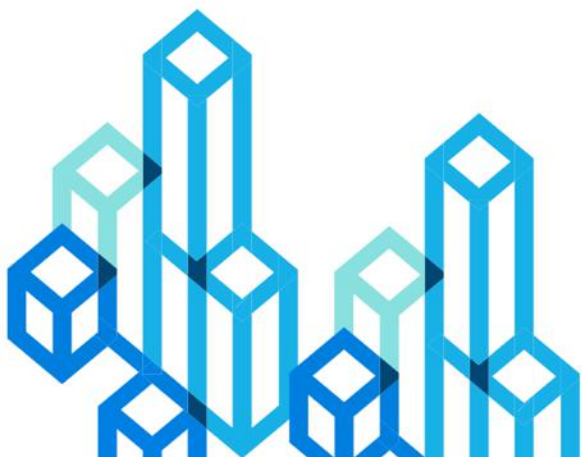


# OPPO广告中的 定向标签挖掘与落地实践

OPPO商业平台部数据挖掘团队





**欢迎扫码一起交流**

## 王平

OPPO商业平台部数据挖掘团队负责人

主要负责广告业务的用户画像、智能定向、召回排序以及搜索广告召回等模块。

硕士毕业于哈尔滨工业大学计算机学院SCIR实验室。从业以来涉猎过的行业包括通用搜索、金融问答、电商搜索和计算广告。2019年加入OPPO广告业务。

践行的技术理念是：不做技术极客，只为业务量身定做。

本分享重在邀请大家一起再走一遍我们  
兴趣标签工作的探索旅程

不秀技术指标，只还原思考过程  
一起出发吧

如果觉得有趣，会后欢迎交流甚至加入我们继续探索

11:00

## 业务背景

- 展示广告中的标签定向、兴趣标签的挖掘与使用

11:10

## 定向标签的挖掘算法

- 行为统计、行业建模、兴趣建模

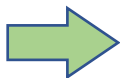
11:30

## 定向标签的落地应用

- 白盒、灰盒、黑盒、无盒

11:50

## 总结、思考和交流



## 业务背景

- 展示广告中的标签定向、兴趣标签的挖掘与使用

## 定向标签的挖掘算法

- 行为统计、行业建模、兴趣建模

## 定向标签的落地应用

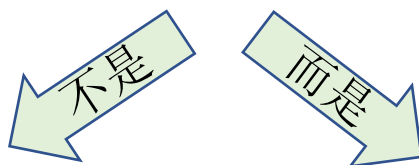
- 白盒、灰盒、黑盒、无盒

## 总结、思考和交流

# OPPO的广告业务：是赚钱卖广告而不是花钱打广告



这里的OPPO  
广告业务

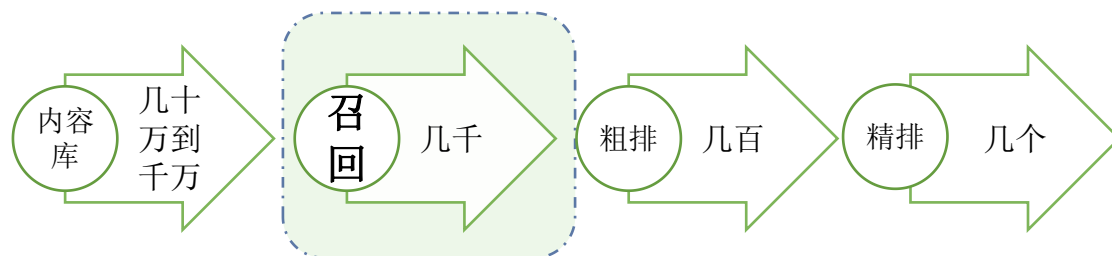


当然如右图  
OPPO Find X3  
广告，自卖自  
夸也是有的



# 展示广告的特殊性决定标签的作用和挖掘亦有特殊性

广告/推荐系统  
通用流程



召回环节  
几种类别  
场景比较

业务类型	用户意图明确性	内容方参与度
内容推荐	★★★★☆	★☆☆☆☆
推荐广告	★★★★☆☆	★★☆☆☆☆
搜索广告	★★★★★	★★★★☆☆
展示广告	☆☆☆☆☆	★★★★★


