

首页推荐策略v1.2版本

1. 最近一次思路转变

首页整体千人千面：

首页除去顶部爆款，全部修改为推荐区域。

召回：运营配置+ “推荐相关度”高的商品（推荐召回范围比较窄）

排序：首页商品按照 “业务数据（销量等）+推荐相关度” 综合排序

推荐模块：

首页爆款和运营单品中间增加推荐模块。

召回：召回和用户有相关性的商品（推荐召回范围非常宽泛）

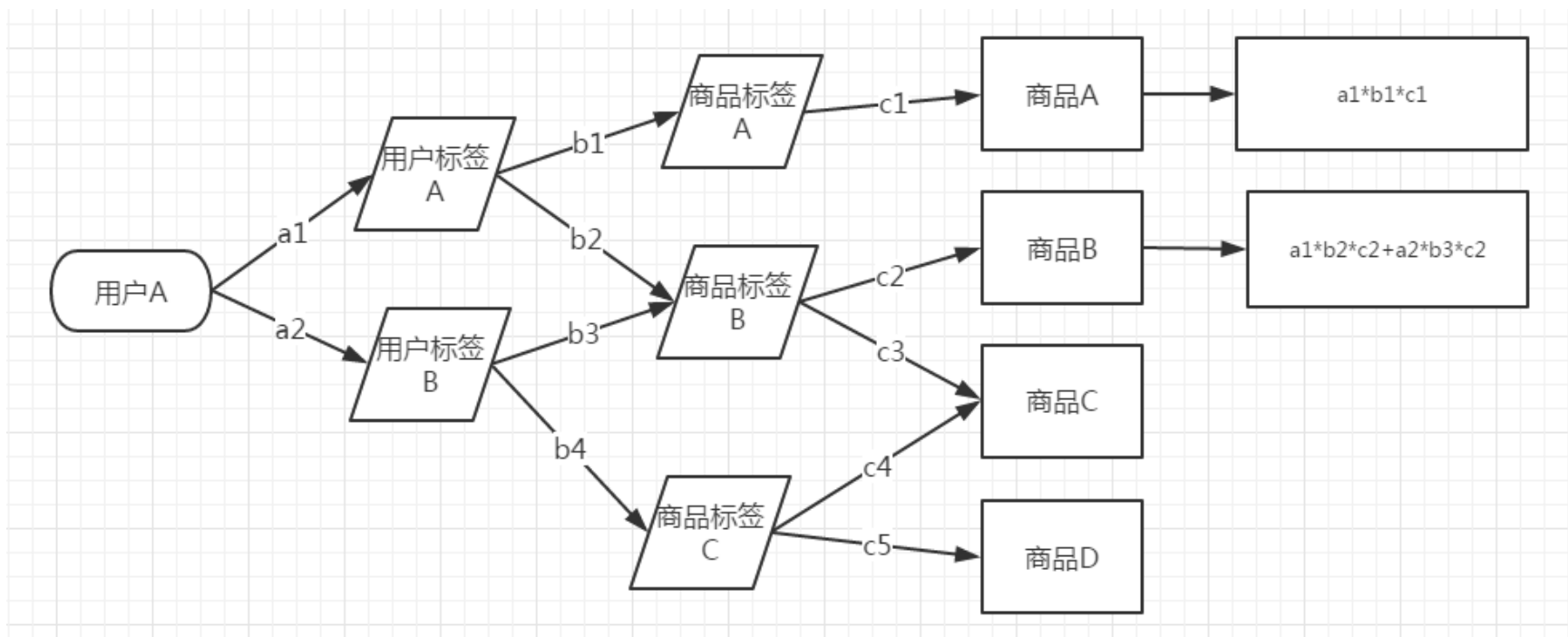
排序：排序业务数据+推荐相关度

- 精准推荐
- 销量不影响召回
- 运营推荐和挖掘长尾



- 宽泛推荐
