

外卖商户数据化运营体系建设

李凯



个人简介



李凯

15年加入美团

15年-16年 美团平台技术部

16年-至今 美团外卖技术部

关注领域：商户运营

系统程序员

数据挖掘爱好者

聊点什么

1. 商户数据化运营一些实践经验
2. 系统程序员如何应用AI技术帮助业务增长
3. 业务和技术问题的拆解和思考方式

目录

- 1

个人简介
- 2

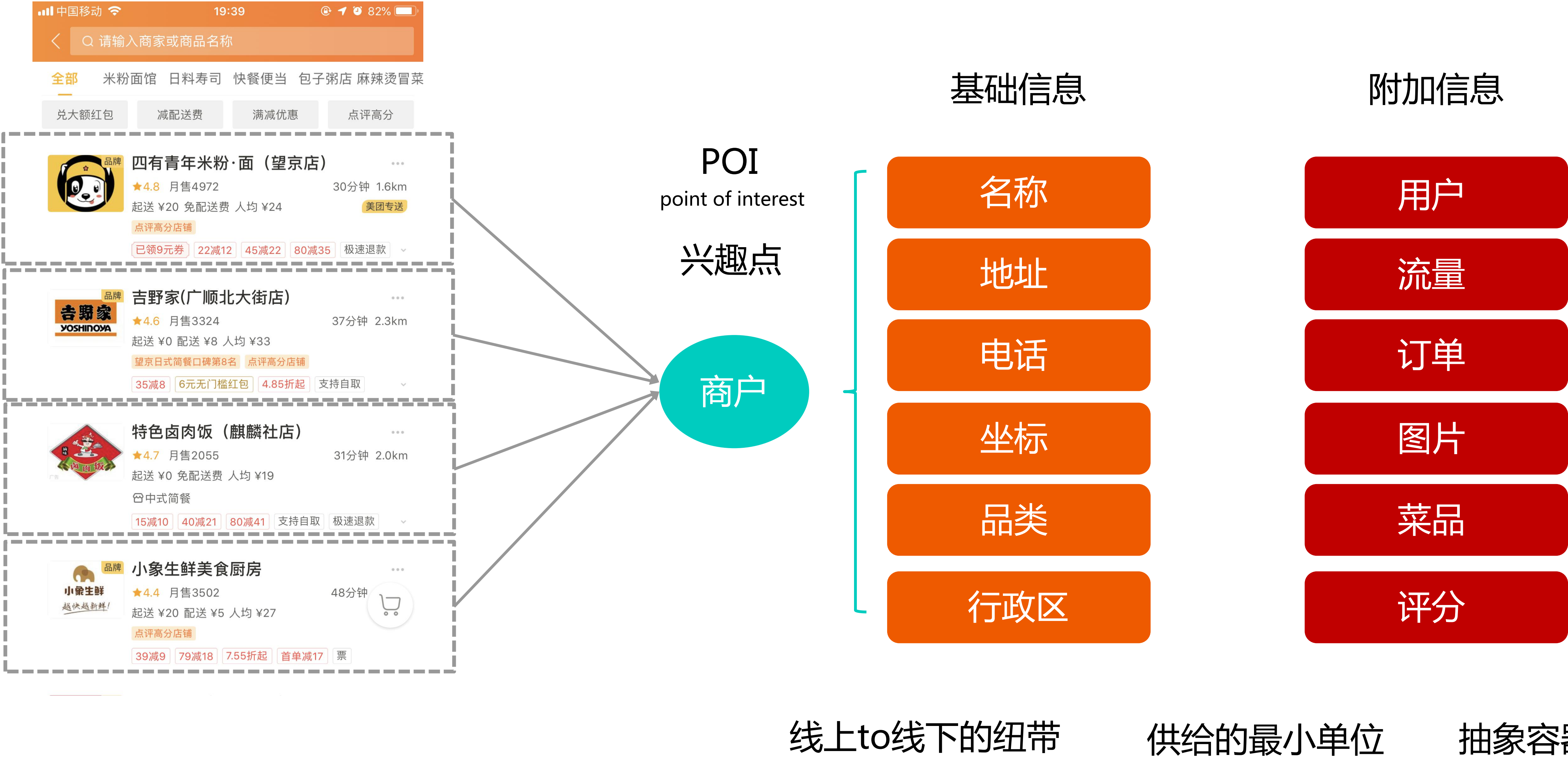
业务背景
- 3

运营体系
- 4

思考总结

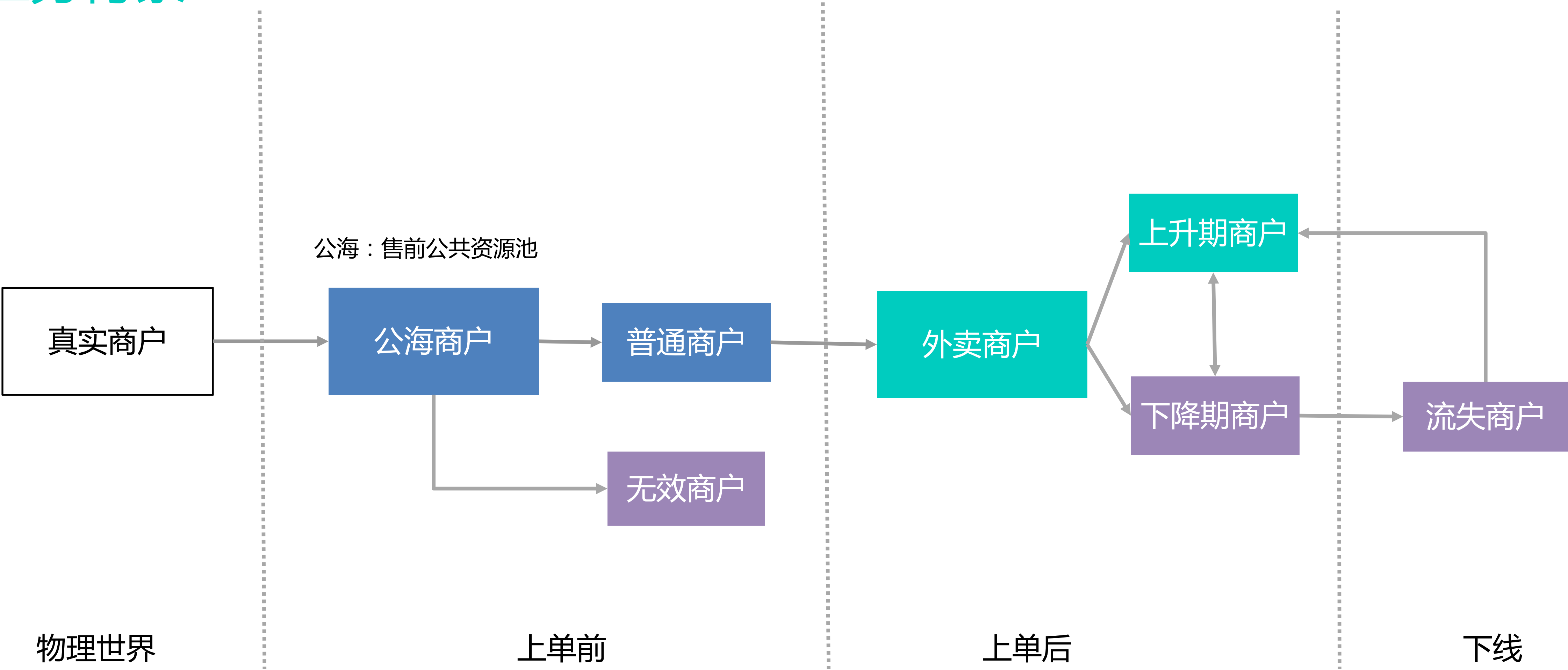
什么是商户

业务背景



什么是商户数据化运营
业务背景

商户数据化运营就是以POI为核心的信息管理和价值挖掘工作



上单：从未合作变成合作的签约过程

阶段划分
业务背景

商户数据化运营体系



目录

- 1

个人简介
- 2

业务背景
- 3

运营体系
- 4

思考总结

商户信息审核

业务拆解

POI运营体系v1.0



商户信息管理

POI运营体系v1.0

审核效率提升

核心挑战

如何在保证数据质量的前提下提升审核效率？

解决思路1

尽可能的避免重复和冗余工作

审核规则自动化

解决思路2

尽可能的避免停顿

业务协调系统

解决思路3

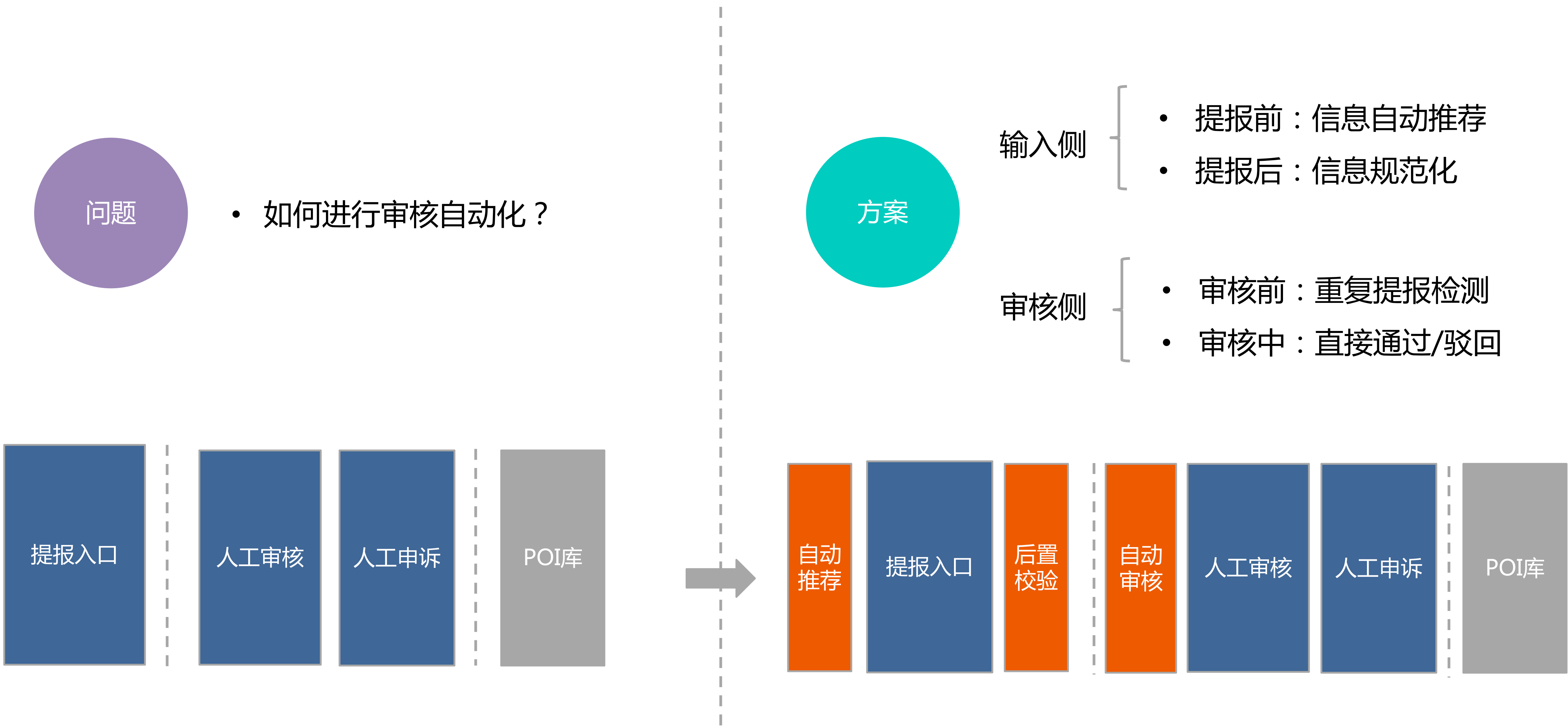
