

反欺诈技术交流

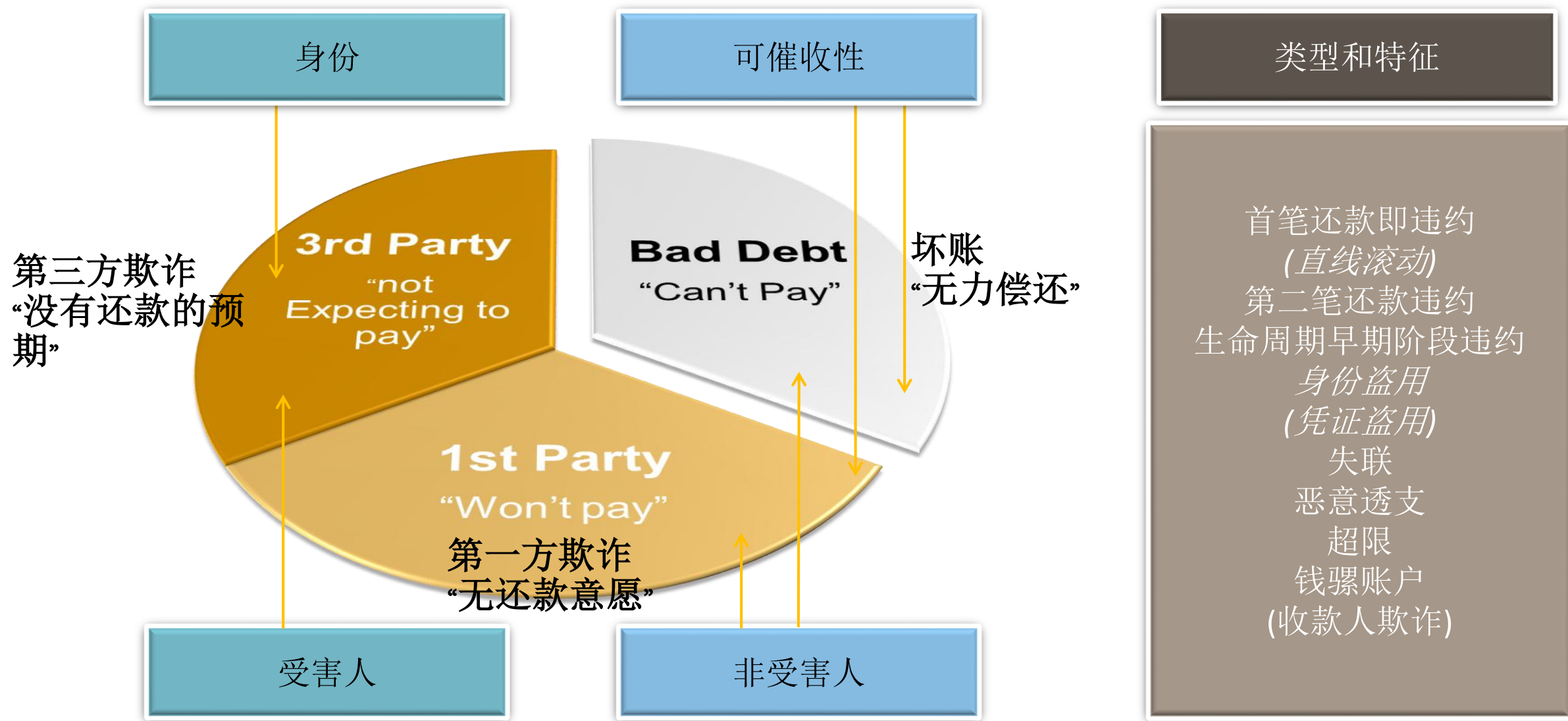
常国珍

本次目标

- 银行贷款申请欺诈简介
- 欺诈检测方法
- 先进案例
- 建模思路

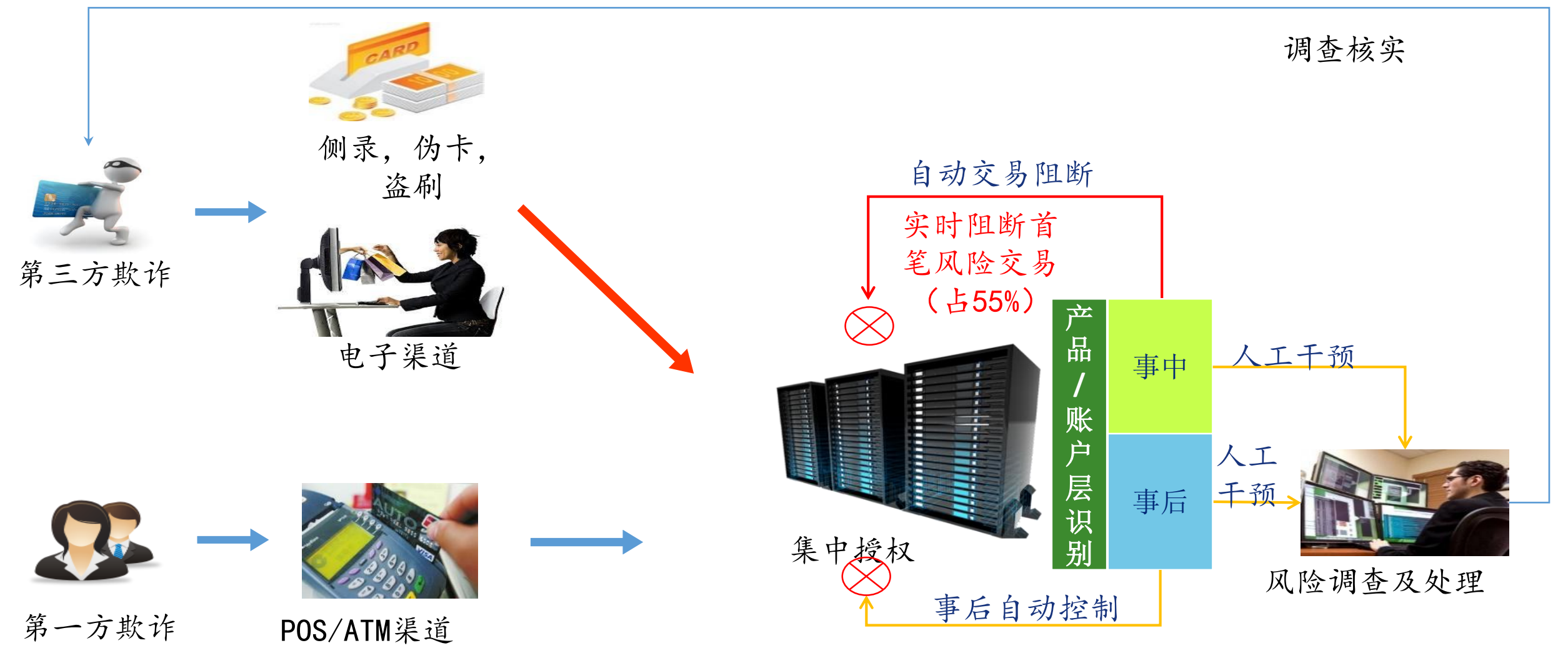
1银行贷款申请欺诈简介

申请欺诈理论上的分类

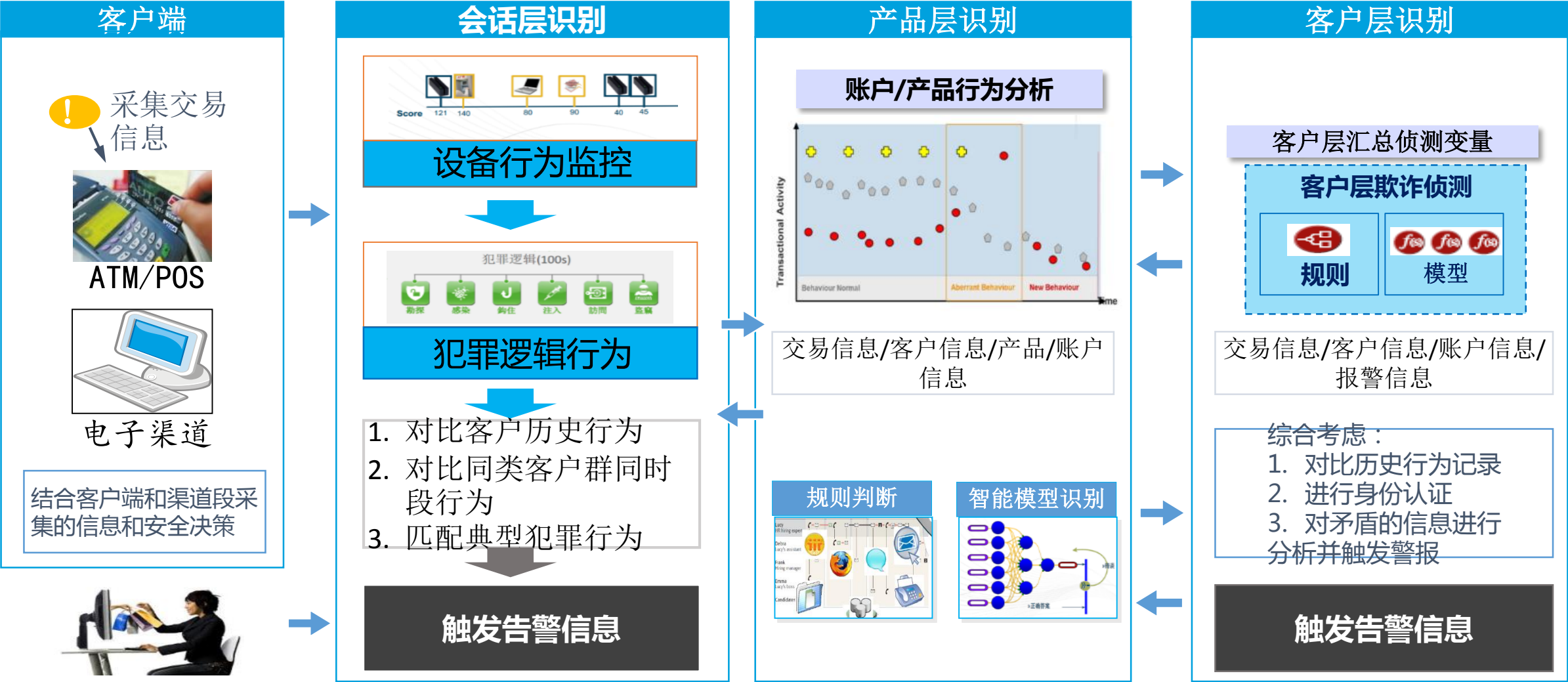


2 欺诈检测方法

银行卡欺诈典型防控场景

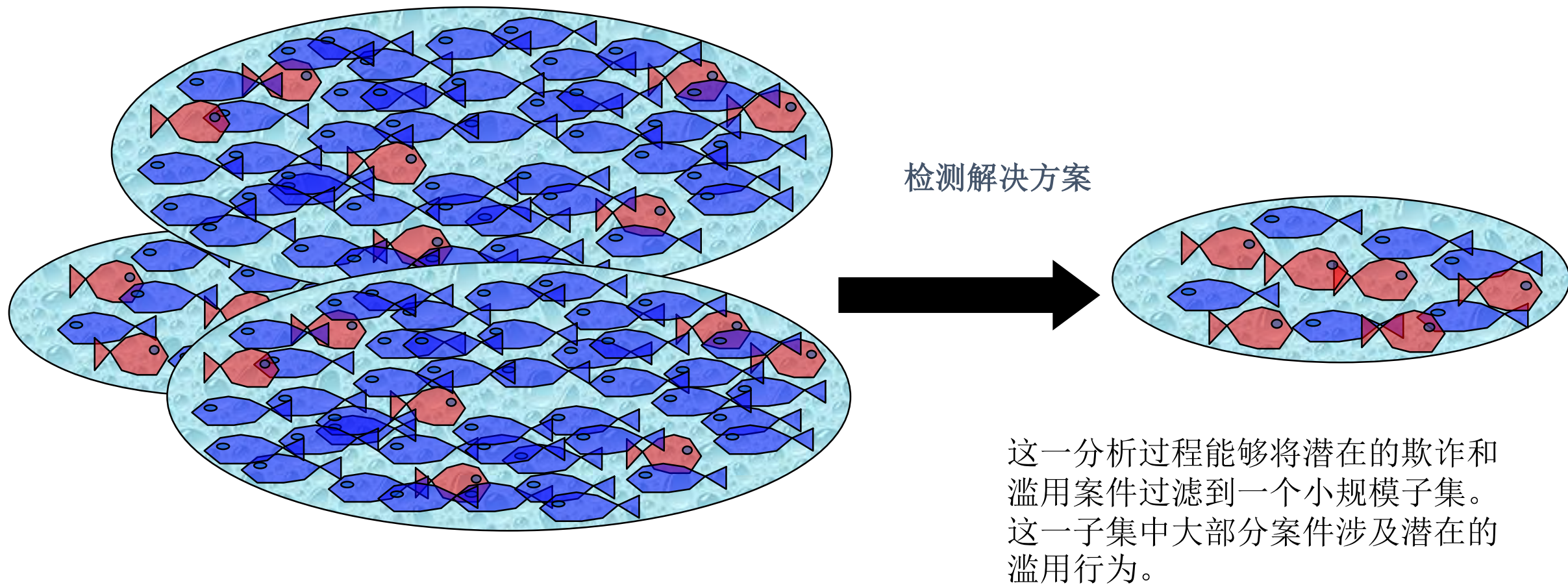


