

# 2023 DataFunSummit

## 投放场景下的问题分析和用户价值预估

演讲人：徐国强—腾讯—数据科学

# Contents

## 目录

1

投放场景下  
的问题分析

2

投放场景下  
用户价值建  
模实践

3

未来工作展望

4

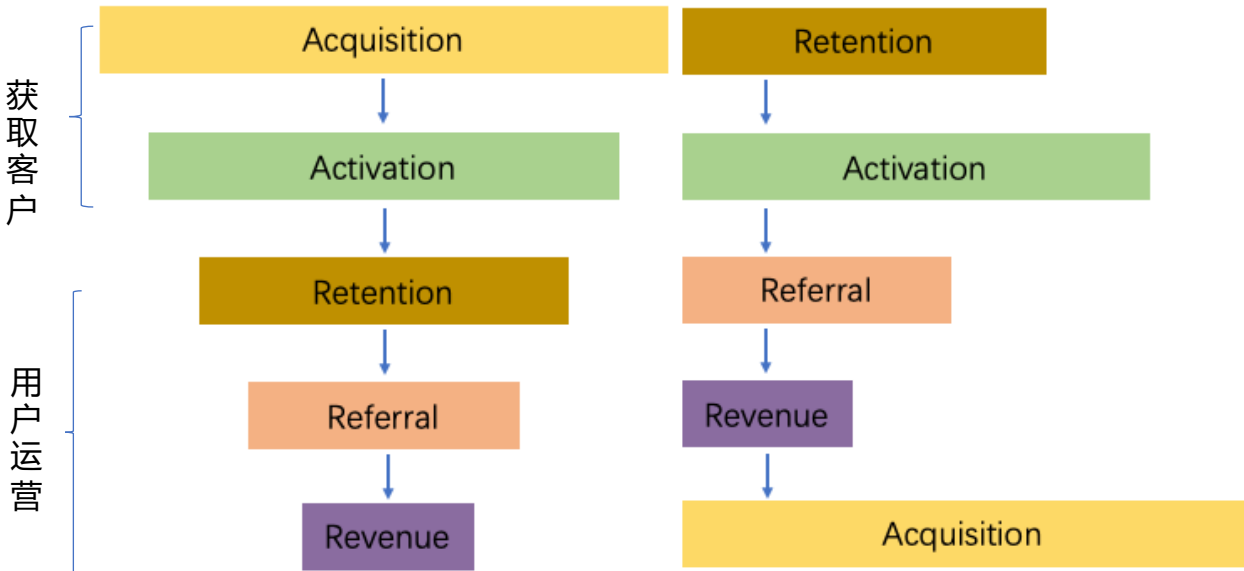
致谢

# 01 投放场景下的问题分析

---

# 用户价值在用户增长的作用

## 用户增长模型



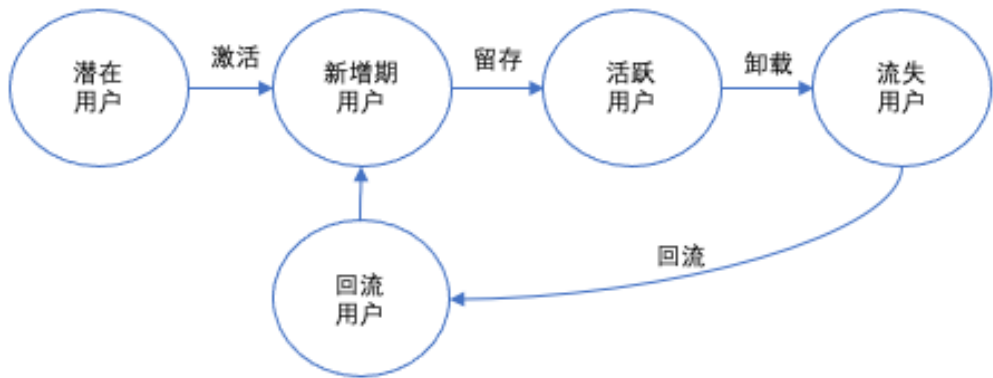
用户增长中的AARRR模型

用户增长中的RARRA模型

降本增效背景下，每个STAGE都需要优化ROI

用户增长技术两阶段：从外部获取新用户、持续激活新增和存量用户

## 用户价值模型是不同用户状态下的用户模型



用户生命周期状态图

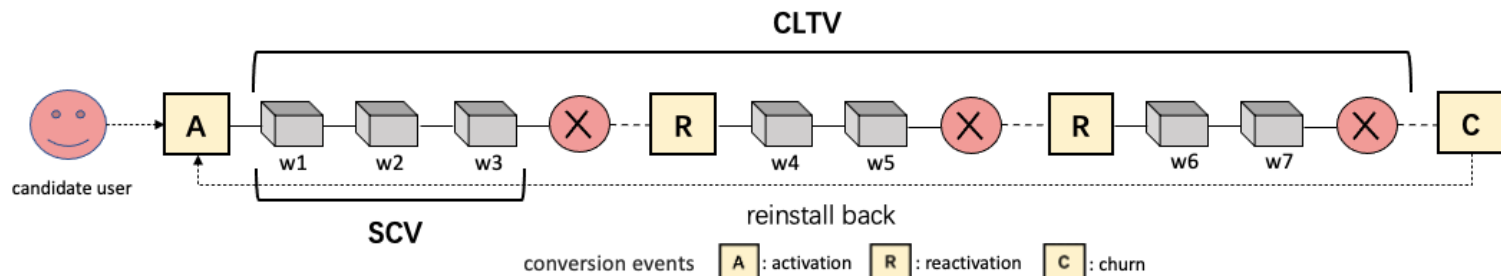
CLTV建模是一项底层的通用能力，其目的是准确预估处于不同用户状态的用户价值。增长模型的各个环节都可以在看清用户价值的基础上，制定出合理的增长策略，从而实现增长目标的达成。

- 投放获客场景：通过CLTV建模预估不同渠道cohort粒度的CLTV，得到每个渠道投放的ROI，调整出价策略和预算分配策略。
- 投放获客场景：通过CLTV建模得到用户粒度的CLTV，在投放过程中对媒体流量进行筛选，提升买量效率和效果。
- 生命周期运营场景：预估运营策略干预下的用户价值弹性，找到对运营策略敏感人群，叠加资源分配策略提升整体干预效果。

