Week 13

Real-Time Bidding

Excerpt from: http://wnzhang.net/slides/ecir16-rtb.pdf

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Conversion attribution
- Learning to bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Conversion attribution
- Learning to bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

Advertising

- Make the best match between advertisers and customers with economic constraints

Wasteful Traditional Advertising

Computational Advertising

- Design **algorithms** to make the best match between the advertisers and Internet users with economic constraints

Sponsored Search

Google | iphone 6s case | Weinan

Web | Shopping | News | Images | Videos | More | Search tools

About 16,900,000 results (0.33 seconds)

iPhone 6s Cases - case-mate.com
Ad www.case-mate.com/iPhone-6s-Cases
4.6 ★★★★★ rating for case-mate.com
Shop The **iPhone 6s Case** Collection. Free Standard Shipping!
Refined Protection · Slim & Tough · Genuinely Crafted · Premium Designs

iPhone 6s
Ad www.apple.com/
The only thing that's changed is everything. Learn more.
A9 chip · Two sizes · Now in rose gold
Pre-order 9.12 - iPhone Upgrade Program - 3D Touch - Cameras

In the news

Speck's iPhone 6s CandyShell + MightyShell cases bring best-of-breed protection to Apple's latest iPhones
9 to 5 Mac - 1 day ago
With the **iPhone 6s** and **iPhone 6s** Plus debuting next week, it's important to start thinking ...

Moshi's iPhone 6s and 6s Plus cases offer premium protection
iMore - 23 hours ago

Top 5 Best Leather iPhone 6s Cases
Heavy.com - 12 hours ago

More news for iphone 6s case

iPhone 6s Cases & Covers from OtterBox
www.otterbox.com/en-us/**iphone-6s-cases** OtterBox
Get protection that inspires confidence with **iPhone 6s cases** and covers from OtterBox.
Demandware SiteGenesis.

iPhone 6s - Accessories - Apple
www.apple.com › iPhone › iPhone 6s Apple Inc.
The essential Apple-designed **cases**, accessories and all-new aluminum docks for **iPhone 6s** and **iPhone 6s** Plus.

Shop for iphone 6s case on Google — Sponsored

Case-mate - Karat Case Fo...	Moshi - Iglaze Armour Case...	Logitech - Protection...	Moshi - Overture Wall...
$49.99	$39.99	$21.99	$49.99
Best Buy	Best Buy	Best Buy	Best Buy
★★★★½ (163)	★★★★½ (161)	★★★★½ (90)	★★★★½ (18)

Case-mate - Brilliance Cas...	Case-mate - Wallet Folio C...	Marc by Marc Jacobs Metalli...	Case-mate - Karat Hard Sh...
$44.99	$54.99	$38.00	$49.99
Best Buy	Best Buy	shopbop	Best Buy
★★★½☆ (294)	★★★★½ (173)		★★★★½ (34)

Search: iphone 6s case

Sponsored Search

- Advertiser sets a bid price for the keyword
- User searches the keyword
- Search engine hosts the auction to ranking the ads

Display Advertising

The New York Times

SUBSCRIBE NOW | SIGN IN | Register

INTERNATIONAL | DEALBOOK | MARKETS | ECONOMY | ENERGY | MEDIA | TECHNOLOGY | PERSONAL TECH | ENTREPRENEURSHIP

Exxon Mobil Investigated in New York Over Possible Lies on Climate

By JUSTIN GILLIS and CLIFFORD KRAUSS
3:30 PM ET

The sweeping inquiry, by the state attorney general, focuses on whether the oil company lied to the public and investors over the risks of climate change.

250 Comments

T. Fallon/Bloomberg, via Getty Images

An Exxon Mobil refinery in Los Angeles, Calif. The New York attorney general is investigating the oil and gas company.

LATEST NEWS

5:01 PM ET	'Grand Theft Auto' Maker Take-Two's Revenue Nearly Triples
5:00 PM ET	United Airlines CEO to Return in Early 2016 After Heart Attack
4:57 PM ET	NY Attorney General Investigating Exxon Over Climate Statements

MARKETS »

At close 11/05/2015

European Union Predicts Economic Gains From Influx of Migrants

By JAMES KANTER
12:10 PM ET

Officials forecast that the three million arrivals expected by 2017 would provide a net gain of perhaps a quarter of 1 percent by that year to the European economy.

