Large Scale Recommendation in a RTB Platform

Suju Rajan

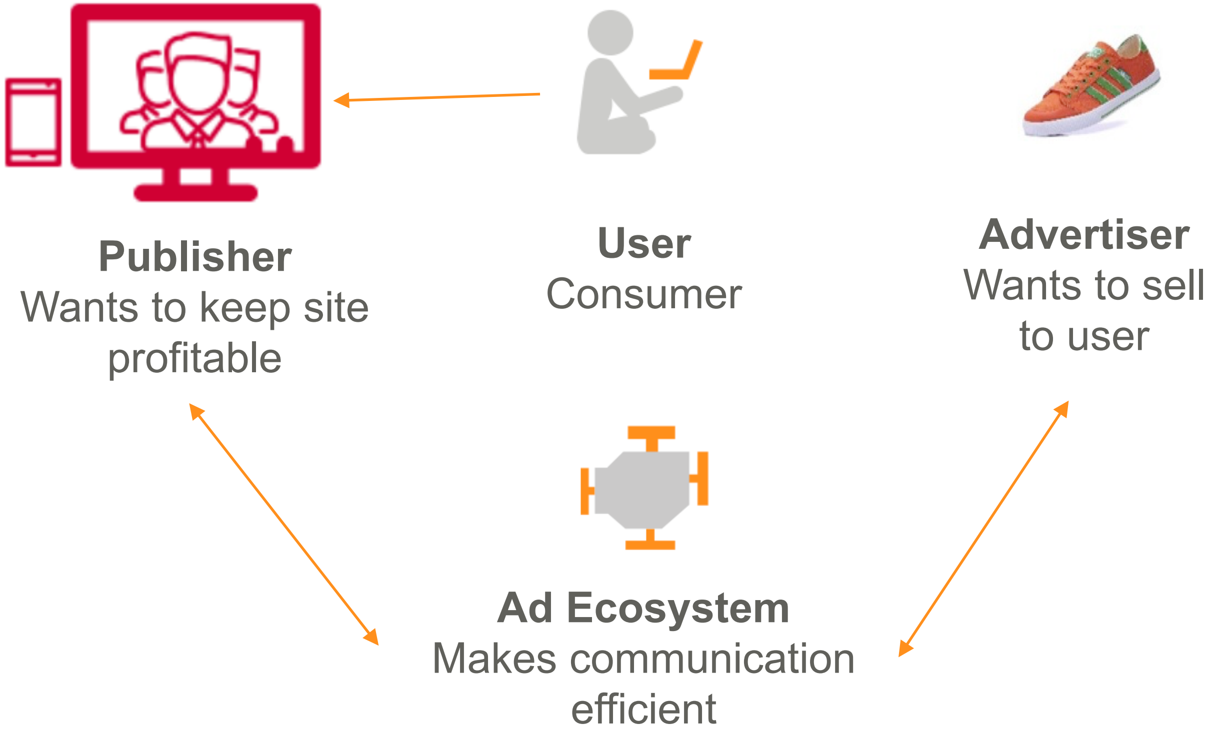
VP, Head of Research
Criteo

Wait, what is RTB?

Why do we need online advertising?

Maintains the free use of internet

Who are the main players?



criteo research

Wait, what is RTB?

Imagine, managing the business of showing an ad every time someone uses an online “site”.

Many different options of which participating in Ad Exchanges is one.

