



DataFunTalk

深度召回在文玩个性化推荐中的实践



赵争超 微拍堂-数据应用部负责人

目录

CONTENTS

01

业务背景及技术挑战

02

解决方案综述

03

基于deepMF和skip-gram的拍品embedding学习

04

跨域学习搜索关键词意图embedding

05

用户多兴趣embedding学习

06

总结与展望

01

业务背景及技术挑战

Subject



微拍堂APP个性化业务场景

关于 微拍堂

- 专注文玩领域的垂直电商，DAU百万+
- 线上直播、竞拍、鉴宝

推荐算法在微拍堂APP的流量分发和用户体验中起着核心主导作用



首页feeds流

逛一逛，发现更多文玩好货



频道页feeds流

个性化直播内容推送



购物路径节点相关推荐

看了还看/买了还买

DataFunTalk

召回侧的技术挑战

拍品理解

- 文玩行业的非标品特征，欠缺表达拍品“卖点”的标准化属性ID
- 浅库存、短生命周期的竞拍属性，既无法直接对拍品id生成embedding，也很难利用历史行为信息

用户理解

- 用户在文玩行业的专业性差异大，导致模型基于用户行为学习消费需求时偏差不确定，embedding质量在人群之间差异大
- 部分用户需求过于聚焦，使得对推荐内容的多样性带来挑战

