

推荐技术分享—— 多业务融合推荐实践与思考

主讲人： 王炜

部门：58同城-TEG-推荐技术部

日期：2020年4月25日

www.58.com



TEG推荐技术团队负责推荐技术基础能力建设，协同业务通过推荐提升用户体验与连接效率，目前支持场景涵盖APP首页帖子信息流推荐，58租房业务，安居客租房业务，Push消息等。

欢迎对推荐感兴趣的工程与算法同学加入，简历邮件至 luojing04@58.com (罗景)

推荐系统开发

- 1、负责推荐系统和算法服务平台等分布式系统的设计、开发、性能调优等工作；
- 2、技术预研和技术攻关，保障系统可用性、稳定性、扩展性等；
- 3、参与业务，快速支持业务策略上线，完善基础架构；
- 4、实现算法策略与系统解耦，提升系统与策略的迭代效率

推荐算法研发

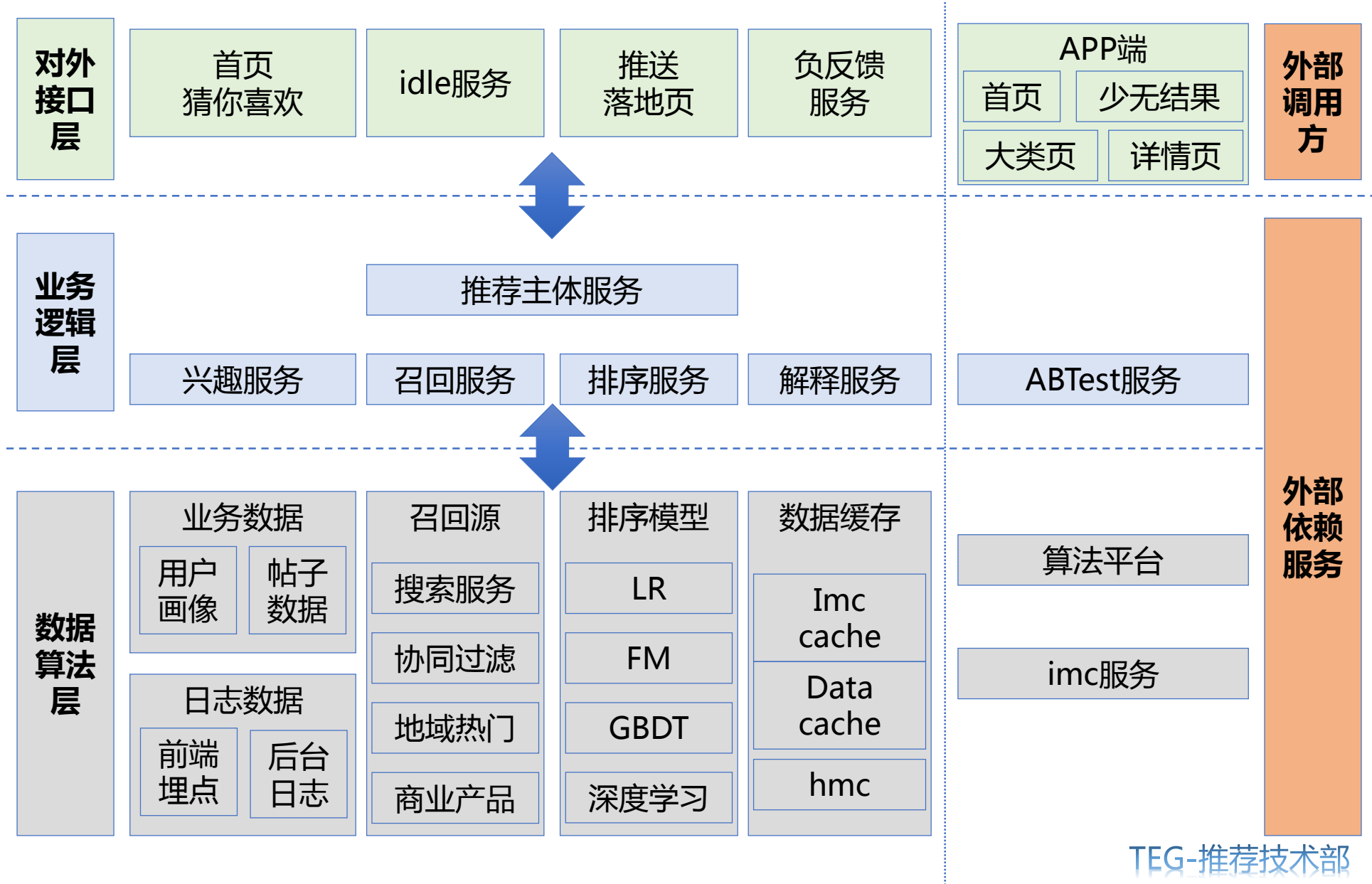
- 1、结合业务，探索预研前沿推荐算法与策略，实现在不同业务场景中的落地与优化；
- 2、针对特定业务场景构建并优化推荐策略，涵盖召回，排序以及创意优化；
- 3、通过机器学习模型解决推荐场景中的排序问题，实现点击转化效果的提升；
- 4、构建机器学习流程，对接召回与排序服务策略层面的开发

