

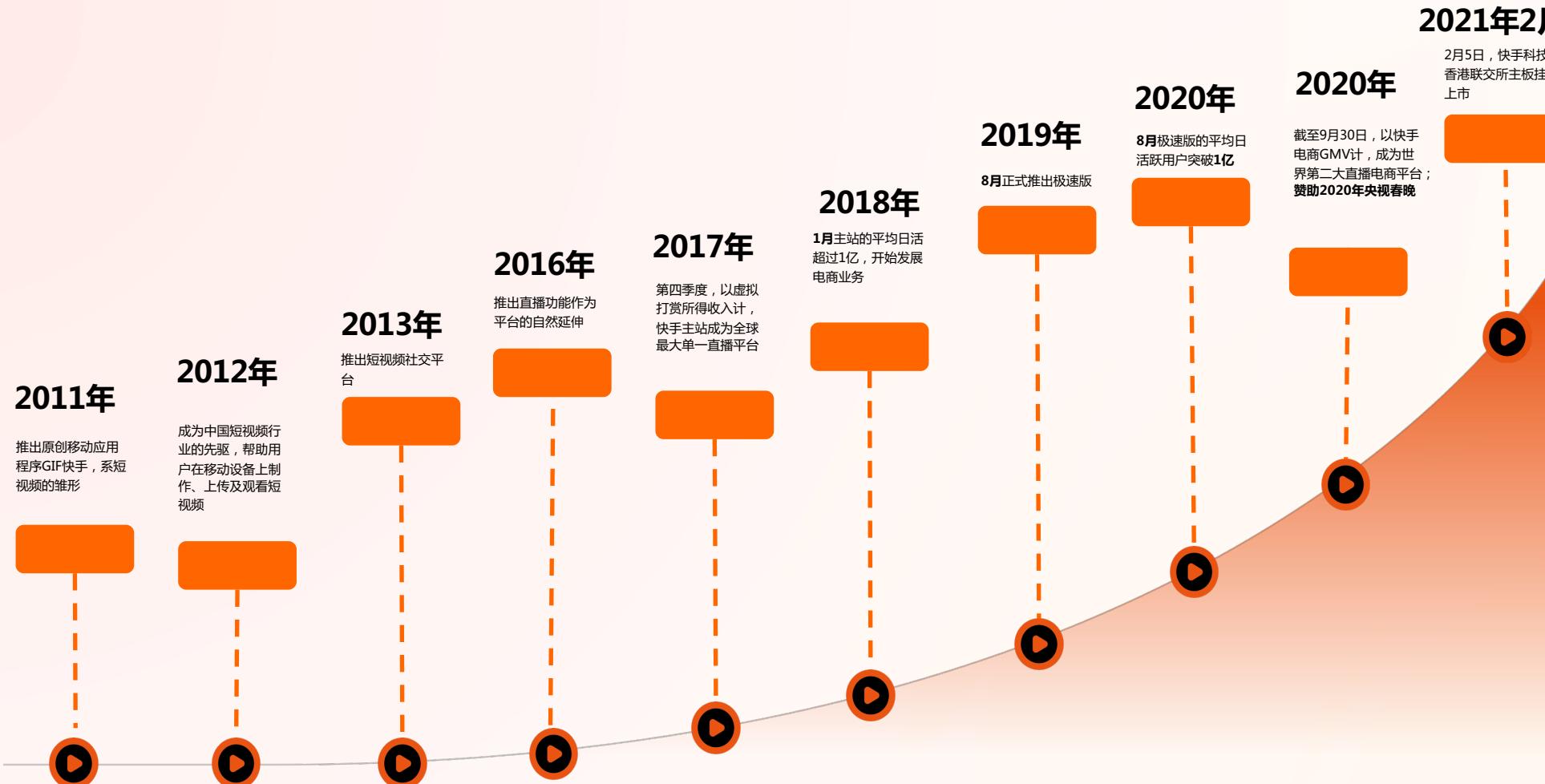


快手AI前沿技术探索及合作布局

演讲人：宋洋 快手科技副总裁



快手：高互动强参与，普惠数字社区



AI技术驱动，打造极致用户体验



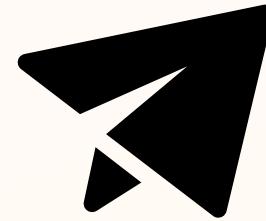
生产

做有**创意**的作品



理解

构**精准**的描述



分发

看**感兴趣**的内容

内容生产：高质量的视频生成

视频特效与创作工具

智能人像美化

数字人与XR互动

语音及文本生成



快手3D数字人
-为蒙牛打造首位数字员工



变老魔法表情

内容理解：大规模语言模型 (LLM)