银行卡欺诈防控体系

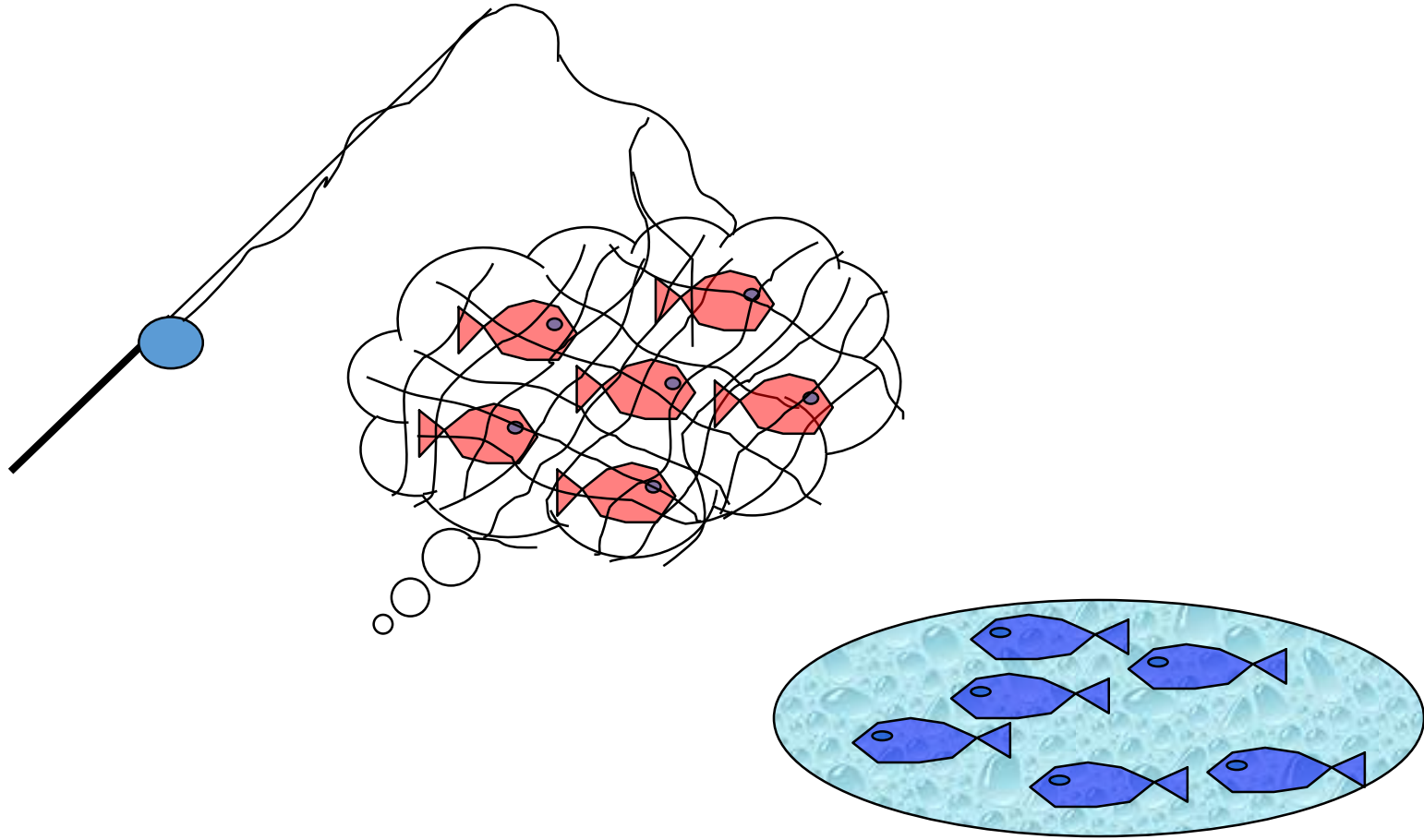


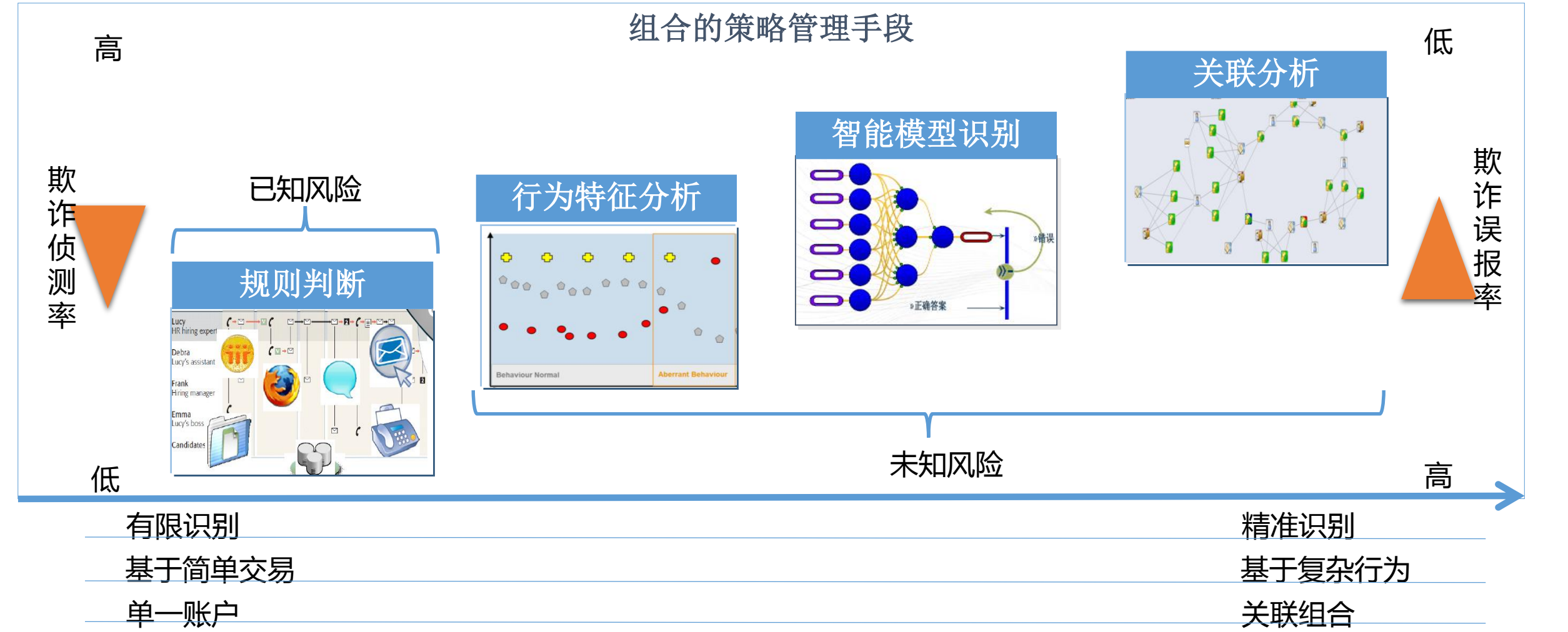
检测思路

- 检测：将潜在的“坏”活动隔离

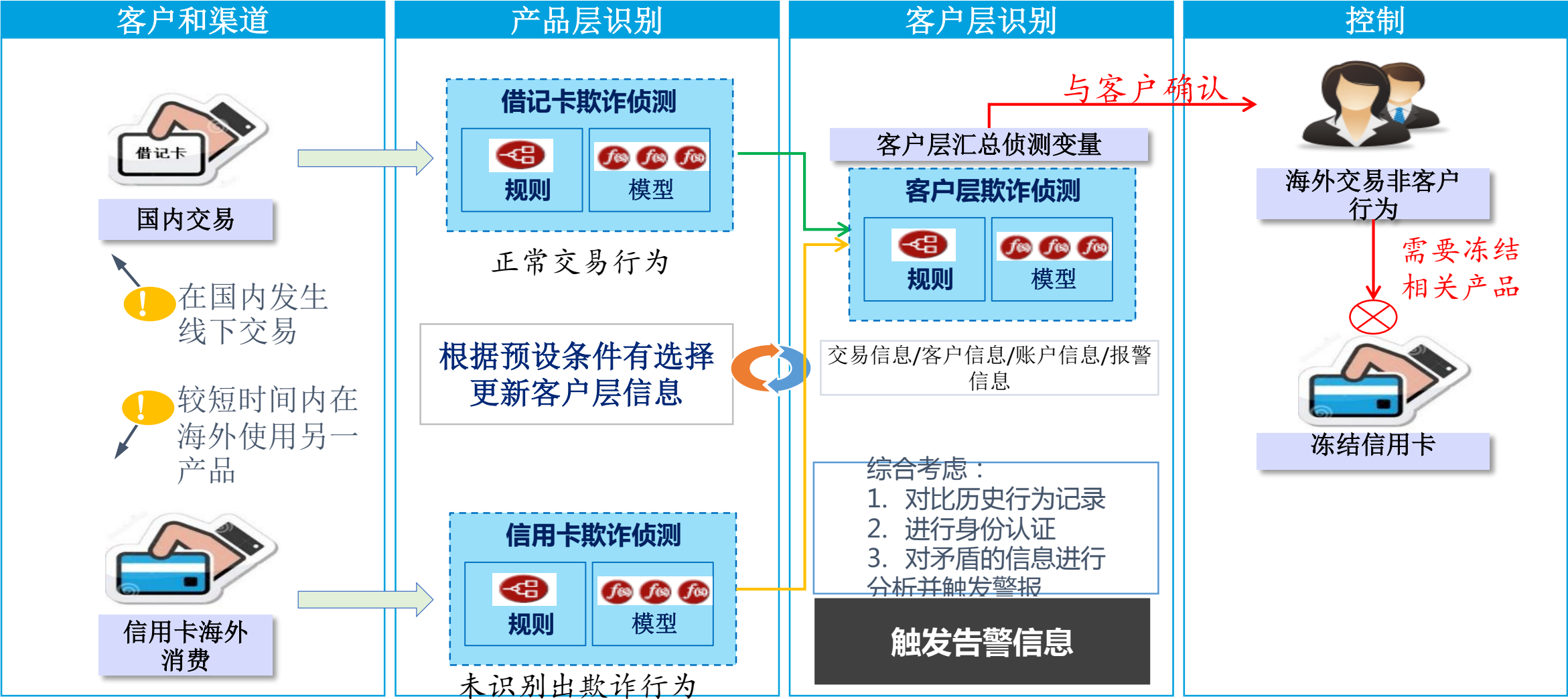


人工审查

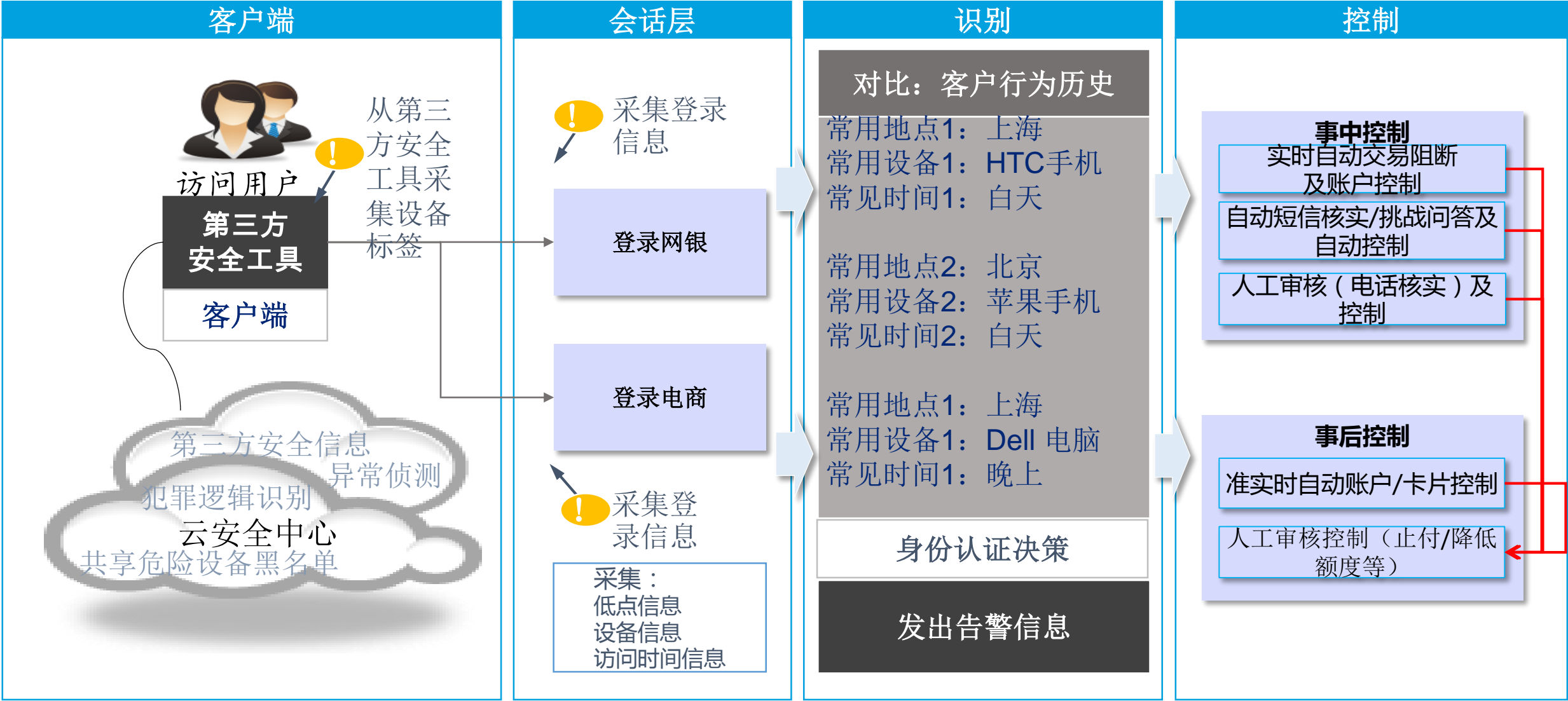




检测技术：规则判断

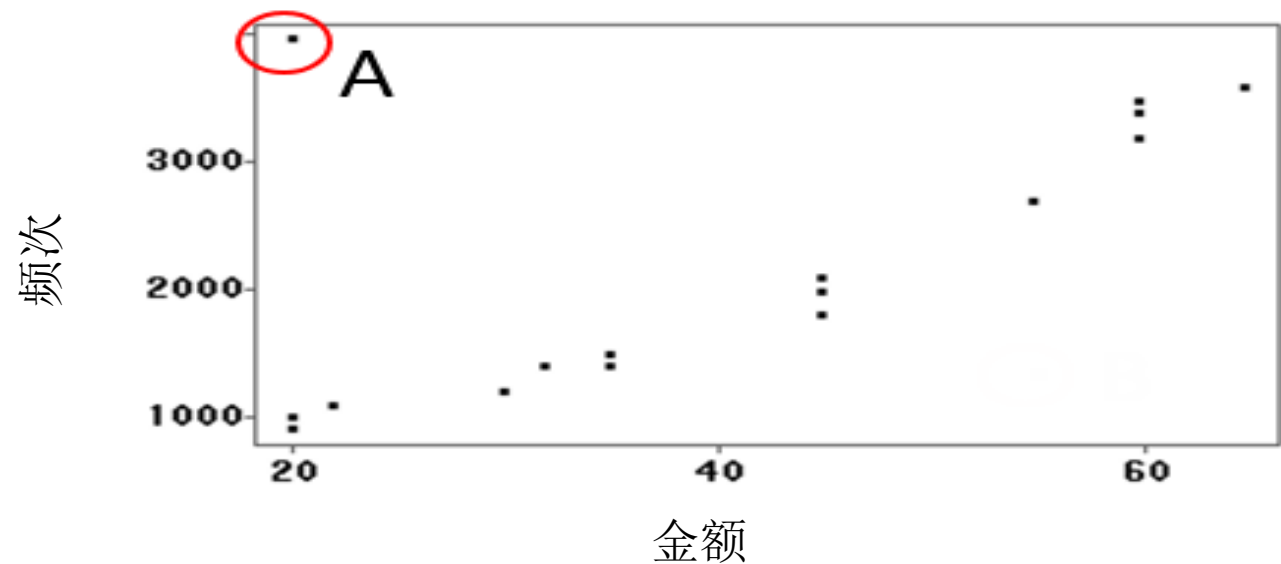


