

腾讯广告 检索技术架构升级演进

郑宇欣 腾讯广告

精彩继续！ 2020 一线大厂创新技术 实践大盘点

QCon 全球软件开发大会

📍 上海站

时间：2020年12月20-21日

地点：上海·宝华万豪酒店



扫码查看完整日程

📍 北京站

时间：2021年01月08-09日

地点：北京·国际会议中心



扫码查看完整日程



极客时间 App

轻松学习 · 高效学习

大会到场福利

新老用户

可获得课程优惠口令 立减10元

QCon66666

新用户

可免费领取一门课

扫码立领 >>>



vanke

万科南方区域

Ad



深圳北站约1.6公里 | 万科新房

龙华人的万科房!

约600米到白石龙站, 建面约83-139㎡装修房, 点击预约参观!

6 hrs ago

..

Purcotton 全棉时代

Ad

终于搞活动了👉
足足2大盒口罩! 100片!
现每片低至0.69元~数量有限【手慢无】



👉 立即购买

8 hrs ago 最近有4.4万人购买

..

广告




RIMOWA 日默瓦

关注公众号

广告投放



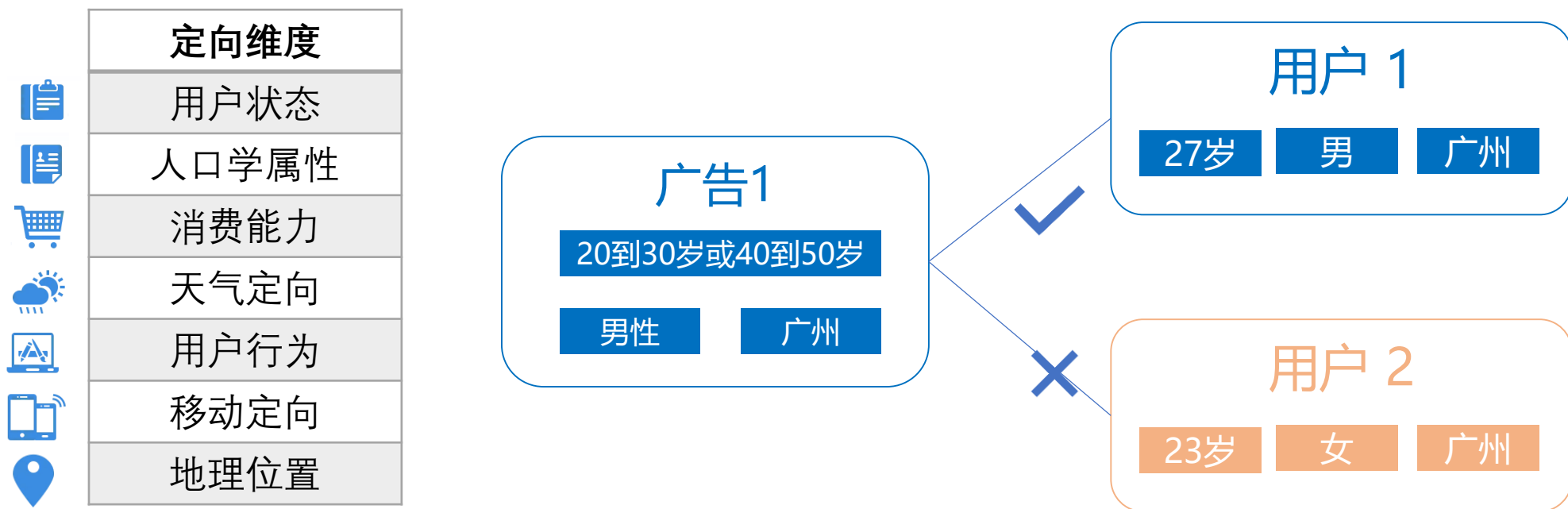
广告投放	What	Why	Who	When	Where	How	How much
广告	广告创意 落地页	推广目标	受众用户	投放时间	投放版位	账户设置 (计划、 出价等等)	预算
房产广告		表单线索	常驻深圳 25 - 50岁	12月1日- 12月31日	朋友圈	出价20元	1千元

广告定向：选择广告可曝光用户，实现精准营销

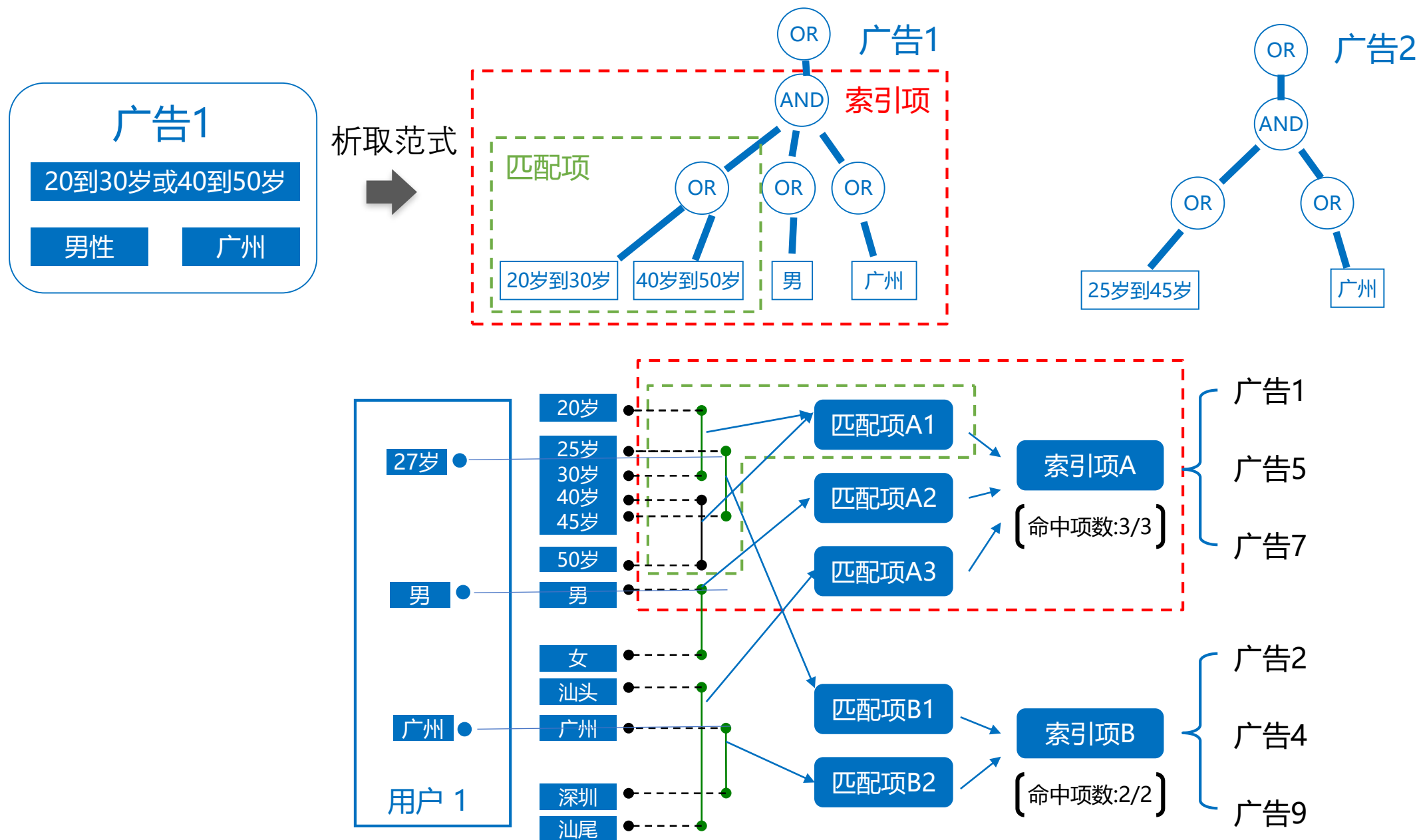
“我知道在广告上的投资有一半是无用的，但问题是我不知道是哪一半。”
--约翰·沃纳梅克