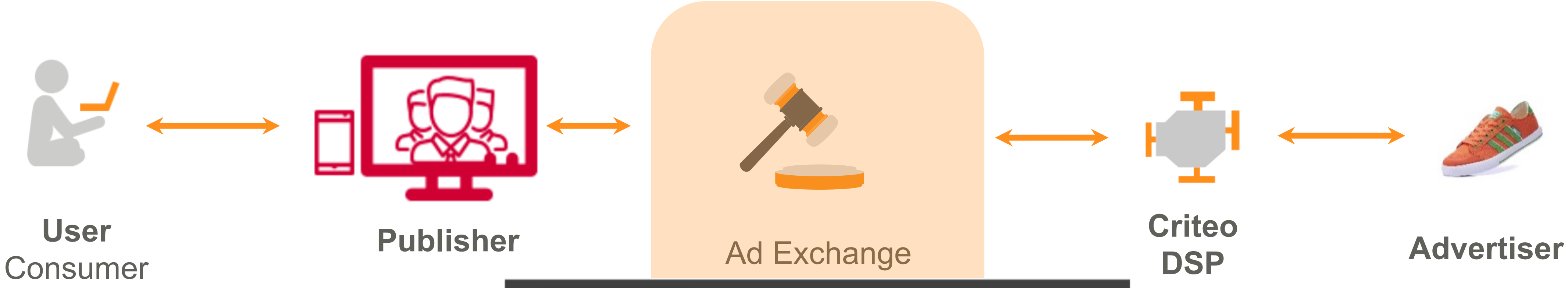
RTB = Real Time Bidding

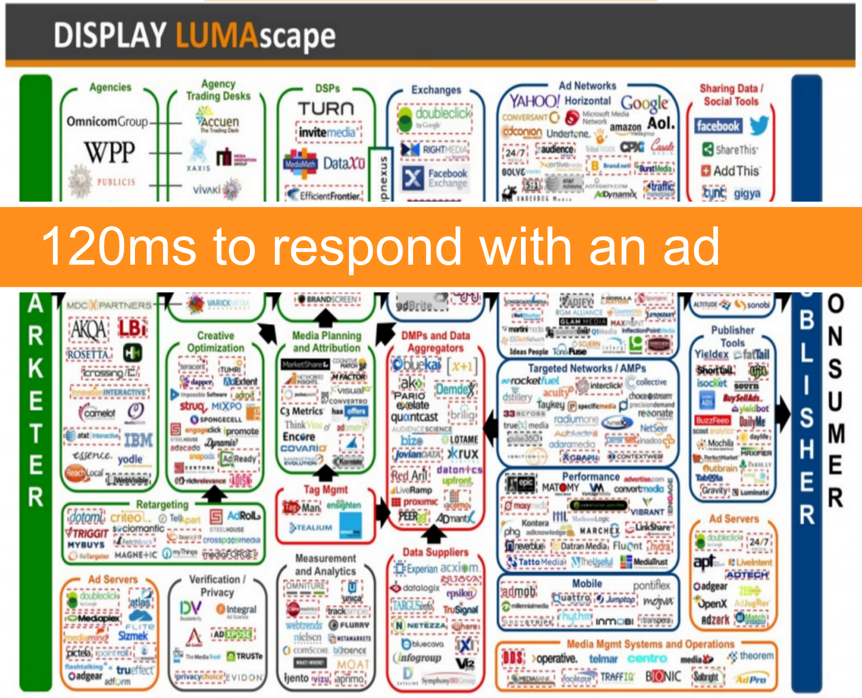
Simply stated:

- Publisher has a display opportunity (user, page, slot, timestamp)
- Display opportunity auctioned into an ad exchange
- Ideally, 2nd price auction winner gets to display an ad to the user

criteo research

As an example, at Criteo:





120ms to respond with an ad

RTB at Scale:

1: REVENUE IN 2016
2: ANNUAL RATE 2016
3: $ OF TURNOVER GENERATED TO OUR CLIENTS
- TURNOVER POST-CLICK WW FROM JANUARY TO
DECEMBER 2016

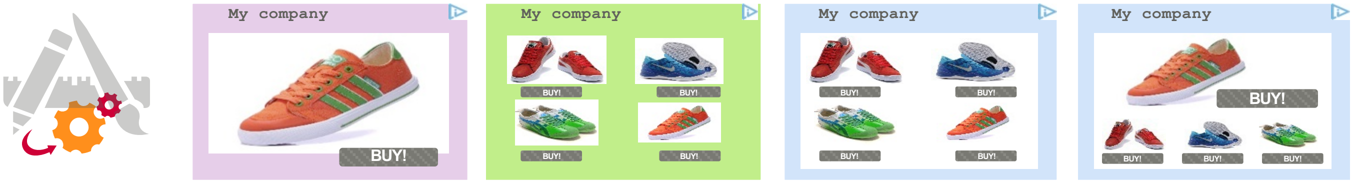
criteo
research

Main questions to answer:

