



SACC

2020 中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2020

架构融合 云化共建

LIVE 2020年10月22日 - 24日网络直播

推荐系统在同程旅行的应用

数据中心—吴磊

<http://www.itpub.net/>

自我介绍

- 2017年，毕业于华东理工大学 研究方向：统计计算
- 2017-2019年，携程旅行网，酒店研发部 资深算法工程师
- 2019年至今，同程旅行网，数据中心 资深算法工程师

1. **推荐系统应用场景——资源及推荐场景介绍**
2. 推荐系统架构设计——架构及其实时性
3. 推荐系统算法设计——功能模块化

推荐系统应用场景

资源类：

酒店
景点门票
国内长线游
周边游
出境游

交通类：

机票
火车票
汽车票·船票

推荐系统应用场景

中国移动 20:30 57%

搜索目的地/景点/酒店等

您好,宸倚陌

北京

豪华酒店 | 低调奢华,难忘今宵

618 促销

酒店

火车票

机票

汽车·船票

景点门票

民宿公寓

旅游

打车

借钱·理财

全部服务

领红包

每日签到

早起打卡

黑鲸会员

您有一个神秘奖品待领取

点击领取

现金红包、视频VIP卡派送中

精选

周边去哪

亲子旅行

住酒店

出发, 轻松游

特价机票

首页

里程商城

行程

客服

我的

中国移动 19:19 95%

国内/海外酒店

猜你喜欢

高档型

独墅湖书香世家酒店 (苏州东振...)

4.8分 棒极了 4,356条评论

工业园区金鸡湖西/苏州中心 | ...

¥321起

经济型

IU酒店 (苏州独墅湖月亮湾地铁...)

4.7分 棒极了 1,345条评论

工业园区独墅湖高教区 | 月亮...

¥212起

经济型

如家酒店 (苏州独墅湖西交利物...)

4.8分 棒极了 443条评论

工业园区独墅湖高教区 | 独墅...

¥151起

舒适型

苏州工业园区智选假日酒店

4.8分 棒极了 1,236条评论

工业园区金鸡湖东/诚品书店 | ...

¥406起

舒适型

全季酒店 (苏州独墅湖店)

4.6分 棒极了 719条评论

工业园区独墅湖高教区 | 独墅...

¥381起

舒适型

周庄漫行居花园美宿

我的订单

支付·退改·点评·返现

我的酒店

收藏·历史·问答

中国移动 16:22 89%

搜一搜, 带你玩转周边

上海

猜你喜欢

上海海洋水族馆

带领孩子探索海洋奥秘

海洋馆 | 上海, 浦东新区

¥105起 ¥110

上海野生动物园

走进动物王国, 与野生动...

动物园 | 上海, 浦东新区

¥8起 ¥19

上海自然博物馆 (上海科技馆分馆)

给予孩子探索自然的空间...

博物馆 | 上海, 黄浦区

¥9.9起 ¥99

上海中心大厦上观光厅

登天空之城, 看魔都的上...

展览 | 上海, 浦东新区

¥9.9起 ¥99

SACC 2020

中国系统架构师大会

中国联通 0.00K/s 16:22

我的订单

点击“...”添加到我的小程序, 查单更方便

全部订单 待付款 待出行 待点评

预订(近-远) 出发(晚-早) 筛选 开发票

特价机票

上海—深圳 10月19日 周六	¥910	无锡—广州 10月19日 周六	¥900
上海—重庆 10月19日 周六	¥660	上海—广州 10月19日 周六	¥830
上海—西安 10月19日 周六	¥520	无锡—昆明 10月19日 周六	¥960