「快意」大模型 (13B/66B) 在绝大部分 Benchmarks 上取得了同等模型尺寸下的 **SOTA 效果**
在2000条评测集合上，相比ChatGPT，快意13B **GSB**达到了0.85，快意66B **GSB**达到了1.05

模型	C-Eval 中文、全科 13.9K个单选题 清华&上交	CMMLU 中文、全科 11.5K个单选题 MBZUAI&上交&MSRA	MMLU(%) 英文、全科 14K个单选题 国外大学与研究所	GSM8K 英文、小学数学 8.5K个多步计算题 OpenAI	HumanEval 英文、代码与算法 168道编程题 OpenAI
通义千问-7B	59.6		56.7	51.6	24.4
百川2-13B	58.1	62.0	59.2	52.8	17.1
LLaMa2-13B	-	-	54.8	28.7	18.3
快意-13B	64.2	62.8	57.8	52.3	46.3
GPT-4	68.7	70.95	86.4	92.0	67.0
ChatGPT-175B	54.4	55.51	70.0	57.1	48.1
GLM-130B	44	-	44.8	-	-
LLaMa2-70B	50.1	-	68.9	56.8	29.9
快意-66B	72.0	70.5	66.5	59.1	56.1

*基于业界同等规模大模型主流benchmark的评测

内容理解：大规模多模态模型

建设媲美MidJourney的视觉内容生成能力，构建“**AIGC+X**”完整技术生态

构建快手AIGC-benchmark
(14个垂类+12个挑战项)

文生图可用率73%，进入行业第一梯队

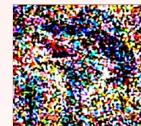
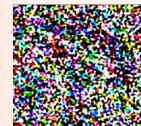
大规模中文概念数据

将Human Feedback引入
文生图链路

大语言模型与CLIP的
融合LLM

分级分辨率提升
技术方案

一只鸵鸟戴着软呢帽子对着麦克风深情唱歌



提示词工程

文本表征模型

T2I扩散模型

图像细节修复

动漫3D风格，超写实油画，五官精致，俊美的脸庞，黑色头发的小女孩，在巨大的中国龙和树木旁边，背后是巨大的守护动物

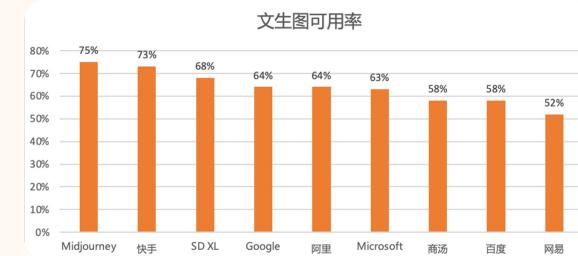
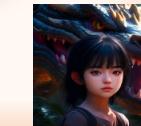
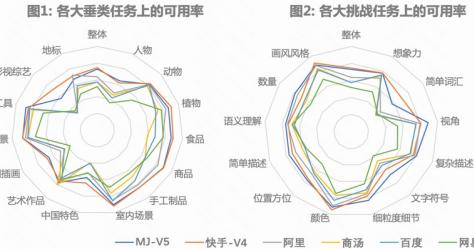