- 销量影响召回
- 兼顾销量和相关度

2. 推荐实现路径>>理想状态模型



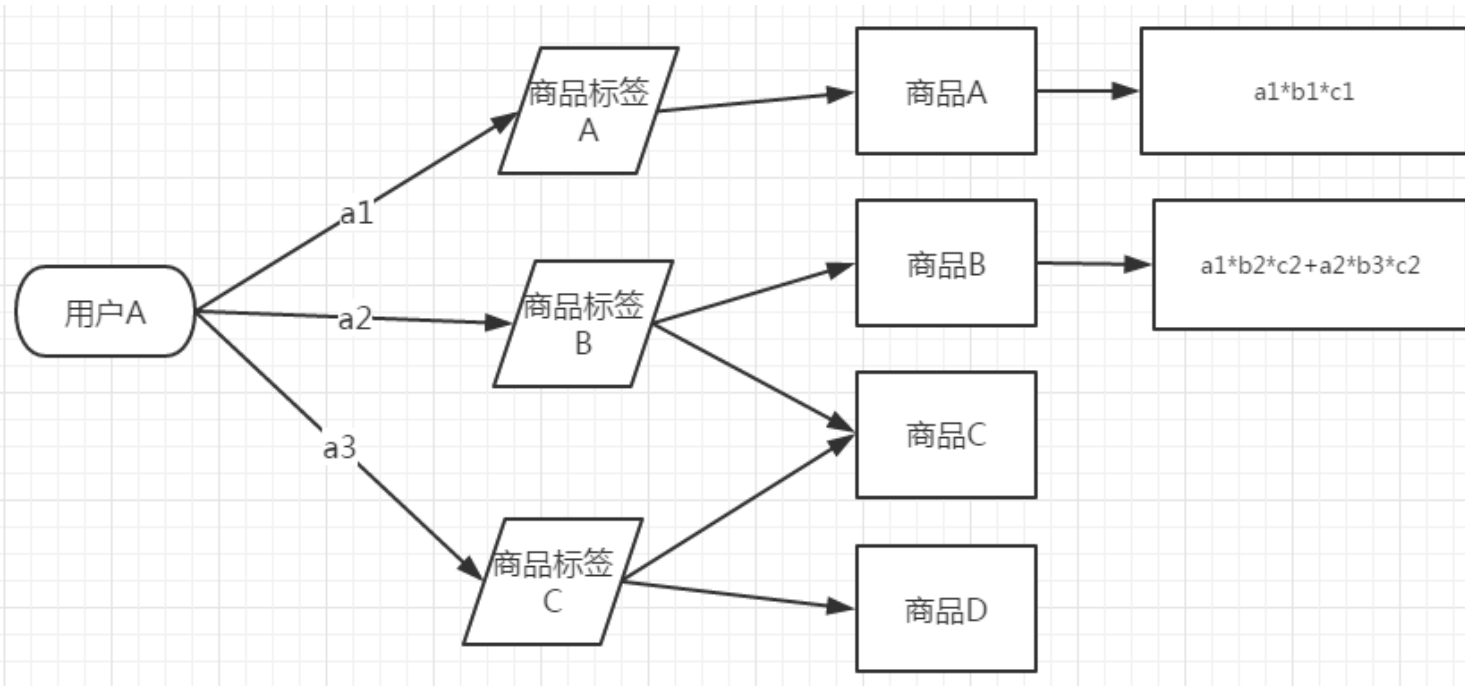
- 机器学习的聚类
- 多维度用户标签挖掘
- 用户标签不等于商品标签
- 标签之间有关联策略
- 考虑时间衰减因素
- 考虑及时响应策略

3. 推荐实现路径中间演化版本—> 基于商品标签的推荐

1. 机器学习的聚类，改为人工理解的聚类
2. 用户标签挖掘：信息来自主动收集和商品数据
3. 用户标签完全等同于商品标签
4. 标签之间没有关联度
5. 时间衰减因素未考虑
6. 未考虑及时响应策略