标签在召回环节连接着广告与用户，  
是平台与广告主之间非常重要的沟通互动语言

广告主对该语言存在巨大的认知偏差，  
最大原因是用户意图在标签挖掘与使用时时空分离



### 业务背景

- 展示广告中的标签定向、兴趣标签的挖掘与使用



### 定向标签的挖掘算法

- 行为统计、行业建模、兴趣建模

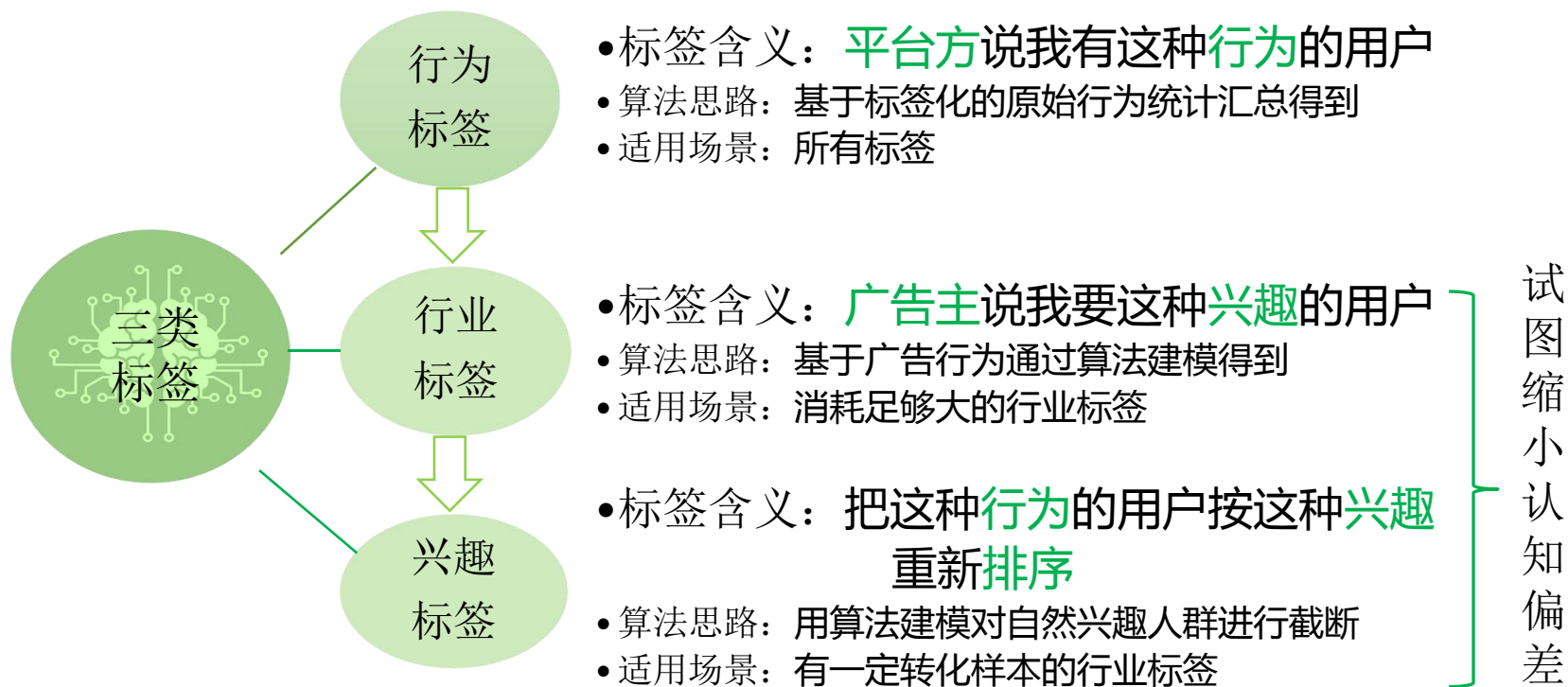
### 定向标签的落地应用

- 白盒、灰盒、黑盒、无盒

### 总结、思考和交流



兴趣标签任务的本质：  
平台方用标签描述一群平台用户，  
广告主用标签表达一群目标用户。



# 行为标签建模：拉长时空全面抽象用户行为

$$S(u, t) = \sum_{d=1}^D \alpha_t^d * \sum_{\substack{i \in I(u, d) \\ b \in B(u, d)}} \beta_b * w_b * w_i * w_{i,t}$$

用户  $u$       时间窗口  $D$       衰减因子  $\alpha_t^d$       行为权重  $\beta_b$       行为强度  $w_b$       Item 权重  $w_i$       标签权重  $w_{i,t}$

标签  $t$        $d$  天前       $i \in I(u, d)$        $b \in B(u, d)$

——> 该用户当天所有的item  
——> 该用户当天所有的行为

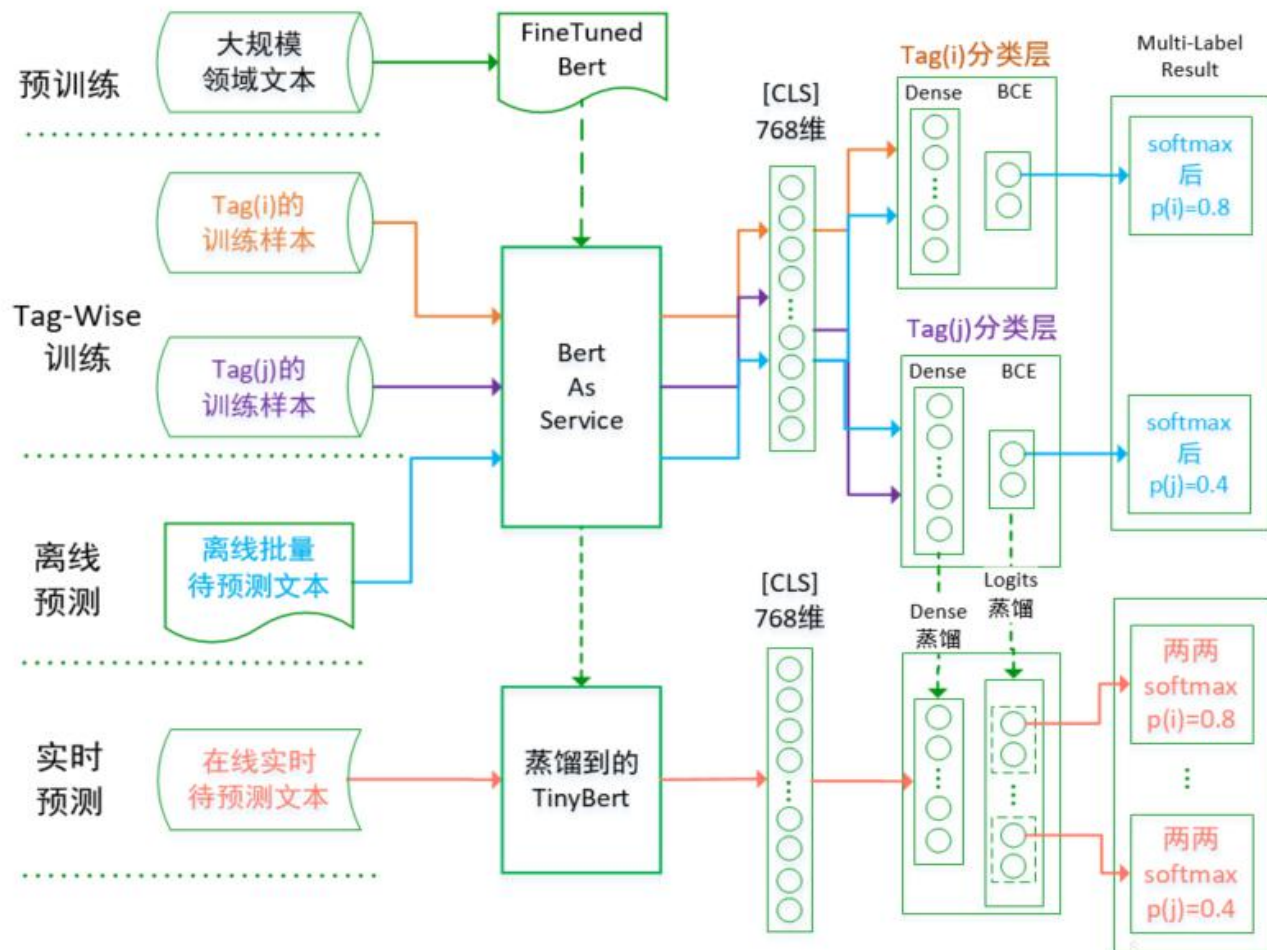
参数符号	含义	如何取值	说明