工程实现

- faiss中，基于用户多兴趣召回时，如何实现“带约束”的最近邻

02

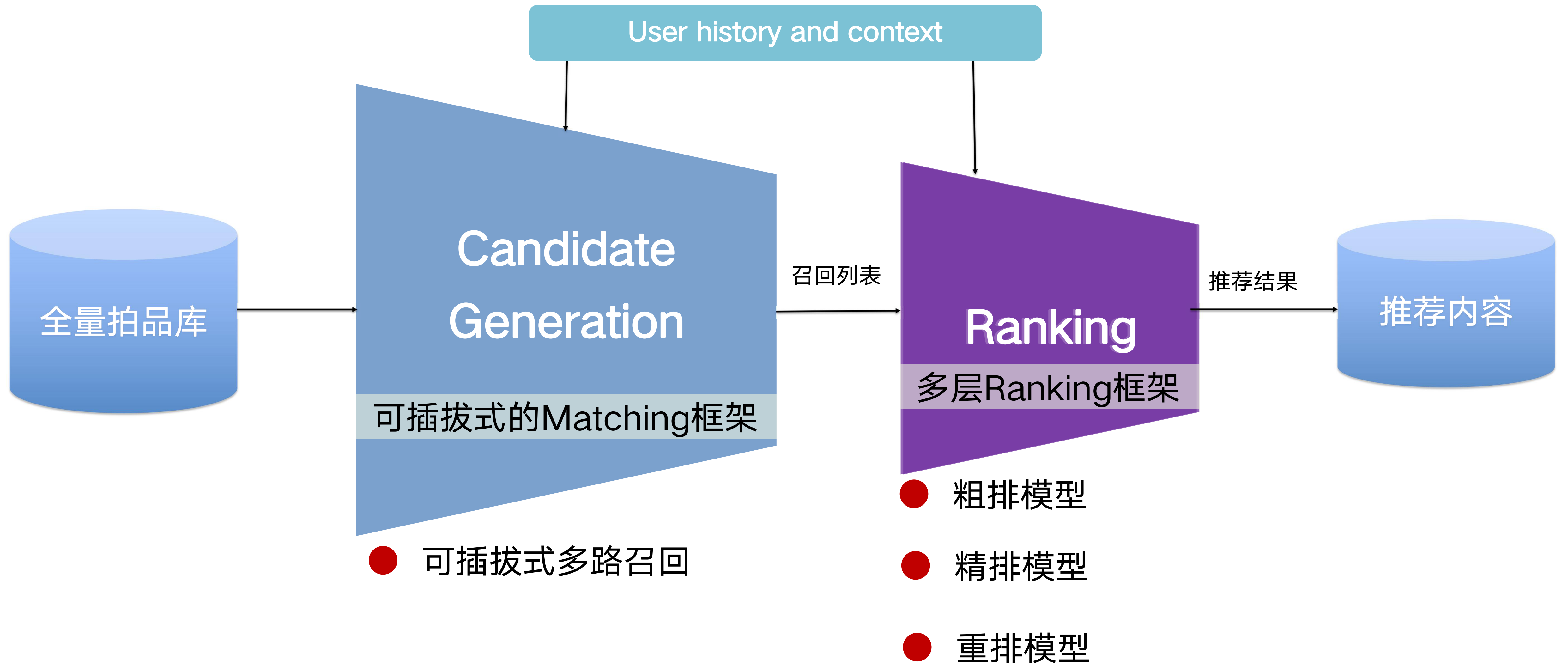
解决方案综述

Subject

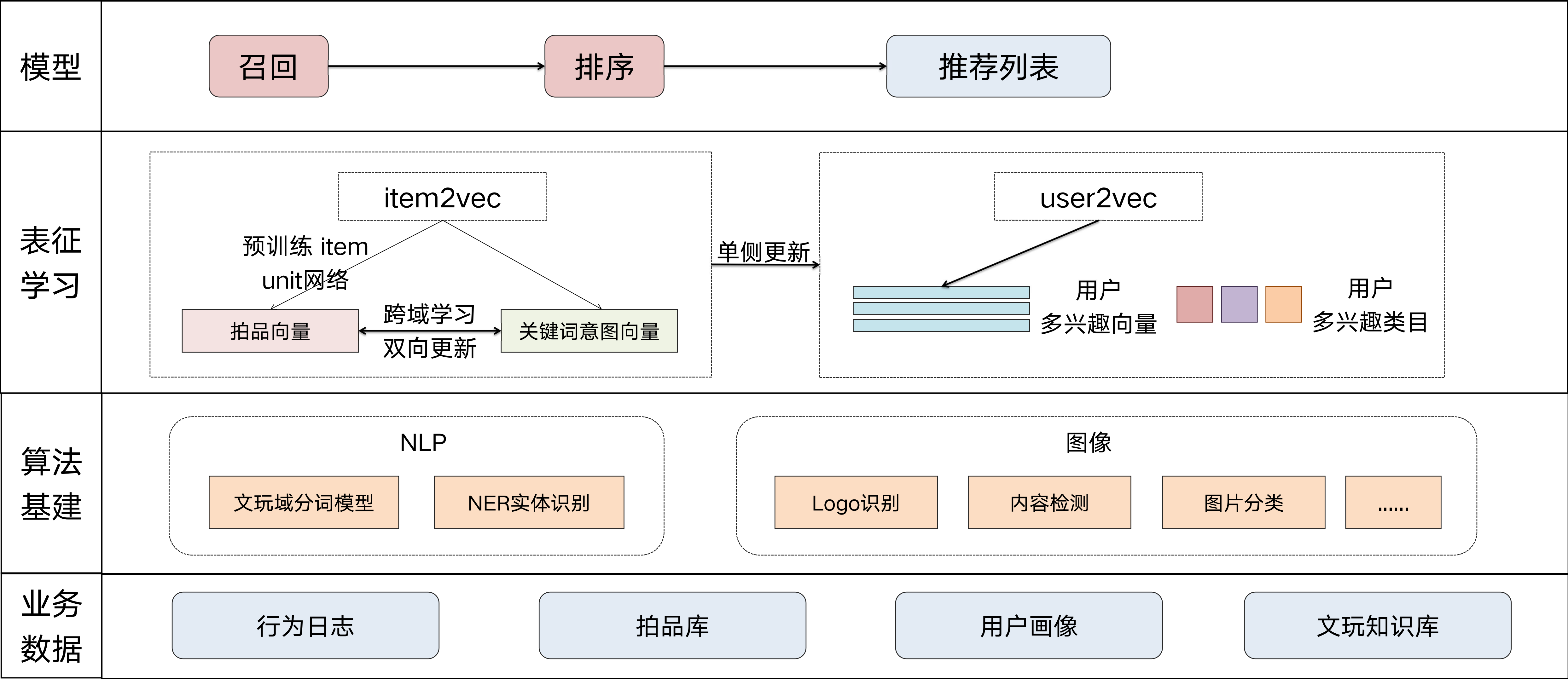


通用的算法框架

Matching+Ranking



召回侧的技术方案



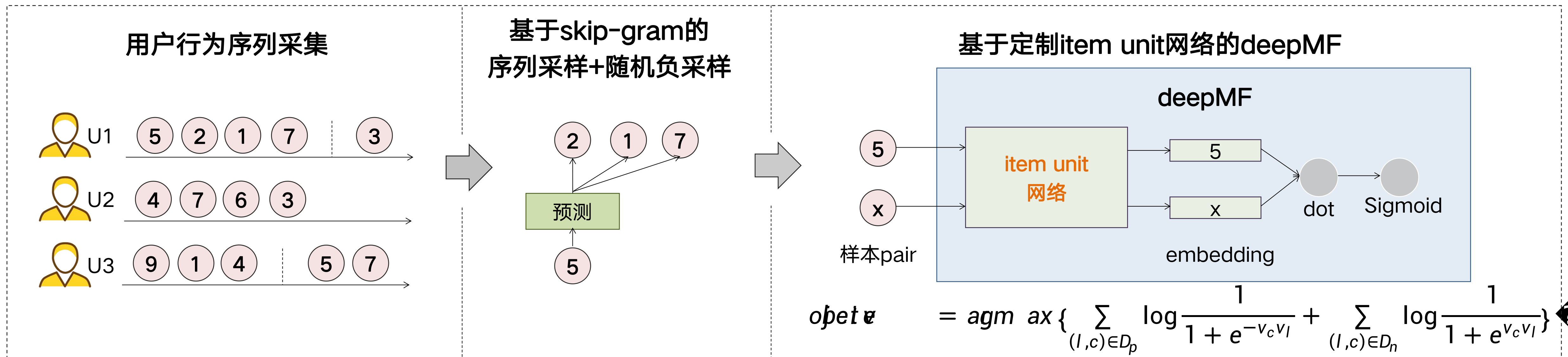
03

基于deepMF和skip-gram的拍 品embedding学习

Subject



拍品embedding学习



样本策略调优机制

- 结合用户文玩购物接触路径的session截断
- 热门拍品打压与负样本选取概率的trade off

$$1 - \sqrt{\frac{10^{-7}}{frequency}} \quad \text{VS} \quad \frac{frequency_j^{0.75}}{\sum_{i=0}^N frequency_i^{0.75}}$$

样本丢弃概率 负样本选取概率

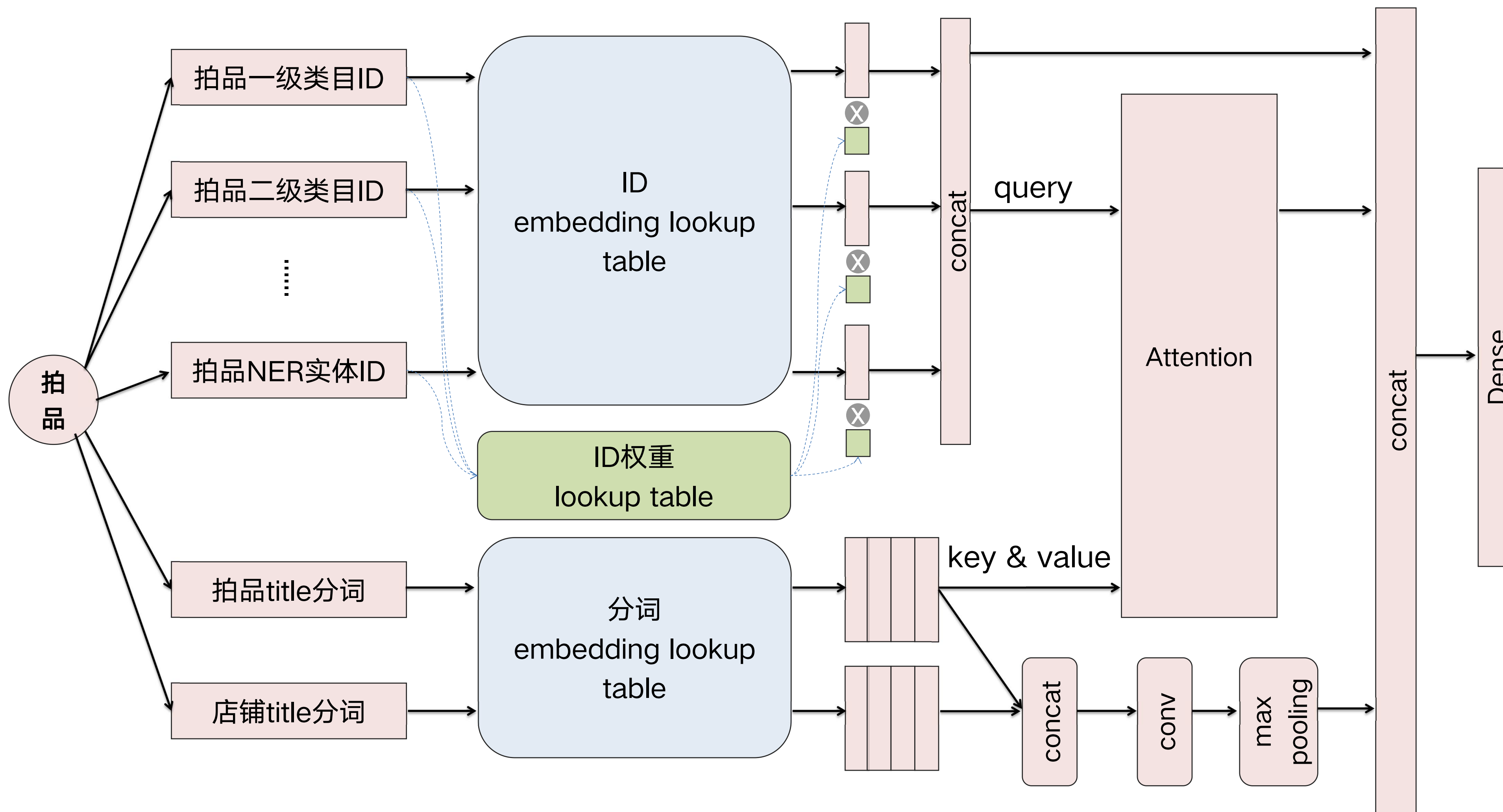
- 样本pair对物理意义的“距离”上的分布合理性调优，让模型既能学到拍品之间跨行业的大的差异，又能学到同类目之间的细微差异

浅库存/短生命周期下的deepMF

- 摒弃直接学习item embedding
- 结合拍品多维度属性id构建权重全局共享的item unit网络来表征item embedding
- 端到端地学习属性embedding及item unit网络参数权重

拍品embedding学习

item unit的网络结构



Item unit的设计思路

- 通过前置的NER实体识别任务，挖掘更多的拍品“卖点”
- 通过学习类目属性、“卖点”的embedding及组合参数来作为整个拍品的向量表达

类目属性、“卖点” 拍品标题”

