Computational Advertising: A Growth Field for Statistics

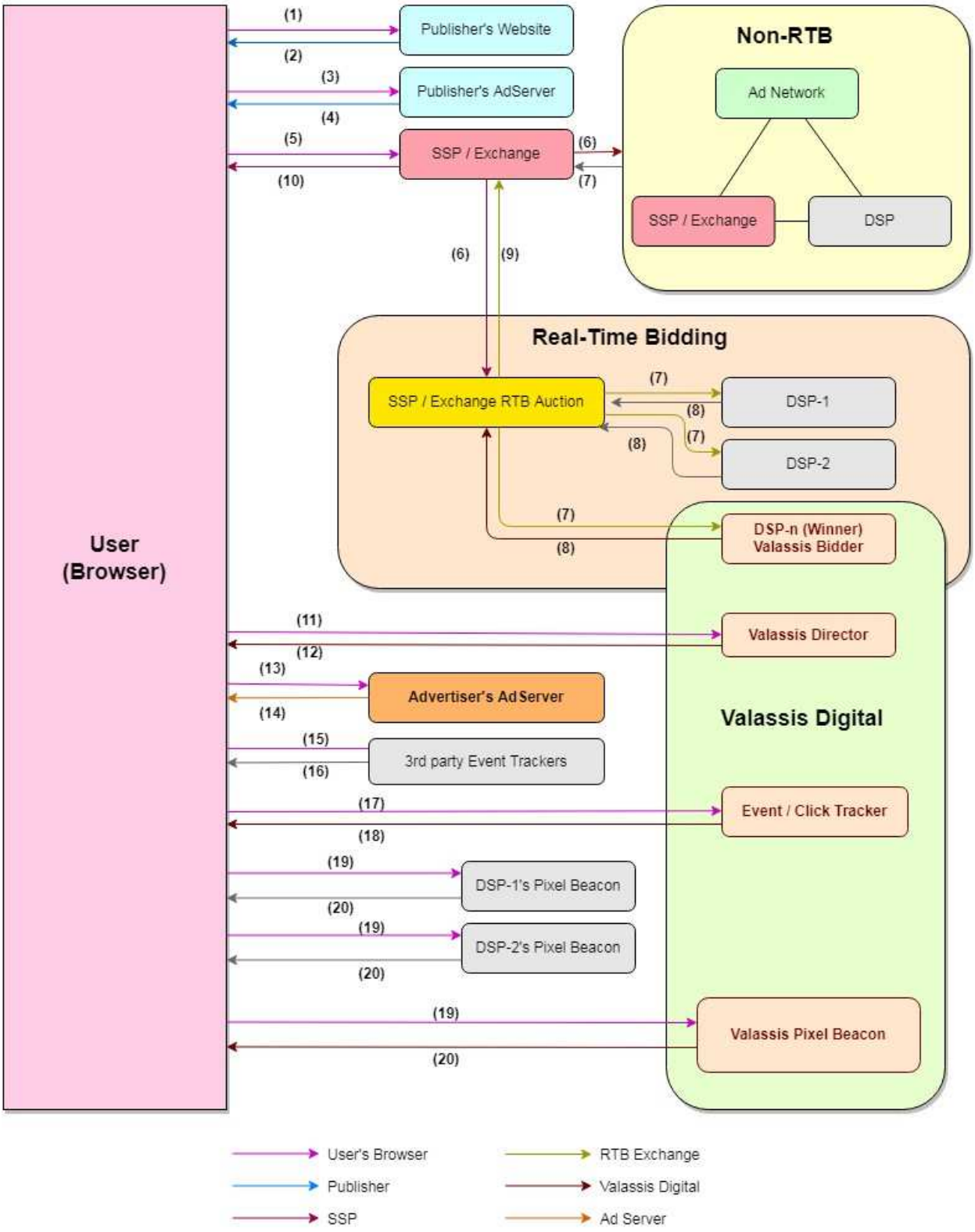
David Banks
Duke University

1. Introduction

Computational advertising (CA) is a young, fast-moving field. And it is highly statistical.

The online advertising market was valued at $309 billion in 2019, and is projected to be valued at $982 billion by 2025. It is the dominant revenue stream for many major IT companies.

