



JIGUANG | 极光
数 据 改 变 世 界

极光在金融反欺诈领域的算法探索

黄鑫

极光大数据CTO

17/7/3

目录

Part

01

传统反欺诈算法概述

Part

02

极光反欺诈模型探索

Part

03

极光反欺诈的未来工作

传统反欺诈算法概述

互联网金融的核心：基于互联网平台，利用大数据技术让资金更快、更有效地流转

互联网金融的难点：欺诈风险是线上信贷平台的最大挑战。

风控的关键



传统线下平台审核方式：

1. 基本资料
2. 工作证明
3. 收入证明
4. 财产证明

传统线下平台优势：

可以针对每项信息进行人工评估，准确率高。

传统线下平台劣势：

审核效率低下，人工成本高，无法规模化。

传统反欺诈算法概述

平台审核：

身份核验：通过调取三方API，实现基本的三维核验

行为核验：通过短信验证码等形式验证基础身份

用户授信：各公司通过信用评分算法计算用户风险分数

1. 三方数据：多头借贷、违约记录、银联记录、运营商数据等
2. 行为数据：用户在APP内的行为、用户当前信息
3. 填写信息：用户提供的电商信息等

为了增进用户体验，以下几步需要在分钟级完成



征信模型：

大部分公司依然在使用信用评分规则引擎

规则引擎优劣势：

简单易懂，容易利用领域知识调整

错误率较高，很难利用组合维度优势

目录

Part

01

传统反欺诈算法概述

Part

02

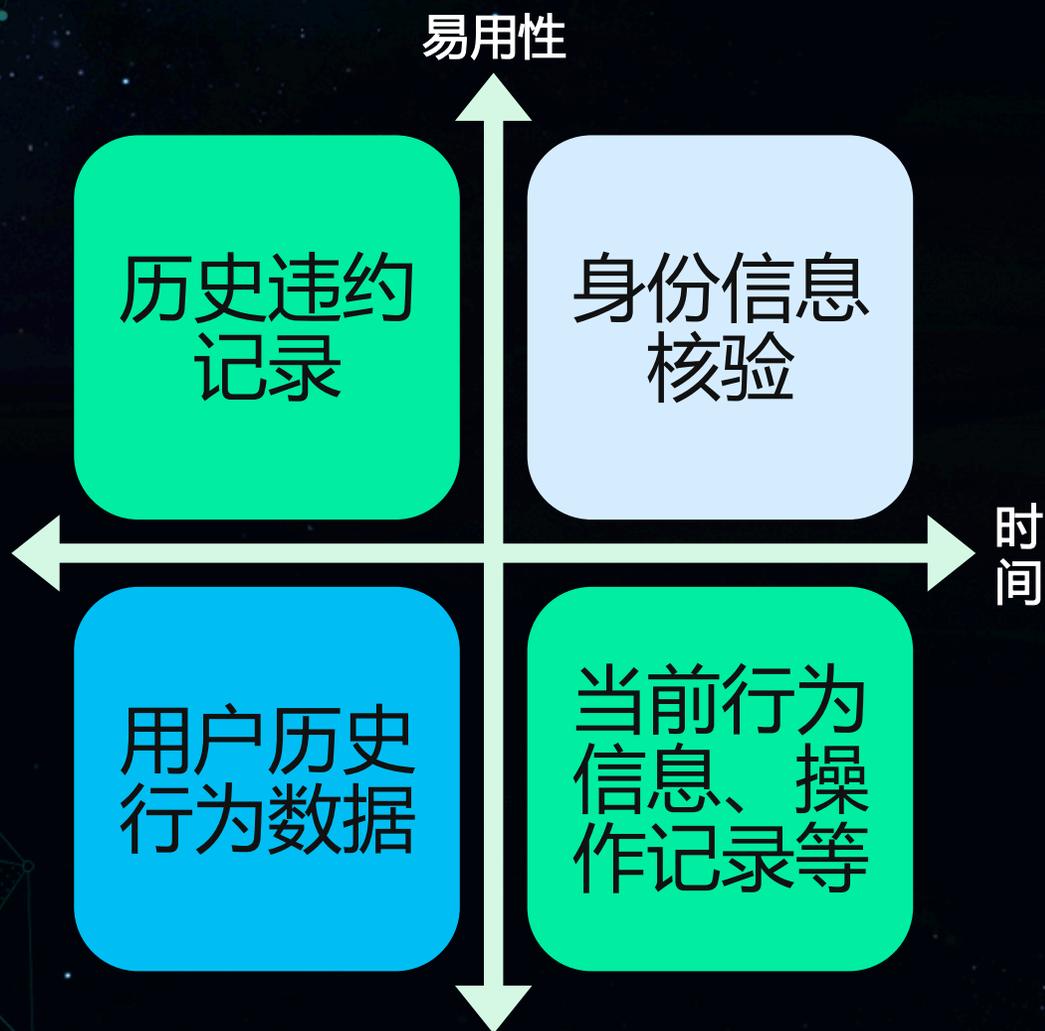
极光反欺诈模型探索

Part

03

极光反欺诈的未来工作

极光反欺诈模型探索



注：

X轴原点右侧是APP本身可以获取的信息(第一和第四象限)

X轴原点左侧是APP需要通过第三方数据补全的信息(第二和第三象限)

Y轴原点上方是基本无需挖掘的可用信息(第一和第二象限)

Y轴原点下方是需要进行挖掘才可使用的信息(第三和第四象限)

极光为什么要做反欺诈：

极光的覆盖量可以保证在第二和第三象限的数据优势

极光的技术可以保证在第三和第四象限的模型优势

关于历史数据，我们可以做什么？

极光的自有数据

应用列表的历史安装卸载状态

0/1变量，可以直接作为特征使用

历史的地理位置轨迹

需要做特征映射，无法直接使用

海量应用的使用行为

部分是连续变量，需要进行一定处理

海量的推送消息及后续行为

需要对消息本身进行文本处理挖掘

极光反欺诈模型探索 — 特征工程

常用的特征选择方法：单变量选择法，GBDT，CNN

单变量选择法：简单高效，但是无法有效特征组合，强依赖领域知识

GBDT：容易解释，依赖树形关系，对于特征组合粒度不够

CNN：当下最流行的特征工程方法，通过深度网络进行特征“深度”组合，但是调参复杂度高，最麻烦的是可解释性差

如何集齐所长？

简单、高效、易添加特征

可解释性、可以深层组合

客户需要什么？

