Name: Sahithi Dodda
Person Number: 50441731
UB Email: sahithid@buffalo.edu

Assignment 2

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

Lecture by Sharat Chikkerur - Computational Advertising

Q1): Describe the market sector or sub-space covered in this lecture.

Ans: Computational advertising is the market sector or sub-space covered in this lecture. It is sometimes referred to as "Online advertising" or "Web advertising," which describes the process of locating the Web adverts that are most pertinent to a given context. Depending on the type of advertising, the context may refer to the content where the advertisement is displayed, the viewer of the advertisement, or the user's social network unlike traditional advertising agencies like WPP, Arnold Worldwide, Publicis Groupe etc., which uses print, television, radio, magazines, and other forms of media to advertise which is basically not targeted advertising. Google, Pinterest, Facebook, Instagram, Twitter, and WeChat are some of the top computational advertising companies because they have access to consumers, deliver customized advertising, and are publishers. It is crucial for advertising because here is where information appears and is widely viewed. Computational Advertising is a field of study which combines areas of data science, machine learning, economics, and computing. To be successful computational advertisers, several issues must be resolved like, Address the web-scale audience; to display these advertising, a mature infrastructure is required, Determine who should see adverts, when they should be shown, and, if you are an advertiser, how much to charge for those ads (value estimate), Determine what kind of advertisements should be displayed for a certain user at 10pm which is Personalized advertising. 90% of a company's income comes from computational advertising, making these commercials more profitable than television commercials.

Q2): What data science related skills and technologies are commonly used in this sector?

Ans: Several data science related techniques and tools are used in computational advertising, and the lecture has covered a variety of techniques used. Linear regression, log-linear regression, logistic regression, Poisson regression, Support Vector Machines, Quantile regression, and recommender systems, stochastic gradient descent, content-based filtering, collaborative filtering, bandit optimization, binary classification, multiclass classification, and logistic regression are used. Additionally, probability, statistical analysis, information retrieval, value estimation, optimization, operations research, distributed computing, and game theory are all used on a regular basis. Machine learning frameworks like Spark ML and Vowpal Wabbit are utilized to handle the enormous quantities of data in this field. Vowpal Wabbit is one of the most well-liked and scalable machine learning frameworks for computational advertising.

Q3): How are data and computing related methods used in typical workflows in this sector? Illustrate with an example.

Ans: Utilizing data and computing techniques, computational advertising optimizes the distribution of advertisements to specific audiences and evaluates their efficacy. Targeted display advertising on websites and social media platforms is a typical example of computational advertising in real life. For instance, if we have recently searched Google for handbags, we could get handbag advertisements when we go to other websites. To find patterns and insights, our browser history, search history, interests, website tracking cookies, demographics and other online behaviors may be used to target advertisements at us when we use social media or explore the web. Using this information, complete user profiles are generated, which are then used to target advertisements to certain audiences. This data is then reviewed for trends and insights that may be used to improve ad targeting using machine learning algorithms. Advertisers may target their adverts to the most appropriate people by using several targeting criteria, including geography, age, gender, interests, and behavior. Computational advertising techniques are employed in addition to targeting to assess ad performance and improve ad delivery. The data from tracking measures like click-through rate, conversion rate, and cost per acquisition may be used by advertisers to plan future ad campaigns.

Q4): What are the data science related challenges one might encounter in this domain?

Ans: Computational advertising relies heavily on data science techniques to optimize the delivery of ads to specific audiences. However, the field also faces several data science-related challenges. One significant challenge is the quality and accuracy of the data used to train machine learning models. Large volumes of user activity and preference data must be gathered and analyzed by advertisers, yet this data may be noisy or incomplete, creating biased or erroneous models. Therefore, before utilizing the data to train machine learning models, it must be carefully chosen and preprocessed to guarantee that it is trustworthy and clean. The enormous amount of data that needs to be handled in real-time presents another difficulty. Real-time bidding (RTB) requires algorithms to quickly analyze large amounts of data to make optimal bidding decisions. The ability to process data in real-time is critical to the success of computational advertising, and advertisers must have robust data processing pipelines to handle the volume of data. Another major issue in computational advertising is privacy issues. The demand for data to target advertisements must be balanced with the requirement to safeguard users' privacy by advertisers. In many places, more privacy laws are being implemented, and advertisers must abide by these laws while still producing successful advertisements. Value estimations, optimization, and pricing are further data science-related challenges encountered in this sector.