规则尽可能快的发布和修改

审核规则配置化

审核流程框架

POI运营体系v1.0

Tip1：避免重复和冗余工作



商户业务协调系统

POI运营体系v1.0

Tip2 : 减少无效等待

问题

- 重复：保留一家门店
- 倒闭：直接下线

- 绑定业务需要迁移
- 下线业务需要确认

线下流程冗长、角色多、等待

方案

核心思路

将多人协作的过程线上化和标准化

业务协调

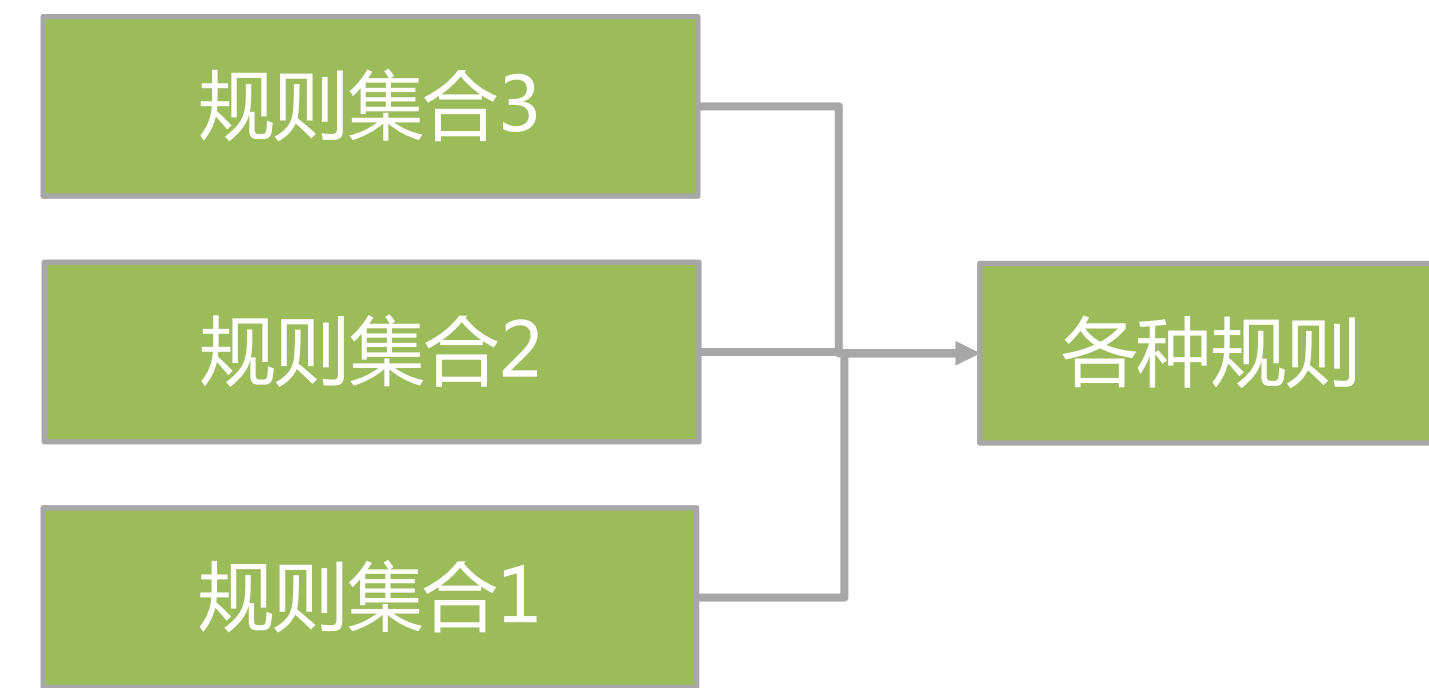
- 实现任务的自动流转
- 基于组织节点的升级
- 提供业务迁移回调接口

商户审核规则引擎

POI运营体系v1.0

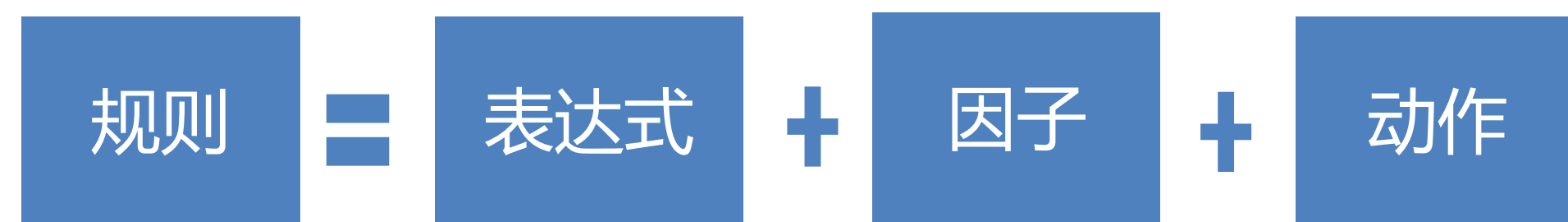
挑战：规则多如何快速上线？

问题



- 上线慢
- 难维护
- 难追溯

数据模型



业务系统

调用层

数据对象
规则场景

服务层

外部服务1
外部服务2
外部服务3

执行层

决策解析
执行动作

配置系统

场景管理

规则1

规则2

规则N

因子X
因子Y

因子Z
因子A

因子B
因子C

自定义回调

决策1

决策2

决策M

总决策

总结

POI运营体系v1.0

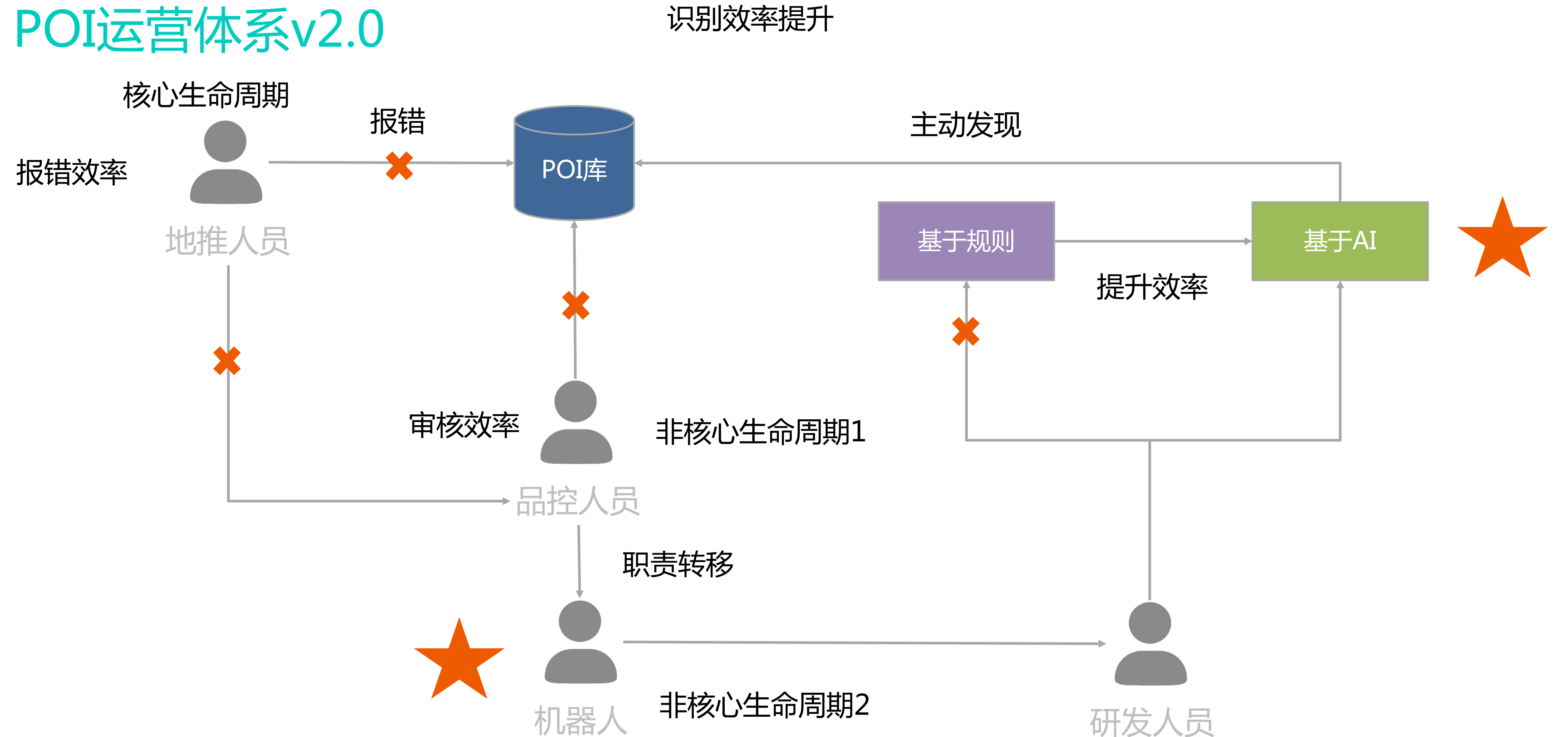
- 1. 业务结果：提升审核效率
- 2. 算法能力：人工+规则处理框架
- 3. 研发效率：规则引擎

