

机器学习在贝壳用户画像 上的实践与挑战

TGO 鲲鹏会

汇聚全球科技领导者的高端社群

📍 全球12大城市

👤 850+ 高端科技领导者

使命
Mission

为社会输送更多优秀的
科技领导者

愿景
Vision

构建全球领先的有技术背景
优秀人才的学习成长平台



扫描二维码，了解更多内容

关于我



郭凯

贝壳找房 数据智能中心-用户策略

阿里巴巴 阿里妈妈-自动机制与智能创意

搜狗科技 营销事业部

哈尔滨工业大学 硕士

1

用户画像概述

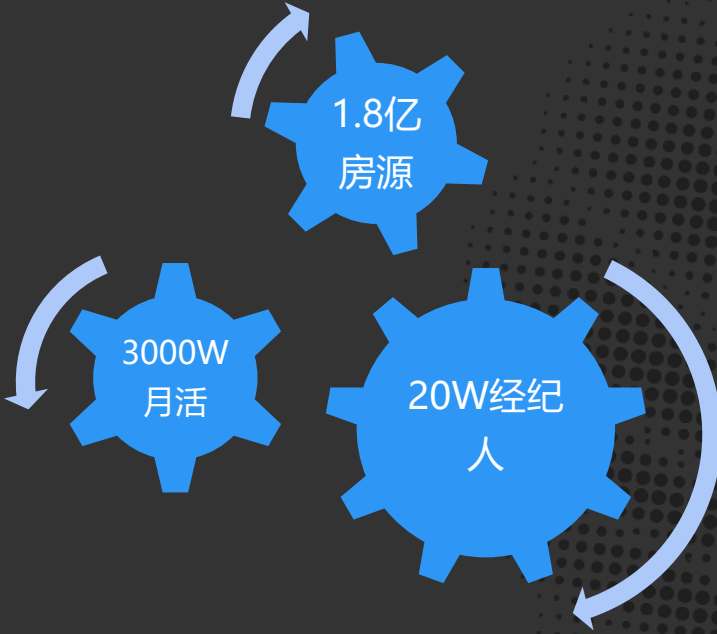
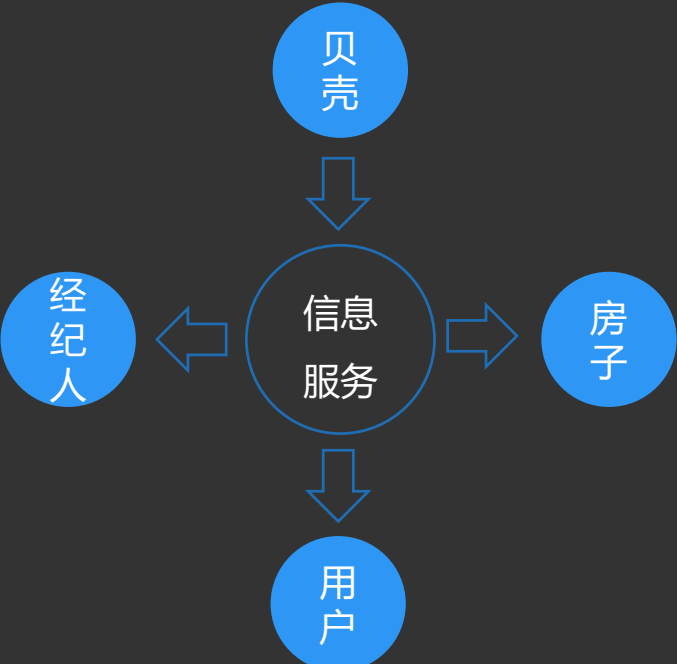
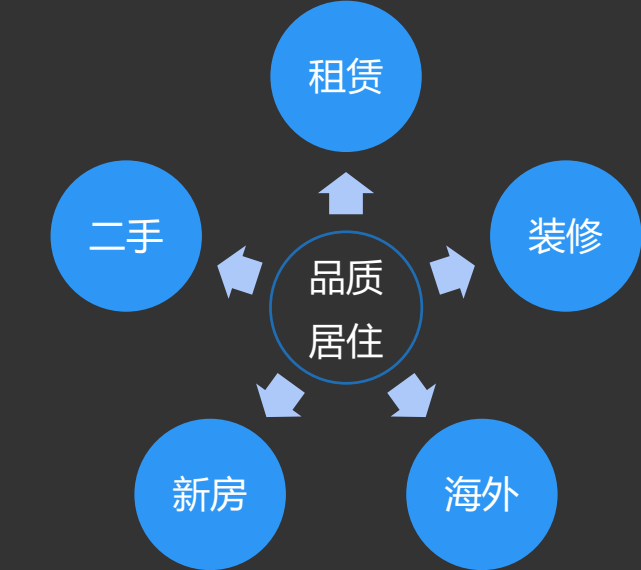
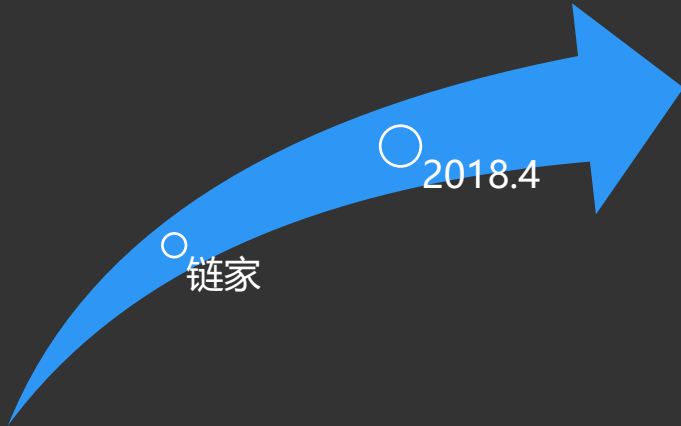
2

