

大数据分析 在征信里的应用

- 征信报告里有什么数据？
- 征信报告如何影响信贷决策？
- 如何构造征信变量与建模？



CONTENTS

1. 信用报告概述
2. 信用报告的内容简介
3. 信用报告的业务应用
4. 度小满对征信报告的解读方法





01

信用报告概述

信用报告概述

信用报告

是个人/企业信用历史的客观记录
记录个人/企业借债还钱、合同履行、遵纪守法等信息



对个人



贷款



信用卡审批



员工录用



任职资格审查



对金融机构

核实客户身份

利用多维度信息
核实客户身份真实性

评估风险水平

结合客户已有负债
情况认知客户风险

判断偿债能力

推测收入水平，评
估客户偿还能力

辅助信贷管理

丰富规则集、提升
模型准确率



概述	<h1>信用报告的产生与应用流程</h1> <div><div>收集信用信息</div><div>形成信用报告</div><div>用户查询/授权查询</div><div>机构查询信用报告</div></div>			
内容简介				
业务应用				
解读方法				
	<h2>主要流程</h2>	<p>信息来自放贷机构、公用事业单位、法院和政府部门等</p>	<p>信用报告存储在中国人民银行征信中心的征信系统内</p>	<ul style="list-style-type: none">• 用户可以自查信用报告• 用户可以授权放贷机构查询信用报告• 金融机构查得最多• 法院和政府部门可依法查询
	<h2>名词解释</h2>	<h3>征信系统</h3> <p>又称为“金融信用信息基础数据库”“企业和个人信用信息基础数据库”；</p> <p>是国家建立的、记录个人/企业信用信息的数据库，由中国人民银行征信中心负责建设、运行和维护</p>	<h3>书面/电子授权</h3> <p>“自本人签署本授权之日起至贷款最终结果确定之日 或偿还之日，授权xx银行通过征信系统查询、打印、保存、使用本人信用报告用于审核贷款申请、担保人资格审查及与贷款申请和贷后管理相关的其他事项。”</p>	

信用报告覆盖面

哪些人/企业有信用报告？

凡是与银行发生信贷关系或者开立了个人结算账户的个人都有自己的信用报告。

截至2020年12月底，征信系统收录情况如下：

11亿自然人

在14亿自然人中，征信系统共收录11亿人的信用信息



6092.3万户企业及其他组织

在近1.4亿户企业中，征信系统共收录6092.3万户企业及其他组织的信用信息



3656.1万户小微企业



1167万户个体工商户



02

信用报告内容简介

征信报告主要内容



基本信息

- 身份信息
- 居住信息
- 职业信息...



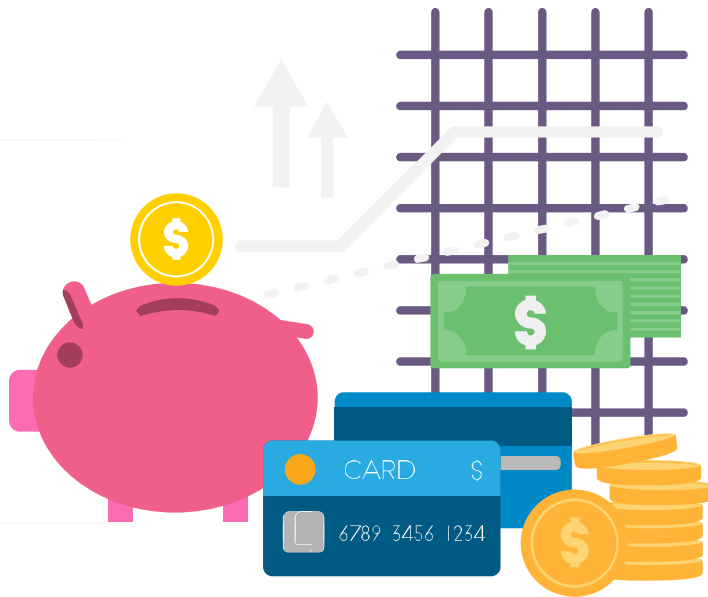
信贷信息

- 借债还钱信息
- 信用报告中最核心的信息



非金融负债信息

- 先消费后付款形成的信息，如电信缴费



公共信息



- 社保公积金信息
- 法院信息
- 欠税信息
- 行政执法信息等

查询信息



- 过去 2 年内，何人何时因为什么 原因查过信用报告

01 报告头



包含标识信息、信息主体的证件信息报告的交付对象信息：

- 报告编号、查询请求时间和报告时间
- 被查询者姓名、证件类型、证件号码
- 查询操作员、查询原因

02 个人基本信息



包含人口统计学类特征：

- **身份信息：**性别、出生日期、学历学位等，**其中，手机号、单位电话、住宅电话均为最新**
- **配偶信息：**姓名、证件类型、证件号码、工作单位、联系电话
- **居住信息：**居住地址、居住状况、信息更新日期
- **职业信息：**工作单位、单位地址、职业、行业、职务、职称、进入本单位年份、信息更新日期