特价酒店

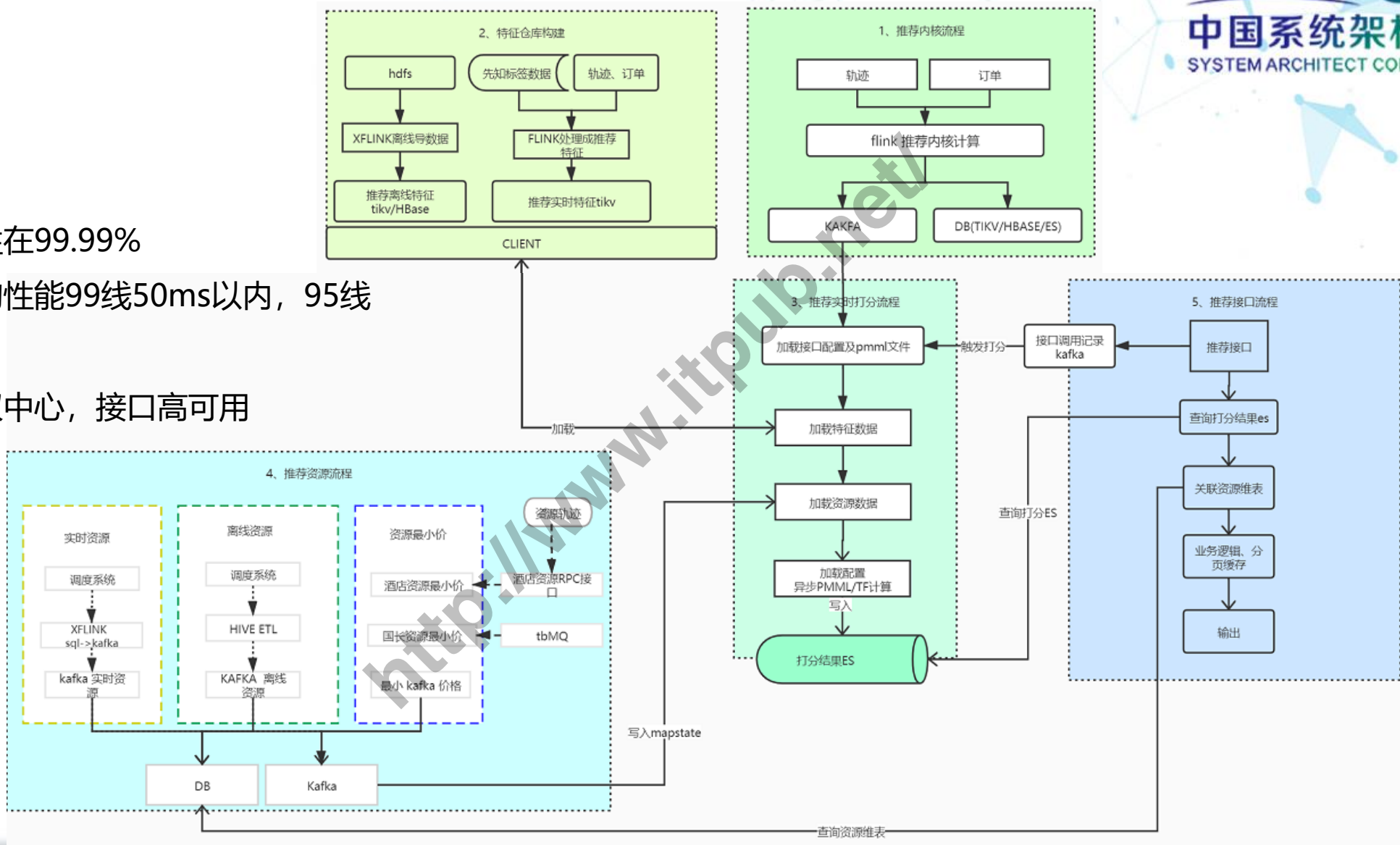
荆州金九龙大酒店 荆州市·荆州区·荆州 高档型 评论数: 205 评分: 4.4	¥408
荆州市·荆州区·万达广场 舒适型 评论数: 507 评分: 4.3	¥278

1. 推荐系统应用场景——资源及推荐场景介绍
- 2. 推荐系统架构设计——架构及其实时性**
3. 推荐系统算法设计——功能模块化

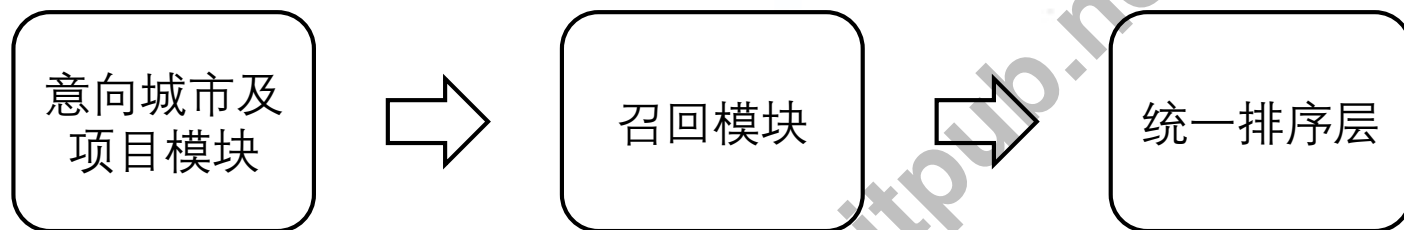
推荐系统架构设计

整体目标:

- 1. 接口稳定性在99.99%
- 2. 接口的查询性能99线50ms以内, 95线20ms以内
- 3. 接口支持双中心, 接口高可用



1. 推荐系统应用场景——资源及推荐场景介绍
2. 推荐系统架构设计——架构及其实时性
- 3. 推荐系统算法设计——功能模块化**



- 意向城市及项目由位点（详情页看了又看、订单页面交叉推荐）或算法确定
- 统一排序层提供对外统一排序功能，供业务直接调用，定制化场景另开发
- 各模块独立运行，功能上透明

