平安银行算法实践

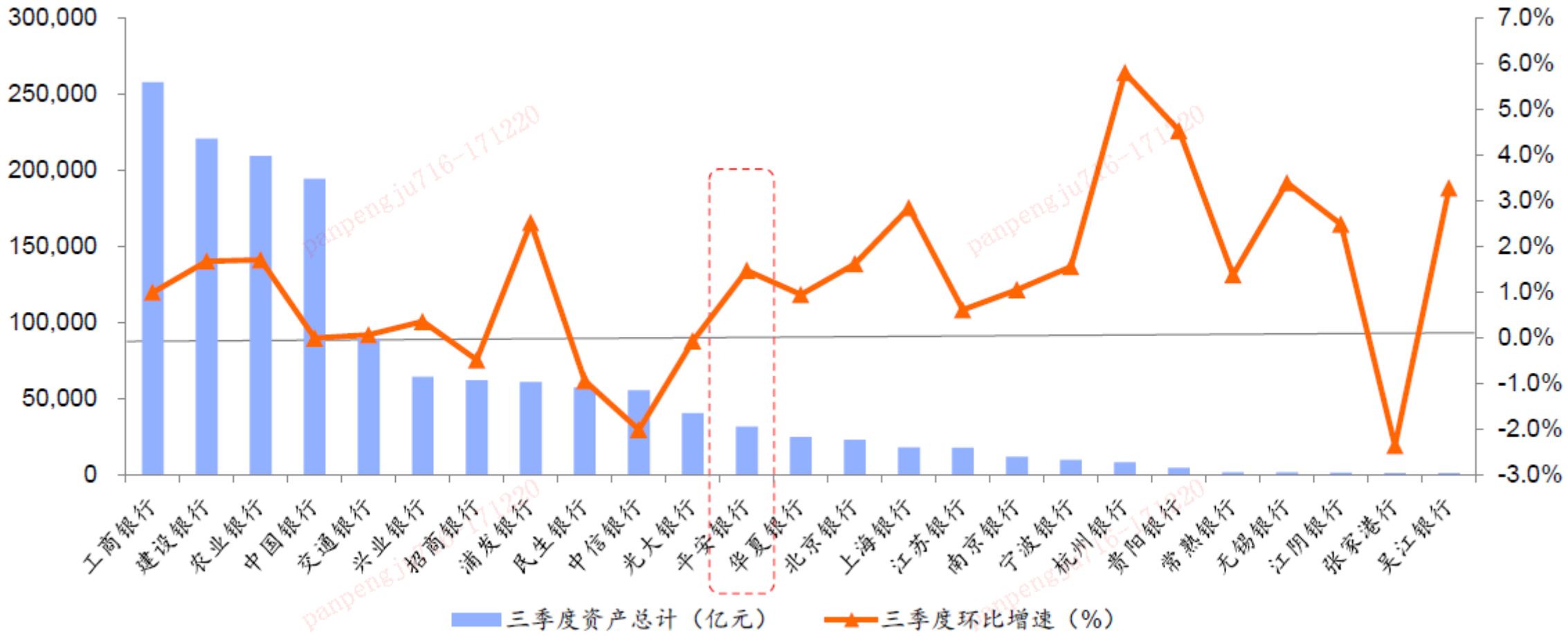
潘鹏举（ppj）- 算法团队负责人

自我介绍-PPJ

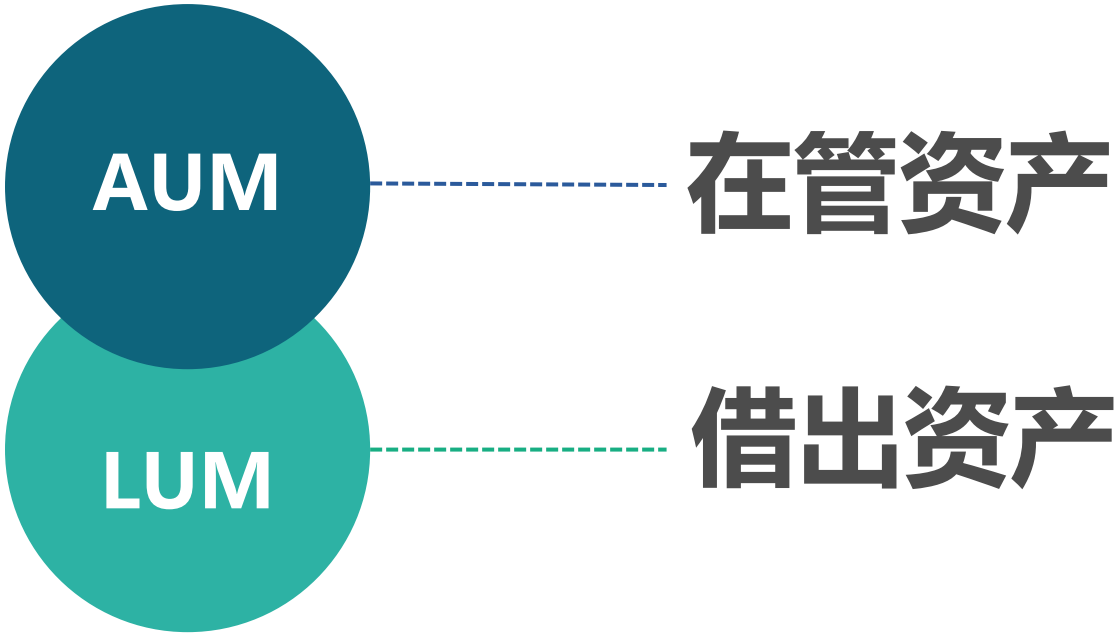
- 平安银行大数据AI算法团队Leader
- 携程酒店技术算法组Leader
- 全栈数据
- 码农+管理