$$At(Q, K, V) = softmax \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

- 类目、卖点作为内在属性，标题作为营销表达，通过attention挖掘两者之间的“加持”作用

拍品embedding学习

服饰 皮具	精品男士腰带	1	服饰 皮具	正品冠琴进口机芯男士全自动机械表夜光防	服饰 皮具	真皮男士皮带自动扣二层牛皮裤带商务男士腰带纯牛
服饰 皮具	精品男士腰带	2	服饰 皮具	正品冠琴新款手表女学生超薄韩版简约女	服饰 皮具	男士经典高档皮带
服饰 皮具	精品男士腰带	3	服饰 皮具	卡西欧?类型	服饰 皮具	印花 休闲牛仔皮带 男宽腰带 针扣时尚牛皮腰带
服饰 皮具	精品男士腰带	4	服饰 皮具	原装正品卡西欧	服饰 皮具	无拼接鳄鱼皮带
服饰 皮具	精品男士腰带	5	服饰 皮具	胸包运动小包	服饰 皮具	正品鳄鱼皮皮带
服饰 皮具	精品男士腰带	7	服饰 皮具	官方正品·新款瑞士正品全自动机芯手表男	服饰 皮具	真皮新款时尚鳄鱼纹胸包头层牛皮
服饰 皮具	精品男士腰带	9	服饰 皮具	正品全自动多功能防水夜光男手表学生韩版	服饰 皮具	男士斜挎包鳄鱼纹时尚韩版潮
服饰 皮具	精品男士腰带	15	服饰 皮具	卡西欧GA110-奥利奥手表	服饰 皮具	瑞士新款全自动机芯
服饰 皮具	精品男士腰带	30	服饰 皮具	【漏】官方正品爆款欧利时绿水鬼石英男款	服饰 皮具	韩版男包休闲包单肩包手提斜挎包潮流商务原创设计
金银饰品	宝格丽小红人同款手链	1	金银饰品	天然和田碧玉转运珠耳坠	金银饰品	施华洛世奇SWA INFINITY浪漫叠加手链
金银饰品	宝格丽小红人同款手链	2	金银饰品	简约通勤·旭平首饰小花简约耳环注	金银饰品	配证书)足银手链女日韩简约个性闺蜜学生森系玫瑰
金银饰品	宝格丽小红人同款手链	3	金银饰品	s925银镶天然A翡翠观音坠子项链	金银饰品	天然蓝宝石手链
金银饰品	宝格丽小红人同款手链	4	金银饰品	草绿色玉髓耳坠	金银饰品	(配证书)足银手链女日韩简约个性森系玫瑰花甜美
金银饰品	宝格丽小红人同款手链	5	金银饰品	手镯女孔雀银镯子送女友妈妈	金银饰品	女款三生三世手环
金银饰品	宝格丽小红人同款手链	7	金银饰品	罗马数字圆环相扣钛钢玫瑰金项链女时尚百	金银饰品	银饰品首饰
金银饰品	宝格丽小红人同款手链	9	金银饰品	永恒星月吊坠	金银饰品	罗泰空心吊坠
金银饰品	宝格丽小红人同款手链	15	金银饰品	复古民族风新娘结婚开口指环镀24K金色	金银饰品	纯银5201314手镯三生三世送女友
金银饰品	宝格丽小红人同款手链	30	金银饰品	【新品】S925纯银镀金景泰蓝祥云平安扣耳	金银饰品	【配证书】纯银手镯999开口光面心经贵妃镯
服饰 皮具	牛皮马丁靴	1	服饰 皮具	加大 外套	服饰 皮具	牛皮马丁靴
服饰 皮具	牛皮马丁靴	2	服饰 皮具	春秋新款牛仔外套男潮牌韩版宽松休闲牛仔	服饰 皮具	韩版马丁靴男厚底黑色英伦青年潮流男生皮靴子秋冬
服饰 皮具	牛皮马丁靴	3	服饰 皮具	富贵鸟外套男秋季韩版潮流帅气上衣休闲百	服饰 皮具	皮鞋男高帮休闲鞋子男韩版潮流马丁靴男鞋冬季加绒
服饰 皮具	牛皮马丁靴	4	服饰 皮具	大爱中国风!中华舞狮大印花秋冬新款男士时	服饰 皮具	OSCO真皮马丁靴男2019秋冬新款潮流中帮皮靴时尚
服饰 皮具	牛皮马丁靴	5	服饰 皮具	春季英伦丝质衬衫男士刺绣亮面长袖韩版修	服饰 皮具	头层皮皮毛一体鞋
服饰 皮具	牛皮马丁靴	7	服饰 皮具	夹克男外套2020秋装上衣男士潮流帅气宽松	服饰 皮具	G5 CONXEGN秋冬季真皮马丁靴男高帮工装鞋中帮
服饰 皮具	牛皮马丁靴	9	服饰 皮具	马甲男秋冬季韩版学生帅气男士保暖背心办	服饰 皮具	全鳄鱼皮豆豆鞋
服饰 皮具	牛皮马丁靴	15	服饰 皮具	雅鹿秋季灯芯绒长袖衬衫男磨毛纯色衬衣中	服饰 皮具	头层牛皮时尚板鞋
服饰 皮具	牛皮马丁靴	30	服饰 皮具	中国风男士短袖t恤夏季薄款印花五分袖T恤	服饰 皮具	全编织懒人鞋纯手工休闲鞋2020新款豆豆鞋皮鞋潮流

拍品

相似拍品
NCF版

相似拍品
新版

单路召回点击率提升10%↑ 出价率提升5%↑

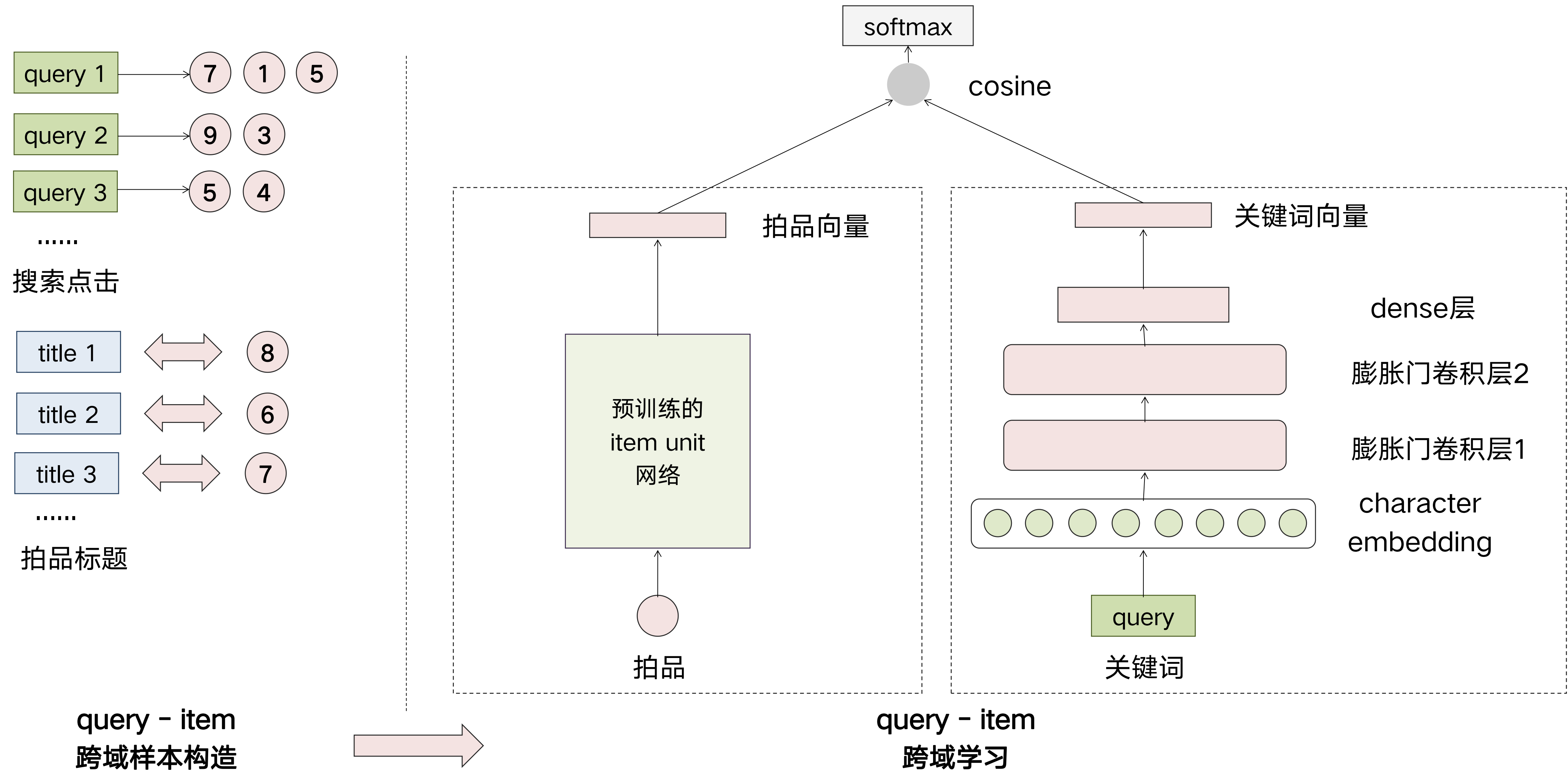
04

跨域学习搜索关键词意图 embedding

Subject



关键词意图embedding学习



■ 关键词embedding学习

非迁移学习版本

银元	0	保粹云南小双旗两角银币
银元	1	孙中山像银币角广东省造
银元	2	年中华民国二十二年;船洋;银币
银元	3	孙像船洋壹圆, 年
银元	5	浪琴表系列, 公价万, 表径。全新全套未使用,
银元	7	二手正品新欧米茄蝶飞黄金精钢自动机械男表
银元	10	浪琴精钢自动机械男表
银元	20	众诚评级奥地利因斯布鲁克冬奥会银币, 包老保真! 原图发
银元	100	玉色龙牌吊坠
书法	0	朱子家训国展高手李成峰书法真迹精品小楷长卷
书法	1	兰亭书家陈美东楷书文房联
书法	2	胡紫桂省书协副主席首届兰亭获奖提名八届国展奖
书法	3	十二届国展杨科云精品行书竖幅书法真迹支持定制
书法	5	杨科云十二届国展楷行草三体入展, 六届兰亭奖对联新作
书法	7	王堂兵中书协理事书法行书横幅真迹支持定制
书法	10	二手正品新欧米茄蝶飞黄金精钢自动机械男表
书法	20	欧利时绿水鬼跨境爆款石英表防水男士手表男表劳力士同款绿水鬼
书法	100	欧米茄超霸系列自动机械男表

迁移学习版本

银元	0	保粹评级广东贰毫银币, 严评入盒, 包老保真! 原图发货
银元	1	孙小头单日开
银元	2	原光地瓜伍角银币
银元	3	原味民国袁大头银元, 公博评级, 原味透骨老包浆, 自然醇
银元	5	云南光绪半圆老龙
银元	7	库存银元低价放漏
银元	10	甘肃壹圆三年大头
银元	20	英国乔治五世印度卢比, 重, 原味非常漂亮, 包老保真包
银元	100	丙午湘撇户点云龙大清十文清代铜元永久保老
书法	0	书法作品尺寸三尺整版米
书法	1	兰亭书家陈美东楷书文房联
书法	2	国展状元汪敦银真迹
书法	3	朱子家训国展高手李成峰书法真迹精品小楷长卷
书法	5	道教会长任法融款书法
书法	7	金新宇老师亲自寄给您
书法	10	徐朝江款书法作品书房小楹联
书法	20	回流字画书法
书法	100	纯手绘国画书法

