

荣耀推荐算法架构演进实践

冯晓东 AI架构师

TGO 鲲鹏会 - 助力企业赢得数字化先机

TGO 鲲鹏会

tgo.infoq.cn

TGO 鲲鹏会是科技领导者同侪学习社区，致力于把技术领导者和专家连接在一起，通过领导力峰会、专属小组活动、闭门沙龙等形式，为所有“孤军奋战”的科技管理者获得自身的成长和职业的发展。

TGO鲲鹏会目前拥有1600+位高端会员。在北京、上海、深圳、广州、杭州、南京、成都、厦门、台北、硅谷、武汉、苏州等全球十二个城市设立分会。

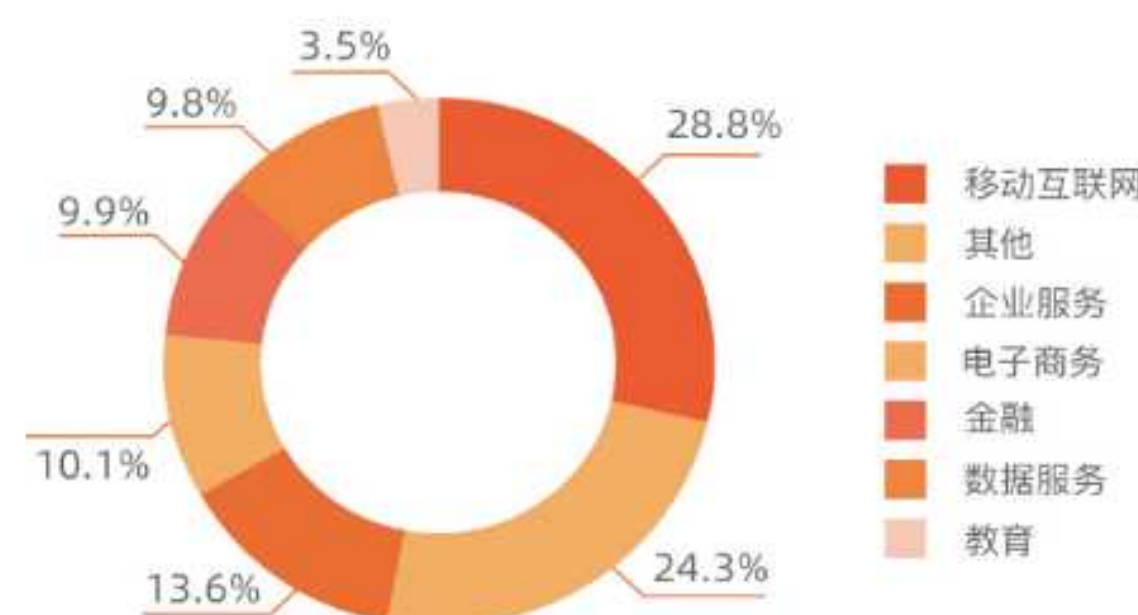


专属 小组活动

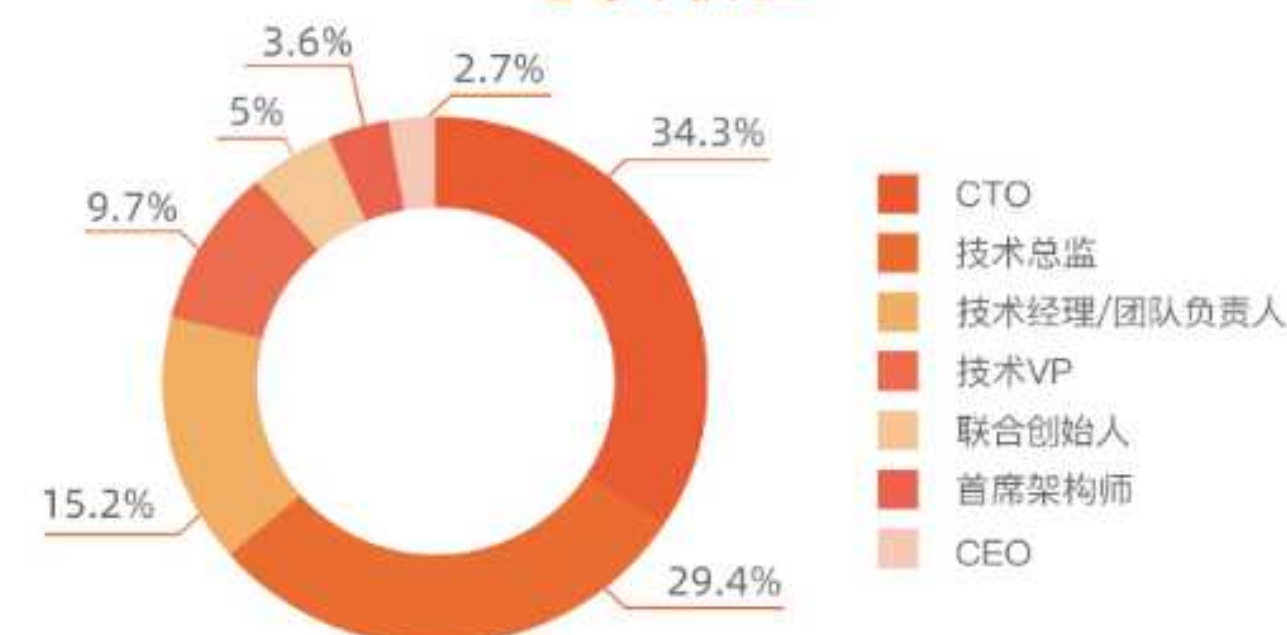
组内6-8位学员的每月专属聚会，通过案例分享战略、技术、管理等实战心得



学员领域



学员职位



目录

- 背景
- 推荐算法的演进
