The Value of Information in Mobile Ad Targeting

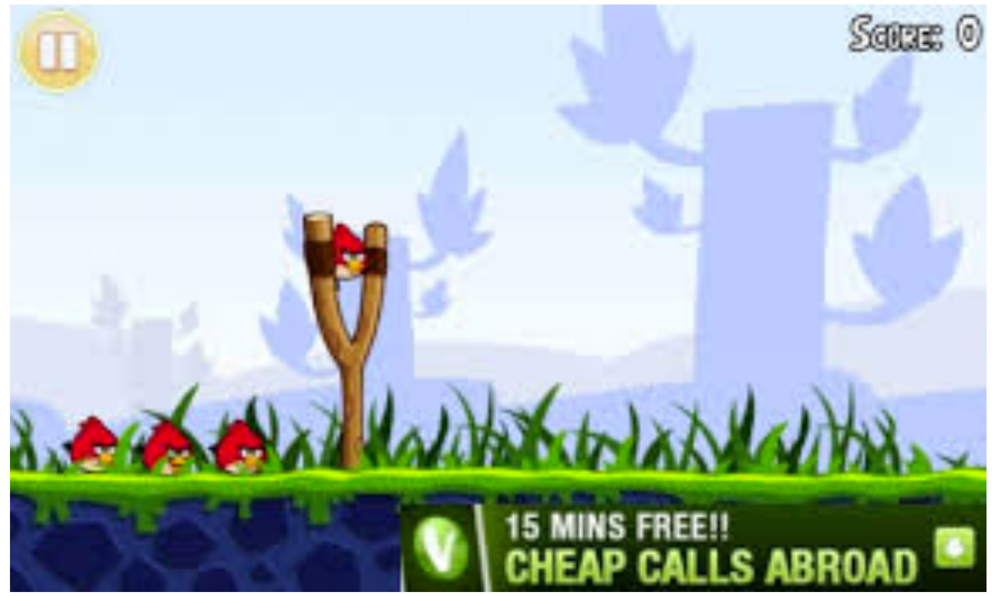
Omid Rafieian
Hema Yoganarasimhan

University of Washington

Smartphone Industry

- Smartphones are increasingly popular worldwide
  - 2 Billion users in 2015
  - Avg. user spends 2.8 hours/day on mobile phones
- “Apps” or Applications usage
  - 25 Billion iOS apps and 50 Billion android apps
- Monetization of Apps
  - Paid model
  - Freemium model (in-app purchases or paid premium)
  - In-app advertising

In-App Advertising



- Mobile ad-spend
    - 13 Billion USD
- Key players
    - Publishers — Host ads
    - Advertisers — Bid and place ads
    - Ad Network — Match publishers and advertisers
- Common goal: increase ad response rates

Targeting to improve ad-effectiveness

- What is targeting?
  - Matching an impression to the best ad available
- How to do effective targeting?
  - Variables
    - Behavioral: what the user did (browse, click history)
    - Contextual: where and when of the impression
  - Data
    - User-level/aggregate
    - Size, length
    - How to combine across sources?

Targeting has privacy implications

Apple's IDFA Crackdown Reverberates Through Mobile Ad Ecosystem

by Judith Aquino // Monday, February 17th, 2014 – 6

Adobe Pitches Marketers On A Cross-Device Data Co-op, But Privacy Is A Snag

by Zach Rodgers // Tuesday, July 28th, 2015 – 5:04 pm

Share:

The competition among cross-device providers is no longer a one, two or even three-horse race, but looks more like a mad steeplechase with 20 animals of various sorts, including

The biggest consumer
ɜn given the best odds,
ɔack.

ɔrk on its own cross-
ɑring cooperative
among customers, AdExchanger has learned.

Adobe has begun actively recruiting co-op members, ahead of a

Mobile, personalized in-app ads present 'a new privacy threat,' Georgia Tech study shows

March 7, 2016 | By Daniel Kobialka

Research Agenda

- Substantive
  - How does targeting improve effectiveness of mobile ads
  - What type of information helps improve targeting and to what extent — Contextual or Behavioral?
  - What is the value of more/better data?
- Methodological
  - What types of methods perform well — Econometric vs. Machine Learning
- Policy and Privacy
  - Would additional privacy regulations (e.g., no tracking ID) worsen ability to target? How much?
  - What are the incentives for data-sharing — between advertisers, between advertisers and the platform?

