Causally Driven Incremental Multi Touch Attribution Using a Recurrent Neural Network

Ruihuan Du[†], Yu Zhong[†], Harikesh S. Nair[*], Bo Cui[†], Ryan Shou[†]

[†]JD.com
[*]JD.com and Stanford University

京东

Thanks To

Xi Chen, PM, JD.com

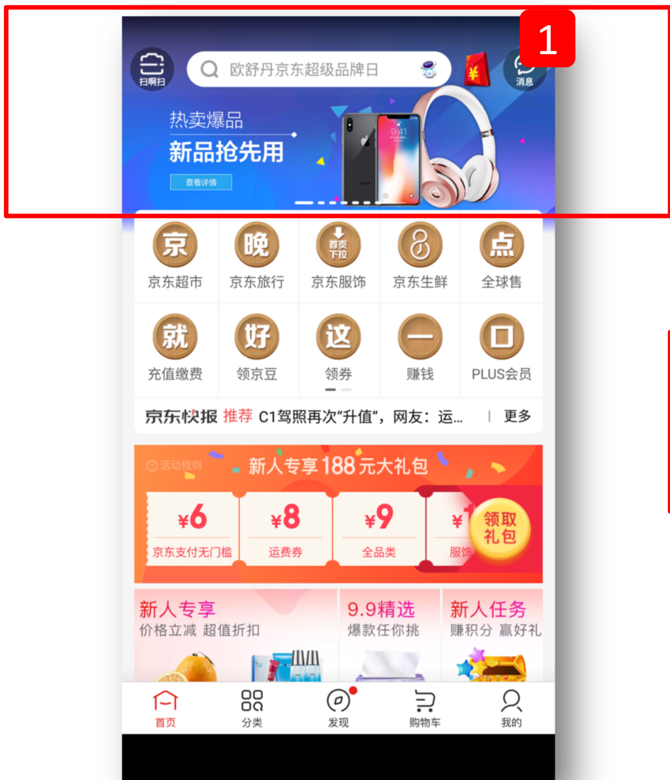
Jack Lin, Sen. Dir., JD.com

Paul Yan, VP, JD.com

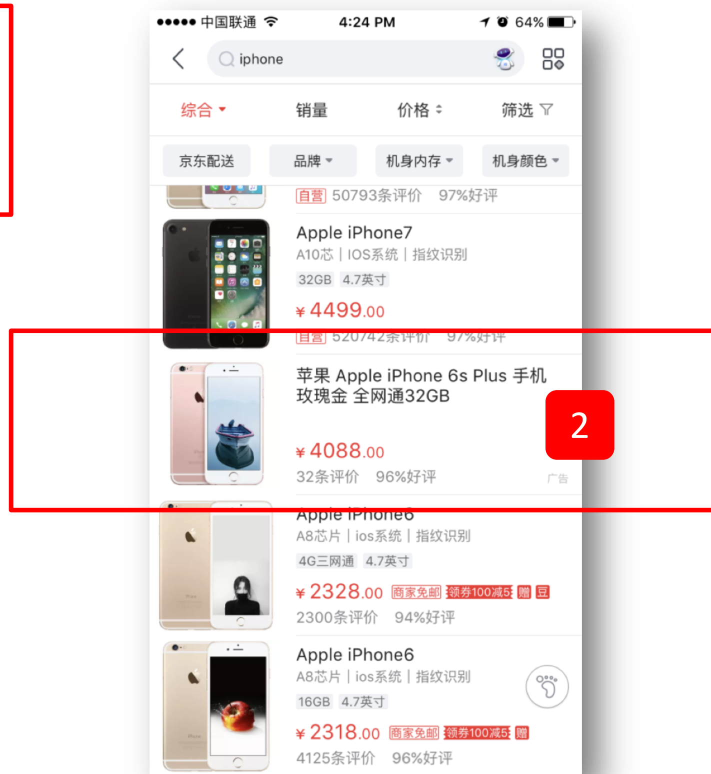
京东

JD is both an eCommerce platform and a publisher of ads

Front facing banner ads



Search ads



SKU recommendation display ads

Brands buy these ads and want us – as the publishing platform – to help them understand how they perform

Multi Touch Attribution Problem: How much of the purchase propensity was due to each ad/touchpoint?