业务背景

平安银行资产排名


300,000
250,000
200,000
150,000
100,000
50,000
0
7.0%
6.0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
-1.0%
-2.0%
-3.0%
工商银行
建设银行
农业银行
中国银行
交通银行
兴业银行
招商银行
浦发银行
民生银行
中信银行
光大银行
平安银行
华夏银行
北京银行
上海银行
江苏银行
南京银行
宁波银行
杭州银行
贵阳银行
常熟银行
无锡银行
江阴银行
张家港行
吴江银行
三季度资产总计（亿元）
三季度环比增速（%）

银行核心KPI



AUM： Asset Under Management

LUM： Loan Under Management

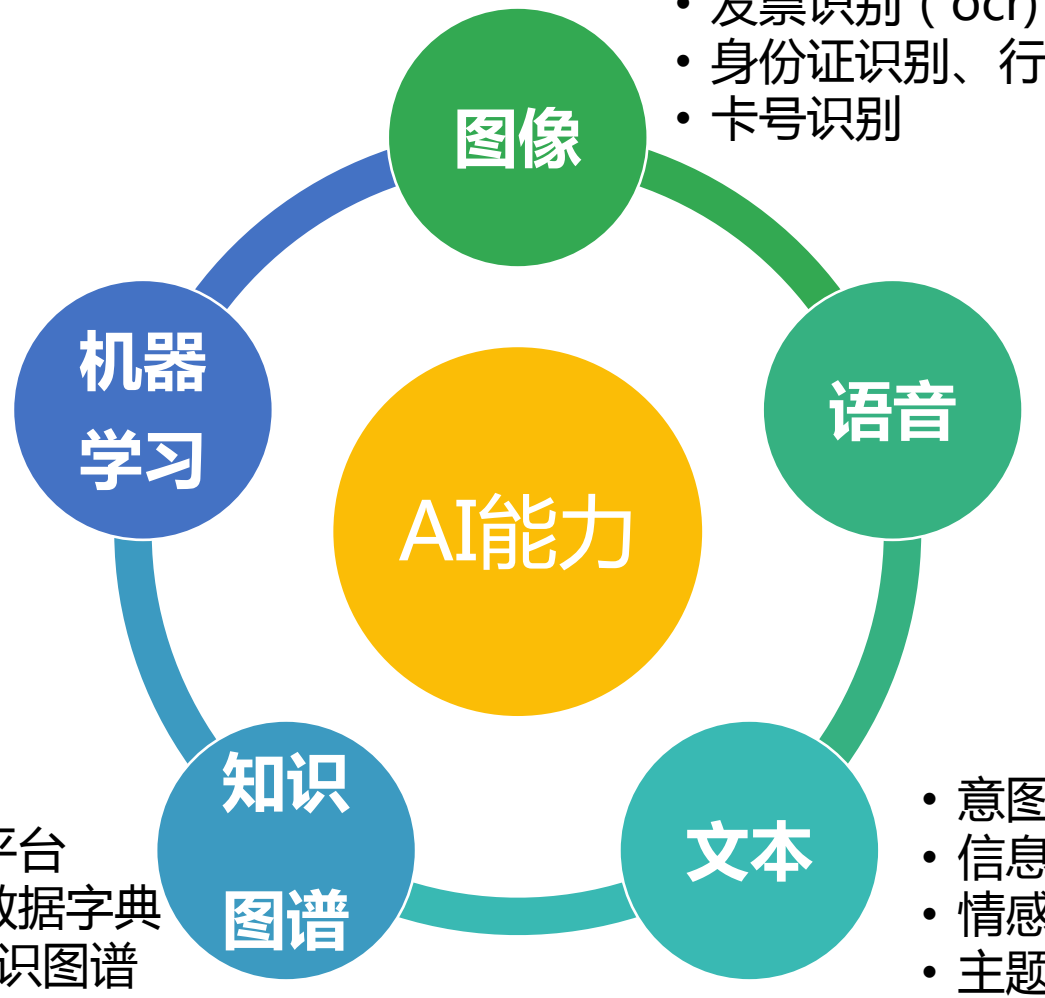
一些挑战

算法实践

总体架构图


口袋A
行员A
代理人A
网银
进线
线下
微信/支付宝...
应用渠道
精准营销
智能风控
搜索引擎
智能客服
可视化平台
实时侦测
征信评分
时空分析
标签平台
知识图谱
Saas
元数据管理
数据仓库层（实时和非实时数据）
Paas
大数据基础能力
AI基础能力
基础服务能力
IAAS