$I(u, d)$ $w_i$ $w_{i,t}$	<ul style="list-style-type: none"><li>用户 <math>u</math> 在 <math>d</math> 天前所有有交互行为的Item</li><li>每个item权重不同</li><li>Item打上某标签的概率也不同</li></ul>	<ul style="list-style-type: none"><li>日志记录</li><li>Item权重按uv算idf</li><li>Item的标签权重：若人工打标统一为1，否则为模型预测值</li></ul>	<ul style="list-style-type: none"><li>Item有App、Query、资讯、网页、广告等；日均处理千万级Item</li><li>如陌陌因为用户量更大，权重小于Soul</li></ul>
$B(u, d)$ $\beta_b$ $w_b$	<ul style="list-style-type: none"><li>用户 <math>u</math> 在 <math>d</math> 天前所有的行为</li><li>每种行为权重不同</li><li>具体到每个用户每次行为的强度也不同</li></ul>	<ul style="list-style-type: none"><li>日志记录，每个行为对应一个item</li><li>行为权重通过LR进行拟合</li><li>行为强度按人工规则</li></ul>	<ul style="list-style-type: none"><li>App有安启卸/注册/付费等；Query为搜索、广告有曝光点击转化；日均处理70亿行为</li><li>如App活跃按时长取log得到强度</li></ul>
$\alpha_t$	标签 $t$ 的衰减因子	按该标签平均行为频次14天衰减到0.1以下反推	如浏览行为比付费行为要衰减快很多

