Practicum Mid-term Report

Anushital Sinha
OMSA, Georgia Institute of Technology
Employer : Moksha.ai
Supervisor : Nabeel Siddiqi

The problem and proposed solution

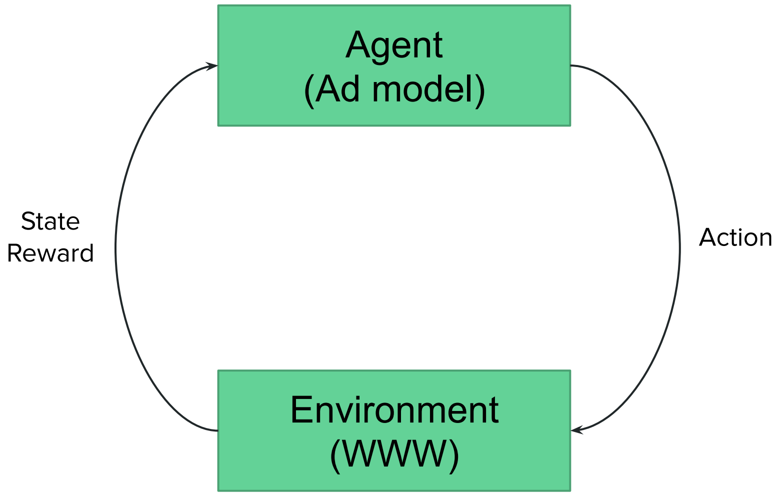
Problem: optimizing ad performance

- E-commerce operators face the problem of determining which digital ads they should continue to invest in versus which ads they should drop
- At every time-step (e.g. an hour or a day) a decision needs to be made which ad creative to run on which destination website given the hour of the day, day of week, week of year, destination website properties, creative properties, etc.

Solution: reinforcement learning

- The problem can be formulated to follow the Markov property, which states that the future is independent of the past given information in the present
- This means that we can formulate the solution to be non-temporal reinforcement learning
- This will also enable intelligent exploration and exploitation of the various options

What is reinforcement learning?



- Reinforcement learning works by having an **agent** (model deciding which ad to run) with the **environment** (the world-wide-web)
- At each **time-step** the **environment** provides the **state** (hour of day, historical ad performance, etc.) then, given the state, the **agent** chooses an **action** (an ad creative + destination website), then the environment returns a **reward** for the given action (click-through-rate of the ad) and the next state, finally the agent updates the model given the reward and chooses the next action given the new state

Why reinforcement learning?

1. Enables dynamic decision making in a changing environment
    - Digital advertising is a constantly changing landscape, an ad that performed well yesterday may not work well today, hence having a model that can change its mind is necessary
2. Enable balanced exploration and exploitation in problems with limited data
    - New ad creatives are introduced all the time, hence the need to minimize regret and figure out intelligently if a new ad will perform well is important, RL enables that to happen quickly thus saving precious time
    - As soon as an ad shows good performance it should be used as much as possible, but this needs to be balanced with the potentially changing landscape (see point 1)
3. If agent model is based in Bayesian techniques, it can incorporate prior knowledge/assumptions
    - Probabilistic programming techniques, when combined with RL, can be effective in situations where a more complex model is not needed
    - Probabilistic programming enables prior knowledge to be incorporated in the form of prior probability distributions or probability parameters

Implementation will consist of two components

Environment for simulation

- Environment will represent the WWW, in that it will determine the performance of an ad and provide information of the state e.g. hour of day, day of week, week of year, etc. and historical performance of destination websites and ads if available
- Environment will be implemented in the OpenAI gym library to enable a familiar format for reinforcement learning

Agent for decision-making

- The agent will decide which ad to run in the next time-step given the state and the reward
- The agent model is required to output not only a prediction but also some form of uncertainty, this is required to enable exploration of other options
- The agent model will be implemented with various algorithms, the current plan is:
    - Thompson Sampling
    - Bayesian Linear Regression
    - Bayesian Neural Networks

Agent models and measurement methodology

Agent Models

- Thompson Sampling: a simple model that ignores the state and only looks at the ad performance, it can perform well when very little information is available
- Bayesian Linear Regression: model of medium complexity that takes into account state but only in a linear relationship
- Bayesian Neural Networks: complex model that can represent very complicated relationship, however may not perform well with little information