“相关词” 推荐的点击率提升10%

05

用户多兴趣embedding学习

Subject

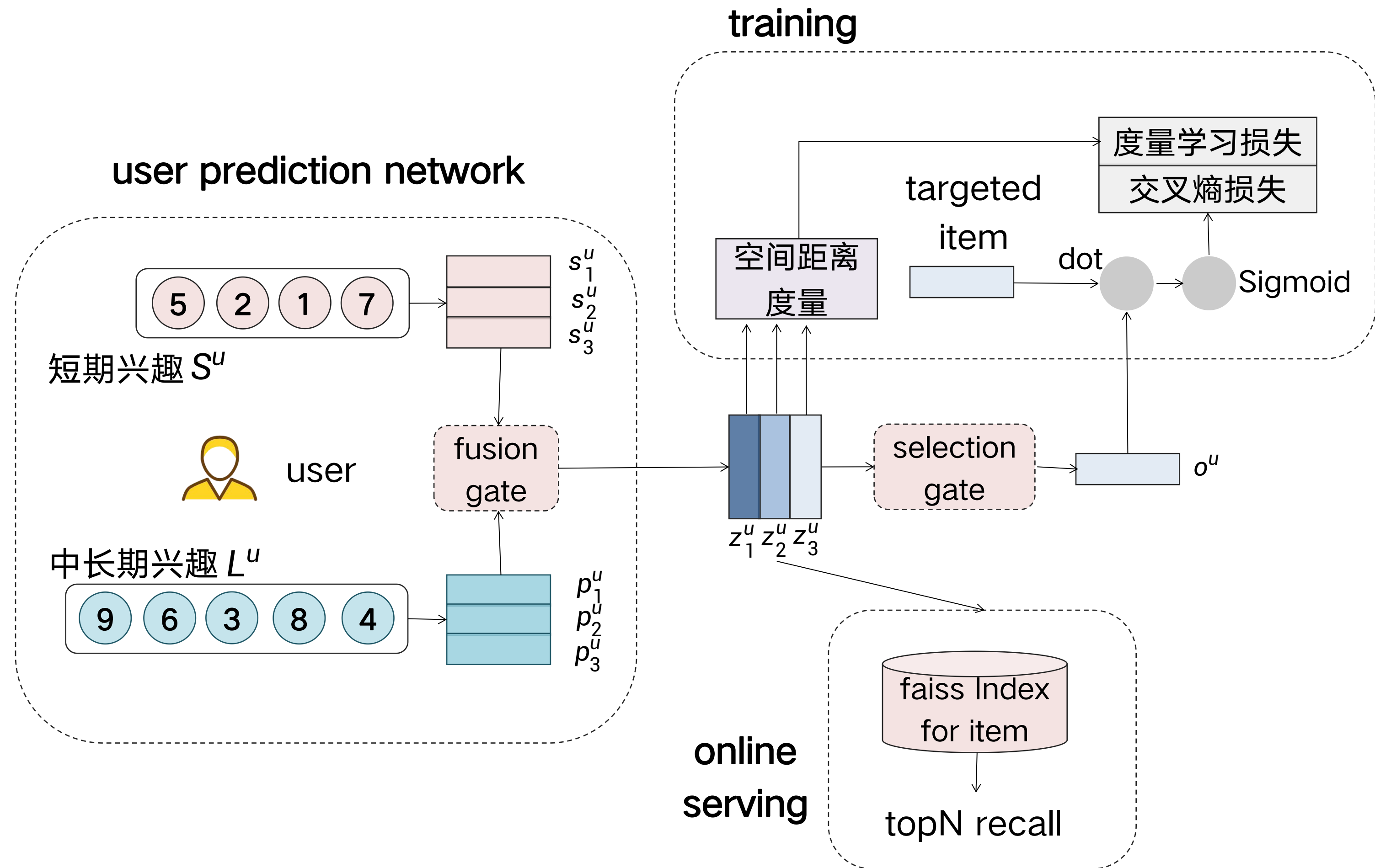


用户多兴趣embedding学习

核心业务问题

- 如何学习用户在一个session中的多峰兴趣
- 如何融合、区隔长短兴趣对后续行为的影响
- 如何更好地学习用户多兴趣之间的差异度

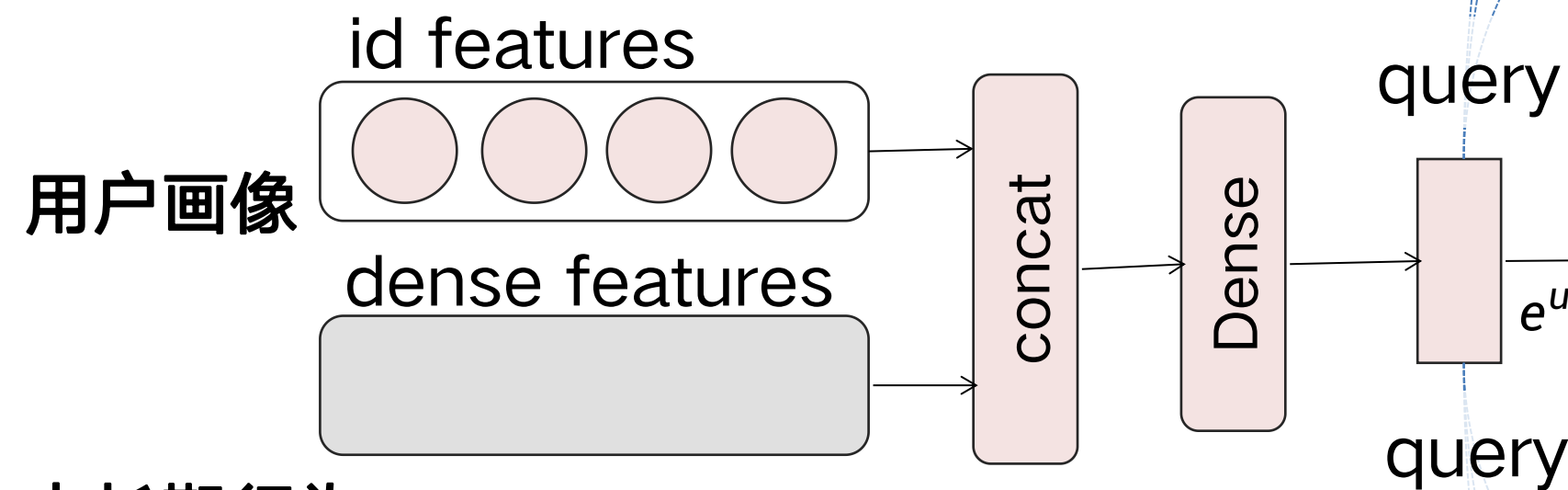
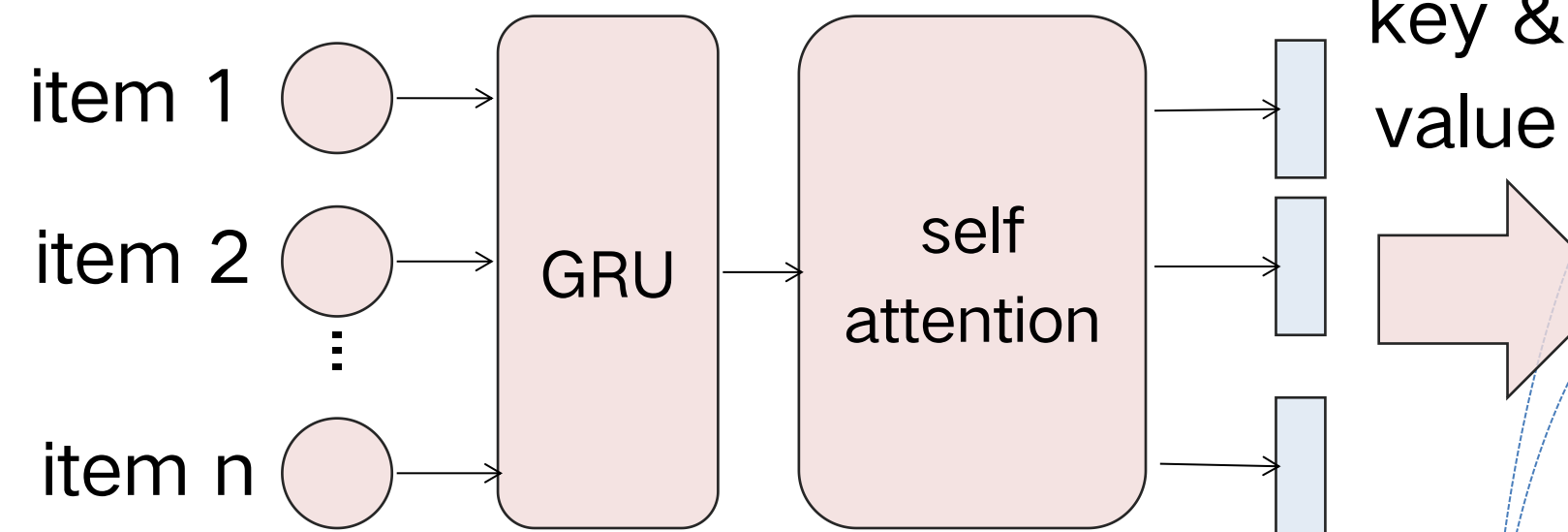
general network