- 基于跨域特征的用户冷启动
- 基于跨域序列化的召回模型
- 总结与展望

背景

智能推荐中心支撑多个业务领域和多种内容类型的推荐。

荣耀商城

商品推荐：

手机
平板
PC
智能穿戴
亲选产品



我的荣耀

信息流推荐：

咨询
玩机技巧
活动
服务



荣耀俱乐部

视频推荐：

评测视频
爱摄影
发布会



背景

应用层



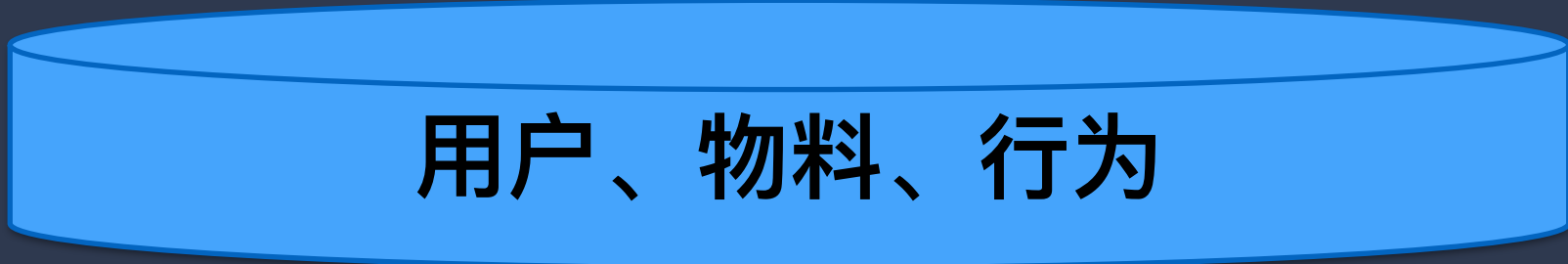
- 领域多：3+
- 目标多：CTR、CVR、留存率

算法层



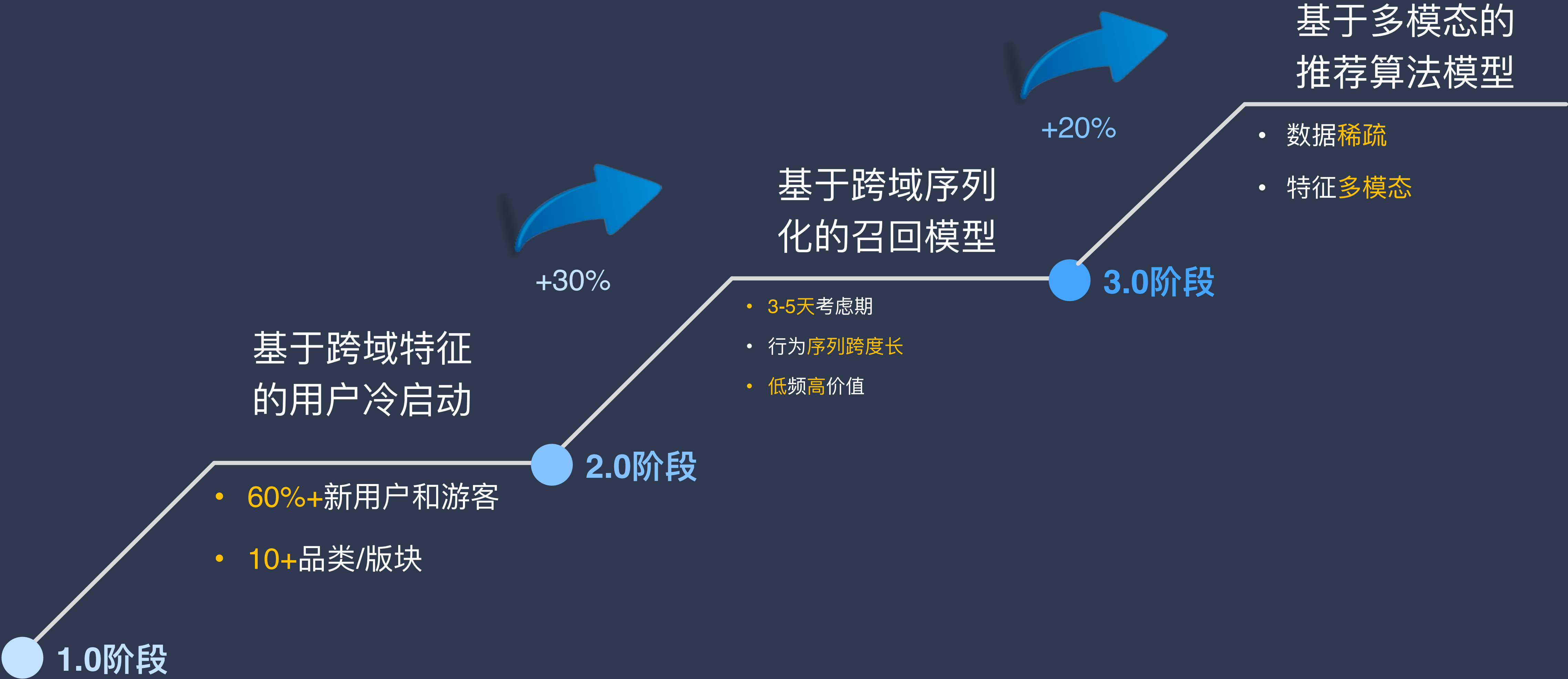
- 噪音大：购买考虑期较长难以挖掘兴趣
- 冷启动：新用户特征难捕捉

数据层

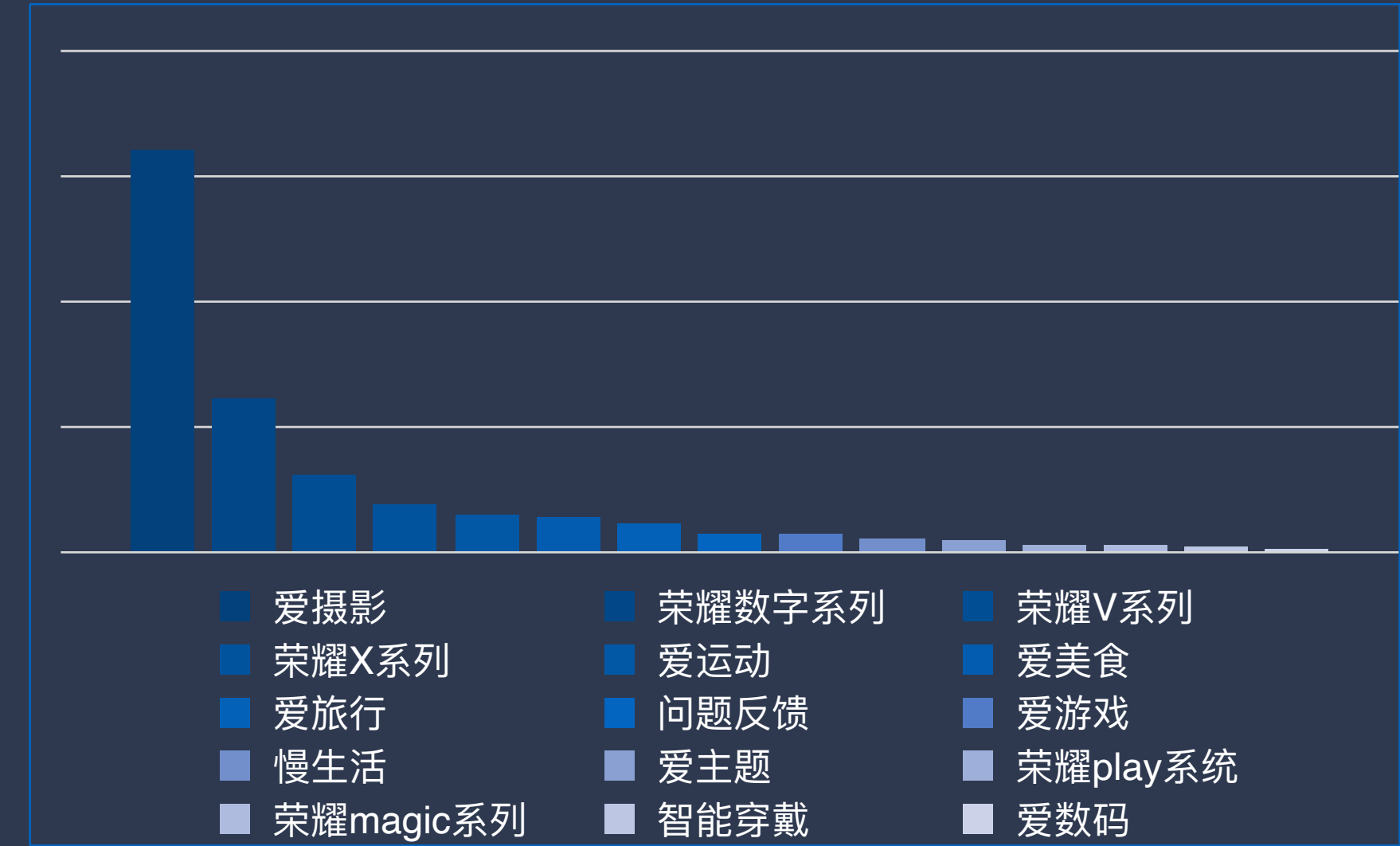
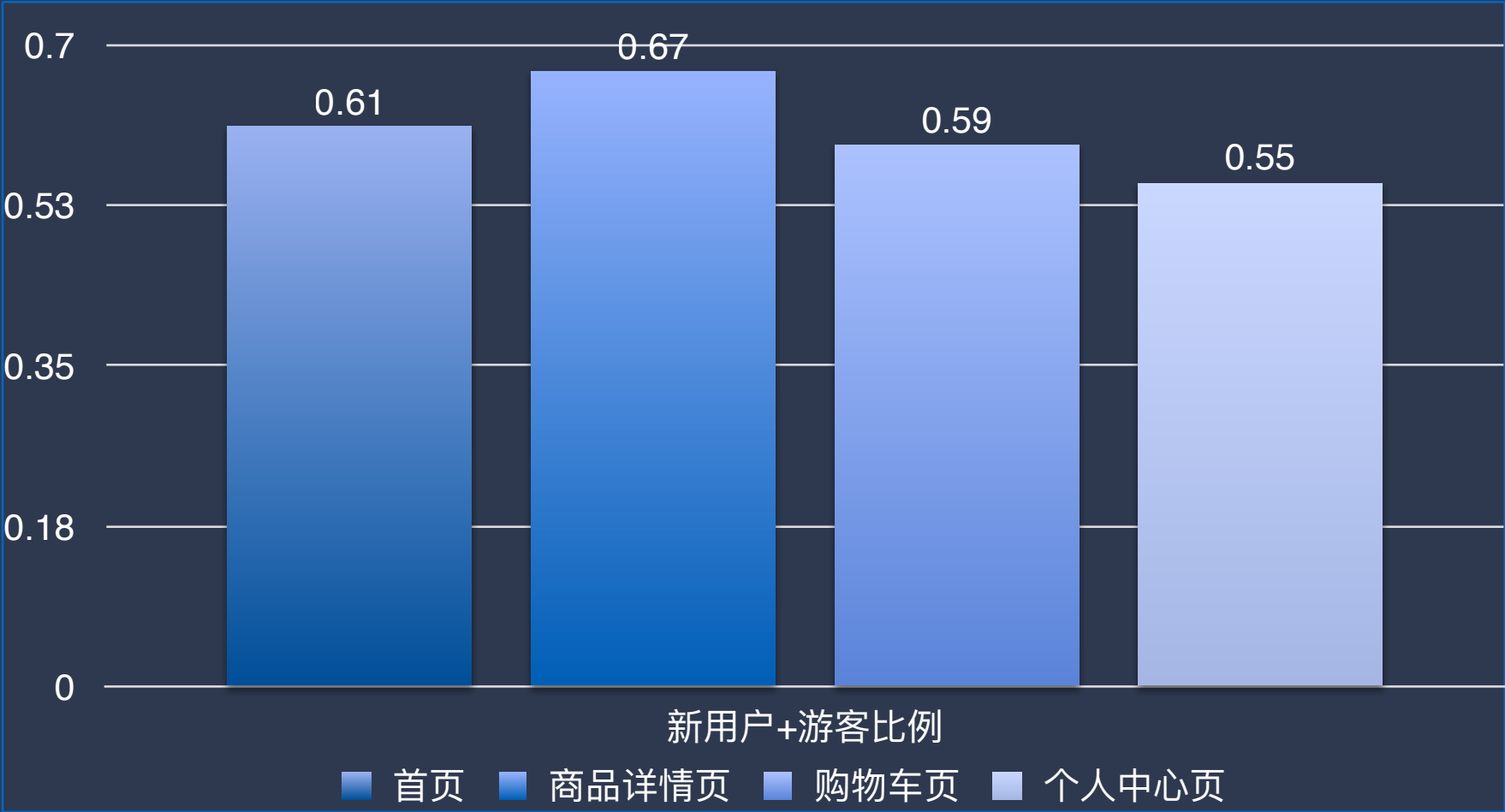


