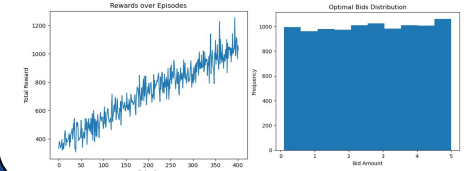


Results

First Iteration Results: The Q-learning agent was trained over 500 episodes, achieving a consistent total reward of 56,618 across all episodes, with an average evaluation reward of 5.6618. The bid distribution appeared uniform, suggesting limited differentiation in state-action mapping or reward function.



Second Iteration Results: Without inherent rule for "correctness", the Q-learning agent using heuristics successfully demonstrates improvement over the training episodes. Although rewards are not excellent, they consistently increase across episodes from 166 to 1029 over 500 episodes, reflecting gradual learning and adaptation to the environment.



Conclusion & Future Work

I implemented Q-learning algorithm for bid optimization in the context of RTB display advertising. The project involved extensive experimentation with various Q-learning parameters, allowing for a thorough exploration of the algorithm's performance under different configurations. Although the RL agent didn't have a good performance, I got a chance to have hands-on experience with RL, especially model-free Q-learning approach.

Challenges:

- Lack of public RTB data
- Lack of computational resources to train agent due to complex and data-intensive nature of RTB
- Model uncertainty and robustness

Future work could focus on some promising directions like:

- Alternative exploration-exploitation approach
- Integrate user profiles such as interests
- Investigate more sophisticated state representations

References

[1] Liao H, Peng L, Liu Z, Shen X. iPinYou Global RTB Bidding Algorithm Competition Dataset. Published online August 24, 2014. doi:10.1145/2648584.2648590

[2] Cai, H., Ren, K., Zhang, W., Malialis, K., Wang, J., Yu, Y., & Guo, D. (2017). Real-Time Bidding by Reinforcement Learning in Display Advertising. ARXIV. doi:10.1145/3018661.3018702

[3] *Real-Time Bidding (RTB) seen through the lens of machine learning research | Numberly*. (n.d.). Numberly. https://numberly.com/en/real-time-bidding-rtb-seen-through-the-lens-of-machine-learning-research/