图1:各大垂类任务上的可用率



内容分发：个性化推荐感兴趣的内容

多任务&多领域

千人千模 长短兴趣结合

打破信息茧房的魔咒

业界首个**万亿**参数排序模型

1.9万亿参数规模

千亿的模型特征量

超过**万级**的用户历史序列

单列精选页



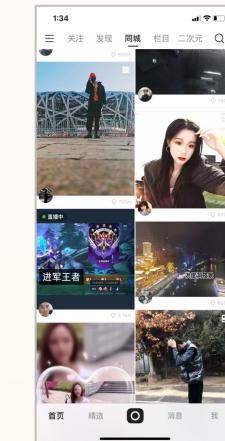
双列发现页



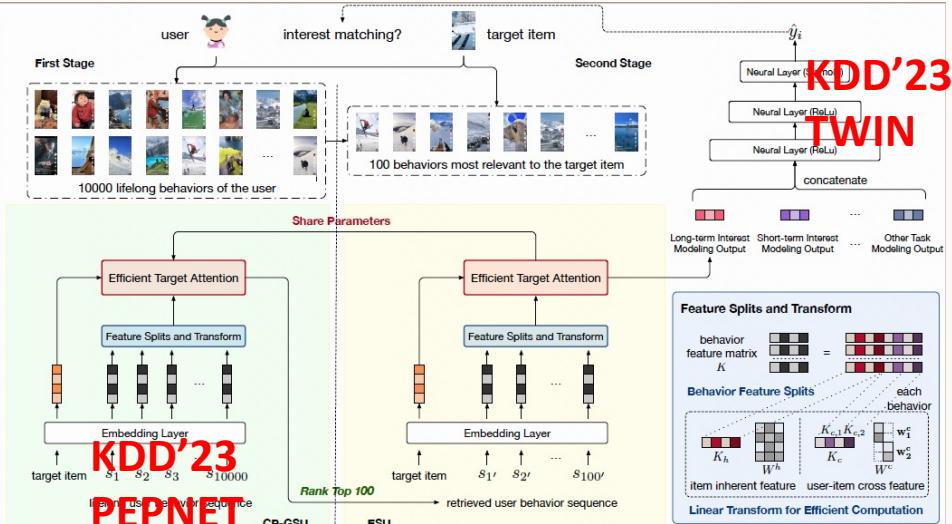
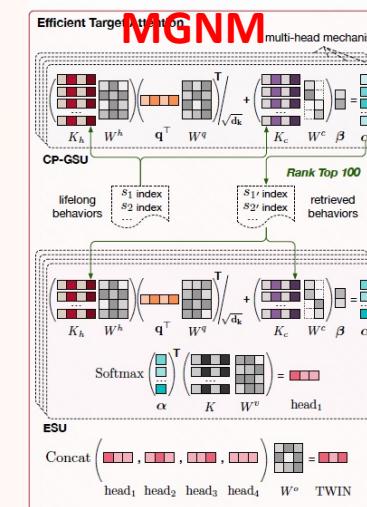
关注页