推荐系统算法设计——意向识别



意向识别：给用户推什么资源？推哪里？

大首页九宫格

30万酒店5折起

出行领好礼
人人可以领

酒店

火车票

机票

汽车·船票

景点门票

民宿公寓

旅游

打车

借钱·理财

更多服务

领红包

每日签到

早起打卡

黑鲸会员

精选

十一去哪玩

自驾游

亲子时

去哪玩 去哪住

上海-北京

¥285起

首页

里程商城

行程

客服

我的

大首页搜索

北京

北京

北京站 T2航站楼... 北京南站 T3航站楼... 故宫 北京协和... 天安门 301医院

北京旅游产品·北京

北京热门景区·北京

上海到北京的全部交通

北京站附近的酒店·北京

频道首页搜索

上海 北京

09月29日 周二

经济舱

搜索

上海-北京往返机票仅需782元起

列表页曝光

上海野生动物园 5A

上海海昌海洋公园 4A

上海中心大厦上海之巅... 4A

上海海洋水族馆 4A

详情页曝光

杭州西湖慢享主题酒店

4.9

1365条点评

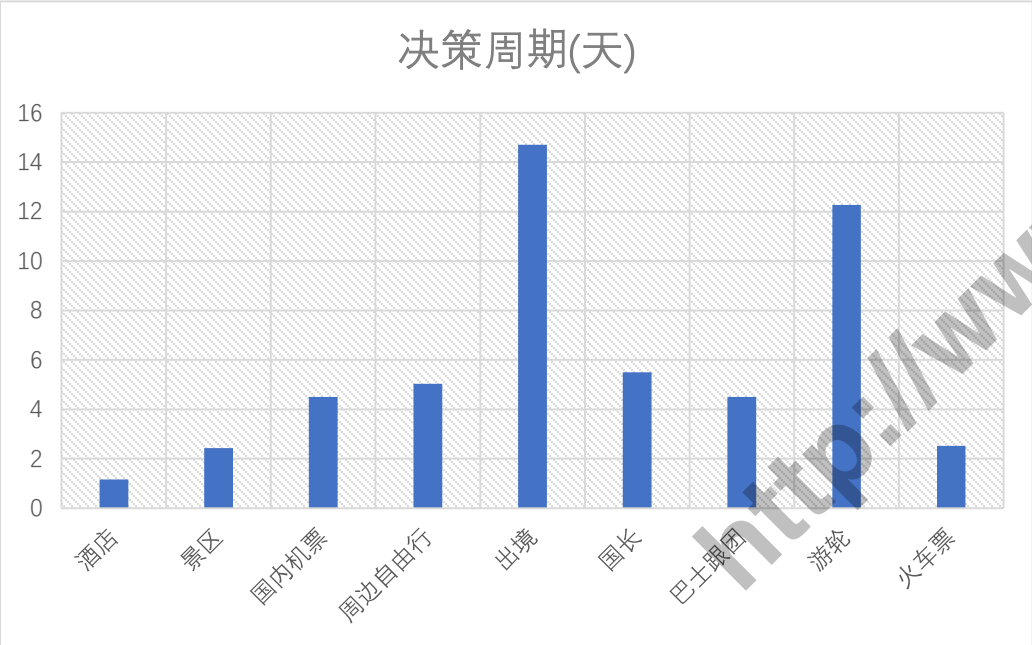
西湖湖滨商圈

预订 金卡特权 酒店享9折起

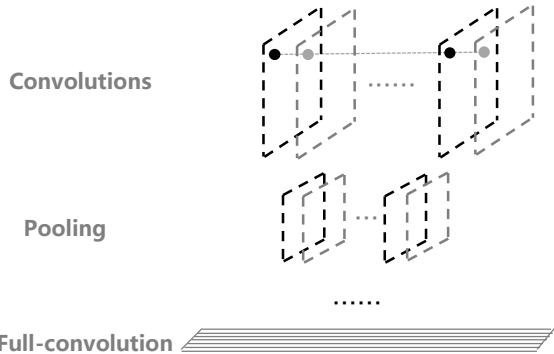
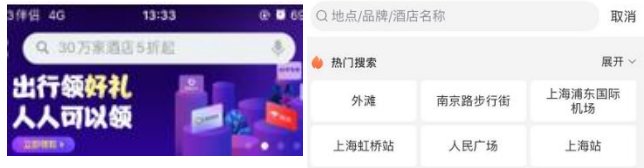
9月28日 今天入住 9月29日 明天离店 共1晚