# 用户增长中的用户价值定义

## CLTV的定义

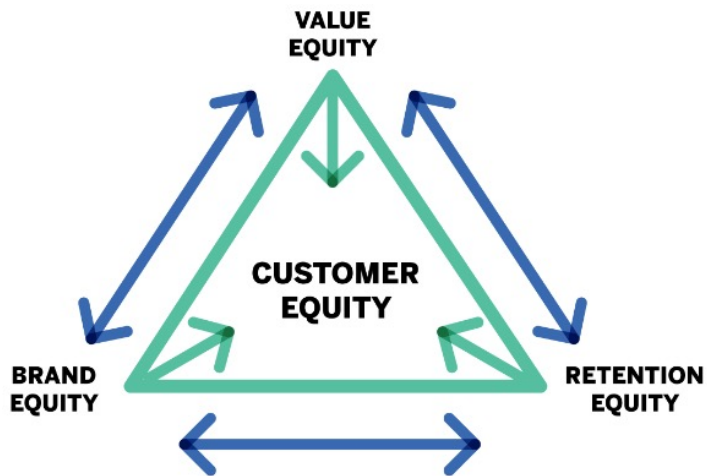
CLTV(Costumer Lifetime Value) 最早为市场营销领域的重要概念，表示的是用户在生命周期内为产品带来的收入总和。在用户增长实践过程中，用户时长、用户活跃等也可以被认为是一类用户价值。我们通常采用的建模目标是SCV (nLTV)，即用户在生命周期的某个session内的价值。



CLTV和SCV的关系

## 客户资产的定义

CE(Costumer Equity) 客户资产被定义为一组用户的CLTV总和。在一些场景下，客户资产最大化是产品的重要优化目标。例如：生命周期运营中，在有限的资源下，最大化运营干预下的用户活跃是最大化客户资产的任务之一。





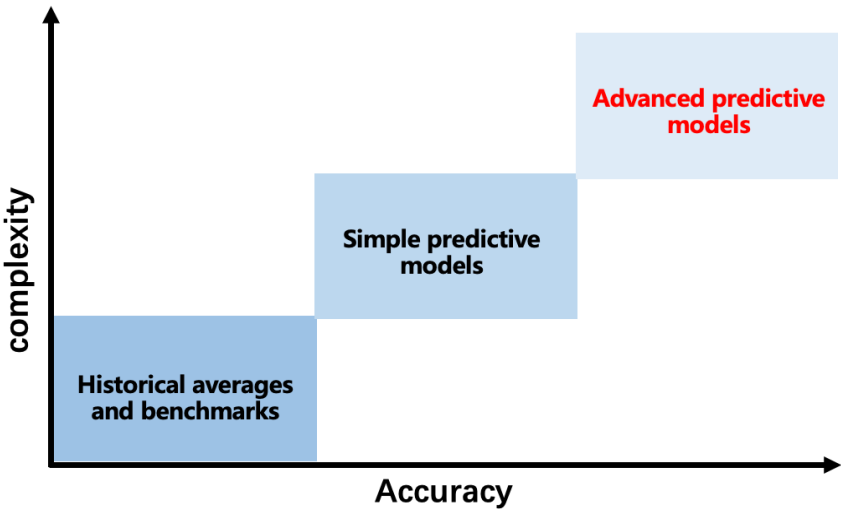
# 用户价值建模行业相关工作

## CLTV建模行业相关工作

- 随着大数据和机器学习的发展，行业相关工作越来越多，主要围绕数据稀疏、数据不平衡、多分布等问题展开。



scholar上能检索到的CLV相关工作文章趋势



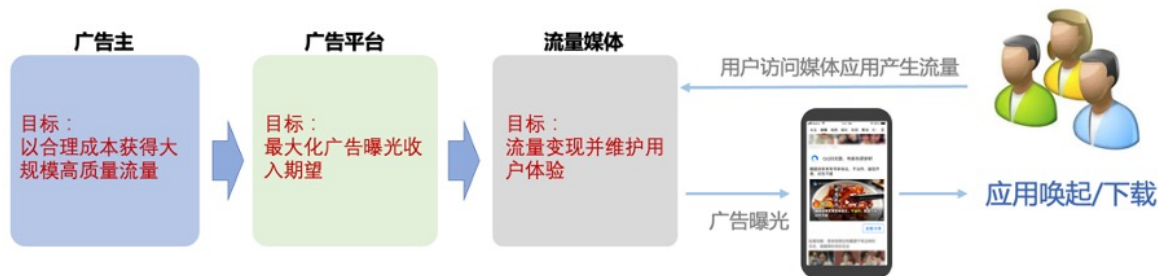
机构	具体工作	解决问题	时间
腾讯 (游戏、浏览器)	MarfNet	LTV建模中特征缺失的补充方法研究； Bidden算法通过动态难例挖掘，解决样本不均衡；	2023
	CDFA	多游戏的用户行为跨域应用，丰富用户刻画能力	2023
	DWT+GAT	通过模型结构设计分别建模时序成分和结构成分，提高LTV预测稳定性	2021
	BST+attention	基于模型结构优化利用用户在多个游戏的统计特征丰富用户刻画	2020
快手	MDME	提出用户LTV存在多种分布的现象，并通过分治思路优化建模； 对GINI评估方法进行分析并针对性改进；	2022
网易	perCLTV	提出在游戏发行领域，用户细粒度行为数据来丰富用户刻画提升预测准确性的重要性	2022
Google	ZILN	发现LTV中的零值膨胀问题，并在cross entropy中引入logNormal损失提高高价值用户建模准确度； 提出通过GINI系数对LTV进行评估；	2019
NetFlix	Markov chains	在拉活场景中，LTV的uplift建模问题及方案	2022
Fraunhofer IAIS	SMOTE	通过“造样本”样本增强方式解决样本不均衡问题	2018
Yokozuna data	DeepLearning in LTV	较早的提出利用深度学习在游戏行业建模LTV的工作	2018
ASOS	User Embedding	通过表征学习思路来替代用户行为序列特征工程，丰富用户刻画	2017
GroupOn	Engagement Feature	对用户全域触达历史（如email点击等）的利用丰富用户刻画，提高电商LTV预测准确度	2016