1. How much should we bid for a given ad space?

2. What products should we recommend/show?

3. What is the best look and feel of the banner?



COMMON OBJECTIVE:

Maximize client (advertisers)'s value

criteo. research

Recommendation at Criteo

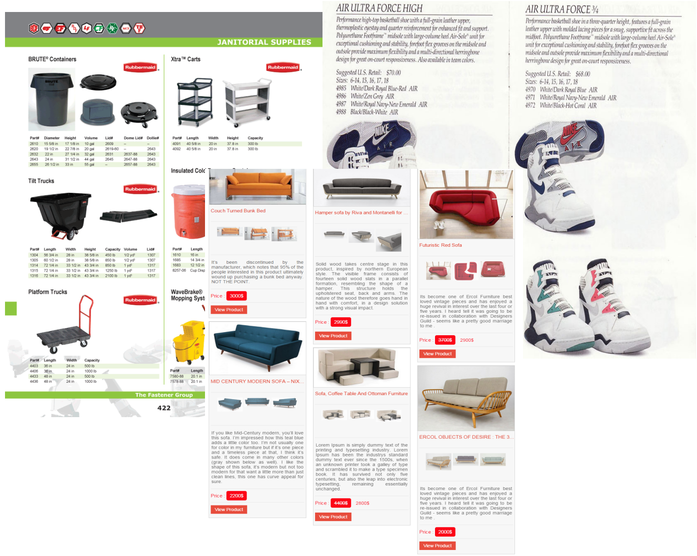
Recommendation Challenge

We have successfully bid and won an ad impression on behalf of an advertiser. Now we need to decide the right product to put in front of the user.

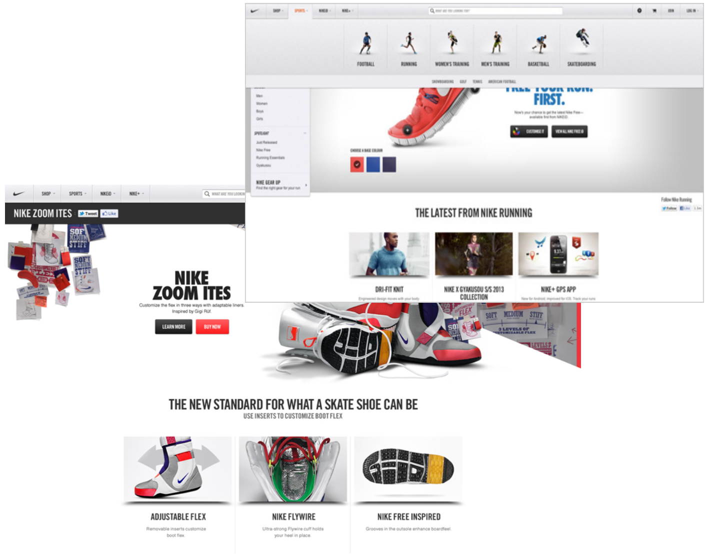
We have less than 100ms to respond.