Measurement Methodology

- Given that the environment is known in this project, we will be using **total regret** as the metric to compare agent models
- Total Regret = Best Reward - Current Reward
  - Best Reward: is the best reward possible, e.g. the CTR of the ad that would have performed the best given the current state
  - Current Reward: the reward received for the action chosen by the agent

Tasks accomplished so far

1. Literature Review on the following two areas:
    - Research on RL applied to Digital Advertising
    - Current practices of advertising and optimizing digital marketing
2. Understanding the concepts and applications of Reinforcement Learning
3. Understanding Probabilistic Programming
4. Understading Bayesian Neural Networks
5. Planning and honing the methodology for implementation and testing

> “Simons et al (2009) discuss that using reinforcement learning for digital advertising is not recommended because RL modes are high dimensional and that determining the complete customer trajectory takes days. This is why we are exploring Thompson Sampling as a low dimensional approach and focusing on CTR, which is immediately available as a metric”

Literature Review

Paper 1 : Luckeciano et al. (2021) discuss that using RL for digital advertising is not recommended because RL mode are high dimensional and determining the complete consumer trajectory takes days[1]

Implication: Due to the above reason, we are using Thompson Sampling as a low dimensional approach and focusing on CTR, which is immediately available as a metric.

Paper 2 : Chatterjee et al. (2003) - “The effect of repeated exposures to banner ads is negative and nonlinear, and the differential effect of each successive ad exposure is initially negative, though nonlinear, and levels off at higher levels of passive ad exposures”[2].

Implication: When simulating the click through rates, we wil leverage the findings from this paper.

Literature Review

Paper 3 : Richardson et al. (2007)- “ We can use features of ads, terms, and advertisers to learn a model that accurately predicts the click- through rate for new ads”[3].

Implication : Accurate CTR prediction is important to building an accurate RL model policy and this paper shows that using ad features improves this

Paper 4 : Zhou et al. (2018) discuss that when an advertiser goes to a web publisher to place an ad, they need to have a prediction of the Click Through Rate(CTR) in place to accurately provide appropriate inventory of ad spots to an advertiser. The paper is using uses a deep neural network to predict the CTR using the customer data.

Implication : For further research, using information about customer can further improve the predictive capacity of the policy model in reinforcement learning by using services that provide customer contextual data in real time.

Next steps

1. Implement the environment with the following features:
    a. Parameterized control over the distribution of ad performance and the distribution parameters
    b. Varying distributions based on factors such as hour of day, day of week, week of year, holidays, website destinations and ad creative
    c. Calculation of total regret
    d. Implementation along the lines of OpenAI Gym
2. Implement the agent with the following details:
    a. Thompson Sampling at the ad creative level
    b. Thompson Sampling at the ad creative and website destination level
    c. Bayesian Linear Regression with state input at the ad creative level
    d. Bayesian Linear Regression with state input at the ad creative and website destination level
    e. Bayesian Neural Network as a universal model with all ads and website destinations

References

1. Melo L, Martins L, Oliveira B, Brandão B, Soares D, Lima T. *PulseRL: Enabling Offline Reinforcement Learning for Digital Marketing Systems via Conservative Q-Learning Deep Learning Brazil Deep Learning Brazil Deep Learning Brazil Deep Learning Brazil*. Accessed June 15, 2022. https://offline-rl-neurips.github.io/2021/pdf/9.pdf

2. Chatterjee P, Hoffman DL, Novak TP. Modeling the Clickstream: Implications for Web-Based Advertising Efforts. *Marketing Science*. 2003;22(4):520-541. doi:10.1287/mksc.22.4.520.24906

3. Richardson M, Dominowska E, Ragno R. Predicting clicks. *Proceedings of the 16th international conference on World Wide Web - WWW '07*. Published online 2007. doi:10.1145/1242572.1242643

4. Zhou G, Zhu X, Song C, et al. Deep Interest Network for Click-Through Rate Prediction. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Published online July 19, 2018. doi:10.1145/3219819.3219823