- 数据稀疏：评分矩阵density < 2%
- 数据分散：100+表

推荐算法的演进

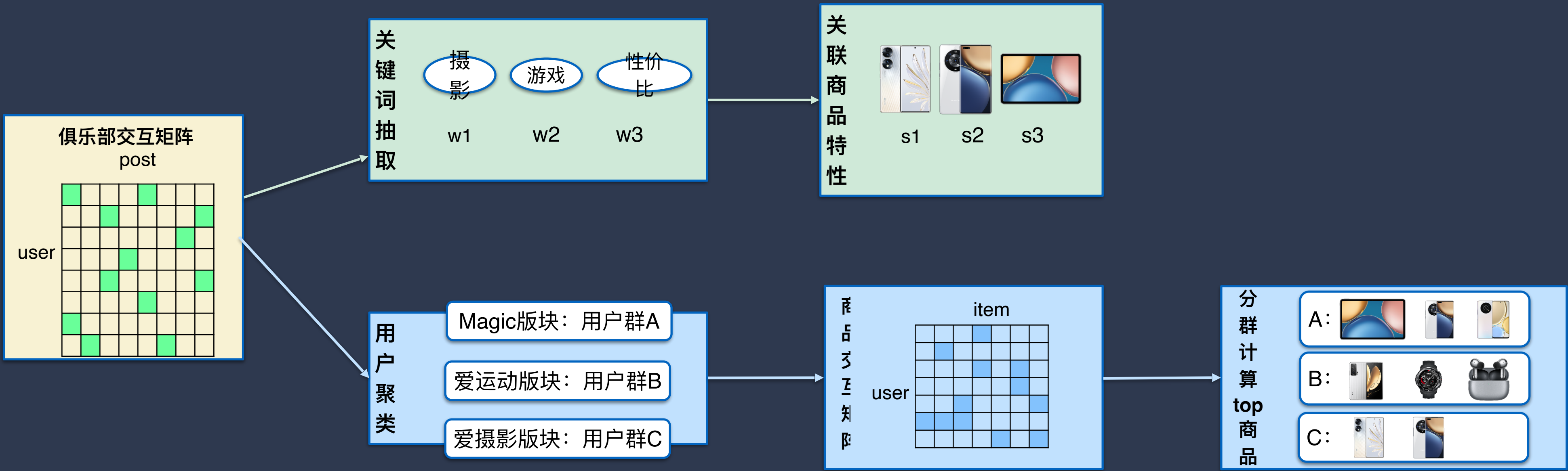


基于跨域特征的用户冷启动



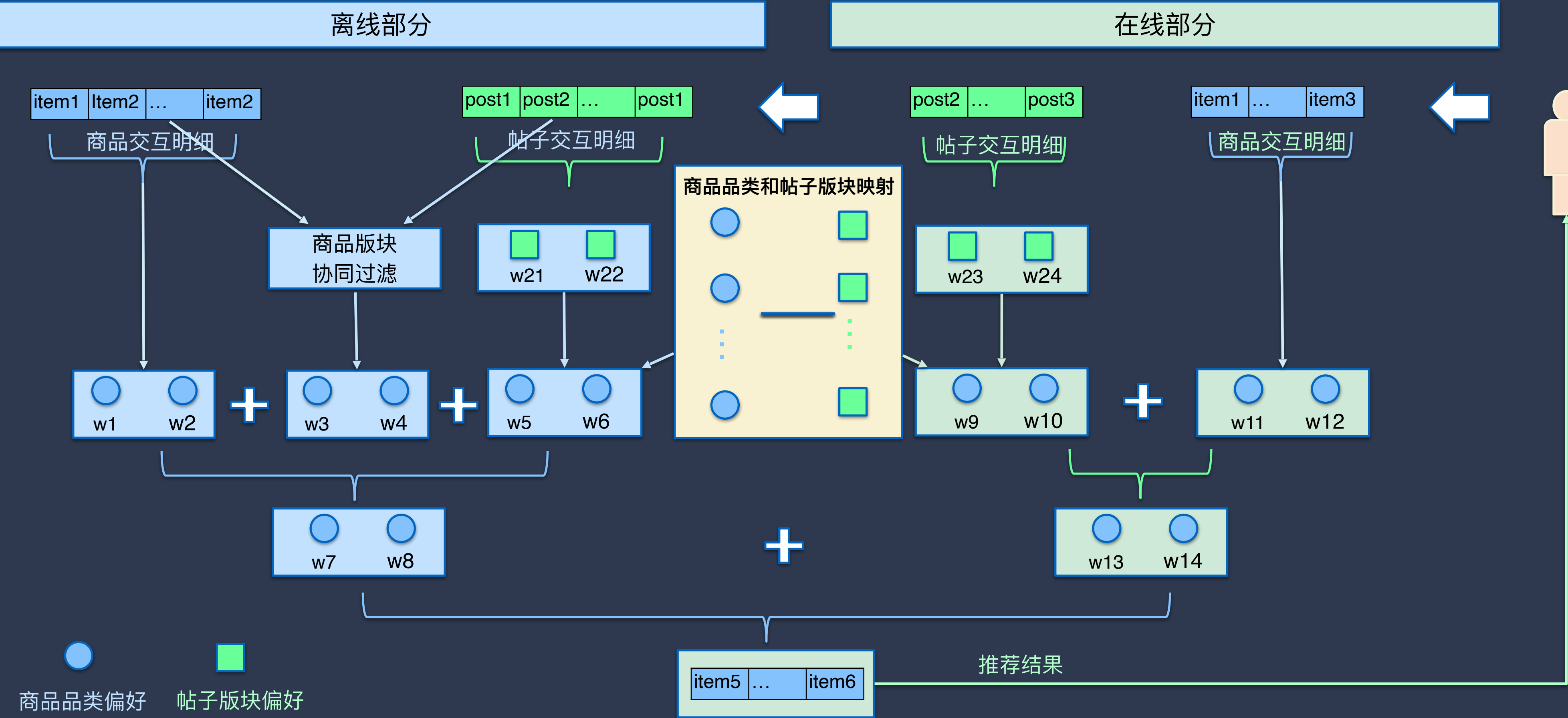
基于跨域特征的用户冷启动

基于用户群和关键词的跨领域推荐冷启动



基于跨域特征的用户冷启动

基于跨域特征的用户长短期偏好推荐