03 信息概要



说明信用报告主要包含哪些内容，信息主体的主要违约情况和负债情况：

- **信用提示：**
 - 十类借贷账户中，各类账户的借贷笔数/时间

- **逾期及违约信息概要**

- **授信及负债信息概要**

04 信贷交易信息明细



本部分是信用报告的核心内容，反映借款人办理信贷业务的详细信息，包括借款人授信及负债情况，借钱和还钱的历史信息

- **资产处置信息：**被资产管理公司接收的不良贷款和信用卡的信贷后形成的债务，相关信息包括违约后拍卖房产、车辆以及公司设备等
- **保证人代偿信息：**借款人逾期后，由担保机构或保险机构代为偿还的信息
- **贷款：**借款人在金融机构处的住房贷款、助学贷款、消费贷款等记录
 - 具体包括发放日期、机构类型、贷款机构、贷款种类、担保方式、还款期数等
 - 不同状态下的贷款展示内容有所不同



05 公共信息明细

报告头

个人基本信息

信息概要

信贷交易信息明细

公共信息明细

查询记录

其他

记录信息主体的社会表现，展示信息主体在社会公共部门所形成的正、负面信息

- **欠税记录：**5年内的欠税情况，包括欠税总额和欠税日期等。
- **民事判决记录：**5年内涉及的已结案的民事诉讼信息，主要由各地法院采集
- **强制执行记录：**5年内涉及的法院强制执行信息或被列为失信被执行人的信息

- **行政处罚记录:** 自生效日到查询日5年内, 来自政府部门公开的行政处罚信息
- **住房公积金参缴记录:** 住房公积金缴纳情况, 属于**正面信息, 一直展示**

06 查询记录



记录信息主体最近两年查询记录明细信息，包括个人查询、贷后管理、信用卡审批查询记录等。短期内，查询次数过于频繁属于信用报告或信用记录查询次数异常；如果查询次数过于频繁得基础上，贷款或信用卡未获批，可能资产情况、信用情况等诸多因素开始恶化

- 查询时间
- 查询机构名称
- 查询原因

07 其他



本人声明

- 报告主体对该账户信息附注的简要说明，由本人提出并对声明的真实性负责。

五 本人声明

编号	声明内容	添加日期
1	本人身份证丢失，2011 年 1 月后的新业务均非本人办理。	2010.10.14

异议标注

- 说明信息主体对信用报告中的哪些信息有异议,以及商业银行对异议信息的核实情况

六 异议标注

编号	标注内容	添加日期
1	中国铁通甘肃分公司报送的固定电话并未欠费，异议正在处理中。	2010.12.16

报告说明

- 展示对报告中专有名词等内容的解释与说明

一二代征信主要差异

一、报告头	一、报告头
二、个人基本信息	二、个人基本信息
(一) 身份信息	(一) 身份信息
(二) 婚姻信息	(二) 婚姻信息
(三) 居住信息	(三) 居住信息
(四) 职业信息	(四) 职业信息
三、信息概要	三、信息概要
(一) 个人信用报告“数字解读”	(一) 信用提示
(二) 信贷交易信息提示	(二) 逾期及违约信息概要
(三) 信贷交易违约信息概要	(三) 授信及负债信息概要
(四) 信贷交易授信及负债信息概要	
(五) 非信贷交易信息概要	
(六) 公共信息概要	
(七) 查询记录概要	
四、信贷交易信息明细	四、信贷交易信息明细
(一) 被追偿信息	(一) 资产处置信息
(二) 非循环贷款	(二) 保证人代偿信息
(三) 循环额度下贷款	(三) 贷款
(四) 循环贷账户	(四) 贷记卡
(五) 贷记卡账户	(五) 准贷记卡
(六) 准贷记卡账户	
(七) 其他	(六) 担保信息
(八) 相关还款责任信息	
五、非信贷交易信息明细	
后付费记录	
六、公共信息明细	五、公共信息明细
(一) 欠税记录	(一) 欠税记录
(二) 民事判决记录	(二) 民事判决记录
(三) 强制执行记录	(三) 强制执行记录
(四) 行政处罚记录	(四) 行政处罚记录
(五) 住房公积金参缴记录	(五) 住房公积金参缴记录
(六) 低保救助记录	(六) 养老保险金缴存记录
(七) 执业资格记录	(七) 养老保险金发放记录
(八) 行政奖励记录	(八) 低保救助记录
	(九) 执业资格记录
	(十) 行政奖励记录
	(十一) 车辆交易和抵押记录
	(十二) 电信缴费记录
七、本人声明	六、本人声明
八、异议标注	七、异议标注
九、查询记录	八、查询记录

二代

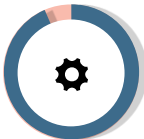
一代

新增授信协议信息



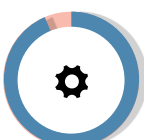
- 非循环贷款(分次放款)、循环额度下贷款、循环贷账户在一定条件下展示授信协议；账户在授信协议下展示。