- 推荐系统整体架构
- 场景介绍
- 重排优化
 - 用户兴趣优化&重排
 - 业务流量分配机制
 - 动态刷新机制
- 总结与后续规划
- Q & A



整体架构

推荐系统-整体架构





场景介绍

58同城App首页场景

1. 首页信息流推荐场景

58同城App首页的下方
推荐tab展示

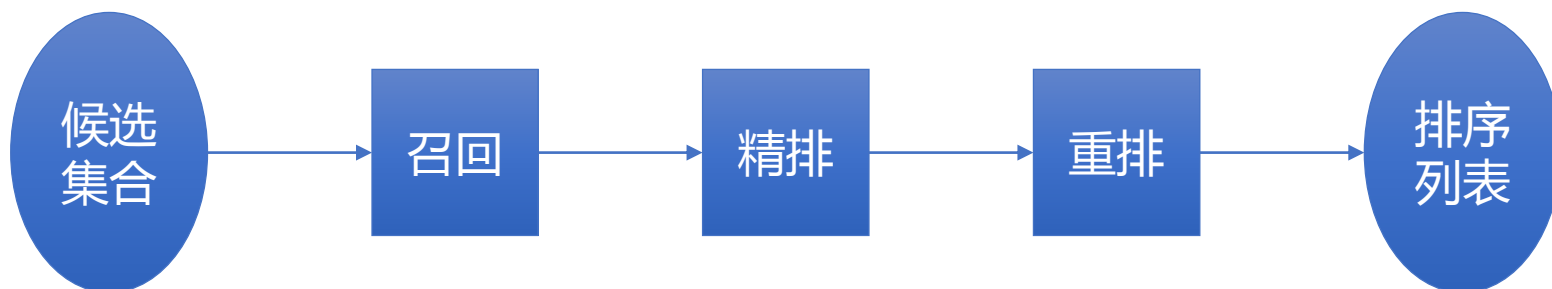
2. 特点与挑战

强兴趣

多业务

推荐感知





召回