通过定向选择用户：



定向检索：高性能布尔表达式检索

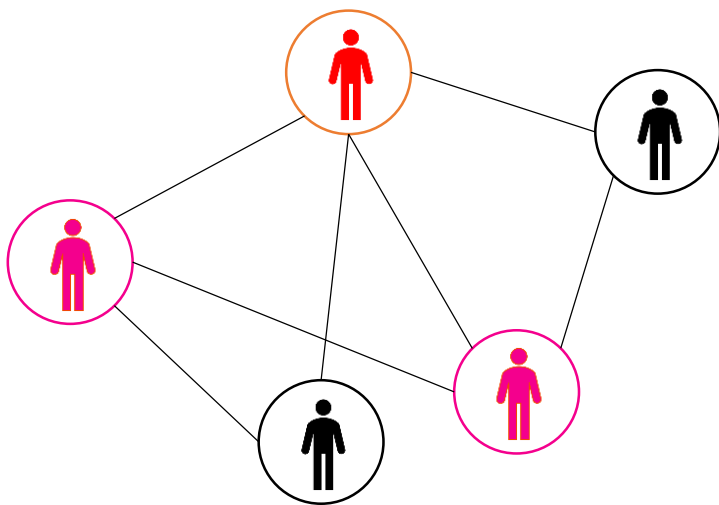


智能定向：解决人工选择定向调优难的问题

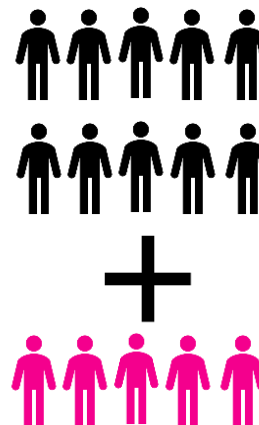
已有转化用户



通过机器学习找到大盘用户中的
高转化用户



智能定向人群



原定向人群

新增智能定向人群

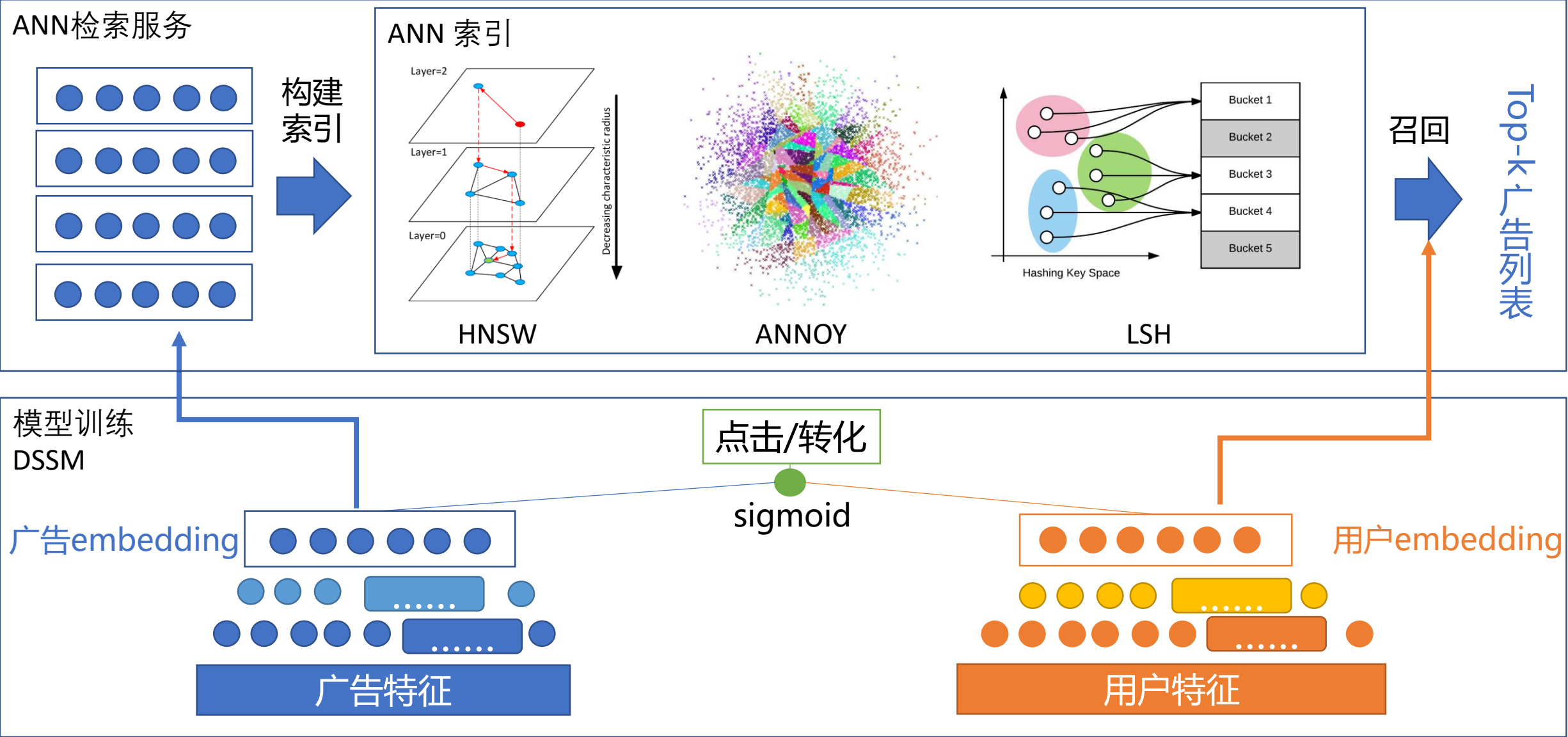
基于 DSSM 的语义近邻检索

1. 曝光到点击预估
2. 点击到转化预估
3. 曝光到转化预估
4. 曝光到点击&转化预估

语义标签检索

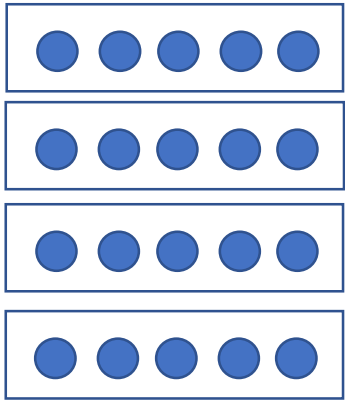
1. 行为定向
2. 兴趣定向
3. 意向定向
4. App 安装定向
5. 后台策略标签定向

基于 DSSM 的语义近邻检索



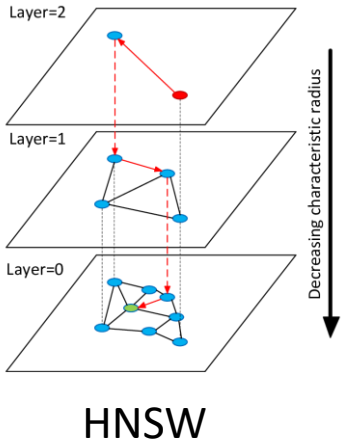
语义近邻检索

ANN检索服务

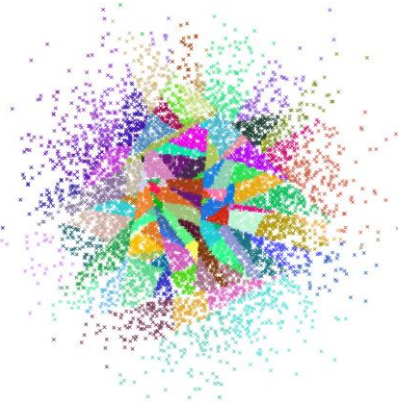


构建索引

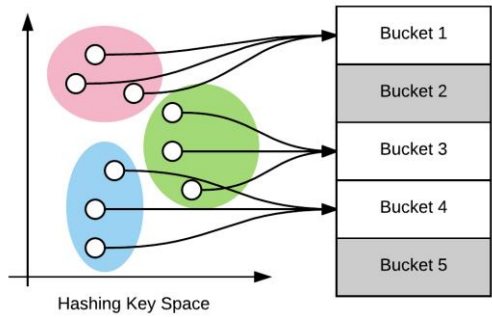