检测技术：行为特征分析



检测技术：异常值发现

- 异常值发现模型是基于行为分析的方法。
- 欺诈行为多多少少都会变现出异常行为。比如小额多频的转账。
- 可以通过使用k-means聚类方法。



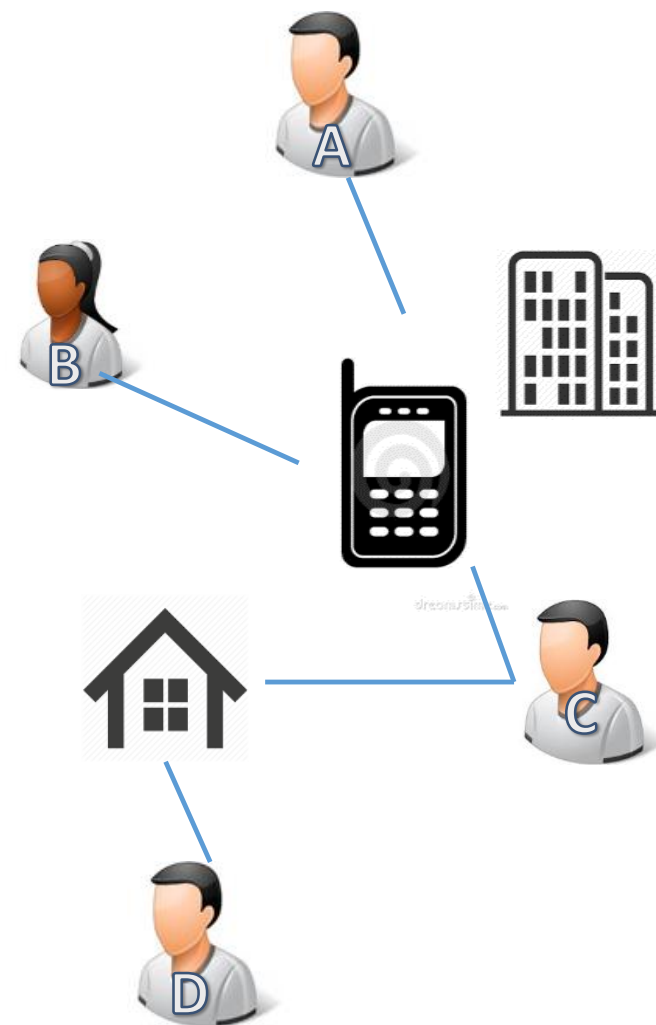
检测技术：神经网络

- 欺诈行为规律性不强，难以识别，因此线性模型难以奏效。目前两大商业欺诈监控软件均选择神经网络作为建模工具。
- 讲申请人、申请机构、申请渠道等方面的信息纳入模型。