criteo research

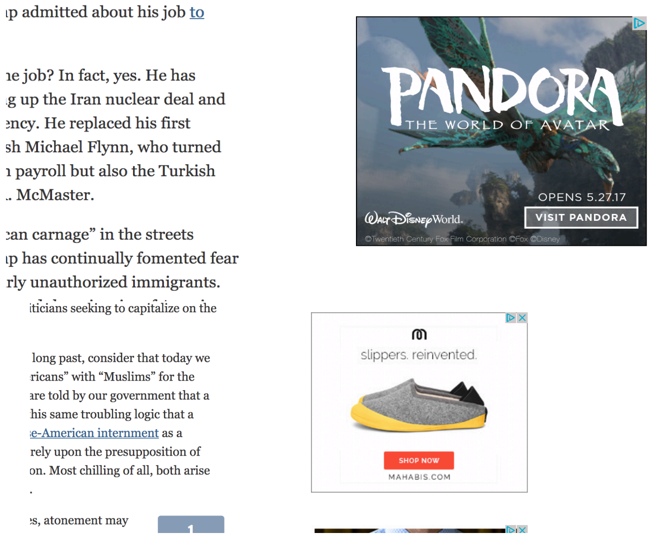
Recommendation Challenge : Data Sources



Advertiser Catalogs
~3B+ products



Advertiser Site Events
~2B+ events/day

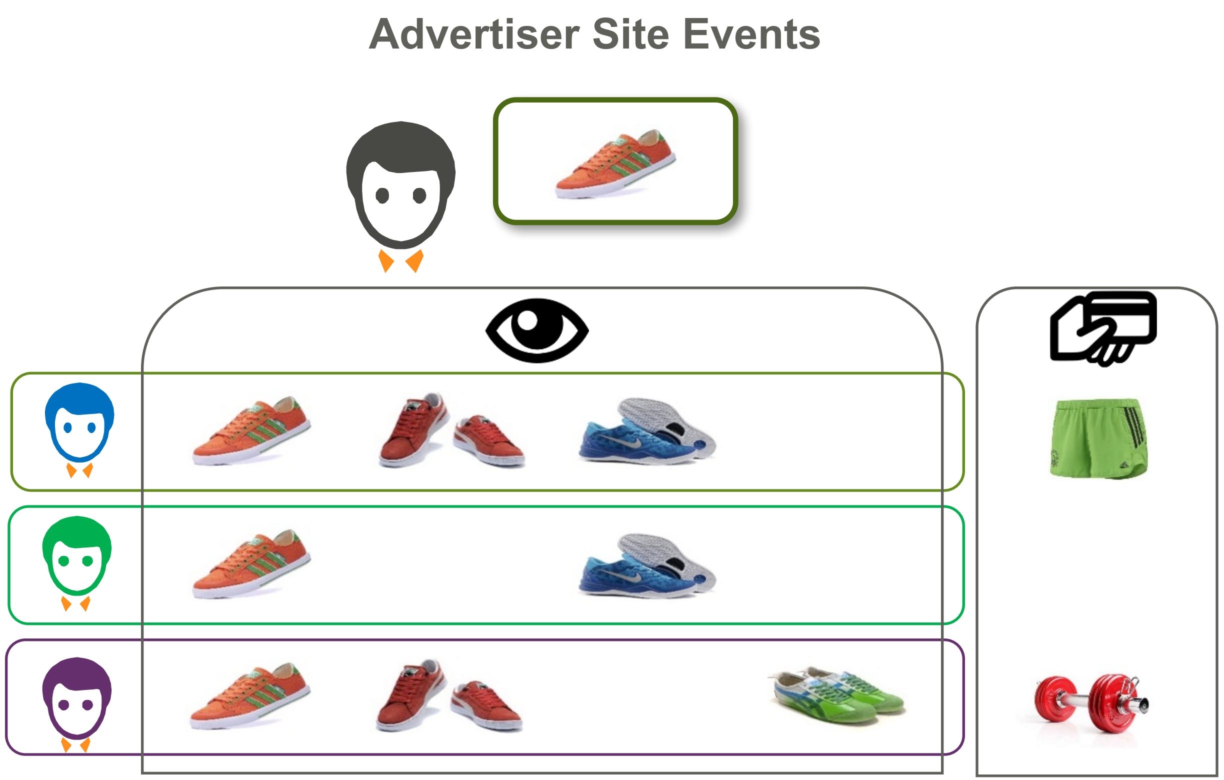


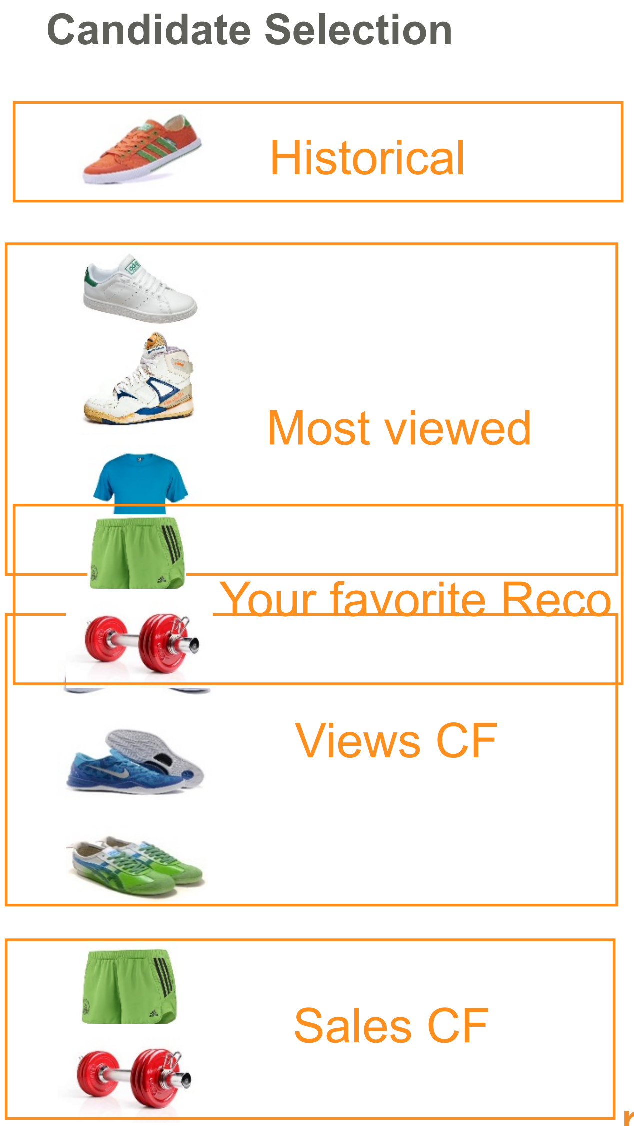
Ad Display Events
~20B+ events/day