1	序号	业务域	特征名称	特征描述说明
3	2	客户信息因素	客户触达稳定指数	反映客户的联系方式、使用设备等变更频次的稳定程度
4	3	客户信息因素	设备价值水平指数	通过客户所持移动设备的品牌、型号、参数及出厂价格等因素综合体现设备的价值水平
7	6	风险业务因素	信用消费频度指数	反映客户对于不同信用消费平台的关注深度
8	7	风险业务因素	信用消费规模指数1	反映客户对于不同信用消费平台的关注广度
9	8	风险业务因素	信用消费规模指数2	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，洞察客户对消费金融平台的兴趣度
10	9	风险业务因素	信用消费规模指数3	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，洞察客户对综合金融理财的兴趣度
11	10	风险业务因素	信用消费规模指数4	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，洞察客户对金融中介平台的兴趣度
12	11	风险业务因素	信用消费规模指数5	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，洞察客户对传统信用消费平台的兴趣度
13	12	互金业务因素	投资理财偏好指数1	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，评价客户对移动信用理财平台的兴趣度
14	13	互金业务因素	投资理财偏好指数2	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，评价客户对金融增值服务平台的兴趣度
15	14	互金业务因素	投资理财偏好指数3	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，评价客户对互联网移动支付的兴趣度
16	15	互金业务因素	投资理财偏好指数4	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，评价客户对传统银行理财平台的关注程度
17	16	互金业务因素	投资理财偏好指数5	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，评价客户对银行综合服务的关注程度
18	17	互金业务因素	互金业务活跃度指数1	通过使用频度、使用时长、使用规模等不同维度数据交叉分析，评价客户对互联网信贷服务的关注程度
19	18	互金业务因素	互金业务关注指数2	反映客户对于各类大额借贷类产品的持续关注程度和稳定使用程度
20	19	互金业务因素	互金业务关注指数3	反映客户对于各类短期借贷类产品的持续关注程度和稳定使用程度
21	20	互金业务因素	互金业务关注指数4	反映客户对于各类小额借贷类产品的持续关注程度和稳定使用程度