Overall Goal

MTA

- Help advertisers understand which ad-inventory performs best
- Help advertisers implement better campaigns on JD
- Input to budget allocation and bidding
- Better ad pricing

- Develop an MTA product that could be provided as an "add-on" for large advertisers or as a paid-service

Ad-Business is “high scale”

Number of ad-positions	300+
Number of ad-impressions per day	5B
Number of advertisers	75K
Number of product categories	180
Number of brands	140K
Number of Users	300M+

Marginality and Efficiency

- Marginality Principle:
  - Credit allocated to a unit should depend only on touchpoint's marginal contributions to total incremental benefit
- Allocative Efficiency Principle:
  - Credit allocations should sum up to total incremental benefit without waste
- Question: figure out a credit allocation policy that satisfies both marginality and efficiency
  - Difficult, because in general, sum of marginal contributions will not add up to total incremental benefit (think increasing or decreasing returns from co-operation)

Remarkable Result: The Shapley Value

- The Shapley Value is the unique allocation rule that satisfies incrementality and efficiency
- Has a lot of theoretical appeal as a fair allocation system

This Project

- Uses the Shapley Value as the attribution mechanism (fairness)
- Defines it appropriately to only allocate incremental benefits from advertising to underlying ad-positions (incrementality)
- Uses the data to learn marginal-effects, which form an input into the computation of the Shapley Values (data-driven)
  - Train an RNN on user-data and obtain marginal effects from this response model

What's New

- Empirical literature on data-driven attribution (response model + credit allocation system)
  - Empirical response models following Shao and Li (2011): data-driven, simple forms of dependence, not focused on fairness + incrementality
  - Dalessandro et al. (2012) was the first to propose using the Shapley value as a credit allocation mechanism for the MTA problem. They call this "causally-motivated" attribution because of the causal interpretation associated with the "marginality" property of the Shapley Value rule. Simple forms of dependence. Yadagiri et al. (2015) presents semi-parametric extension.
  - This paper: explicit consideration of role of sequence
    - RNN (nothing new in terms of methods), but first application to MTA problem in combination with credit allocation to our knowledge. Some new aspects in specification to ad-response.
    - Shapley Value computation – allocation at ad-position-day level, plus exact aggregation. Efficient MapReduce algorithm for application at scale.
- Theory: Agarwal et al. (2009); Wilbur and Zhu (2009); Jordan et al. (2011); Hu et al. (2016); Berman (2018) propose efficient contracts

Selection & Confounding

A limitation of our approach and of all the response models
cited previously, is the lack of exogenous variation in user
exposure to advertising. Extant papers that have trained ad-response
models on data with full or quasi-randomization
have done so at smaller scale, over limited
number of users and ad-types due to the cost and complexity
of such randomization, and have not considered the corresponding
credit-allocation problem. The approach adopted
here is to include a large set of user features into the model
as synthetic control, so that by including these, we convert
a “selection on unobservables” problem into a “selection
on observables” problem. Controlling flexibly for these observables
mitigates the selection issue somewhat, albeit not
Perfectly.

Overall Architecture

Training step

1 — Train a model for ad-response on historical data

Attribution step

2.1 — Pick all observed orders for brand on a day

2.2 — Find all ad-exposures by brand to that user over past *T* days

2.3 — Use the trained model to compute marginal effects and Shapley Values to attribute order to those ad-exposures

2.4 — Aggregate to desired level and report to advertiser

In paper

1. Define incremental benefit from a brand's ads to an observed order
2. Define a Shapley Value over ad-position-day tuples that occurred before order, such that they sum to this incremental benefit
3. Develop a scalable algorithm that mixes Monte Carlo simulation and exact computation to compute Shapley Values in few hours (for daily reporting)

Overall Architecture

Training step

1. Train a model for ad-response on historical data

Attribution step

2.1 Pick all observed orders for brand on a day

2.2 Find all ad-exposures by brand to that user over past *T* days

2.3 Use the trained model to compute marginal effects and Shapley Values to attribute order to those ad-exposures

2.4 Aggregate to desired level and report to advertiser

We need..