增加信贷交易信息分类



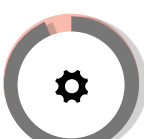
- 各类信贷业务按照被追偿信息、非循环贷款、循环额度下贷款、循环贷账户、贷记卡账户、准贷记卡账户、其他业务、相关还款责任展示

增加相关还款责任信息



- 除包括担保(保证)责任外，还包括其他相关还款责任；
- 除包括有相关还款责任的个人借款，也包括有相关还款责任的企业借款。

其他新增



- 大额专项分期信息
- 特殊事件说明信息
- 已结清贷款、转出贷款和呆账:24个月还款状态
- 未结清/未销户借贷业务:最近5年历史表现信息



03

征信报告的业务应用

征信刻画用户画像

经营画像

用户满意度相关



竞品额度

商业银行类竞品额度最小值
商业银行类竞品额度均值



竞品利率

所有贷款借据日利率均值
所有贷款借据日利率加权均值

用户借贷需求



查询类

最近3/6/9/12/24个月信用卡审批查询次数
最近3/6/9/12/24个月查询机构数
最近3/6/9/12/24个月总查询次数



风险画像

还款能力

额度相关

所有贷款授信额度总和
商业银行结清贷款总额度



期限结构

结清贷款最近12个月总期数均值
未结清贷款剩余还款期数最小值



明细借据

结清贷款数量
按月归还贷款次数



还款意愿

逾期相关

贷款账户近3个月最高逾期期数
贷款最长逾期月份

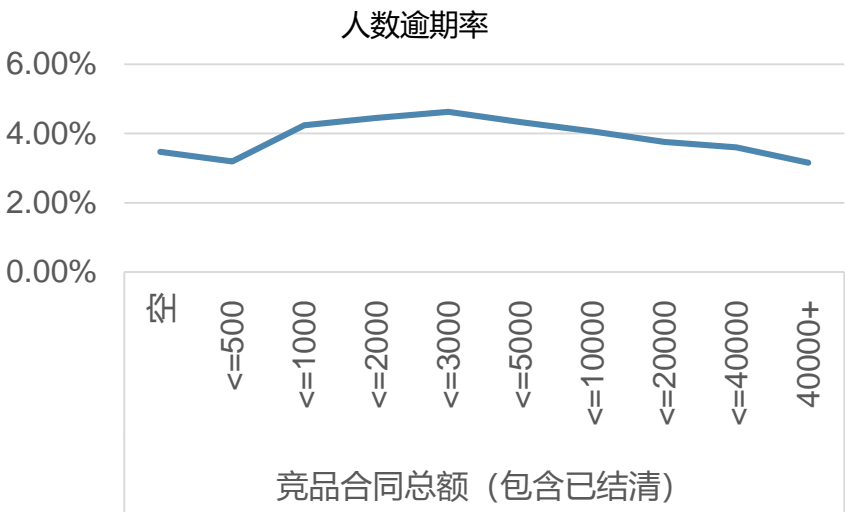


征信在风控中的应用

风险策略应用

- 挑选 **征信特征** 做策略：**
- 1、挑选区分度极好（高iv），单调性好，稳定性高，解释性好，用于做人群划分等规则
 - 2、挑选覆盖度低，但逾期率区分明显的特征适合做黑灰强规则拦截

字段举例



- 1、合同总金额超过3000后，数值越高，用户信用风险越低。
- 2、整体趋势是先上升，后下降。合同金额在1000以内的之前是白户的可能性较大，可以挖掘这部分用户的特征，作为运营触达白户用户的筛选。

风险模型应用

征信特征是否入模对贷前通过率的影响：
如果以A等级作为准入门槛，预估该客群人数逾期率为3%。则 **带征信数据的模型** 比不带征信数据的模型能多通过5%的用户。
同理可得B、C。

征信增益

贷前等级	有无征信数据	人数逾期率	ks	通过率
A等级	有	3%	27%	19%
	无		19%	14%
B等级	有	6%	22%	48%
	无		20%	37%
C等级	有	7%	11%	69%
	无		10%	59%