Challenge 1: Need high predictive accuracy

- Econometric models focus on causality, not prediction
- Causality: Given a model, derive consistent estimates
    - Goal: is to make counterfactual recommendations
    - Challenge: endogeneity concerns
- Prediction: No assumptions on underlying model
    - Goal: High out-of-sample predictive accuracy
    - Challenge: search space is over models (bias-variance trade-off)

Challenge 2: Large number of attributes with complex interactions

- Usually, we assume a fixed functional form and infer parameters — gives mediocre results
- Need to infer both the functional form and the parameters
- Difficult problem with ~38 features and unknown non-linear interactions
  - approx. 1600 variables with just two-way interactions

Related Literature

- Targeting
  - Analytical: Chen et al. (2001), Iyer et al. (2005), Levin & Milgrom (2010)
  - Empirical: Rossi et al. (1996), Ansari & Mela (2003), Chatterjee et al. (2003), Manchanda et al (2006), Ghose & Yang (2009), Yao & Mela (2011)
  - Online ads: Goldfarb & Tucker (2011b), Lambrecht & Tucker (2013), Goldfarb (2014)
- Privacy and data intermediaries
  - Pancras & Sudhir (2007), Goldfarb & Tucker (2011b, c, d), Johnson(2013), Tucker (2014)
- Mobile marketing and advertising
  - Luo et al. (2013), Ghose et al. (2012), Hui et al. (2013), Andrews et al. (2015), Sahni & Nair (2016a, b), Narang & Shankar (2016)
- Methodology: click prediction for online ads and MART
  - Methods: Friedman (2000, 2001, 2002), Friedman et al. (2001)
  - Applications: McMahan et. al. (2013), He et al. (2014)
- Machine learning in marketing
  - Toubia et al. (2003, 2004), Evgeniou et al. (2005, 2007), Huang & Lou (2016), Liu & Dzyabura(2016), Yoganarasimhan (2016)

Outline

- Introduction
- Research Agenda and Challenges
- Related Literature
- Setting and Data
- ML Framework
- Results

Setting

- Major app-store and in-app advertising platform
  - One of top three IT companies in Iran
  - Over 17 million actives users
  - Over 50 million ads served daily
  - Apps — 25,000 apps and 250 ads
- Our data and sampling
  - Focus on top 50 ads and top 50 apps (approx. 80%)
  - Sample 727,000 users from 3 days for training & test
  - 17.7 million impressions in training, 9.6 million in test
  - History of 135 million impressions to make features

Data

- For each impression:
  - Advertising ID (user-resettable, device specific)
  - App ID
  - Ad ID
  - Geographical Location (IP address)
  - Click indicator
  - Time-stamp

Outline

- Introduction
- Research Agenda and Challenges
- Related Literature
- Setting and Data
- ML Framework
- Results

Problem Definition

- Problem: Accurately predict the probability that impression i, by user U, in app P, for ad A, at time T, with global history H, will lead to a click

- Goal: Devise an algorithm that takes as input a set of pre-classified data and generates an output probability $p_i(U, P, A, T, H)$, as close as possible to the true click probability observed in test data

Machine Learning Framework

- Evaluation metric
- Feature set
- Classifying algorithm or supervised learning algorithm

Evaluation Metric

- LogLoss

$$-\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\left(p_i\right) + \left(1 - y_i\right)\log\left(1 - p_i\right)\right)$$

Prediction p_i and click indicator y_i

- Relative Information Gain

$$NE = \frac{-\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\left(p_i\right) + \left(1 - y_i\right)\log\left(1 - p_i\right)\right)}{-\left(y_i \log\left(p\right) + \left(1 - y_i\right)\log\left(1 - p\right)\right)}$$