信息审核

商户质量识别

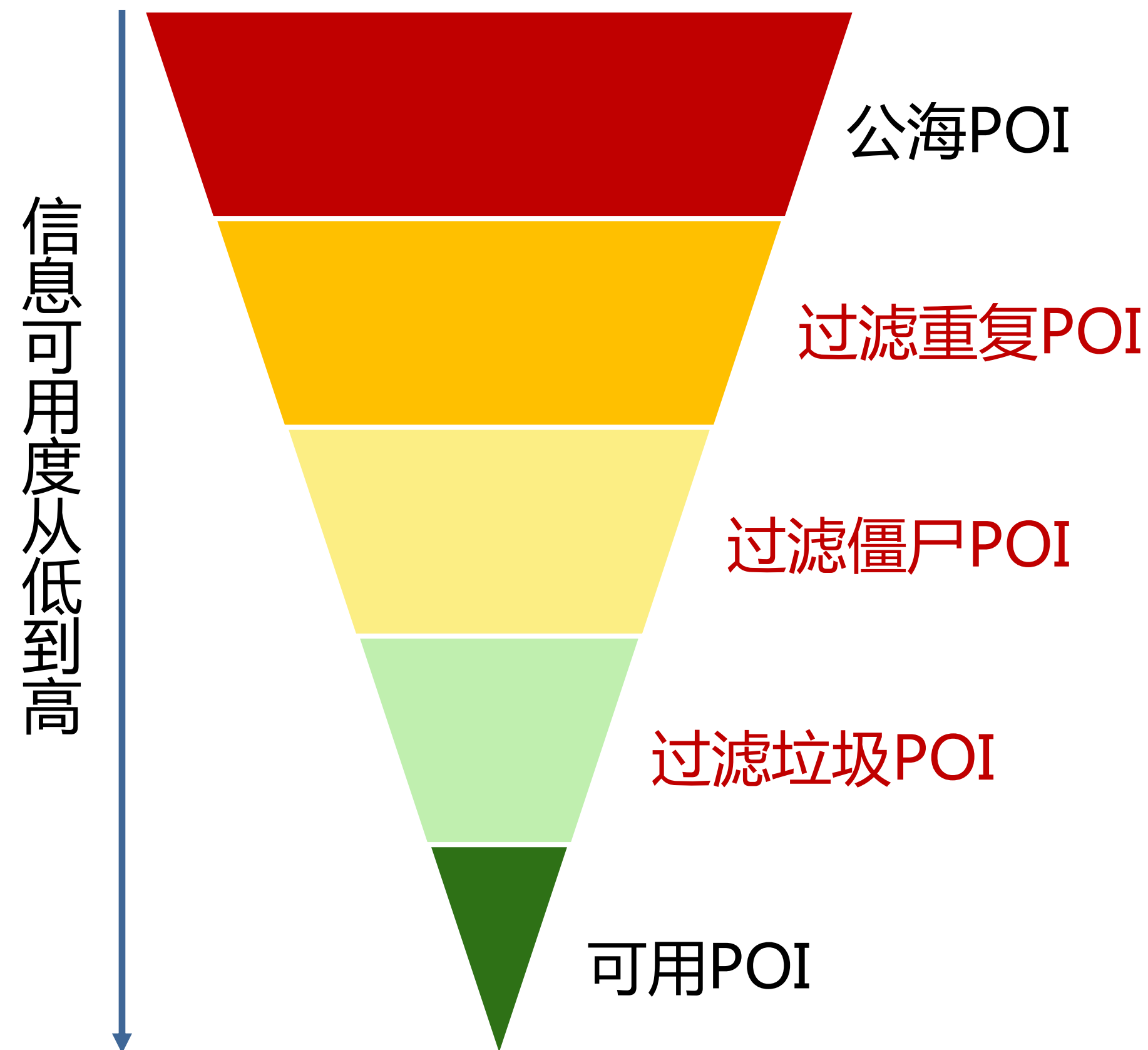
业务拆解

POI运营体系v2.0



质量体系建设

POI运营体系v2.0



标签化

重复

僵尸

垃圾

基于规则的效果差

分类问题

使用机器学习模型

核心问题

给定条件下是A还是B的概率

评测结果明确

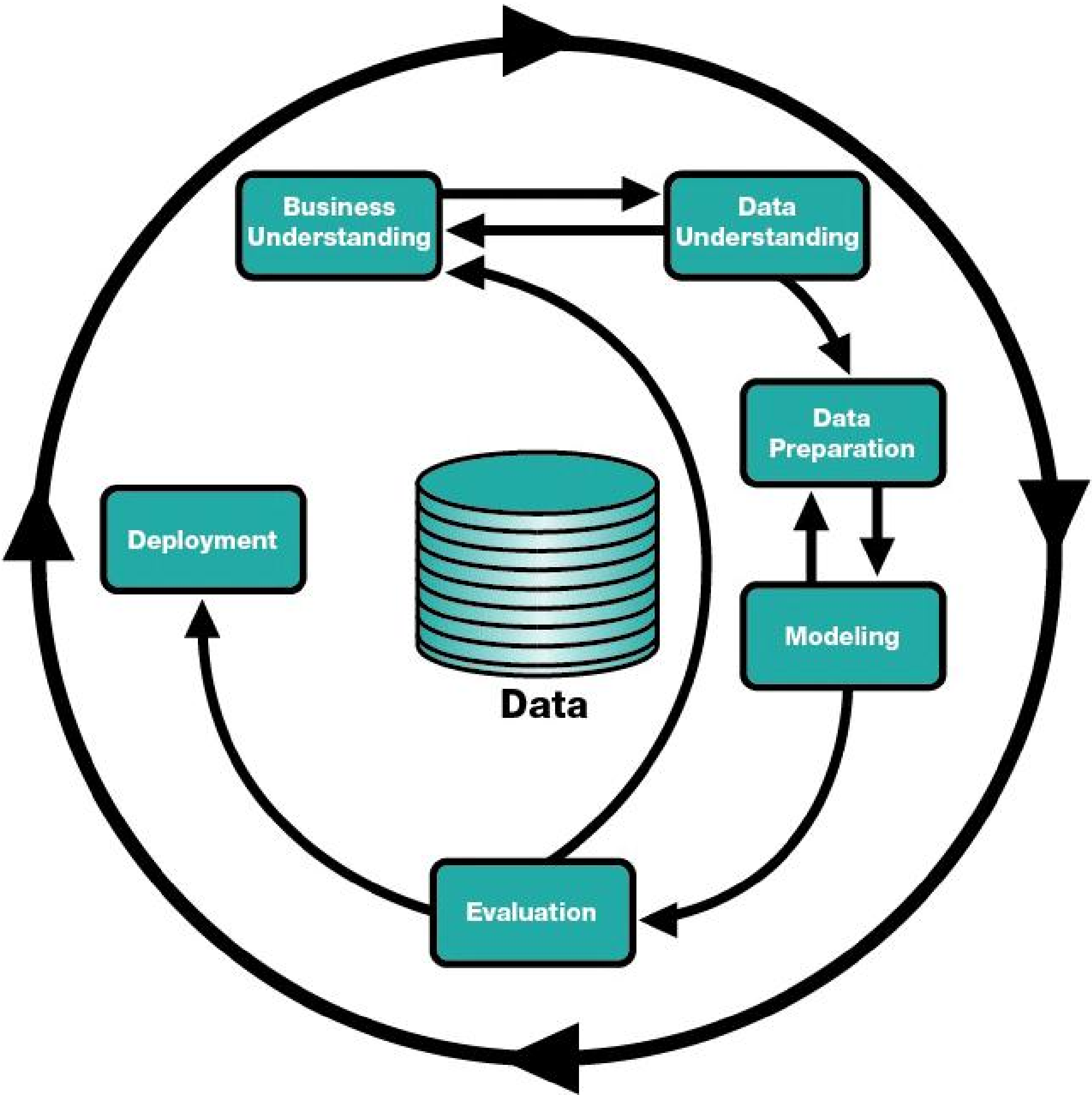
挑战

确定性问题的算法研发能力

跨行业的数据挖掘标准流程(CRISP-DM)

POI运营体系v2.0

第一步：了解标准流程



- 1.业务理解
 - 这个问题适合用机器学习来解吗？
 - 是分类问题还是回归问题？
- 2.数据理解
 - 知道数据在哪儿，数据特点和分布情况
- 3.数据准备
 - 样本数据是从哪里来？
 - 特征如何选取？
- 4.模型搭建
 - 使用什么模型？线性模型/决策树模型 /神经网络模型？
 - 使用这些模型怎么解决泛化的问题
- 5.模型评估
 - 模型好坏的指标是什么？AUC/准确率/召回率
 - 离线评测还是在线评测
- 6.模型发布
 - 使用什么方式导入？