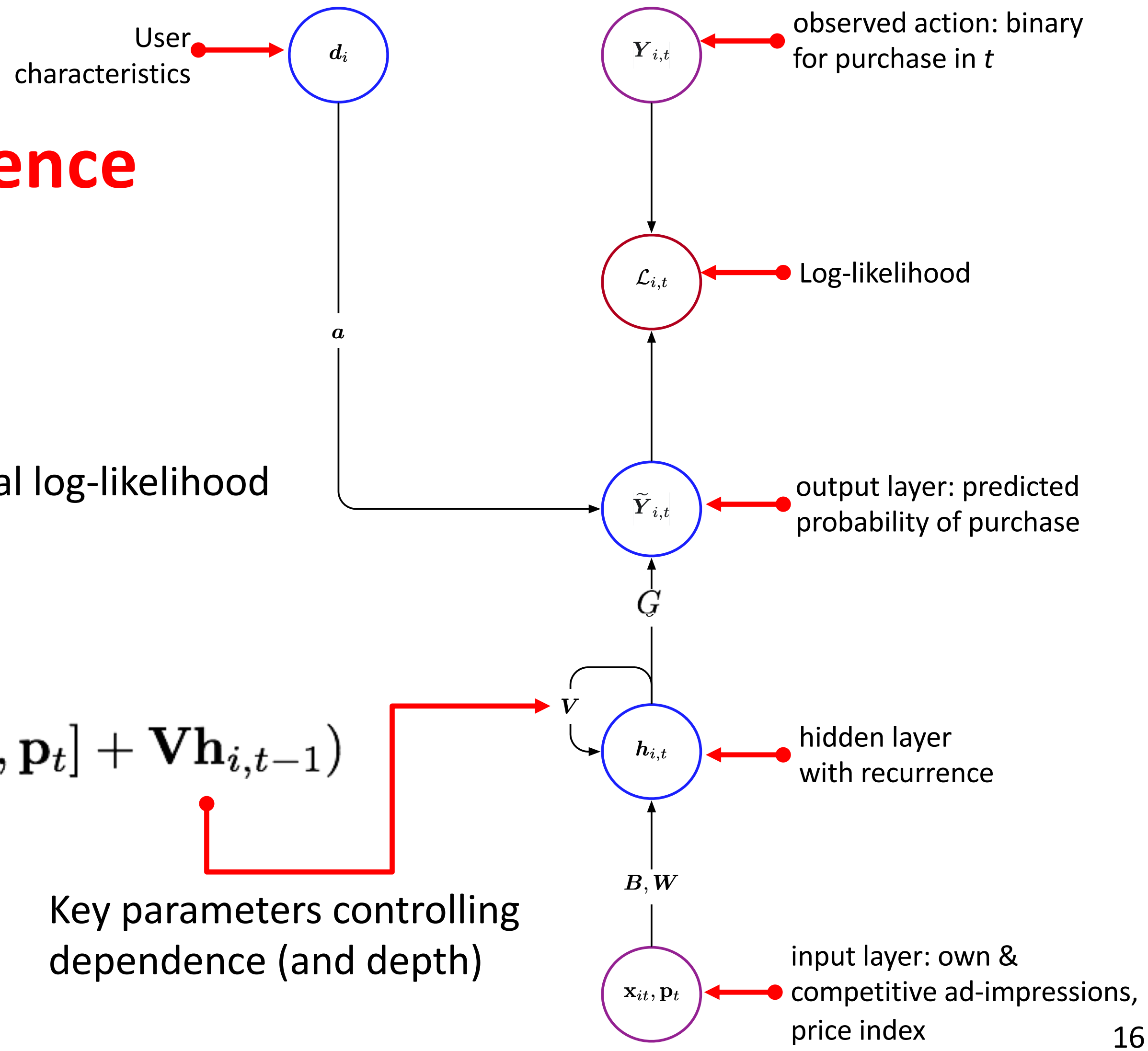
- A response model that takes the sequence into account
  - *Intensity* of ad-exposure
  - *Timing* of ad-exposure
  - *Stock effects* of ad-exposure
  - *Competitive* ad-exposure effects
  - *User heterogeneity* and (limited) selection into exposure
  - Accommodates dimension: roughly 140,000+ brands across roughly 180 categories