行业公司LTV相关的机器学习研究中的代表工作

# 投放场景下的问题分析和用户价值的应用

## □ 付费投放获客是多方博弈下对公域流量的利用

- 付费投放是用户增长领域最重要的用户触达工具。
- 多个参与方目标不完全一致，是博弈关系。



## □ 媒体广告平台侧现状

- 广告平台越来越开放，希望引入更多一方数据，优化广告效果

## □ 付费获客广告主侧现状

- 越来越希望通过数据的深加工，更加自主的提升投放的ROI
- 广告主需要一个**统一的对接方案**，解决渠道严重的割裂现象

### 渠道分布越来越割裂

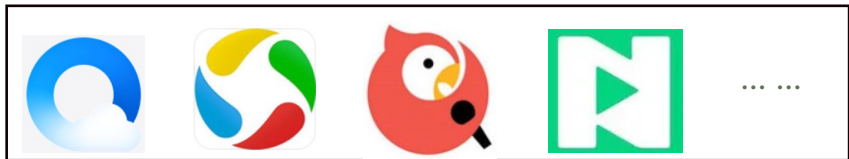


流量供给



流量优选

获客广告主更加关注ROI



广告平台开放RTA



# 投放场景下的问题分析和用户价值的应用

□ 提升获客的LTV是提升ROI的关键抓手

$$ROI = \frac{LTV}{CAC}$$

1.提升LTV  
2.降低CAC

- CAC优化中，降低出价可带来分母快速下降，但严重影响拿量能力。
- LTV优化中，对媒体流量的优选是潜客拉新中投放端最有效抓手。

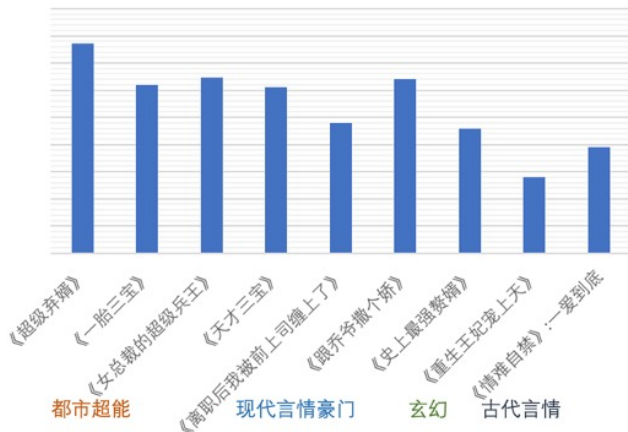


# 投放场景下的问题分析和用户价值的应用

## 提升LTV的解决方案

随广告平台、数据能力的支持和自身能力不断探索，提升LTV的手段不断进化。

**内容选品:** 分析建模“品类”带来的人群LTV差异，优化投放品类。



**回传LTV建模:** 广告平台深度出价能力+浅层指标回传，实现“平台助力质量建模与优选”。

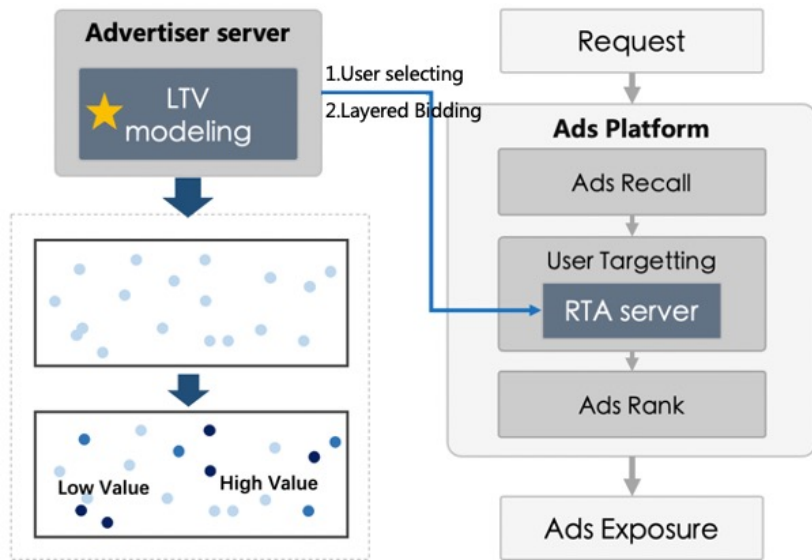
$$eCPM = \text{SmartBid} * pCTR * pCVR * \frac{pDCVR}{DCVR}$$

- 基础排序公式
- pDCVR为平台根据广告主回传次留、浅层转化Label进行建模预估，实现质量过滤

问题：粒度太粗，数据稳定性差。

问题：行业统一建模，难以深入解决业务特有问題。

**RTA(Realtime API):** 自主建模个性化LTV表征用户价值，并通过RTA实时流量优选/分层出价。实现方案统一，适合平台化。



优势：个性化粒度，自主灵活建模，具备ABTest。

## 02 投放场景下用户价值建模实践

---

# 大禹投放平台多业务用户价值建模实践

## 大禹投放平台

大禹投放平台是腾讯PCG内部的一站式广告投放平台，向BG范围内产品的增长业务线提供**素材创意、广告投放、RTA策略、效果分析**等多维度的能力，让接入产品更低成本、高效率的落地广告投放业务。

大禹投放平台已经服务于手机QQ浏览器、腾讯应用宝、全民K歌、腾讯动漫等十余个业务和产品，并且为这些业务的付费获客投放的ROI带来的巨大的提升。



## 多业务下的用户价值建模面临的挑战

付费场景	场景举例	收入构成
Subscription Services	腾讯视频、QQ音乐、腾讯动漫、腾讯课堂	会员充值
Game	应用宝	游戏充值
Live Streaming	全民K歌、Now直播	直播打赏充值
Content Consumption	手机QQ浏览器、应用宝	广告收入 (CPD/CPM)

接入产品多样化

投放场景	场景举例
展示广告	全民K歌、应用宝、腾讯课堂、Now直播、腾讯动漫
SEM广告	应用宝、手机QQ浏览器
应用商店	腾讯动漫、手机QQ浏览器
内部导流	-