同城页



SIGIR'22



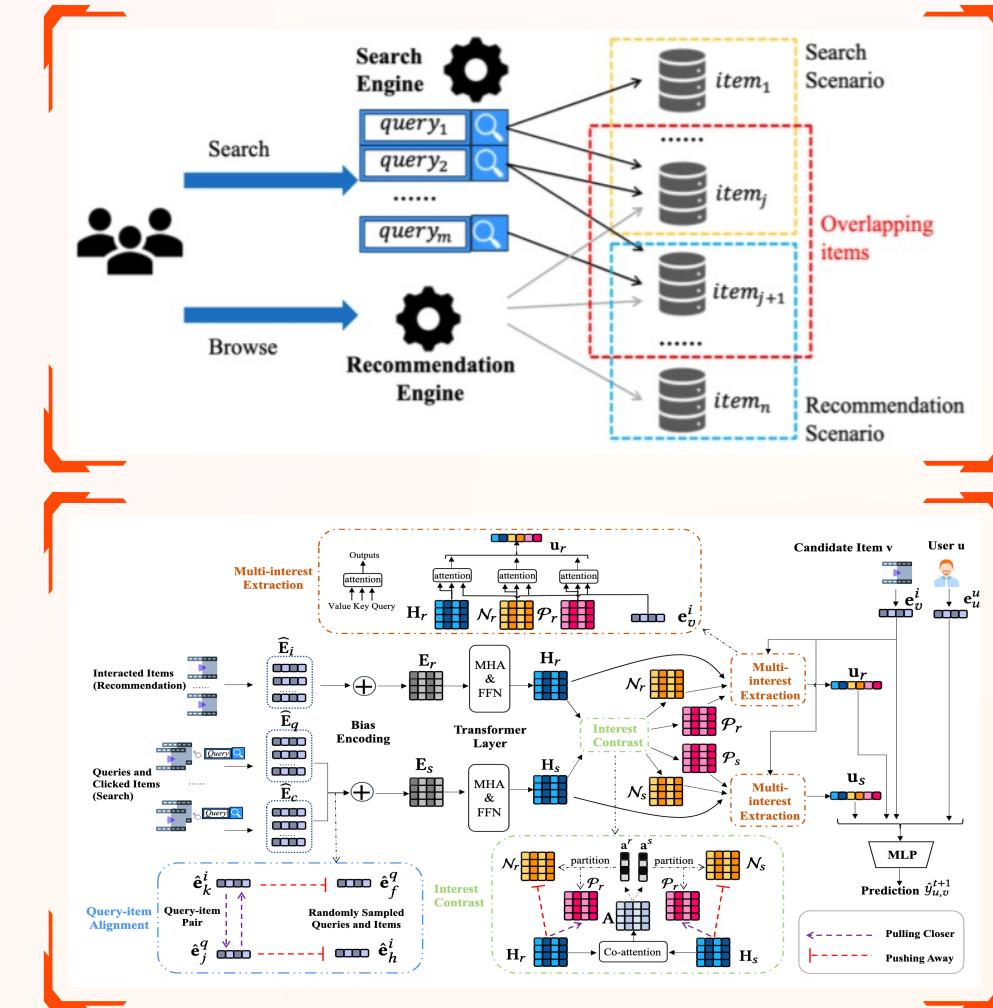
搜推联合：联合建模搜索&推荐兴趣

三Q 搜索促推荐

- 因果推断消除推荐系统bias
- 对比学习区分搜索和推荐兴趣的异同

推荐促搜索

- 序列建模预测搜索行为



A Model-Agnostic Causal Learning Framework for Recommendation using Search Data, WWW' 22.
Enhancing Recommendation with Search Data in a Causal Learning Manner, TOIS 23.

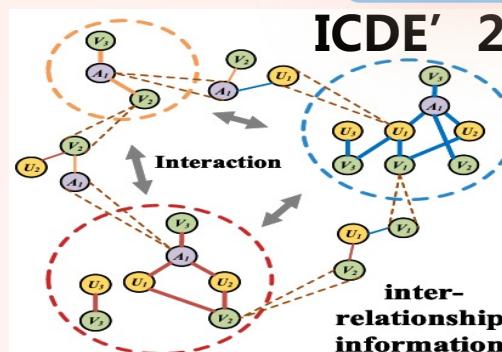
When Search Meets Recommendation: Learning Disentangled Search Representation for Recommendation. Sigir 23.