RNN with
Hidden-Hidden Recurrence

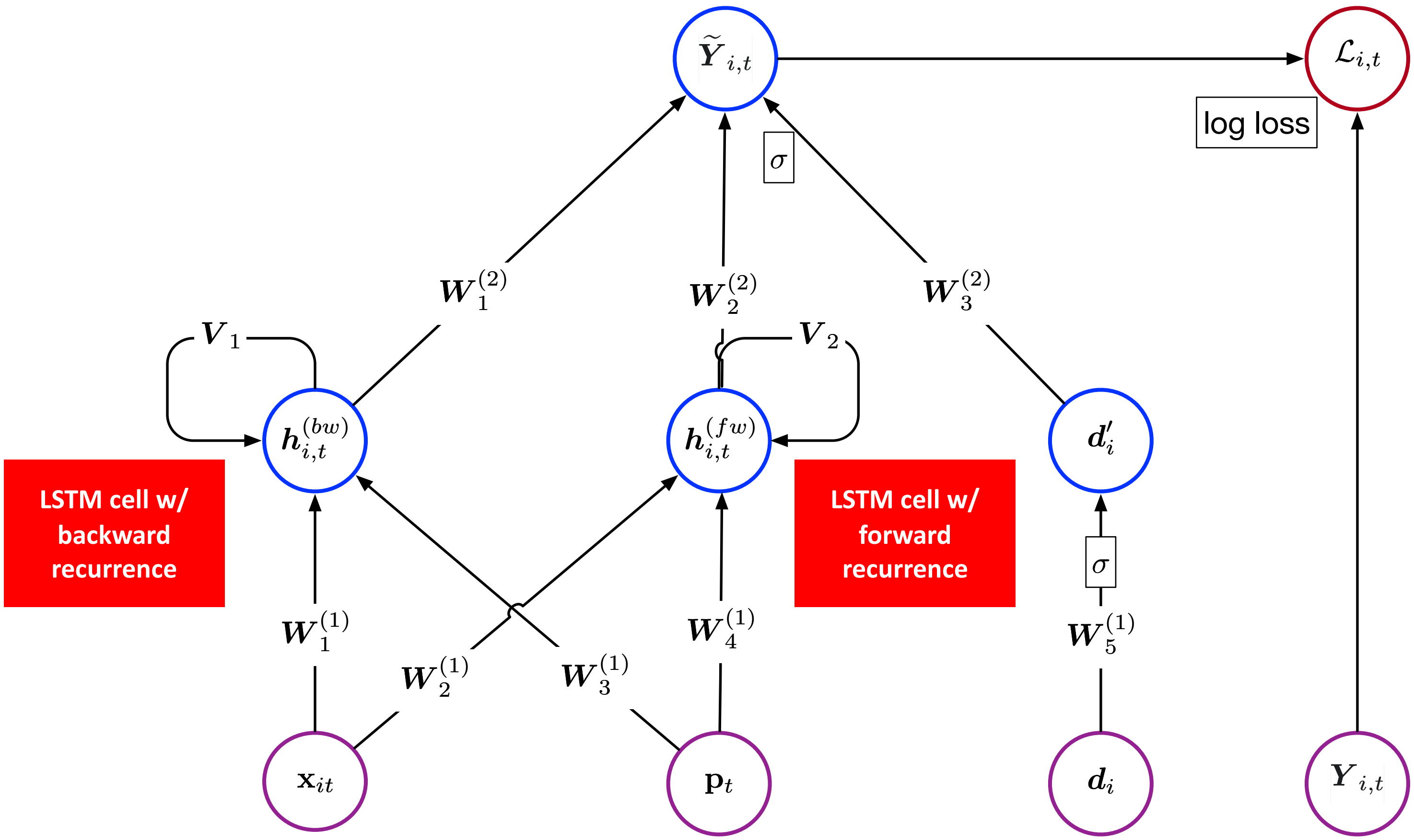
$$\mathcal{L}_i = \sum_{t=1}^{T} \mathcal{L}_{it}$$ Overall loss = total log-likelihood