ANN 索引



HNSW



ANNOY



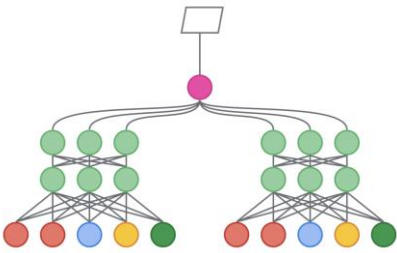
LSH

召回

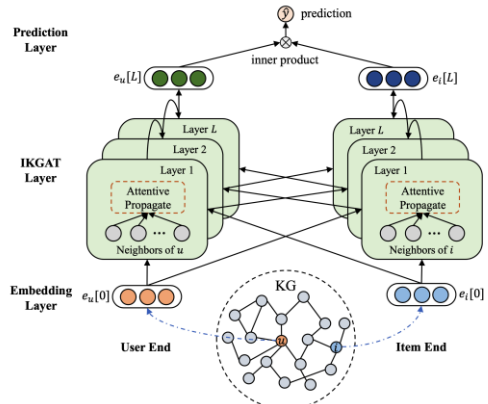


Top-K 广告列表

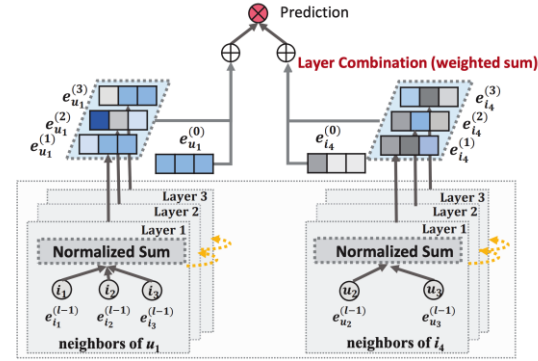
支持更先进模型



DSSM



Graph Attention Network



LightGCN

ANN Recall 过低：距离必须是可度量的(Metric)

- 距离函数定义：

设 X 为一个非空集合，他的元素叫做点。 R 是全体实数的集，若函数 $d : X \times X \rightarrow R$ 对于任意 $x, y, z \in X$ 合乎条件：

(a) 若 $x \neq y$ ，则 $d(x, y) > 0$ ，并 $d(x, y) \geq 0, d(x, x) = 0$ ；（称作正定性）

(b) $d(x, y) = d(y, x)$ ；（称作对称性）

(c) 对于一切 $z \in X$ ， $d(x, y) \leq d(x, z) + d(z, y)$ ；（称作三角形不等式） [2]