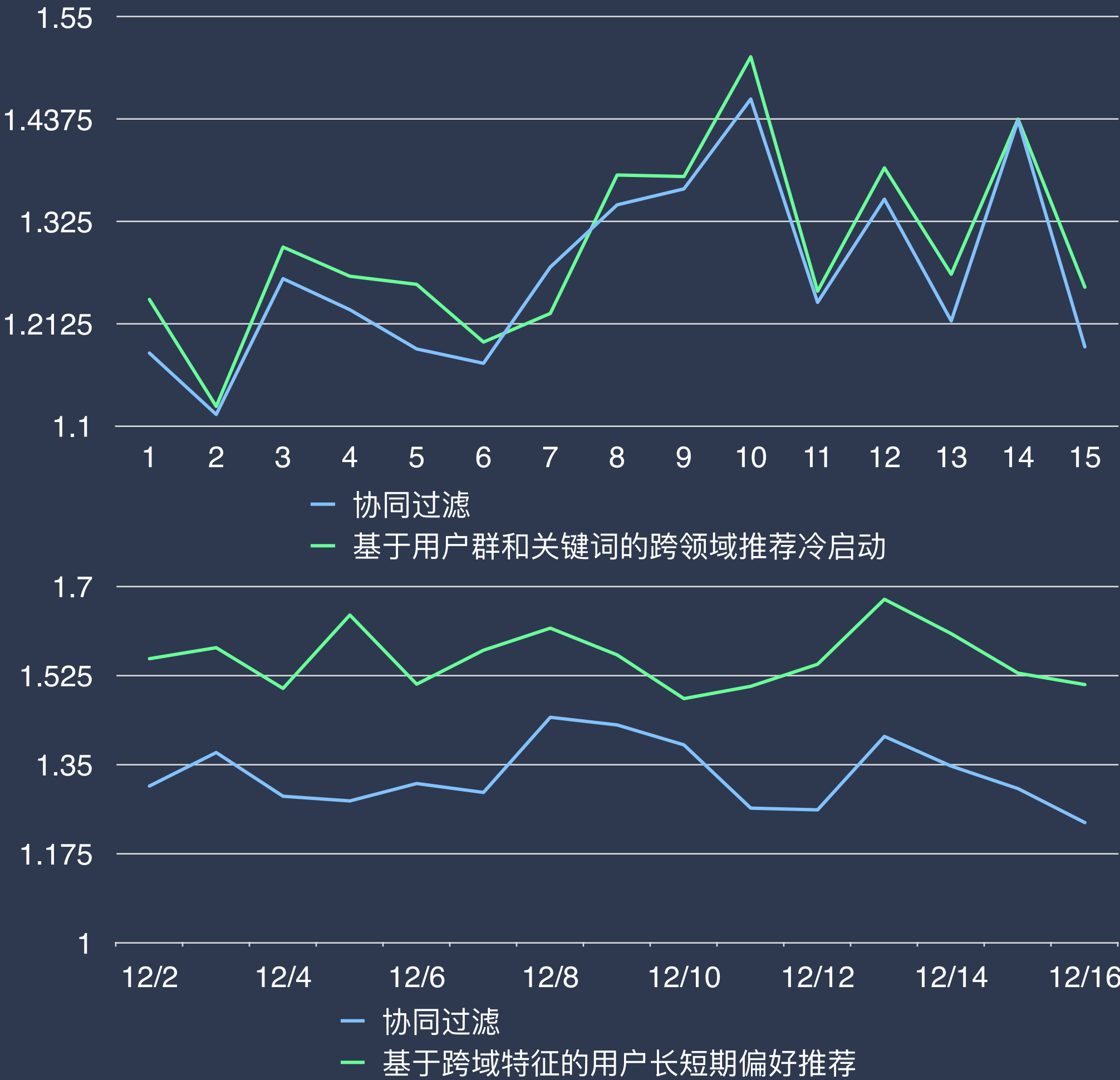
基于跨域特征的用户冷启动

基于跨域特征的用户冷启动效果分析

- CTR: 分别提升**2.3%**和**10%**
- 多样性: 曝光品类增加**3.5%**

存在的问题

- 前期跨域用户较少
- 未考虑行为序列



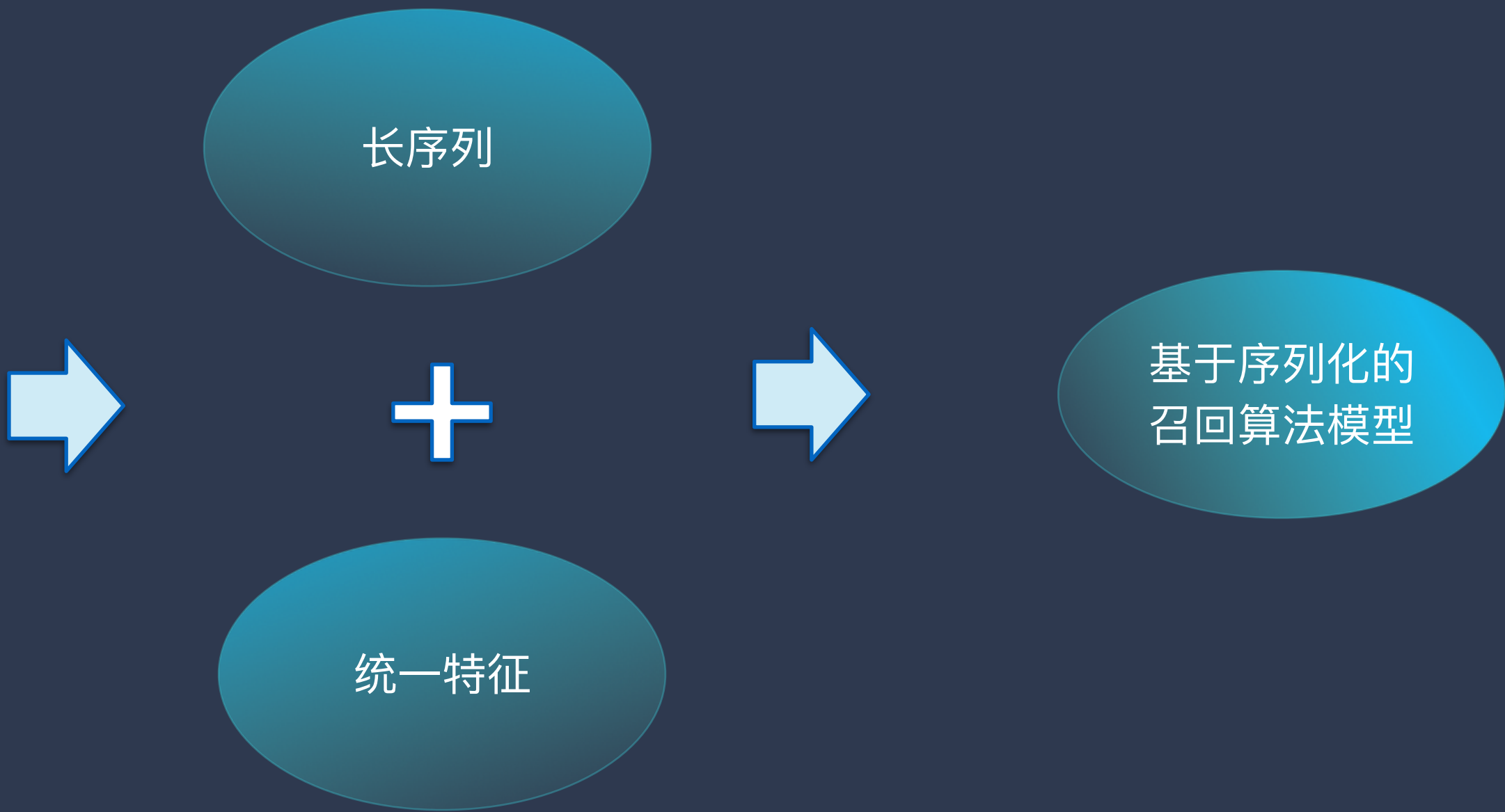
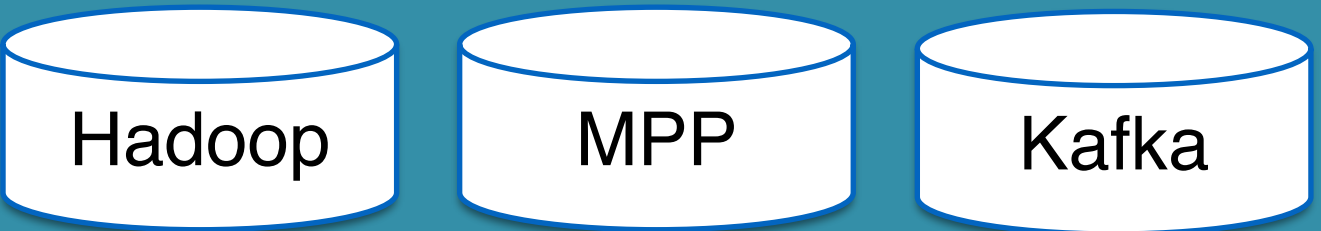
基于跨域序列化的召回模型