应用举例：去重

POI运营体系v2.0

POI相似度

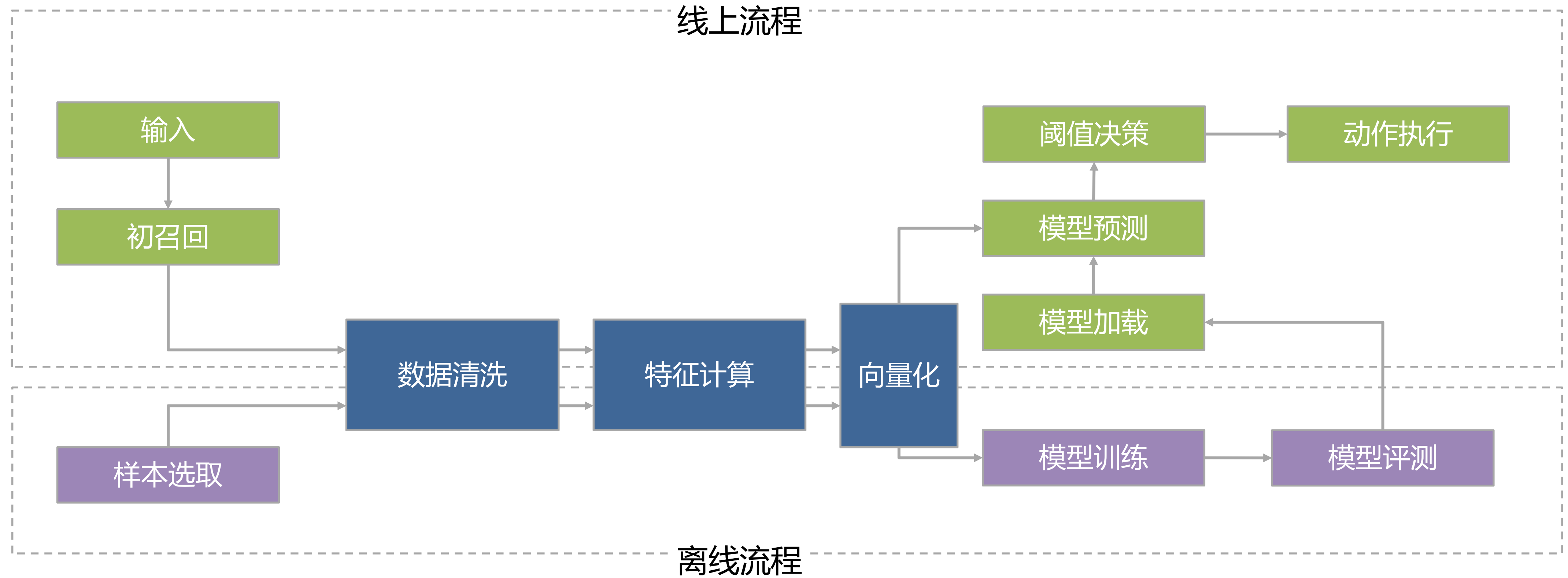


淮南牛肉汤面（安德利美食城店） 军二西路21号安德利广场A321号商铺 159xxxxxx81 快餐小吃	18米	百世牛肉汤（安德利美食城店） 军二西路21号安德利广场A321号商铺 159xxxxxx81 快餐小吃	否
一心龍点（拱桥店） 城中区四会中学东围墙8-9卡 134xxxxxx99 包子/粥	402米	一心龍点（港式包点） 城中区沿江路滨江新苑四座11号（首层） 134xxxxxx99 其他小吃快餐	是

相似度整体框架

POI运营体系v2.0

POI相似度整体流程框架



样本管理

POI运营体系v2.0

Tip1：不要以为样本采集是一次性工作

问题

- 人工审核的时候有电话输入
- 有噪音的人为纠错

举例

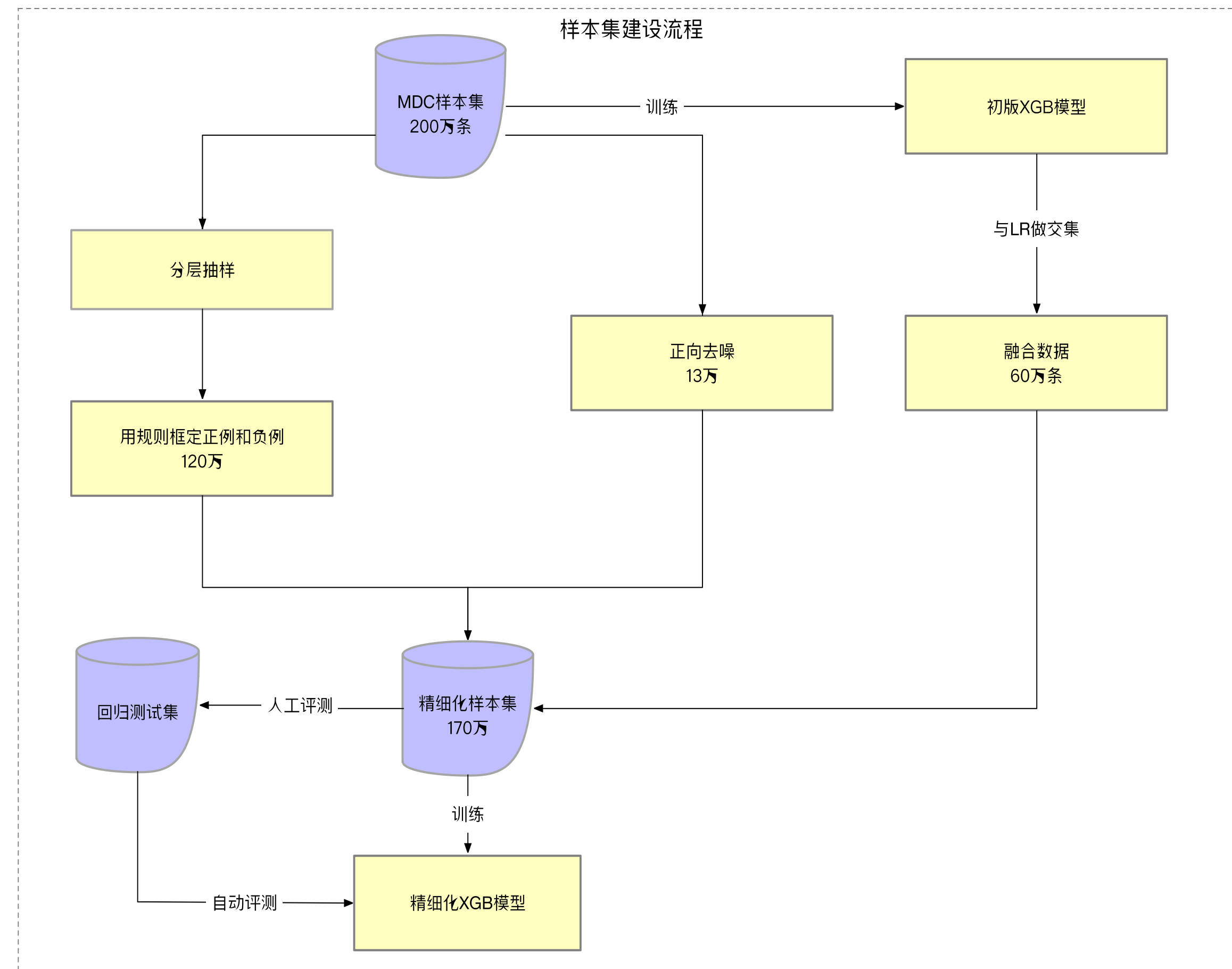
- 距离相差10000KM

样本来源

人工审核结果作为输入

样本提纯

正向建模
逆向剔除
多模型比对



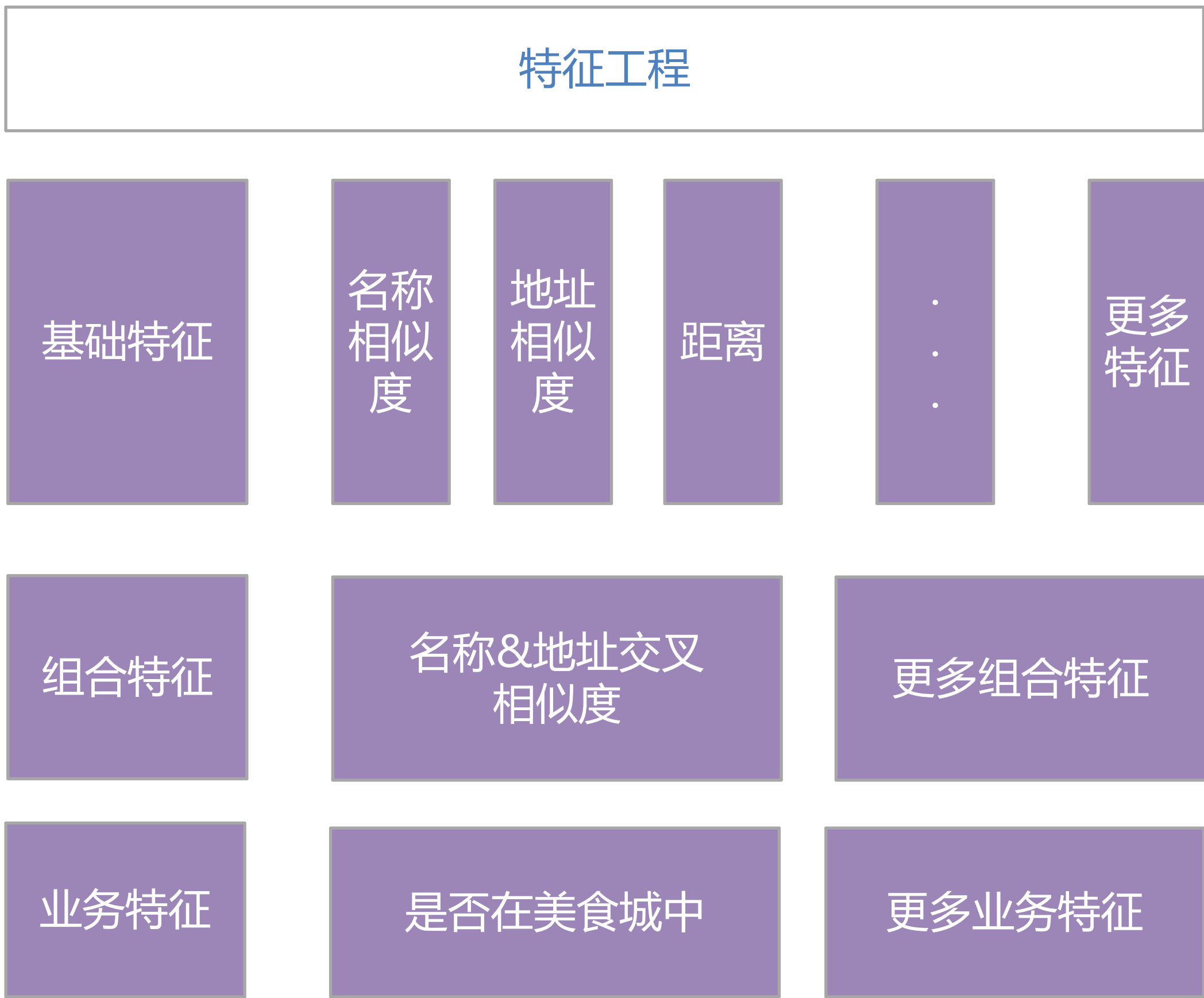
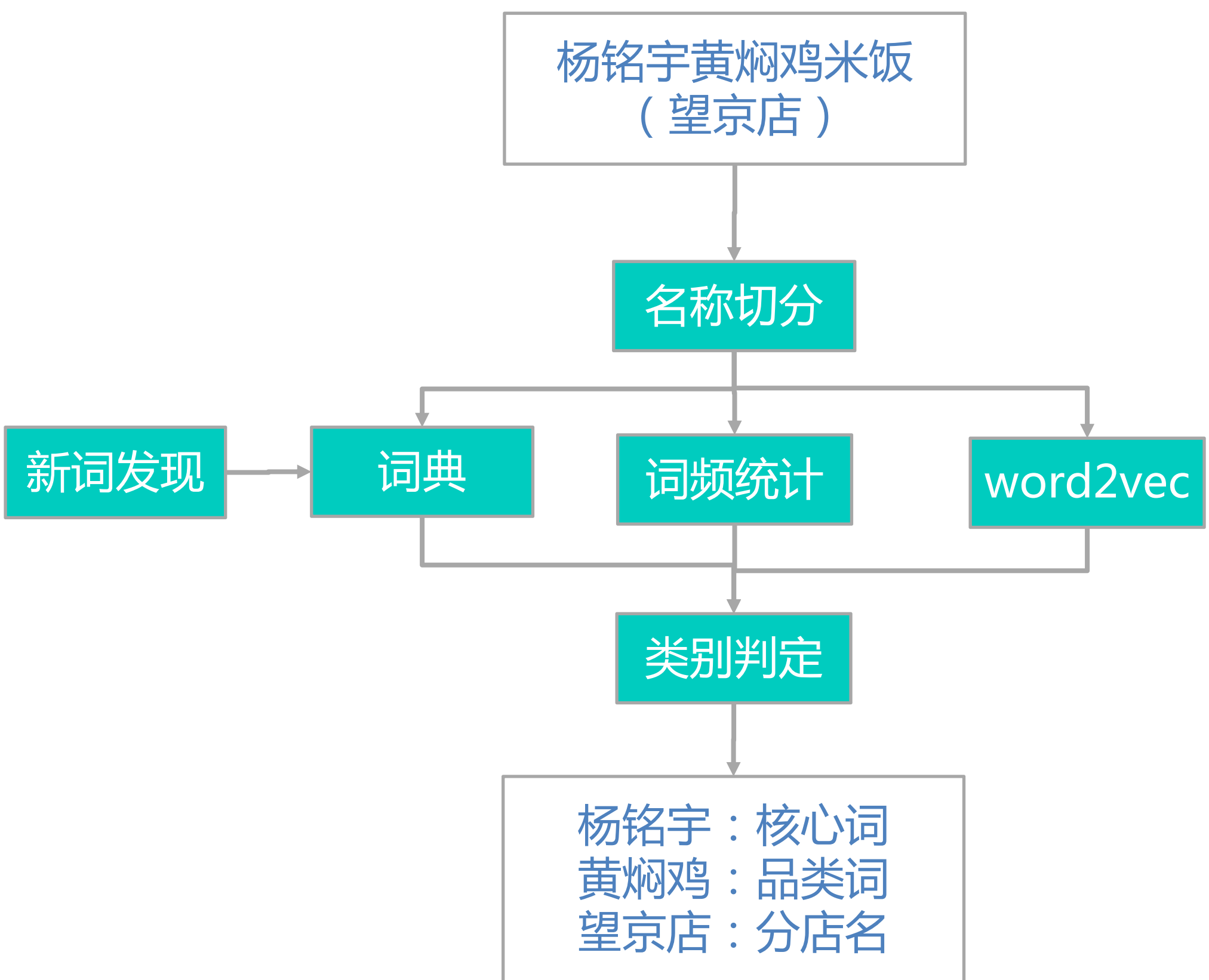
特征工程