$$\tilde{\mathbf{Y}}_{it} = \texttt{sigmoid}\left(\mathbf{a}\mathbf{d}_i + \mathbf{G}\mathbf{h}_{it}\right)$$

$$\mathbf{h}_{it} = \texttt{sigmoid}\left(\mathbf{B} + \mathbf{W}\cdot[\mathbf{x}_{it}, \mathbf{p}_t] + \mathbf{V}\mathbf{h}_{i,t-1}\right)$$



Bi-directional LSTM RNN



RNN has ~18M “parameters”

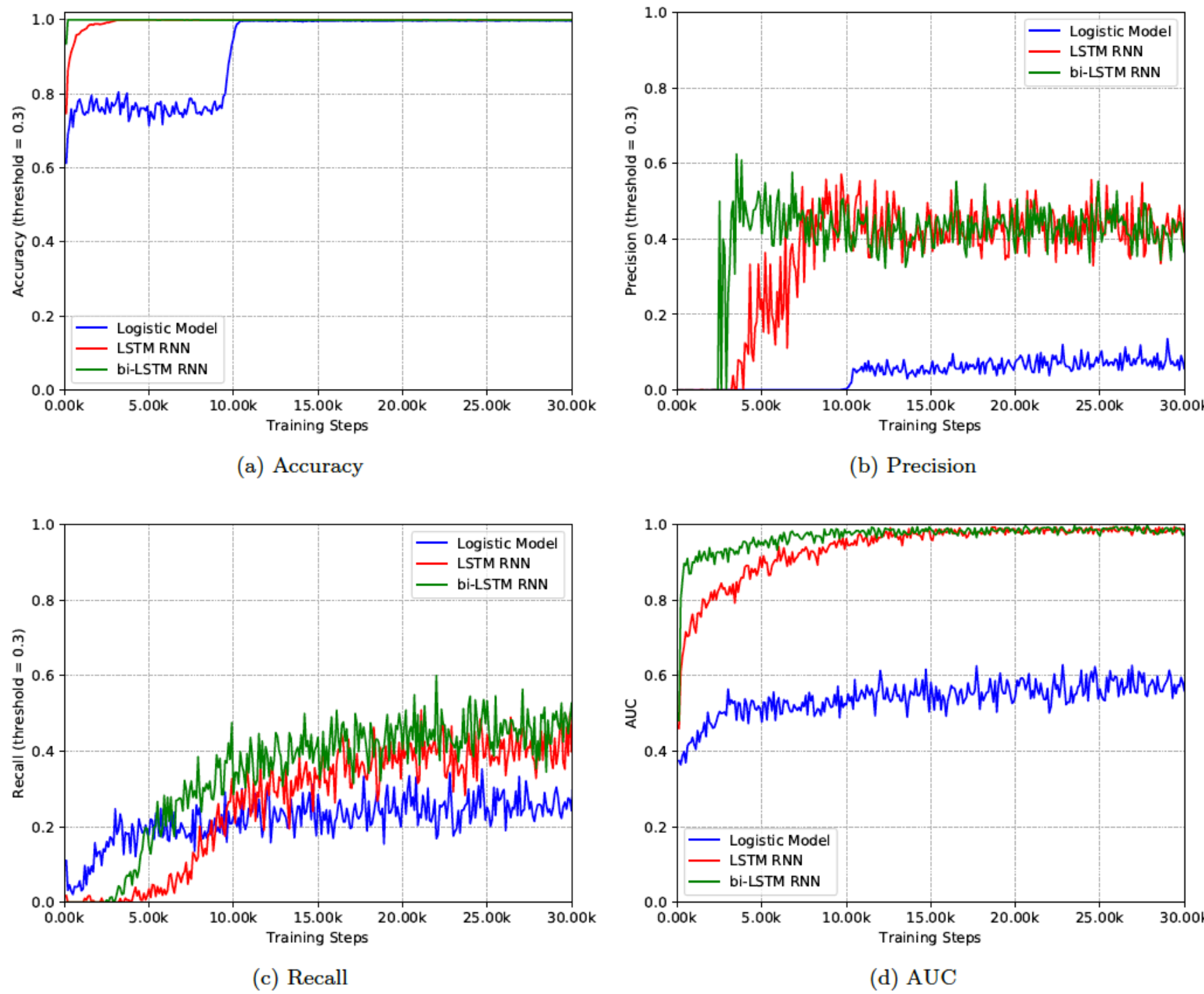
TensorFlow Code: Simulates data, trains RNN, computes Shapley Values