主要问题与挑战

- 购买考虑期：购买距离最后一次浏览该商品

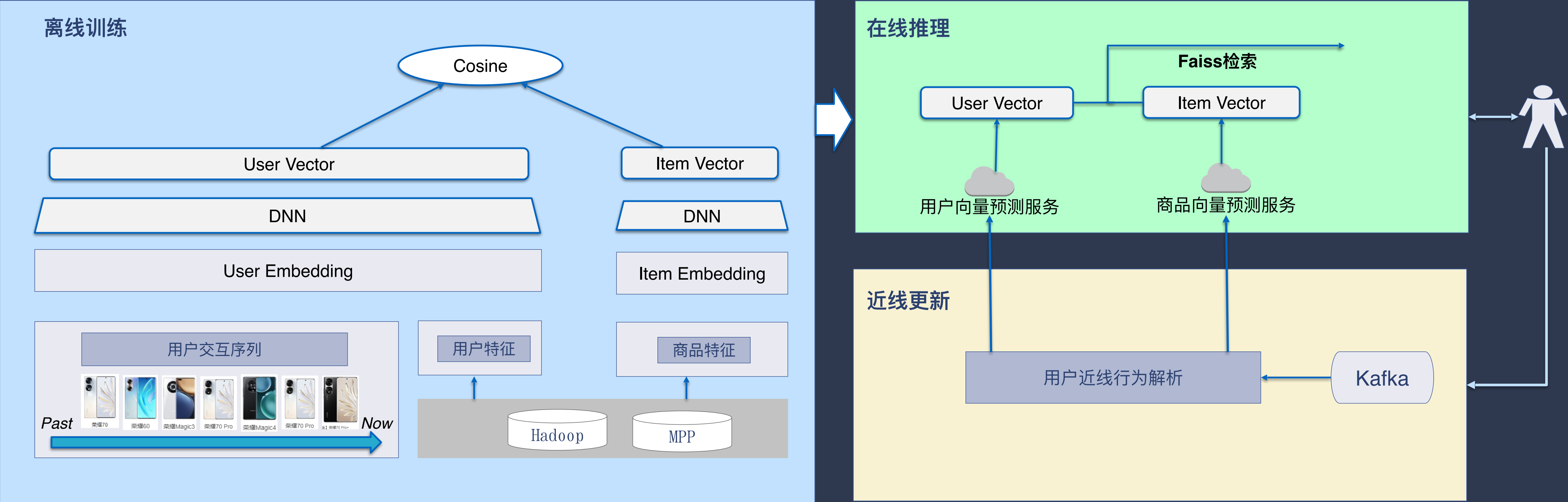


- 数据表：多且分散



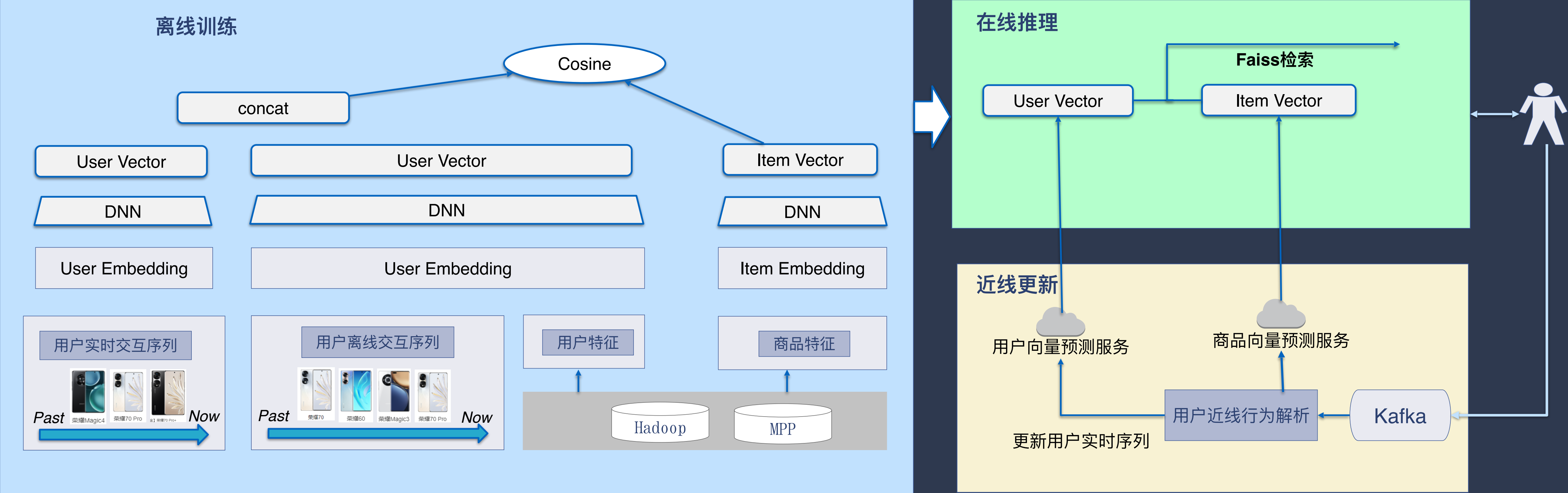
基于跨域序列化的召回模型

基于实时行为序列的双塔召回



基于跨域序列化的召回模型

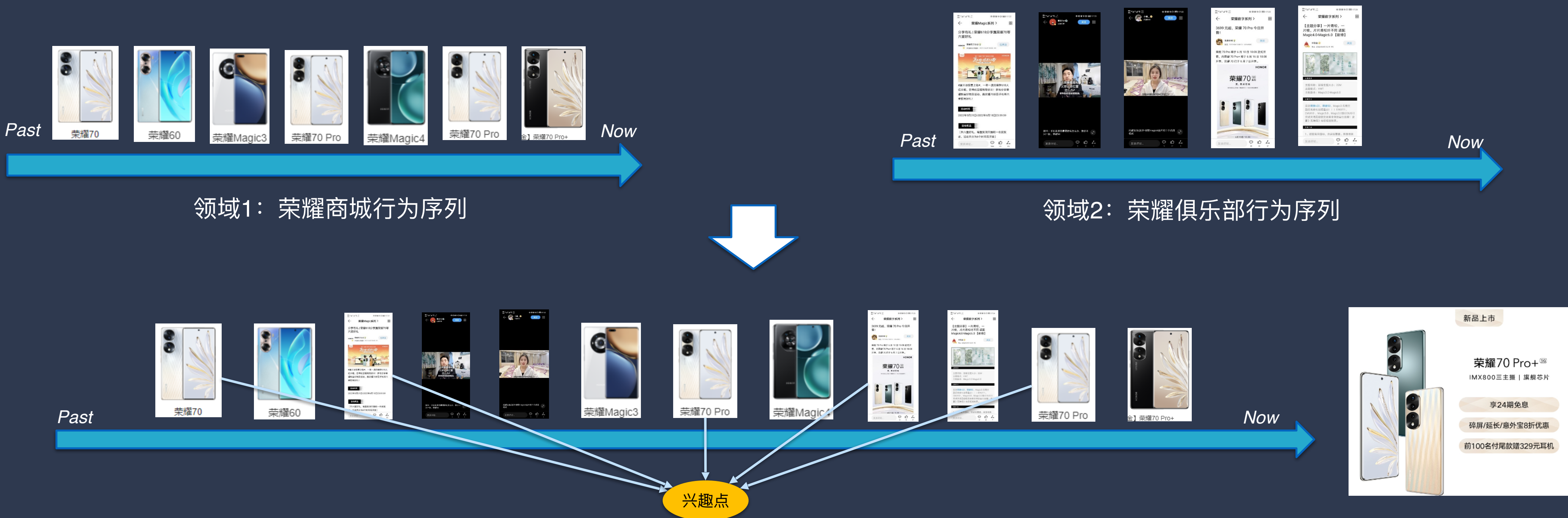
基于实时行为序列的双塔召回



基于跨域序列化的召回模型

实时双塔存在的问题

- 单次登录行为序列短：3-5个
- 多次对比不同商品：真正兴趣难以捕捉
- 关联兴趣挖掘：其他兴趣挖掘



基于跨域序列化的召回模型

基于跨域序列化的召回模型建模思路