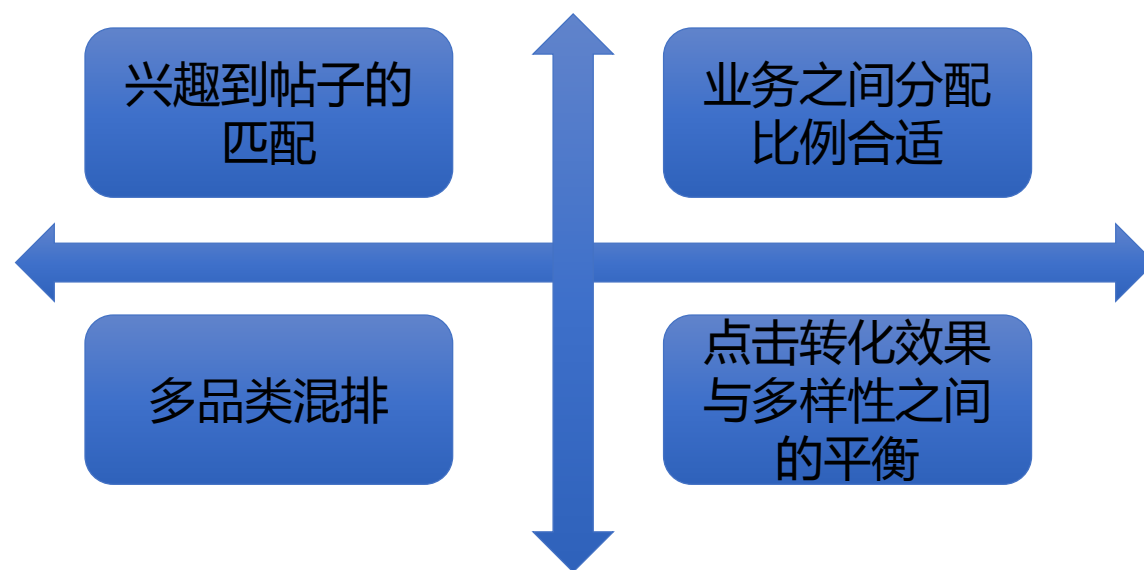
用户兴趣召回 + 向量化召回

排序

深度学习模型

首页场景推荐需求

多品类 与 单一品类 的推荐需求不同



结合业务的具体实现

- (一) 兴趣策略的优化；
- (二) 业务分配策略的优化；
- (三) 动态化刷新机制的优化。

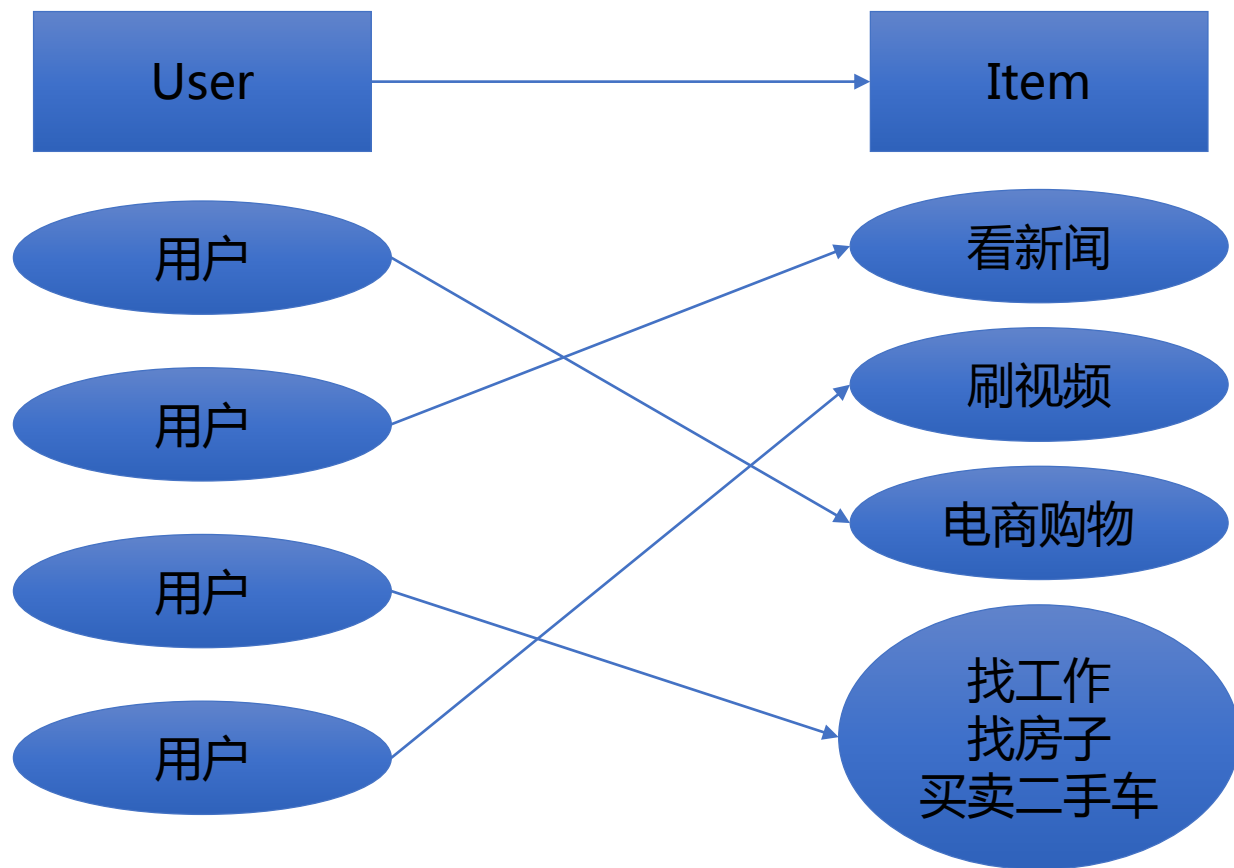


重排优化

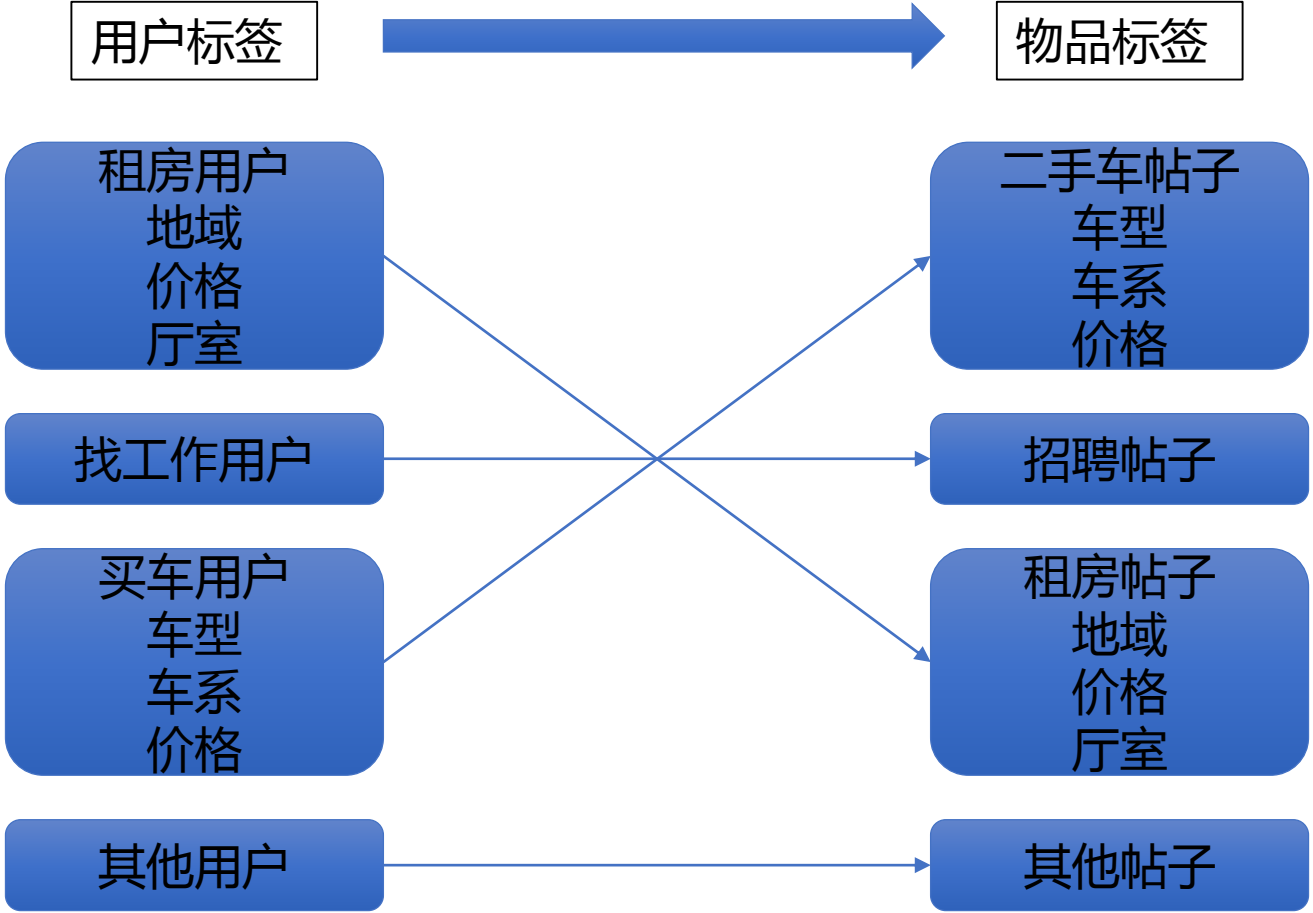
(一)