接入媒体渠道多样化



### 建模基本问题

- 投放场景下的用户价值建模目标选择问题
- 如何合理评估投放场景下的用户价值建模效果

### 业务场景中的数据问题

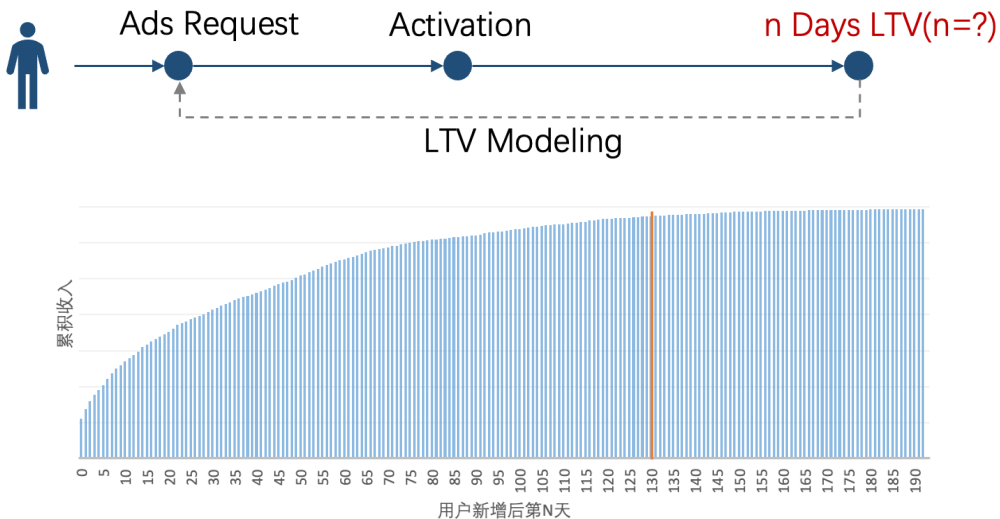
- 投放样本数据非常稀疏，建模用户价值时如何解决
- 如何解决样本多成分多分布的问题
- 渠道之间数据如何复用

多业务场景下LTV建模面临的关键问题

## 投放场景下CLTV建模目标选择问题

### 投放场景CLTV建模如何量化选择建模目标

cohort粒度CLTV在激活后随时间增加而不断增长，选择多长窗口期作为建模目标是个权衡问题。



- 窗口期设置过长 ➡
1. 样本等待期长，与大部分场景在初期即有需求形成矛盾。
  2. 决定了线上ABTest时观察周期偏长，严重影响迭代。

RTA流量优选场景必须要选择短期的、敏捷的指标进行建模和观察。

### 基于Pearson相关性选择建模目标

	LTV1	LTV3	LTV7	LTV14	LTV30	LTV60	LTV90	LTV120
LTV1	1	0.86	0.8	0.77	0.63	0.55	0.47	0.45
LTV3	0.86	1	0.95	0.87	0.73	0.63	0.53	0.5
LTV7	0.8	0.95	1	0.93	0.86	0.68	0.57	0.55
LTV14	0.77	0.87	0.93	1	0.86	0.81	0.72	0.76
LTV30	0.63	0.73	0.86	0.86	1	0.88	0.84	0.83
LTV60	0.55	0.63	0.68	0.81	0.88	1	0.96	0.95
LTV90	0.47	0.53	0.57	0.72	0.84	0.96	1	0.99
LTV120	0.45	0.5	0.55	0.76	0.83	0.95	0.99	1

LTVn之间的Pearson相关性分析

- 较高的线性相关性对替代建模会更优把握。
- $Pearson(Ltv14, LTV120)$ 在0.7~0.8左右，基本属于强相关性。
- 14天样本等待期和观察期对多数场景基本可接受。
- 综合看，LTV14是比较可行的建模指标。

# 投放场景下LTV建模效果评估方法

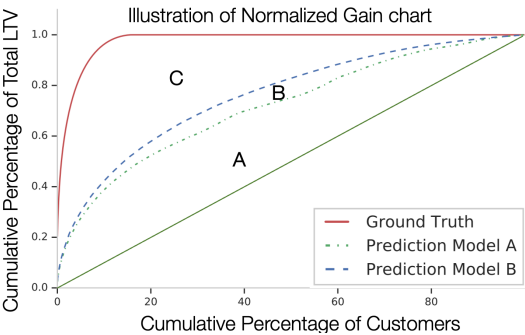
## 常见模型评估方案

行业大部分工作及我们针对渠道粒度LTV建模评估均采用nMAE、nMAPE评估方法，但在投放应用场景中存在问题。

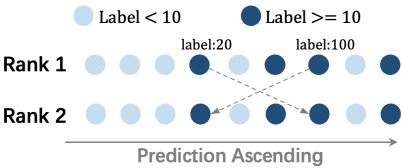
$$nMAE = \frac{1}{n} \sum_{i=0}^n \frac{|\hat{y}_i - y_i|}{\bar{y}} \quad nMAPE = \frac{1}{n} \sum_{i=0}^n \frac{|\hat{y}_i - y_i|}{y_i}$$

流量优选场景下，LTV建模更偏向于**Discrimination**问题，而不是**Calibration**问题。  
样本的**Ranking**比样本的**精确值**更重要。

- 方法1：将头部 x% 样本为正样本，其他或末尾 y% 样本为负样本，评估AUC
- 方法2：采用Normalized Gini Coefficient [Google ZILN 2019]



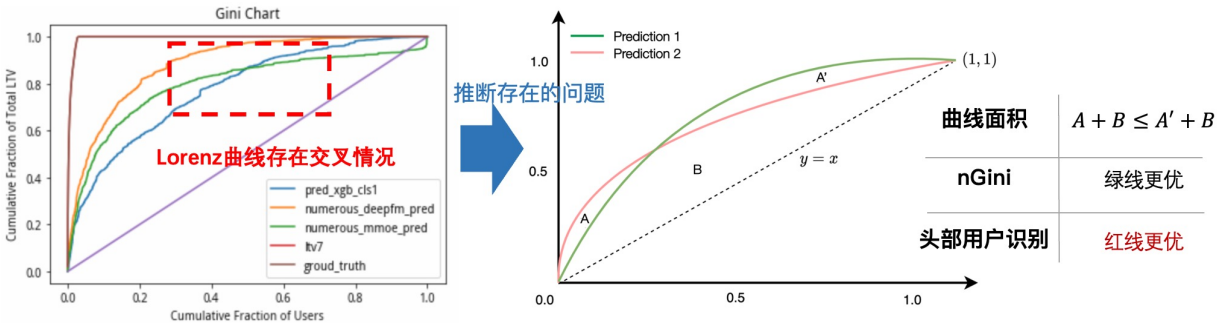
nGini系数在Gain chart基础上计算曲线下面积。模型B的Gini系数为: (A+B)/(A+B+C)



AUC	Norm Gini
Rank1 = Rank2	Rank1 > Rank2

优先采用nGini系数

## nGini模型评估方案拓展



当投放场景主要为头部高价值用户的筛选或出价时，需要补充关注头部用户的Precise-Recall等指标。

## ABTest实验上线前效果前测

ABTest上线前，根据业务LTV提升目标或规模缩减程度，通过关系趋势表推算应用阈值，估算实际提升效果。

Score阈值	0.049	0.615	0.689	0.755	814	0.879
尾部用户过滤比例	0%	10%	20%	30%	40%	50%
LTV14提升比例	0%	6.88%	14.79%	23.49%	33.34%	47.95%