Users
~1B

criteo.
research

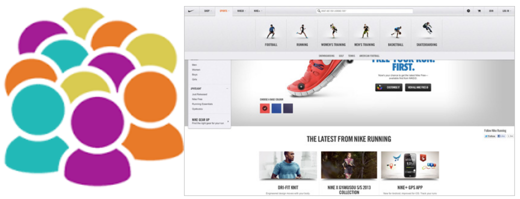
Recommendation Stage 1: Candidate Selection





criteo research

Recommendation Stage 1: Candidate Selection



Users
~1B

50B Events

Multiple times
a day

Source 1

Source 2

Source 3

Source n

~500M
entries

criteo research

Recommendation Stage 2:Ranking

We select a thresholded number of products from each source

De-dupe the sources

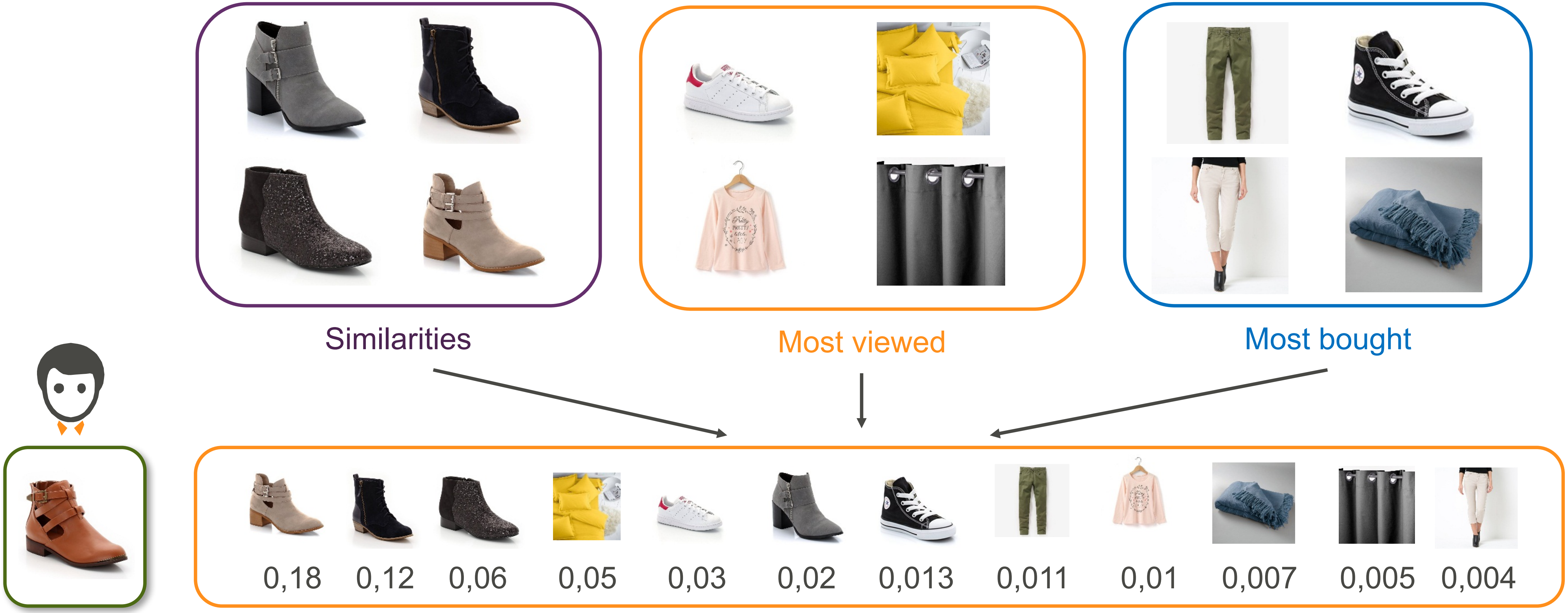
The reduced set of products are then ranked by a LR model that tries to maximize the probability of sale of a product