贝壳用户画像实践

3

赋能业务应用

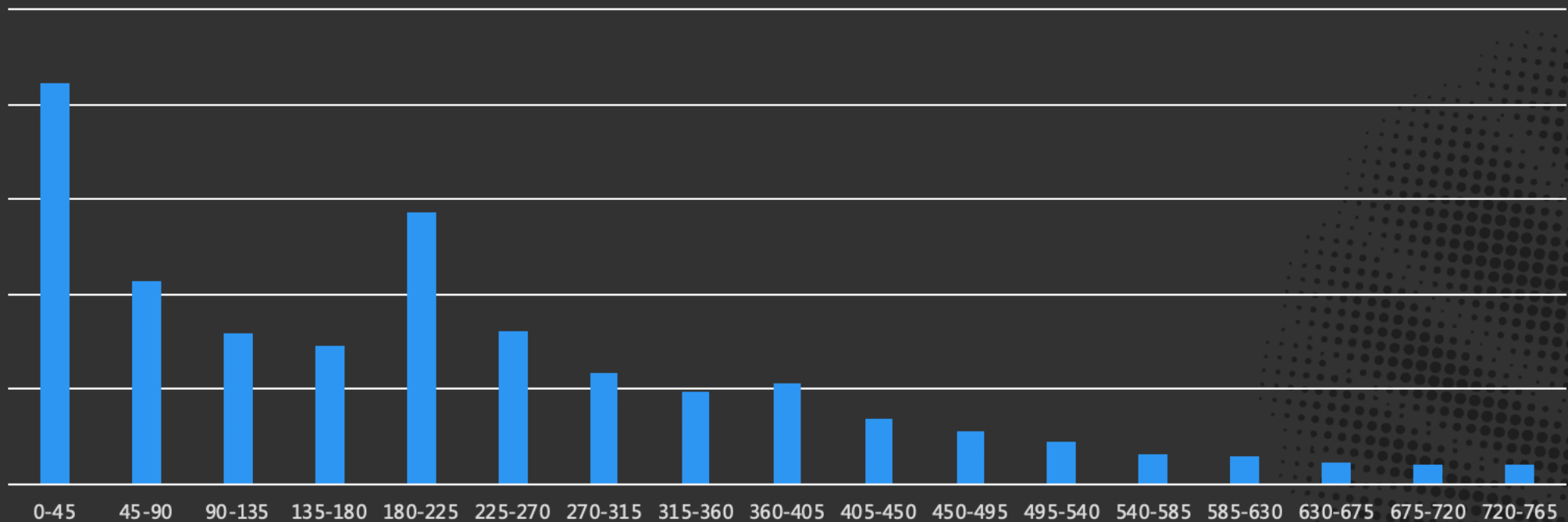
关于贝壳



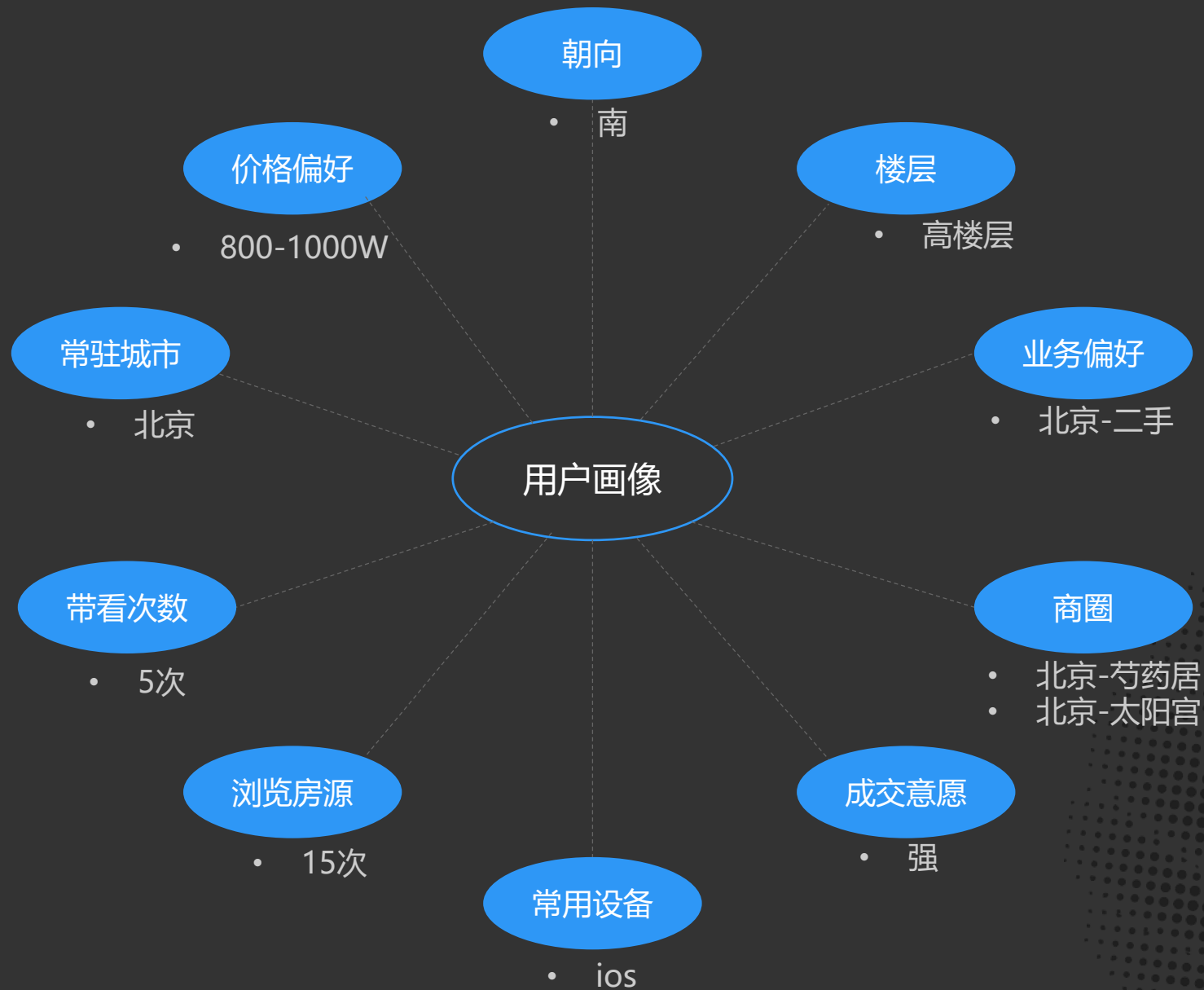
数说用户



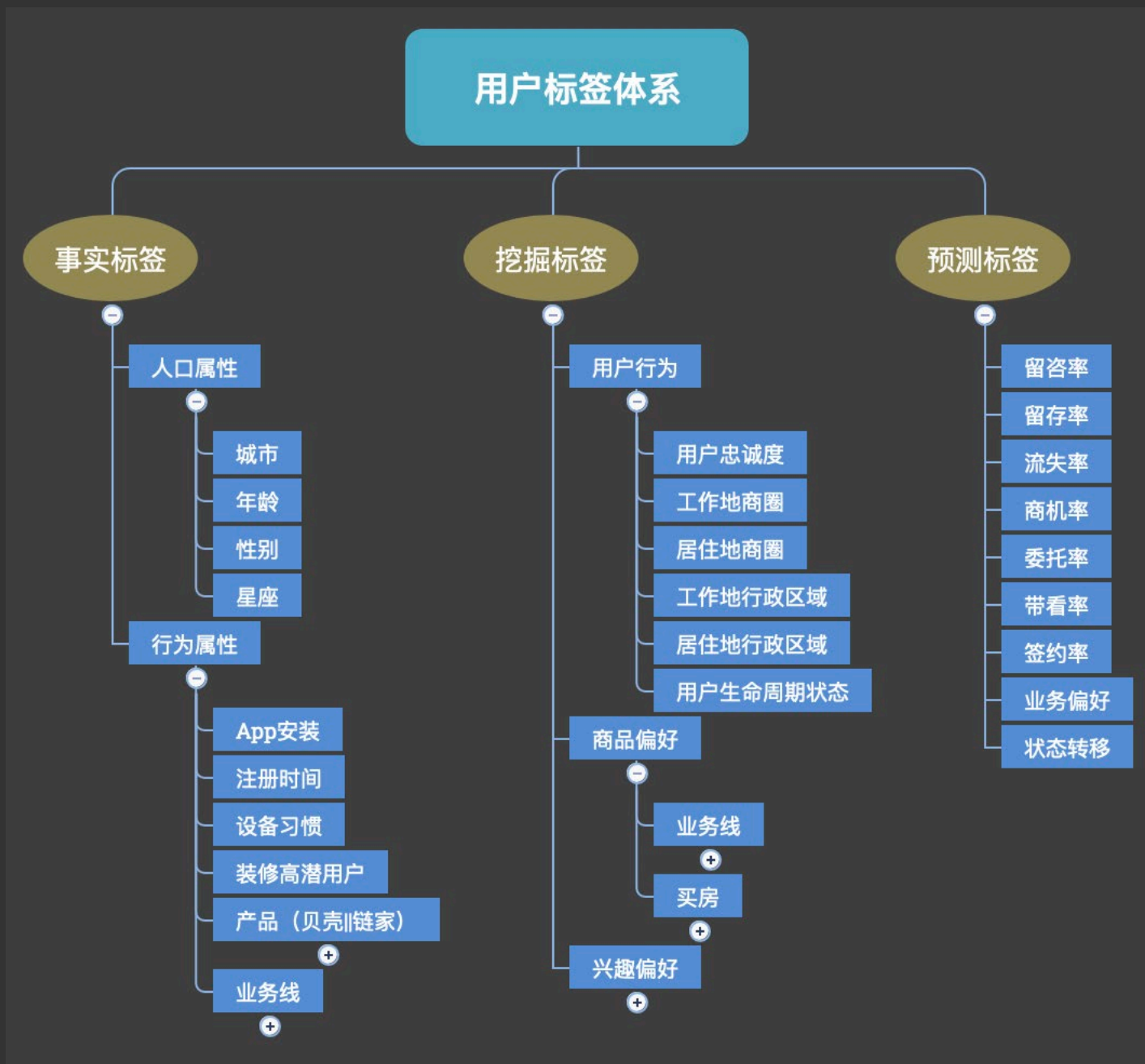
成交用时分布



画的什么?