- 内积不是距离函数：转化成距离函数

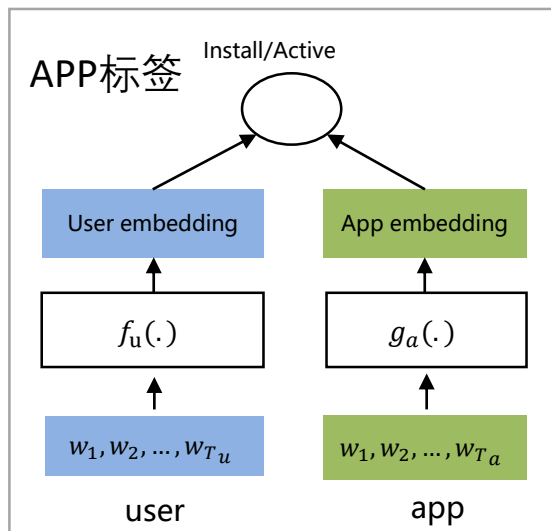
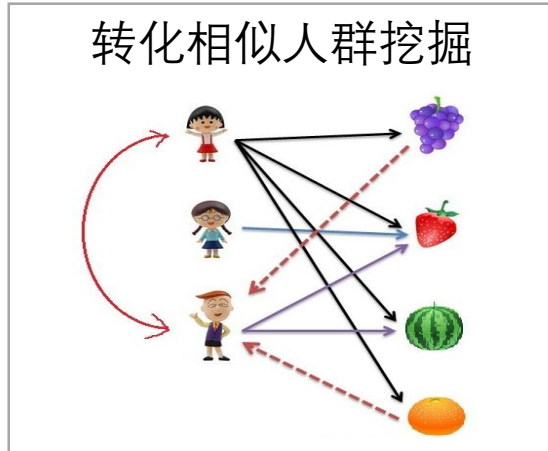
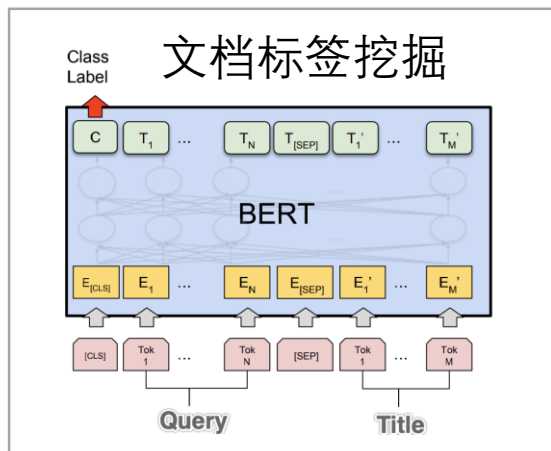
数据库点转换	$\hat{x} = (x, \sqrt{1 - \ x\ ^2})^T$
查询点转换	$\hat{q} = (q, 0)^T$

- Recall 60% -> 90%+

语义标签检索

离线挖掘广告/用户的关系

标签挖掘



人群包挖掘

兴趣标签挖掘

.....

用户

行为定向

兴趣定向

意向定向

App 安装定向

自定义人群

...

创意

AppID

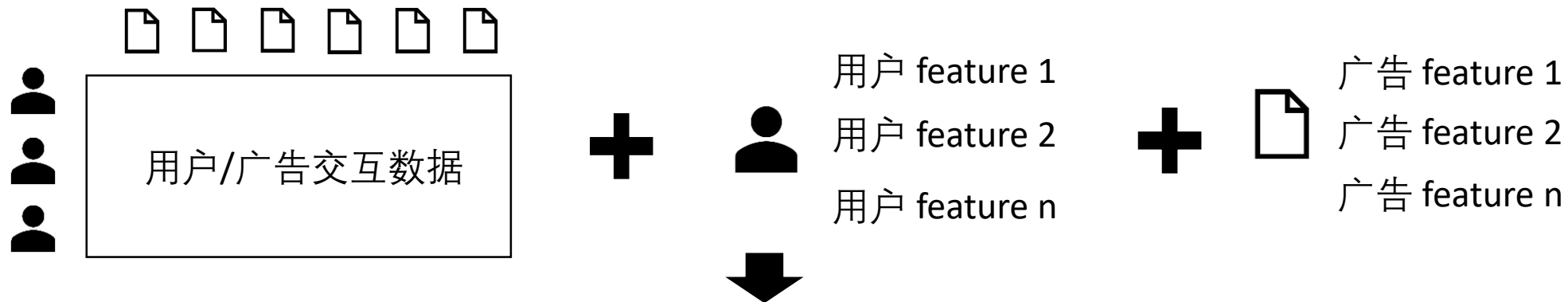
相关转化人群

账户uid

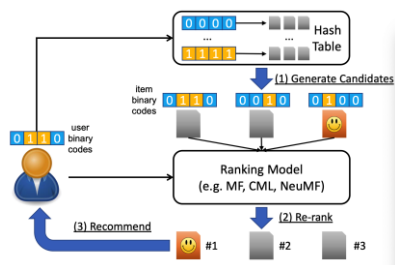
...

广告

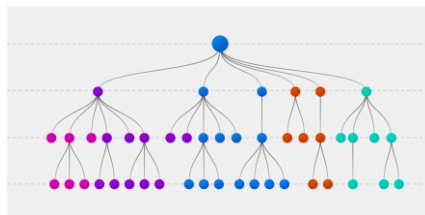
算力提升->在线使用更先进模型



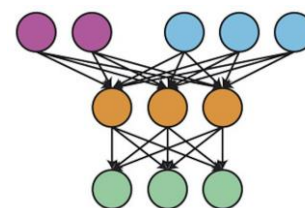
召回模型：基于表示学习(representation learning)的模型召回