POI运营体系v2.0

Tip2：每一个字段都有很多文章可以做

需要拆得很细

核心词+品类词+分店名

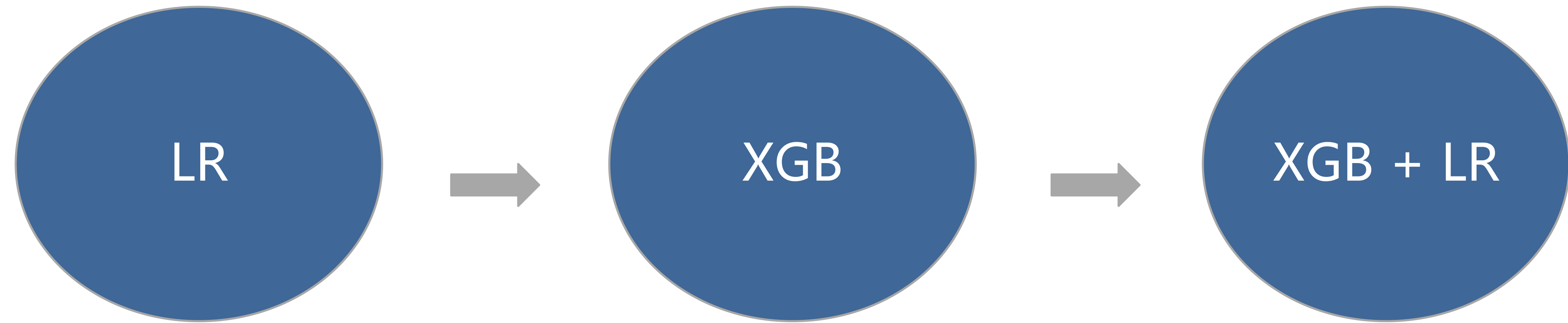


融入组合特征+业务属性

模型选择

POI运营体系v2.0

Tip3：根据业务特性决定模型



优点：简单、平滑
缺点：泛化性能不够

初始阶段好实现

优点：泛化性能高
缺点：不可解释

符合人判定的过程

优点：平滑结果
缺点：部署复杂

适应有排序的场景

机器学习四要素
POI运营体系v2.0

机器学习四要素



- 获取样本
- 样本提纯



- 数据不等于特征
- 多层特征组合



- 与业务特性匹配
- 是否考虑解释性



- 是否有线上反馈
- 评测标准和指标

自建样本库
(规则合成)