兴趣策略

业界常见的推荐系统



58首页推荐系统的主要目标



三、兴趣策略优化

1. 为什么要持续优化兴趣策略？

原始兴趣效果不理想

深入挖掘潜在信息

2. 本次优化以前，原策略在基于用户兴趣的策略上有以下问题：

实时性

兴趣分层

兴趣去噪

向量化兴趣

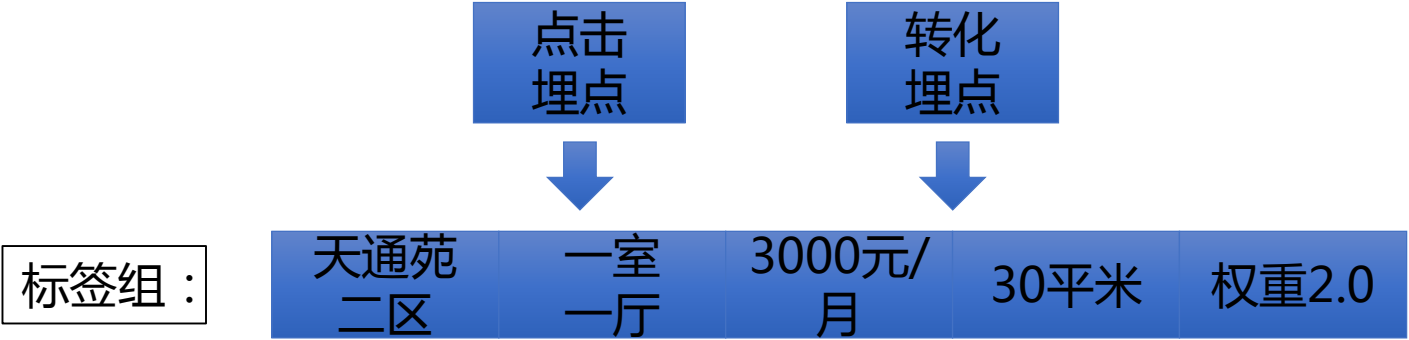
3. 推荐系统的主要子系统之一：用户画像系统

1. 目标：抽象出用户画像，或者说给用户打标签。

2. 用户兴趣的整体流程

1. 数据抽取
2. 数据清洗
3. 兴趣分类
4. 兴趣计算
5. 兴趣衰减
6. 兴趣合并
7. 兴趣聚合
8. 兴趣去噪
9. 兴趣扩展
10. 兴趣排序

1. 数据抽取



2. 数据清洗

重复埋点



无关场景



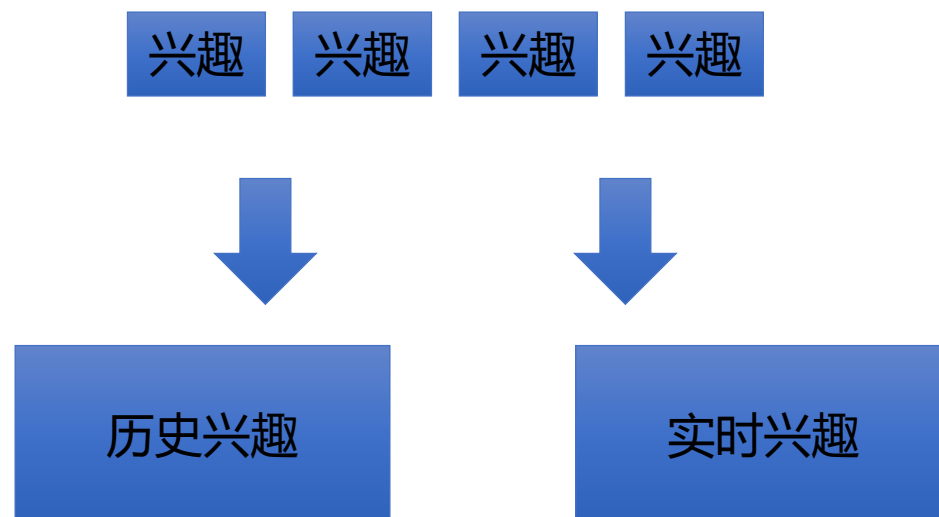
必要字段缺失



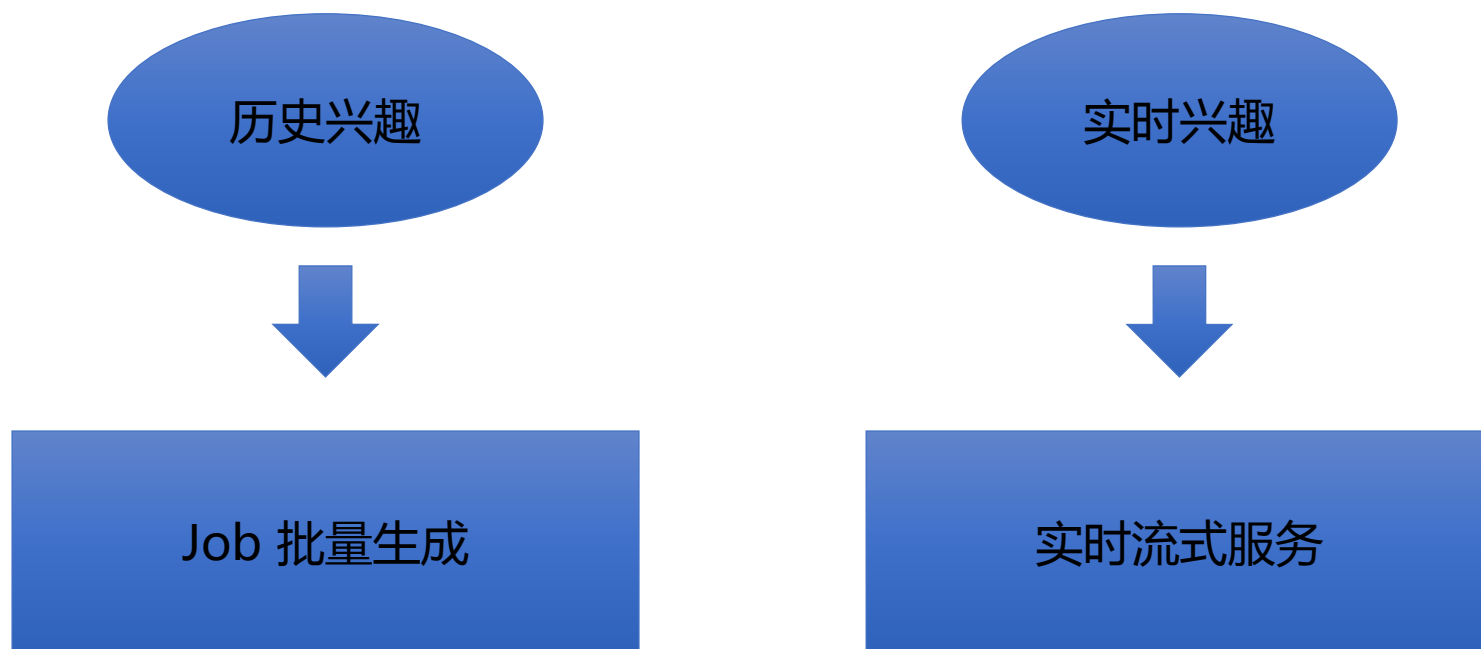
关键字段错误



3. 兴趣分类



4. 兴趣计算