Learned hash

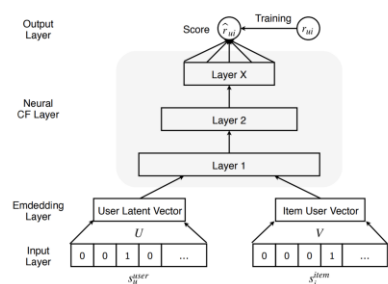


Learned tree

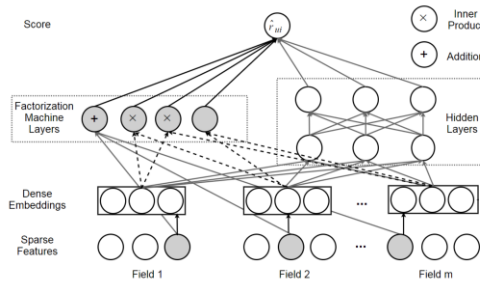


Learned path

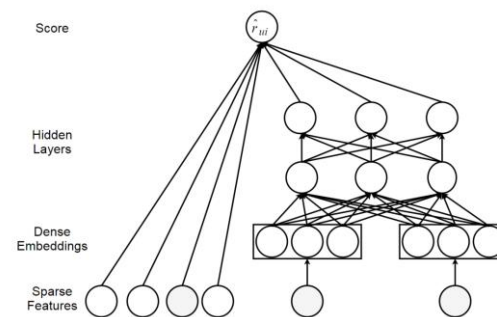
粗排模型：表达能力更强的模型



NeuralCF

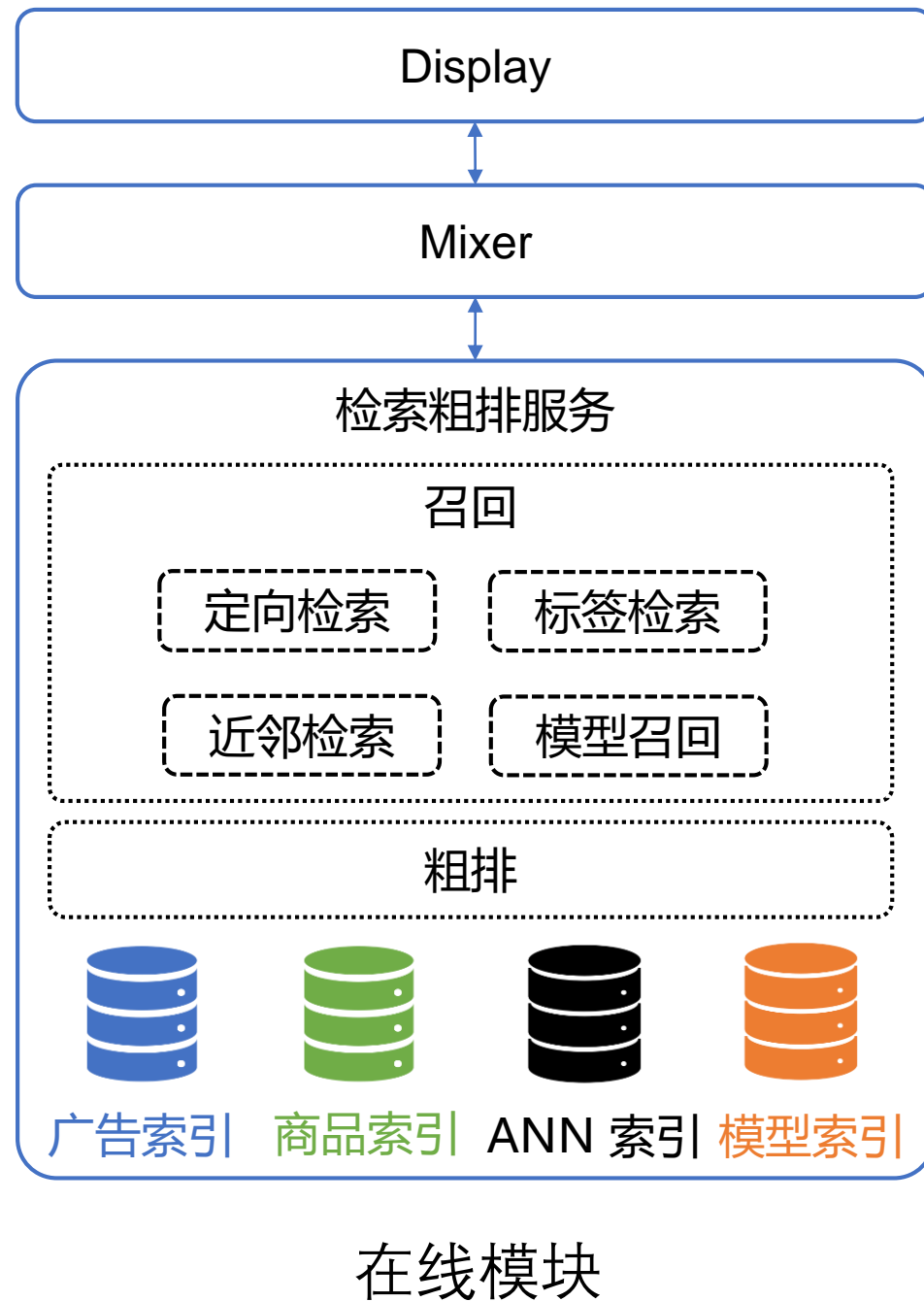
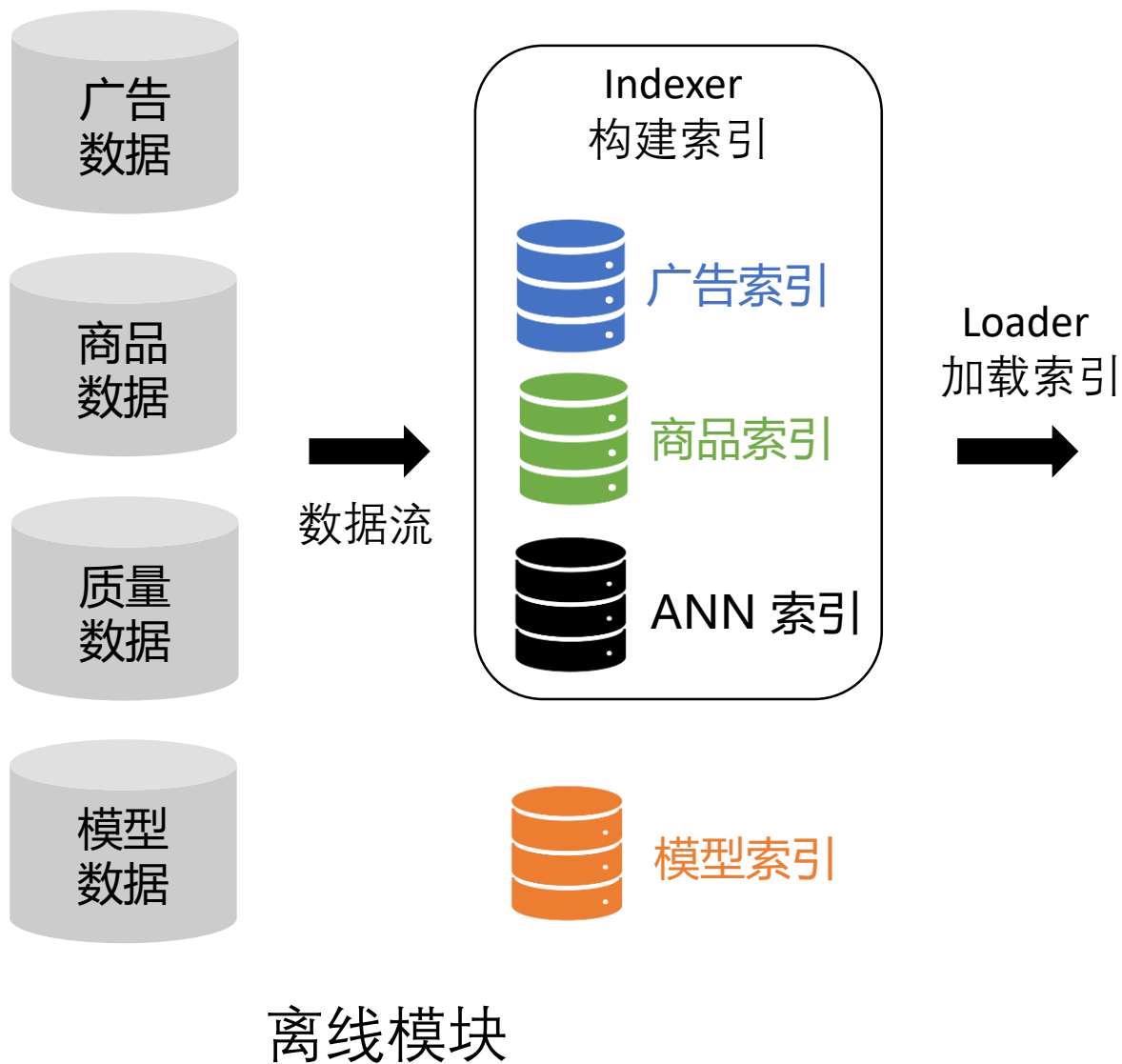