名称	地址	坐标
电话	品类	行政区

LR

人工判定

举例：垃圾

优化模型上线效率

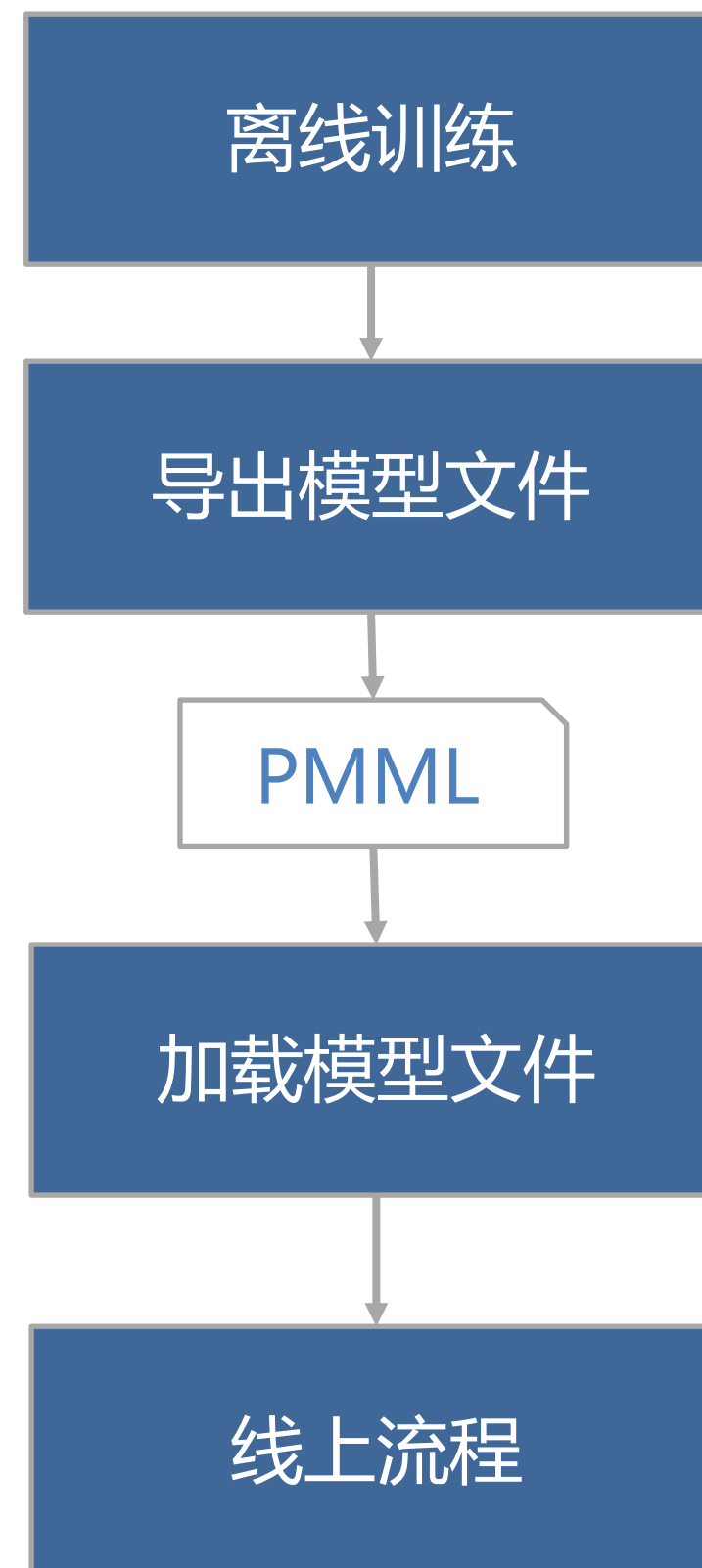
POI运营体系v2.0

挑战：多模型多版本的平滑上线

POI相似度应用场景



模型与代码解耦



多模型多场景的管理系统



总结
POI运营体系v2.0

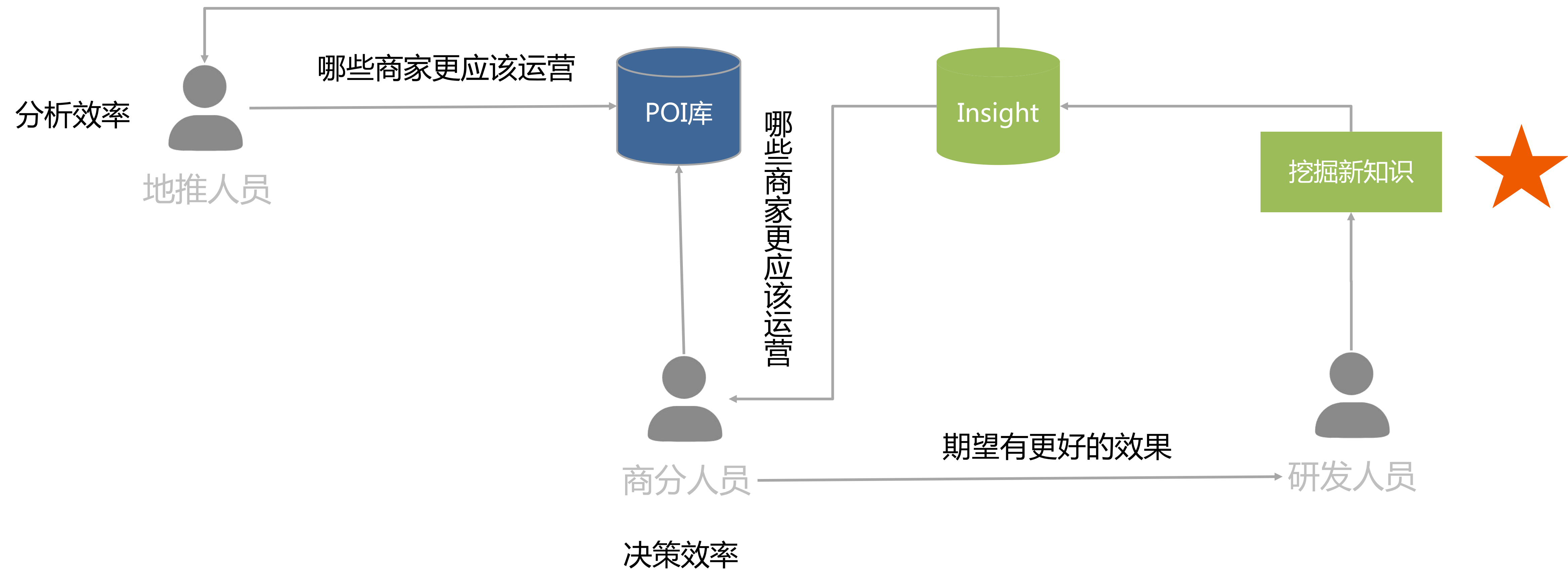
- 1. 业务结果：提升自主识别效率
- 2. 算法能力：确定性分类能力强化
- 3. 研发效率：模型管理