5. 兴趣衰减

衰减系数 α : $\alpha = e^{-0.1*d}$, d表示兴趣距今的天数
新权重 w' : $w' = \alpha * w$

相对于不做时间衰减的兴趣策略而言：

	1天	4天	7天	10天	14天
点击效果提升	+0.7%	+1.3	+2.7%	+1.1%	+0.3%

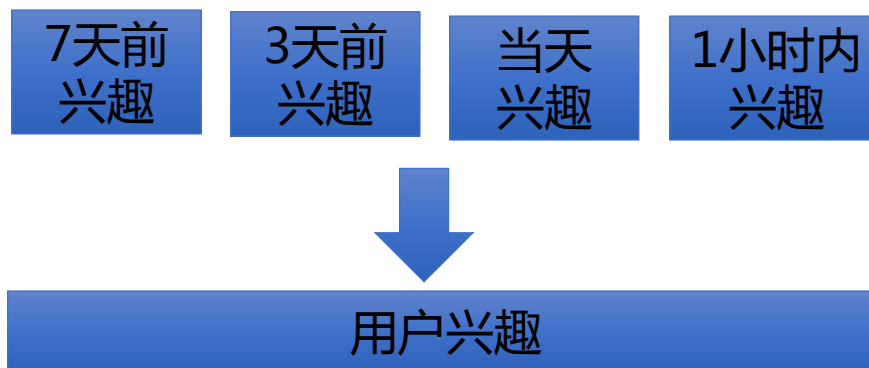
5. 兴趣衰减

其它思路：

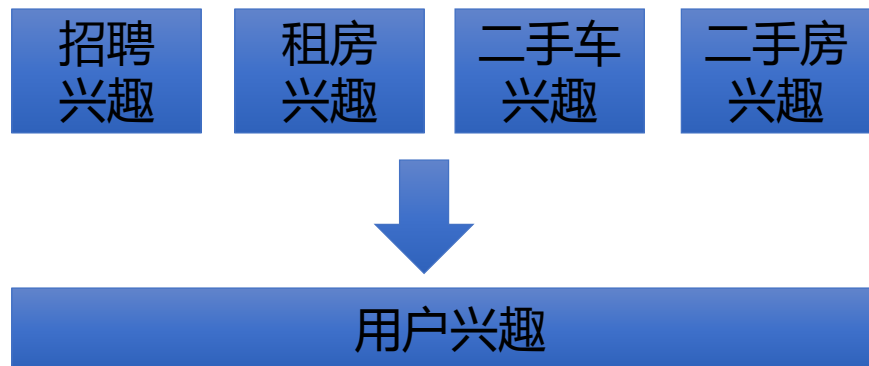
- （1）不同业务需求，衰减公式或参数不同，衰减快慢不同；
- （2）实时与历史兴趣，短期与长期兴趣，衰减快慢不同。

6. 兴趣合并

时间维度的横向合并：



品类维度的横向合并：



7. 兴趣聚合

目的：针对特定的业务，通过价格区间等条件进行聚合。

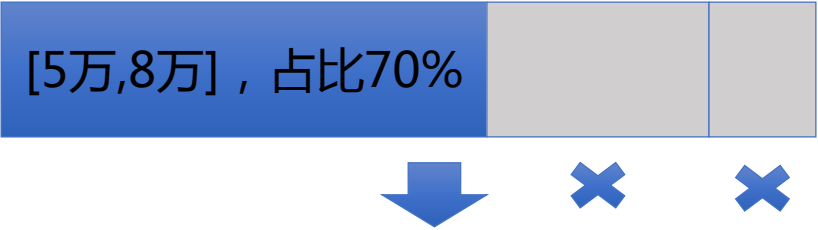
二手车价格区间：
[1万以下]，[1万,3万]，[3万,8万]，[8万,12万]。。。[40万以上]