图学习：多行为异构表征学习

挑战：

视频存在**时效性**，用户喜好**变化快**，
行为存在**噪音**

行为间关系探索

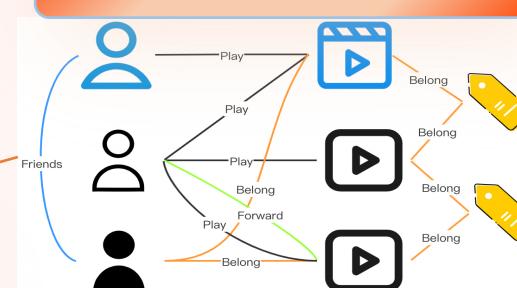


异构图元路径采样

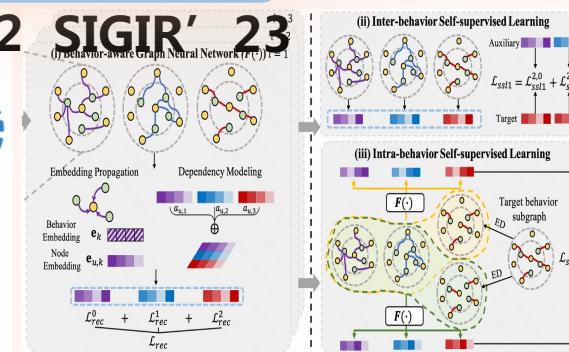
学习混合表征

多行为图对比学习

短视频场景异构图

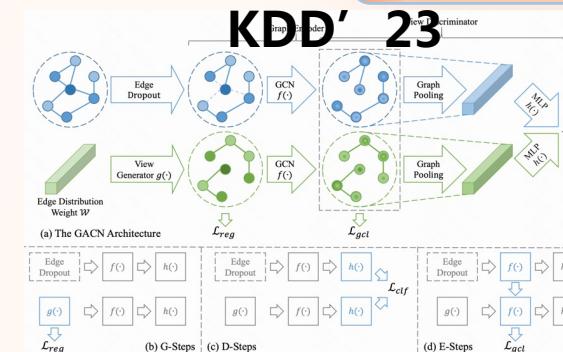


图结构学习



学习混合表征

KDD' 23



基于GAN生成图的对比学习带行为置信度的采样机制

去噪声

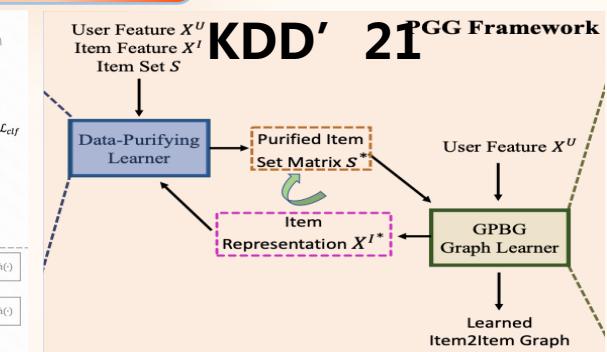
Graph Contrastive Learning with Generative Adversarial Network. KDD' 23
Multi-behavior Self-supervised Learning for Recommendation. SIGIR' 23

Instant Representation Learning for Recommendation over Large Dynamic Graphs. ICDE' 23

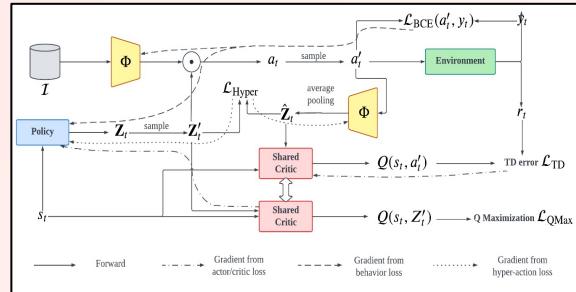
HybridGNN: Learning Hybrid Representation for Recommendation in Multiplex Heterogeneous Networks. ICDE' 22

Purify and Generate: Learning Faithful Item-to-Item Graph from Noisy User-Item Interaction Behaviors. KDD' 21

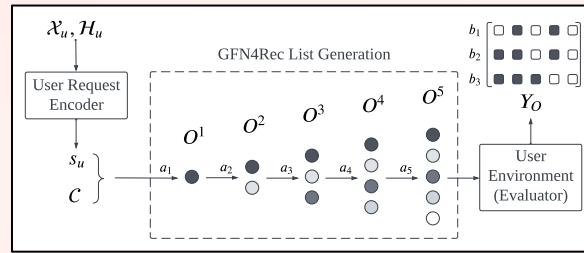
Concept-Aware Denoising Graph Neural Network for Micro-Video Recommendation. CIKM' 21



