

# 2017

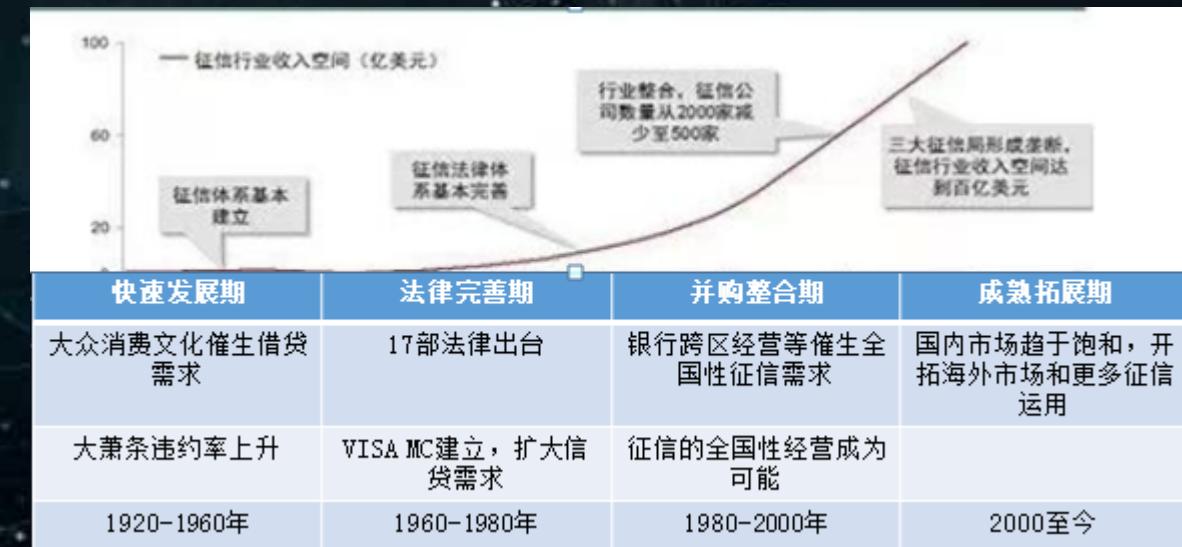
银行业反欺诈思路与方法

---

### 美国银行业反欺诈成绩斐然

- 刷卡欺诈交易占总量的约0.02%。信贷坏账率只有1-2%
- 推荐算法成功，用户忠诚度高

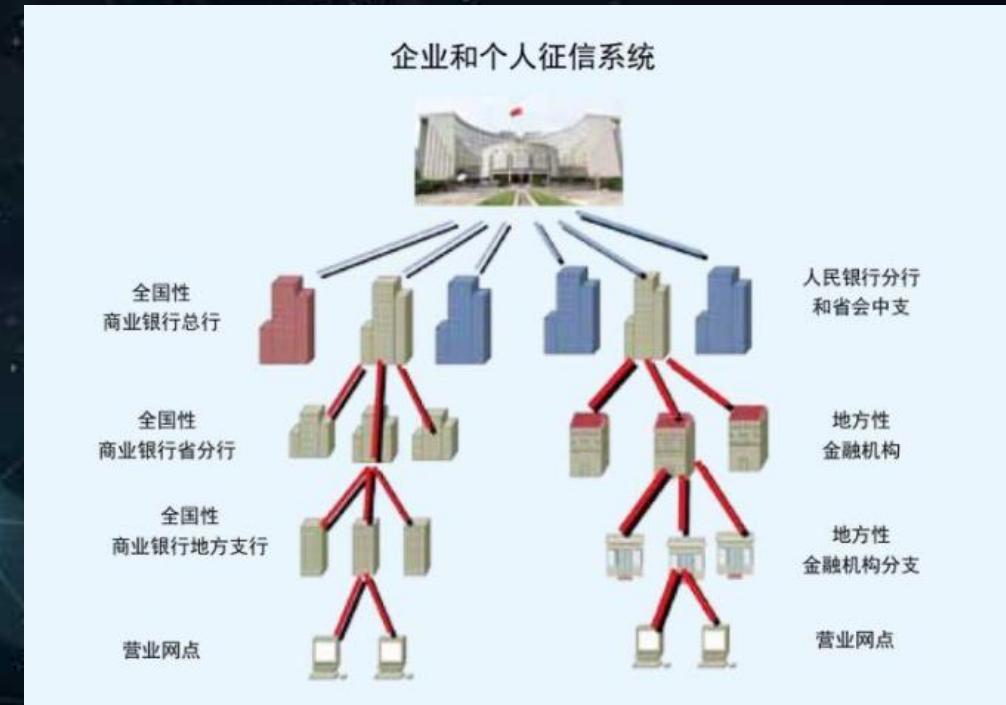
### 美国征信行业发展成熟



## 国内征信业概述

征信（credit reporting）：是指依法收集、整理、保存、加工自然人、法人及其他组织的信用信息，并对外提供信用报告、信用评估、信用信息咨询等服务，帮助客户判断、控制信用风险，进行信用管理的活动。