FICO（费埃哲）的Falcon产品	SAS公司的SFM产品
实时的交易欺诈侦测	SAS Connector接口。可以实现与建行的授权、支付等核心系统的消息和事务处理
Profiling 技术	评分引擎支持。
部署神经网络模型	反欺诈神经网络
	趋势分析
灵活的规则管理	灵活的规则管理
企业级反欺诈支持	企业级应用
	强大案件管理
全球联盟模型知识库	支持落地的客户自主模型开发

有组织的欺诈

- 各申请人之间有一些关联，此类欺诈通常有其“标志性特征”。
 - “标志性特征”可以是申请人类型和产品类型。
 - “标志性特征”可以包括位置。
 - 可通过雇主、地址、电话号码或命名的个人相关联。
- 需要借助关联分析工具来识别有组织的欺诈“网络”。
 - 申请人C与申请人B和A通过共同的雇主电话相关联；同时申请人C与申请人D通过共同的地址（模糊匹配）相关联；至此形成一个网络。

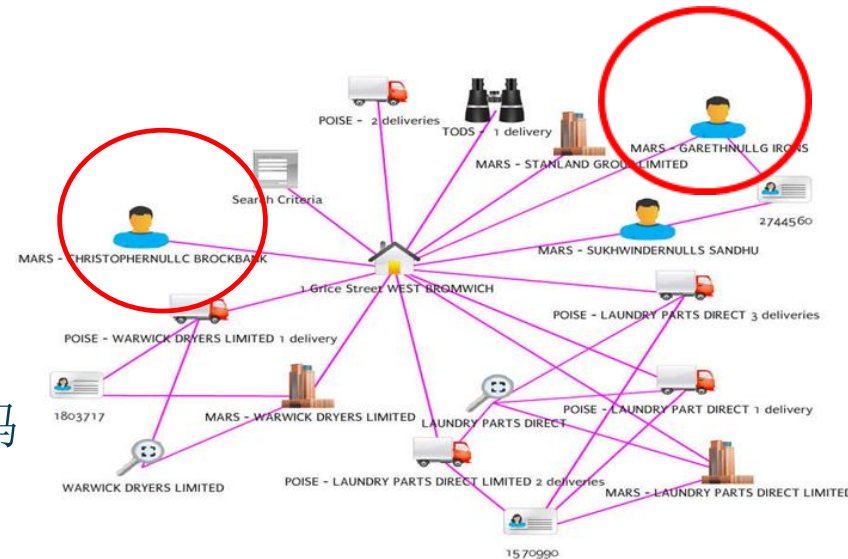


检测技术：关联分析

- 关联分析通过追踪“实体”之间的“关联”，发现那些“不明显的”关系。
- 关联分析的主要对象是“有组织的欺诈”。
- 关联分析通常使用“模糊匹配”来创建关联。
 - 姓名匹配、地址匹配等等