应用概览



- 发票识别（ocr)
- 身份证识别、行驶证、驾驶证
- 卡号识别

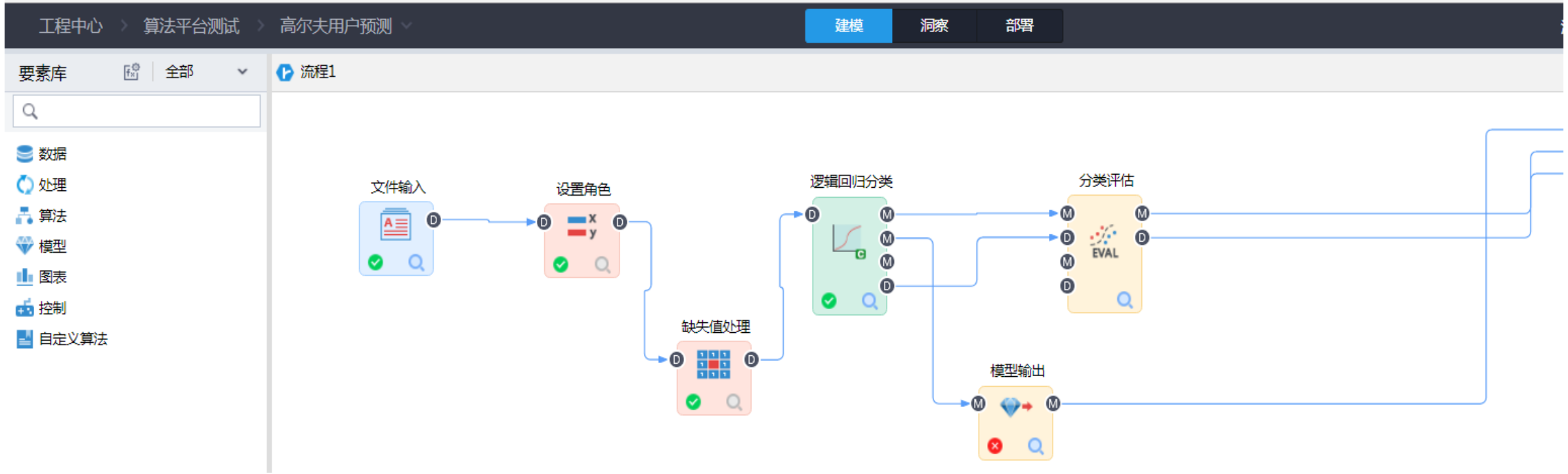
- 客服质检
- 声纹识别（噪音消除）
- ASR语音转文本（去噪、增强）
- TTS语音合成