征信在用户运营中的应用 – 满意度预测

满意度预测

背景

探测在贷结清或即将结清用户对当前额度或利率是否满意，有针对性调整额度和利率。

数据需求

需要拿到用户最近在竞品的利率和额度以及借贷情况数据

目的期望

辅助模型和策略判断用户当前期望的额度和利率以及是否有借款需求。

征信竞品利率推断

报告提供入参数：
竞品期数
竞品剩余期数
竞品月还款额
竞品本金



出参：
还款方式、
竞品利率



step1 | 判断是否有任何贷款、信用卡



step2 | 判断剩余期数和期数的关系



step3 | 判断当月应还与本金的关系



step4 | 判断本金余额是否等于本金



step5 | 推断竞品利率



04

度小满对征信报告的解读方法

以上征信特征、征信模型如何构造？

征信解读技术发展趋势

征信报告中，存在丰富的结构化信息与非结构化的信息；
对这些数据的利用，经历了以下五个阶段



征信特征衍生：人工经验

 **1000+** 特征

用户还款能力

- 公积金月缴存额
- 学历、地理位置稳定性指标
- 房贷车贷总期数的均值
- 用户负债总额

用户还款意愿

- 贷款账户近3个月最高逾期期数
- 贷款逾期至今天数
- 连续逾期期数
- 贷款最长逾期月份数

用户满意度

- 等额本息利率
- 信用卡额度最大值
- 所有贷款授信额度总和

用户借贷需求

- 信用卡审批查询次数