Item based CF&基于标签的推荐

因为用户和某些商品发生了关系，基于商品标签，推荐了很多有相同标签的商品



4. 数据来源&标签类型

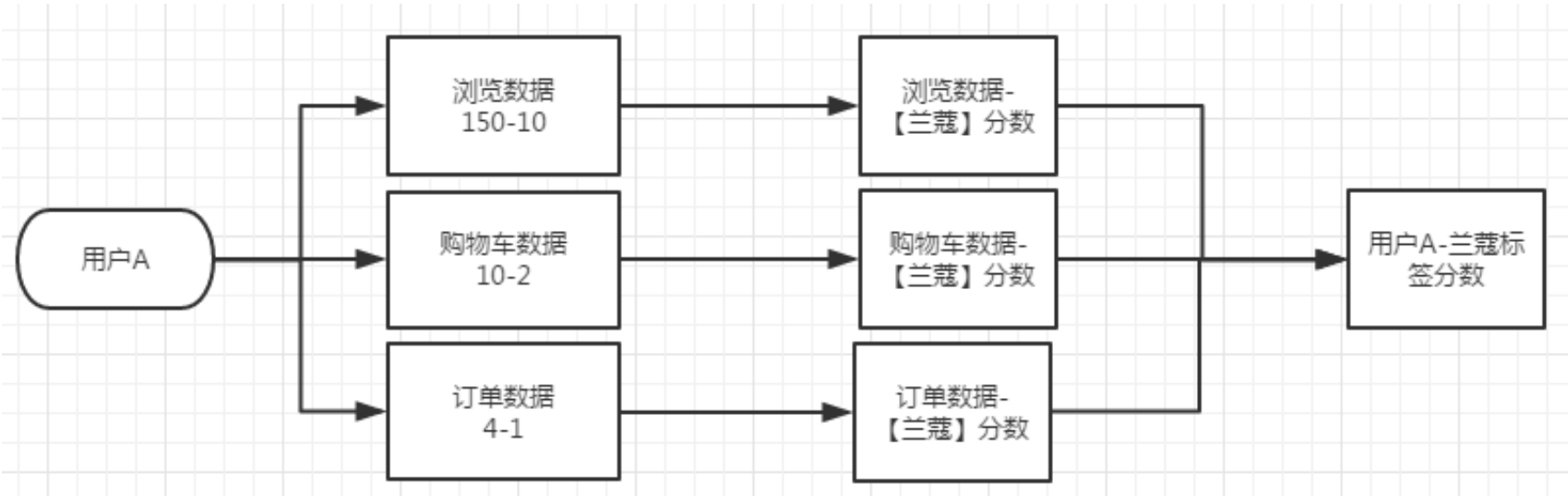
数据来源：

- 【P0】浏览：对最近90天浏览数据，映射为标签。
- 【P0】订单：对最近一年的订单，分析商品中的标签。
- 【P0】引导：主动收集用户数据，映射为标签。
- 【P1】购物车：对最近90天加入购物车的商品，分析商品标签。

标签类型：

- **普通标签【召回，排序】**（如：兰蔻。一个标签关联有限商品，一个标签标签关联部分用户）
商品四个级别的分类；品牌；商品产地。
- **特殊指标【排序】**（一个标签关联全部用户和商品但是分数不同）
消费指数：用户消费指数匹配商品消费指数
- **业务数据：商品数据销量、流量数据**
- **待讨论：品牌调性**作为特殊指标还是普通标签？有哪些可增加的标签？

5. 普通标签打分方案



有特定标签商品数据，占总数据的比例

$$s_{ij} = \sqrt{\frac{u_{ij}}{n_j}} \cdot \lg(n_j + 1)$$

数据量越大，分数越高

标签业务权重

$$s_i = \frac{\alpha_i}{\sqrt{\sum \gamma_j^2}} \cdot \frac{1}{\lg(X_i + 1)} \sum_j \gamma_j \cdot s_{ij}$$