where p is the baseline CTR

Framework for Feature Generation

Feature No.	Feature Name	Feature Class	Feature No.	Feature Name	Feature Class
1	Impressions (*user*, _, _, _)	F_B	20	CTR (_, *app*, _, _)	F_C
2	Impressions (_, *app*, _, _)	F_C	21	CTR (_, _, *ad*, _)	F_C
3	Impressions (_, _, *ad*, _)	F_C	22	CTR (_, _, _, *time*)	F_C
4	Impressions (_, _, _, *time*)	F_C	23	CTR (_, *app*, *ad*, _)	F_C
5	Impressions (_, *app*, *ad*, _)	F_C	24	CTR (*user*, *app*, _, _)	F_B, F_C
6	Impressions (*user*, *app*, _, _)	F_B, F_C	25	CTR (*user*, _, *ad*, _)	F_B, F_C
7	Impressions (*user*, _, *ad*, _)	F_B, F_C	26	CTR (*user*, *app*, *ad*, _)	F_B, F_C
8	Impressions (*user*, *app*, *ad*, _)	F_B, F_C	27	CTR (*user*, _, _, *time*)	F_B, F_C
9	Impressions (*user*, _, _, *time*)	F_B, F_C	28	AdCount (*user*, _)	F_B
10	Clicks (*user*, _, _, _)	F_B	29	AdCount (_, *app*)	F_C
11	Clicks (_, *app*, _, _)	F_C	30	AdCount (*user*, *app*)	F_B, F_C
12	Clicks (_, _, *ad*, _)	F_C	31	AppCount (*user*, _)	F_B, F_C
13	Clicks (_, _, _, *time*)	F_C	32	AppCount (_, *ad*)	F_C
14	Clicks (_, *app*, *ad*, _)	F_C	33	AppCount (*user*, *ad*)	F_B, F_C
15	Clicks (*user*, *app*, _, _)	F_B, F_C	34	TimeVariability (*user*)	F_B
16	Clicks (*user*, _, *ad*, _)	F_B, F_C	35	AppVariability (*user*)	F_B
17	Clicks (*user*, *app*, *ad*, _)	F_B, F_C	36	Entropy (*user*, _)	F_B
18	Clicks (*user*, _, _, *time*)	F_B, F_C	37	Entropy (_, *app*)	F_C
19	CTR (*user*, _, _, _)	F_B	38	Entropy (*user*, *app*)	F_B, F_C

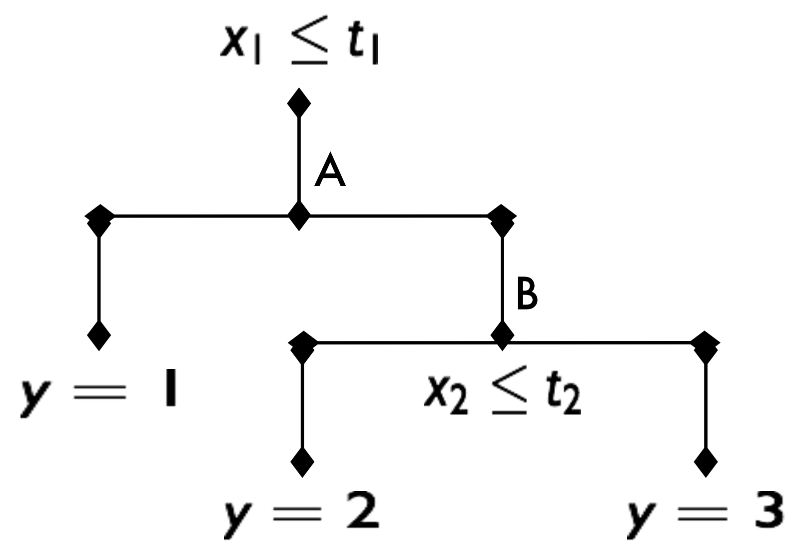
- Parsimonious functions — take as input User, Ad, App, Time
- Classify features as “behavioral” or “contextual” (time, ad, app) or both

Classifying algorithm

- OLS
- Logistic Regression
- Boosted Trees (MART)
  - See Yoganarasimhan (2016) for application
  - Chapter on ML methods in Marketing (Dzyabura and Yoganarasimhan 2016)

Multiple Additive Regression Trees

- Example of CART



- MART
  - Boosted combination of multiple CARTs
  - Can infer both optimal functional form and parameters

- Advantages of MART
  - Automatic variable selection, scalable to big data
  - Can incorporate nonlinear combinations of hundreds of features
  - Empirically shown to be the best classifier in the world

Outline

- Introduction
- Research Agenda and Challenges
- Setting and Data
- ML Framework
- Results

Results

RIG over baseline

Method	User	Ad	App	Ad-App	User-Time	User-Ad-App	All
MART	0.093	0.007	0.044	0.051	0.112	0.132	0.152
Logistic Regression	0.068	0.008	0.042	0.044	0.079	0.094	0.100
OLS	0.066	0.009	0.044	0.046	0.078	0.091	0.095
Ad-App CTR	0.049	0.049	0.049	0.049	0.049	0.049	0.049

- ML methods (MART) perform much better than baseline, Logit model, and OLS
- Behavioral targeting more useful than Contextual (ad, app, time)
    - Of course, combining both is even better
    - app-specific features more valuable than ad-specific features
- Overall model performance is very good; 15.2% improvement in predictive accuracy

Policy Questions and Consumer Privacy

- Strengthen privacy regulations:
  - What if we get rid of Advertising ID?
- Weaken privacy regulations:
  - What if we allow the platform to share data with advertisers at different levels of granularity?
  - What if we allow advertisers access to own data and allow data-sharing among them?