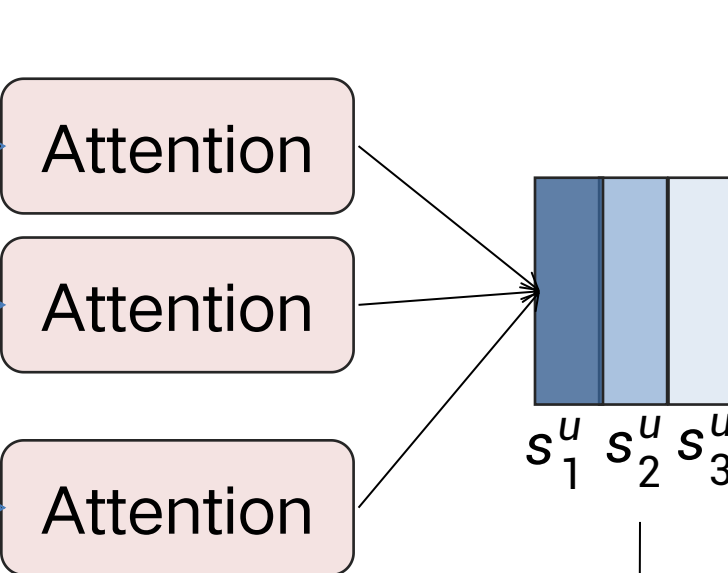
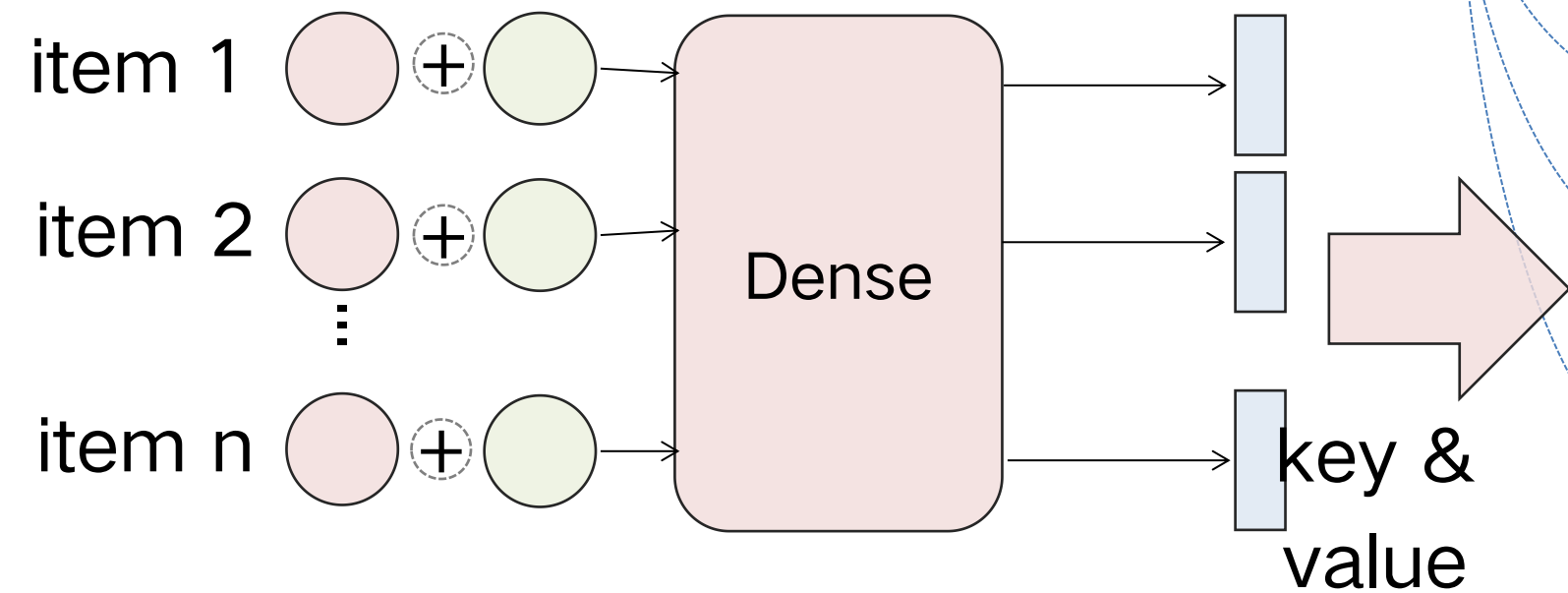
用户多兴趣embedding学习

fusion gate + select gate达成用户兴趣的准确性和独立性

短期行为



中长期行为



fusion gate

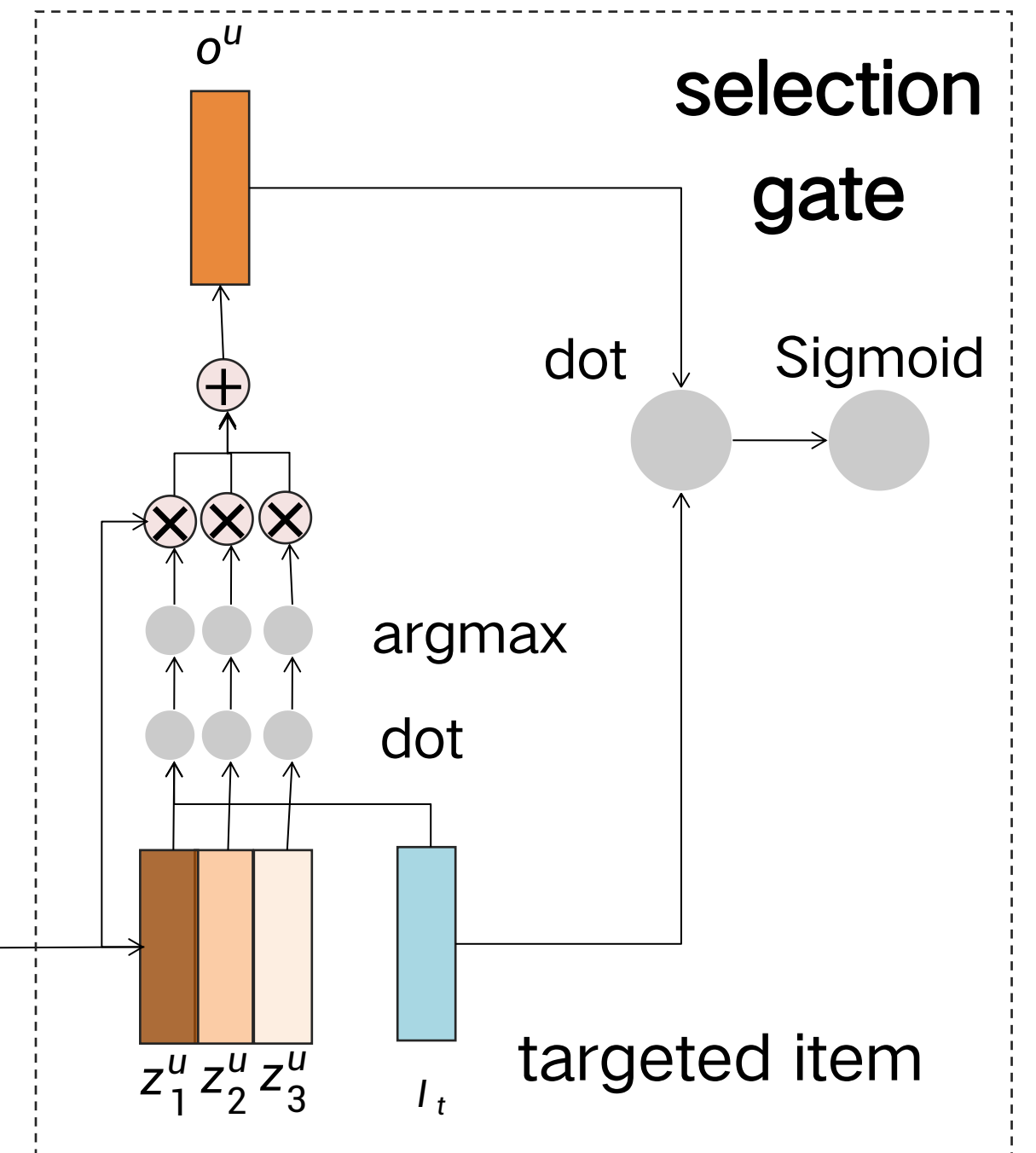
$$z_i^u = (1 - G^u) \odot s_i^u + G^u \odot p_i^u$$
$$G^u = \sigma(w^1 e^u + w^2 s^u + w^3 p^u)$$

selection gate

$$o^u = G'_1 \odot z_1^u + G'_2 \odot z_2^u + G'_3 \odot z_3^u$$
$$g'_i = I_t \cdot z_i^u$$
$$G'_i = \begin{cases} 1 & g'_i = \text{argmax} \{g'_1, g'_2, g'_3\} \\ 0 & g'_i \neq \text{argmax} \{g'_1, g'_2, g'_3\} \end{cases}$$