① 行为序列表达

② 单领域行为关系建模

③ 跨域行为关系建模

④ 多兴趣建模

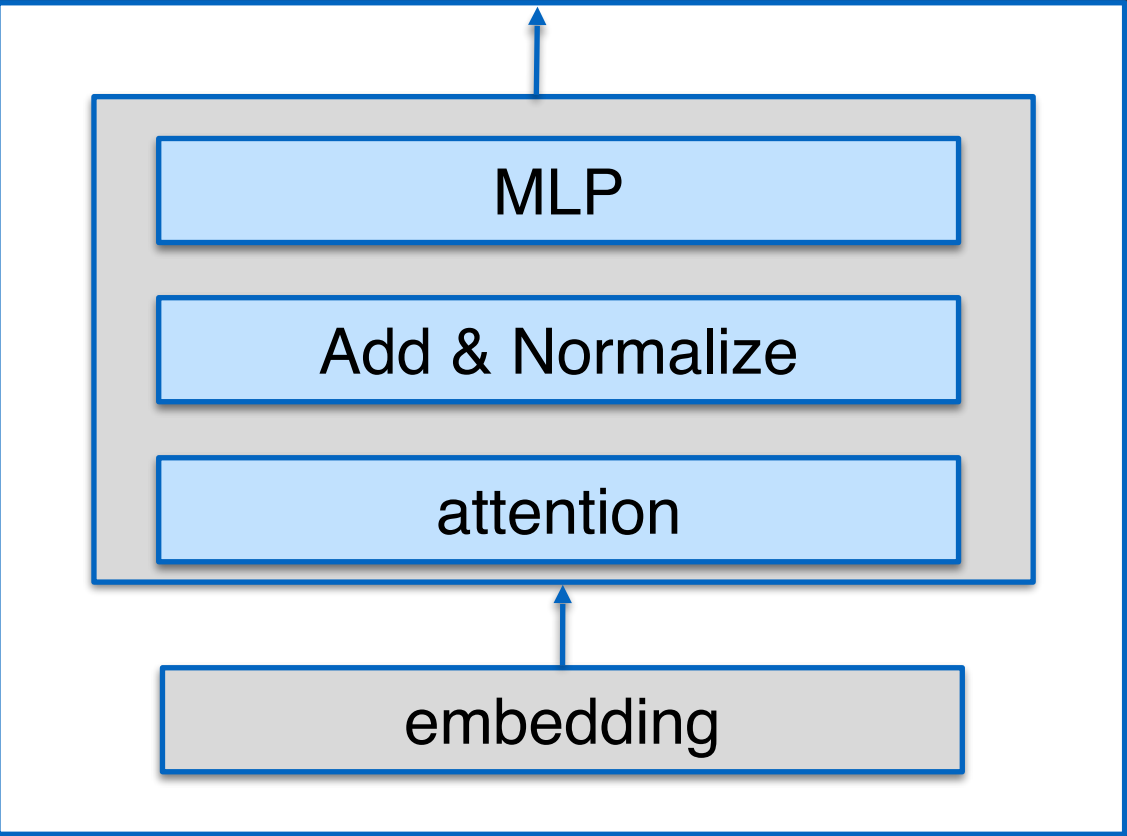
Item: 价格、上市时间、评论、行为统计

Post: 版块、发帖时间、图章、行为统计

Self attention

Multi-head attention

Target attention



基于跨域序列化的召回模型

跨域兴趣建模

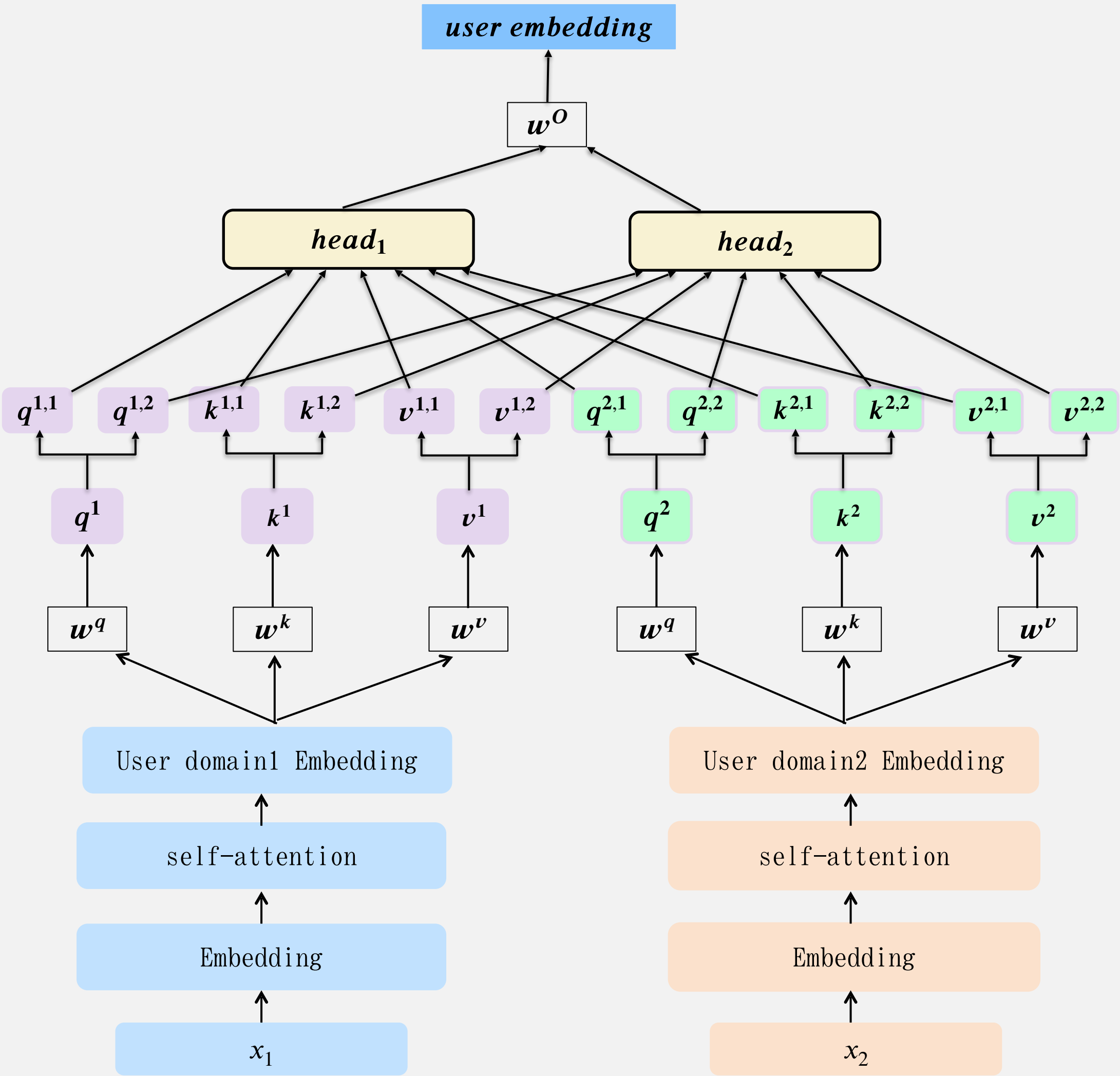
1、不同域的序列化向量表征

2、跨域的注意力机制构建

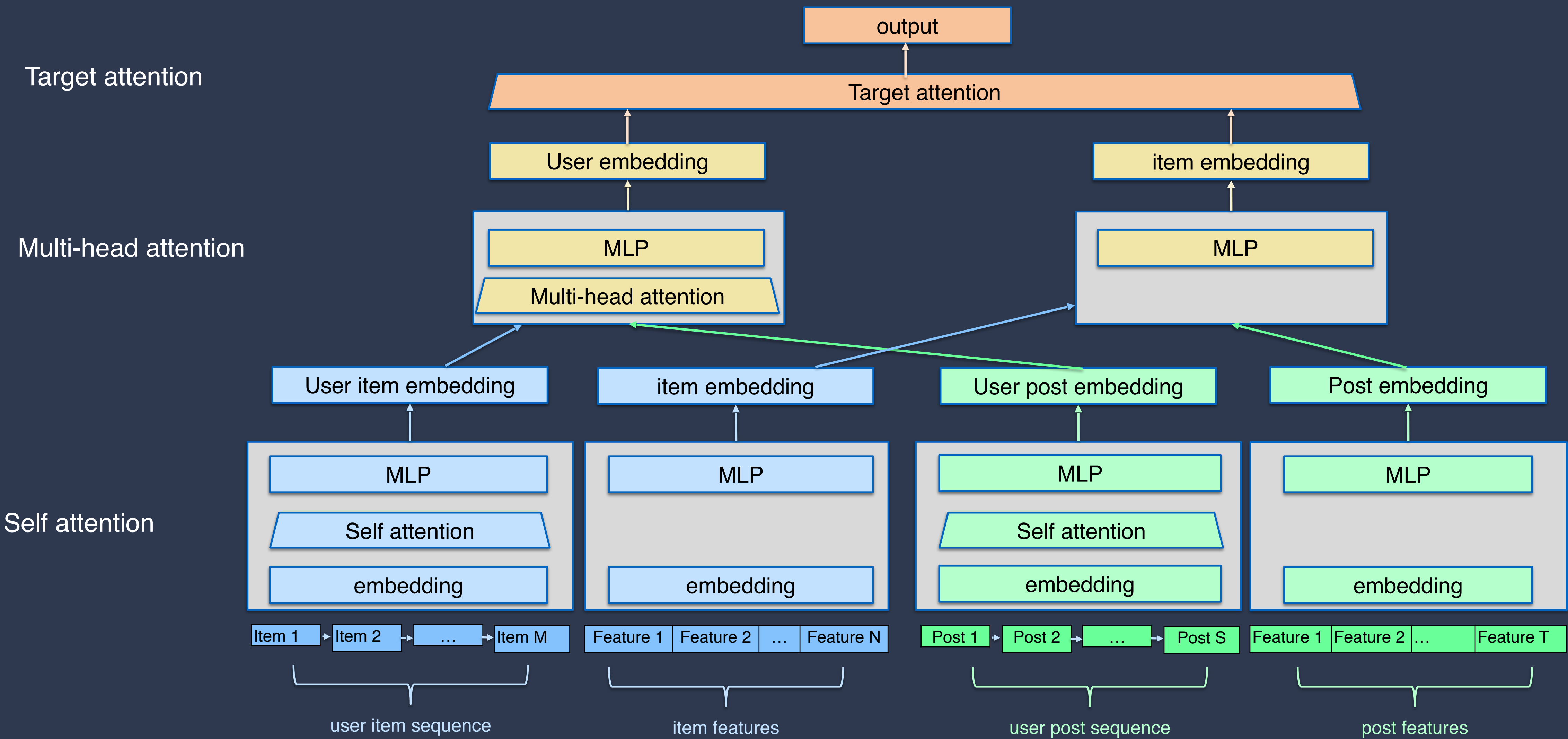
例： x_1 为商品序列， x_2 为帖子序列