推荐系统算法设计——意向识别

样本标签构造：用户意向是有一个决策周期的，针对不同的资源，该决策周期也不同；在利用订单构建标签时需要设定不同的时间窗口



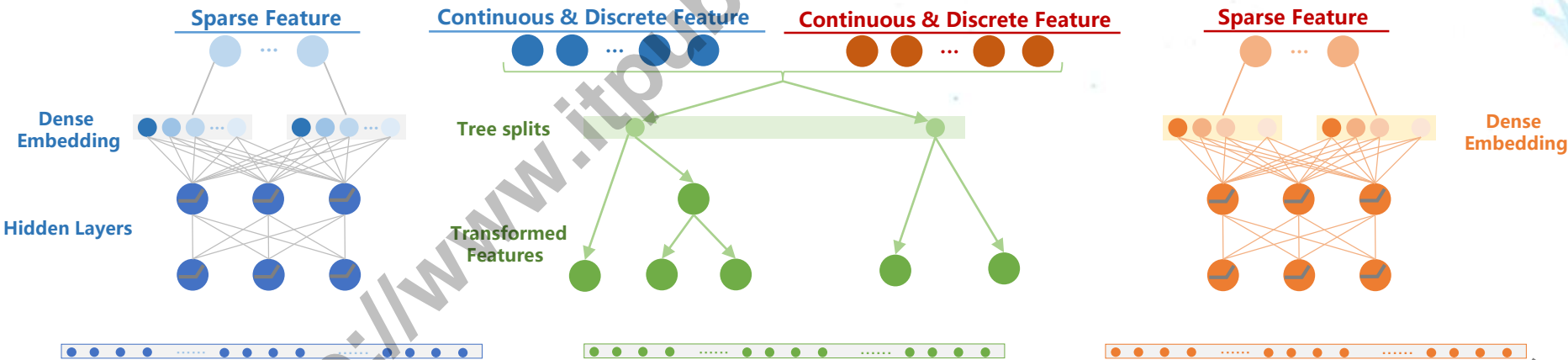
- 为保证公共位点在各意向资源的可比性，采用了统一建模打分的方式；
- 意向模型的主键为：
 $uid + productid + cityid$
- 负样本如何采样？