Product-specific | User-specific | User-product interactions | Display-specific

Feature Space

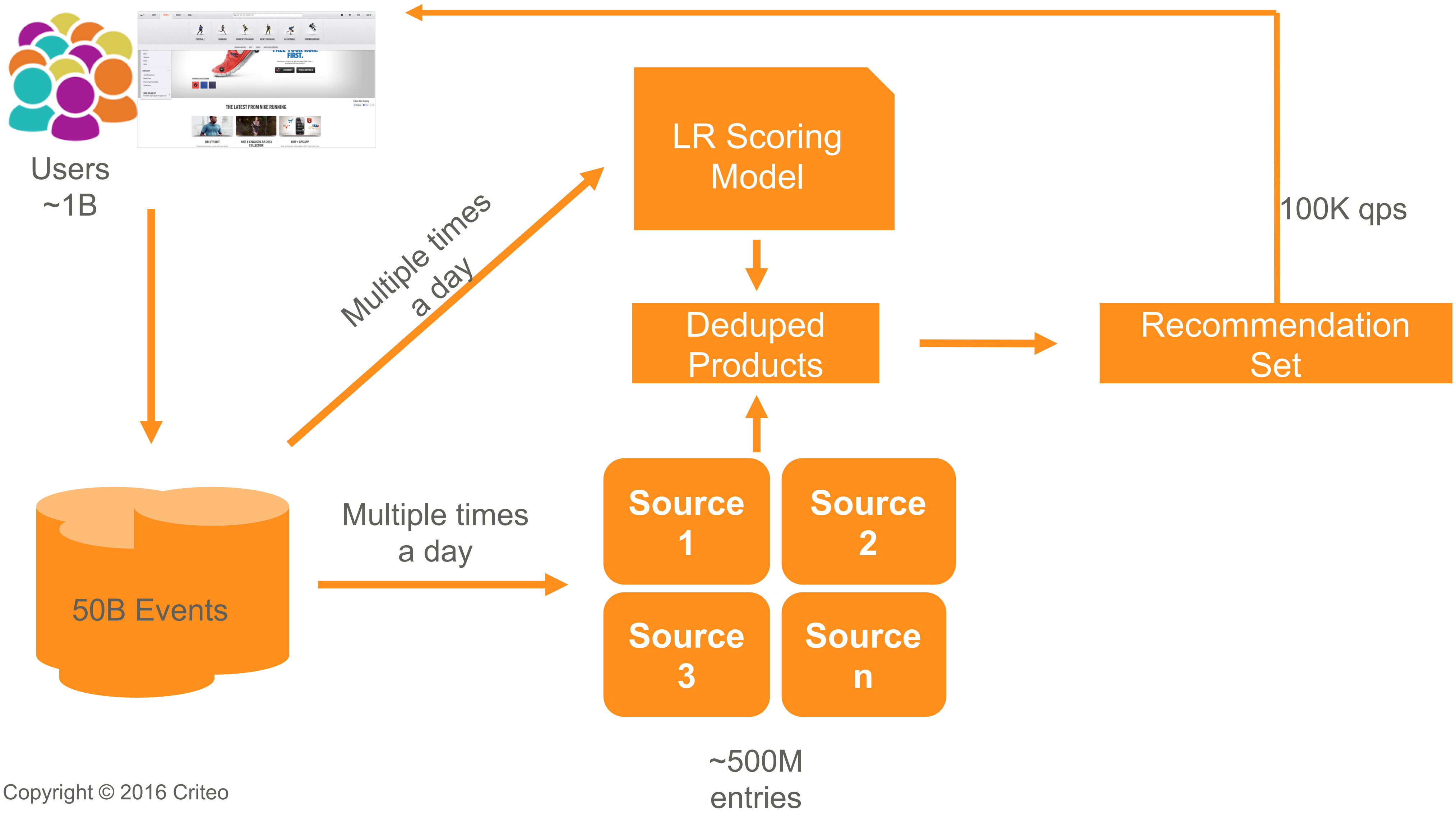
criteo research

Recommendation Stage 2:Ranking



criteo research

Recommendation Stage 2: Ranking

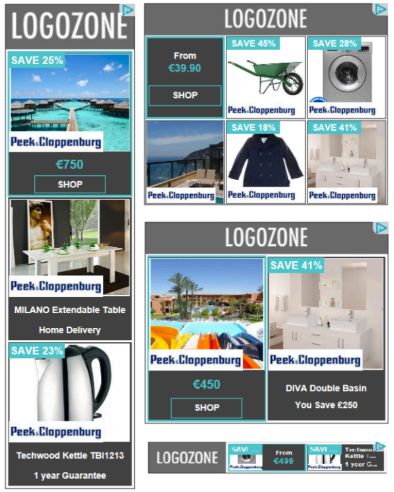


Enabling counterfactual analysis

We don't want to only display top-k products selected by LR

No exploration limits the performance of models learned from the biased data

Instead we sample the products to be shown from a multinomial distribution defined by the LR scores (f_p)



$$P(slot1 = p) = \frac{f_p}{\sum_{\{p' \in P_c\}} f_{p'}} \qquad P(slot2 = p' \mid slot1 = p) = \frac{f_{p'}}{\sum_{\{p^\dagger \in P_c \wedge p^\dagger \neq p\}} f_{p^\dagger}}, \quad \text{etc.}$$

criteo research

Dataset released for evaluation of Policy Learning Algorithms

Has ~100M ad impressions

8500+ banner types (Top 10 = 30% of impressions)

Upto 6 displayed products with a candidate pool that is 10 times the number of displayed products