除了行为权重  $\beta_b$ ，最重要的是Item如何打上标签，即  $w_{i,t}$

# Item理解：文本是重要信息载体，文本分类是核心技术

重点解决两个问题：

- ✓ 可增量新增类目，而不影响现有类目的效果
- ✓ 弹性算力投入，缓解技术和资源瓶颈



网络结构：

一个公用Encoder，  
每个Tag一个MLP分  
类层

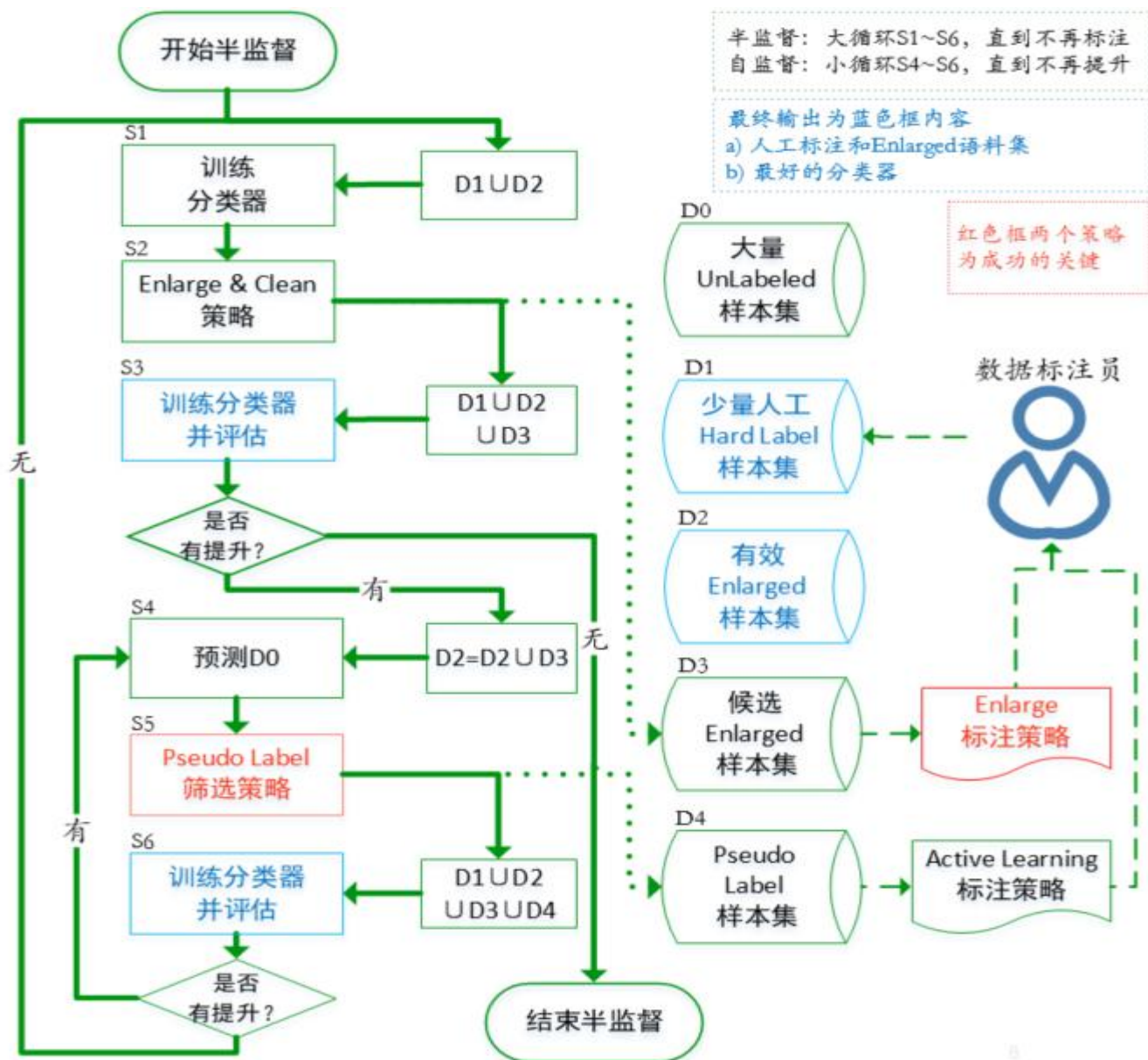
离线预测：

时延要求低，使用  
RoBERTa，结果  
Cache到线上

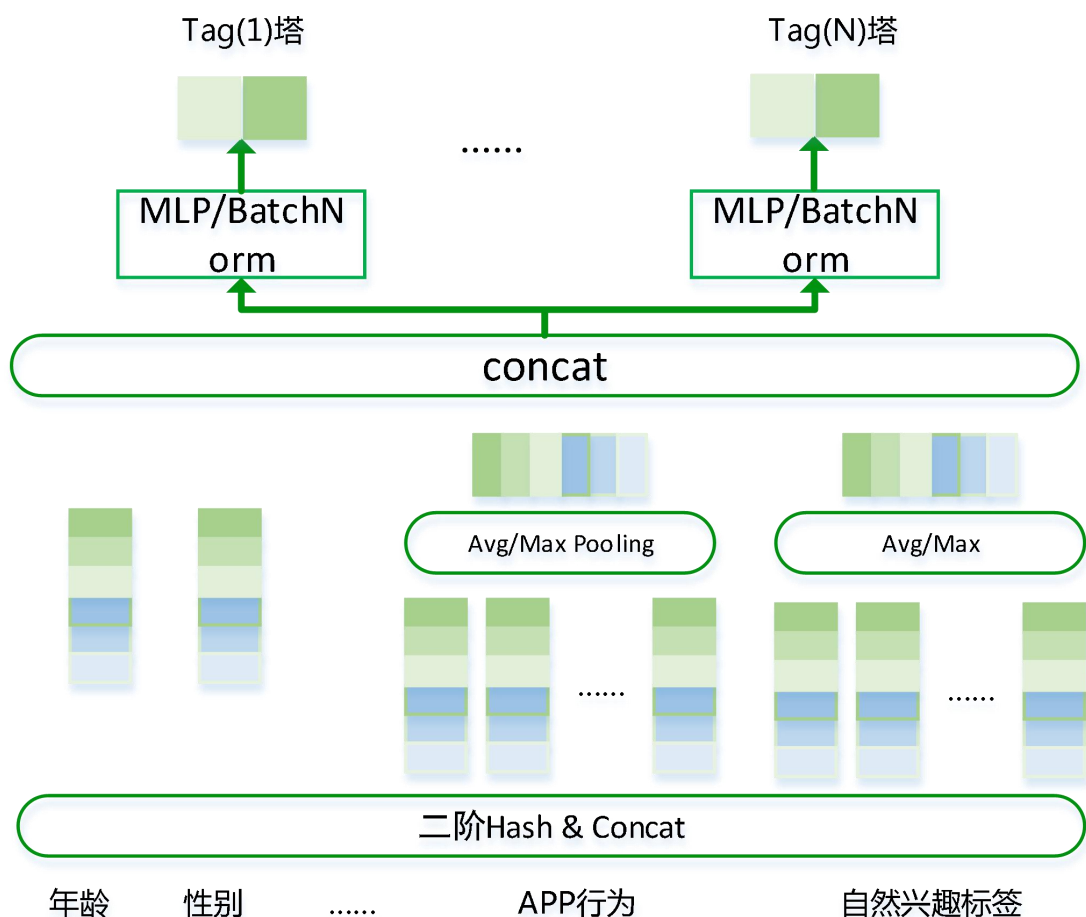
在线预测：

时延要求高，使用  
TinyBert，体现弹性  
算力思想

# 半监督与自监督：标注人员都可以训练分类器



形式和流程上，仍然是给用户打上兴趣标签  
本质和思想上，应该是为一类广告圈选人群



## 要点：

- 样本的艺术  
召回环节大盘随机一定对吗？消耗大涨转化率大降！
- 多目标  
每个标签一个塔，底层特征共用；对小样本标签特别管用
- 流量特征  
标签分别用在信息流和联盟场景是预先知道的，将其作为特征加入效果特别显著
- 多阶Hash



并且我们需要维持行为标签的本来物理含义

解决方案

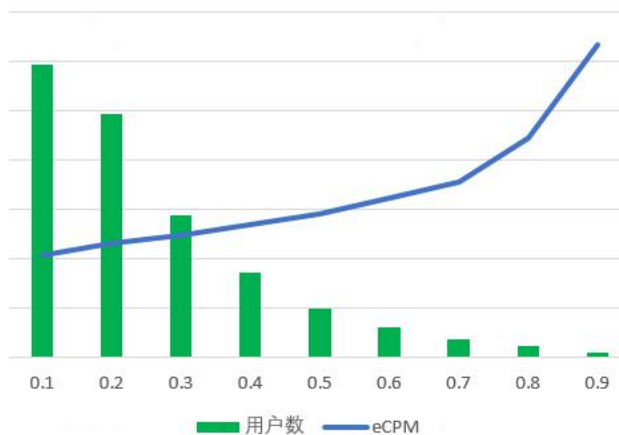
将行为标签按行业标签思想，以转化率为目标进行算法建模  
确实太少（少于1000正样本）则仍维持统计

算法	特征	训练集	预测集	超参
<ul style="list-style-type: none"><li>• XGB</li><li>• 每个标签一个模型</li></ul>	<ul style="list-style-type: none"><li>• 融合时间但分行为的统计标签</li></ul>	<ul style="list-style-type: none"><li>• 正样本：出价层转化，不足补上一层</li><li>• Hard负：曝光未转化</li><li>• Easy负：大盘随机</li></ul>	<ul style="list-style-type: none"><li>• 所有当天有本标签行为的用户</li></ul>	<ul style="list-style-type: none"><li>• 树深度要依据样本多少调整</li></ul>