资源维度信息

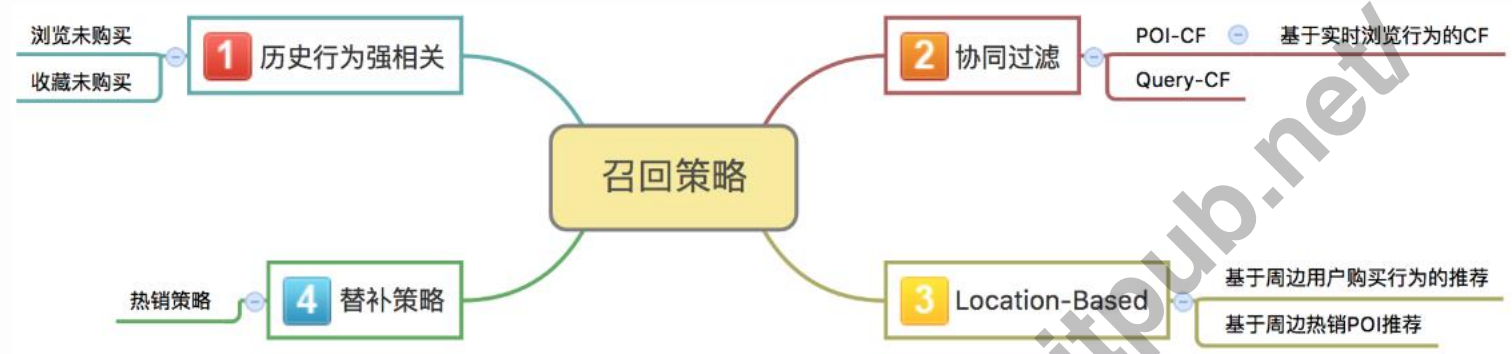
用户-资源Pair

用户维度信息



Layer

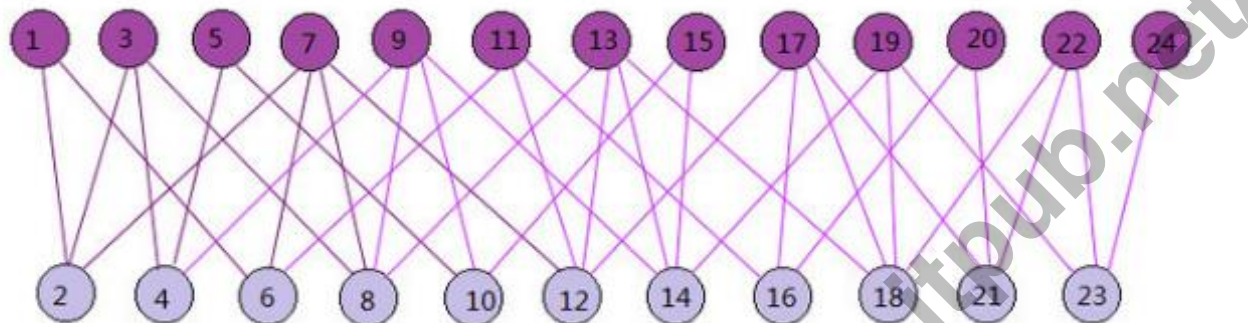
OUTPUT



召回策略	超参数设置
历史行为强相关	20
协同过滤	20
周边策略	30
热销策略	50

推荐系统算法设计——资源相似度

SimRank算法：取过去 7 天用户点击的有效资源详情页数据做二部图，基于Spark Graph框架做迭代



(1) 当 $a = b$ 时, $s(a, b) = 1$.

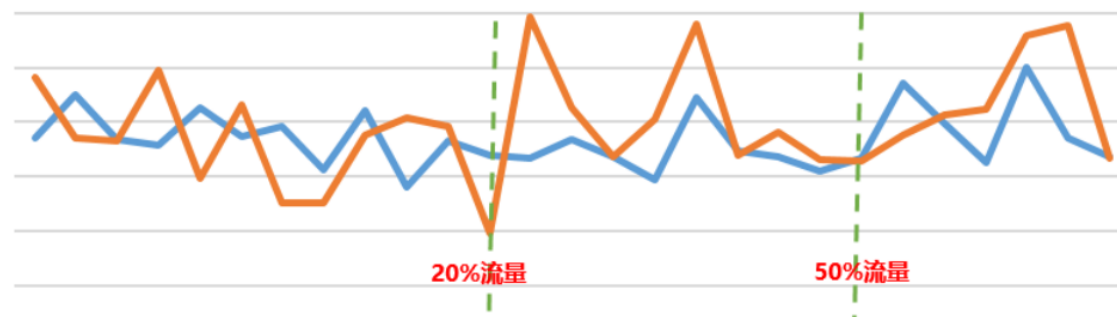
(2) 当 $\mathcal{I}(a) = \emptyset$ 或者 $\mathcal{I}(b) = \emptyset$ 时, $s(a, b) = 0$.

(3) 其他情况下,

$$s(a, b) = \frac{C}{|\mathcal{I}(a)| |\mathcal{I}(b)|} \sum_{i=1}^{|\mathcal{I}(a)|} \sum_{j=1}^{|\mathcal{I}(b)|} s(\mathcal{I}_i(a), \mathcal{I}_j(b))$$

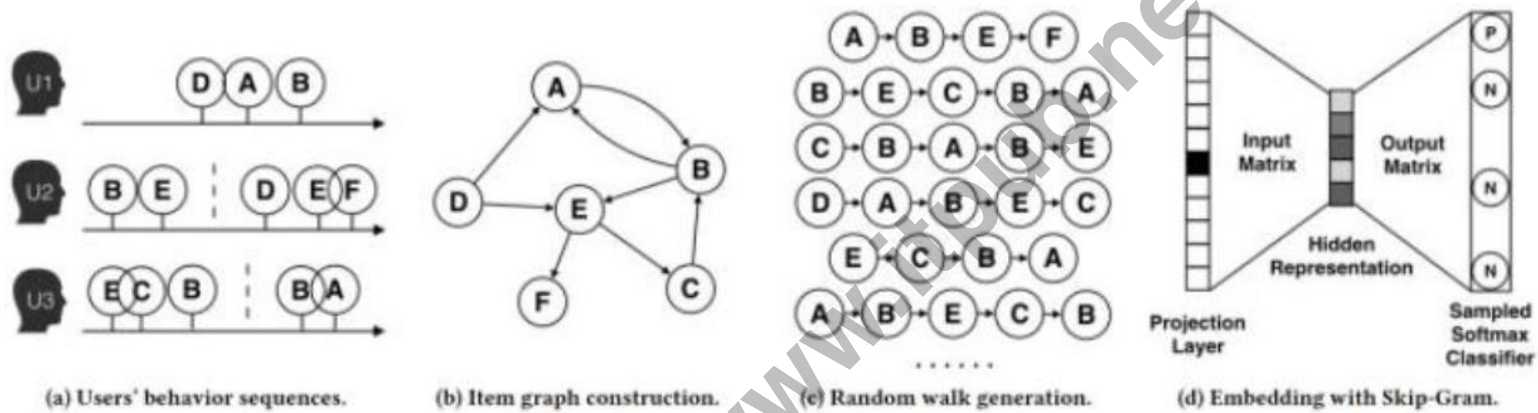
其中, $0 < C < 1$ 是一个阻尼系数.