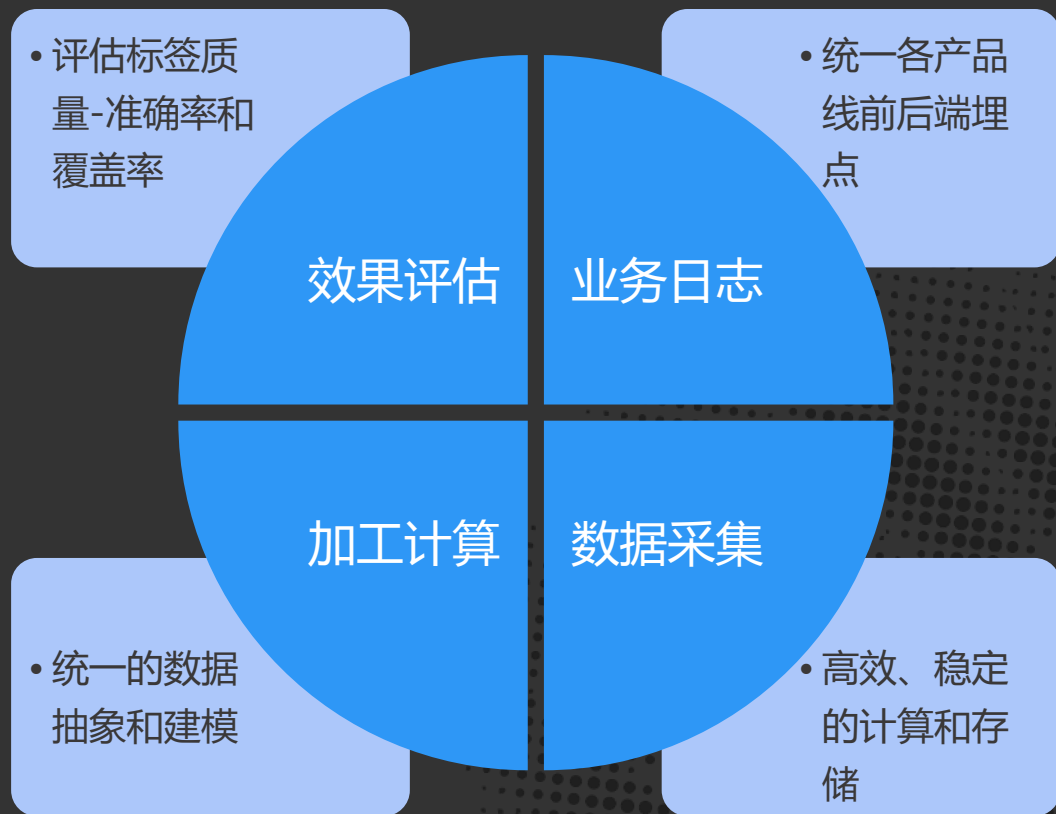
标签体系



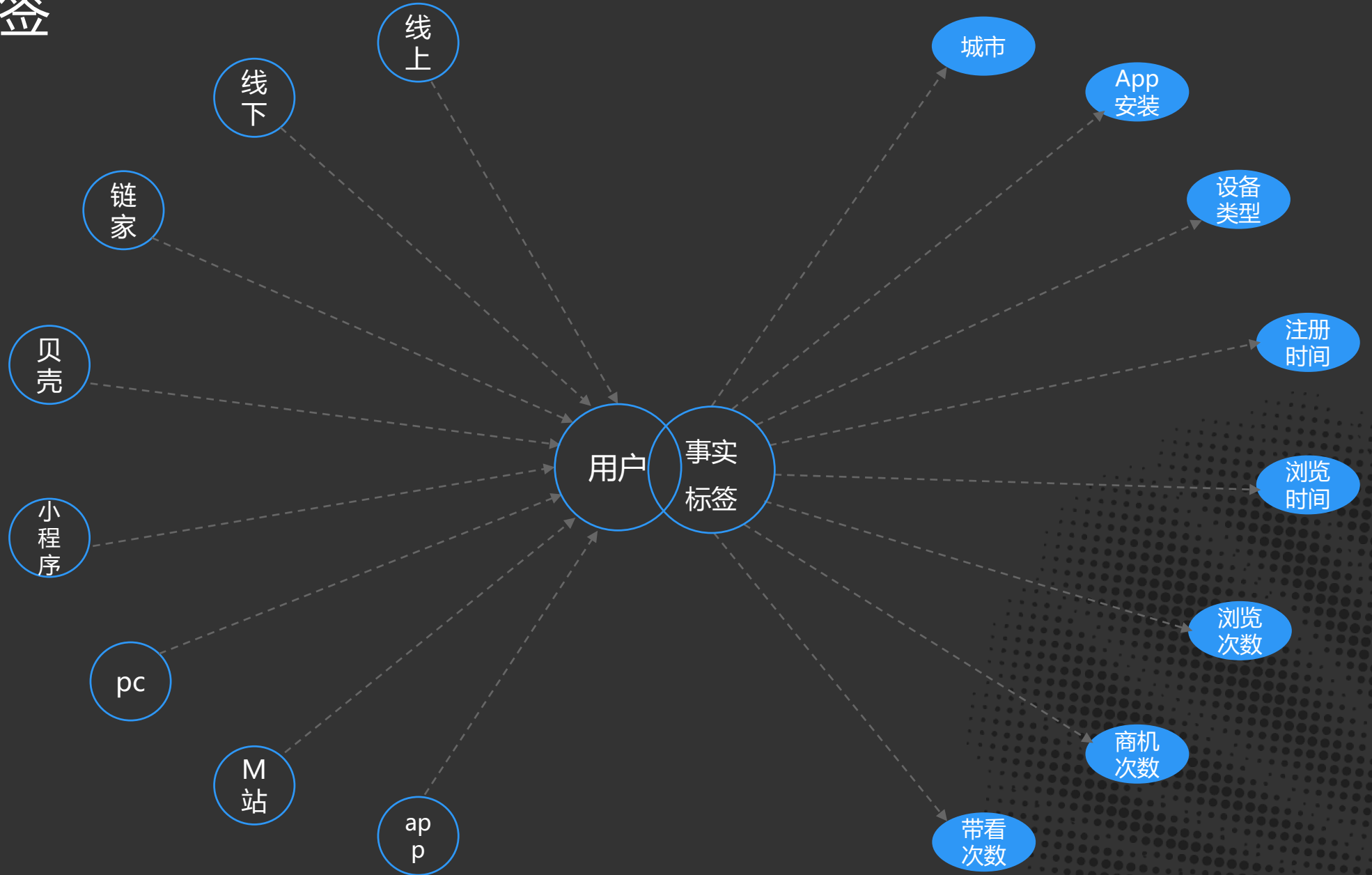
关键难点



◆ 精准刻画用户关键难点



