# Computational advertising landscape

Sharat Chikkerur
sharat@alum.mit.edu

# About me

- Google, Senior Software Engineer
  - AdWords pCTR modeling (McMahan et al. [2013])
  - UI and position effect modeling
- Microsoft, Senior Software Engineer (AzureML [2016])
  - Azure machine learning
- Nanigans, Principal data scientist
  - Value modeling for Facebook/Twitter/Instagram
  - Modeling stack for Real time display (RTB)
- Microsoft, Principal data scientist lead
  - AI applications

# About the class

- Computational advertising landscape
- Vowpal Wabbit: Scalable machine learning library
- Value estimation: Generalized linear models
- Topic modeling: Latent dirichlet allocation

# Introducing the landscape

# Traditional advertising

Ogilvy & Mather

dentsu AEGIS network

SAATCHI & SAATCHI

WPP

PUBLICIS GROUPE

ARNOLD WORLDWIDE

# Computational advertising

# What is computational advertising

Computational advertising is a new discipline that spans areas of computing, economics and machine learning

| **Feature** | **Discipline** |
|---|---|
| Web-scale audience | Distributed computing |
| Value estimation | Machine learning |
| Targeted delivery | Information retrieval |
| Personalized content | Machine learning |
| | Economics (Explore Exploit) |
| | Recommendation system |
| Dynamic pricing | Game theory |

# Economic landscape

# Economic scale: worldwide

- 11% annual growth rate
- 260B by 2020

# Economic scale: US

- 17% annual growth rate, 100% annual growth for mobile

# Economic scale: formats

# Market landscape

# Entities

- Advertiser
- Publisher
- Demand side platform
- Supply side platform
- Ad networks
- Exchanges
- Data aggregators

# SEARCH LUMAscape

LUMA

Denotes acquired company

Denotes shuttered company

© LUMA Partners LLC 2016

# DISPLAY LUMAscape

# Delivery landscape

# Search ads

- Intent is well known
- Search engine acts as a publisher, exchange and data aggregator
- Targeting includes keywords, demographics, geo and user history

# Display ads : Retargeting

- Targets user who have almost converted elsewhere.
- Users are tracked through cookie syncing between advertiser and exchange
- Higher intentionality, time sensitive and very profitable

# Display ads: Prospecting and brand

- Typically used for brand advertising and optimized for reach
- Primarily targeted using demographics
- Low intentionality

# Native ads

- Sponsored content is mixed with native content
- Very popular within social networks and portal sites
- Requires good recommendation scheme

# Multiformat

- Sponsored content is mixed with organic content
- Ads from other networks also displayed

# Market landscape

# Marketplace

- Grauranteed delivery (Futures contract)
  - Direct advertiser publisher contract
  - Publisher guarantees delivery of impression to a given audience at a fixed price.
    - Publisher on the hook for shortfall
    - forecasting is the key to make a profit
- Header bidding/First look (Options contract)
  - Advertiser pays a premium to get the first look at buy at a lower price than auction
  - Publishers get a better margin than the open exchange

# Marketplace(cont.)

- Non guaranteed delivery (Auctions)
  - Each impression goes through an auction in one or multiple exchanges
  - Publisher chooses the best bid
  - Less expensive

# Pricing type trend

# Challenges

# The challenge of advertising

- Find the best match between a given **user** in a given **context** and a suitable **advertisement** (when the advertisement is relevant in the context) [1]

**Context**

| Channel | Context |
|---|---|
| Search | Keywords, Query, Ads, Geolocation, user behavior |
| Display | Page content, user behavior, geolocation, demographics |
| Native | Page content, user behavior, social connections |

[1] Introduction to computational advertising `https://web.stanford.edu/class/msande239/`