AB实验：对比基于物品的协同过滤，点击率提升17.1%



推荐系统算法设计——资源相似度

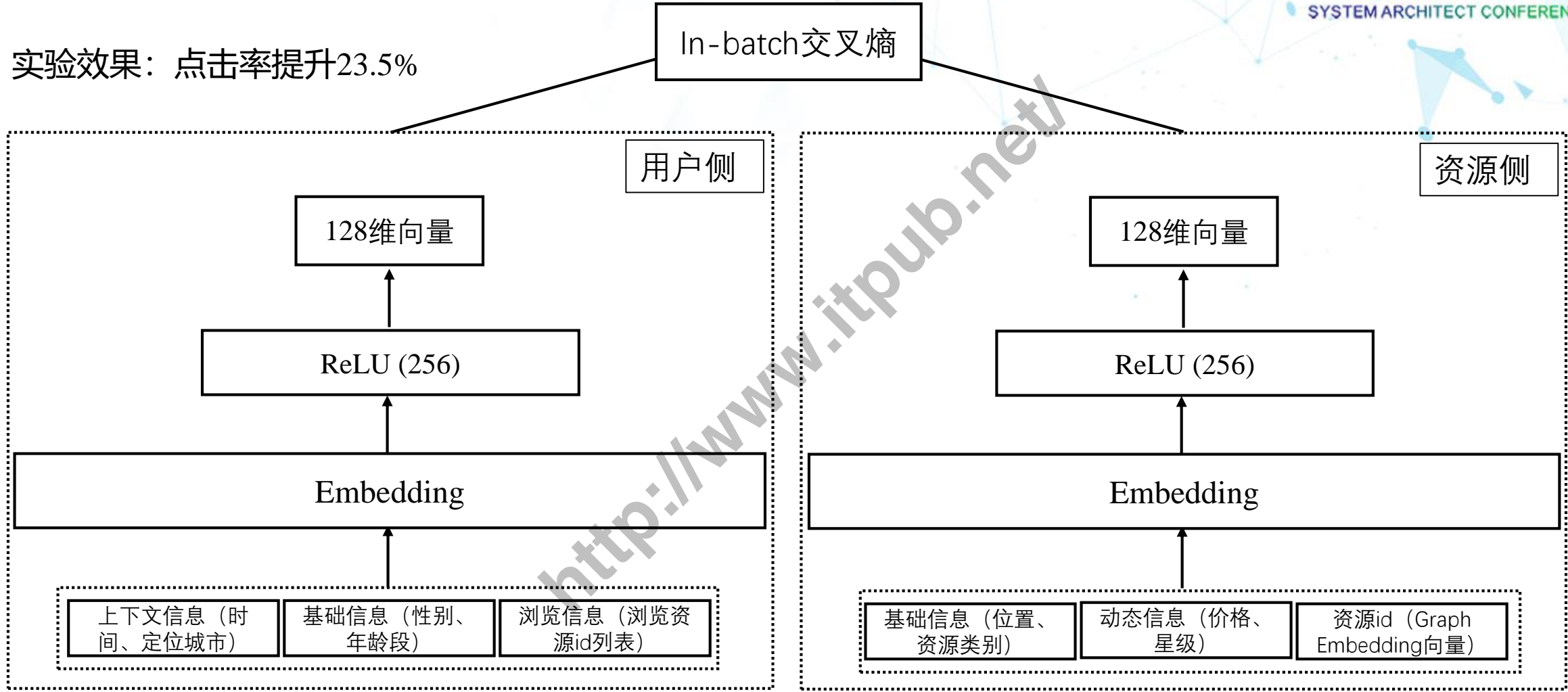
Graph Embedding算法：取用户过去 7 天详情页浏览数据， 构建资源关系的图表示， 通过随机游走算法（step=12， time=30） 采样获取资源的分布。 采用 Item2Vec 算法计算资源的向量表示



algorithm	label	precision	recall	cover	algorithm	label	precision	recall	cover
SimRank	top 5	0.46%	1.33%	3952	SimRank	top 20	0.50%	5.88%	13940
GraphEmbedding	top 5	1.03%	3.03%	3590	GraphEmbedding	top 20	0.56%	6.61%	11860
SimRank	top 10	0.58%	3.40%	7545	SimRank	top 50	0.44%	12.86%	28952
GraphEmbedding	top 10	0.79%	4.61%	6611	GraphEmbedding	top 50	0.30%	8.79%	24458