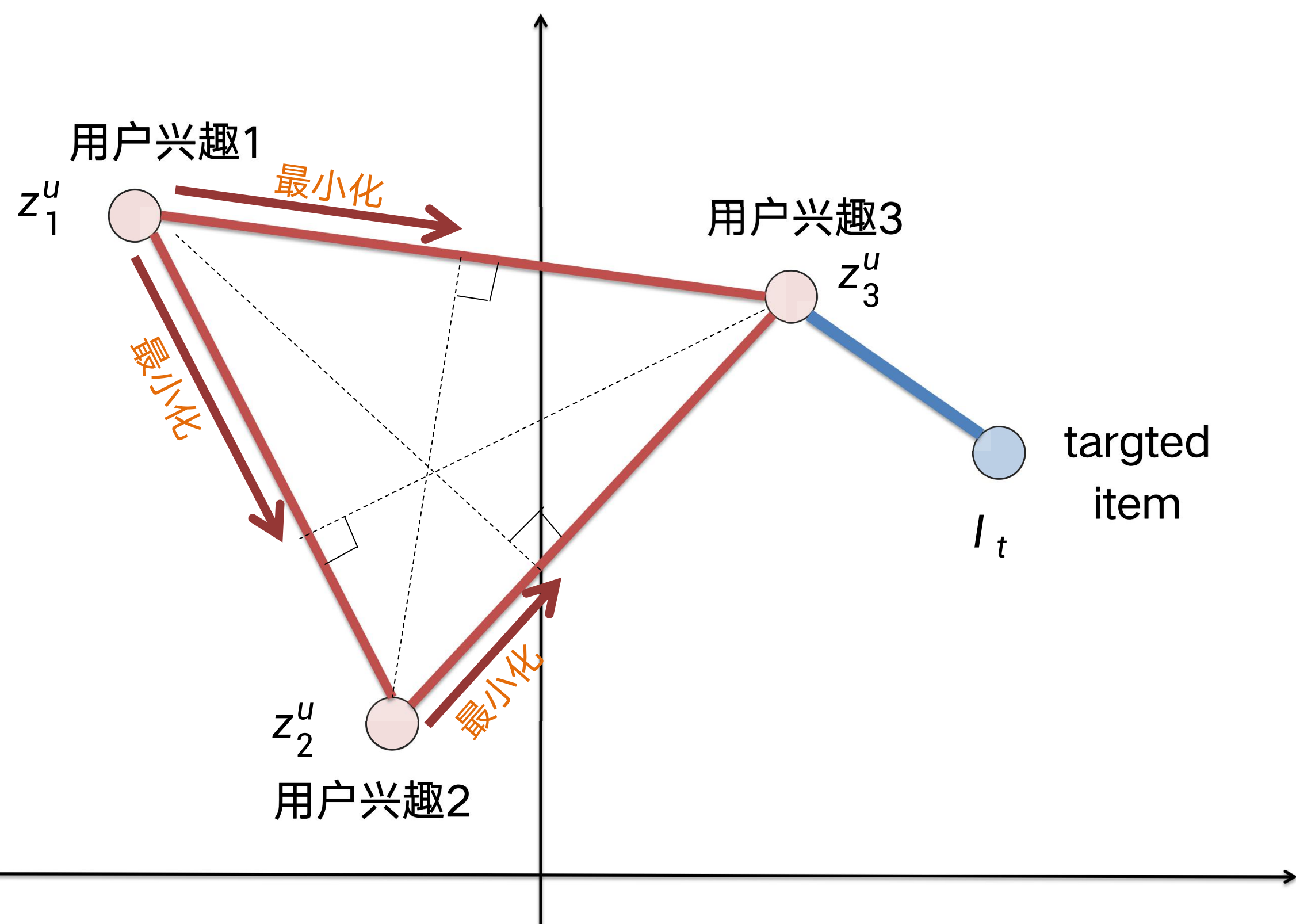
fusion gate

selection gate



用户多兴趣embedding学习

通过度量学习，增加用户多兴趣之间的区隔度



why metrics learning?

- 上一步的selection gate机制会导致与target item兴趣不一致的其他向量训练不充分？
- 通过度量学习人为增加用户多兴趣之间的区隔度
- 在学习上一步的交叉熵损失的同时，3个用户兴趣向量之间的距离尽量远

$$L_{tr} = \sum_{u \in U} \sum_{i \neq j} z_i^u \cdot z_j^u = \sum_{u \in U} \sum_{i \neq j} \|z_i^u\| \cdot \|z_j^u\| \cos \theta(i, j)$$

$$L = L_0 + \lambda \cdot L_{tr}$$

用户多兴趣embedding学习



兴趣向量

● 普洱茶 其他茶叶 紫砂 白酒 岩茶 陶器 白茶 当代瓷 0.016

● 古钱币 奢侈品 观赏鱼 铜元 木质珠串 岫岩玉 木质把件 工艺刀剑 0.011

● 园艺用品 铜元 金银饰品 奢侈品 翡翠原石 松石 核雕 | 核桃 银元 0.011

兴趣向量

● 普洱茶 其他茶叶 白酒 紫砂 白茶 岩茶 陶器 当代瓷 0.017

● 葡萄酒 岩茶 白酒 翡翠原石 普洱茶 洋酒 滋补营养品 白茶 0.018

● 贵重宝石 奢侈品 金银饰品 纪念币/章 岫岩玉 工艺刀剑 白茶 纸钞 0.012

通过 度量学习

- 提升了多兴趣之间的区隔度
- 提升了单个兴趣的聚焦程度

度量学习版本

普洱茶	冰岛销魂砖	0.850805
普洱茶	老班章砖克拍	0.853364
普洱茶	一提七饼年老班章	0.857160
普洱茶	年老班章	0.859862
普洱茶	一提七饼老班章	0.861713
普洱茶	文革砖克起拍	0.863543
普洱茶	老班章回馈新老客户	0.866646
普洱茶	捡漏砖冰岛古树生茶砖	0.869893
普洱茶	珍藏老茶	0.876492
普洱茶	年龙园号大班章熟茶	0.881735
普洱茶	活动超捡漏福利勐海经典老茶砖	0.889409
普洱茶	年代虫屎茶砖	0.904079
普洱茶	一提饼年早春磨烈生饼	0.909909

非度量学习版本

普洱茶	斤班章黄	普洱茶	冰岛销魂砖	0.862766
普洱茶		普洱茶	尝鲜价一提七饼拍多少算多少名称老班章	0.868427
普洱茶		普洱茶	尝鲜价一提七饼拍多少算多少	0.869896
普洱茶		普洱茶	尝鲜价一提七饼拍多少算多少	0.870883
普洱茶		普洱茶	一提七饼一起拍拍多少算多少	0.871561
普洱茶		普洱茶	年老班章	0.876402
普洱茶		普洱茶	一提七饼一起拍拍多少算多少	0.877787
普洱茶		普洱茶	一提饼老班章熟茶克尝鲜价一提七饼拍多少算多少	0.883293
普洱茶		普洱茶	一提七饼老班章	0.889305
普洱茶		普洱茶	砖一起拍勐海经典放漏卖了	0.896175
普洱茶		普洱茶	一提七饼冰岛古树茶	0.903272
普洱茶		普洱茶	饼元起拍年老班章生茶	0.916761
普洱茶		普洱茶	饼冰岛乔木开拍	0.922517
普洱茶		普洱茶	饼放漏老班章纯料古树茶	0.927847