质量识别

商户价值挖掘

业务拆解

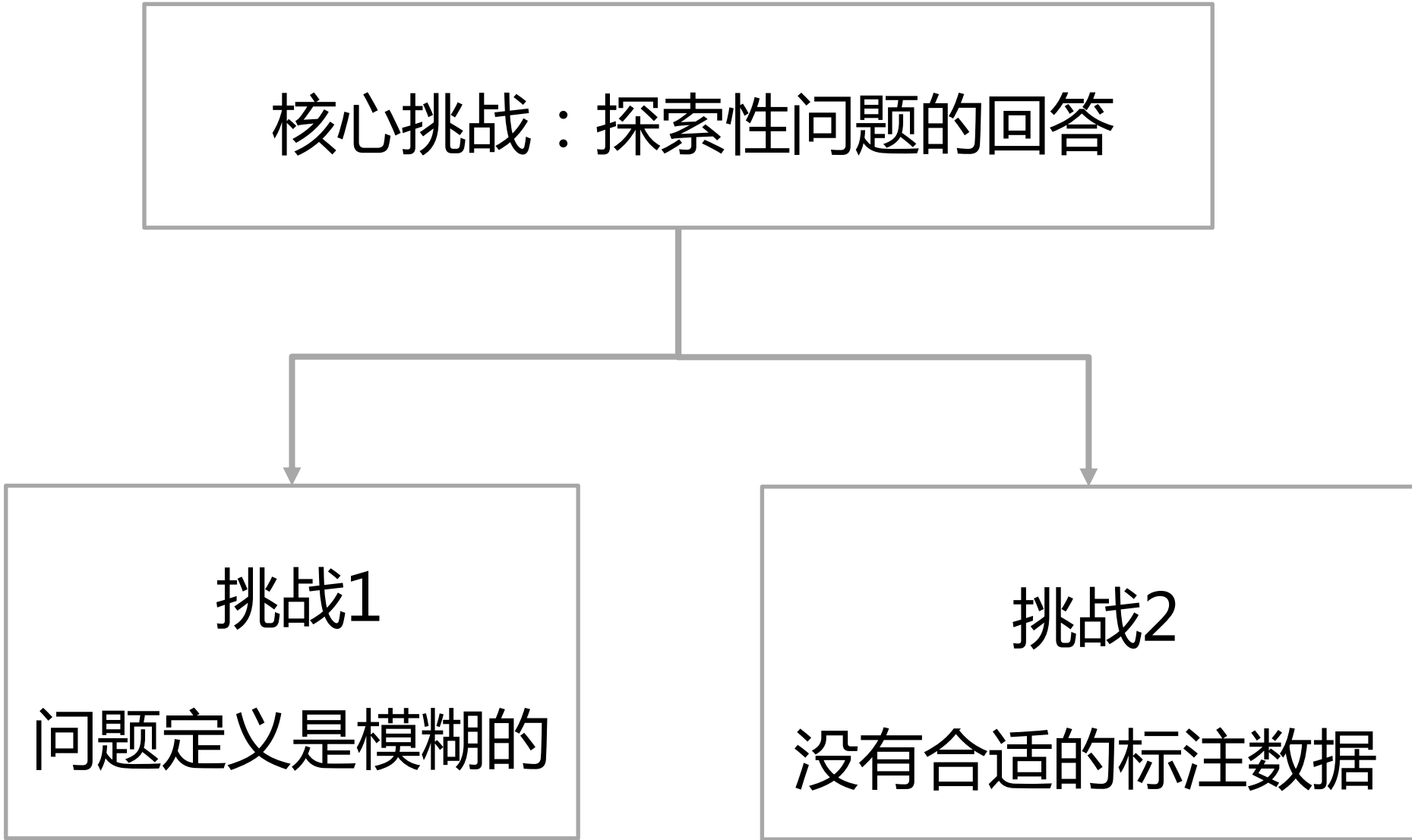
POI运营体系v3.0



整体视图

POI运营体系v3.0

提升决策效率



问题定义优化

POI运营体系v3.0

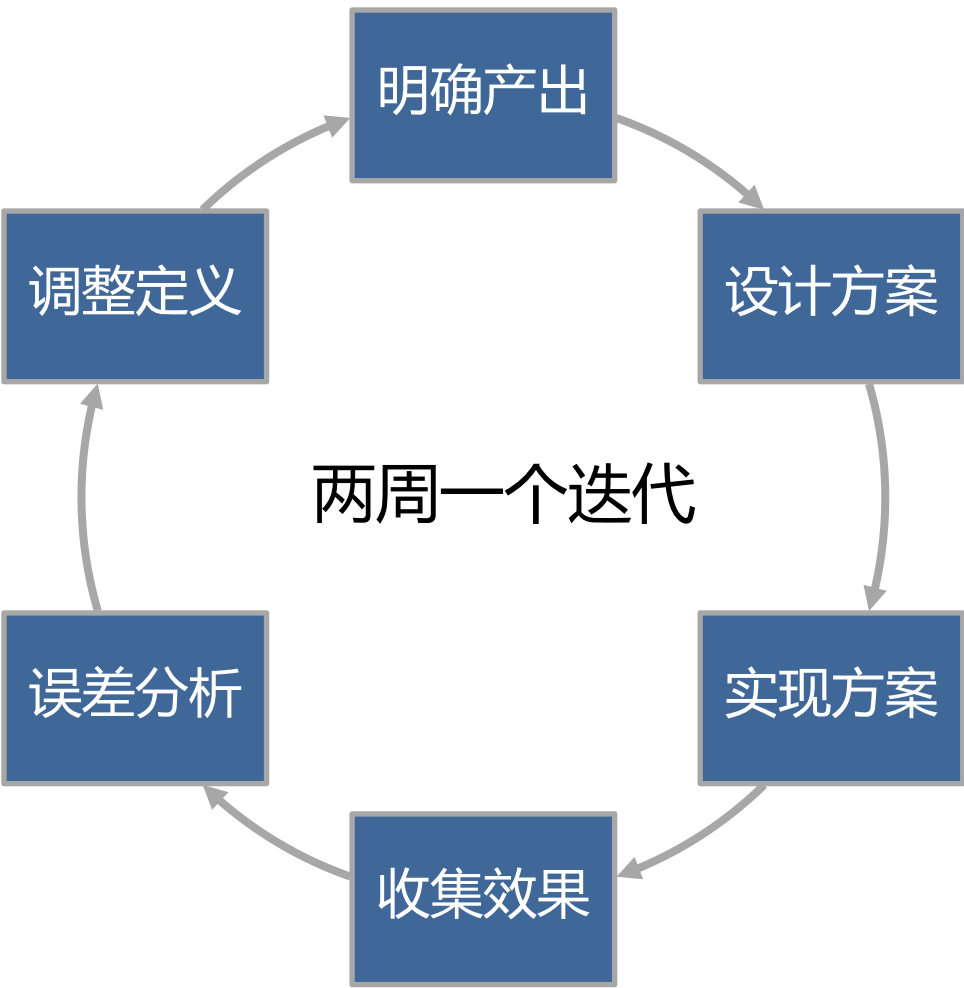
Tip1：问题定义需要不断去细化



定义模糊



- 概念拆解与细化
- 找一个近似模型
- 快速迭代评估改进

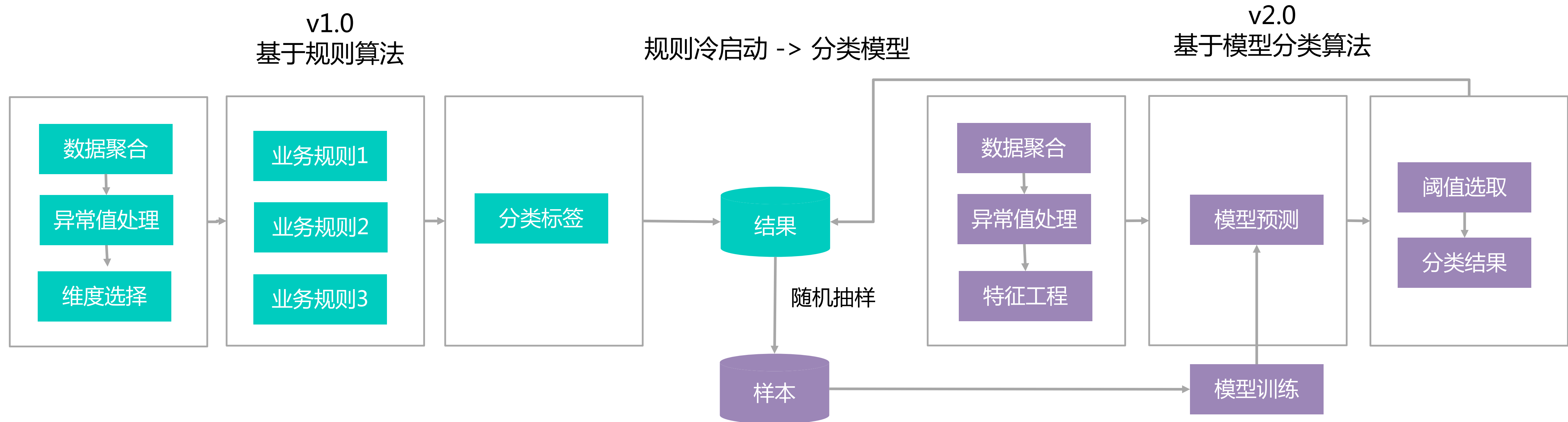


原则：最小可行性试错（ MVP ）

算法迭代优化

POI运营体系v3.0

Tip2：无标注的处理方法



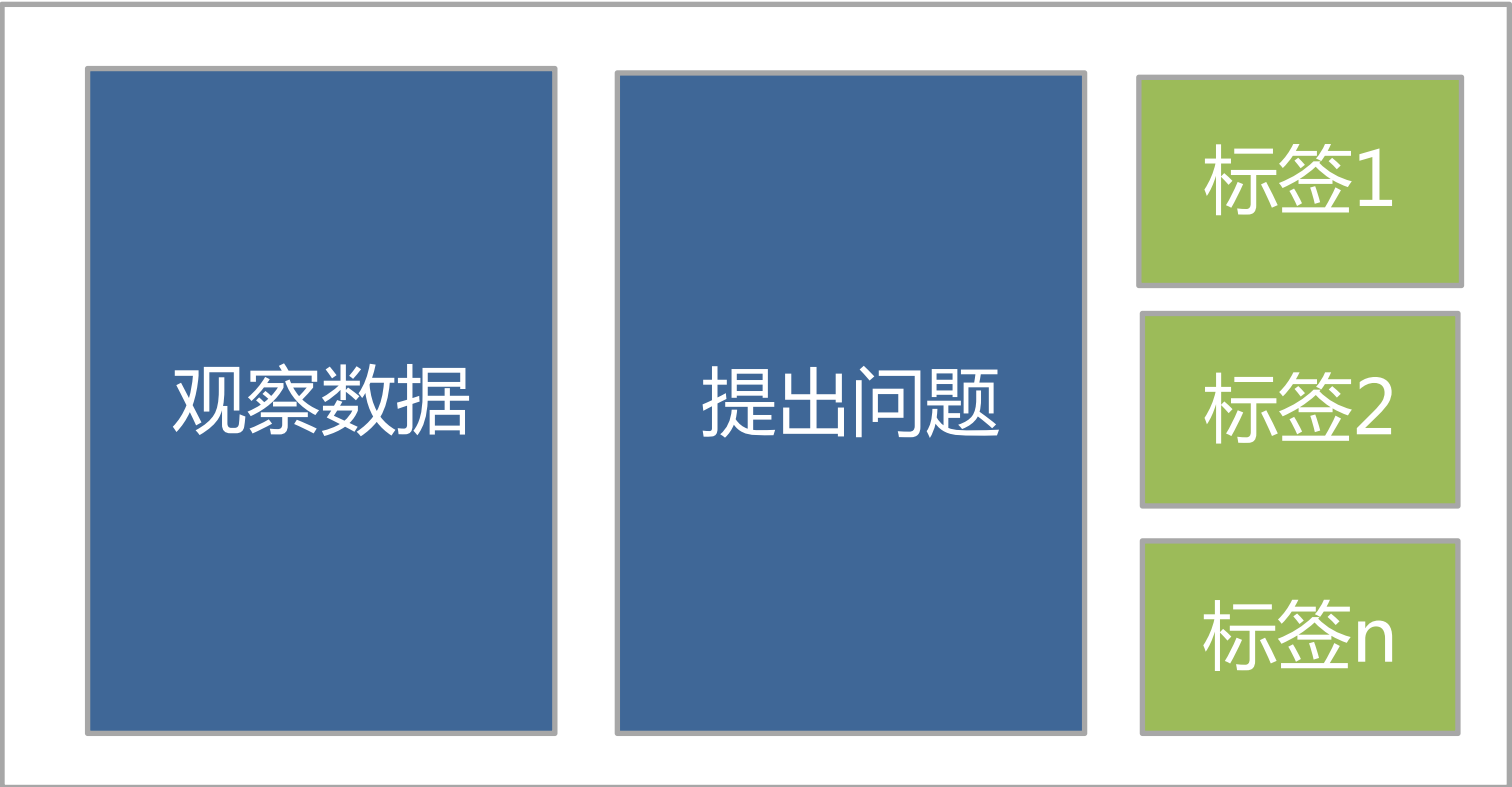
举例：打分问题转化成两两比较

评测效率优化
POI运营体系v3.0

挑战：大量的线下评测工作

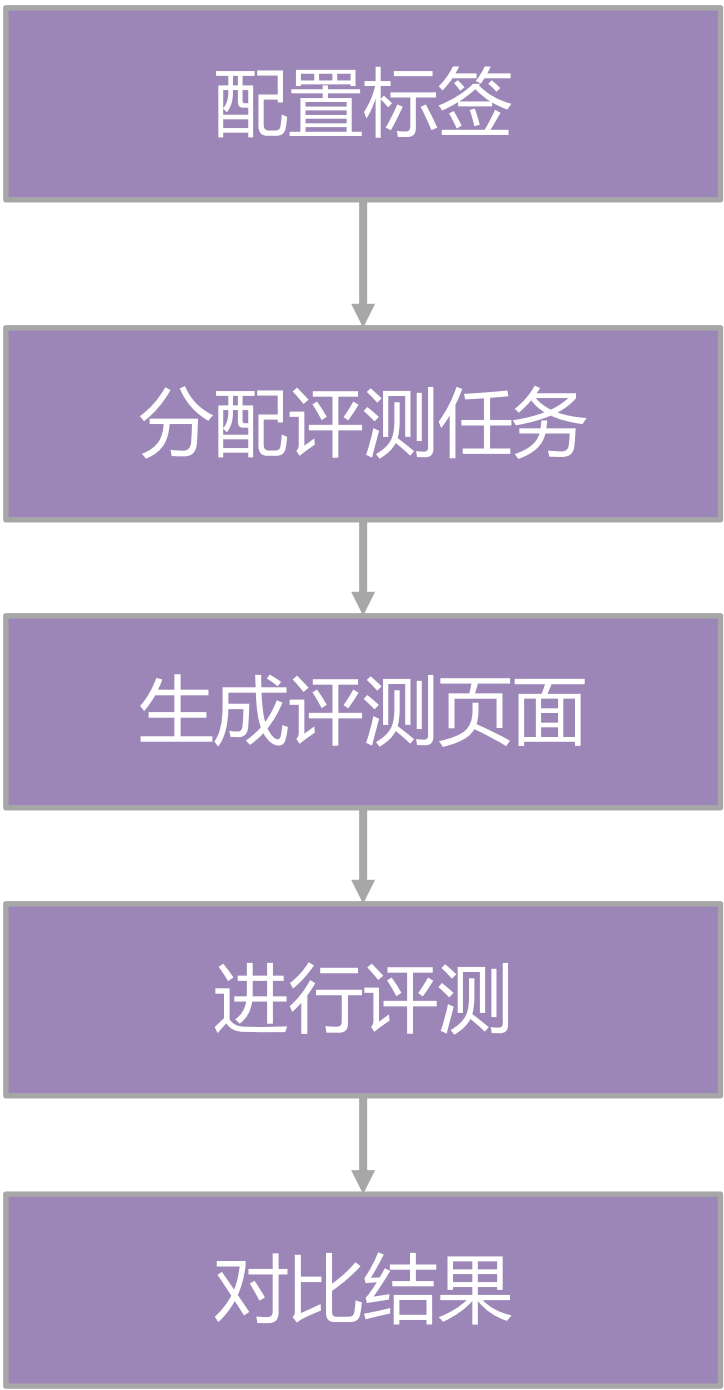
研发效率

评测模型

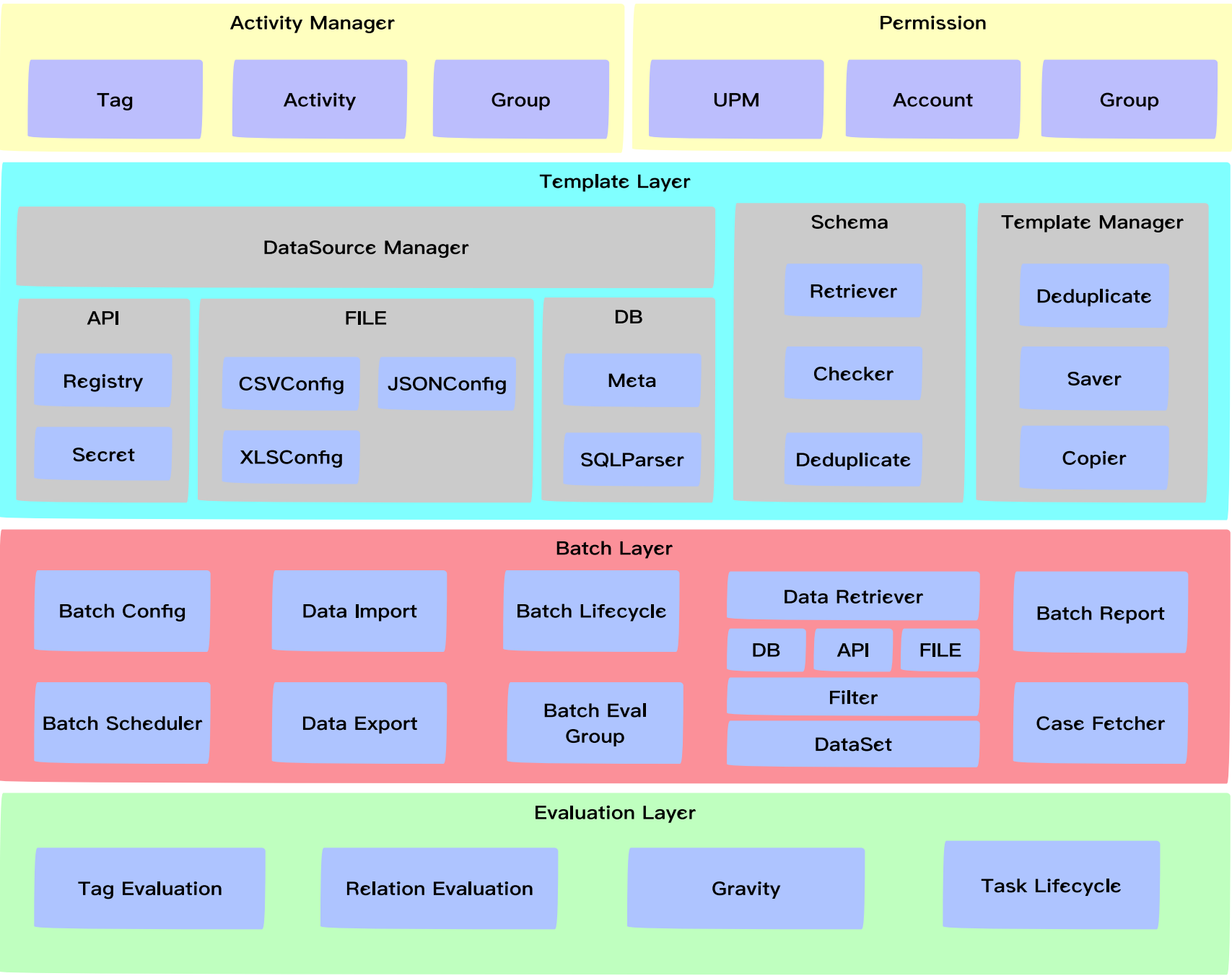


举例：相似度&优质&分级

评测流程



评测效率：3d -> 2h