# 样本收集成本高带来的数据稀疏问题和解决思路

## 用户价值建模中的样本稀疏问题

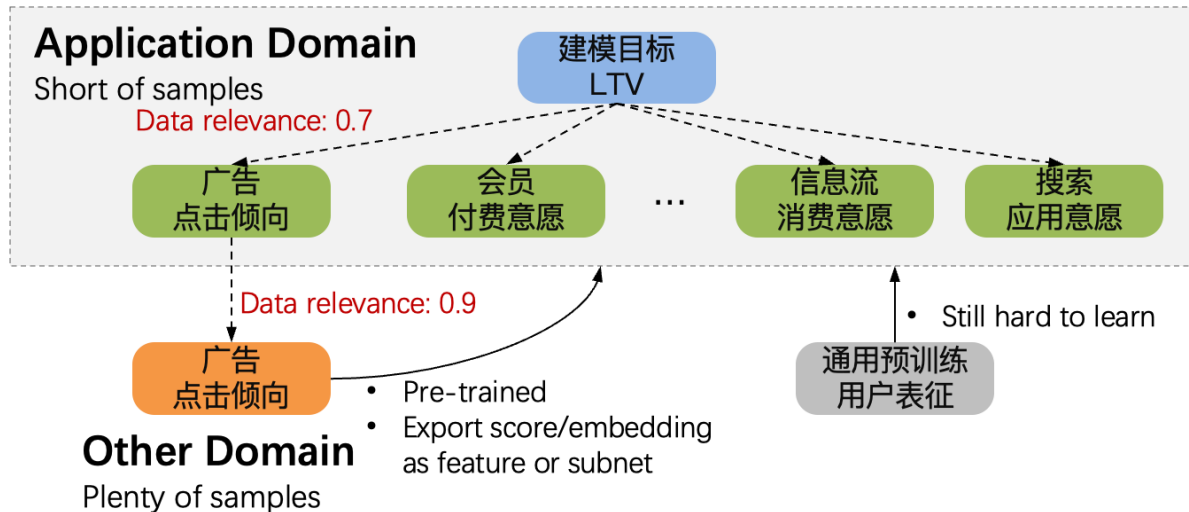
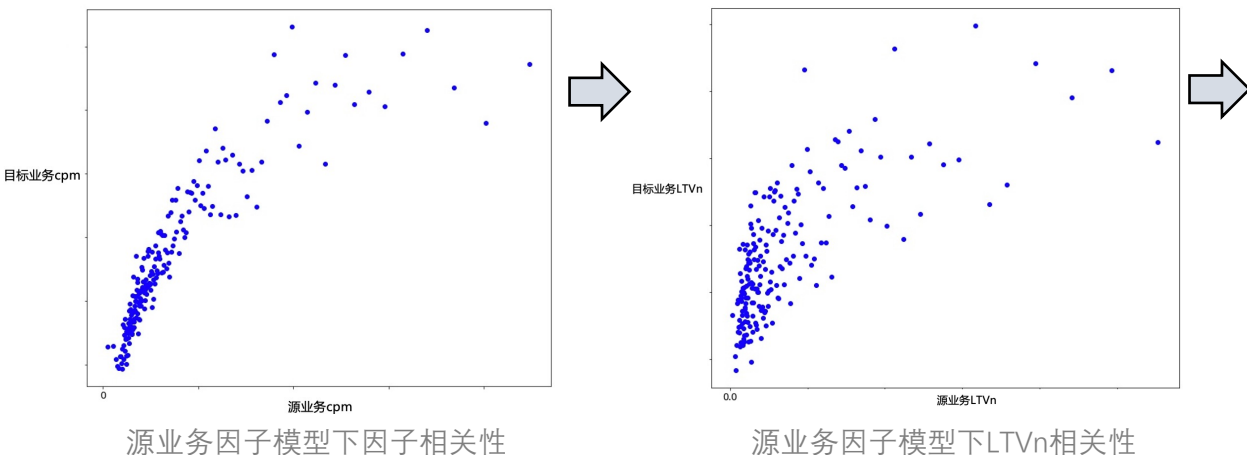
为提升个性化建模的准确性，在搜索和推荐场景中通常采用超高维用户特征和海量样本的方案。付费投放因为预算条件，获取海量样本几乎不可能。