可以通过以下特征将两个申请人相关联：

- ❑ 共同的地址（一级关联）
- ❑ 共同的电话号码和地址（二级关联）
- ❑ 共同的电话号码、地址和另一个共同的电话号码（三级关联）



检测技术：汇总

商品更新频率，货物购买速度

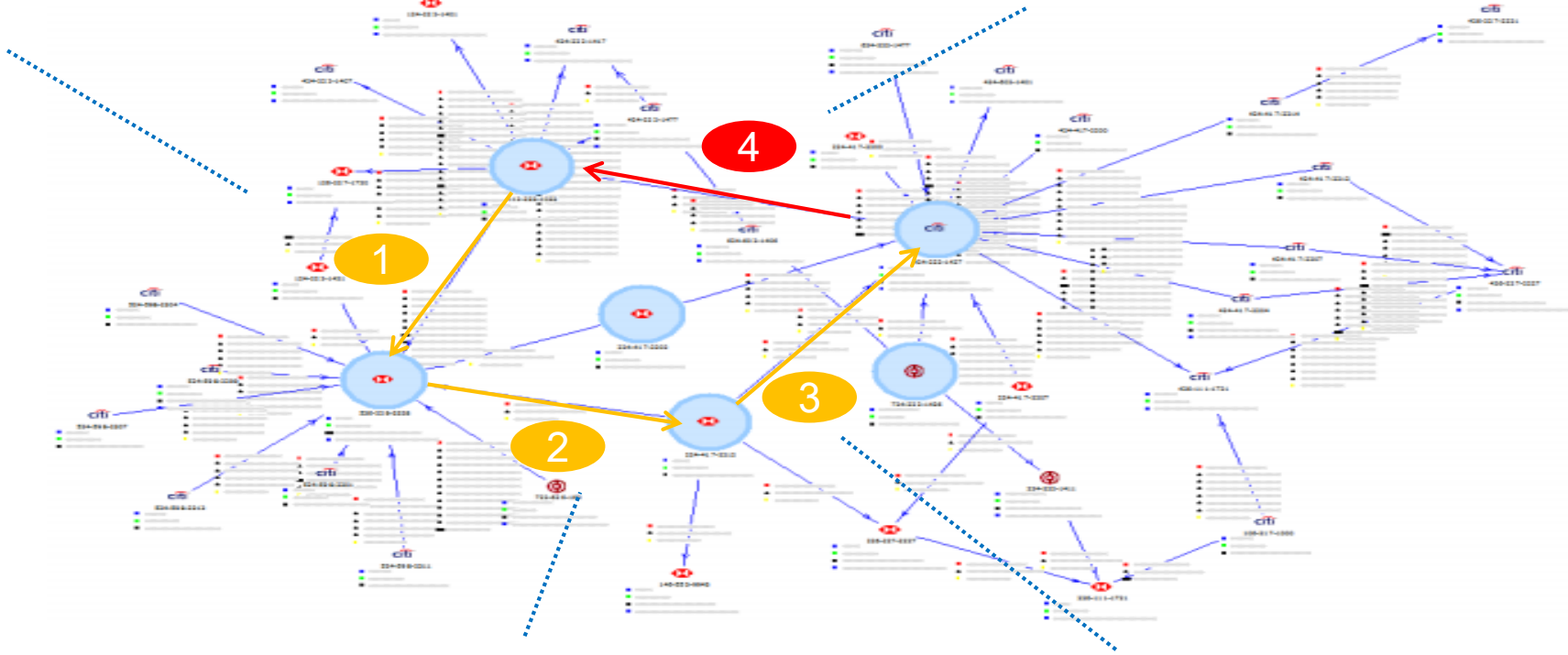
- 1. 监控监控商品更换速度。
- 2. 对比商业活动频率

实际控制人的身份调查，周期性客户调查

- 1. 社会信息关联，发现实际控制人
- 2. 周期性对其客户进行背景调查

对比商户
行为档案

可疑活动
监控



基于风险
的客户
背景调查

资金流向
关联分析

固定交易对手，不正常交易份额

- 1. 固定交易对手在交易份额中的占比
- 2. 固定交易对手的交易频率

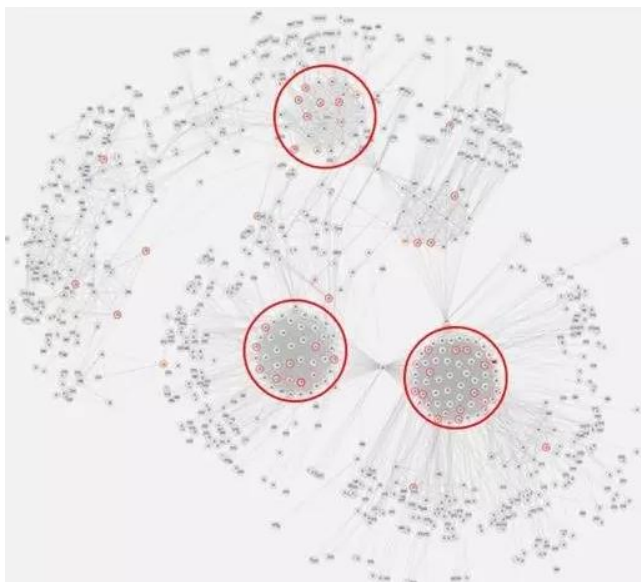
追踪资金走向，还原交易关系

- 1. 跟踪资金走向，侦测循环资金流
- 2. 对高速，大量的资金流动进行重现

3 先进案例

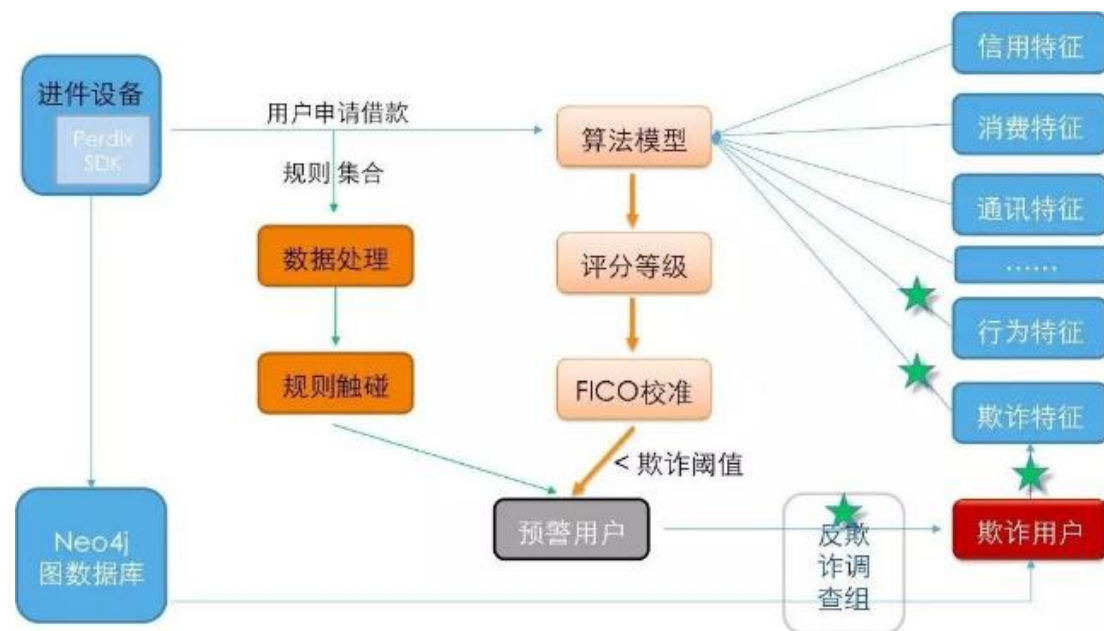
先进案例

在贷前审批上的运用



众安依照“物以类聚、人以群分”的思路，发现欺诈高发的群体；
执行“射人射马、擒贼擒王”的方法，排除黑中介。
腾询构建聊天人群的复杂网络，用于进行传销等事件的侦测。