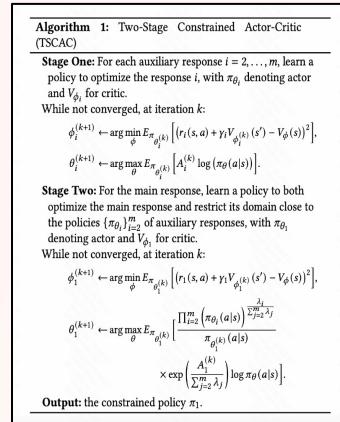
强化学习：可持续推荐满足用户长期交互需求



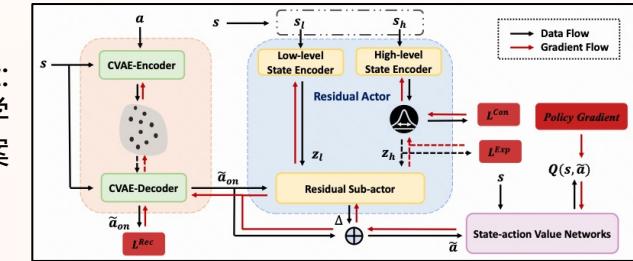
WWW 23'
HAC:
动作隐性表达、探索
与正则化



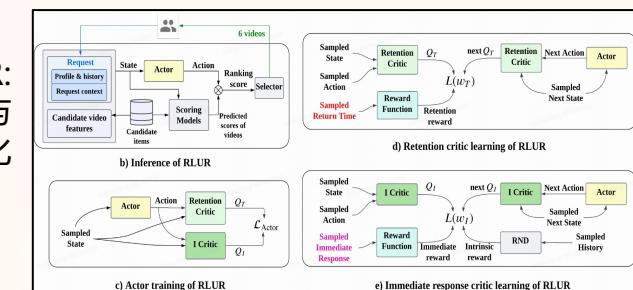
KDD 23'
GFN4Rec:
Listwise 生成流模型



WWW 23'
TSCAC:两阶段解决主辅
目标分歧



ICLR 23'
ResAct:
基于动作重构与残差学
习的离线强化学习方案

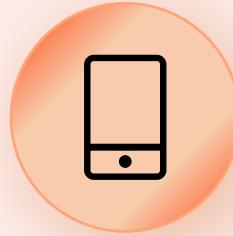


WWW 23'
RLUR:
用户留存建模与
优化

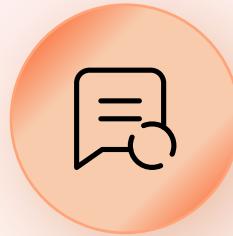
返回时间/用户留存