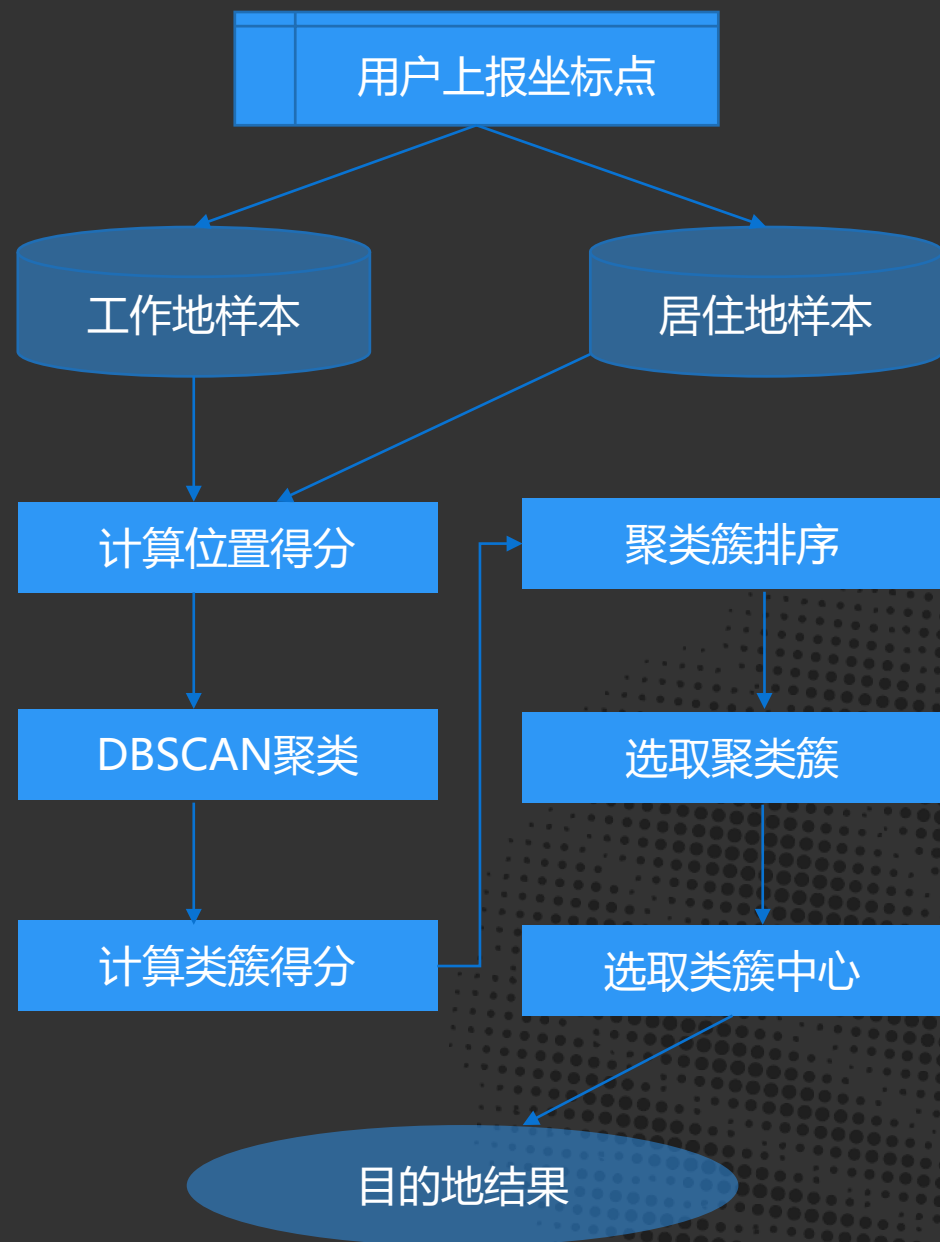
事实性标签



挖掘标签-工作地&居住地

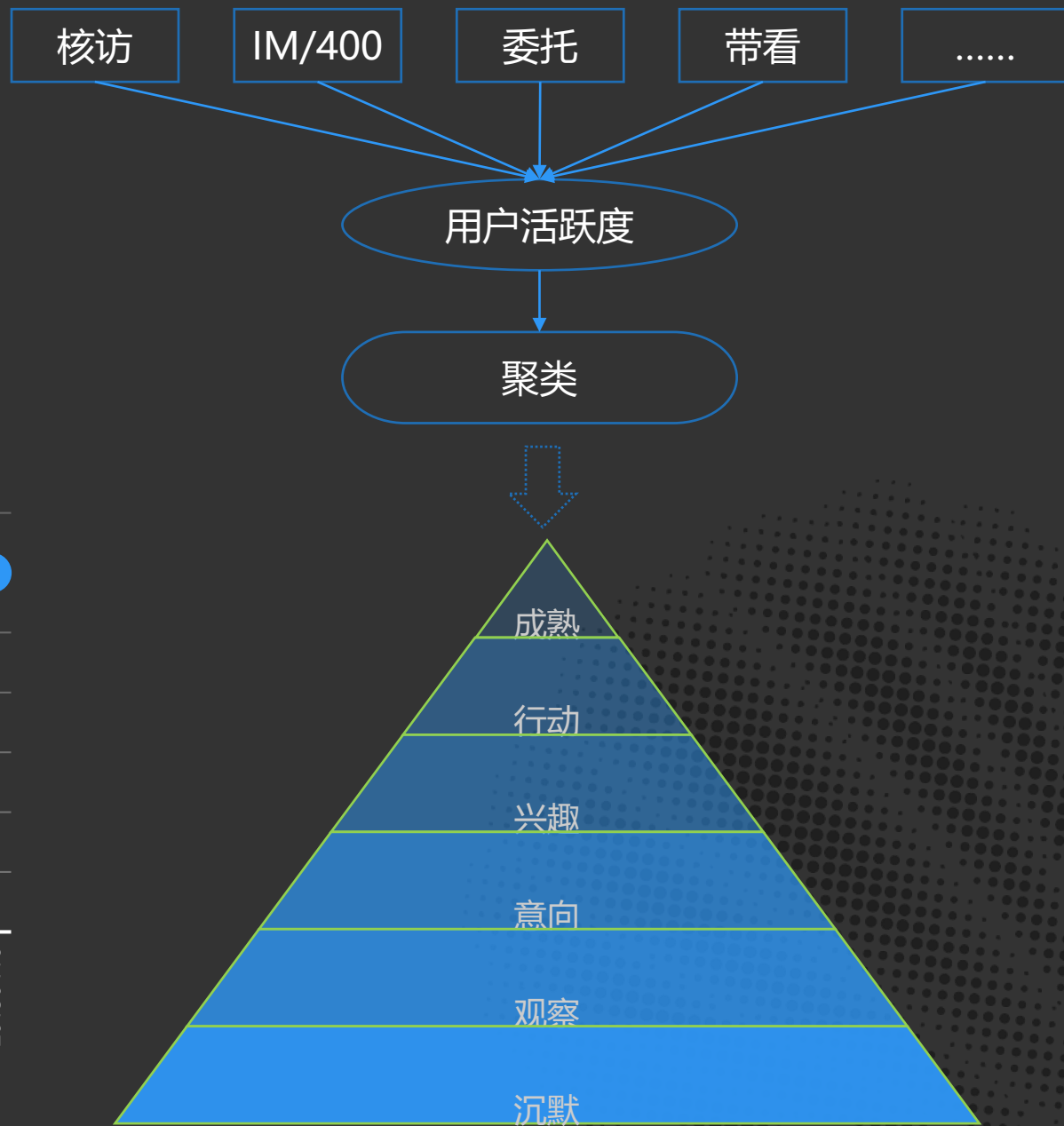
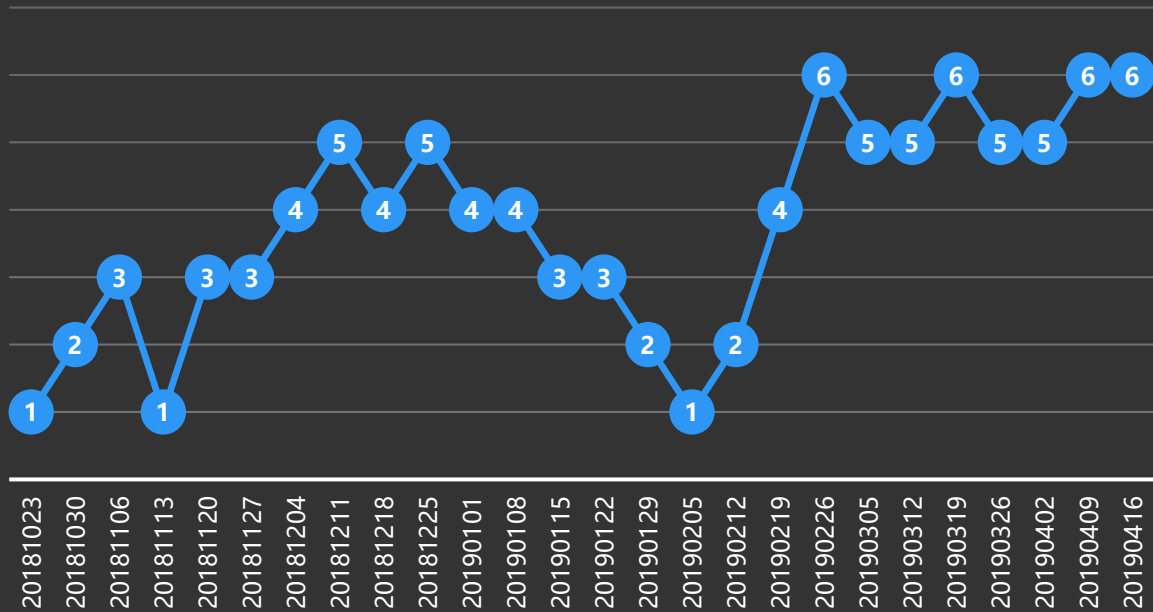