加权求和的归一化参数

各种类型数据加权求和

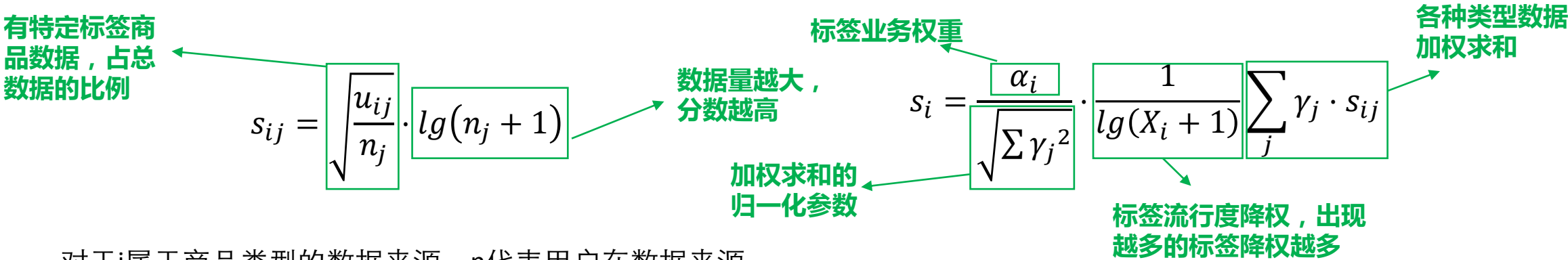
标签流行度降权，出现越多的标签降权越多

对于j属于商品类型的数据来源，n代表用户在数据来源类型j中的数据样本量， $\lg(n_j + 1)$ 为数据可靠性提权， u_{ij} 为在数据来源类型j中、命中标签i的商品数。 γ_j 代表数据类型j的权重。

X_i 代表某个标签命中的用户数， $\sqrt{\sum \gamma_j^2}$ 为归一化参数， $\lg(X_i + 1)$ 表示对流行度高的标签进行降权， α_i 代表对于标签i给出的业务权重。

6. 普通标签打分方案>>算例

数据类型	类型权重(γ_j)	数据量(n_j)	兰蔻商品数(u_{ij})	特定数据标签得分 (Sij)	分数
浏览数据	1	150	10	0.562	
购物车数据	5	10	2	0.447	
订单数据	2	4	1	0.301	
兰蔻标签用户数为10000 业务权重为5				求和分数($\sum_j \gamma_j \cdot s_{ij}$)	1.757
				综合分数	0.401



对于j属于商品类型的数据来源，n代表用户在数据来源类型j中的数据样本量， $\lg(n_j + 1)$ 为数据可靠性提权， u_{ij} 为在数据来源类型j中、命中标签i的商品数。 γ_j 代表数据类型j的权重。

X_i 代表某个标签命中的用户数， $\sqrt{\sum \gamma_j^2}$ 为归一化参数， $\lg(X_i + 1)$ 表示对流行度高的标签进行降权， α_i 代表对于标签i给出的业务权重。

7. 消费指数计算

订单内商品	所属品牌商品平均价 ($\overline{p_B}$)	所属分类的平均价 ($\overline{p_C}$)	所属该品牌品类的平均价 ($\overline{p_{C \cap B}}$)	商品价格 (p_i)	商品消费指数 a_i	
商品A	500	100	200	100	1.600	
商品B	100	200	100	120	1.040	
商品C	200	400	250	200	1.798	
用户消费指数 a_u						1.51

品牌品类的高端程度

$$a_i = \frac{(\overline{p_{C \cap B}})^2}{\overline{p_C} \cdot \overline{p_B}} \cdot \lg(p_i + 1)$$

价格程度

$$a_u = \sqrt{\frac{\sum_{i \in N(u)} a_i^2}{N_u}}$$

订单商品加权求和

$$\theta = \frac{\lg(N_u + 1)}{|a_u - a_i| + \beta}$$

数据量的大小

消费指数吻合程度

商品*i*的分类为C，品牌为B， $\overline{p_{C \cap B}}$ 为品牌B且分类C的商品平均价格， $\overline{p_C}$ 代表分类C的商品平均价格， $\overline{p_B}$ 代表品牌B的商品平均价格。 p_i 代表商品*i*的价格。

$N(u)$ 表示用户订单内所有商品的集合。 N_u 代表用户订单内的商品数。 β 代表业务参数，用于调整权重。

8. 最终排序策略

- 召回商品设立阈值（销量非常差不召回，完全不相关不召回）
- 最终对商品的打分依赖于业务数据。

$$score = \left(\underbrace{\varepsilon \lg(M)}_{\text{业务数据：销量，浏览等}} + \underbrace{\sqrt{\sum_{i \in N(i)} s_i^2}}_{\text{标签匹配分数}} \right) \cdot \underbrace{(1 + \zeta \theta)}_{\text{特殊指标：消费指数}}$$

- 母婴用户：命中母婴标签高权重，可以保证母婴标签得分比较高的用户，推荐母婴商品且排序靠前
- 高端用户：消费指数可以保证消费指数相近的商品有更好的权重，高端商品可以排序靠前
- 无数据用户：排序算法退化为按照业务数据排序，选取最热门商品推荐。
- 数据量要求：少数商品数据因为有足够跨度的商品标签，可以召回大量商品计算相关性。

The end