Q5): What do you find interesting about the nature of data science opportunities in this domain?

Ans: One interesting aspect of computational advertising is the sheer volume and variety of data that is generated and analyzed. There are many different sources of data that can be used for ad

targeting and optimization, including website tracking cookies, social media profiles, and third-party data providers. This data is often unstructured and complex and requires sophisticated machine learning algorithms and statistical models to extract meaningful insights. The real-time nature of the bidding process in computational advertising is another intriguing feature. Real-time bidding (RTB) includes bidding on ad inventory in real-time and necessitates making quick, correct decisions based on a range of variables, such as user behavior, advertiser targeting criteria, and budgetary restrictions. This requires advanced algorithms and computational infrastructure to make optimal bidding decisions in real time. Additionally, the constantly evolving nature of the advertising industry and consumer behavior presents ongoing challenges and greater opportunities for data scientists in this domain.

(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. (10 pts of the 80 C+R points in the rubric)

Demand Side Platforms: Advertisers and agencies use demand side platforms to buy ad inventory from various sources and manage ad campaigns. They use data science to optimize ad targeting, ad delivery for clients and they use machine learning algorithms to analyze vast amounts of data about users and their behavior to identify patterns, insights that can be used to improve ad targeting.

Supply Side Platforms: Supply Side Platforms are used by publishers to manage the sale of ad inventory on their websites or mobile apps. They use data science to evaluate ad requests, determine the optimal price and ensure that ads are delivered in a timely and efficient manner. They use machine learning algorithms to analyze data about users and match the right ads with the right users, and to optimize ad delivery based on user behavior.

Ad Networks: Ad Networks act as intermediaries between advertisers and publishers. They connect advertisers with publishers, offering ad inventory from multiple publishers to advertisers. Ad networks use data science to analyze user data and behavior to optimize ad targeting and delivery, in a similar way as DSPs.

Ad Exchanges: Ad Exchanges are marketplaces where buyers (DSPs and advertisers) can bid on ad inventory from sellers (SSPs and publishers). Ad exchanges use data science to determine the optimal price for ad inventory based on factors related to user's data as well as the advertiser's targeting parameters and budget. They use machine learning algorithms to match the right ads with the right users, and to optimize ad delivery based on user behavior.

Data Science Role in online advertising: Overall, data science plays a crucial role in online advertising, enabling these different entities to collect and analyze large amounts of user data to deliver more relevant and effective ads. By using machine learning algorithms and other data science techniques, advertisers and publishers can optimize ad targeting, ad delivery, and ad pricing, ultimately improving the overall user experience and driving better results for advertisers.

(ii): Comment on the role of stochastic gradient methods in ML applications. (10 pts of the 80 C+R points in the rubric)

Ans: Stochastic gradient methods play a critical role in a wide range of machine learning applications, including computational advertising. It is an optimization technique which determines the ideal function weights to reduce the cost or loss function. By calculating the gradient of the loss function and changing the weights using a learning rate that relies on time, it is frequently used to solve any convex loss function. Each sample goes through this process again until convergence is attained. When there is a lot of data, like in computational advertising, Stochastic gradient Descent is a variation of gradient descent that is especially helpful. The longer convergent time of Stochastic gradient descent is one of its drawbacks, though. Stochastic Gradient descent is also employed in a variety of other domains, including Deep learning, Natural Language Processing, Logistic Regression, Lasso and Ridge Regression, Linear Regression, Support Vector Machines etc. In the context of computational advertising, SGD is often used in the training of machine learning models for ad targeting and optimization. By leveraging data from ad exchanges, demand-side platforms, and supply-side platforms, advertisers can use stochastic gradient methods to train models that can make predictions about which ads will be most effective for which users. These models can then be used to optimize ad delivery and improve advertising effectiveness. Overall, the use of stochastic gradient methods is critical for effective machine learning in computational advertising and other domains.

(iii): Also, answer the following multiple-choice questions: You can list the question number and the letter corresponding to the correct choice as Answer in your report, (2x5 = 10 pts of the 80 C+R points in the rubric)

Ans:

Q1) E
Q2) D
Q3) D
Q4) A
Q5) E