- 意图识别
- 信息标准化(地址、电话)
- 情感分析
- 主题提取
- 句法分析

- 风控平台
- 智能数据字典
- QA知识图谱
- POI知识图谱

- 潜客识别
- 流失预测
- 业务预测

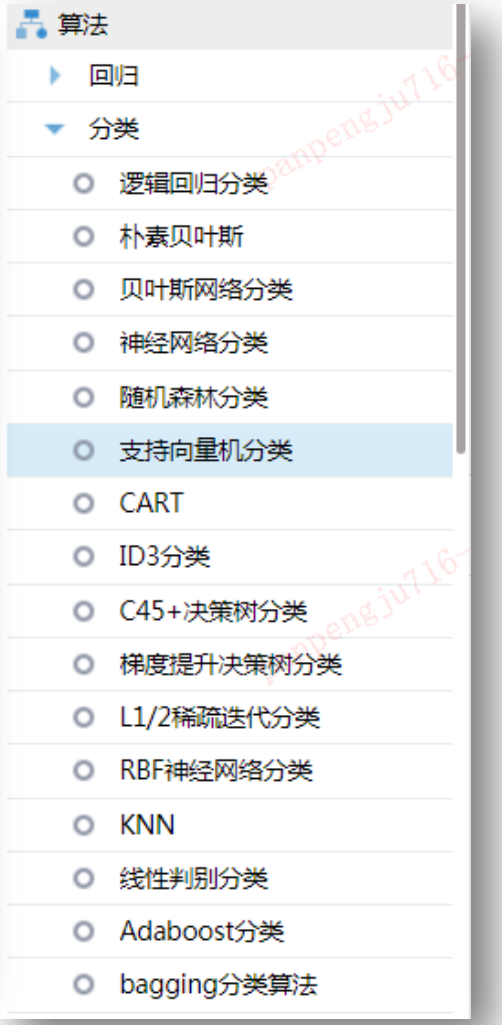
算法平台

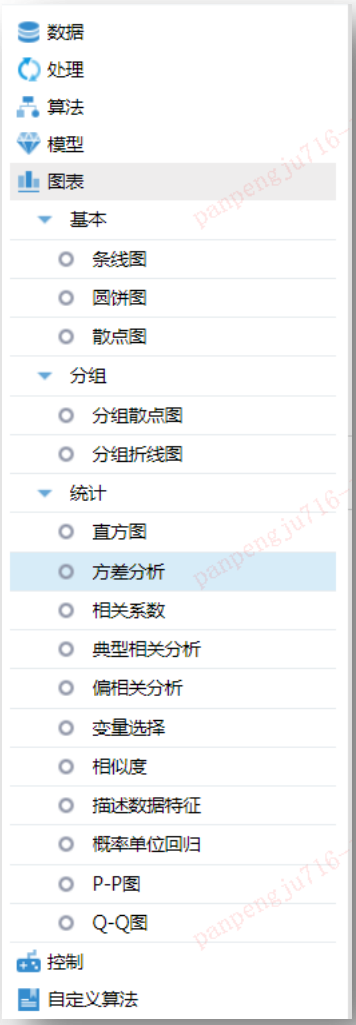

工程中心
算法平台测试
高尔夫用户预测
建模
洞察
部署
要素库
全部
流程1
数据
处理
算法
模型
图表
控制
自定义算法
文件输入
设置角色
缺失值处理
逻辑回归分类
分类评估
EVAL
模型输出