One component of CA is on-line ad clicks. When you type "pizza" into their browser, it triggers a virtual auction that lasts a few milliseconds. Domino's and Papa John's and Pizza Hut bid for your eyeballs. The highest "qualified" bids are displayed, with the highest bidder getting the top position. But the process is actually much more complex.


User
(Browser)
(1)
Publisher's Website
(2)
(3)
Publisher's AdServer
(4)
(5)
SSP / Exchange
(6)
(7)
(10)
Non-RTB
Ad Network
SSP / Exchange
DSP
(6)
(9)
Real-Time Bidding
SSP / Exchange RTB Auction
(7)
(8)
DSP-1
(7)
(8)
DSP-2
(7)
(8)
DSP-n (Winner)
Valassis Bidder
Valassis Digital
(11)
(12)
Valassis Director
(13)
(14)
Advertiser's AdServer
(15)
(16)
3rd party Event Trackers
(17)
(18)
Event / Click Tracker
(19)
(20)
DSP-1's Pixel Beacon
(19)
(20)
DSP-2's Pixel Beacon
(19)
(20)
Valassis Pixel Beacon
User's Browser
Publisher
SSP
RTB Exchange
Valassis Digital
Ad Server

1. A person's browser contacts the publisher's website (e.g., CNN.com).
2. The publisher's website sends back content including placements that will need to be fulfilled though an AdServer.
3. The browser contacts the publisher's AdServer to fulfill placements that will not go up for auction. If the user has an in-app or a browser ad blocker this interchange will not happen.
4. The publisher's AdServer sends back predefined ad content.
5. For other placements the browser will contact an Exchange with placement information and an indication on whether the placement should go out for bid, or if another private (guaranteed) deal has been setup for the placement.

6. If the placement is marked for real-time bidding (RTB), an auction is set up by the Exchange.

7. Once the auction is initiated the demand side platforms (DSPs) are simultaneously contacted to participate in the auction. This all runs in parallel.

8. If a DSP decides to bid its bid offer and a director tag (for tracking) are returned to the auction.

9. The winning bid from the auction and information on the bidder, the DSP, is sent to the Exchange

10. The Exchange returns the DSP wrapped ad tag that contains everything needed to track the bid, impression, and director log joins. It's essentially a link to the director so one can obtain user information and pass back the actual creative ad tag.

11. The browser contacts the DSP's director to obtain the AdServer tag.

12. The browser receives the AdServer tag.

13. The browser then uses the AdServer tag to contact the advertiser's AdServer requesting the impression.

14. The advertiser's AdServer returns all of the impression assets and creative content back to the browser.

15. If the campaign/line item is using a third party for tracking it is contacted with the impression and user information.

16. The third party tracker sends back a 1×1 pixel as a verification or handshake token.

17. The ad tag (sent in #10) contains the javascript that the browser/app will load when the impression renders. These signals are passed back ONLY AFTER the creative content is loaded which is why these actions/behaviors are so far down the chain.

18. The DSP sends back a 1×1 pixel as a verification/handshake.

19. If the DSPs are tracking pixel fires this information is sent from the browser.

Even this is an oversimplification. The economic ecology of ecommerce is evolving. For example, it used to be that nearly all auctions were second-price; now they are moving towards first-price auctions.

CA touches on many aspects of statistics. Important topics include design of experiments, causal inference, recommender systems, predictive inference, and time series modeling.

But CA can also engage with text and sentiment analysis, dynamic network analysis, probabilistic ad contracting, spatio-temporal processes, censored data analysis, and so forth.

This is a good opportunity for academic research, since most companies that work in this space are not interested in proving theorems, but rather eking out an extra half percent of profit or putting out urgent fires.

2. Active Recommender Systems

A particularly fun application is active recommender systems. If someone asks for a movie or book recommendataion, I typically ask them about other movies or books that they like, and base my recommendation on their reply.

These are rather like playing "20 Questions" but with special features:

- personalized priors
- complexity constraints
- non-standard feature selection.

In principle, one would build a proximity matrix for books or movies or music.

Hypothetically, Amazon could calculate a distribution over the probability of me buying any book they have on offer. This distribution would be based upon previous purchases I have made, and some model for "nearness" in book space. Collaborative filtering and content-based filtering are default methods.