渠道	厂商	AMS	快手	头条
样本规模	100w	10w	20w	25w

投放中采用高维用户特征建模个性化，但样本规模远小于特征空间

## 引入预训练手段解决样本稀疏问题

用户价值由多种因素共同作用，某些因子属于用户特质，引入外部数据**强化此类信息**的**用户表达**，可以在业务间迁移。

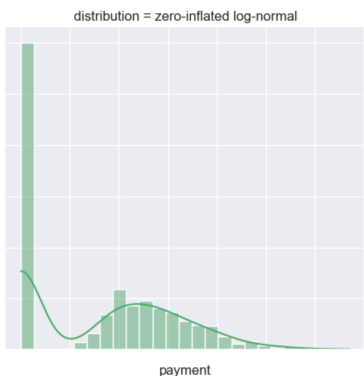


预训练因子模型解决方案

# 用户价值建模中的多成分多分布问题和解决思路

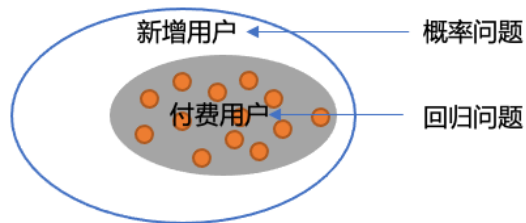
## 增值付费场景零值膨胀问题

- 增值服务类产品LTV分布非常极端，付费用户占比极低，回归预测效果较差。

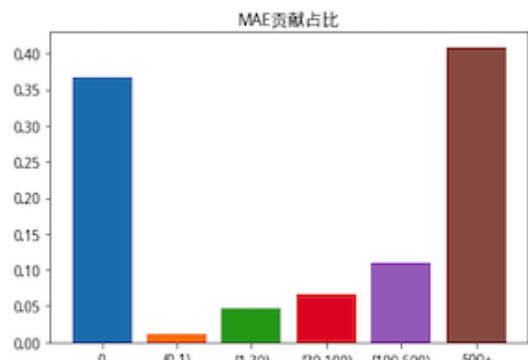


增值付费商业模式下Ltv分布

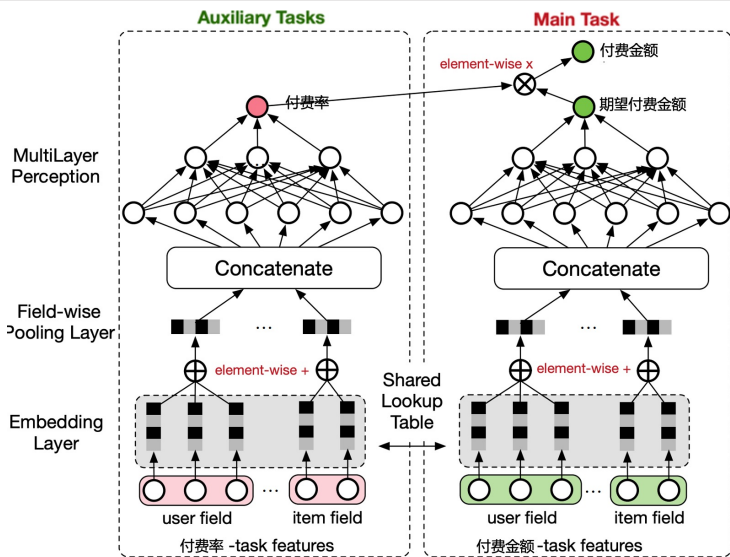
付费用户占比<1%，样本不平衡问题严重



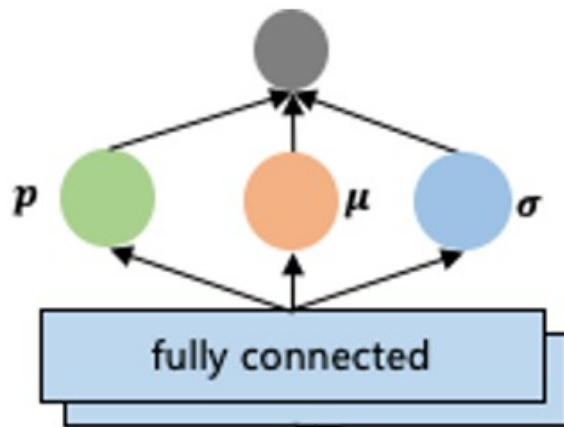
只建模金额



## 零值膨胀分布解决方案



付费率和付费金额联合建模[ESMM 阿里 2019]



ZILN：零值膨胀下LTV建模的方案[Google 2019]

$$pred = p * \mathbb{E}(Y) = p * \exp(\mu + \frac{\sigma^2}{2})$$

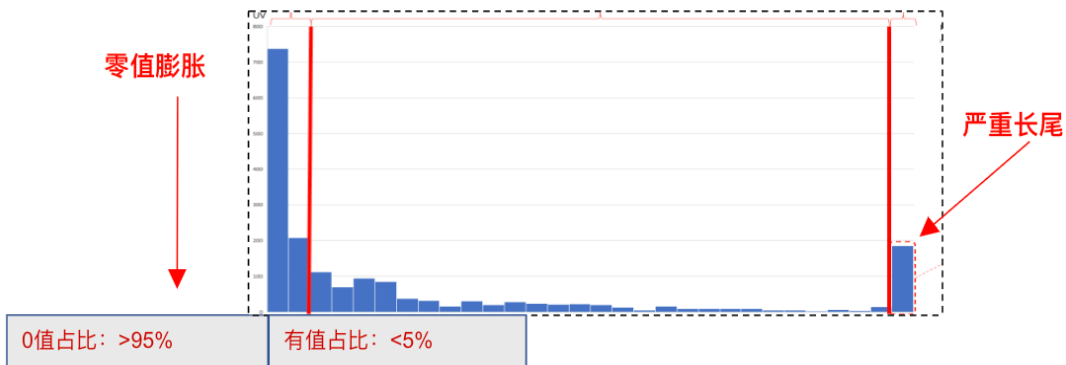
$$L_{ZILN}(x; p, \mu, \sigma) = L_{CrossEntropy}(\mathbb{1}_{\{x>0\}}; p) + \mathbb{1}_{\{x>0\}} L_{Lognormal}(x; \mu, \sigma).$$



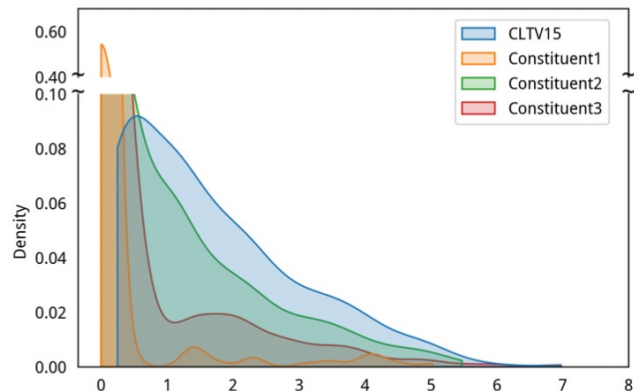
# 用户价值建模中的多成分多分布问题和解决思路

## CLTV多分布和多成分问题

- 很多场景不满足Lognormal分布，本质上是**多种分布的叠加**。
- 从微观角度来看，用户收入是由多种成分构成，形成**多成分多分布问题**。

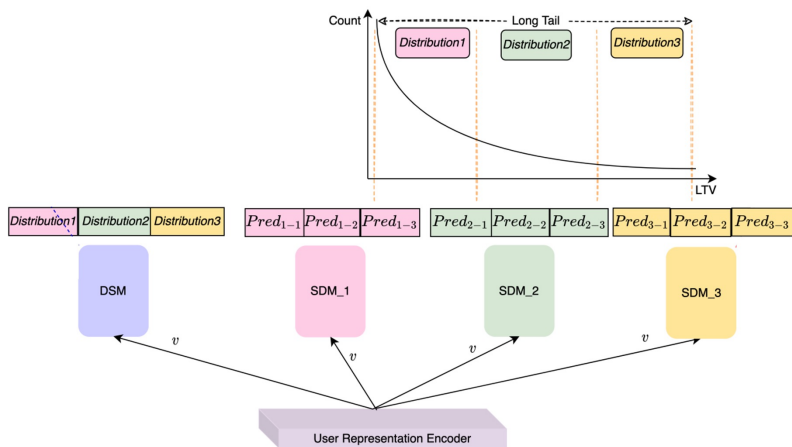


LTV多分布示例

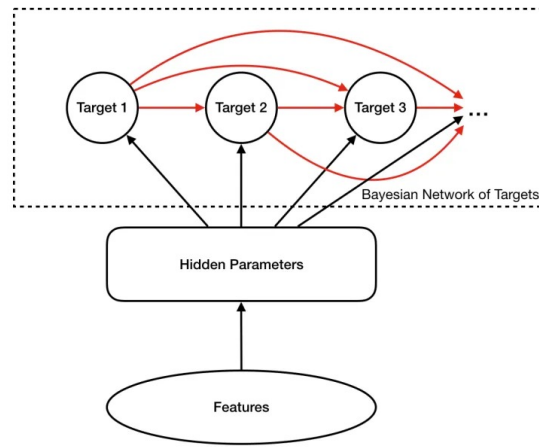


LTV多成分多分布示例

## 多成分多分布问题解决方案



MDME多分布LTV模型 [快手 2022]

