EAS 504: Applications of Data Science – Industrial Overview – Spring 2023

-Lecture by Sharat Chikkerur

Name: Tananki Ranga Sai Saran Rohit

UB ID: 50441793

UB Email: rangasai@buffalo.edu

Ques 1 : Describe the market sector or sub-space covered in this lecture :

The market sector or sub-space covered in this lecture are Computational advertising. Computational Advertising (CA) is a scientific subscience at the intersection of information retrieval, machine learning, statistical modeling, optimization, large-scale search, and text analysis. The objectives of CA are kind of similar to traditional advertising (print, television, and other forms of media) , in computational advertising ,we achieve a more efficient allocation of advertising resources through better targeting and to improve the relevance and personalization of advertisements. It involves data analysis and algorithms that target the right audience with the right ads, at the right time and on the right platform. This process involves collecting user data, identifying patterns and using predictive modeling to optimize ad delivery and increase the effectiveness of advertising relevant set of audience. Computational advertising is a hybrid mixture of areas like computing, economics and machine learning. In order to extract a particular feature, we follow specific routines such as Web-scale audience, we use Distributed computing, for Value estimation and Personalized content – we use Machine learning and Recommendation system , for Targeted delivery -we use information retrieval , for Dynamic pricing – Game theory is used . Google, Pinterest ,WeChat, Facebook, Kayak,Instagram, Twitter, and Snapchat are some of the top advertising companies as they have access to people. They are multiple entities involved like Advertisers (like Mercedes, BMW), Publishers (yahoo,Google etc…) , Ad networks ,Ad Exchanges , SSP’s (similar to e-trades – which are at selling side) , ATD’s , DSP’s

(Demand side platform -optimises the spend) .Back in 2016, Computational advertising exceeded television advertising ,it has 11 percent growth rate.it is projected that , it meets 260B$ mark by 2020.

Ques 2 : What data science related skills and technologies are commonly used in this sector?

In Computational advertising , data science related skills and technologies plays an very important role such as developing machine learning models that pull user behavior information to find patterns. Machine learning and automation are also used to test the effectiveness of marketing campaigns, analyze market segmentation, automate advertisements and other forms of communication, and change user response to a brand, product, company, or advertising campaign. Vowpal Wabbit is an open-source fast online interactive machine learning system library and program developed by yahoo and currently maintained by Microsoft which supports out of core execution with in memory model. Recommender systems, Linear regression, stochastic gradient descent, logistic regression, Poisson models, and other machine learning methods were employed. Vowpal Wabbit provides fast, powerful and flexible web-based machine learning techniques for reinforcement learning, supervised learning and more. Important features of Vowpal Wabbit for machine learning are continuous learning (online learning), dimensionality reduction and interactive learning. Vowpal Wabbit is also a solution to problems when model information does not fit in memory.it has tricks like feature hashing ,caching , MPI style AllReduce , Namespaces and interactions.

Ques 3 : How are data and computing related methods used in typical workflows in this sector? Illustrate with an example.

In general, data and computational methods are critical to the success of computational advertising flows, from data collection and analysis to ad optimization and performance analysis. The ability to work with large data sets, use machine learning and predictive modeling, and understand network analytics and optimization tools is key to success in Computational advertising industry. Computational advertising flows typically begin by collecting and analyzing data about user behavior, demographics, and interests. This information is used to identify patterns and make informed decisions about targeting and ad placement.

Machine learning and predictive modeling are used to analyze data and predict user behavior, such as which ads they are likely to click on or which products they are likely to buy. Once user data is analyzed, advertisers can use more precise targeting methods to reach the right audience with the right ads. They can even target the audience on basis of previously viewed, purchase history , demographic information, marital status, age etc. ..... For example – Facebook ads, Google ad sense.

Ques 4 : What are the data science related challenges one might encounter in this domain?

In Computational advertising , huge amounts of data is getting continuously generated from various other sources like different social media platforms, browsers, plug-ins, extensions, mobile phones, applications etc… Some of the data science challenges that can be encountered are data variability - different types of data , quantity - data size, speed - processing speed where data can often be complex and difficult to process and data can contain missing values, which are some of the obstacles that can be encountered in Computational advertising. Data protection is taken care of in computer advertising and new regulations such as GDPR and CCPA have been introduced to protect user data. This can make it difficult to collect and use the information needed for effective advertising. Machine learning algorithms can be biased based on the data used to train them, which can lead to unfair targeting and discrimination. Careful analysis and testing is required to identify and reduce algorithm bias. Fraudulent clicks and impressions can increase advertising costs and decrease campaign effectiveness, requiring sophisticated fraud detection methods.