# 兴趣分如何使用：效果与数量的权衡

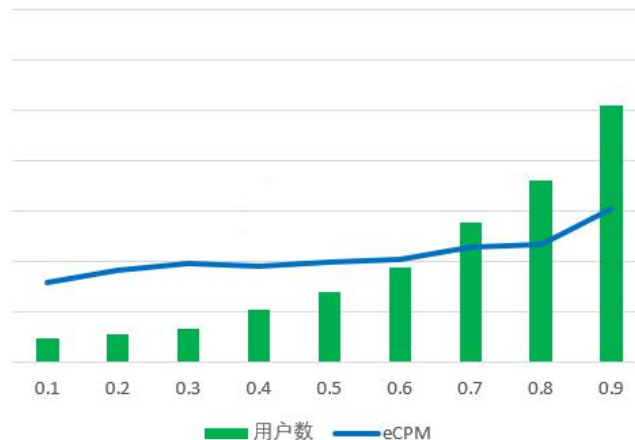
模型分与用户数、eCPM关系



好处：分数对应了效果  
问题：高分段人数太少

用百分位点  
为得分

分位分与用户数、eCPM关系



好处：分数与人数对应；标签横向可比  
新问题：eCPM和稀泥了，与分数关系弱

## 解决方案

分位点	5	10	15	25	35	45	60	75	90
归一分	9	8	7	6	5	4	3	2	1

## 进一步思考

小众标签和大众标签各超过50%的人，得分应该一样吗？  
标签小众，是确实兴趣小众，还是仅因为我们无数据或没挖掘？




## 业务背景

- 展示广告中的标签定向、兴趣标签的挖掘与使用

## 兴趣标签的挖掘算法

- 行为统计、行业建模、兴趣建模



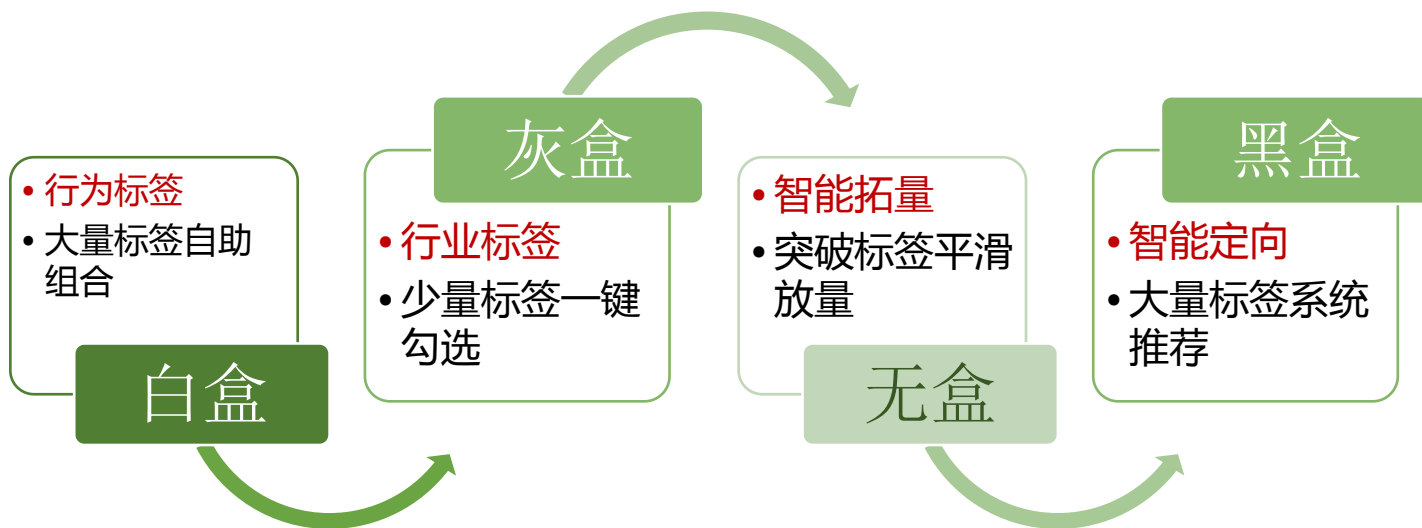
## 兴趣标签的落地应用

- 白盒、灰盒、黑盒、无盒

## 总结、思考和交流

标签挖掘的完成，不是工作的结束，  
甚至不是结束的开始，  
而仅仅是开始的结束。

把标签能力当成是一个产品，  
把广告主当成客户去运营，  
是中小广告平台标签工作发挥影响力的关键



我们落地工作的探索过程

## 从0到1打造标签挖掘系统的经验碎碎念

**专门团队规划**：真正需要标签的业务会首先千方百计东拼西凑、方法上从行为统计开始；若还是痛点才会组建专门挖掘团队，团队要**业务意识+算法建模+数据分析兼备**

**标签体系规划**：先要全面规划避免重构（4级共500+类目）、充分**与使用方沟通、参考业界**（腾讯、头条，不要参考国外FB、Google之类）、查阅百科垂站（游戏、小说、音乐、视频等）、重点内容从多角度规划且每个角度都可以打一个标签（游戏有题材、玩法、画风）；要留个主题类（如游戏有超休闲和游戏内付费；电商有二类电商）；**类目ID至少要有两套**（分别用于内容组织和界面呈现）

**数据标注规划**：算法人员要花大力气进行**种子标注**、阶段验收，减少返工；不要将完整类目体系扔给标注，对不重要细分类目**进行剪枝**，后续需要时再细分；最好**培养一名固定标注人力**

**挖掘系统规划**：整个流程要与眼下有多少标签、多少数据源无关，方便随时按需新增数据和标签；要**同步考虑实验数据流**；要抽样部分用户进行全流程追踪打标过程（借鉴广告归因，我们称为**标签归因**）

**效果评估规划**：非客观标签不要谈准确率；Item-Tag评估；**第一和第三视角评估**；AB实验评估

## 需求访谈

Q 你为什么不用标签定向？

你们标签太少了，没有适合我的

A

以上为3个月前的消息

Q 你为什么不用标签定向？

你们标签太多了，不知道哪些适合我

A

以上为3天前的消息

Q 那你还需要我们吗？

要要要！你帮我挖个人群包吧

A

Q 好的！你想我们怎样给你挖人群？

不知道啊，反正我要找对小额借款有兴趣的人

A1

我有经验！要挖掘有房贷或工资1万以上、有3张以上信用卡、装有3款以上贷款App... 的人，需要的话我提供App列表

A2

以上为3分钟前的消息

Q 好的！我想想（静静）...

## 需求思考

- ✓ 我们标签体系有固有组织逻辑，广告主自有他的特殊需求和想法
- ✓ 任何标签体系都不能完备
- ✓ 广告主的逻辑完全不管或完全迁就都会失败
- ✓ 因此我们提出“行业精选”标签概念，本质是提炼了一类广告主的需求、以标签形式满足人群包需求，内部技术上，我们称为商业兴趣标签
- ✓ 提炼广告主诉求并提供行业精选后，兴趣标签的消耗提升50%以上

## 需求访谈

Q 你为什么不用标签定向？

效果不够好，是不是标签量级太大了啊？好几千万！

A

以上为1个月前的消息

Q 你为什么不用标签定向？效果好像还可以了

效果是可以，量级太小，消耗不出去

A

以上为3分钟前的消息

Q 好的！我想想（静静）...