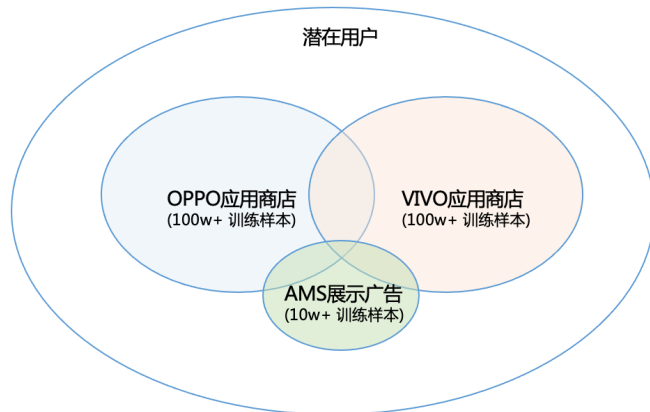
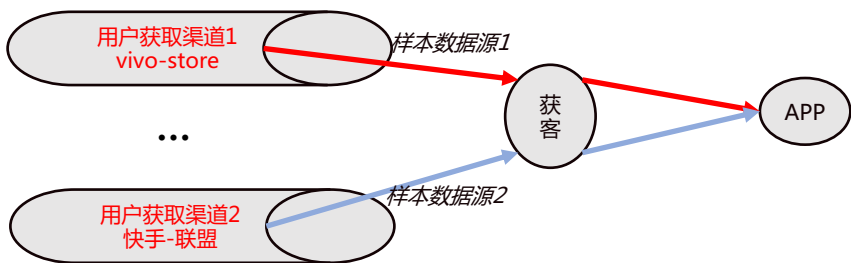


贝叶斯子成分依赖模型[阿里 DBMTL 2019]

# 用户价值建模中的多渠道问题和解决思路

## 投放渠道之间数据复用问题

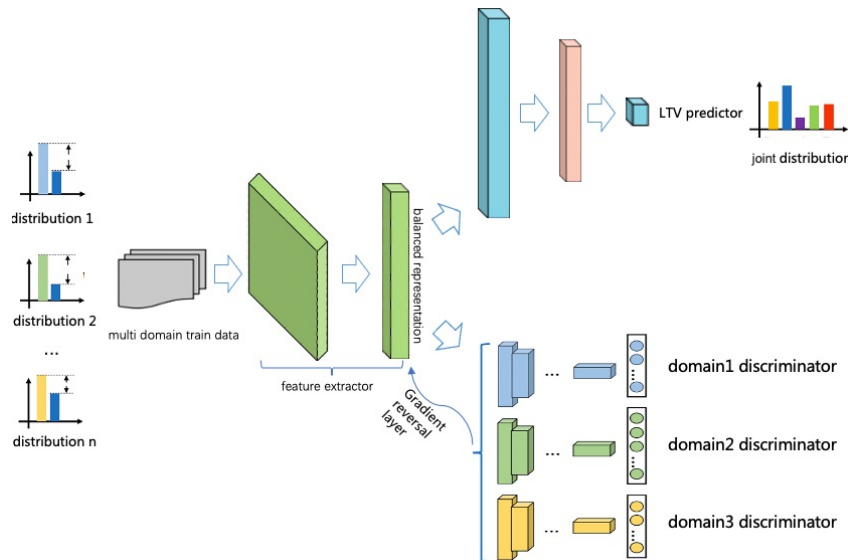
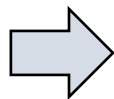
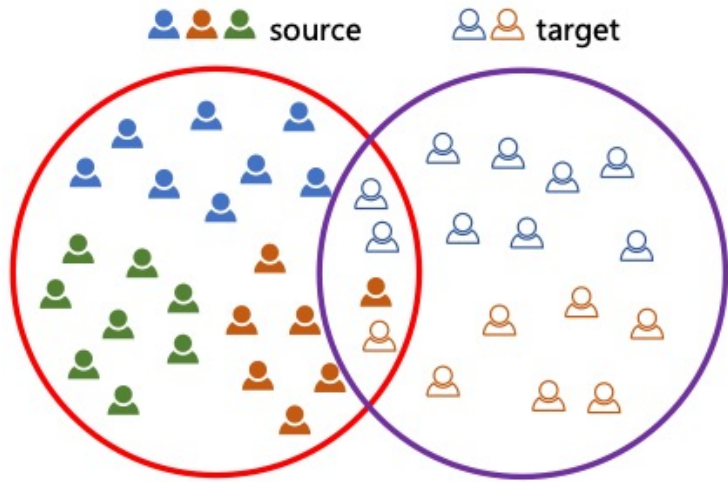
- 大禹投放平台目前接入的渠道10个以上