21M impressions for 1-slot and over 14M for 6-slot banners

Has a subset of product features

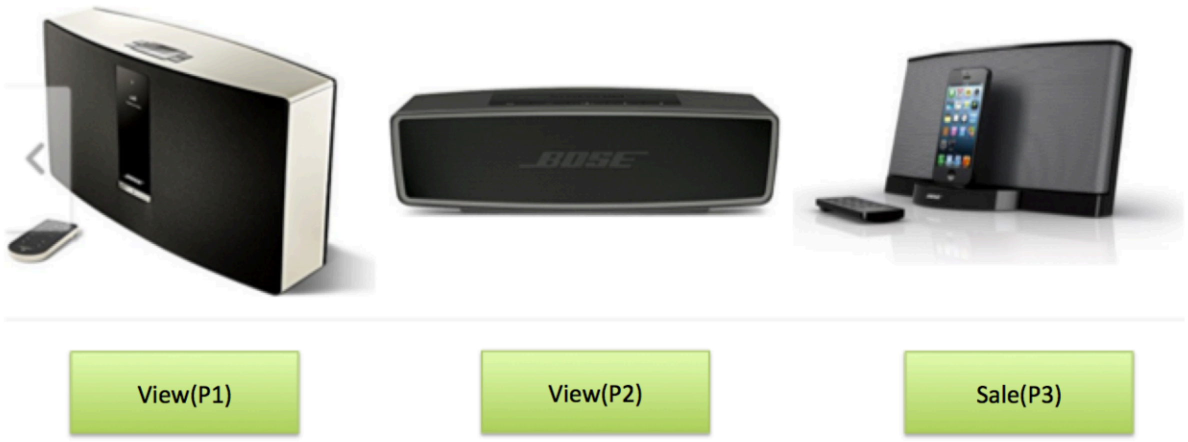
Dataset pointer at research.criteo.com

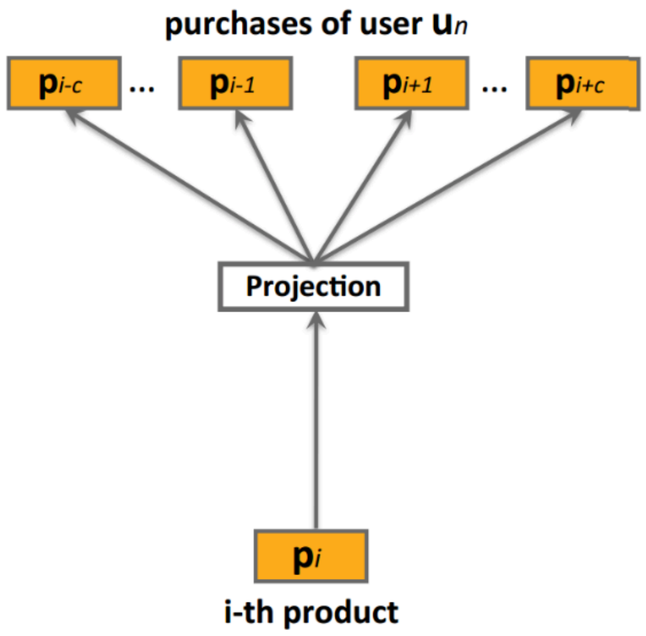
Hey, where’s the Deep Learning?!!

How about a little Prod2Vec first? [Grbovic et al., WWW 2015]

Word2Vec: Words that appear in similar context get embedded into a space where they are closer

Apply the same idea to user & product interaction sequences: Prod2Vec





criteo research

MetaProd2Vec [Vasile et al., RecSys 16]

Prod2Vec + Product Meta Data (Example: Categories, Brands, etc.)

Place additional constraints on product co-occurrence based on meta data

Helps create noise-robust embeddings specifically in cold-start cases

criteo research

MetaProd2Vec [Vasile et al., RecSys 16]

$$L_{MP2V} = L_{J|I} + \lambda \times \left(L_{M|I} + L_{J|M} + L_{M|M} + L_{I|M}\right)$$

M	metadata space
λ	hyperparameter that expresses importance of extra constraints
$L_{M\|I}$	constraint #1
$L_{J\|M}$	constraint #2
$L_{I\|M}$	constraint #3
$L_{M\|M}$	constraint #4

Constraint 3: Product plausibility given brand

Constraint 4: Next brand plausible given the current one

criteo research

MetaProd2Vec [Vasile et al., RecSys 16]

Dataset: 30Music Dataset

- *Playlists data from Last.fm API*
- *Sample of*
- *Resulting v*

Task: Next ev

- *Hit Ratio @ K*
- *NDCG*

Method: Cold Start	HR @20 (Pair freq=0)	HR@20 (Pair freq<3)
Rank by Popularity	0.0002	0.0002
Prod2Vec	0.0003	0.0078
MetaProd2Vec	0.0013	0.0198

criteo research

WIP: Content2Vec [Nedelec et.al, Under Review]

Take into account all product signal (image, text, co-occurrences etc).

Assume final task is one of predicting “co-event”, such as, co-view

1. Find the representation that optimizes P(co-event)

2. Merge the representations from different signals

Showing promising results on cold-start case improving over individual models

criteo research

WIP

Causal embeddings

Modeling attribution to recommendations

Deeper user profiles

Prospecting or user cold start

Thanks!

criteo
research