- 查询机构数
- 最近两年贷后管理查询次数
- 最近借款频率

01 衍生逻辑不统一

不同建模人员设计的衍生变量逻辑不同，管理成本高

02 衍生变量不深入

个人征信变量衍生不成体系，可能遗漏重要信息

03 衍生效率低下

人工经验有限，难以快速衍生出大量特征，且难以获取反映新趋势的特征

征信特征衍生：CTF范式化特征生成框架

1、获取征信报文

2、JSON转化

3、数据校验

4、特征解析与清洗

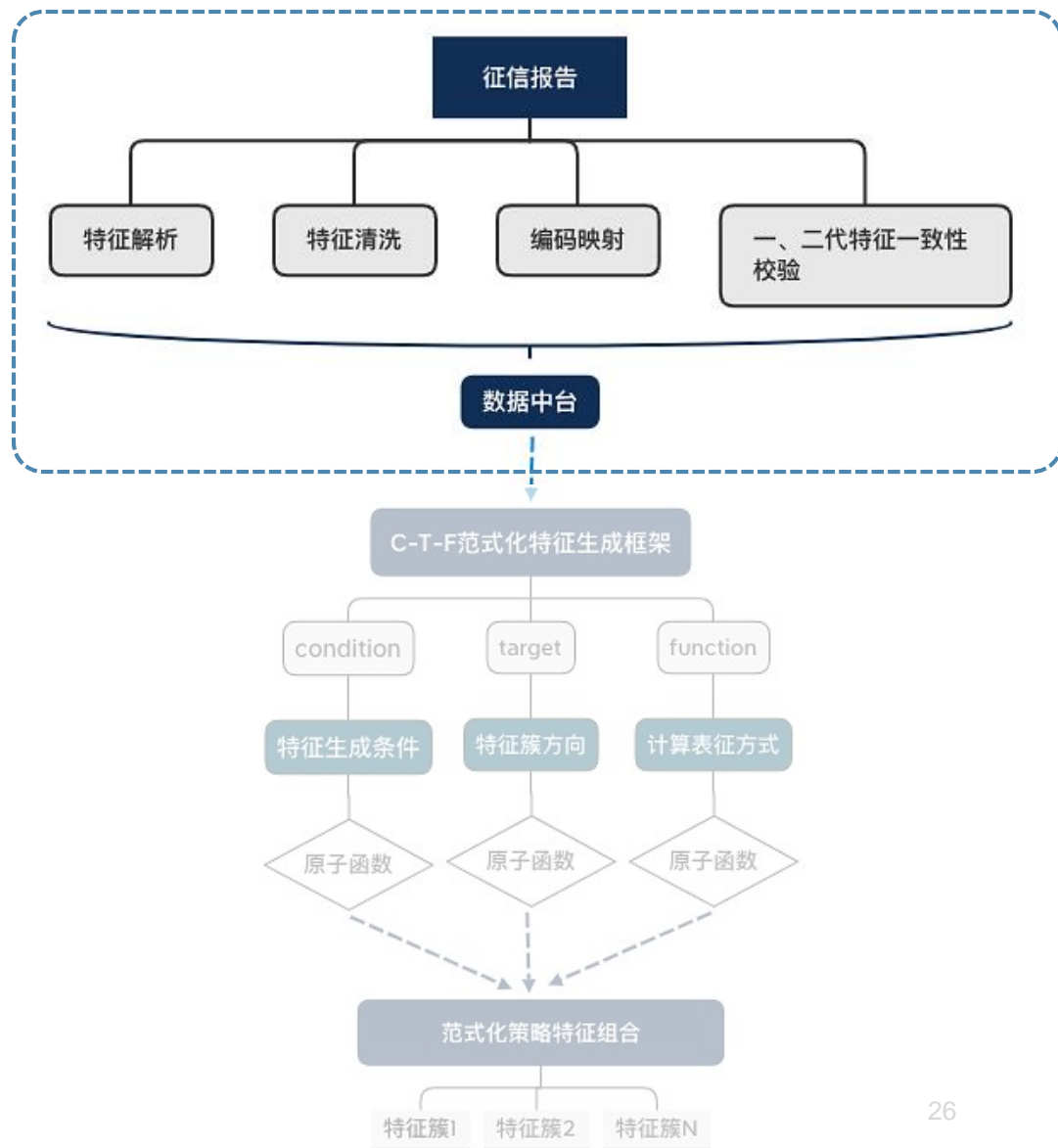
5、编码映射

将所有枚举类型的变量抽取出来，统计每个类别出现的次数，再按从大到小排序，给每组类型特征统一 ID 化编码。

6、一二代特征一致性校验

将一、二代征信报告中名称不一致的变量名统一映射成唯一的变量编码。

筛选一致率较高的中台特征做为后续的特征簇生成源。



征信特征衍生：CTF范式化特征生成框架



Condition

贷款状态、贷款机构类型
贷款类型、时间周期



Target

本月应还金额、本月实还金额
贷款金额、贷款期数

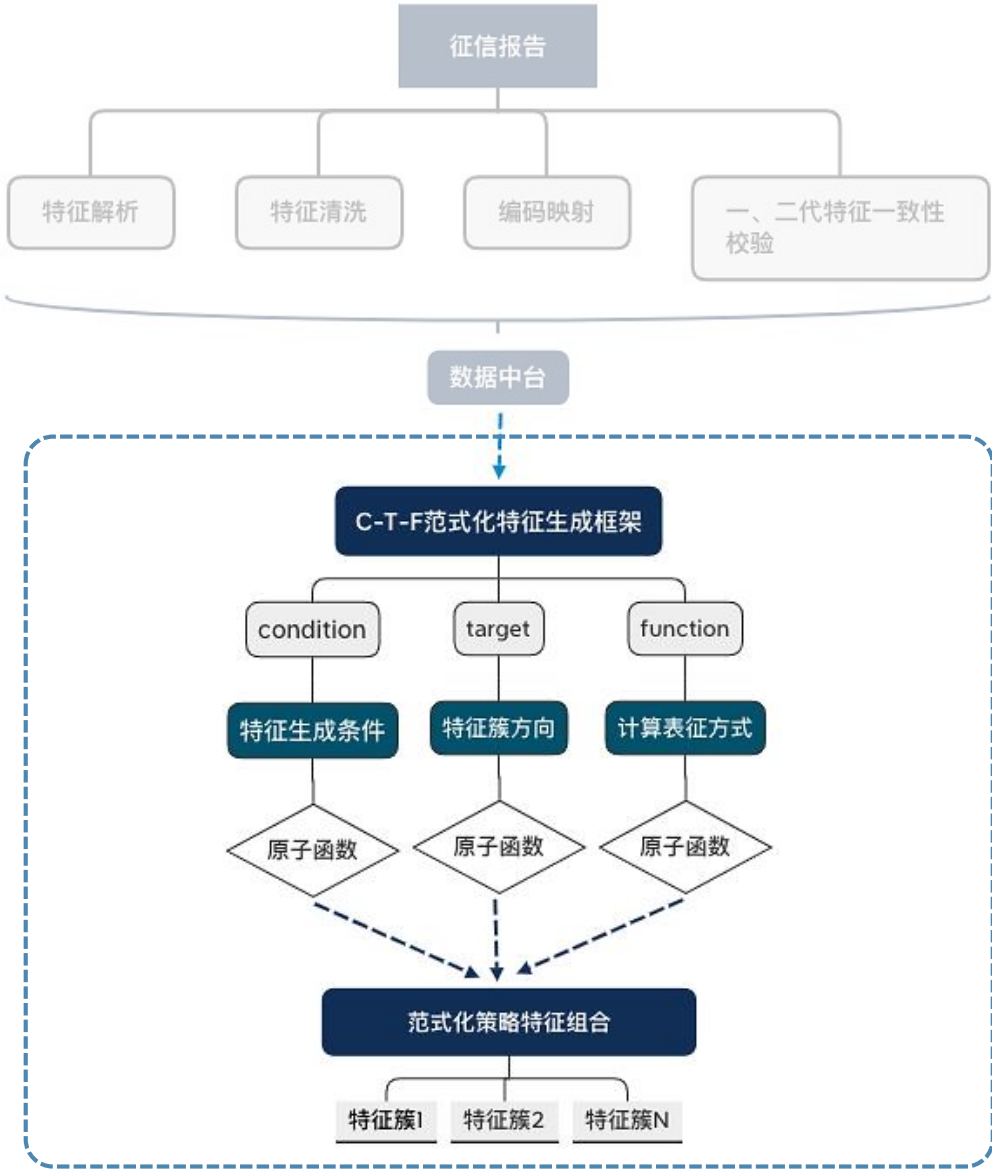


Function

Sum\max\min\average

举例:

模式: condition (贷款状态+还款类型) * target (还款金额) * fun (均值)