Value of User Identifiers: Ad ID vs. IP

RIG over baseline

Optimization model	Advertising ID	IP
MART	0.143	0.092
Logistic Regression	0.092	0.066
OLS	0.092	0.062

- Significant loss in targeting ability with IP
  - Low persistence: moving from one network connectivity to another changes IP.
  - Masking: VPNs and masked IPs lead to many users falling under the same IP.
  - Automatically reset: Ad ID needs to be actively changed by user, whereas IP changes automatically.

If platform is allowed to shares data with advertisers

- Arrangement between advertisers and platform
  - Scenario 1: ad-specific CTR

- Consider four counterfactual scenarios
  - Scenario 2: access to app-ad specific CTR
  - Scenario 3: access to individual-level data for own ads
  - Scenario 4: access to full feature-set, but individual-level data only for own ads
  - Scenario 5: access to all the data

Value of data to advertisers

RIG over baseline

	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Large	0.037	0.098	0.152	0.155
Medium	0.046	0.071	0.098	0.104
Small	0.038	0.063	0.105	0.119

- While least privacy-preserving arrangement is first-best, we can get very close to it while preserving ad-user privacy!
  - Scenario 4 is only marginally inferior to Scenario 5
- Large advertisers benefit most, followed by small and medium
  - Size of the data helps
  - Controlling for size, advertisers with higher variation in the data (higher CTR) benefit more

If we allow advertisers to share data?

- We compare the value of sharing data among pairs of advertisers (i, j) or (receiver, giver)
- What affects gains of *i* from sharing data with *j*?
  - Larger advertisers gain less from sharing
  - Better when both advertise in common contexts (apps)
- Incentives of sharing pairs is not perfectly aligned
  - Need an incentive-compatible payment system
  - Positive implications for privacy

Conclusion

- Targeting is an important decision in mobile advertising
- From industry perspective
  - How to measure the returns to targeting?
  - What type of information is valuable?
  - What kind of models perform well?
  - Benefits to data-sharing? Who benefits and how much?
- From consumers' perspective
  - Significant privacy concerns
- Some answers
  - Behavioral targeting is more valuable than contextual
  - ML models outperform even the best Logit/OLS models
  - We don't need complete individual-level data for targeting
  - Players' incentives are not aligned in data-sharing

Thank You!

Advertiser's Problem

$$\frac{d\pi\left(x,y(z)\right)}{dx} = \frac{\partial\pi\left(x,y(z)\right)}{\partial x} + \frac{\partial\pi\left(x,y(z)\right)}{\partial y(z)}\frac{dy(z)}{dx}$$

- x is the bid
- y(z) is the probability of clicking conditional on z
- $\pi\left(x,y(z)\right)$ is the profit from bid x and click prob. y(z)
- G(x, z, y(z)) is the probability of winning

Advertiser's Problem

$$\frac{d\pi\left(x,y(z)\right)}{dx}=\frac{\partial\pi\left(x,y(z)\right)}{\partial x}+\frac{\partial\pi\left(x,y(z)\right)}{\partial y(z)}\boxed{\frac{dy(z)}{dx}}=0$$

- x is the bid
- y(z) is the probability of clicking conditional on z
- $\pi\left(x,y(z)\right)$ is the profit from bid x and click prob. y(z)
- G(x, z, y(z)) is the probability of winning

Advertiser's Problem

$$\frac{d\pi\left(x, y(z)\right)}{dx} = \frac{\partial\pi\left(x, y(z)\right)}{\partial x} + \frac{\partial\pi\left(x, y(z)\right)}{\partial y(z)} \boxed{\frac{dy(z)}{dx}} = 0$$

- x is the bid
- y(z) is the probability of clicking conditional on z
- $\pi\left(x, y(z)\right)$ is the profit from bid x and click prob. y(z)
- G(x, z, y(z)) is the probability of winning

$$\pi\left(x, y(z)\right) = (V - x)G(z, x, y(z))y(z)$$

$$-xG(z, x, y(z))y(z) + (V - x)G'(z, x, y(z))y(z) = 0$$

Advertisers need a good predictive model of y(z)