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$
$$q^i = e_i w^q; k^i = e_i w^k; v^i = e_i w^v$$
$$a_{ij} = q^i k^j / \sqrt{d};$$

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h) W^o$$
$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

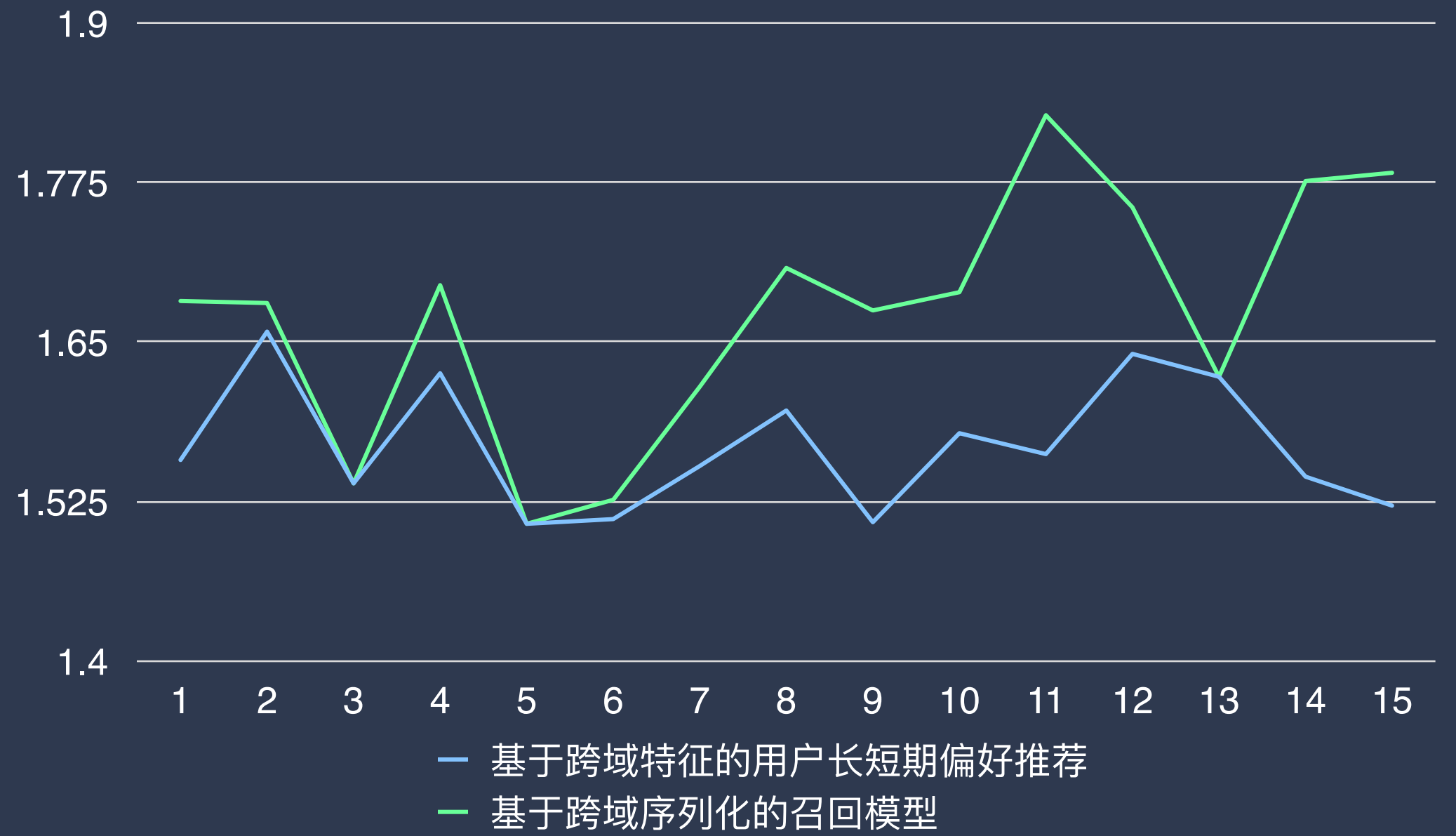
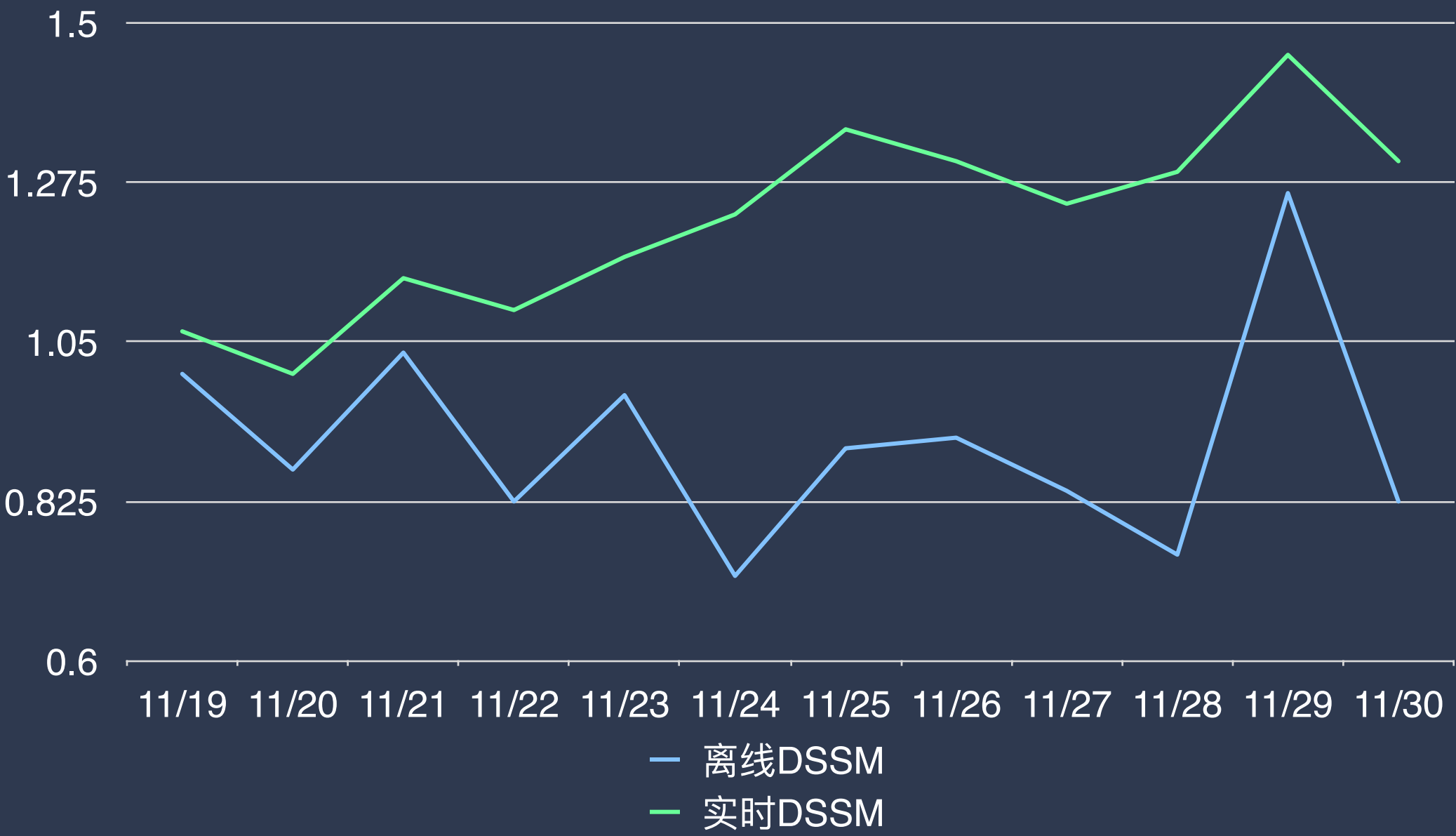