# Problem domains

Advertising can be thought off as a sequential decision process

- Targeting:
  - **Who** do we show the advertising to and **when**
- Ranking/Matching/Recommendation
  - Which **set of ads** do we show
- Bidding
  - How much is this opportunity worth to the **advertiser** ?
- Optimization
  - What's the objective of this ad campaign?
- Pricing
  - How much is this opportunity worth to the **publisher** ?
- Budgeting & Pacing
  - When should we spend and how much to maximize yield

# Challenge: Targeting and ranking

# Targeting: audience selection

- Targeting is usually manually specified based on campaign goals
  - Demographic (age, gender, income bracket)
    - User supplied
    - Credit card data
    - Shipping address etc.
    - Inferred demographics
- Audience targeting can be framed as search/ranking task given (context, ad) pair
  - contrast with Ad selection = search(context, user)

# Ranking: ad selection

- Ad selection can be framed as ranking problem given (ad, display context, user) features
- Common metric: $E[revenue] = p(Click) * CPC$

## Position dependence

- Click through rates depend on display position.

  - Factored click through rate model (Richardson et al. [2007])

    $p(Click|ad, context) = p(view|ad, context) * p(click|view, ad, context)$

  - Factorization machine (Rendle [2012])

    $$f(ad, context) = b + \sum_i (w_i^a a_i) + \sum_j (w_j^c c_j) + \sum_i \sum_j \langle v_i, v_j \rangle a_i c_j$$

# Challenge: Bidding and value estimation

# Bidding: Factored value formulation

- Bidding value is the optimal strategy (Vickrey [1961])
- Value is factored based on a conversion funnel
- View $\rightarrow$ Click $\rightarrow$ Convert $\rightarrow$ Value

Factored formulation for downstream (revenue) optimization

$$VPI = E[V|A] * P(A|Click) * P(Click|Impression)$$

| | | |
|---|---|---|
| $E[V\|A]$ | Value per action | Linear/Log-Linear/Poisson regres |
| $P(A\|\text{Click})orE[A\|\text{Click}]$ | Actions per click | Logistic/Poisson Regression |
| $P(\text{Click}\|\text{Impression})$ | CTR | Logistic regression |

Fortunately, all of these regression can be implemented using a general setup.

# Value estimation: Generalized linear models

A generalized linear predictor specifies

- A linear predictor of the form $\eta(x) = w^T x$
- A mean estimate $\mu$
- A link function $g(\mu)$ such that $g(\mu) = \eta(x)$ that relates the mean estimate to the linear predictor.

This framework supports a variety of regression problems

| | |
|---|---|
| Linear regression | $\mu = w^T x$ |
| Log-linear regression | $\log(\mu) = w^T x$ |
| Logistic regression | $\log(\frac{\mu}{1-\mu}) = w^T x$ |
| Poisson regression | $\log(\mu) = w^T x$ |

# Challenge: Pricing (Exchange side)

# Challenges: Pricing (Auctions)

- First price **sealed bid**
  - Highest bidder wins
  - Unstable
- Second price **sealed bid**
  - Highest bidder wins but pays second price
  - Stable
  - Incentive to bid true value
- English auction (ascending price)
  - Seller increases price until a single bidder remains
- Dutch auction (descending price)
  - Seller decreases price until a single bidder accepts it

# Second price auction

- Assume value is 100
- Optimal strategy: bid 100, payout price cannot be affected
  - If you bid 100, net gain is 100 - cost
  - If you bid $> 100$, increase chances that cost $> 100$, does not affect chance of price $< 100$
  - If you bid $< 100$, reduce chances of winning

**Vickrey-Clark-Groves (VCG) auction**

- Maximizes total social value
- Everybody pays their externality (TSV w/o bidder participation) - (TSV w/ participation)
- Optimal strategy: all bidders bid their true value

# Generalized second price auction

- Rank by *bid* : $cost(i) = bid(i+1)$
- Rank by revenue: $cost(i) = bid(i+1) * ctr(i+1)/ctr(i)$
  - Bidder always pays less than bid
  - Better CTRs generate discount
- Not identical to VCG
- Revenue may be higher than VCG