总结

POI运营体系v3.0

- 1. 业务结果：提升决策效率
- 2. 算法能力：不确定性问题探索
- 3. 研发效率：评测平台

价值挖掘

目录

- 1

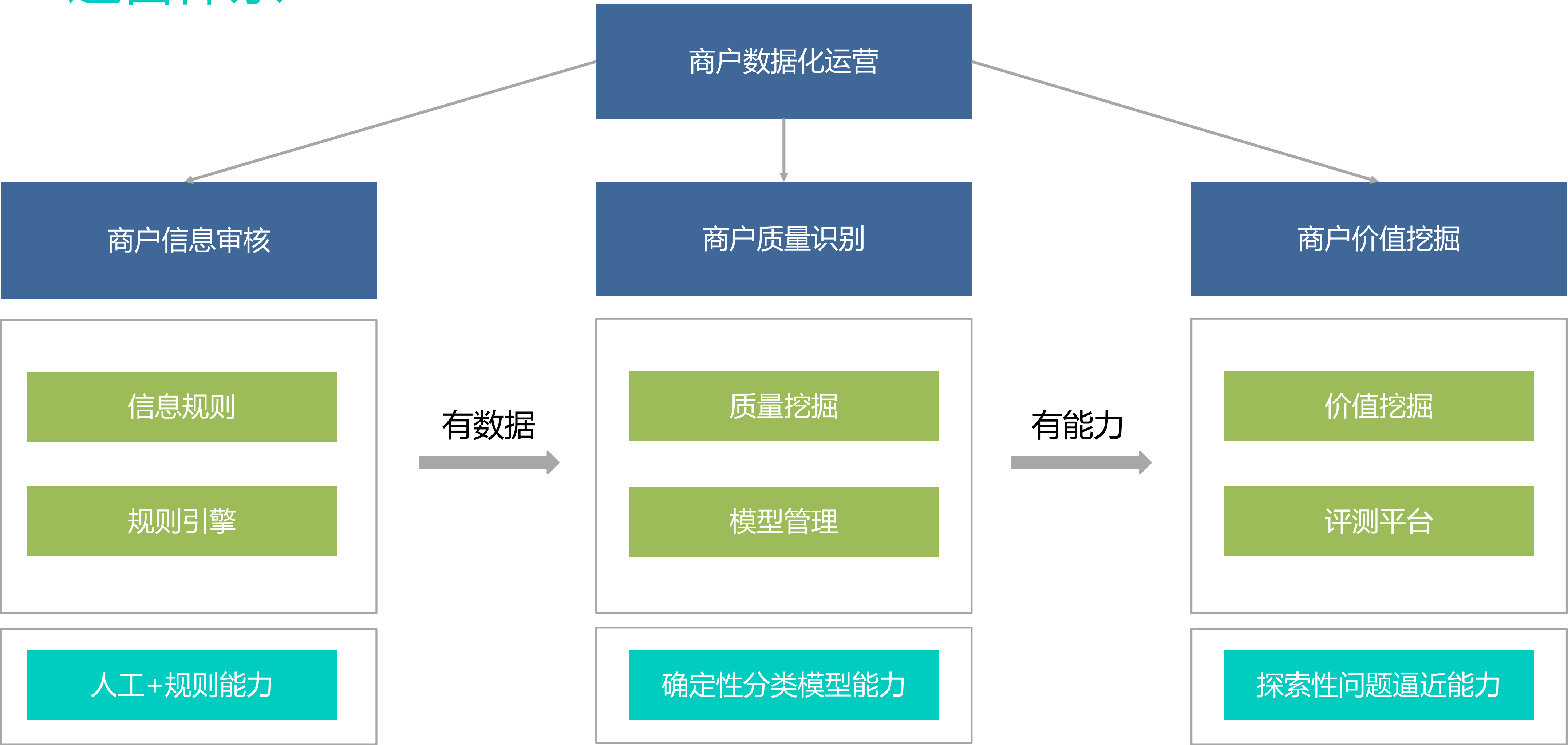
个人简介
- 2

业务背景
- 3

运营体系演进
- 4

思考总结

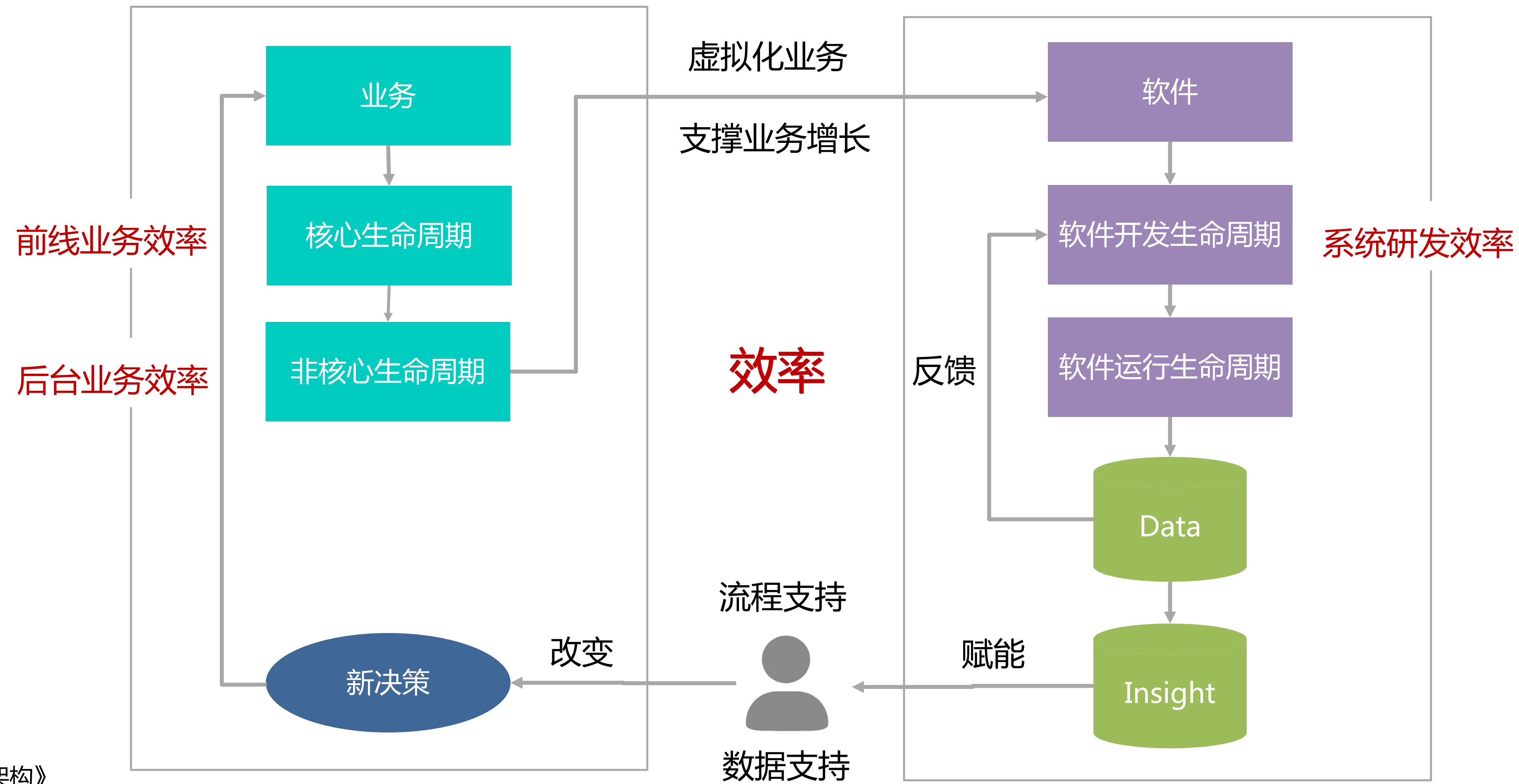
演进
POI运营体系



对架构的理解

业务与架构的拆解

关键问题：什么是当前的效率瓶颈？



总结

系统转AI之路的几点建议

1

建立人工和机器混合处理框架

- 从**规则算法**开始
- 算法部分拆解成单独的服务
- 开始**收集人工反馈数据**

2

找一个简单而明确的分类问题入手

- 掌握**CRISP-DM**流程
- 机器学习四要素：**样本+特征+模型+评测**
- 不断优化这四部分的研究效率

3

尝试去解决探索性问题

- 基于规则的冷启动到模型算法
- 不断**迭代对问题理解**
- 收集各种反馈优化模型

参考文献

1. 机器学习，周志华（西瓜书）
2. Machine Learning Yearning，吴恩达
3. 聊聊架构，王概凯

Q&A