INSIGHT & ANALYSIS

COMMON SENSE

Dewey Jury's Deadlock Exposes a System's Flaws

By JAMES B. STEWART
3:06 PM ET

One reason for the mistrial in the Dewey & LeBoeuf criminal case may have been the requirement for a unanimous decision.

http://www.nytimes.com/

Internet Advertising Frontier:

Real-Time Bidding (RTB) based Display Advertising

What is Real-Time Bidding?

- Every online **ad view** can be evaluated, bought, and sold, all **individually**, and all **instantaneously**.
- Instead of buying keywords or a bundle of ad views, advertisers are now **buying users** directly.

	DSP/Exchange	daily traffic
Advertising	iPinYou, China	18 billion impressions
	YOYI, China	5 billion impressions
	Fikisu, US	32 billon impressions
Finance	New York Stock Exchange	12 billion shares daily
	Shanghai Stock Exchange	14 billion shares daily

	Query per second
Turn DSP	1.6 million
Google	40,000 search

Shen, Jianqiang, et al. "From 0.5 Million to 2.5 Million: Efficiently Scaling up Real-Time Bidding." Data Mining (ICDM), 2015 IEEE International Conference on. IEEE, 2015.

Suppose a student regularly reads articles on emarketer.com

eMarketer | Research Topics | Products | Why eMarketer | Customer Stories | Articles

Advertisers Continue Rapid Adoption of Programmatic Buying

By 2017, advertisers will spend more than $9 billion on RTB

Nov 26, 2013 | Share | Print | Email

Advertisers are spending more than expected on real-time bidding, which is expected to account for a significant share of all display ad spending in the US

advertising—which includes RTB—continues its rapid transition from infancy to a well-established display purchase method in just a few years.

eMarketer projects RTB digital display ad spending in the US will account for 29.0% of total US digital display ad spending by 2017, or $9.03 billion. In 2013, it will account for 19.0%, or $3.37 billion. These estimates are revised slightly upward from our previous forecast in August

Latest from eMarketer

Latest Articles | Latest Webinars

- Hispanic Gen Xers Lead in Daily Tablet Usage
- Chrysler's Multichannel Approach to Online Video Gets Greater Recall
- Android Rules UK Smartphone Sales

More Articles » | eMarketer Daily Newsletter »

Contact Sign-Up | Contact Sign-Up | Contact Sign-Up | Contact Sales

Content-related ads

He recently checked the London hotels

(In fact, no login is required)

Relevant ads on facebook.com

Real-time Bidding System

Figure 1. The Business Process of RTB Ad Delivery

①An Internet user visits the webpages of a publisher.

②If there is one or more ad slots to be sold in the RTB markets, the publisher will send an ad request to the SSP and AdX, containing the information of the user, ad slots, and reserve prices.

③ After receiving the ad request from SSP, AdX will forward all the information to eligible DSPs.

④Each DSP then parses the information in the ad request, asks the DMP about the necessary information of this user (e.g., its geographic position, gender, age, historical behavior, shopping interests and intentions, etc.), and starts an auction to matched advertisers. The ad of the winner advertiser will be fed back to the AdX, and each DSP must response within a specific time.

⑤ AdX starts an auction to winner advertisers from each DSP, and determines the highest bid among these winner advertisers. If this bid is lower than the publisher's reserve price, this RTB process will be terminated with the ad slots left empty or reassigned to non-RTB channels. Otherwise, the advertiser with the highest bid will finally win the ad impression.

⑥ AdX announces the auction result to all DSPs, and send the ads of the final winner advertiser to SSP.

⑦ SSP helps display the ads to the ad slots on the publisher's webpages in front of the users.

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Conversion attribution
- Learning to bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

Auctions scheme

Modeling

- n bidders

- Each bidder has value v_i for the item
  - "willingness to pay"
  - Known only to him – "private value"

- If bidder i wins and pays p_i, his utility is $v_i - p_i$
  - In addition, the utility is 0 when the bidder loses.

- <u>Note:</u> bidders prefer losing than paying more than their value.

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Conversion attribution
- Learning to bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

Predict how likely the user is going to click the displayed ad.

User response estimation problem

- Click-through rate estimation as an example

- Date: 20160320
- Hour: 14
- Weekday: 7
- IP: 119.163.222.*
- Region: England
- City: London
- Country: UK
- Ad Exchange: Google
- Domain: yahoo.co.uk
- URL: http://www.yahoo.co.uk/abc/xyz.html
- OS: Windows
- Browser: Chrome
- Ad size: 300*250
- Ad ID: a1890
- User tags: Sports, Electronics

→ Click (1) or not (0)?