◆ 北京地区覆盖率78%，准确率72%

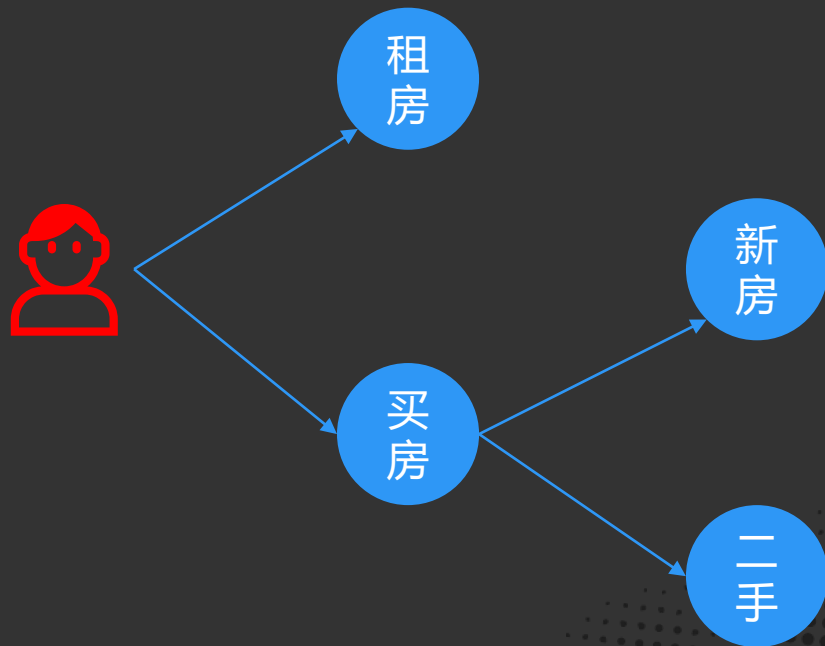
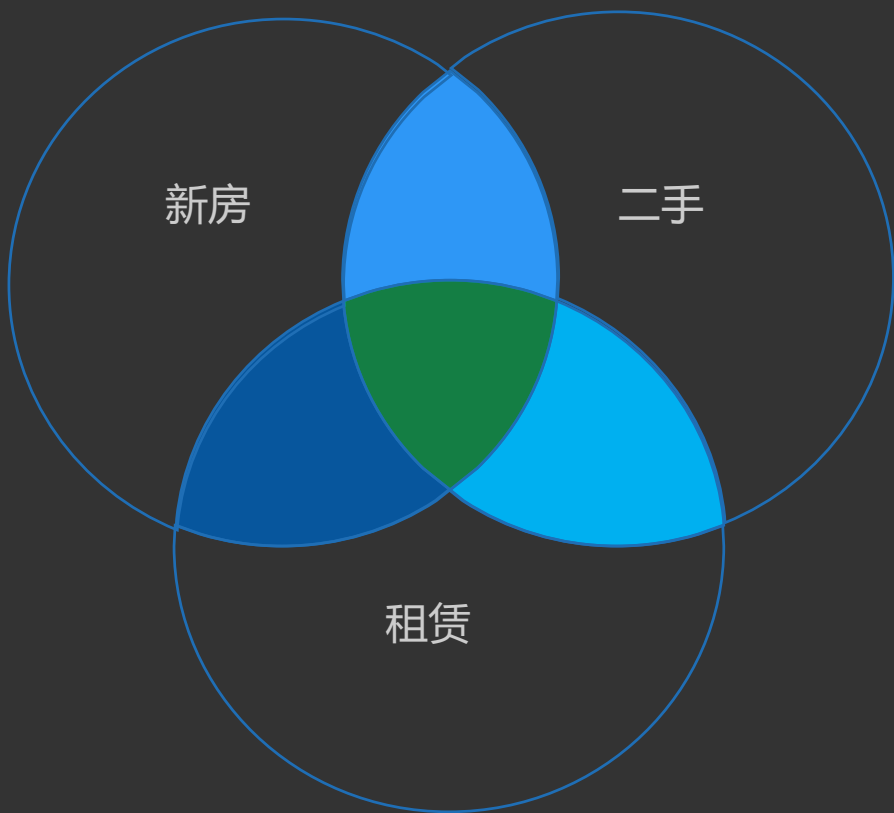