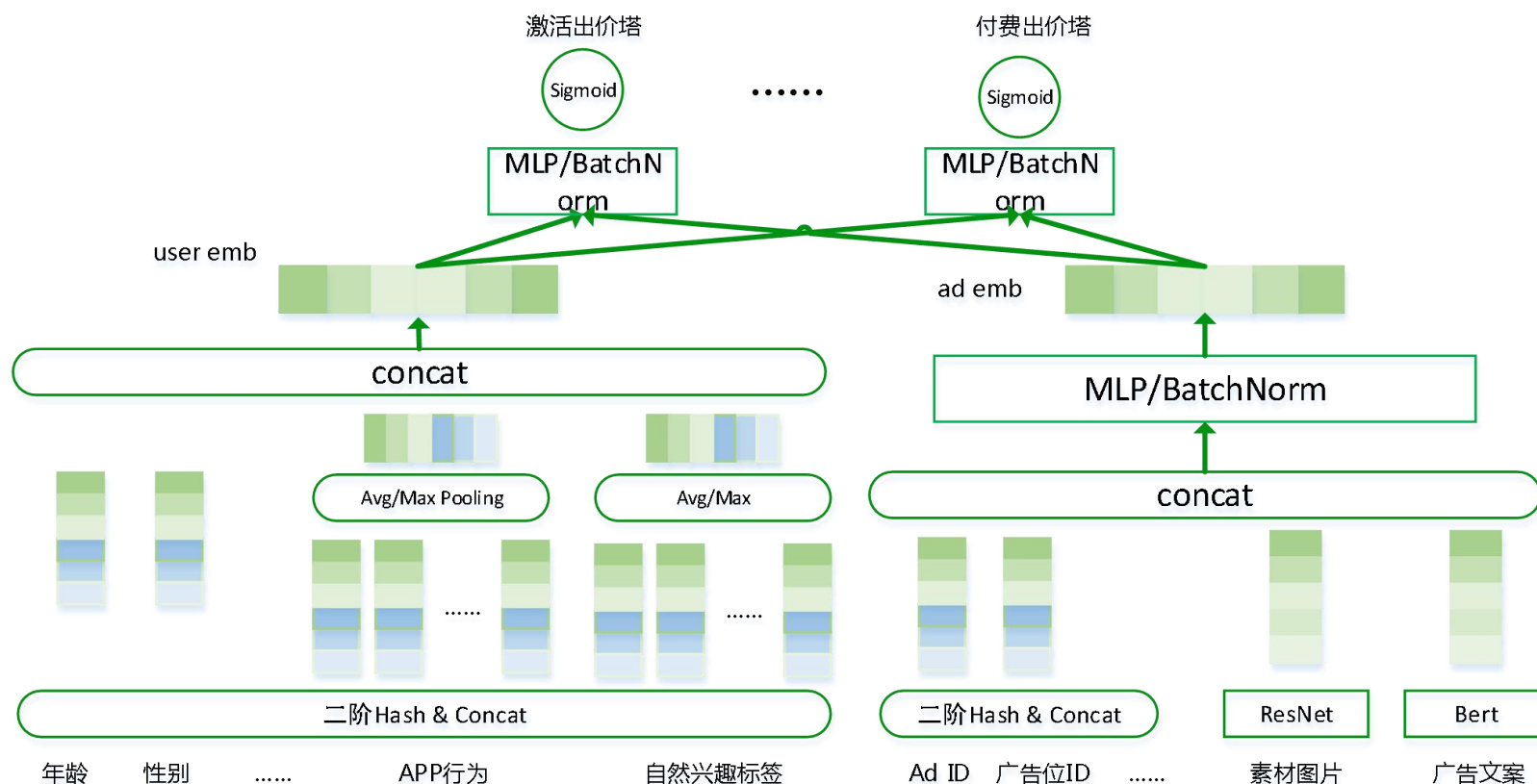
## 需求思考

- ✓ 表面上广告主诉求不一，实际上是一贯的：既要质又要量
- ✓ 能否手动放开标签定向？实践表明突然放开，会导致oCPX模型无法适应，成本激增，广告快速衰亡
- ✓ 智能拓量为正解：与oCPX配合控成本的前提下自动平滑拓量

引入智能拓量能力后，整个定向能力定位为致力打造  
**标签起量 + 拓量起飞 + oCPX控成本**  
的投放模式，并主要服务于中长尾广告主

# 智能拓量模型：经典的双塔模型

- ✓ 自动拓量模型为典型的多塔模型，user塔与商业兴趣建模基本一致
- ✓ 在线策略引入熔断机制以**确保成本**
- ✓ 重点仍然是样本的艺术
- ✓ 特别注意利用类似广告组的样本进行**样本增强**



## 需求访谈

Q 你为什么不用智能拓量？

之前用了，根本拓不了量

A

以上为1天前的消息

Q 我们后台排查发现，你开拓量的广告，没有选标签定向，没有可拓量的空间哦

所以说到底我还得先学会用标签定向，我再看看吧.....

A

以上为1个月前的消息

Q 你为什么不用智能拓量？

我按照你的要求选了定向，然并卵.....

A

以上为1天前的消息

Q 哎，你是金融的广告，为啥选教育类的标签呢-\_-! 原定向就超成本了，没法再拓量了哦

教育类人群相对有经济实力吧.....

A

Q 哦！明白了，我想想（静静）...

## 需求思考

智能拓量运营几个月后，消耗占比很快到了天花板，原因分析：

- ✓ 真金白银面前新功能容易背锅
- ✓ 用错了，不满足拓量条件
- ✓ 最大原因有二：a) 不用标签，无法拓量；b) 不会用标签，原定向不足以进入拓量阶段

因此，一键自动为广告主勾选标签组合，才是真正的智能定向。  
这个能力我们称之为“系统推荐”

白盒还是黑盒？我们已经“不相信”广告主了，也为了完全隔离广告主影响，选择了黑盒



# 标签推荐建模：一个“没有”参考文献的难问题

## 问题描述：

为广告求解标签组合，使得量级和效果最优

## 基于统计：

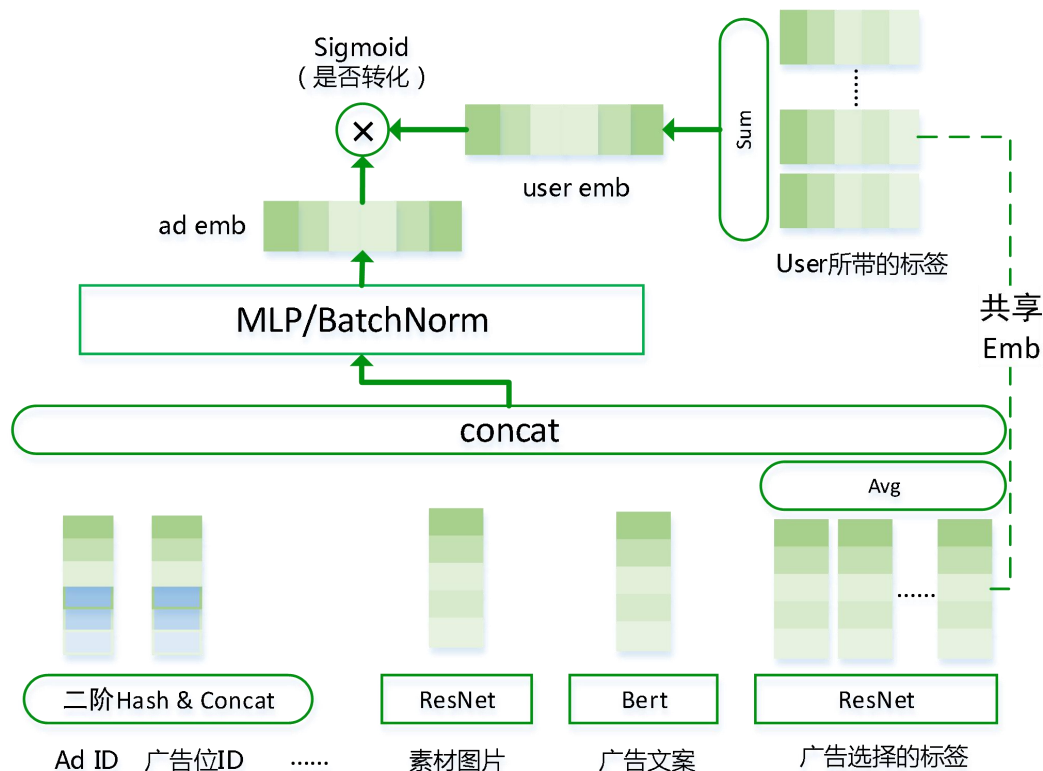
类似TGI思想，根据同类广告的历史表现，找出效果最好的Top N个标签；上线效果明显优于广告主自选