用户兴趣价格区间组数占比统计表：

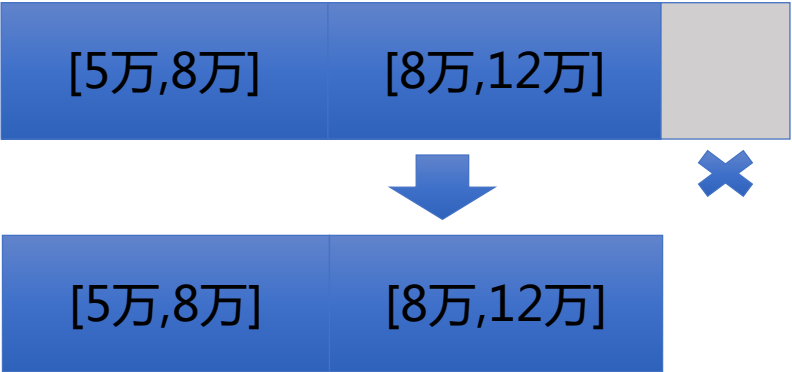
	1组	2组	3组	4组及以上
占比	62%	21%	10%	7%

7. 兴趣聚合

优势组直选策略：



长尾组淘汰策略：

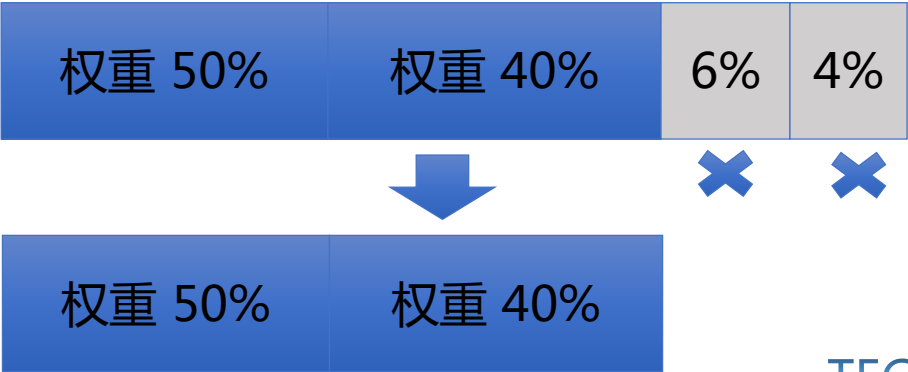


8. 兴趣去噪

前5大权重累加和占比的平均值统计表：

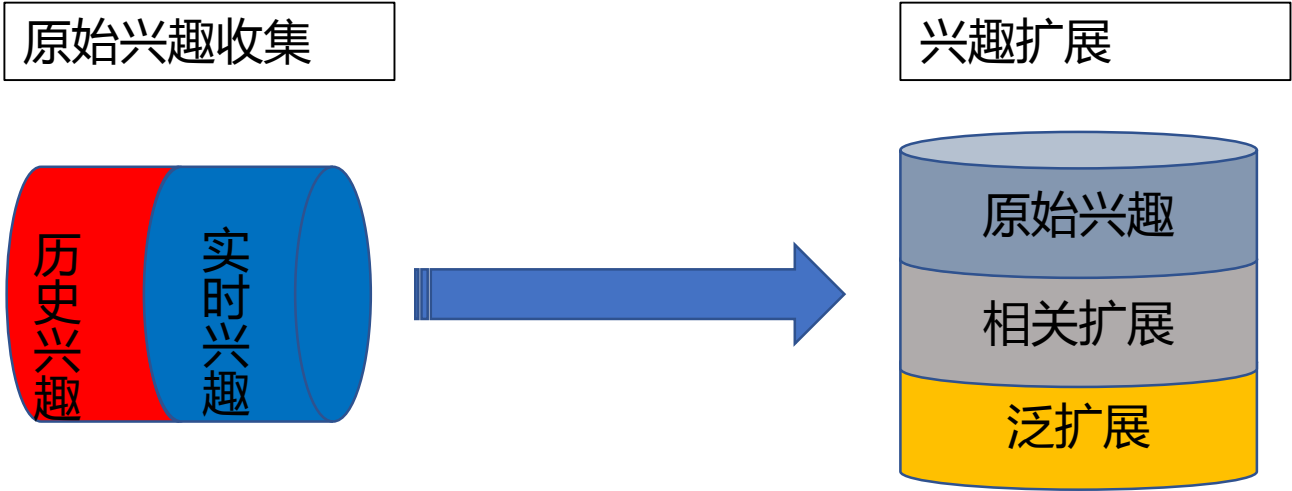
累加至	第1大	第2大	第3大	第4大	第5大
	42%	57%	69%	79%	87%

长尾淘汰策略：



9. 兴趣扩展

目的：扩大召回来源，但效果变差。



	原始兴趣	相关扩展	泛扩展
点击效果 相对于 原始兴趣	N/A	-2.2%	-16%

9. 兴趣扩展

品类内部标签的纵向扩展

天通苑 二区	一室 一厅	3000元/ 月	30平米
-----------	----------	-------------	------



天通苑周 边5km小 区	一室一厅 或大一居	2000~ 4000 元/月	20~40 平米
--------------------	--------------	----------------------	-------------

跨品类的迁移扩展

整租



整租	合租	品牌公寓
----	----	------

招聘品类扩展，从最小类目扩展到二级类目，点击效果提升+2.1%。

10. 兴趣排序

对最终选取的兴趣按照权重排序

权重大的兴趣在召回、过滤、排序、重排、分配流量等环节上优先考虑。

重排优化

(二)

