

June 11, 2024

1 Introduction

The global digital advertising market is worth approximately \$602 billion today. Due to the increasing rate of online participation since the COVID-19 pandemic, this number has been rapidly increasing and is expected to reach \$871 billion by the end of 2027 (eMarketer, 2024). Many of the major Ad platforms such as Google, Facebook and Amazon operate on a cost-per-user-engagement pricing model, which usually means that advertisers get charged for every time a user clicks on an advertisement. This means that these platforms are incentivized to make sure that the content shown to each user is as relevant as possible in order to maximize the number of clicks in the long term. Attaining accurate Click-Through Rate (CTR) prediction is a necessary first step for Ad personalization, which is why study of CTR prediction methods have been an extremely active part of Machine Learning research over the past through years.

Initially, shallow prediction methods such as XGBoost (Cite), Factorization Machines (Cite) and Field-Aware Factorization Machines (Cite) have been used for CTR prediction. However, these methods have often been shown to be unable to capture the higher order feature interactions in the sparse multy value categorical Ad Marketplace datasets (Cite). Since then, Deep Learning methods have been shown to show superior predictive ability on these datasets. The focus of my reasearch project is therefore to explore the different Deep Learning Architectures for Click through rate prediction, and to incorporate the best of these in a Reinforcement Learning framework.

In the following report, I explore the relevant datasets and simulations that I will be using throughout my research projects. In the first section, I perform an exploratory data analysis on three widely adopted benchmark CTR prediction datasets; the KDD12 (Aden, 2012), Avazu (Wang and Cukierski, 2014) and Criteo (Tien et al, 2014) datasets. In the second section, I then explore possible ways of simulating the ad marketplace environment in order to test the reinforcement learning framework.

2 Datasets

2.1 KDD12

2.2 Avazu

2.3 Criteo

3 Simulatinons

4 Summary of findings

5 Suggested Future Research

References

* eMarketer. (2023). Digital advertising spending worldwide from 2021 to 2027 (in billion U.S. dollars) . Statista. Statista Inc.. Accessed: June 09, 2024. <https://www-statista-com.iclibezp1.cc.ic.ac.uk/statistics/237974/online-advertising-spending-worldwide/>

* Aden, Yi Wang. (2012). KDD Cup 2012, Track 2. Kaggle. <https://kaggle.com/competitions/kddcup2012-track2>

* Steve Wang, Will Cukierski. (2014). Click-Through Rate Prediction. Kaggle. <https://kaggle.com/competitions/avazu-ctr-prediction>

* Jean-Baptiste Tien, joycenv, Olivier Chapelle. (2014). Display Advertising Challenge. Kaggle. <https://kaggle.com/competitions/criteo-display-ad-challenge>