基于跨域序列化的召回模型

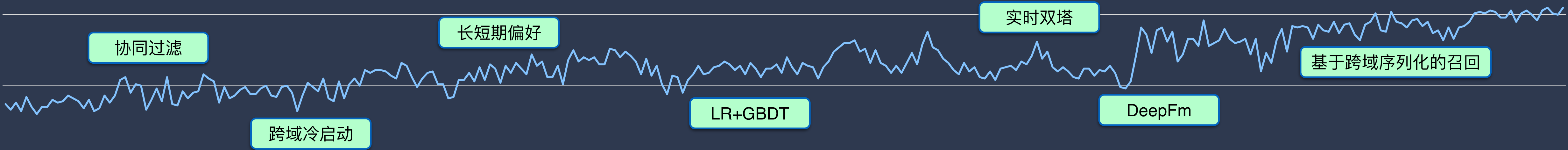


基于跨域序列化的召回模型

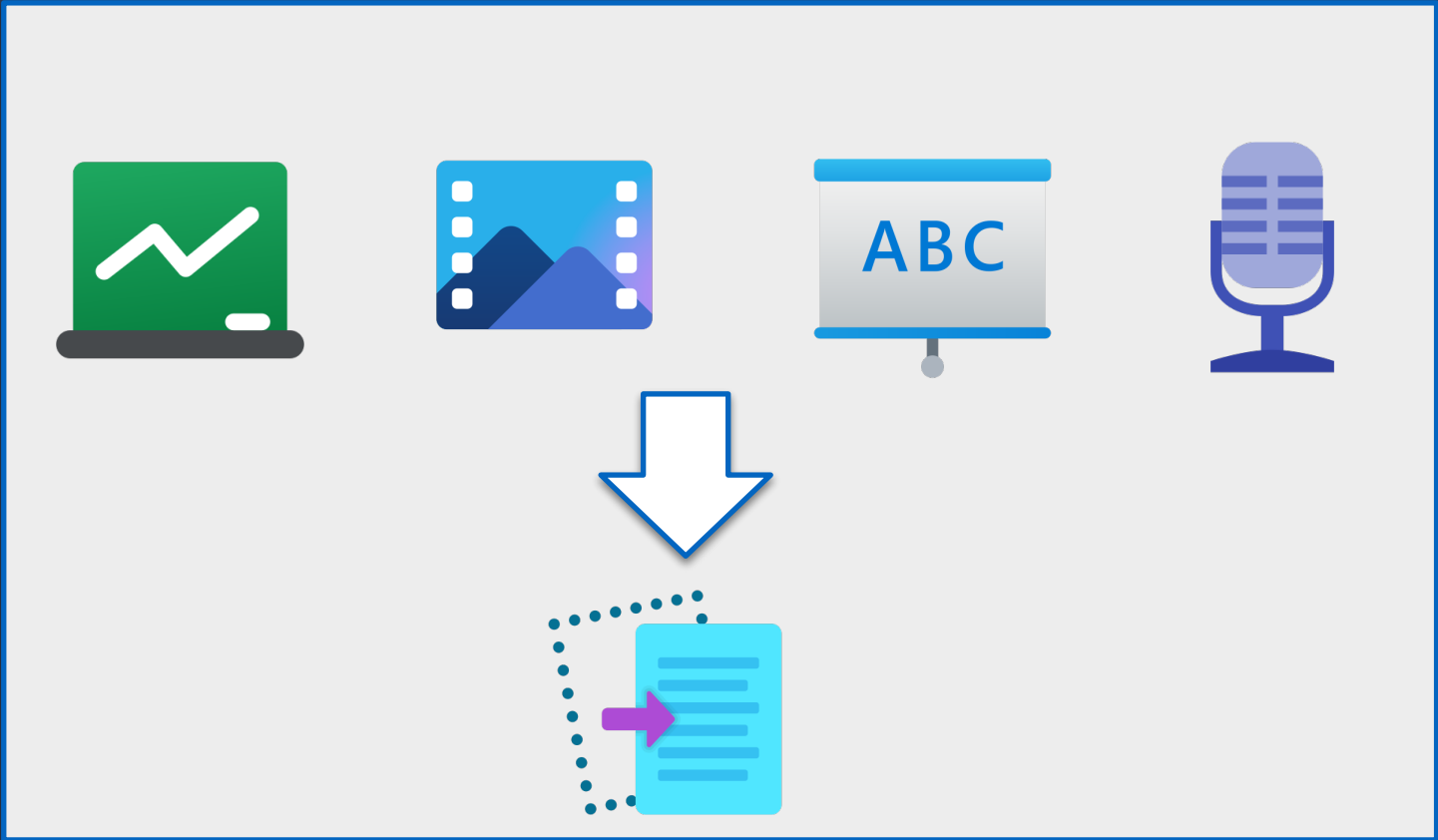
效果与分析



总结与展望



1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97 101 105 109 113 117 121 125 129 133 137 141 145 149 153 157 161 165 169 173 177 181 185 189 193 197 201 205 209 213 217 221 225 229 233 237 241 245 249 253 257 261 265 269 273 277 281 285 289 293 297



用户跨领域/多模态推荐

内容跨领域/多模态推荐

基于跨领域和多模态注意力机制的推荐

精彩继续！ 更多一线大厂前沿技术案例

📍 广州站

QCon

全球软件开发大会

时间：2022年7月31-8月1日

地点：广州·万富希尔顿酒店

扫码查看大会
详情>>



📍 北京站

GITC

全球大前端技术大会

时间：2022年8月

地点：北京·国际会议中心

扫码查看大会
详情>>



📍 北京站

QCon

全球软件开发大会

时间：2022年9月

地点：北京·国际会议中心

扫码查看大会
详情>>



想一想，我该如何把这些
技术应用在工作实践中？

THANKS