业务流量分配策略

1. 为什么要持续优化流量分配策略？

首页推荐哪些业务线的帖子需要为分流这个重要目的服务。

2. 本次优化以前，原策略对各业务的分配策略有以下问题：

（1）品类之间平均分配，未考虑权重；

（2）大品类与小品类之间的冲突、重叠。

3. 原始策略

原始策略一：

直接采用精排结果序列。

原始策略二：

假设58平台流量在大的品类上，占比如下：

招聘：租房：二手车：本地服务 = 40：30：20：10。

每个用户都按上面的比例分配。

4. 新策略

（1）用户兴趣的权重分配

招聘，60%	租房，20%	黄页，20%
--------	--------	--------

（2）召回帖子的权重分配

招聘，65%	租房，25%	黄页 10%
--------	--------	-----------

（3）帖子数量不足的品类通过其它品类补充。

招聘，65%	租房，25%	黄页 10%	其它
--------	--------	-----------	----

4. 新策略

（4）推荐业务帖的流量分配比例与58平台整体流量占比基本一致。

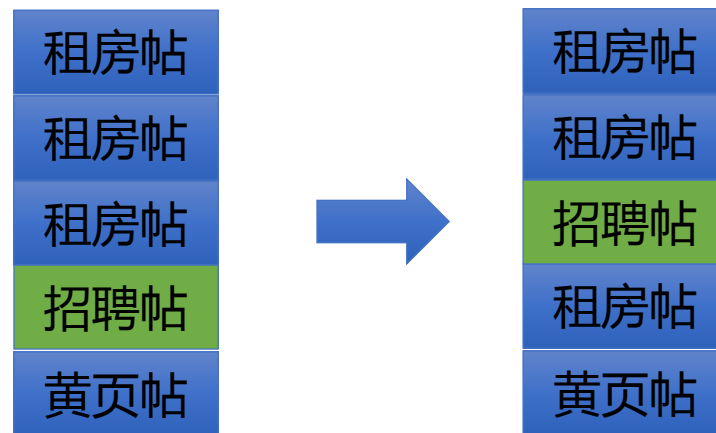
采用按兴趣权重分配流量的策略，点击效果提升+1.1%。

招聘，50%	租房，20%	车辆，20%	黄页，10%
--------	--------	--------	--------

4. 新策略

（5）品类之间打散策略

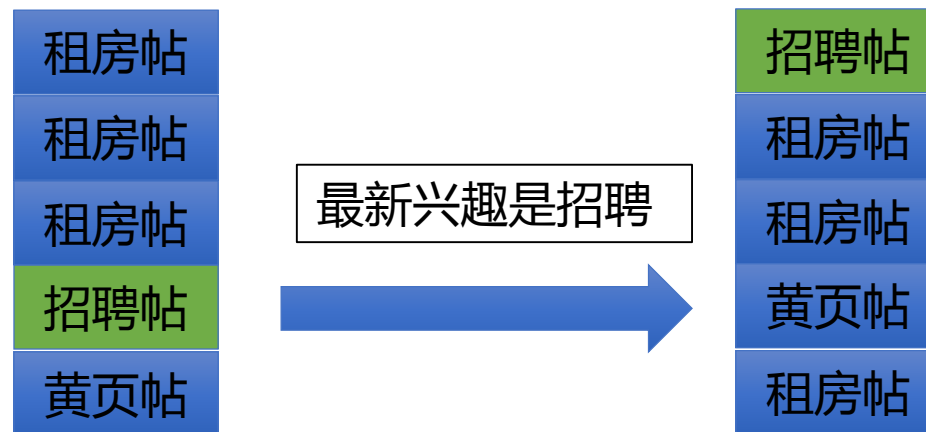
打散策略，点击效果提升+0.4%。



4. 新策略

（6）最新兴趣直选头条

该策略强调了实时兴趣，点击效果提升+0.7%。



5. 整体优化效果

综合兴趣策略和业务分配策略的整体优化效果，
上线当周，点击效果比基准提高+11.0%。

重排优化

(三)