新的框架？

算法组件

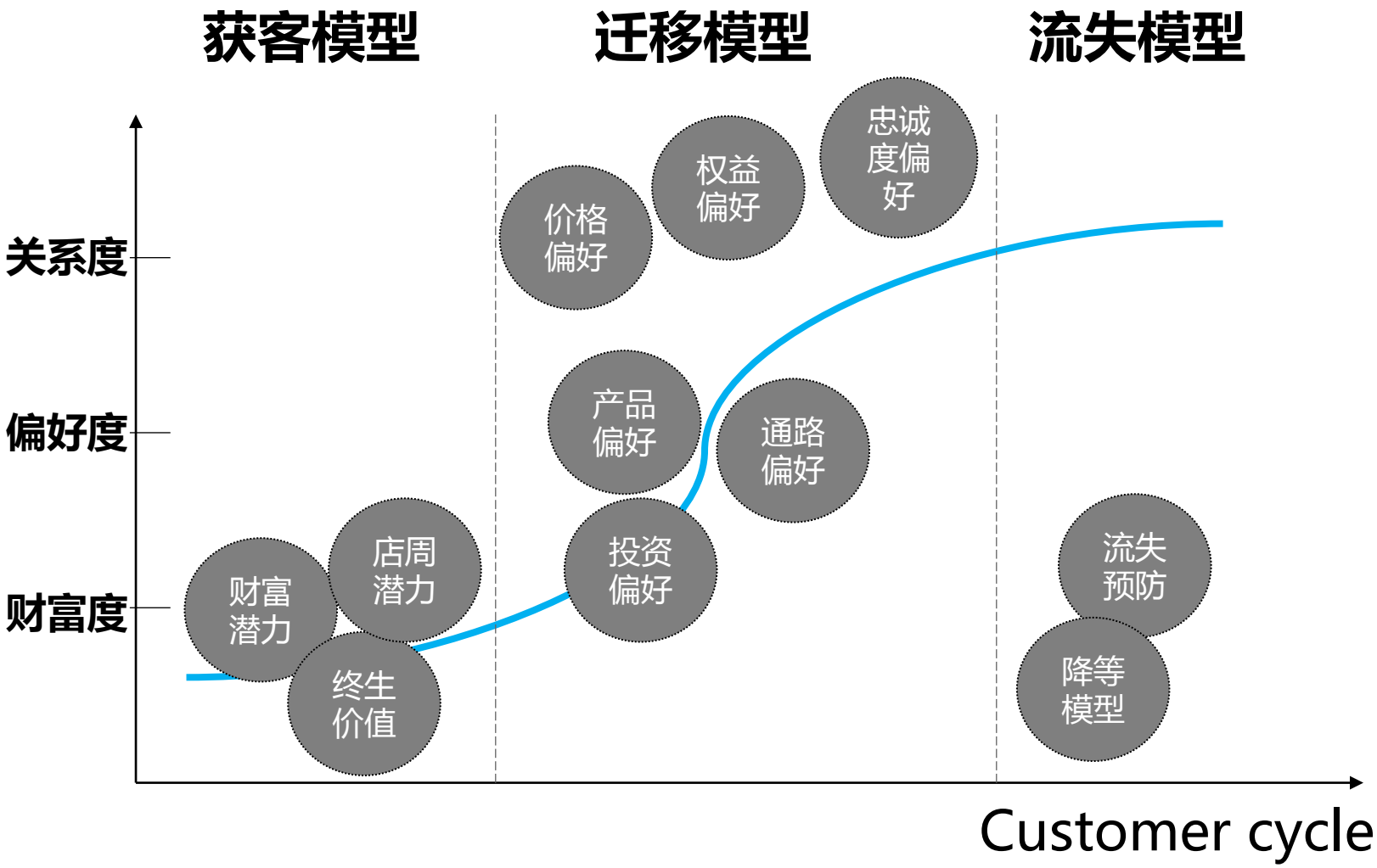
- 数据源->处理->算法->评估->部署





Case:客群管理

- 基于生命周期进行管理

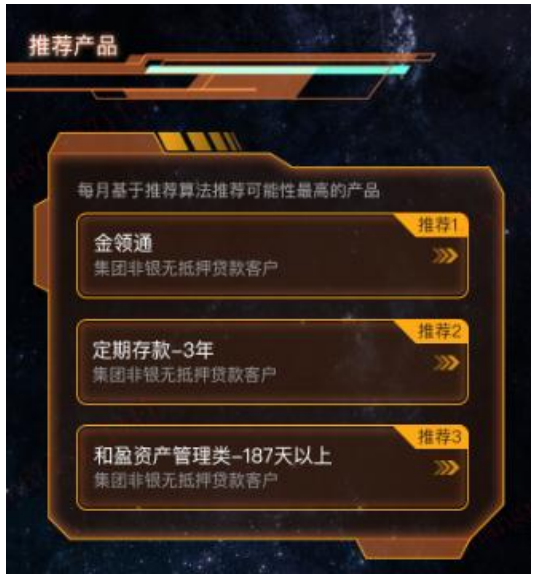


模型	运用方式
财富潜力	预测客户现行财富
终生价值	预测客户未来财富
店周潜力	寻找分行周边的潜力客户
产品偏好	预测客户的购买特定产品的机率
流失预防	预测可能流失的客户