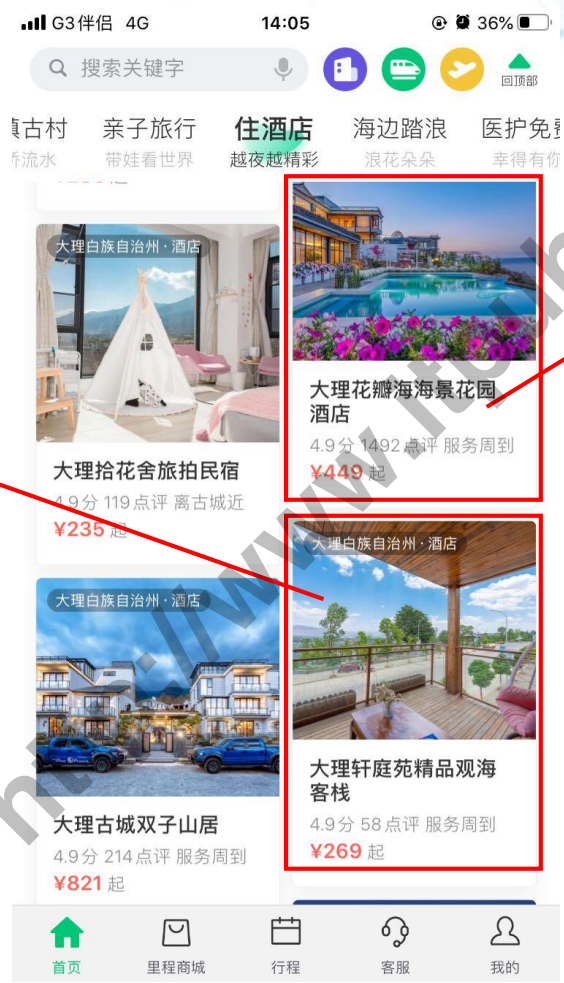
推荐系统算法设计——双塔召回

实验效果：点击率提升23.5%



推荐系统算法设计——列表排序

样本标签构造：同一次请求下，根据曝光、点击、下单作三级相关度标签



推荐系统算法设计——列表排序

特征工程：

用户维度特征：

- 过去1年用户购买资源的平均价格、平均星级
- 用户点击过资源的平均价格、星级

资源维度特征：

- 星级、价格、位置、点评分信息
- 过去7天的点击/转化率
- 过去7天曝光量

用户-资源交叉特征：

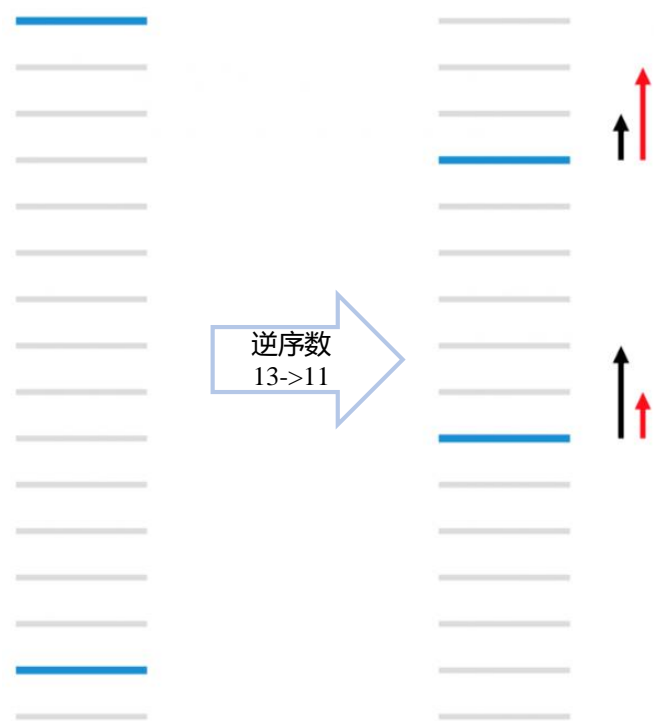
- 过去1小时该资源点击次数
- 该资源与用户点击过资源的价格比、星级比

上下文场景特征：