征信特征衍生：CTF范式化特征生成框架

四大特征维度

30w+ 特征

↑5% 区分度指标

基础类特征

- 是否性别男_年龄在I~J之间_未婚_大专学历

贷记卡类特征

- 开立距今N月_账户状态正常_授信金额_最大值

贷款类特征

- 开立距今N月_账户状态正常_商业银行_消费贷_合同金额_平均值

查询类特征

- 最近N个月_查询次数_和
- 最近N个月&查询原因为贷款审批&查询机构为消费金融公司_查询次数_和

01 特征生成自动化

通过程序化的原子函数，自动化高效组合生成特征簇，有效提高征信特征簇生成效率

02 特征簇覆盖率高

依赖业务经验生成的特征簇可能会遗漏部分重要特征，而组合化生成的特征簇能够覆盖、穷尽所有原子函数的排列组合

03 特征之间易比较

特征簇的底层数据都是同一批原子函数，在数据定义、数据口径上不存在歧义

04 特征未来易拓展

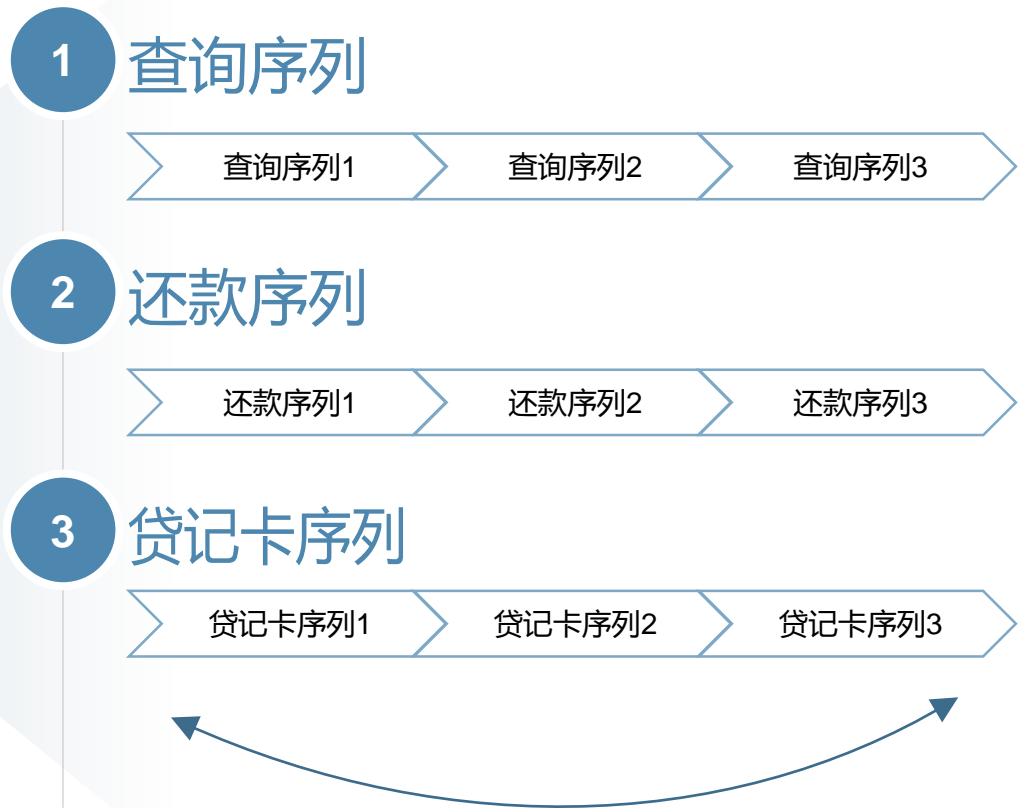
通过增加原子函数、改变原子函数组合拓展现有特征

征信特征衍生：趋势类特征

**Condition**
贷款状态、贷款机构类型
贷款类型、时间周期

**Target**
本月应还金额、本月实还金额
贷款金额、贷款期数

**Function**
Sum\max\min\average



举例:

近N年银行发放的消费贷_近M月还款金额相比近N年还款金额_比值
近N年银行发放的消费贷_刚发放头M各月的还款金额相比近N年_比值

计算序列中由近至远（或由远至近）指定数量月份中趋势的数值变化与比值。

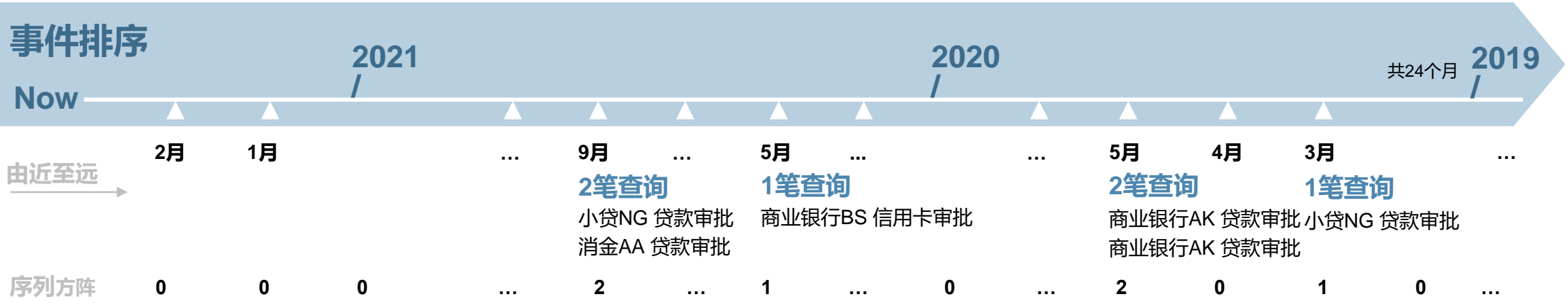
概述
内容简介
业务应用
解读方法

征信特征衍生：趋势类特征-以查询为例

查询记录明细		
查询日期	查询机构	查询原因
2019/3/25	小贷NG	贷款审批
2019/5/9	商业银行AK	贷款审批
2019/5/24	商业银行AK	贷款审批
2020/5/2	商业银行BS	信用卡审批
2020/9/9	小贷NG	贷款审批
2020/9/22	消金AA	贷款审批

查询序列

- 我们将序列设计成 24 个月的时间窗口
- 由近至远按月份对齐每个查询事件，形成序列方阵
- 对对应维度的值进行加和计算
- 得到最终的 24 维度序列值



举例

SUM

近2年查询总次数：3次
近2年商业银行查询总次数：1次

Std. Dev.