Case:客户智能分析

对接大数据标签库，集成客户产寿信用卡信息

基于模型计算客户和盈购买倾向和可支配资金区间

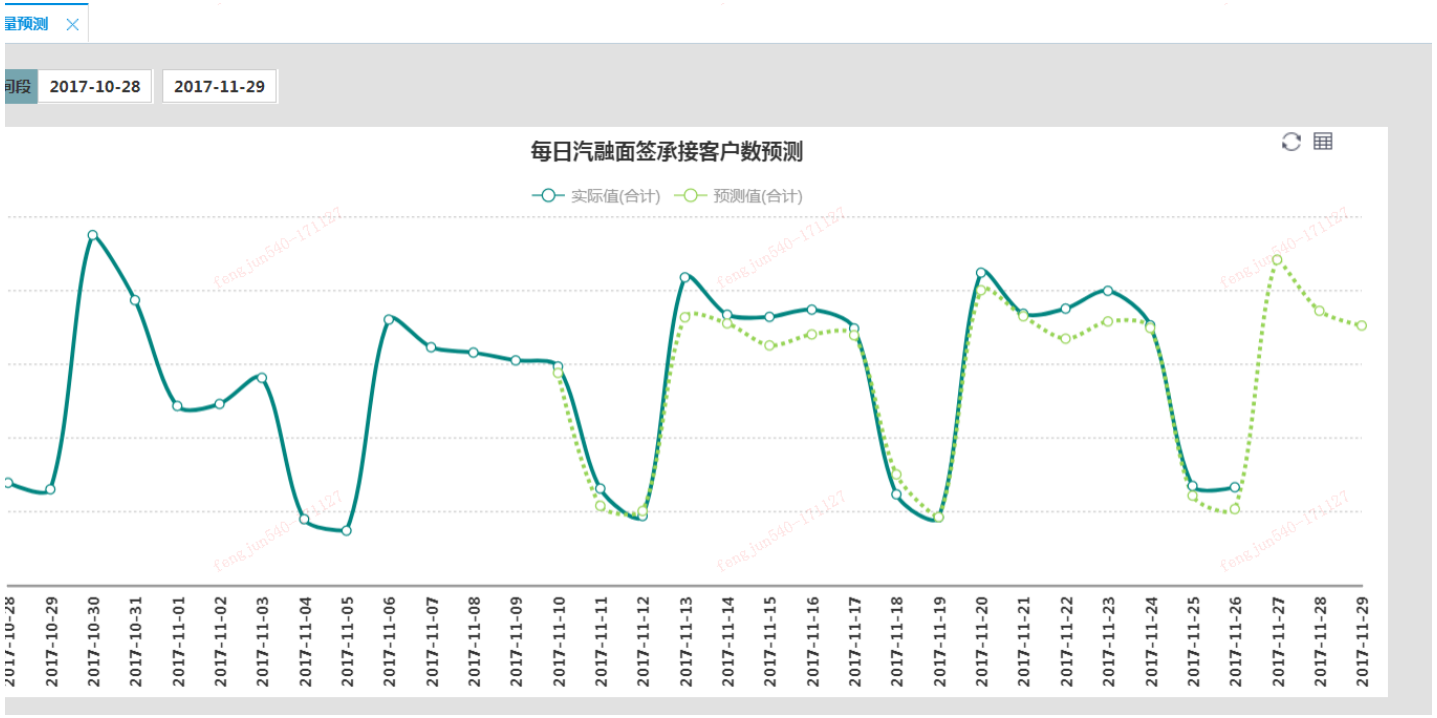


对接推荐产品，支持具体推荐规则展示，直接跳转详情页，产品购买闭环。

计算客户未来两个月流失概率，挖掘异常流失因子，指引接触动作。

Case:业务预测

某业务进件预测，提前预测D+1~D+3，绝对误差<9%



训练数据：不到半年

预测方法：ensemble

- Arima
- Tar
- 指数平滑
- 历史同天平均值
- 上月同一天

回归

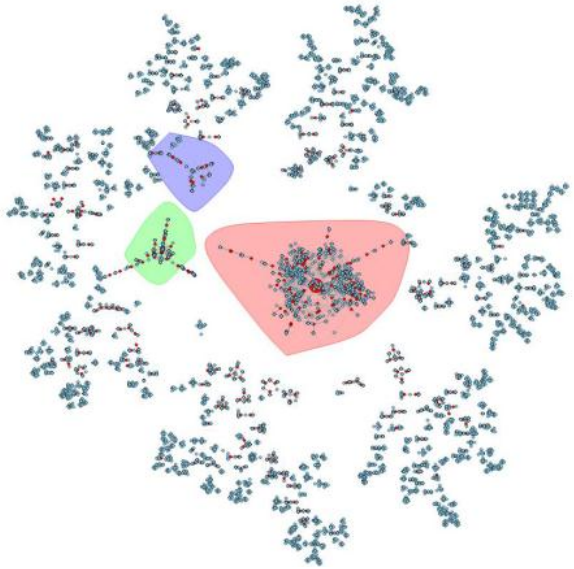
规则

最终结果

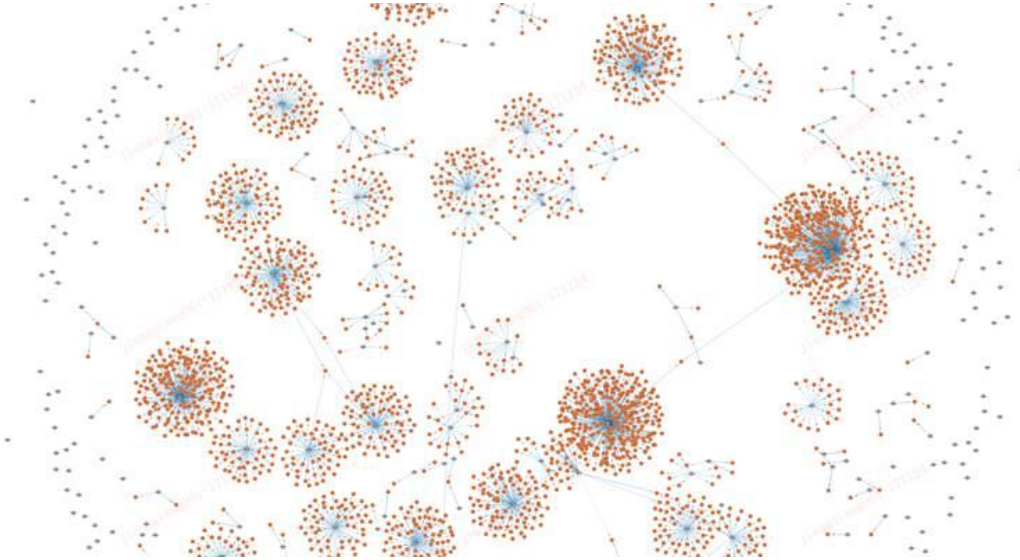
图谱

知识图谱在复杂风控中的应用：风险聚类，案件探索，复杂规则挖掘。

疑似羊毛党

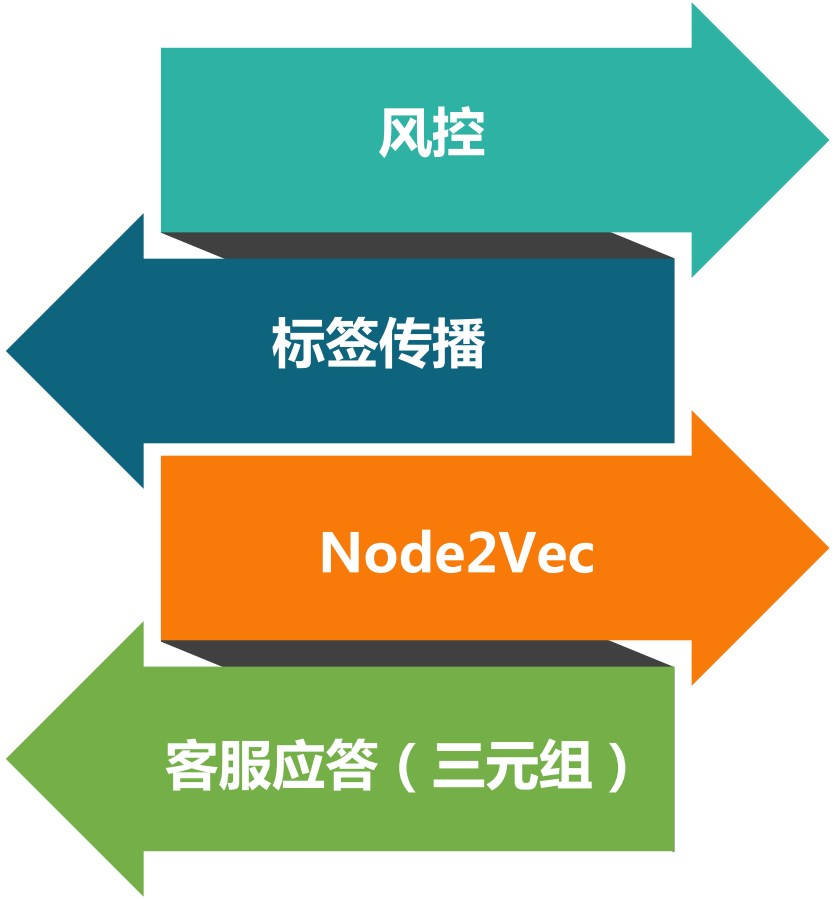


账户、卡号、电话、进件的关系



上亿的节点、10+种的实体关系

图谱应用



新进件联系人和历史黑名单重合三角关系

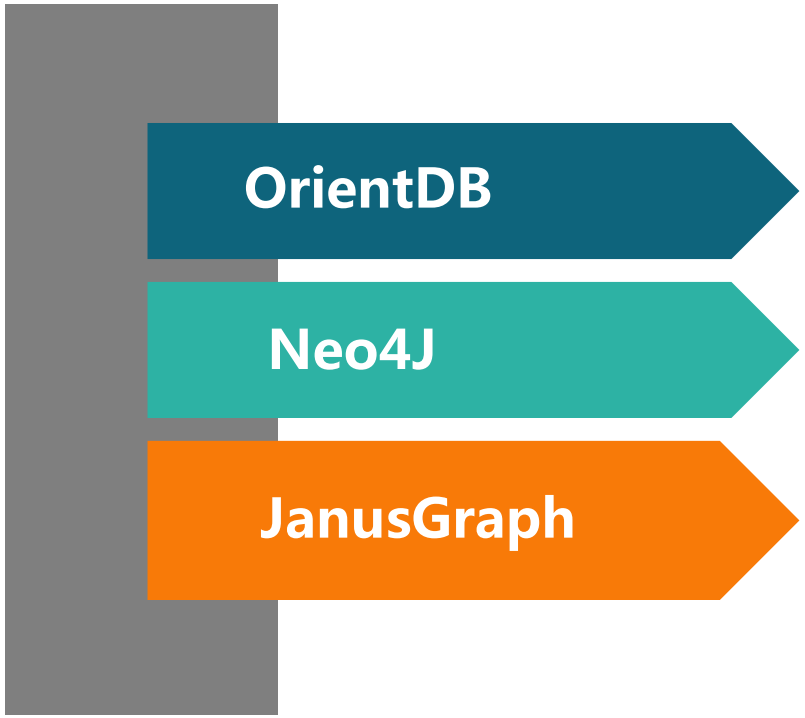
坏人传播算法

创造新feature

语义理解

Node2vec: https://github.com/aditya-grover/node2vec