That model for nearness should be complex and peronally tailored. Some people follow authors, others follow genres, others follow the New York Times book review section. Probably the model for nearness is non-Euclidean, and so one might need isomap or paramap.

Next, Amazon needs to learn what questions to ask that will let it learn the most about the book(s) it will recommend.

Unfortunately, the best questions it should ask are things like "On the whole, do you like these 100 books more than this other list of 150 books?" And that is impossible for someone to cognitively process.

Therefore there needs to be a complexity penalty on the questions that are asked. Defining such a penalty is an area for research, but it can be viewed as a kind of statistical regularization.

Ideally, the questions should be ones for which Amazon's prior gives a 50-50 chance of me responding "yes". That means Amazon will learn at the fastest possible rate.

But there is a hidden optimization problem. The first question Amazon asks might have a goog 50-50 split, but, depending on the answer, subsequent questions might be 90-10 splits.

Therefore, to optimize, one seeks a "question tree" that has lots of near 50-50 questions in the follow-ups. So an initial question with a 60-40 split that has lots of subsequent 60-40 questions would be better.

This is a general class of problems and I am not aware of any previous literature that addresses such cases. But optimal learning, under various complexity, memory, and computational constraints, should be an interesting area of study.

3. Ad Contracts

One financial model is that a company contracts to show a client's ads to, say, 100,000 women between the ages of 20 and 40 who live in California between January 1, 2022 and June 1, 2022.

A second contract promises to show a different set of ads to 200,000 people between the ages of 20 and 50 in northern California between February 1, 2022 and September 1, 2022.

These audiences overlap, and so the company can seek to optimize its revenue by adaptively allocating the ads to people browsing the Internet.

Also, some of this the characteristics (age, gender) may need to be inferred, and the contract must specify how that will be done.

One can imagine agreements being written that require an ad to be shown to 100,000 people who have probability 0.9 of being women between the ages of 20 and 40 who live in California.

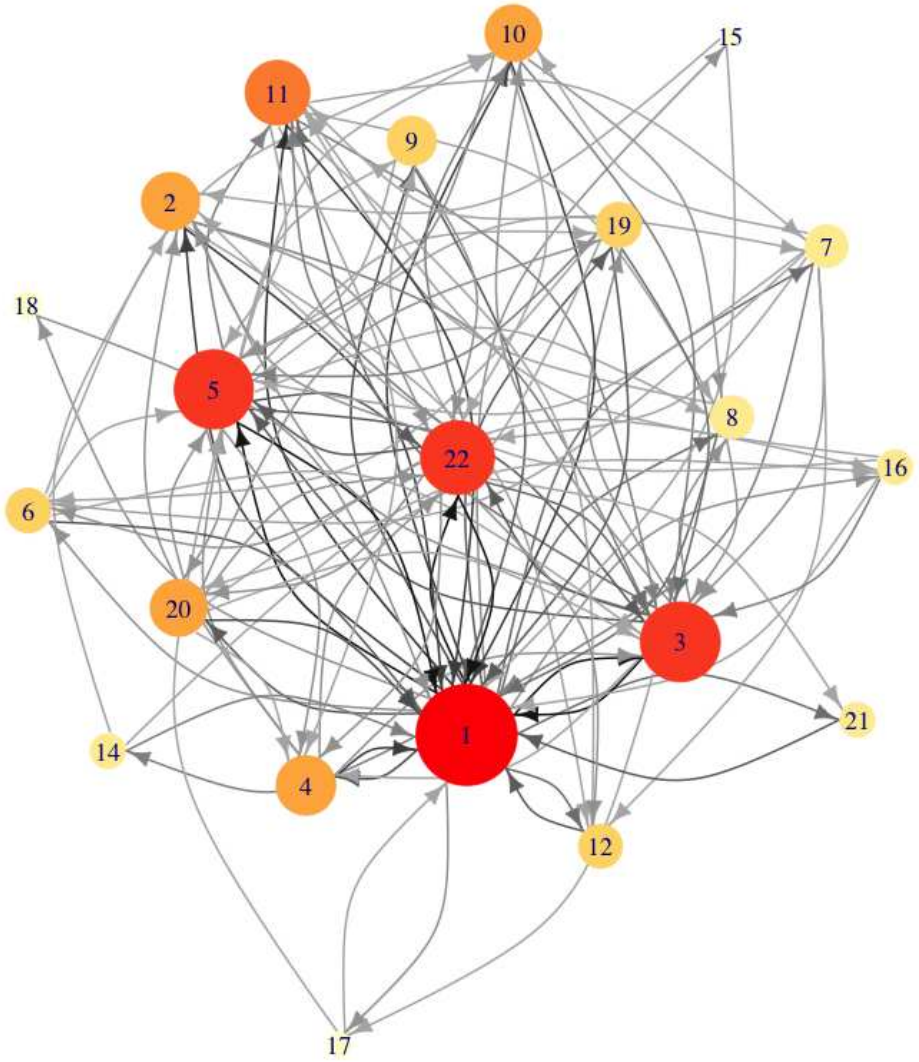
Such contracts would be novel, and will require some sort of third party verification of the probability estimates. Those probability estimates would be based on whatever information is knowable about the person using the browser.