- 业务LTV建模时直接累加数据复用存在数据冲突

多渠道样本规模示例

## 基于领域迁移的多渠道建模解决方案



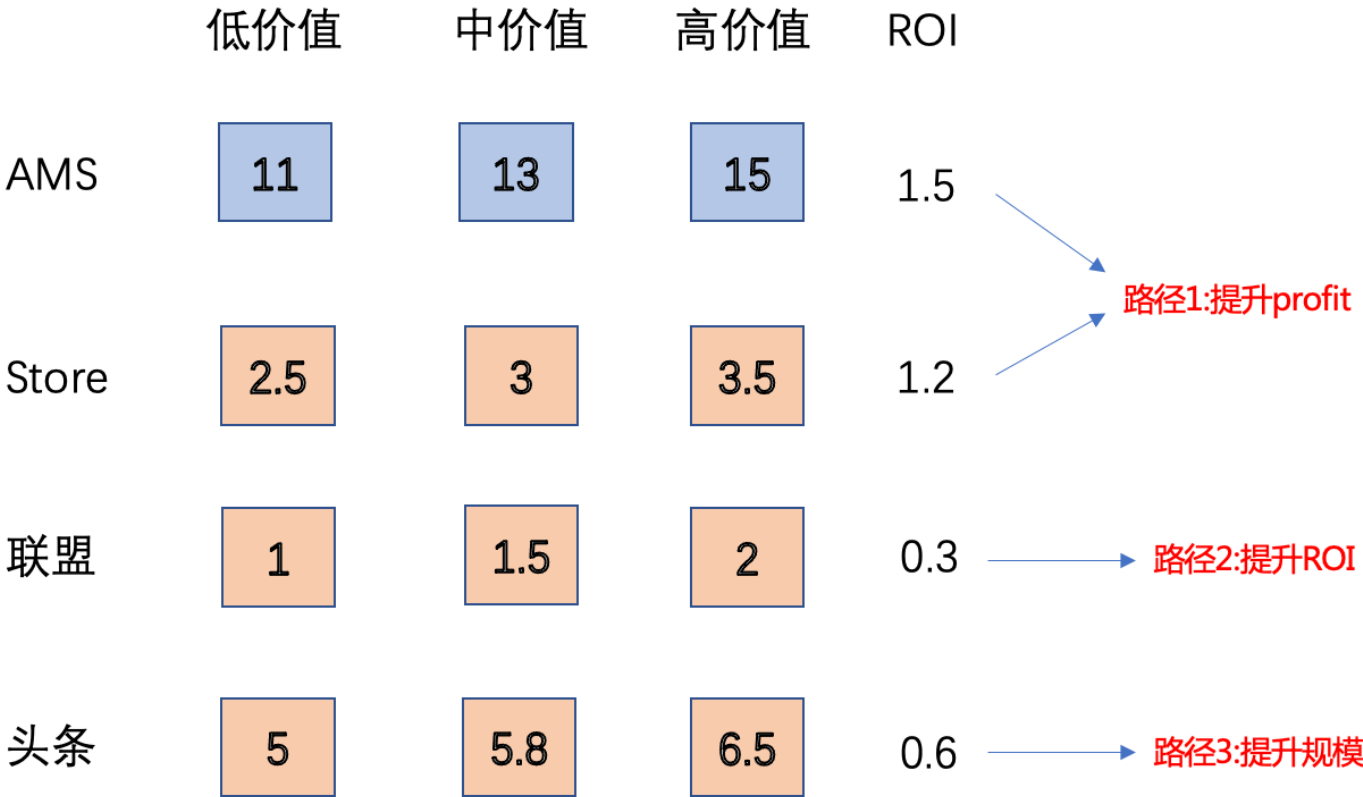
增加域适应机制（例如[DANN, 2016]），提升多个渠道之间的数据复用效率

# 03 未来工作展望

---



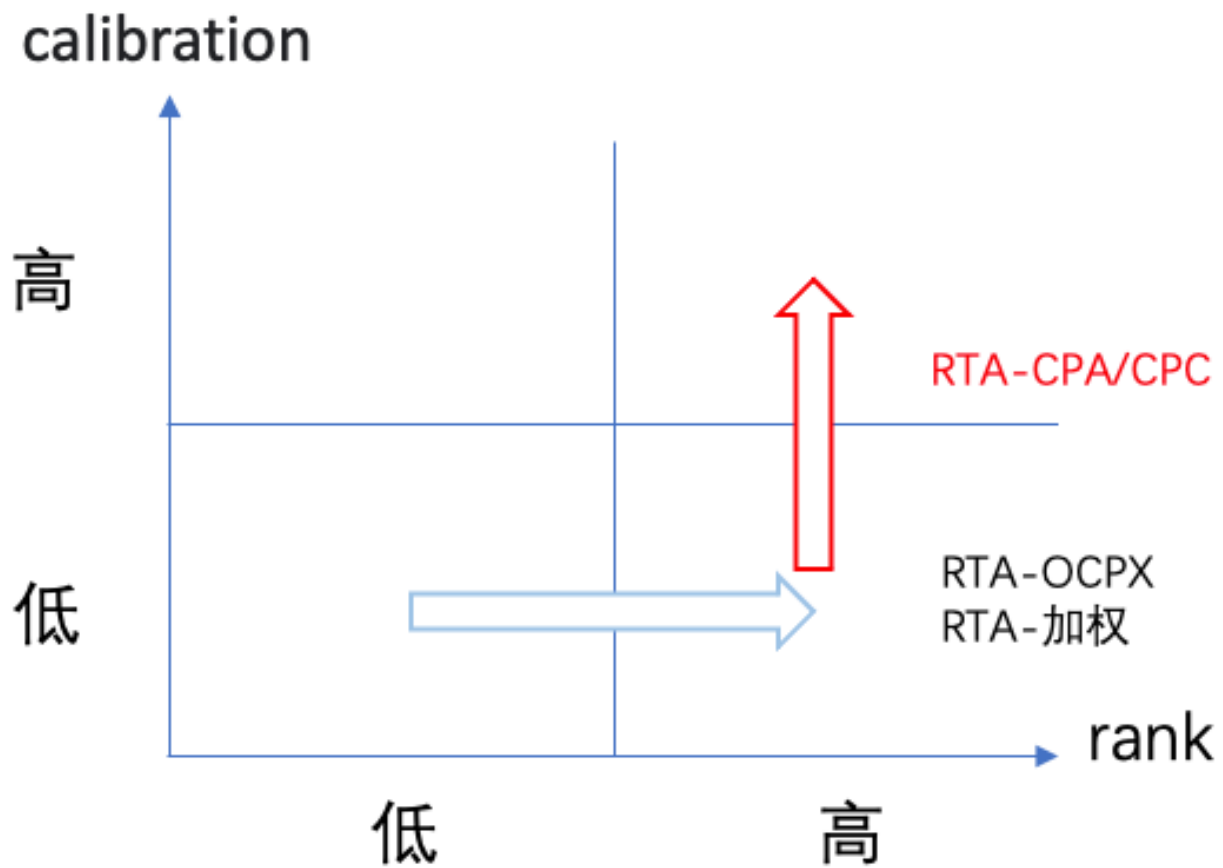
□ 基于客户资产模型优化用户价值分层，通过渠道间预算分配策略达成客户资产最大化。



总预算和ROI约束下，求解规模最大化的全渠道分层投放策略



- 持续迭代用户价值模型，从序准往值准优化，探索更精准的投放策略。



# 2023 DataFunSummit

— THANKS —

感谢您的观看