Predicted CTR (0.15)

Clickthrough rate (CTR): Definition

A ratio showing how often people who see your ad or free product listing end up clicking it. Clickthrough rate (CTR) can be used to gauge how well your keywords and ads, and free listings, are performing.

- CTR is the number of clicks that your ad receives divided by the number of times your ad is shown: clicks ÷ impressions = CTR. For example, if you had 5 clicks and 100 impressions, then your CTR would be 5%.
- Each of your ads, listings, and keywords have their own CTRs that you can see listed in your account.
- A high CTR is a good indication that users find your ads and listings helpful and relevant. CTR also contributes to your keyword's expected CTR, which is a component of Ad Rank. Note that a good CTR is relative to what you're advertising and on which networks.
- You can use CTR to gauge which ads, listings, and keywords are successful for you and which need to be improved. The more your keywords, ads, and listings relate to each other and to your business, the more likely a user is to click on your ad or listing after searching on your keyword phrase.

support.google.com

Feature Representation

- Binary one-hot encoding of categorical data

x=[Weekday=Wednesday, Gender=Male, City=London]

x=[0,0,1,0,0,0,0 0,1 0,0,1,0…0]

High dimensional sparse binary feature vector

Linear Models

- Logistic Regression
  - With SGD learning
  - Sparse solution

ML Framework of CTR Estimation

- A binary regression problem

$$\min_{\boldsymbol{w}} \sum_{(y,\boldsymbol{x})\in D} \mathcal{L}(y,\hat{y}) + \lambda\Phi(\boldsymbol{w})$$

  – Large binary feature space (>10 millions)

  – Large data instance number (>10 millions daily)
  – A seriously unbalanced label
    - Normally, #click/#non-click = 0.3%
    - Negative down sampling
    - Calibration

Negative downsampling

Posted in ieee-fraud-detection 2 years ago

148

At the previous fraud competition the winner team used negative downsampling to reduce data size and create big sets of features without quality lost.
Negative downsampling = downsampling only transactions where isFraud==0 and use all transactions with isFraud==1

```
# Example: Negative downsampling

train_pos = train[train['isFraud']==1]
train_neg = train[train['isFraud']==0]

train_neg = train_neg.sample(int(train.shape[0] * 0.2), random_state=42)
train = pd.concat([train_pos, train_neg]).sort_index()
```

It could speed up training process by 4-5 times without losing the quality. There is very low difference between sample (0.9489), full data (0.9496) and LB (0.9496) in my CV.

It gives us ability to check many different hypotheses fast and to find best features within short time. When we find best features/parameters we could use full data to get maximum score.

Logistic Regression

- Prediction

$$\hat{y} = \frac{1}{1 + e^{-\boldsymbol{w}^T \boldsymbol{x}}}$$

- Cross Entropy Loss

$$\mathcal{L}(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y})$$

- Stochastic Gradient Descent Learning

$$\boldsymbol{w} \leftarrow (1 - \lambda)\boldsymbol{w} + \eta(y - \hat{y})\boldsymbol{x}$$

[Lee et al. Estimating Conversion Rate in Display Advertising from Past Performance Data. KDD 12]

Logistic Regression with SGD

$$\boldsymbol{w} \leftarrow (1-\lambda)\boldsymbol{w} + \eta(y - \hat{y})\boldsymbol{x}$$

- Pros
  - Standardised, easily understood and implemented
  - Easy to be parallelised
- Cons
  - Learning rate η initialisation
  - Uniform learning rate against different binary features

Online learning

Booking.com

Online learning set-up.

Online learning with streaming data, when the distribution of the incoming examples changes over time: the examples are not identically distributed. Suppose you have a learning module for predicting the click-thru-rate of online-ads, the distribution of query terms coming from the users are changing during the year dependent on seasonal trending. The query terms in summer and in Christmas season should have different distribution.

Ref: https://stackoverflow.com/questions/13058379/example-for-non-iid-data

The data

- Data is **no longer i.i.d.**
- Sequence of observations is **ordered**

$$\mathcal{O}_t = (x_1, y_1), (x_2, y_2), \ \dots \ , (x_{t-1}, y_{t-1})$$

The objective

- Predict the next y_t given the next x_t as well as all observations $\mathcal{O}_t$ so far:

$$\hat{y}_t = h(x_t, \theta_t)$$

Online learning.