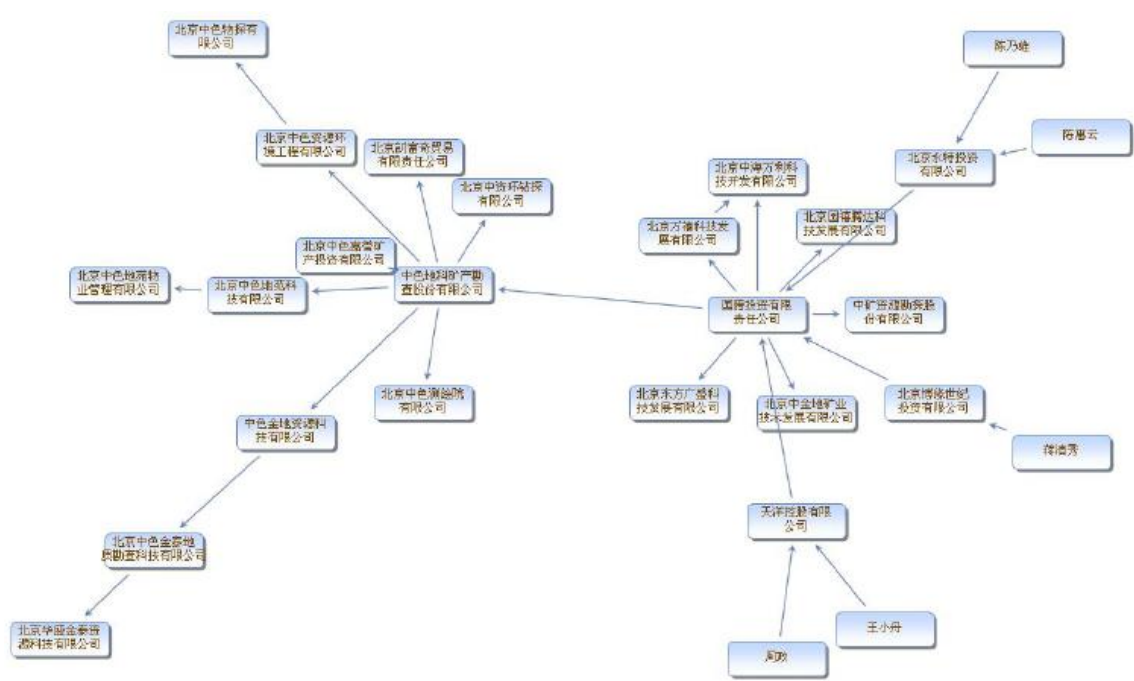
宜人贷将传统反欺诈技术与知识图谱、复杂网络相融合。



参考：<http://www.docin.com/p-1957163916.html>

先进案例

在贷后监控上的运用



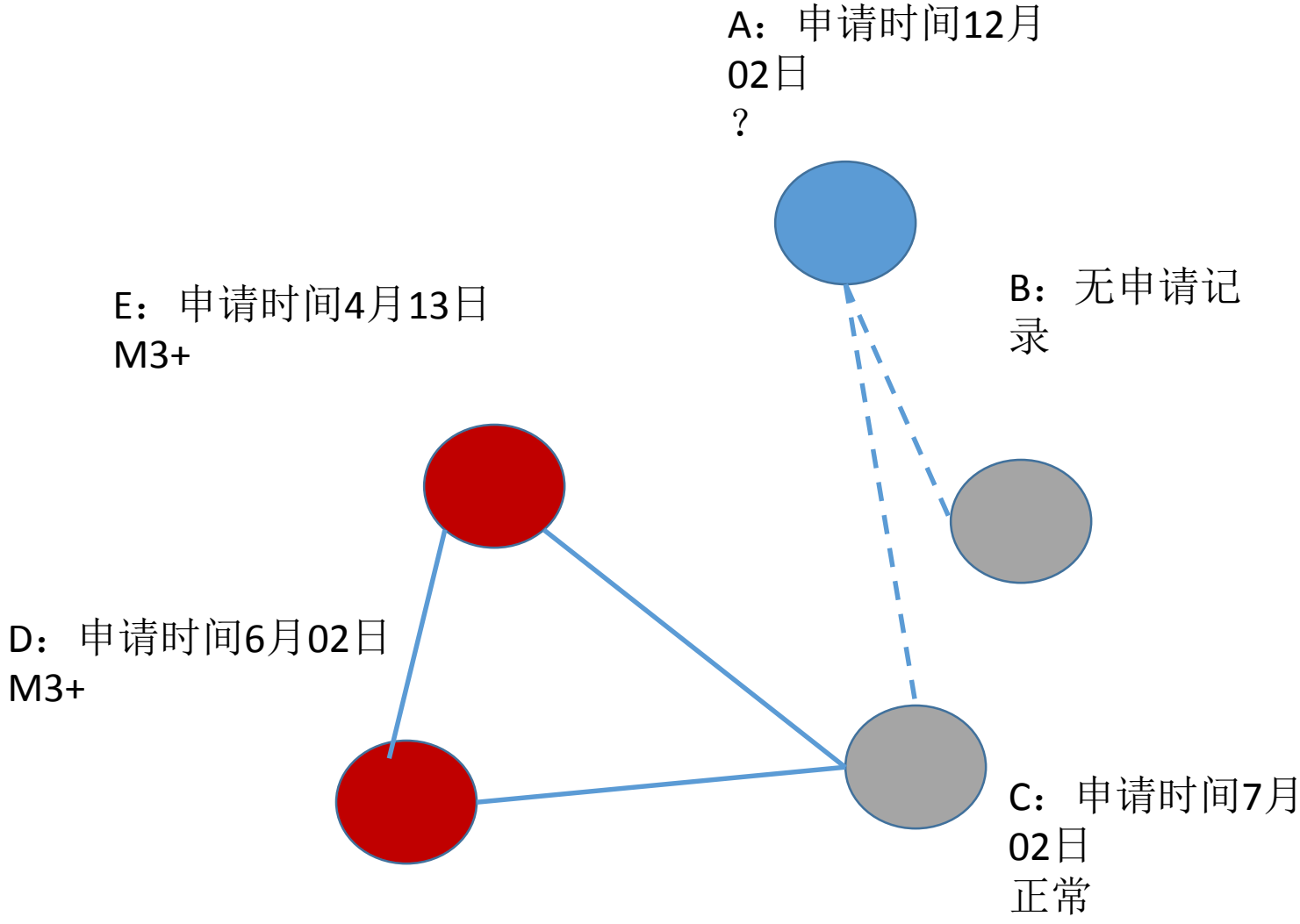
企业关联知识图谱在业界已经广泛运用于贷前审批和贷后监控，并以后者的运用为主。其理论依据是传播模型，也就是“近朱者赤、近墨者黑”原则的量化运用。