■ 用户多兴趣embedding学习

多兴趣向量的进一步发散：潜在兴趣向量

业务问题聚焦：如何扩大用户的兴趣半径？

- 在文玩行业，不同于综合电商，大部分用户的兴趣聚焦在1-2个类目，给推荐多样性带来极大挑战

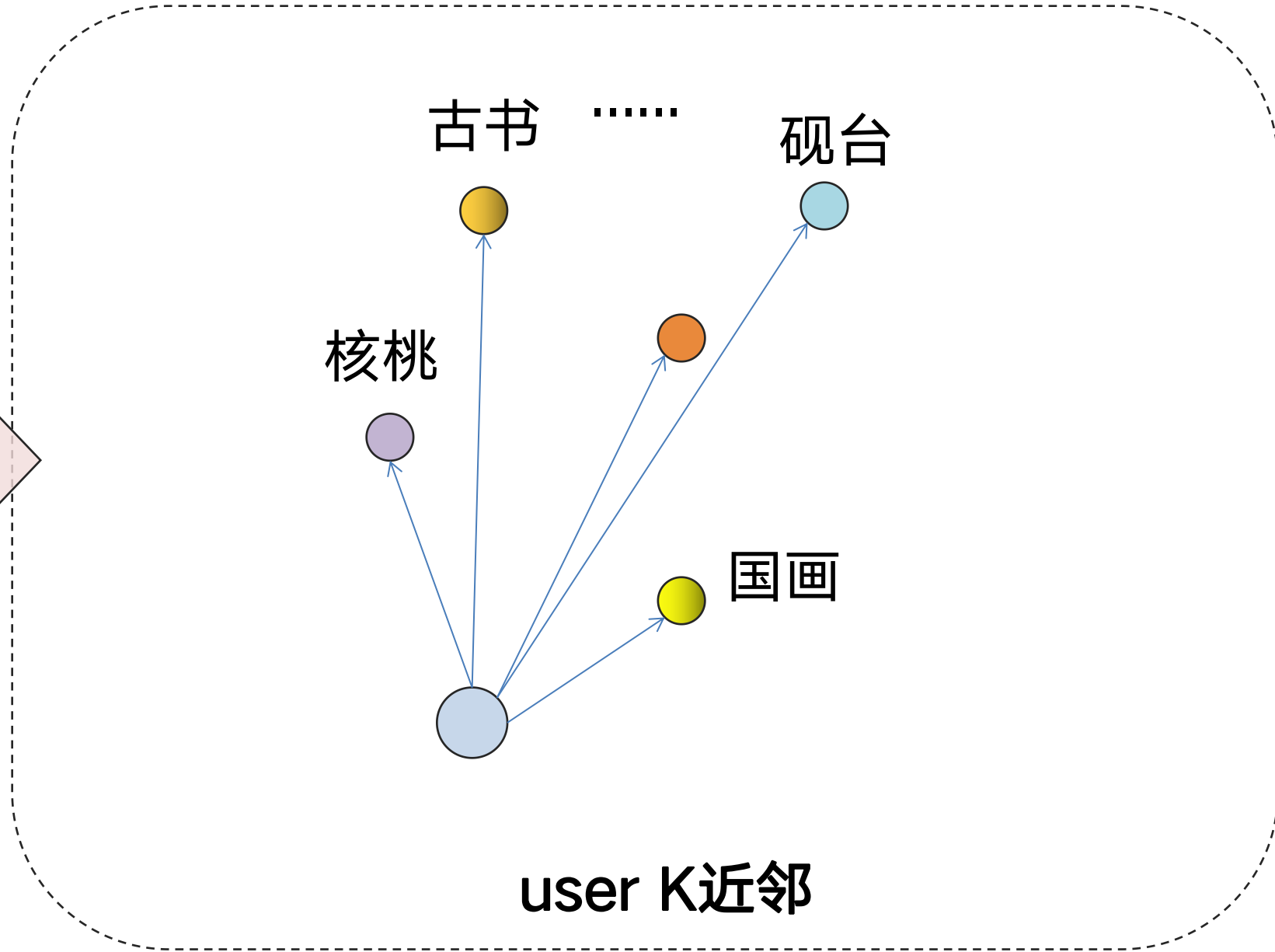
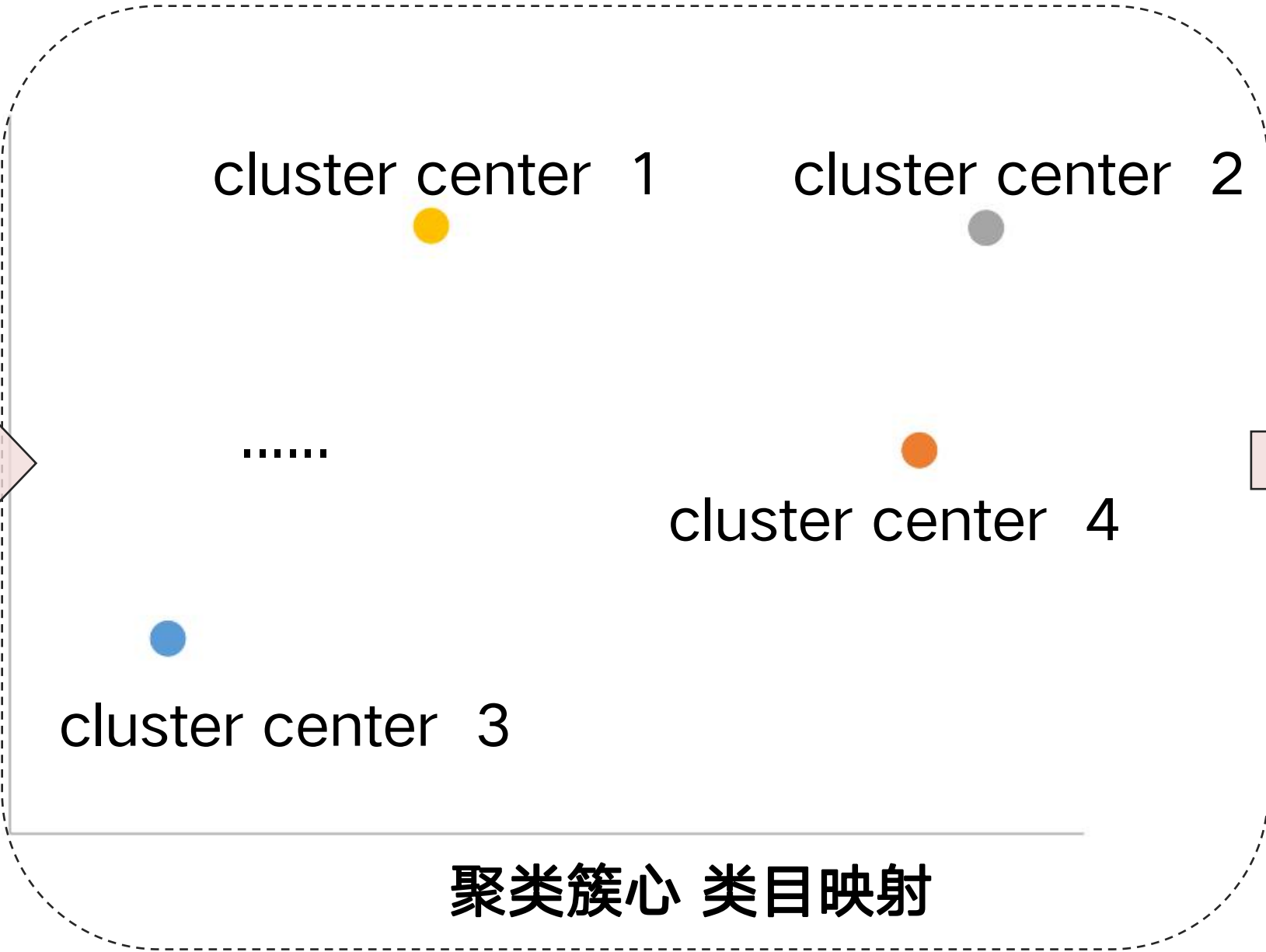
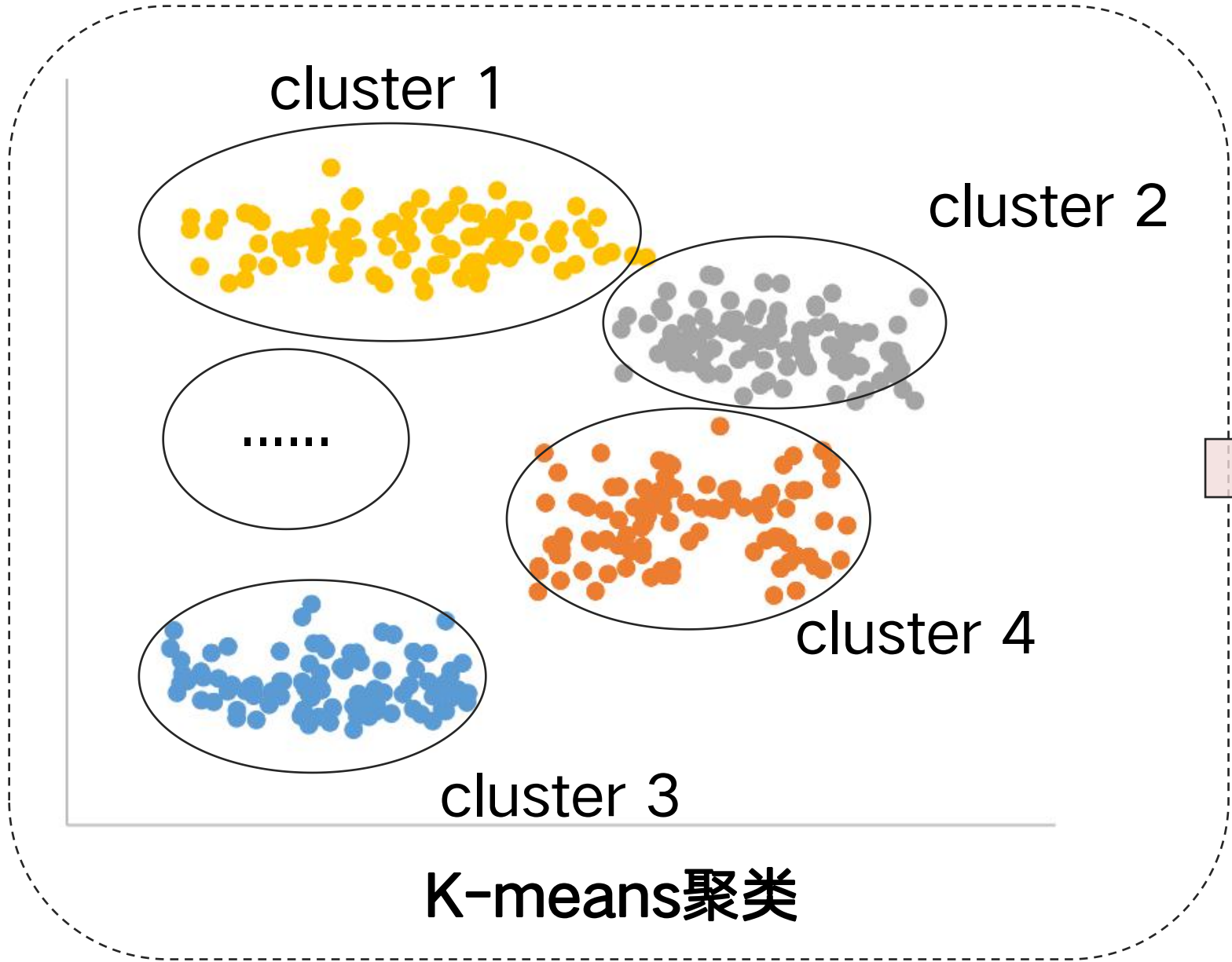
$$ILS(L) = \frac{\sum_{b_i \in L} \sum_{b_j \in L, b_j \neq b_i} S(b_i, b_j)}{\sum_{b_i \in L} \sum_{b_j \in L, b_j \neq b_i} 1} > 0.85$$