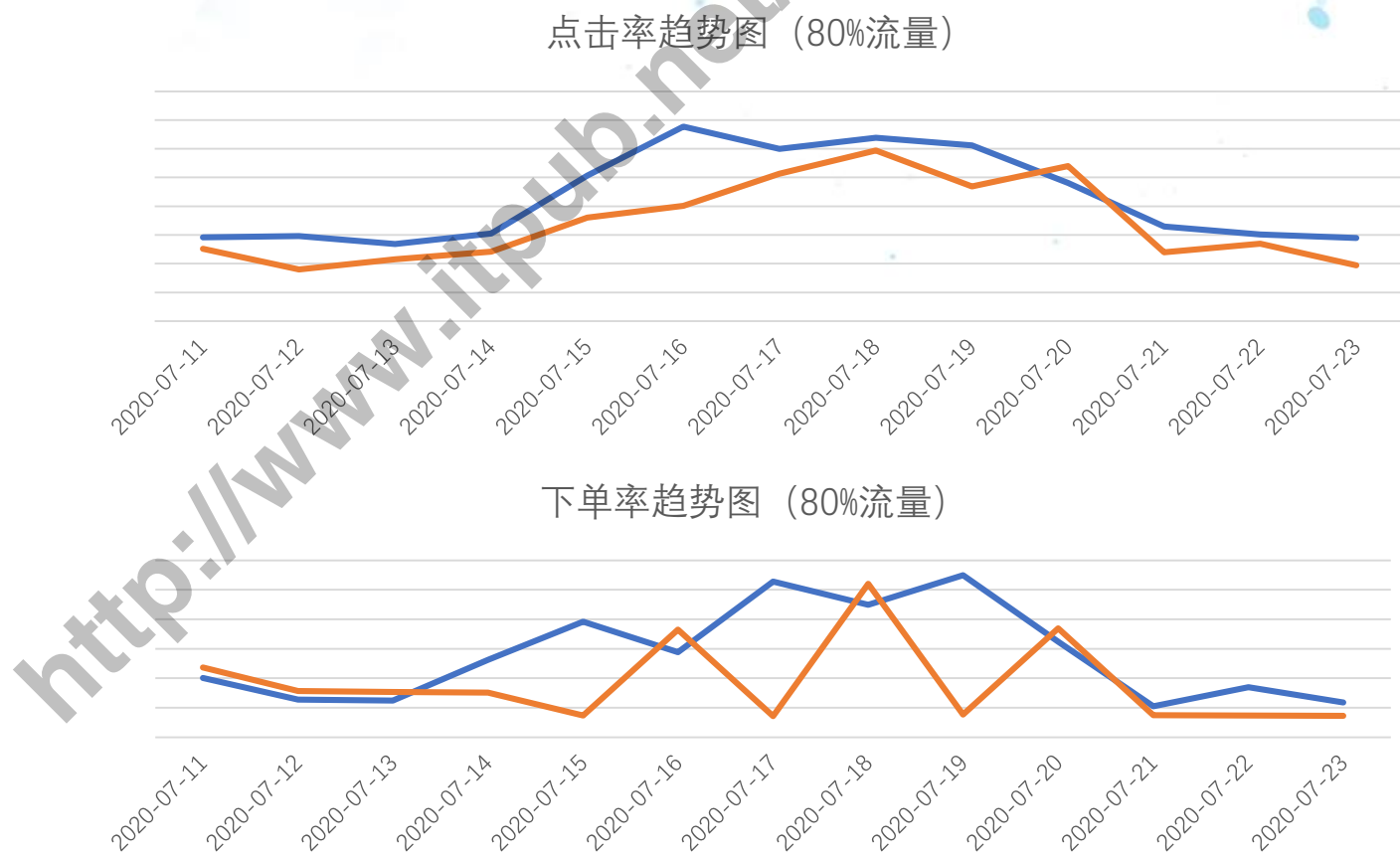
- 提前预定天数
- 当前时间、位置

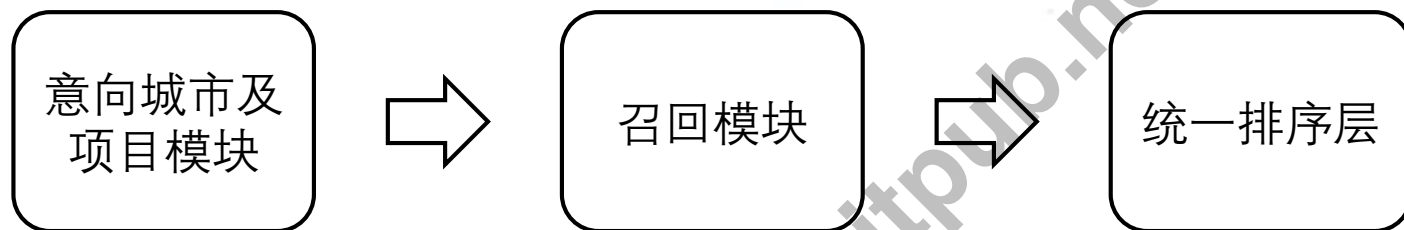
推荐系统算法设计——列表排序

LambdaMART：以逆序对数为优化目标没有考虑位置的权重，这种优化方式对AUC这类评价指标比较友好，但实际的排序结果更加注重头部的相关性，排序指标应选用NDCG这一类指标



注：假如经过两轮迭代实现上图所示的顺序优化





- 意向城市及项目由位点（详情页看了又看、订单页面交叉推荐）或算法确定
- 统一排序层提供对外统一排序功能，供业务直接调用，定制化场景另开发
- 各模块独立运行，功能上透明

Thanks

<http://www.itpub.net/>