挖掘标签-用户活跃度

- 购房意愿会波动
- 意愿的波动会体现在行为的变化上
- 通过短期行为刻画当前意愿

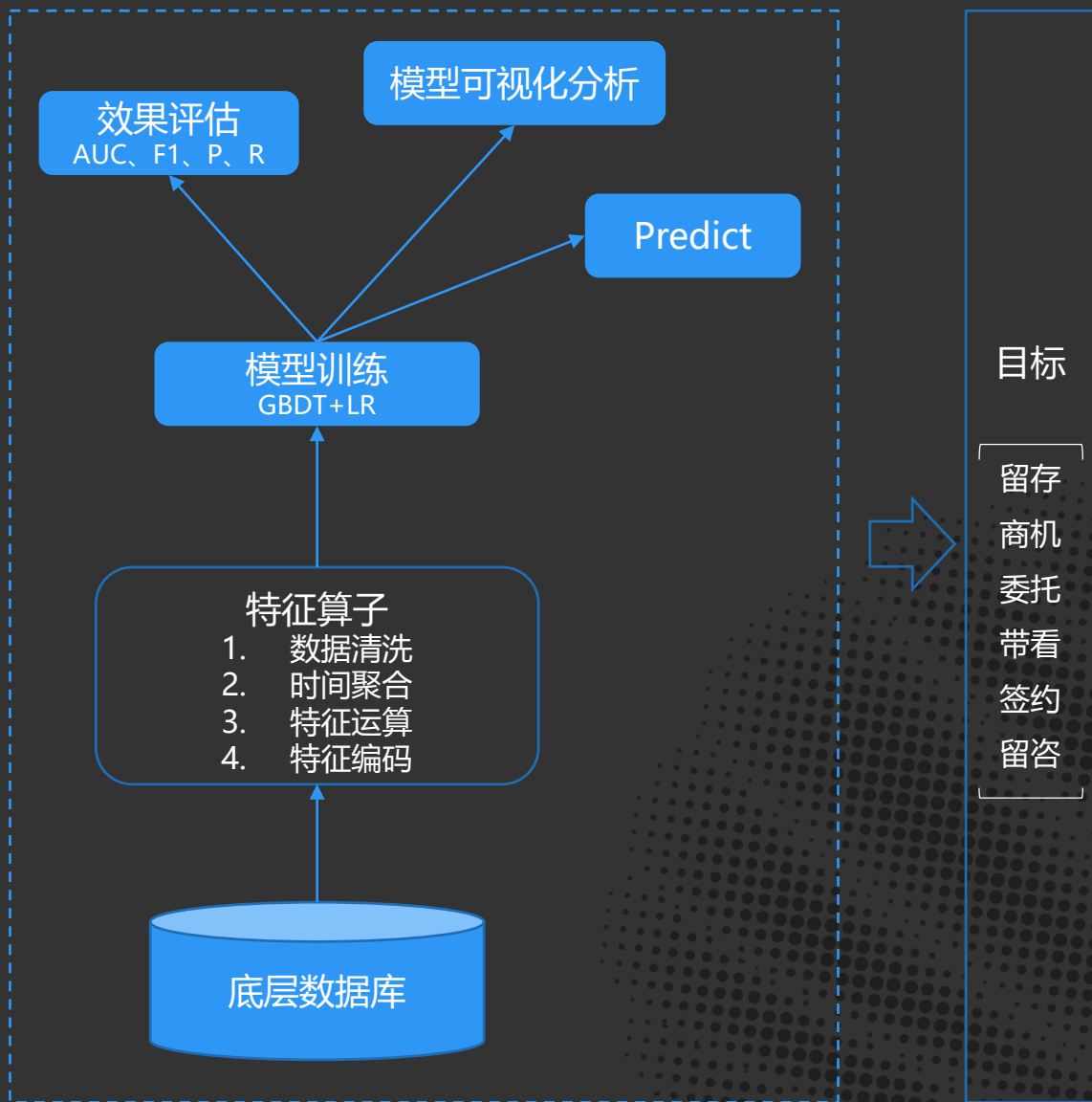
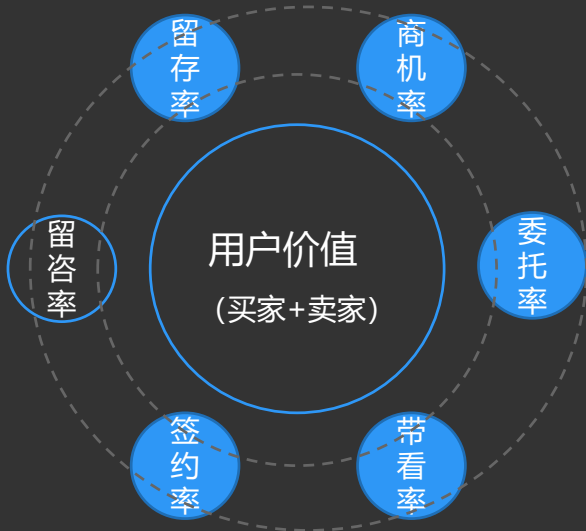
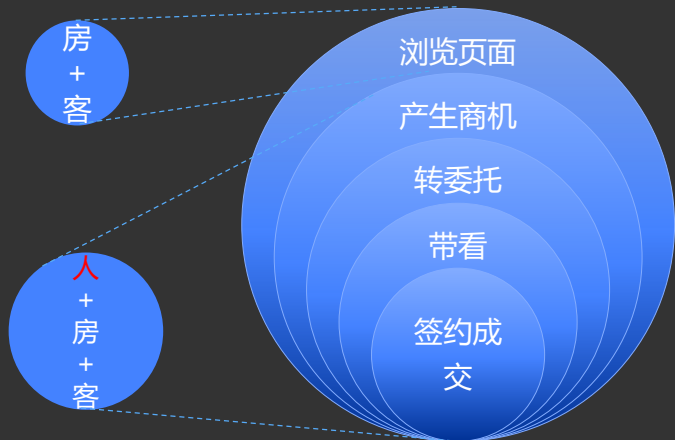