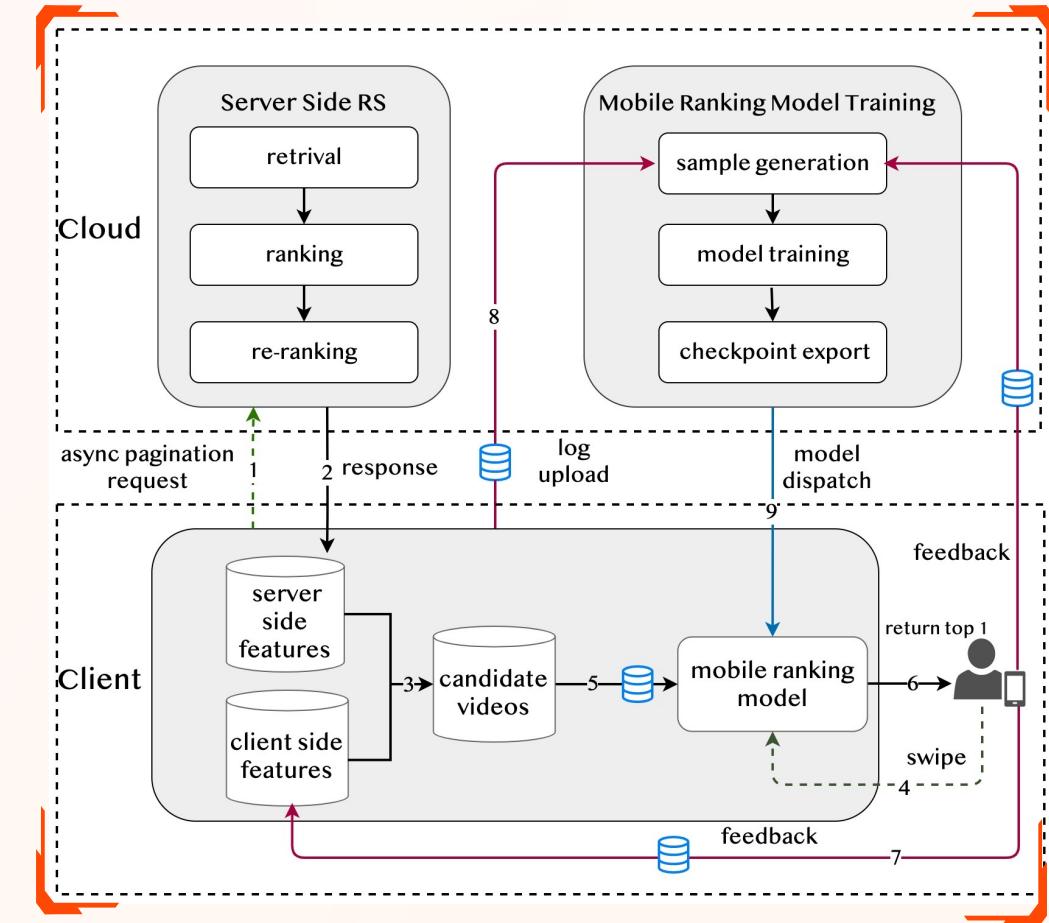
端智能：实时兴趣感知和内容推荐



客户端部署轻量级模型，直接利用端上信号，消除数据延迟



用户显式/隐式行为触发重排，响应实时兴趣变化

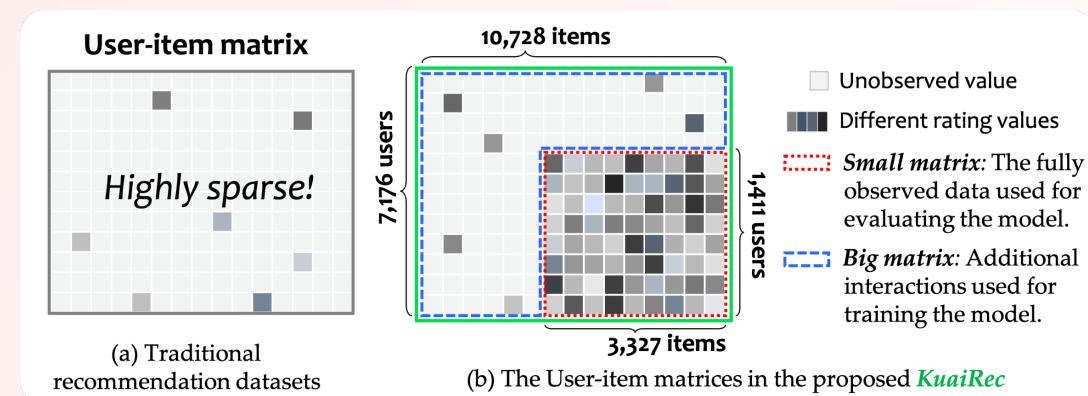


社区贡献-数据

推荐

KuaiRec

首个百万量级交互的密集推荐数据集：
99.6%稠密度



KuaiRand

首个基于随机曝光流量的无偏推荐数据集：2.7万用户，
3千万视频



搜推联合

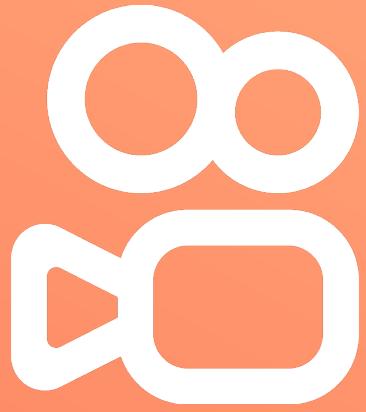
KuaiSAR

首个真实包含搜索和推荐行为的数据集：
2.5万用户，680万视频，1.9千万交互



- ✓ 搜索行为的来源，主动搜/被动搜
- ✓ 搜索和推荐服务之间的转换
- ✓ 用户真实互动标签





打造最有温度
最值得信任的在线社区