DeepFM

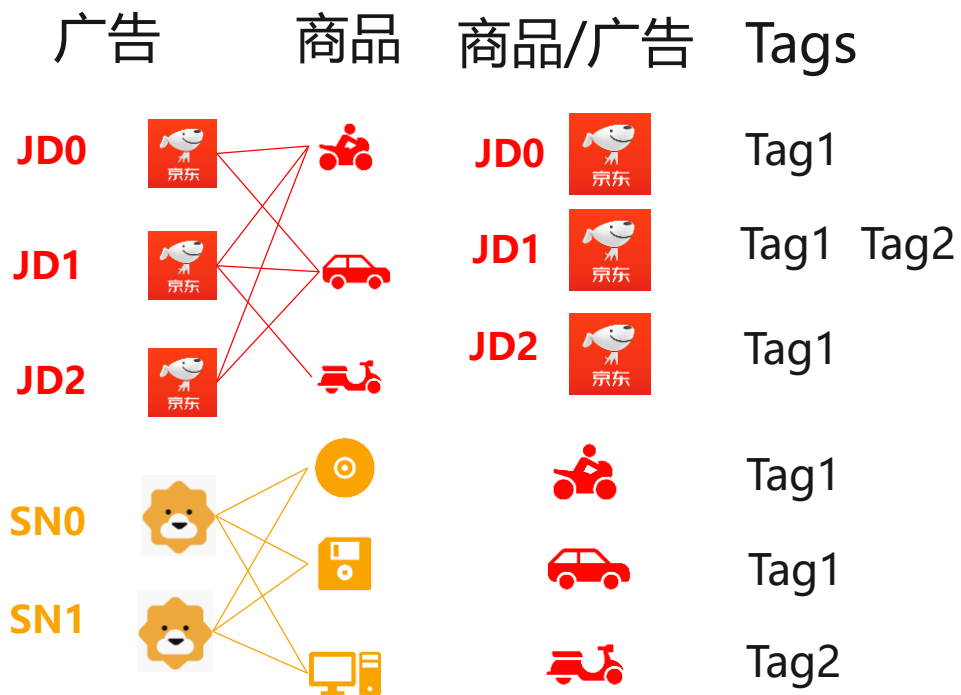


Wide & Deep Learning

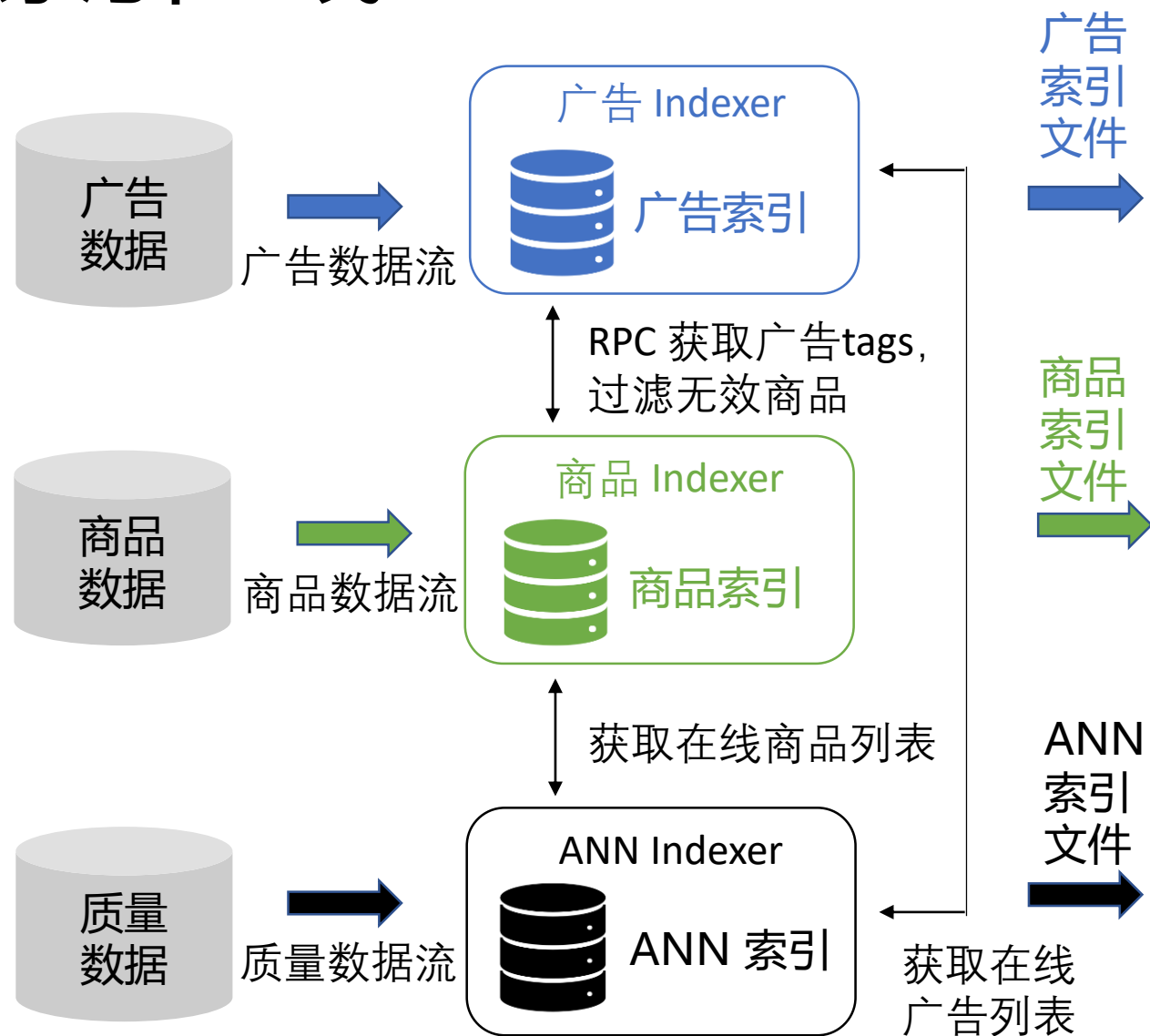
广告检索系统架构



索引独立构建：加快索引上线



离线挖掘商品与广告的匹配关系，
并通过tag标记



检索初排分离：合理利用资源，更灵活的分片策略

检索粗排服务

召回

定向检索

标签检索

近邻检索

模型召回

粗排



广告索引



商品索引



ANN 索引



模型索引

旧架构

检索服务

召回模块

定向检索

标签检索

近邻检索

模型召回



广告倒排



商品倒排



ANN 索引