近2年查询次数标准差：0.51
近2年商业银行查询次数标准差：0.37

Ratio

自有贷款记录以来头3个月查询次数占比：0.5
近3个月查询次数占比：0

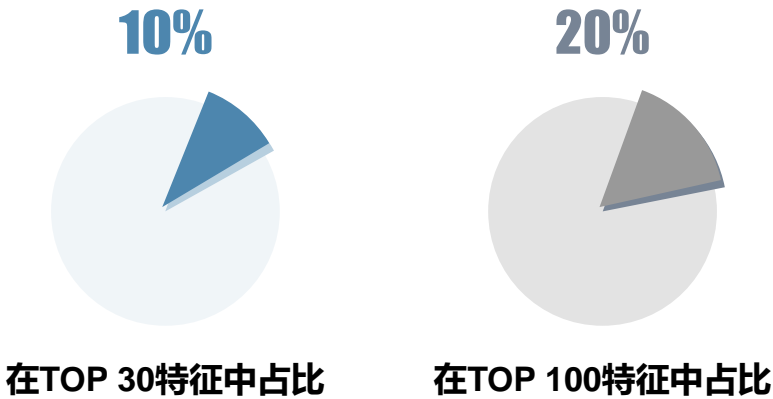
自有贷款记录以来商业银行查询的事件中头3个月占比：2/3

征信特征衍生：趋势类特征

60w+ 衍生特征数量总计60万维

↑9% 区分度指标提升9.1%

趋势类特征重要性



Top趋势类特征

近X个月“信用卡审批”查询次数的标准差

在近X个月中，近N个月“担保资格审批”查询次数占比

在近X个月中，近N个月“贷款审批”查询次数占比



多头查询类

贷款机构类型趋势

贷款机构数趋势

查询次数趋势

查询类型趋势



贷款趋势

额度趋势

总期数趋势

机构类型趋势

机构数目趋势

贷款类型趋势

还款状态趋势



贷记卡趋势

额度趋势

次数趋势

机构数目趋势

还款状态趋势

深度学习 挖掘时序特征

深度学习时代对征信报告的解读

通过时序模型挖掘征信报告中的统计特征、文本信息、时序信息，实现：

- 记忆历史细节，捕获全生命周期信息变化趋势
- 挖掘类别信息及多维数值信息的组合，识别隐藏信息
- 挖掘细粒度的趋势类信息



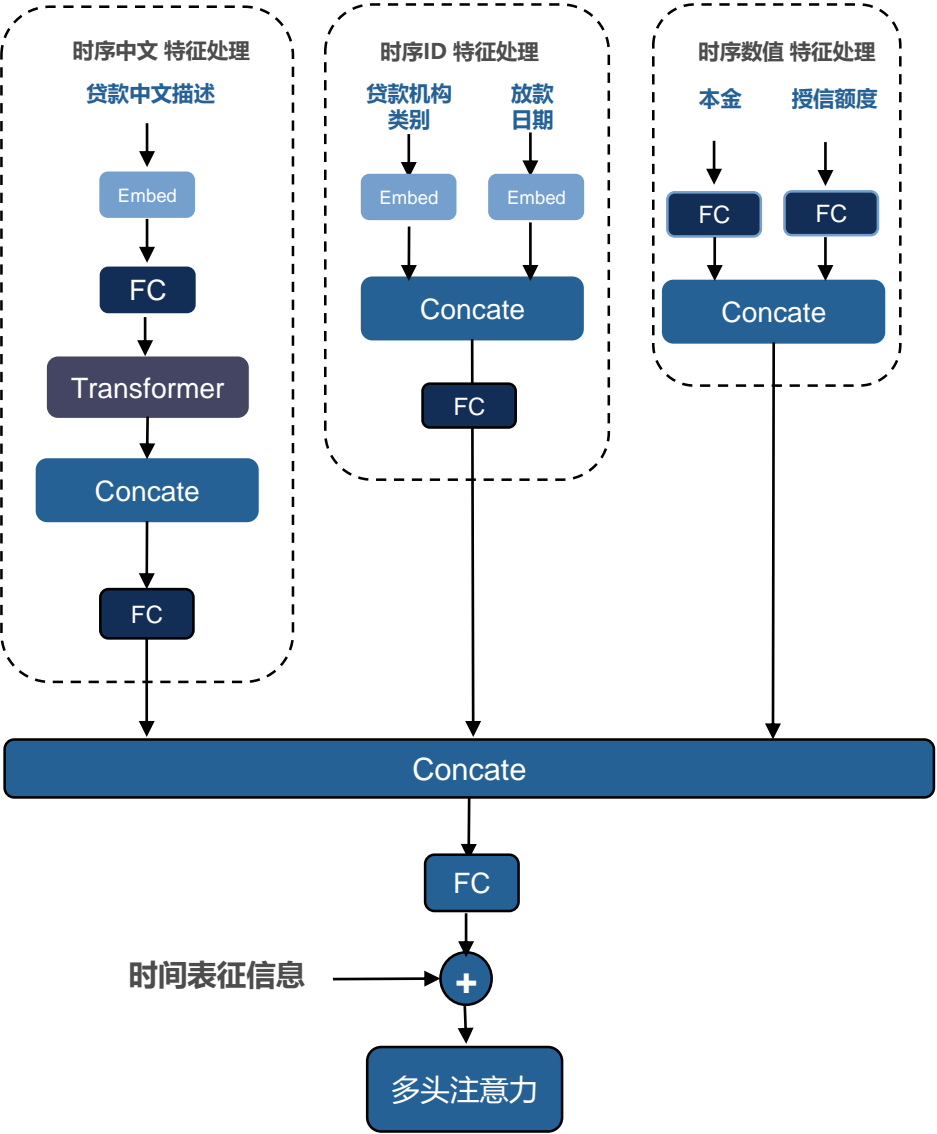
统计特征



文本信息



时序信息



图机器学习 挖掘征信关联信息

图机器学习时代对征信报告的利用

深度学习中的时序模型对征信报告的挖掘已经达到瓶颈，而征信报告中大量的关联信息还有待利用；图机器学习能够很好地识别、表达、利用关联信息

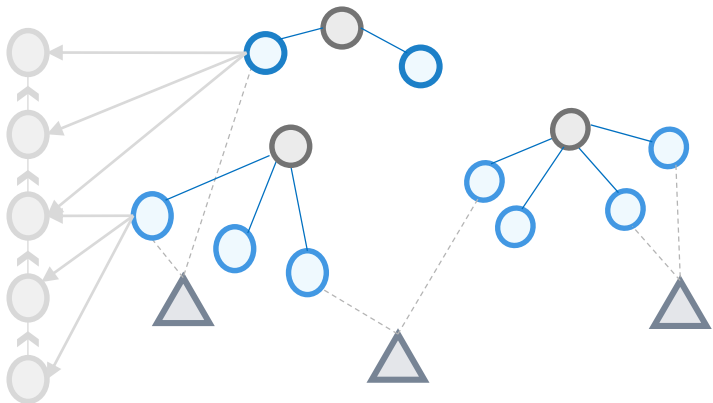


关联信息

关联同一公司
关联同一贷款机构

征信报告中的关联信息

1 一份征信报告内部关联



△ 机构聚合

用户在同机构发生了多项的借贷/查询信息

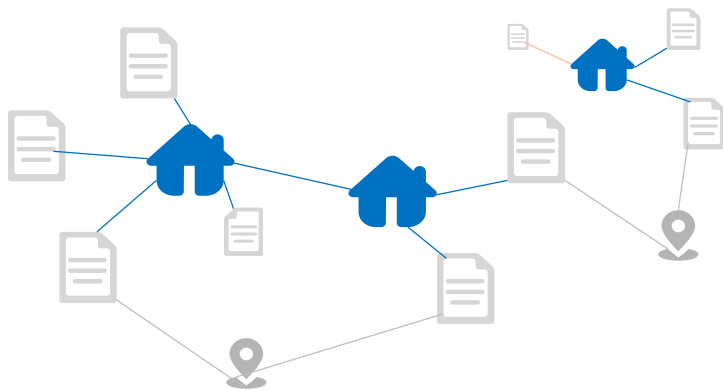
○ 借款/查询聚合

相同借款/查询/还款等行为的聚合

○ 时间聚合

时间序列行为中隐藏着依赖关系

2 不同征信报告之间关联



地理聚合

工作地址/居住地址关联关系



公司聚合

工作单位有关联

Thank you
for watching.