众安科技依照个人通信关联网络，以“近朱者赤、近墨者黑”为原则，预测信用恶化客户。

人行征信研发了基于企业贷款担保的知识图谱体系，用于发现过度担保等问题，避免大规模风险扩散。国开行研发了基于工商总局的股东投资数据的关联图谱，结合人行的风险预警系统，用于及早发现贷后恶化。民生银行研发了基于行内交易数据、工商总局的股东投资数据、人行的担保数据、银联的转帐数据的多类型知识图谱，综合的运用与贷款审批、贷后监控、反洗钱等业务场景。

4 建模思路

贷款申请者的联系人复杂网络构建规则的示例

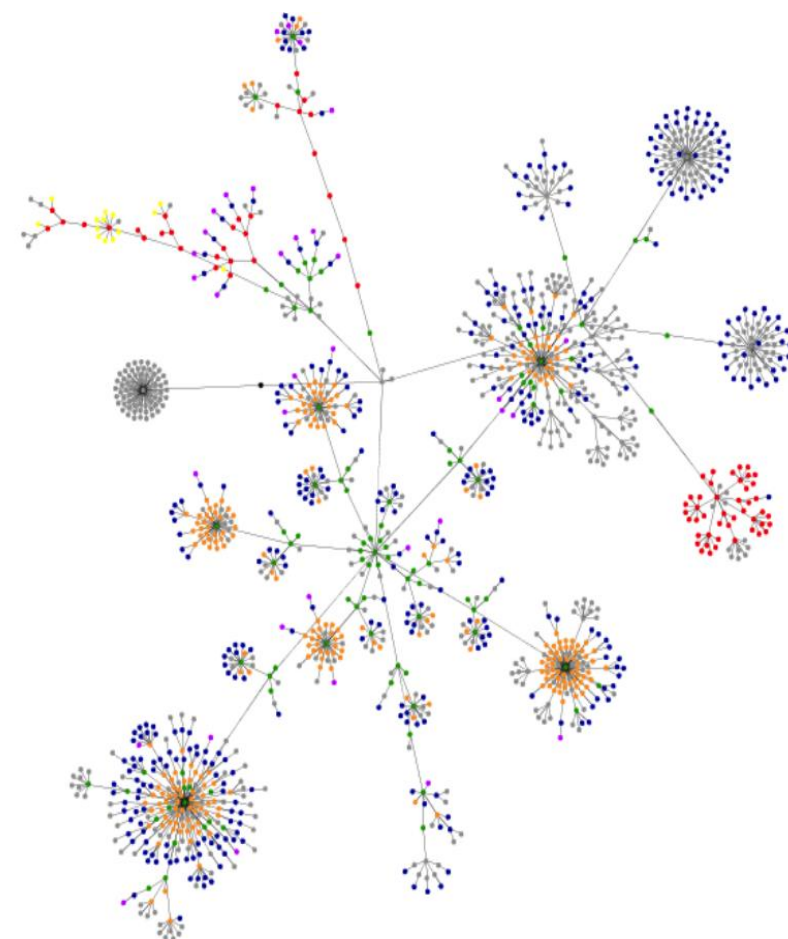


客户A为当前贷款申请客户，他只填写了B和C作为联系人，其中联系人B无申请记录，而联系人C有过申请，而且其与客户DE构成联系人网络。

在已有客户网络中，CDE网络有两个人出现M3+情况，因此申请者A预计会加入一个违约渗透率为2/3的联系人网络。

特征工程-复杂网络特征

图特征	图特征解释
出入度	正常/异常的人的手机号、IP地址出入度不同
子图大小	正常/异常的人所在最大连通子图规模不同
社区大小	正常/异常的人形成的社区大小不同
社区/子图	正常异常的社区与子图大小比例不同
平均近邻 出入度	欺诈者处于欺诈者团伙负责人的周围
pagerank	团伙欺诈的网络中存在重要性较高的节点 而正常网络中一般不会有
最大近邻 pagerank	欺诈者处于欺诈者团伙头目的周围
饱和度	欺诈者所用手机号、IP地址之前有过欺诈历史 饱和度大于0



模型选择

1

- 为使模型有好的可解释性和易用性，考虑基于权重的分类模型和基于决策树的集成学习算法

2

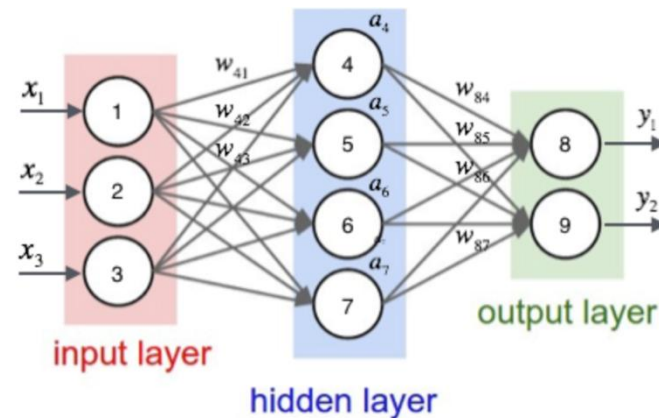
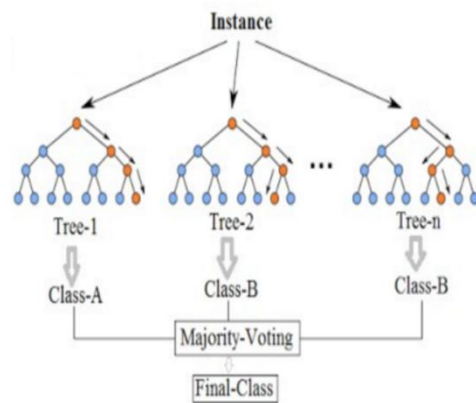
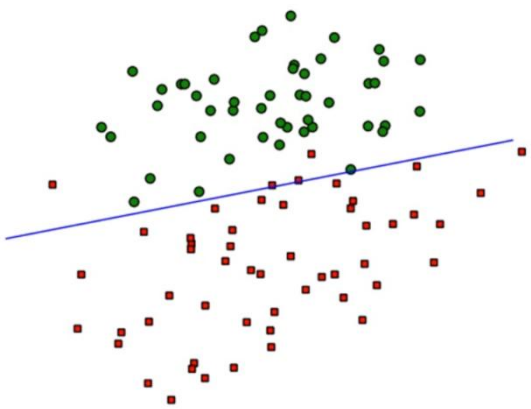
- 基于权重的分类模型，比如逻辑回归模型过于简单，不易发现非线性；比如神经网络，数据量要求大，训练要求时间长。

3

- 基于决策树的集成学习算法，比如 Random Forest, GBDT, XGBoost都有比较好的结果。

4

- 本次建模最终选择的是XGBoost模型。



欺诈团伙分析-训练集欺诈网络可视化