信用评分：是在信息主体信息的基础上，运用统计方法，对消费者或中小企业未来信用风险的一个综合评估。





银行数据仓库

# CRISP-DM数据挖掘标准流程

商业理解	数据理解	数据准备	建立模型	模型检验	结果部署
确定商业目标	数据的初步采集	选择数据	选择建模技术	评估结果	规划部署方案
形势评估	数据描述	数据清洗	生成检验设计	过程回顾	规划监控和维护方案
确定数据挖掘目标	数据探索性分析	数据构建	监理模型	确定下一个步骤	生成的最终报告
项目实施计划	验证数据治理质量	整合数据	评估模型		项目回顾
		数据格式化 重整数据			
点点点					

➤ 欺诈模式多样性

➤ 比对

信用卡使用两种情况：网上支付和刷卡，同一信用卡在相邻很近的时间内（比如5分钟）在相距较远的区域（上海、广东）完成两笔交易

➤ 比较

➤ 模型法：

采集欺诈客户属性和交易行为属性，得出欺诈规则集或者欺诈可能性评分。

# 数据异常检测

## In the add directory



## Benford定律

## K-means



## Anomaly detection

## 异常检测



➤聚类迁移

➤记录较少的群组



数据集过小或过大 (Too much or too little data)

数据集较大时，可以分为不同的子集，分别进行训练，然后再合成分类器。

数据集过小时，可使用自举技术 (bootstrapping)，从原样本集有放回的抽取 $m$ 个子集，训练 $m$ 个分类器，进行集成。

# 评分模型在信贷周期管理中的作用

审批客户	贷后管理	催收管理
批准/拒绝 定价 确定信用额度 简化流程 复核决策	提高/降低信用额度 白名单 确定影子额度 重新定价 激活/挽留 坏账/催收? 续发信用卡?	优化催收策略 合理分配催收资源
信贷生命周期局侧管理		
申请风险评分 征信风险评分 收益评分	行为评分 交易欺诈风险评分 行为收益评分 行为流失倾向评分 征信风险评分 收益评分	交易欺诈风险评分