`https://github.com/jd-ads-data/jd-mta`

Benchmarks

Cell-phone ad exposure and purchase data, 15-day window, 2017, JD.com

Number of users in Overall Sample	75,768,508
Number of users in Positive Sample	2,100,687
Number of users in Negative Sample	73,667,821
Num: of ad-impressions in category	7,153,997,856
Num: of orders in category	3,477,621
Number of orders made on day $T = 15$	175,937
Number of brands (B)	31
Number of ad-positions (K)	301



(a) Accuracy

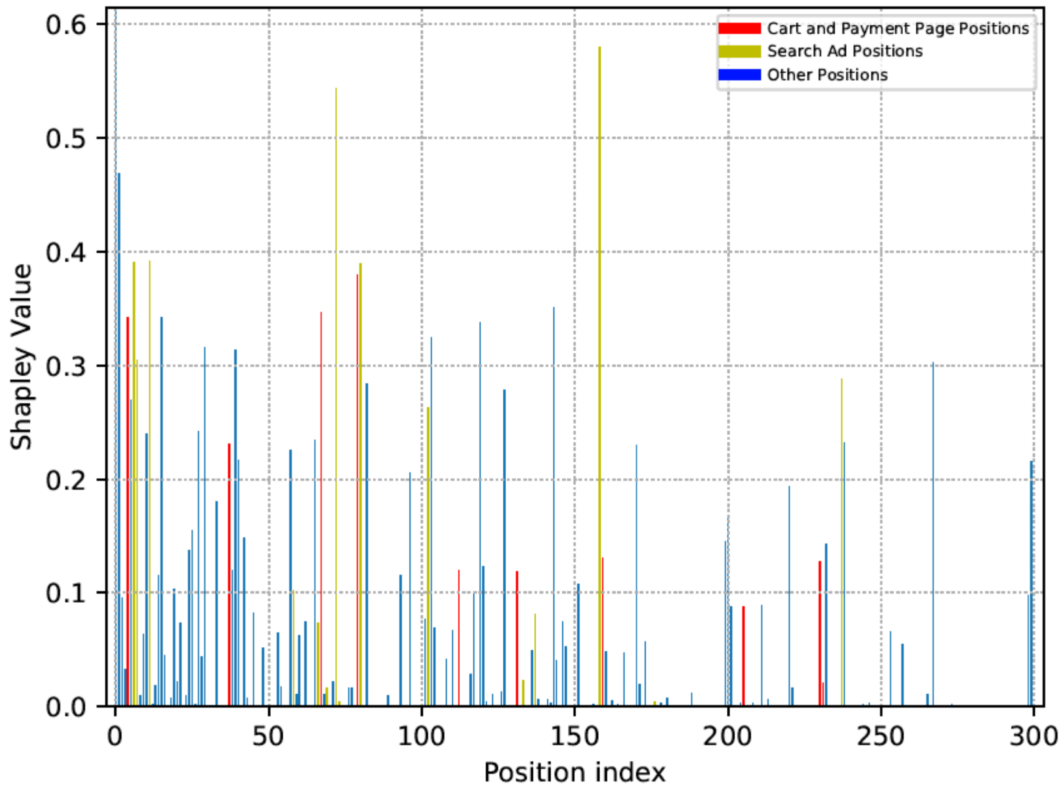
(b) Precision

(c) Recall

(d) AUC

Notes: The figures benchmark the accuracy, precision, recall and AUC (area under the curve) statistics of the bi-directional RNN, against a unidirectional LSTM RNN and a logistic model.

Figure 8: Shapley Values from RNN Model at each Ad-position, Averaged Across all Orders in which that Ad-position was the "Last-clicked"



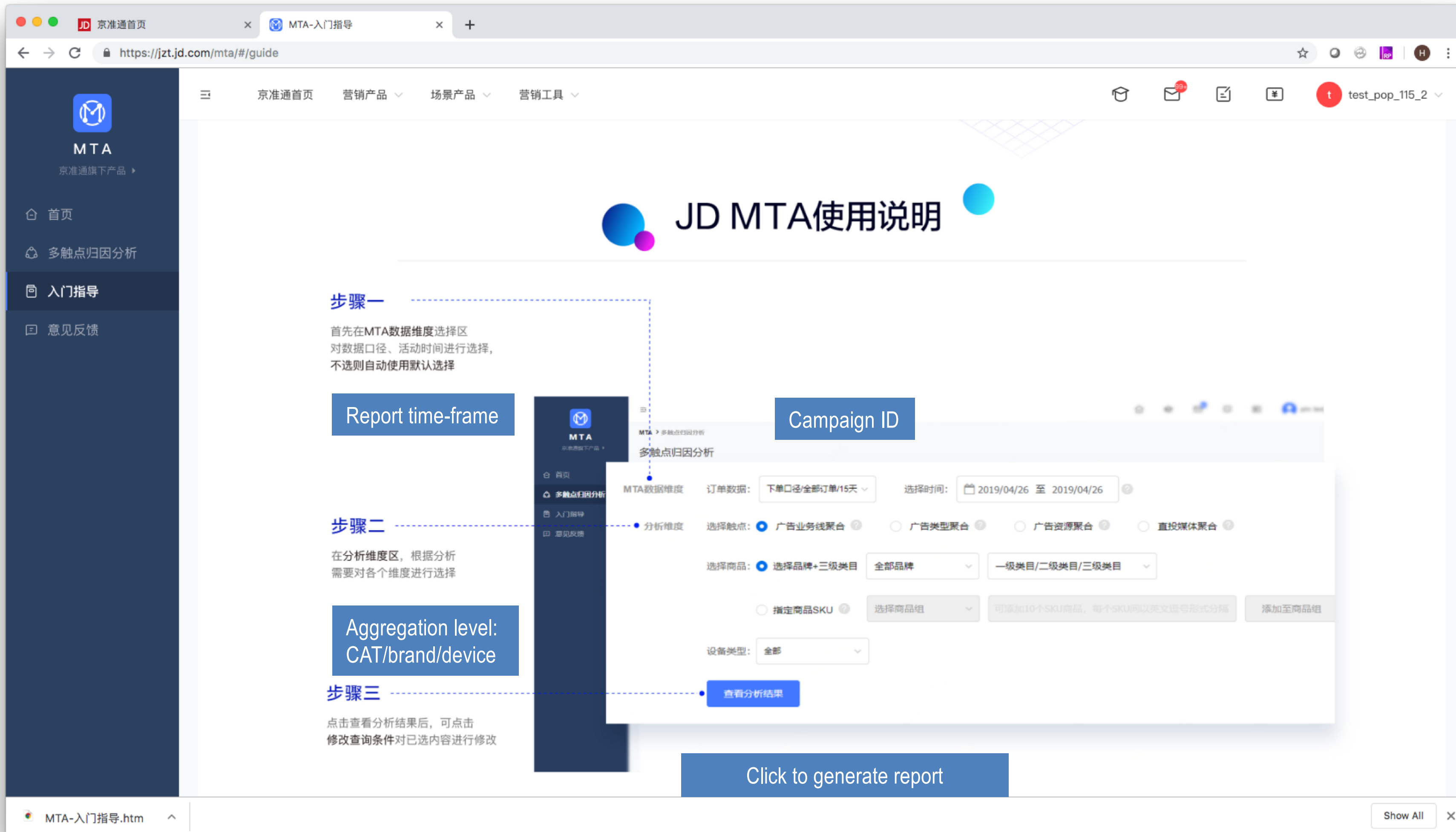
Notes: The Figure shows the Shapley Value from the RNN model at each ad-position indexed on he x-axis, average across all orders on day-15 for which that position was the last clicked. This allows benchmarking the Shapley Value based attribution against "last-click" attribution, which allocates 100% of the credit for the order to the last-clicked ad-position. The Shapley values are all seen to be <1, showing that under the model, the last-clicked ad-positions do not obtain full credit. To the extent that the Shapley values are all less than 0.6, the RNN model suggests that last-clicked ads contribute upto a maximum of 60% to the incremental conversion generated by advertising. Finally, cart and payment page positions, which may get a lot of credit under "last-click" or "last-visit" attribution schemes on eCommerce sites, are seen to not be allocated a lot of credit by the model.