Ques 5 : What do you find interesting about the nature of data science opportunities in this domain?

Increasing demand in the use of data is exponentially growing from past few years in online shopping,use of mobile phone apps , social media ,tech devices etc…. For instance, the amount of data in 2010, which was being created every two days, and in 2021 it was being created just for every 40 minutes. The amount of data ,which is being created by applications, sensors ,electronic devices, virtual assistants , smart phones, shopping websites, AI devices etc… had become massive and the

need for use of this data to enhance customer experience, take business decisions lead to huge demand in data science and analytics field, particularly in Computational advertising. Computational advertising is a rapidly evolving field, with new technologies and methods emerging regularly. This creates exciting opportunities for data scientists to stay on the cutting edge of the field and explore new and innovative approaches to advertising. The success of advertising campaigns can have a significant impact on business results, which is why the work of data scientists in this area is especially rewarding. Effective computational advertising requires collaboration between multiple functions, including marketing, data science and design to gain more profits, to target particular audience.Please discuss the roles of Demand Side Platforms, Supply Side Platforms, Ad Networks and Ad Exchanges and how data science plays a role in online advertising.

(i) Please discuss the roles of Demand Side Platforms, Supply Side Platforms, Ad Networks and Ad Exchanges and how data science plays a role in online advertising.

Demand-side platforms (DSPs):

Demand-side platforms (DSPs), are a digital alternative to buying and bidding on online advertising space. Demand-side platforms are platforms that allow advertisers to manage their digital advertising campaigns across multiple ad exchanges and delivery sources. DSPs use data science to analyze and segment audiences based on demographics, behavior and interests, allowing advertisers to target their campaigns more effectively. Machine Learning is also used for ad placements in real time.

A supply-side platform (SSP):

A supply-side platform (SSP) is a system software that allows publishers to offer their available inventory to ad exchanges and demand-side platforms . SSPs are platforms used by publishers to manage the sale of ad space to advertisers. SSPs use data science to analyze user behavior, demographics and interests to help publishers better understand their audiences and sell their inventory at the best possible price. Machine learning is used to optimize the sale of ad inventory, maximizing publisher revenue.

Ad networks:

Ad networks connect advertisers and publishers by aggregating ad slots from multiple sources and selling them to advertisers. Ad networks use data science and analytics to analyze user behavior and target ads to the desired audience, also optimize ad placements and bidding strategies to maximize advertisers' profits.

Ad exchanges:

Ad exchanges are platforms that facilitate the buying and selling of ads in real-time auctions. Ad exchanges use data science to analyze user behavior and match advertisers with the right ad mix, as well as optimize bidding strategies and ad placements to maximize profits for buyers and sellers.

The Role of Data Science in online advertising:

Data Science plays a very crucial role in online advertising. By analyzing user behavior, demographics and interests, these platforms can target ads more effectively and optimize ad placements to increase conversion rates in Computational advertising. Machine learning algorithms are also used to automate and optimize bidding strategies and ad placements in real time, enabling more effective and efficient advertising campaigns.

(ii) Comment on the role of stochastic gradient methods in ML applications

Stochastic gradient methods are very efficient because they update the model parameters at each iteration based on a small subset of the training data. This allows them to process very large data sets without excessive computing resources. Stochastic gradient methods have been shown to converge to a local minimum or saddle point under certain conditions. It is a variant of slope calculation that is used when there is a huge amount of data, such as in calculated advertising. One of the disadvantages of stochastic gradient descent is that it takes longer to reach a conclusion. Here, one of the uses of Stochastic gradient is in training models for deep learning, where we deal with large amounts of training data sets. Stochastic gradient Is also used in natural language processing, Support Vector Machines, Lasso and Ridge Regression.

(iii) Also, answer the following multiple-choice questions: You can list the quetsion number and the letter corresponding to the correct choice as Answer in your report.

Q1) Ans: E

Q2) Ans: D

Q3) Ans: D

Q4) Ans: A

Q5) Ans: E