### 变量转换

剔除相关系数大的冗余变量  
连续变量Bining处理  
离散变量的类别合并  
IV  
WOE



### logistics

人工选择变量回归  
自动选择stepwise



### 跟踪与检验

Roc k-s

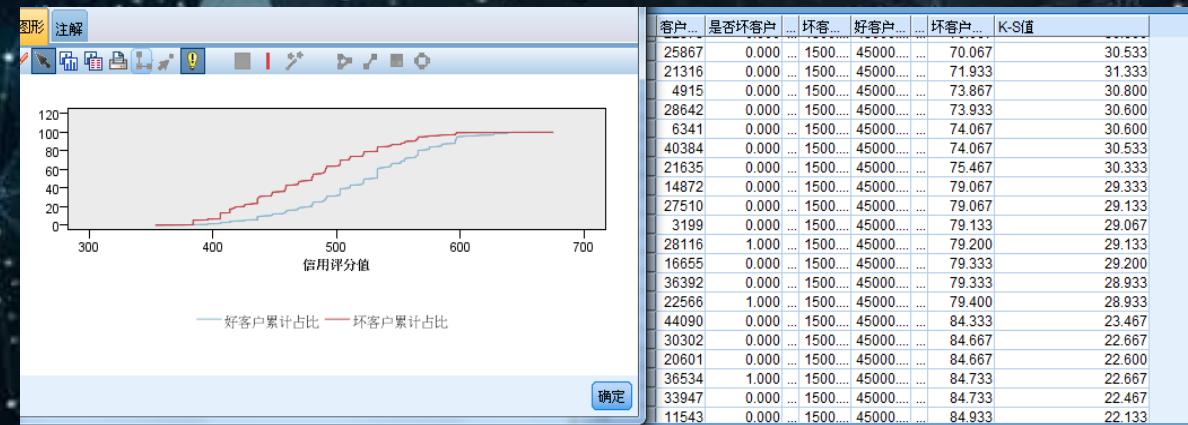


### 评分转化

量表编制

# 模型跟踪与检验

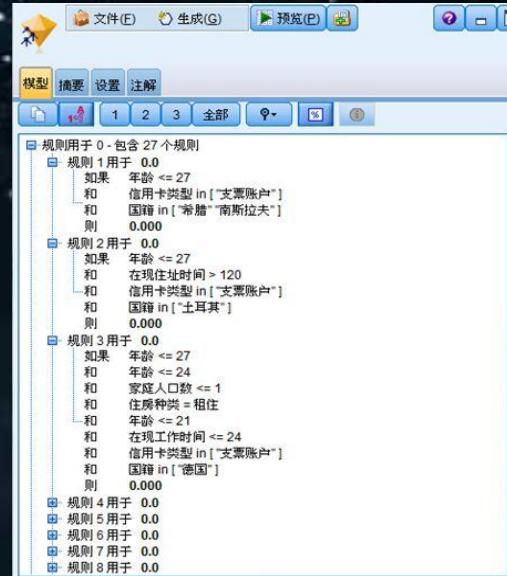
In the add directory



# 单分类器与集成学习

## In the add directory

单分类器：  
如还可以用C5.0决策树：



集成学习：  
基于boosting  
基于bagging  
梯度递增算法

## ➤ Bagging:

Bagging又叫自助聚集，是一种根据均匀概率分布从数据中重复抽样（有放回）的技术。每个抽样生成的自助样本集上，训练一个基分类器；对训练过的分类器进行投票，将测试样本指派到得票最高的类中。每个自助样本集都和原数据一样大

有放回抽样，一些样本可能在同一训练集中出现多次，一些可能被忽略

## ➤ Boosting:

算法的主要思想是在T次迭代中，每次迭代对分类错误的样本加大重采样权重，使得在下一次的迭代中更加关注这些样本。这样训练的多个弱分类器进行加权融合，产生一个最后的结果分类器，提高了该弱分类算法的准确率



# 智能反欺诈交易评分模型

## In the add directory

信用卡账号、持卡人的国家编码和邮政编码  
商户编码 商户的国家编码、  
交易日期和时间、  
交易金额 交易种类（消费、转账、提现等）、  
信用卡额度和可支配余额  
货币代码、邮政编码、  
商户密码核对结果、  
CVV核对结果 信用卡过期日  
支付方式。 . . . .

实时交易授权

历史交易授权

交易登记

比如交易的变量提取：  
以时间/事件为基础

# 智能性的反欺诈交易评估与授权策略

In the add directory

发现一个欺诈账户对于无辜账户的负面影响率

比如说，每抓住一个真欺诈，成本是影响了10个无辜账户，实时拒绝交易，先由客服确认

账户伪阳比例

价值发现命中率

收益和成本

正确判定欺诈交易发现总额除以欺诈交易总额比例

实时实施（销售终端）  
在线实施（上一次风险很低）  
下线实施

# 用户画像

In the add directory

## 标签分类

实时类记录标签		数据挖掘类标签		预测类标签	
客户基本属性	用户自身价值	客户兴趣偏好	客户价值信息	客户风险信息	客户兴趣偏好
人口统计 地理信息 房车自有情况 其他生活信息	生活关联信息 金融关联信息 社交网络信息	金融产品偏好 非金融产品偏好 公司内部渠道偏好 公司外部渠道偏好	用户自身价值 用户对于本公司价值	用户风险评价 黑名单信息	近期需求

## 案例：手机跑分的反欺诈识别 In the add directory



## 案例:智能消灭羊毛党

### In the add directory



构造指纹系统

签名数据库  
分析IP来源  
设备关联

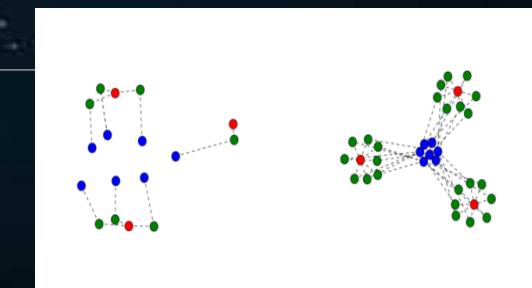
行为细分

用户账户关联  
历史行为偏好

羊毛党典型特征

实时响应

网络图谱模型



# 2017

珍爱信用记录 享受幸福人生