- 期望推荐用户从来没看过，但是又和兴趣有一定相关的拍品，增加惊喜感
- 基于群体智慧的“协同过滤”方法不适用于文玩行业，原因是文玩类目之间缺少必然的相关性

用户多兴趣embedding学习

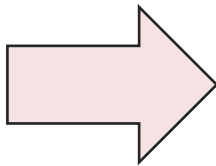
用户潜在偏好类目K近邻挖掘

用户潜在兴趣类目={向量K近邻} - {最近有行为类目}



茶器	侘寂之美茶舍静物
其他	玉杯花瓶日本
茶器	侘寂之美茶舍静物
其他	玉杯花瓶日本
茶器	侘寂之美茶舍静物
其他	玉杯花瓶日本

用户行为



文房器	文房周边	江户晚期明治桐木围棋盒那智石车渠后配棋盘
文房器	其他	日本回流围棋套组
香料 香炉	其他	日本大发制线香原木盒
香料 香炉	香料	日本包邮日本制香名家鸠居堂未使用新品两盒

潜在偏好

用户多兴趣embedding学习

如何在faiss中带类目约束地召回最近邻

我们的方法

- 对每个类目的向量，找到一种独特的距离度量和坐标变换的方法，使得用户对目标类目的所有向量的距离远小于其他类目的向量

1) 拍品和用户向量的标准化

$$V^* = \left[\frac{x_1}{\|V\|}, \frac{x_2}{\|V\|}, \dots, \frac{x_n}{\|V\|} \right]$$

- 将向量转化为模为1的单位向量后，余弦距离等价于欧式距离

$$V_1 = [x_1, x_2, \dots, x_n] \quad V_2 = [y_1, y_2, \dots, y_n]$$

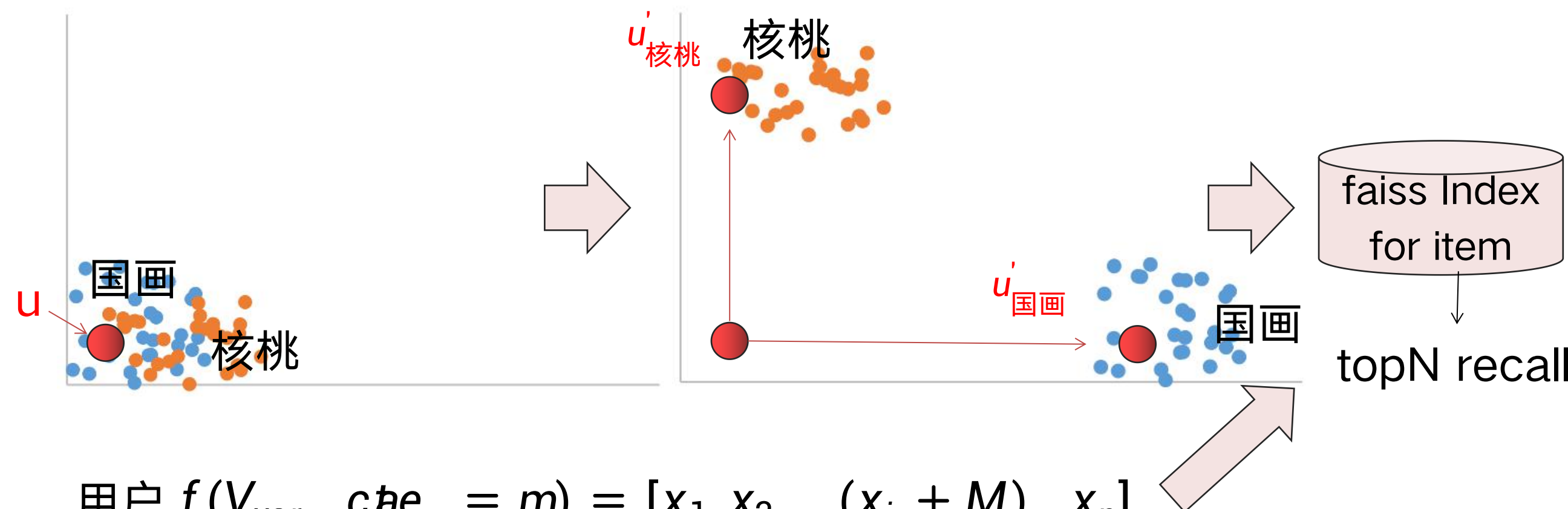
内积 $V_1 \cdot V_2 = \sum x_i * y_i = \|V_1\| * \|V_2\| * \cos \langle V_1, V_2 \rangle$

余弦距离 $1 - \cos \langle V_1, V_2 \rangle$

欧式距离 $\sqrt{\sum (x_i - y_i)^2} = \sqrt{\sum x_i^2 + \sum y_i^2 - 2 \sum x_i * y_i}$
 $= \sqrt{1 + 1 - 2V_1 \cdot V_2} = \sqrt{2 - 2\cos \langle V_1, V_2 \rangle}$

2) 拍品及用户向量的坐标变化

拍品 $f(V_{item}, cate = m) = [x_1, x_2, \dots, (x_i + M), \dots, x_n] \quad M \gg 2$



用户 $f(V_{user}, cate = m) = [x_1, x_2, \dots, (x_i + M), \dots, x_n]$

潜在兴趣向量的召回方式，对比直接的类目之间的协同过滤，单路点击率提升20%↑

06

总结与展望

Subject



■ 总结及未来展望

总结

在线文玩拍卖，是最近几年新兴的电商垂直市场，它将线下文玩交易活动，以竞拍的交易模式搬到了线上。文玩拍卖的个性化推荐，对比淘宝、京东等综合性电商，有着其独特的商业化领域问题，如 文本的非标品属性、竞拍模式的短生命周期等等，最近半年来，我们通过深度学习不断探索和尝试着业务问题的数学解法：

- 通过NER任务，及文玩知识库的构建，挖掘拍品的卖点
- 通过属性、卖点ID组合item unit网络充分利用历史数据来学习新品的embedding表达
- 通过双门网络、度量学习来更好地挖掘用户的多峰兴趣

展望

未来，我们期望在更加深入理解文玩市场、传统文化、竞拍机制下的商业模式的基础上，能够找到AI驱动用户购物体验、网站UV价值提升的更加fancy的打开方式：

- 结合文玩知识的graph模型，更好地表征拍品及核心价值
- 结合强化学习，结合用户的竞拍心智，更好地挖掘用户的潜在需求
- 结合图像美学元素的多模态CTR模型
-



THANKS!

今天的分享就到这里...

Ending