## 问题延伸：

统计方法难以迭代，但似乎没有相关模型论文

## 基于DNN：

偶尔读到微信看一看为用户打标的算法UPTM，透过现象看本质，将文章和用户分别对应我们的用户和广告，得到如下算法



## 核心思想：

每次曝光有Ad和User，User用其所带标签来表示，即得到了AD和标签的转化意义上的相关性

## 在线预测：

模型输出为Ad塔模型以及所有标签的Emb，在线生成Ad Emb后计算出最相似的Top N个标签即可。

实验表明相比统计基线，转化率/转化成本大幅提升；而且网络结构还有很大改进空间

# 定向标签实验：无评估勿迭代，但ABTest用对了吗？

## ABTest用对了吗：

ABTest众所周知，但少为人知的是标签定向场景下，需要对实验组和对照组进一步拆分对比

## 为何要拆分对比：

一般实验中，实验策略会作用于实验组的所有用户，而标签实验并不是；若不细分，实验效果将混然于众人，很难突显

## 如何细分流量：

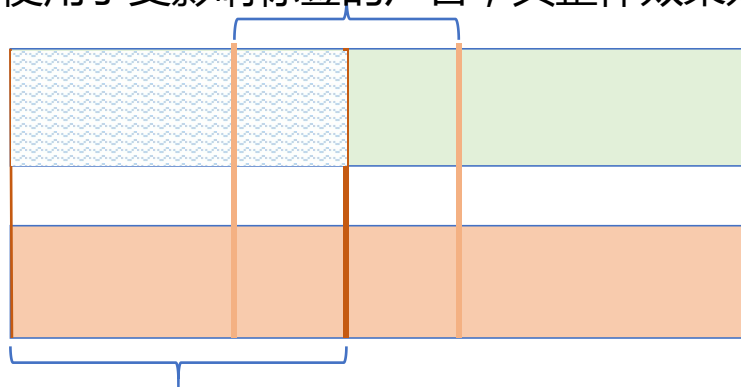
划分为平台测和广告主侧分别各自关注的流量

广告主关注的流量：

使用了受影响标签的广告，其整体效果如何？

对照组

实验组



平台侧关注的流量：

标签改变了的用户，其整体效果如何？

## 实操注意：

- 蓝色斜线部分的用户，作为对照组时不要改变
- 小流量时一定要有 AATest

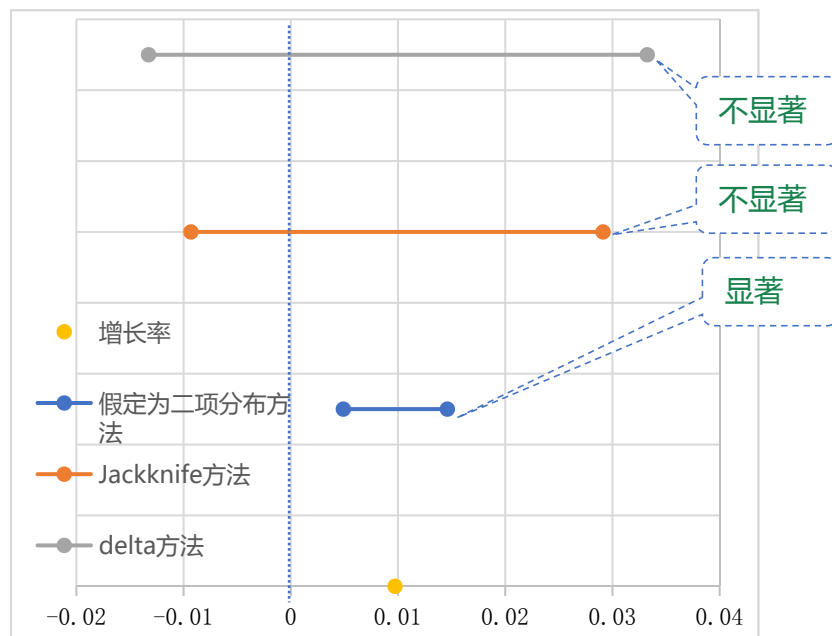
# 定向标签实验：你看显著性了吗？显著性算准确了吗？

为何计算置信区间：必须统计上显著的提升，才是上线的依据