召回模型

RPC

粗排服务：多路并行

shard 1

shard 2

shard n



广告正排



商品正排



排序模型

新架构



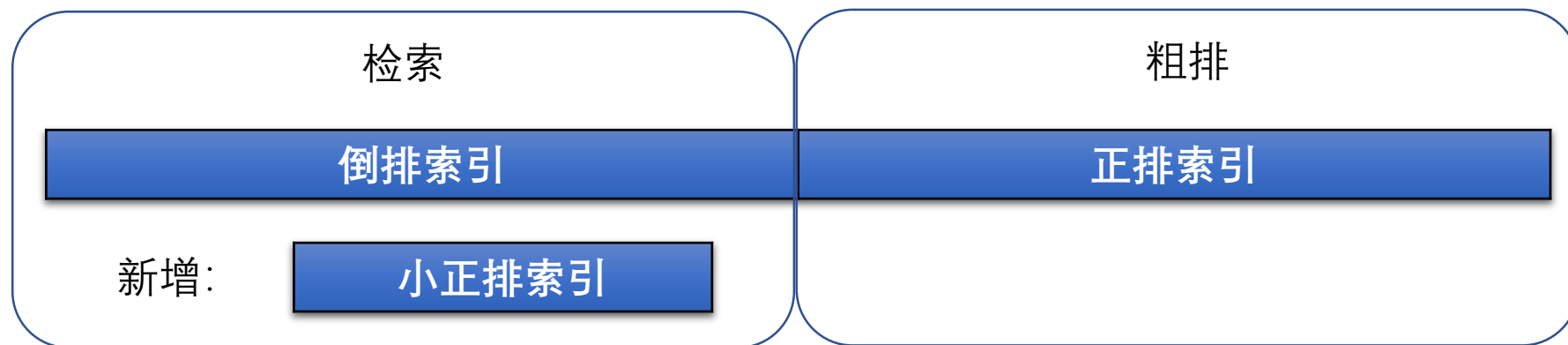
模块



服务

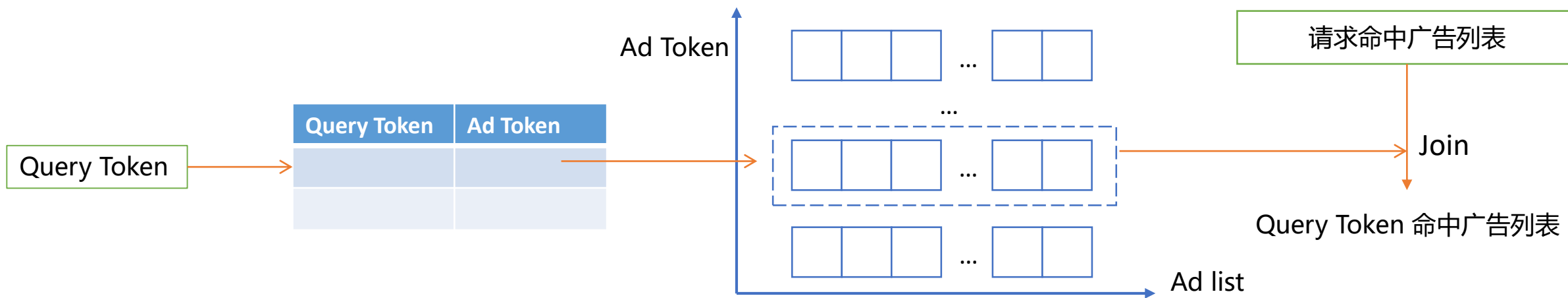
检索粗排分离挑战：模块解耦

- 数据解耦

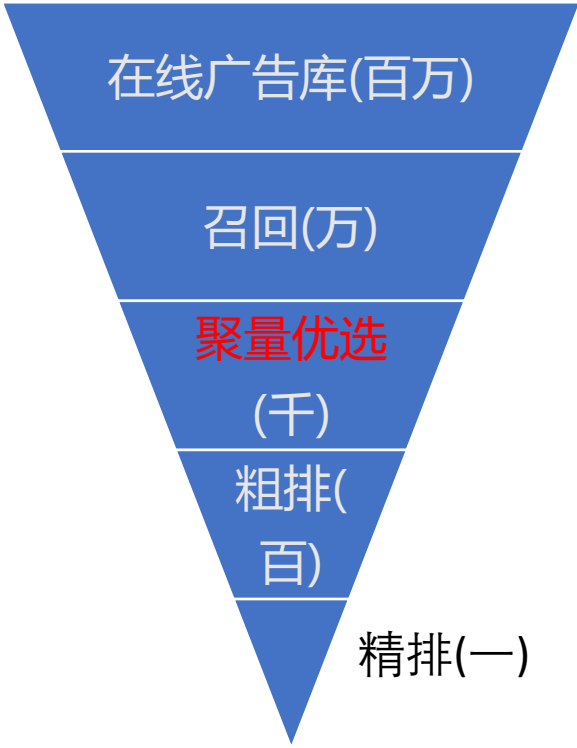
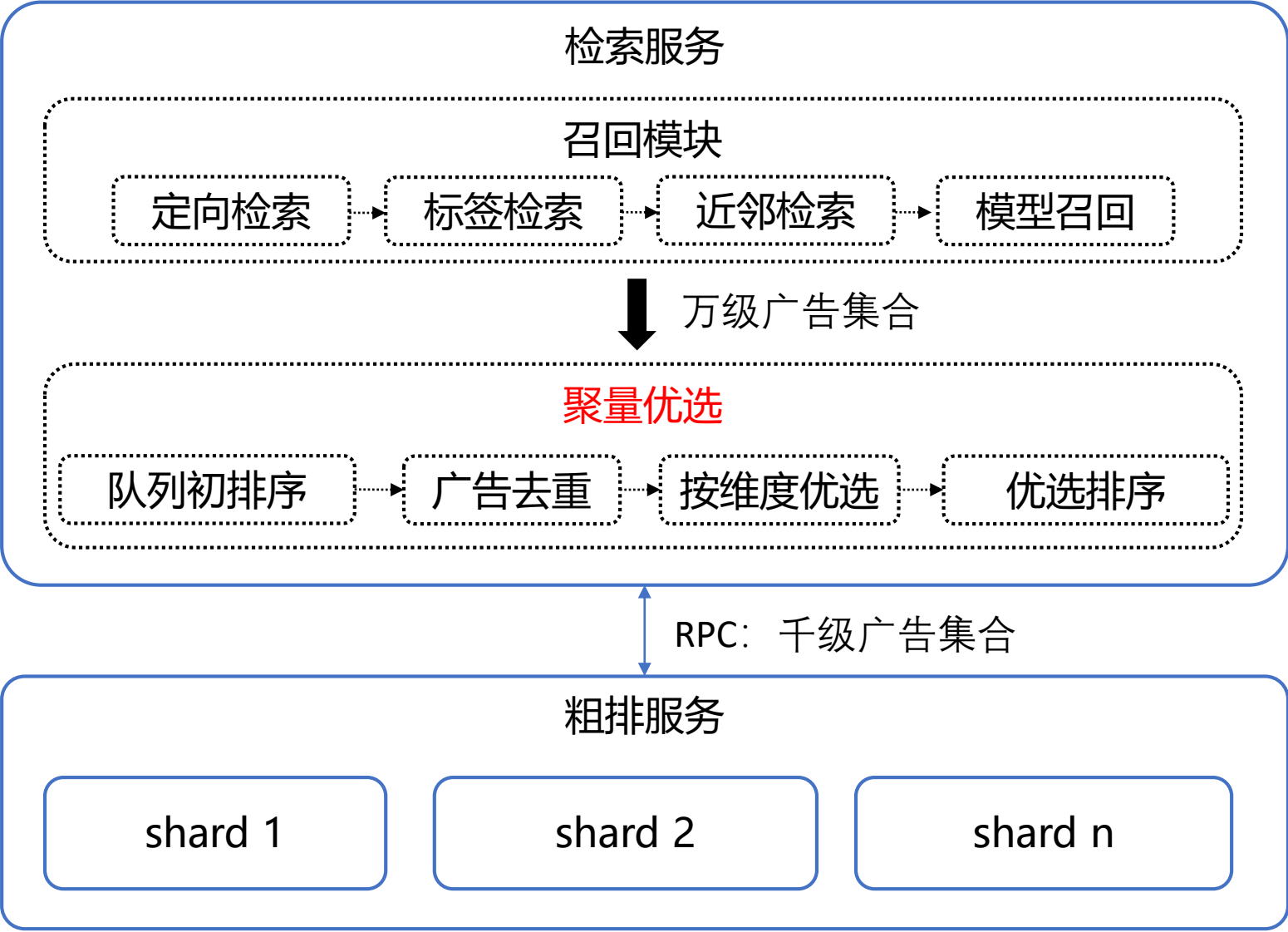


- 逻辑解耦

- 粗排的计算依赖检索的中间结果 query token -> 命中广告
- 方案一：RPC 带上中间结果-> 传输数据太大
- 方案二：使用的地方放到检索服务->检索逻辑引入太多业务逻辑
- 方案三：粗排重建命中关系-> 时间换取空间（上线方案）



聚量优选：融入漏斗模型，解决召回队列过长



检索服务并行召回：降低端到端时延

检索模块：召回算法间并行

定向检索
分片并行

shard 1

shard 2



广告倒排索引



商品倒排索引

标签检索



广告标签索引

近邻检索
模型级并行

模型
1

模型
2

模型
n



ANN 索引

模型召回
模型级并行

模型
1

模型
2

模型
n



召回模型

召回结果多路合并

聚量优选

队列粗排序

广告去重

按维度优选

优选排序



THANKS



THANKS