图数据库



企业级应用，可定制开发

扩展性低，独立索引体系

存储：HBase，索引：ES

GraphX的定位？

地址标准化？

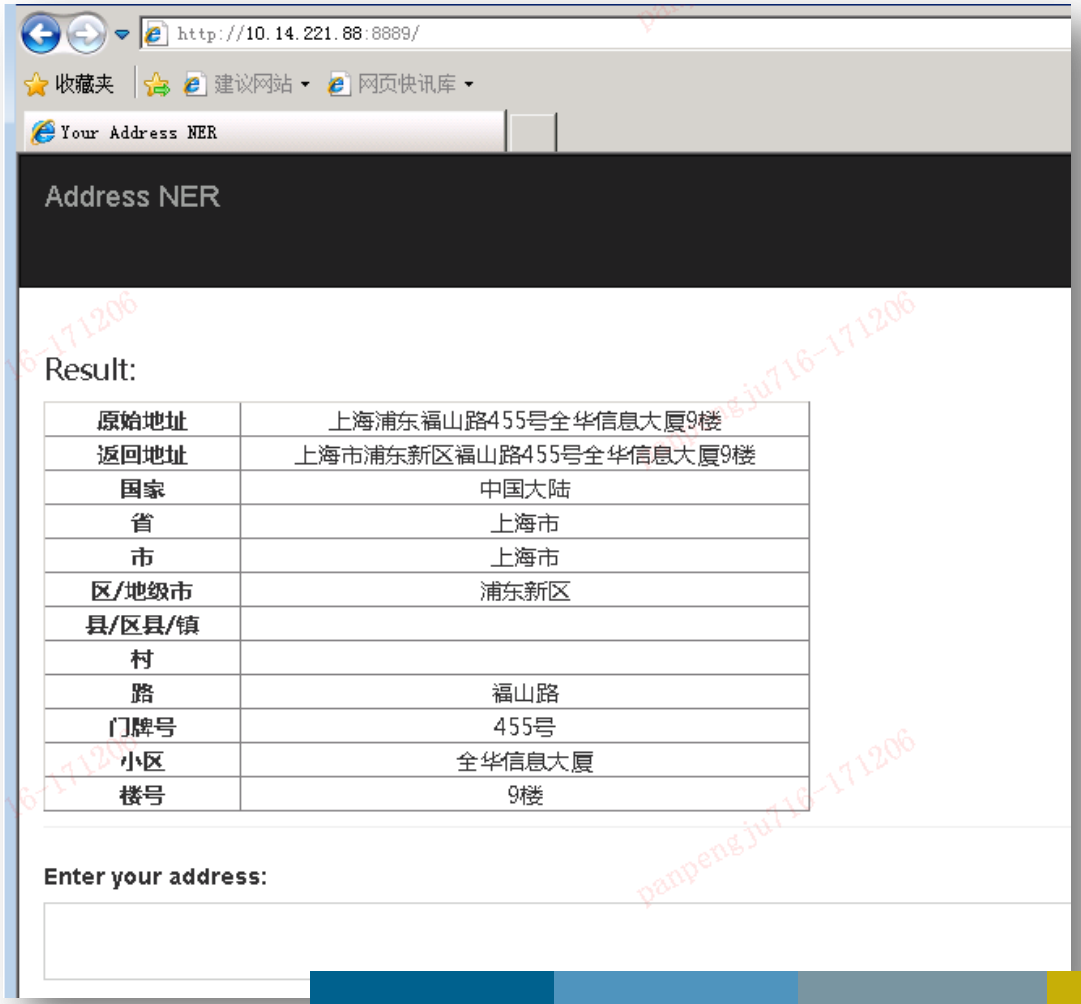
ADDR: 上海浦东福山路455号全华信息大厦9楼

问题	例子
全角地址	重庆市九龙坡区红狮大道１０号３幢２６－６
重复地址	深圳市宝安区观澜镇大水坑村鸿观科技园深圳市宝安区观澜镇大水坑村鸿观科技园
近义词	重庆九龙坡红狮大道10号3栋26-6 重庆市九龙坡区红狮大道10号3幢26-6
填写的区域不是官方的行政区域	['成都市', '高新', '西区', '合作', '路', '888', '号']（高新西区不是行政区）
道路的识别	['沈阳市', '苏家屯区', '金桔', '二路', '8', '-', '1', '号', '3', '2', '室']] ['湖南', '长沙', '雨花', '挂花', '路', '187', '号']]
地址填写顺序不一致，缺少部分行政层级	[' ', '吕', '城镇', '圣旨', '西路', '41', '号']（缺少省和市） ['沈阳市', '苏家屯区', '金桔', '二路', '8', '-', '1', '号', '3', '2', '室']（缺少省）
直辖市有多一级或少一级行政区域的情况	['广东', '深圳', 'null', '罗湖区', '华裕', '花园', 'A19B'] ['广东', '深圳', '罗湖区', '华裕', '花园', 'A19B']
同一地址多种表述	['广东', '深圳', 'null', '南山区', '西丽镇', '腾飞', '苑', 'B座', '303'] ['广东', '深圳', '南', '山西', '丽新围', '村', 'B', '栋', '303']
缺少省、市	东直门外大街天恒大厦8层
地址纠错	

地址标准化！

Restful API:

http://10.14.221.88:8889//api/v1000/single_address?address=上海浦东福山路455号全华信息大厦9楼



算法

规则

+

CRF

Bi-LSTM+CRF?

地址相似度

Restful API:

http://10.14.221.88:9090/rest/nlp/v1/address_similarity?addr1='上海福泉路458号'&addr2='上海福山路456号'

算法 规则 + 图谱 + LR

AI产品经理

中国平安 保险·银行·投资

平安银行 PING AN BANK

欢迎加入平安银行

NLP算法工程师
图像算法工程师
语音算法工程师
机器学习工程师
AI产品经理
数据产品经理
大数据架构师
各种数据人才~~~

潘鹏举

immortalness

Scan the QR Code to add me on WeChat

– THANKS –