Remember batch loss:

$$\mathrm{Loss}(\theta) = \frac{1}{N}\sum_{n=1}^{N} \ell(x_n, y_n, \theta)$$

Instead, we **minimize Regret**

$$\mathrm{Regret}(\vartheta, \theta_*) = \sum_{t=1}^{T} \Big(\ell(x_t, y_t, \theta_t) - \ell(x_t, y_t, \theta_*) \Big)$$

where

$$\vartheta = \{\theta_1, \ldots, \theta_T\}$$

$$\theta_* = \arg\min_{\theta} \sum_{t=1}^{T} \ell(x_t, y_t, \theta)$$

Online learning.

Simplest approach: **Online Gradient Descent**

$$\begin{aligned} \theta_{t+1} &= \theta_t - \eta\, g_t \\ g_t &= \nabla_\theta\, \ell(x_t, y_t, \theta_t) \end{aligned}$$

where η can be simple learning rate schedule:

$$\eta = \frac{\alpha}{\sqrt{t}}$$

or adaptive per-coordinate learning rate (AdaGrad):

$$\eta_i = \frac{\alpha}{\sqrt{\sum_{s=1}^{t} g_{i,t}^2}}$$

AdaGrad

Edit

AdaGrad is a stochastic optimization method that adapts the learning rate to the parameters. It performs smaller updates for parameters associated with frequently occurring features, and larger updates for parameters associated with infrequently occurring features. In its update rule, Adagrad modifies the general learning rate η at each time step t for every parameter θ_i based on the past gradients for θ_i:

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} g_{t,i}$$

The benefit of AdaGrad is that it eliminates the need to manually tune the learning rate; most leave it at a default value of 0.01. Its main weakness is the accumulation of the squared gradients in the denominator. Since every added term is positive, the accumulated sum keeps growing during training, causing the learning rate to shrink and becoming infinitesimally small.

Image: Alec Radford

Online learning.

When to use online learning?

- when consecutive observations are not i.i.d.
- when you care about errors you make early in the learning process
- underlying “state of the environment” changes over time

  *e.g. binary classification for **spam detection***

Follow the (regularized) leader

Why do we expect online learning to work for batch learning?

Booking.com

$$\ell_s(\theta) = \ell(x_s, y_s, \theta)$$

Follow The (Regularized) Leader.

The idea

- Adjust course "now", but in a way that would **also minimize prior loss**
- Update based on **total loss so far**

$$\theta_{t+1} = \arg\min_{\theta} \sum_{s=1}^{t} \ell_s(\theta)$$

- Add regularization:

$$\theta_{t+1} = \arg\min_{\theta} \sum_{s=1}^{t} \ell_s(\theta) + \frac{1}{2\eta}\|\theta\|^2$$

arxiv.org/abs/1403.3465

Shalev-Shwartz survey

Follow The Regularized Leader.

Online Gradient Descent from FTRL

$$
\begin{aligned}
\theta_{t+1} &= \arg\min_{\theta} \operatorname{Regret}_t(\theta, \theta_*) \\
&\approx \arg\min_{\theta} \sum_{s=1}^{t} \theta \cdot g_s + \frac{1}{2\eta}\|\theta\|^2 + \text{const} \\
&= -\eta \sum_{s=1}^{t} g_s \\
&= \theta_t - \eta\, g_t
\end{aligned}
$$

Follow The Regularized Leader.

In short,

- FTRL attempts to minimize Regret *explicitly*, albeit greedy
- FTRL is a general framework that encompasses many algorithms (incl. OGD)
- FTRL picks weights at each round to get closer to the optimum θ_*
  (thus good candidate for batch optimization)

Neat examples

Booking.com

Paid search results on Google.

Paid search results on Google.

Context

- Compute bids for Google AdWords
- We have a **very large number** of distinct keywords
- Runs daily

Old setup

- ML framework: pyspark.ml
- Training time: **90 minutes**
- Prediction time: **2 hours**

Current setup

- ML framework: Vowpal Wabbit
- Training time: **10 minutes** (single instance)
- Prediction time: **2 minutes** (parallelized on Hadoop)

VOWPAL WABBIT

https://github.com/VowpalWabbit/vowpal_wabbit

Vowpal Wabbit
Get started
Features
Tutorials
Blog
Research
Wiki
GitHub
Your go-to interactive machine learning library
Vowpal Wabbit provides a fast, flexible, online, and active learning solution that empowers you to solve complex interactive machine learning problems.
Get started
Tutorials