极光反欺诈模型探索 — 关于样本补充

互金行业样本的稀缺性：

样本收集是长期行为，而且样本的数量极少

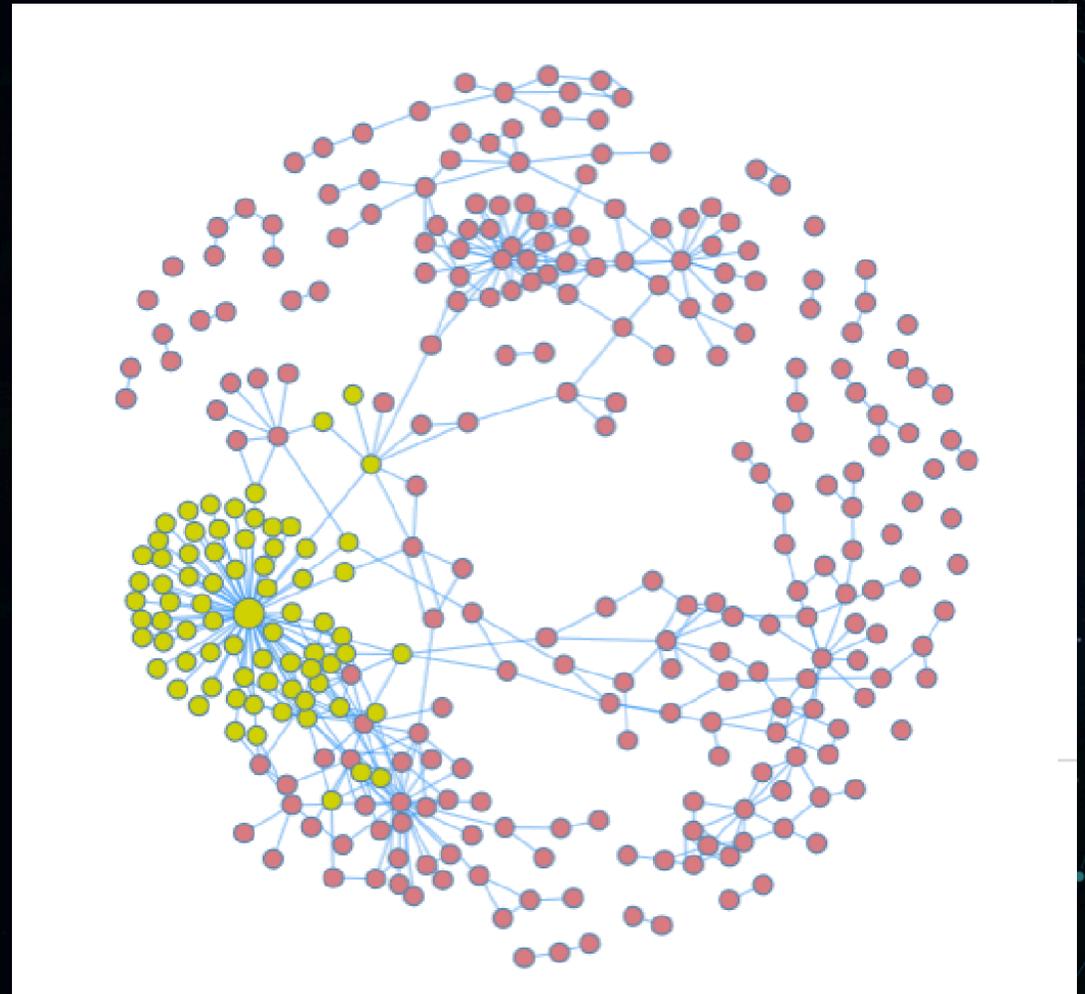
样本噪音较多，不同公司对于违约的定义不同

样本如何补充：

社交关系分析是效果最好的补充方式

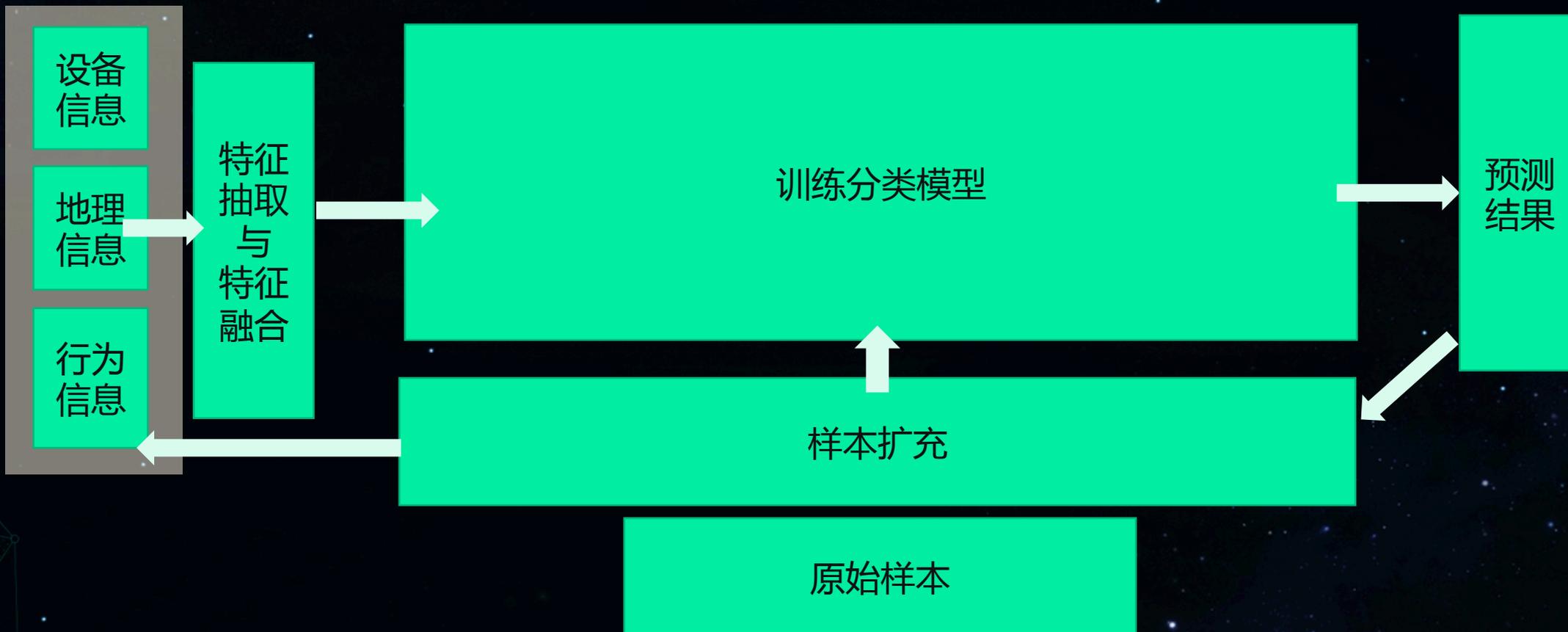
数据增强一定程度上也可以补充数据稀缺的问题

无法追溯历史数据？其实一样可以解决



极光反欺诈模型探索

将特征选择、分类模型、样本补充作为一个完整流程循环迭代，互相补充



目录

Part

01

传统反欺诈算法概述

Part

02

极光反欺诈模型探索

Part

03

极光反欺诈的未来工作

极光反欺诈的未来工作

问题一

不同用户在不同场景下的风险程度和风险损伤不同

解决方案

将问题转换为经典的广告问题，不同的广告产生的收益不同

极光反欺诈的未来工作

问题二

用户的样本难以获取，众多特征都是弱关联关系

解决方案

尝试迁移学习在反欺诈问题上的应用

极光反欺诈的未来工作

问题三

用户第一次不违约并不代表未来不会违约

解决方案

尝试增强学习和RNN在反欺诈问题上的应用

极光反欺诈的未来工作

问题四

是否有更好的特征工程方案

解决方案

需要逐步探索中

END

T H A N K S



数 据 改 变 世 界

黄鑫 极光大数据CTO

Call: 176-0761-0625

Email: huangxin@jiguang.cn