- Interpret as: the smaller the Shapley Value at an ad-position, the more it deviates from "last-clicked" attribution
- Max Shapley is 0.6 => last-clicked never contributes more than 60% incrementally
- Deviations least for search
- Deviation high for cart and payment

Pooled Model

Number of categories in the model	180
Number of brands in the model	140 thousand
Number of impressions in 15 days	80 billion
Size of training dataset	187 million
Number of advertisers we can serve per day	75 thousand
Total model training time	10 hours
Models' average recall	85%~95%
Models' average precision	60%~70%
Time of Shapley value computation	3 hours
GPU cards	130

https://jzt.jd.com


京准通首页
MTA-入门指导
https://jzt.jd.com/mta/#/guide
MTA
京准通旗下产品
首页
多触点归因分析
入门指导
意见反馈
京准通首页
营销产品
场景产品
营销工具
test_pop_115_2
JD MTA使用说明
步骤一
首先在MTA数据维度选择区
对数据口径、活动时间进行选择，
不选则自动使用默认选择
Report time-frame
Campaign ID
多触点归因分析
MTA数据维度
订单数据：
下单口径/全部订单/15天
选择时间：
2019/04/26 至 2019/04/26
步骤二
在分析维度区，根据分析
需要对各个维度进行选择
分析维度
选择触点：
广告业务线聚合
广告类型聚合
广告资源聚合
直投媒体聚合
选择商品：
选择品牌+三级类目
全部品牌
一级类目/二级类目/三级类目
指定商品SKU
选择商品组
添加至商品组
设备类型：
全部
Aggregation level:
CAT/brand/device
步骤三
点击查看分析结果后，可点击
修改查询条件对已选内容进行修改
查看分析结果
Click to generate report
MTA-入门指导.htm
Show All

京准通首页
MTA-入门指导
https://jzt.jd.com/mta/#/guide
MTA
京准通旗下产品
首页
多触点归因分析
入门指导
意见反馈
京准通首页
营销产品
场景产品
营销工具
test_pop_115_2
JD MTA分析结果展示说明
整体效果分析
Overall
By cross-tab
CAT benchmark
三大分析模块
整体效果
交叉分析
品类对比
整体投放数据
筛选数据聚合方式
Aggregation level
GMV
ROI
CPM
CPC
可筛选数据指标
可视化图表
Day of campaign
MTA-入门指导.htm
Show All

Thanks!