# Challenge: Budgeting and pacing

# Pacing

- Purpose of pacing
  - Budget pacing - distributes spend over time period
  - KPI pacing - achieves target KPI (control vs. optimization)
- Method of pacing
  - Throttling (allocation) - useful when bid adjustments are not allowed
  - Bid modification
  - Feedback control

# Pacing types (Lee et al. [2013])

(a) Premature Stop

(b) Fluctuating Budget

(c) Uniform Pacing

(d) Traffic Based Pacing

(e) Performance Based Pacing

# Feedback control (Zhang et al. [2016])

Figure 1: The instability of CPM (cost per mille), AWR (auction winning ratio), eCPC (effective cost per click), and CTR (click-through rate) for two sample campaigns without a controller. Dataset: iPinYou.

# Explicit feedback control

Figure 2: Feedback controller integrated in the RTB system.

The bid is computed using

$$b_a(t) = b(t)exp\{\phi(t)\}$$

Here $\phi(t)$ is the control signal. Other approaches such as $b_a(t) = b(t)(1 + \phi(t))$ were found to be ineffective.

# Controller

- PID controller:

$$
\begin{aligned}
e(t_k) &= x_r - x(t_k) \\
\phi(t_k + 1) &= \lambda_P e(t_k) + \lambda_I \sum_{j=1}^{k} e(t_j)\Delta t_j + \lambda_D \frac{\Delta e(t_k)}{\Delta t_k}
\end{aligned}
$$

  Here, $x_r$ is the reference/desired KPI, $x(t_k)$, KPI at time $t_k$. $\lambda_P, \lambda_I, \lambda_d$ are conotroller gains. All control factors remain the same between updates.

- Water-level based controller

$$
\phi(t_k + 1) = \phi(t_k) + \gamma(x_r - x(t_k))
$$

# Performance

Figure 5: Control performance on eCPC and AWR.

# Challenge: Subjective quality

# Subjective quality

## Ads blindness

- Repeated exposures leads to learned blindness/adblock
- We need to predict when not showing ads are better
- Approaches
  - Thresholding based on scores
  - Utility function (lifetime value)
  - Learning from data

## Ads diversity

- Similar search results lead to redundant information
- Maximize overall utility (hard)
- Maximize incremental utility (easier)
- Tradeoff marginal utility with a diversity metric (easy)

# Summary

- Computational advertising invovles multiple subfields of computer science, machine learning, operations research, economics and game theory
- Entire industry is driven exclusively by data
- Diverse eco-system with lots of challenging data science problems

# Thanks!

# References

Alekh Agarwal, Oliveier Chapelle, Miroslav Dudík, and John Langford.

AzureML. Azureml: Anatomy of a machine learning service. *Journal of Machine Learning Reserach*, 50, 2016.

Nikos Karampatziakis and John Langford. Online importance weight aware updates. *arXiv preprint arXiv:1011.1576*, 2010.

Kc Lee, Ali Jalali, and Ali Dasdan. Real Time Bid Optimization with Smooth Budget Delivery in Online Advertising. *arXiv preprint arXiv:1305.3011*, pages 1–13, 2013.

H Brendan McMahan, Gary Holt, D Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. Ad click prediction: a view from the trenches. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1222–1230, 2013. doi: 10.1145/2487575.2488200.

Steffen Rendle. Factorization machines with libFM. *ACM Trans. Intell. Syst. Technol.*, 3(3):57:1–57:22, May 2012. ISSN 2157-6904.

Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*, pages 521–530. ACM, 2007.

Stéphane Ross, Paul Mineiro, and John Langford. Normalized online learning. *arXiv preprint arXiv:1305.6646*, 2013.

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.

Weinan Zhang, Yifei Rong, Jun Wang, Tianchi Zhu, and Xiaofan Wang. Feedback control of real-time display advertising. Technical report, 2016.