如何计算置信区间：核心是**计算方差**；调用3种方法，在性能与精度间取得平衡



不同方差计算方法对比



注：以某次实验在20191111号一天数据为样本，得出CTR提升1%。在IID假设下是显著的（置信区间不包括0），但其它两种方法均显示并不显著

### 业务背景


- 展示广告中的标签定向、兴趣标签的挖掘与使用

### 兴趣标签的挖掘算法

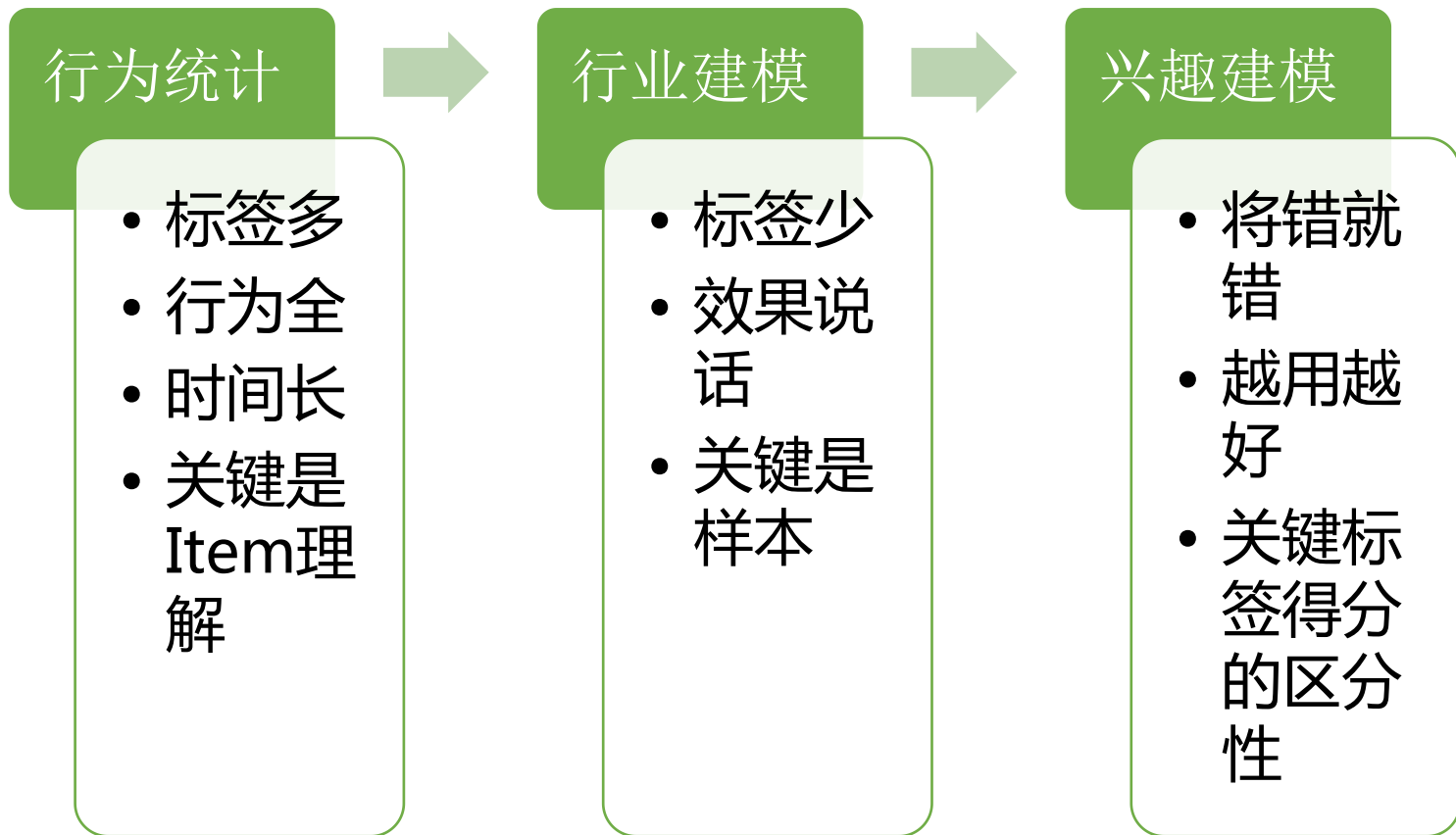
- 行为统计、行业建模、兴趣建模

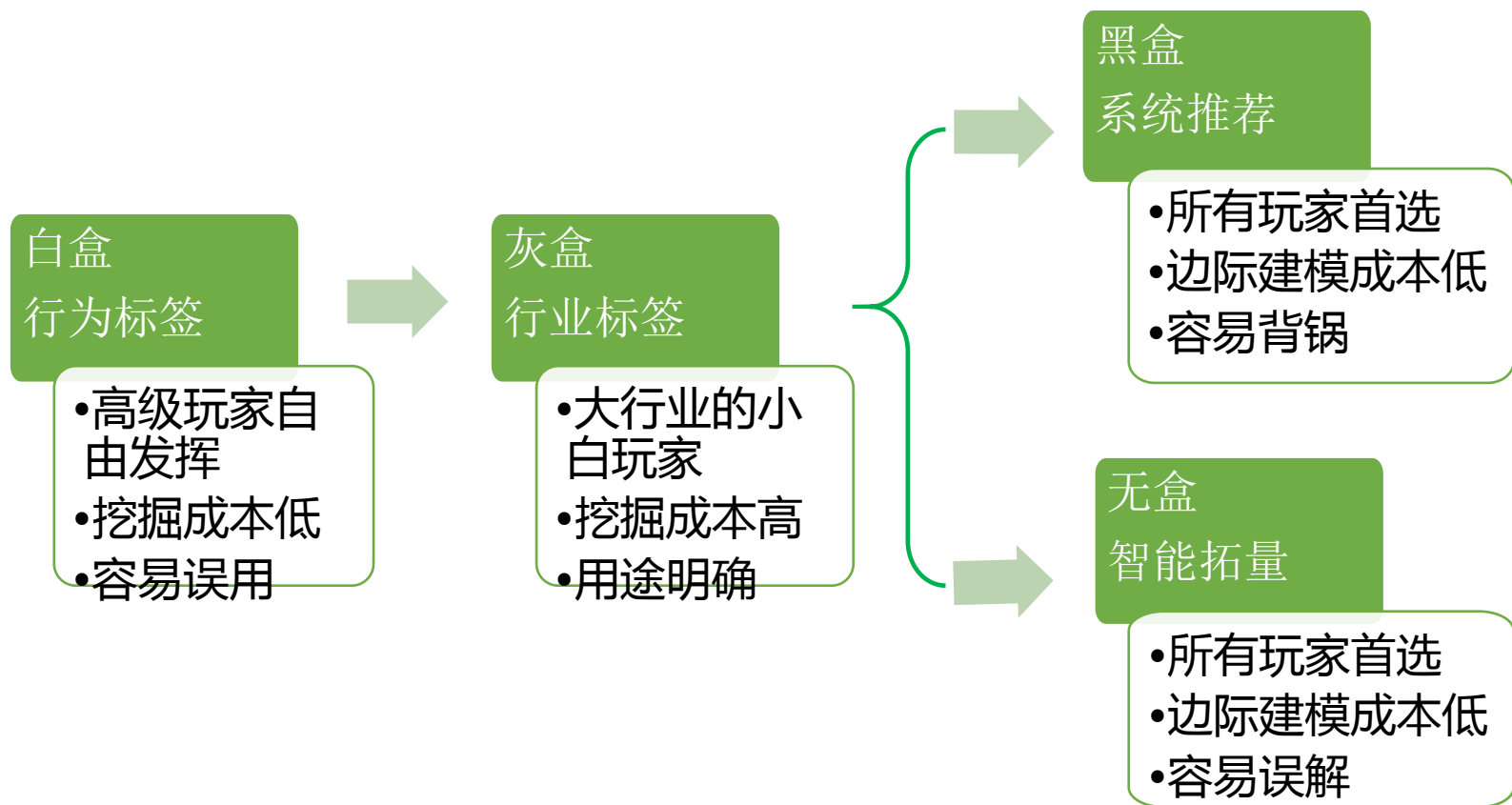
### 兴趣标签的落地应用

- 白盒、灰盒、黑盒、无盒



### 总结、思考和交流





**语言**本质是定义了一套**双方能一致理解**的符号和规则

**标签**自身要**准确明确**，还要确保**易用好用**

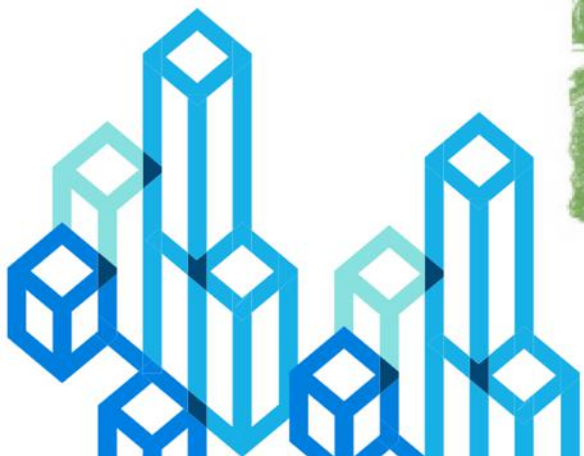
P.S. 本项目获得公司事业群级的《极致体验奖》



# Q & A



扫码交流





麦思博(msup)有限公司是一家面向技术型企业的培训咨询机构，携手2000余位中外客座导师，服务于技术团队的能力提升、软件工程效能和产品创新迭代，超过3000余家企业续约学习，是科技领域占有率第1的客座导师品牌，msup以整合全球领先经验实践为己任，为中国产业快速发展提供智库。



高可用架构公众号主要关注互联网架构及高可用、可扩展及高性能领域的知识传播。订阅用户覆盖主流互联网及软件领域系统架构技术从业人员。高可用架构系列社群是一个社区组织，其精神是“分享+交流”，提倡社区的人人参与，同时从社区获得高质量的内容。