预测标签-业务偏好

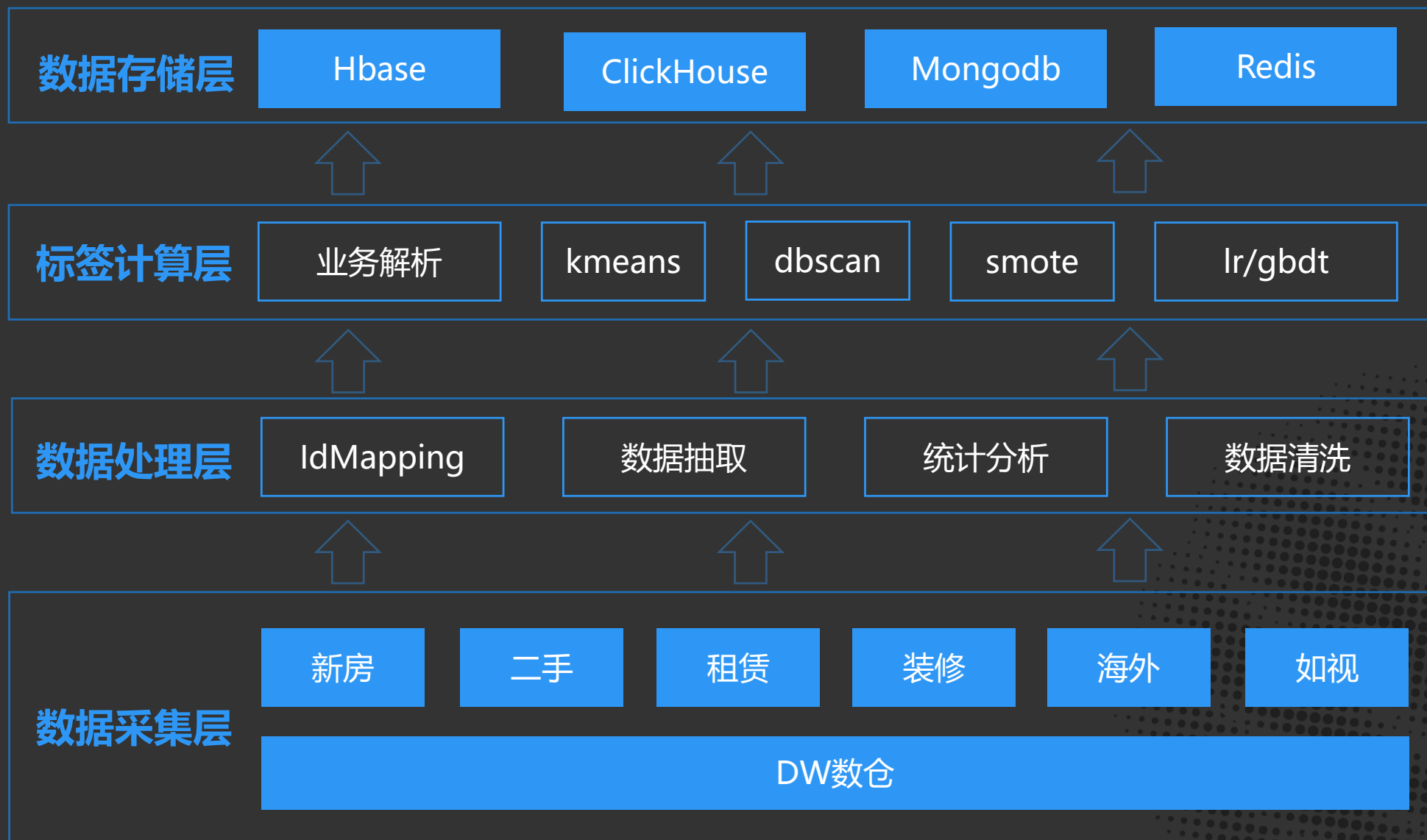


- 如何选取label, 确定偏好
- 如何让样本覆盖不同阶段的用户
- 如何优化样本分布的不均衡性

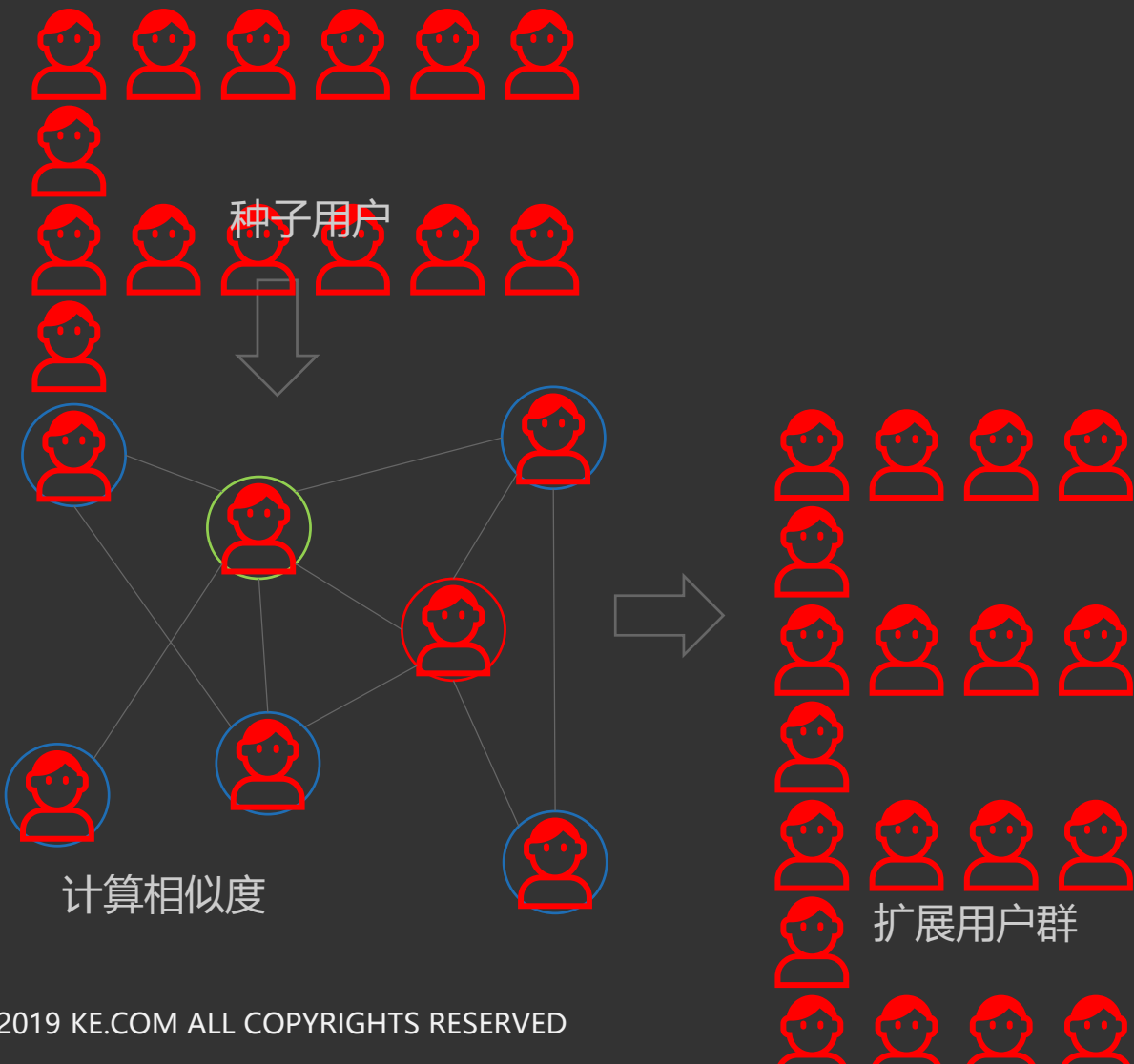
预测标签-LTV



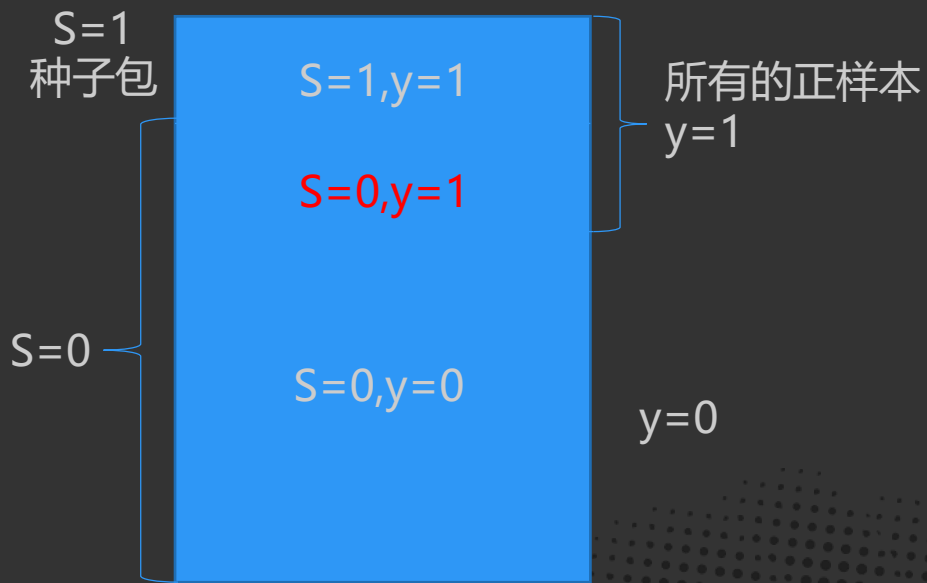
实现框架



典型应用一



Pulearning



$$p(y=1|x) \Leftrightarrow p(s=1|x)$$

$$p(s=1|x) = p(s=1, y=1|x) = p(s=1|y=1, x)p(y=1|x)$$

假设s是从y中随机采样出来的,跟具体的x无关

$$p(s=1|y=1, x) = p(s=1|y=1) = c$$

典型应用二

三端用户状态分布 ?

用户状态 (%)		TGI (%)
其它	0.10	22.90
流失用户	11.99	16.81
沉默用户	5.23	57.83