Google guesses users' age, gender, marital status, income bracket, and personal interests, apparently with pretty good accuracy. Other companies do this too.

4. Dynamic Network Flows

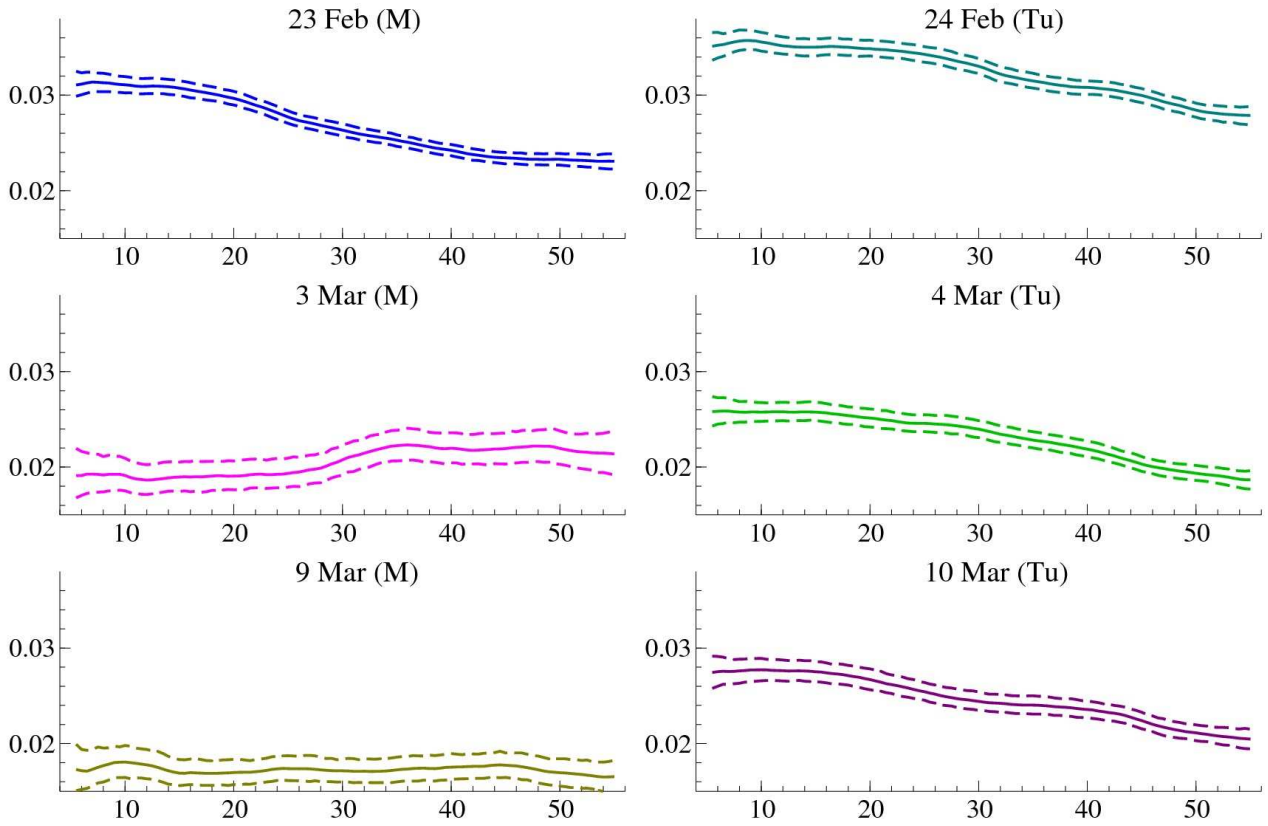
Viewers log in to the Internet and traverse the website. In the same way that groceries put candy and magazines at the checkout stand, computational advertising firms know that certain points in the browsing trajectory are better times to push ads.