动态刷新机制

1. 为什么需要动态化刷新机制？

目的：解决内容多样性问题。

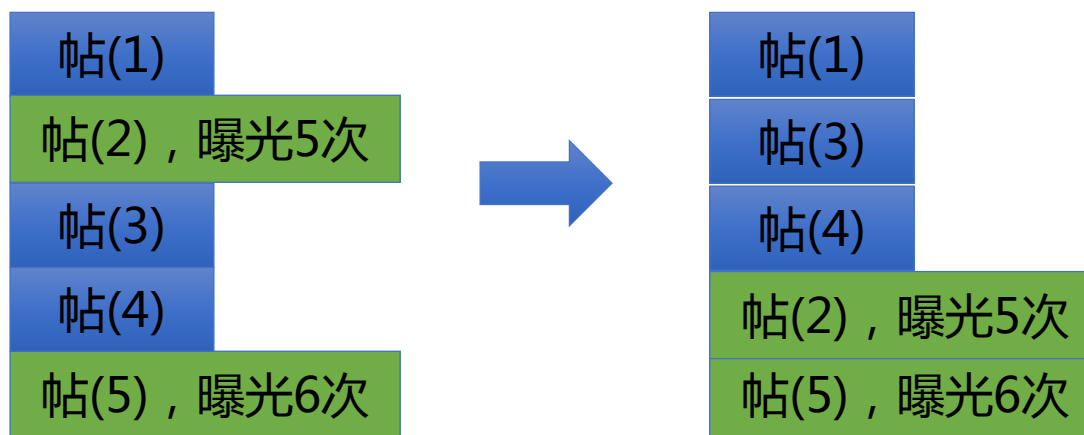
2. 优化思路：

可以通过协同过滤、向量化召回等多路召回的办法增加多样性。

但在召回集合基本不变的前提下，我们优先考虑在集合内部的排序上进行动态调整。

3. 迭代路径：

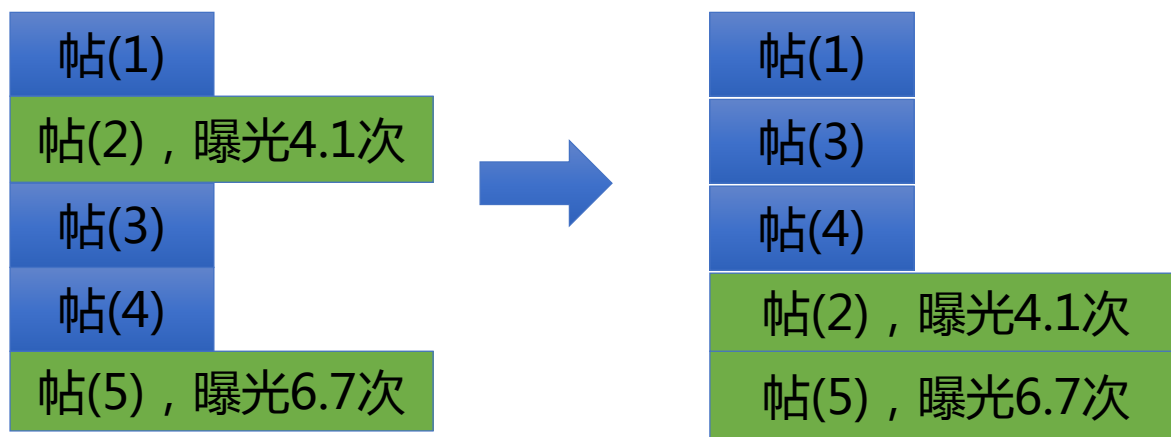
- (1) 曝光次数较多的帖子，重排时后移。
点击效果比未动态化基准提升+4.1%。



3. 迭代路径：

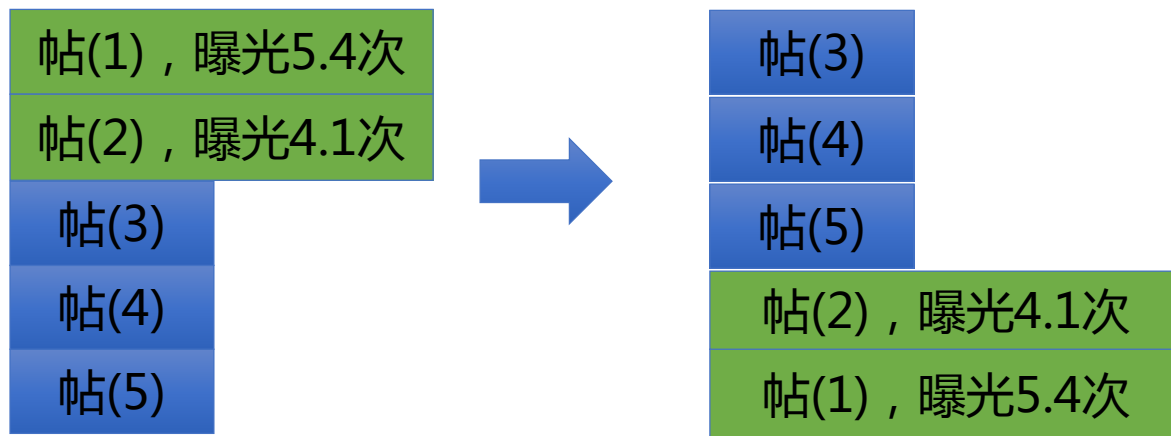
(2) 引入时间衰减。

点击效果比未动态化基准累计提升+6.5%。



3. 迭代路径：

(3) 引入服务端的曝光日志。
点击效果比未动态化基准累计提升+7.0%。



4. 整体效果：

（1）以上3步迭代的整体效果：点击率比基准提升+7.0%。

（2）同一用户最新访问的帖子集合较前一次访问的集合的动态变化比率全天平均达到60%。

5. 近期规划：

（1）引入全站的曝光日志，进一步优化刷新机制。

（2）引入用户的点击日志，通过再营销等策略，进一步优化效果。

重排优化 小结

简单、直观

整体效果可观

供相似场景的同学参考



总结



总结

1. 关键是理解业务。
2. 优化策略之间的冲突、叠加。
3. 多目标之间的平衡。
4. 算法与策略之间的关系。



后续规划

后续规划

1. 动态化刷新机制的进一步优化
2. 首页推荐理由的进一步优化
3. 冷启动的进一步优化
4. 基于feed流机制的访问模式
5. 深度学习模型的持续优化

Q & A

1. 冷启动的优化

(1) 全新用户：

全局热门、地域热门、兴趣试探、新内容试探。

(2) 兴趣不足：

兴趣迁移、标签扩展、召回扩展。

2. 负反馈的优化

(1) 直接屏蔽：物品、兴趣

(2) 厌倦惩罚：迭代兴趣标签、权重。

(3) 独立特征，影响排序。

3. 多样性的优化

目标：减少马太效应，兼顾效果。

(1) 协同过滤等多召回通道；

(2) 全局或个性化的热门内容降权；

(3) 兴趣试探、新内容试探。

Thank you!

—