活跃用户	82.07	437.75
新增用户	0.61	138.00

APP用户业务偏好分布 ?

业务线 (%)		TGI (%)
其它	10.66	12.20
二手	81.86	1256.26
租赁	2.77	92.84
新房	4.71	148.82

事件分析 ?

业务方: ?

大C端

分析类型:



事件



session

事件类型 +

A

所有事件

触发次数

维度选择 +

按

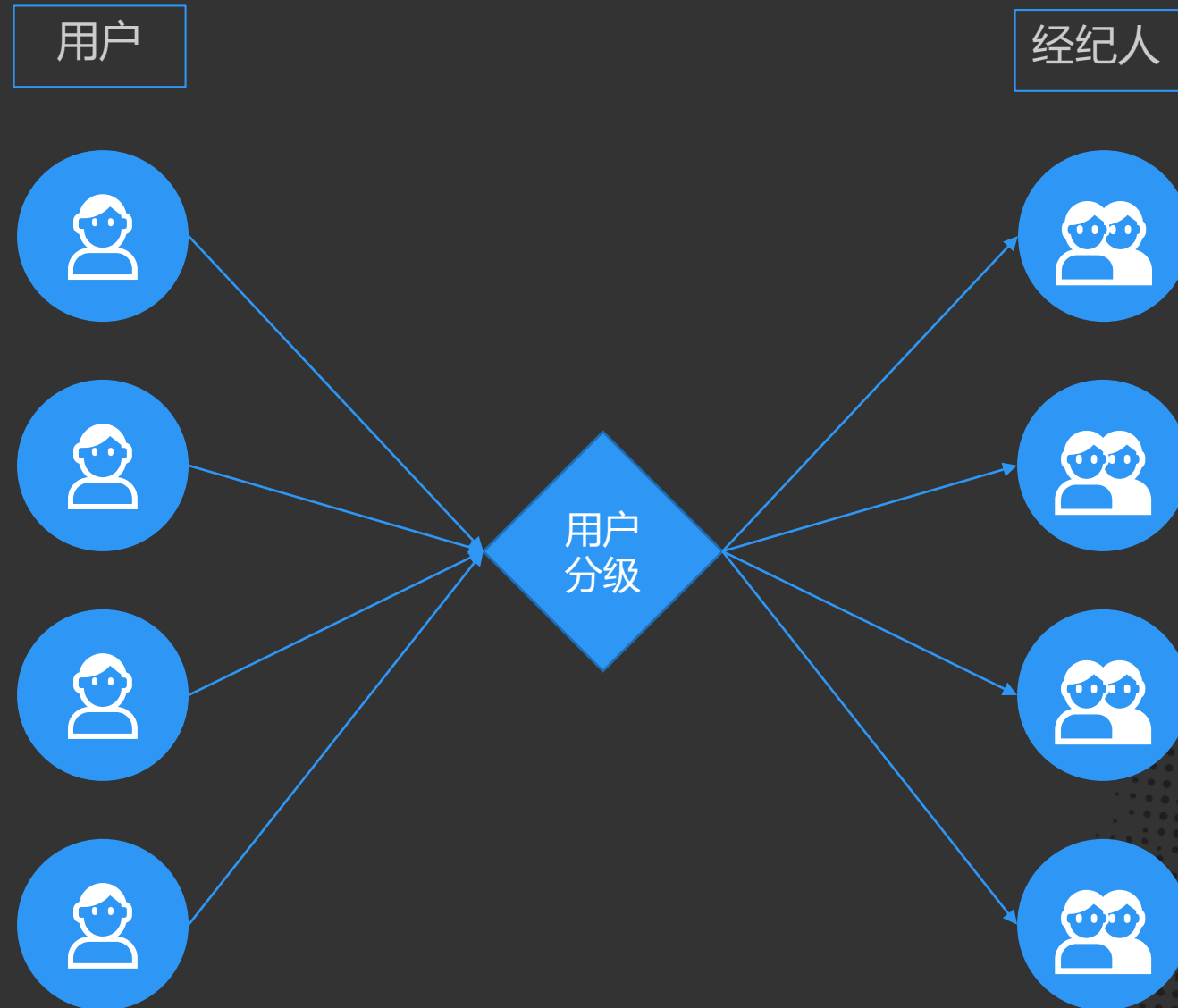
总体

典型应用三



- 老客唤醒成本降低28%
- 商机成本下降8%
- 推荐ctr提升20%+
- 搜索ctr提升5%

典型应用四



用户体系





全球技术领导力峰会

Geekbang> | TGO 鲲鹏会
极客邦科技

500+ 高端科技领导者与你一起探讨 技术、管理与商业那些事儿



🕒 2019年6月14-15日 | 📍 上海圣诺亚皇冠假日酒店



扫码了解更多信息

THANKS