As part of previous work with a computational advertising company, we analyzed flow data within the Fox News website. A decoupled-recoupled technique was used to find Bayesian time varying estimates of traffic, essentially a dynamic gravity model ("Scalable Bayesian Modeling, Monitoring and Analysis of Dynamic Network Flow Data," Journal of the American Statistical Association, 113, 519–533).



1 homepage
2 politics
3 us
4 opinion
5 entertainment
6 tech
7 science
8 health
9 travel
10 leisure
11 world
12 sports
13 shows
14 weather
15 category
16 latino
17 story
18 on-air
19 video
20 nation
21 magazine
22 others

The tools in our paper allow one to do process monitoring and changepoint detection.



These are the estimated transition probabilities from the Fox News homepage to the Entertainment page on six different days.

Process monitoring is also important in tracking an ad campaign, in order to decide how much ad buy is appropriate over time. This is a tricky modeling problem since market competitors are also running their own campaigns—the analysis must take account of their success or failure too.

Tools such as a time-varying Cox proportional hazards model or adaptations of Mark Glickman's models for sports teams would be a way to structure this. These could track changes in the clickthrough rate—at a certain point, the ad becomes stale and must be refreshed.

Deep neural networks, especially generative adversarial networks, would be a potential tool for creating new ad content. Style transfer and other techniques could help.

Style transfer example: Picasso/Van Gogh/Seurat

5. Online Controlled Experiments

Google, Amazon, Facebook, Apple and Microsoft run 10,000+ experiments per year, engaging with millions of users. LinkedIn reportedly runs more than 400 simultaneous experiments per day.

Companies use tools such as Optimizely, Google Optimize, Mixpanel, VWO, AB Tasty, and Split.io to run and manage their experiments.

There are several recent books about online controlled experiments. And Netflix recently posted an job ad for someone who could "support internal research into new methodologies for experimentation as well as adapt existing methods such as response surface methodology to online A/B testing."

Lyft used a 2^2 + centerpoint response surface design to decide what percentage discount to offer customers who had not booked for awhile, and how many such rides would be eligible.

Companies experiment to improve: User retention mechanics, user acquisition funnels, user engagment, email promotions, website layout, aesthetic features, checkout experience, branding, ad campaigns, and many other properties.

The workhorse for online controlled A/B experiments are getting much more sophisticated than the name suggests. Adaptive multiarmed bandits and sequential experimentation have become common.

Often effect sizes are small, but they translate into millions of dollars.

In 2009, Google did a designed experiment to compare 41 shades of blue for its hyperlinks.

In 2008, Obama's campaign increased donations by $60 million by using a factorial design for the donation page's layout.

Open problems in online controlled experiments:

- OCEs usually run for two weeks. How can one estimate longer term effects such as primacy (when a change that is good over the long run temporarily degrades performance) and newness (when a bad long-range change looks good at first).
- One wants to estimate heterogeneous treatment effects, by region or device or gender.
- Network interference can violate the Stable Unit Treatment Value Assumption (SUTVA).
- How to manage information flow when a company is running 100s of experiments each day.

6. Conclusions/Other Topics

CA touches on nearly every aspect of modern applied statistics.

- Cluster analysis and multidimensional scaling should drive market segmentation and collaborative filtering.
- Latent Dirichlet allocation, sentiment analysis, Word2Vec, and NLP are helpful to CA in many ways—one doesn't want to bid to display an McDonald's ad on a PETA website, despite good vocabulary matches.
- Statistical tools for anomaly detection are relevant to cybersecurity. And cybersecurity is essential if people are clicking on ads.
- Spatio-temporal models can indicate when a coupon should be sent to the cell phone of a customer walking through a store, or find times of day when a person in a given location is likely to click on an ad.