



大数据提升互联网金融风险管控能力的应用实践

传统银行业如何借力移动互联网寻求突围，如何在大数据时代持续保持稳健发展的竞争力，降低“互联网+”时代带来的各种金融风险，将是互联网金融近几年来需要不断探索和实践的课题。



江苏银行股份有限公司信息科技部总经理 葛仁余

著名咨询公司波士顿认为，当量级庞大、实时传输、格式多样的全量数据通过某种手段得到利用并创造出商业价值，且能够进一步推动商业变革时，大数据就诞生了。过去，人们在处理数据时受到数据量和数据处理手段的制约，为了确保分析结果的准确性，人们偏向于收集可量化的、准确的数据。

而在大数据时代，人们可以获得海量的、非结构化数据，并且利用内存分析、流处理等新兴技术，大幅度提升了对海量数据的处理能力，人们得以从新的视角重新审视数据的利用和挖掘。大数据的出现，降低了数据分析的成本门槛，实现了从数据到价值的高效转化。

要加快政府数据开放共享，推动资源整合，依托政府数据统一共享平台，大力推进法人单位信息资源库等国家基础数据资源，以及金税、金盾等信息系统跨部门、跨区域共享。

如上海市已开放交通大数据，南京市打造了以社保、住房公积金、车辆违章等为基础的大数据平台并在一定范围内向社会机构开放。据国家发展和改革委员会透露，政府数据统一开放门户有望2018年以前建成，实现面向社会的政府数据资源一站式开放服务。

除了政府公共数据资源逐步开放外，《纲要》中还特别指出要发展新兴产业大数据，大力培育互联网金融等新业态。金融行业在发展大数据能力方面具有天然优势：在开展业务的过程中积累了海量的高价值数据，其中包括客户信息、交易流水等数据。有数据显示，中国大数据应用投资规模以五大行业为最高，其中第一是互联网行业，占28.9%，第二是电信行业，占19.9%，第三是金融行业，占17.5%。而金融行业中银行业又是重点，占41.1%。

面对着数据宝藏，每一家银行都需要回答这样的问题：如何充分利用外部开放数据和银行自有数据，让数据资产迸发出能量。

江苏银行在选择一个合适的大数据技术平台之后，积极采用“应用驱动”、“业务与科技紧密协同”、“迭代式开发和小版本发布”的项目管理方法，结合各个业务条线的业务拓展思路、客户管理和产品设计策略，以开放的思维引入先进的专家经验和汇聚本行创意，通过打造金融大数据应用平台，实现大数据弯道超车的目标，促进业务创新和管理创新。

一、互联网金融面临的风险和应对措施


移动互联网使商业服务和金融服务得以无形地嵌入到人们生活方方面面，为互联网金融的创新和广泛应用提供了基础。

我国互联网金融的发展现状从推进主体来看，互联网金融业态可以分为3种类型：

一是由金融机构进行的互联网交易，即传统金融交易方式的网络化和电子化，如手机银行等业务；

二是由互联网企业推出的金融业务，如第三方支付、P2P和众筹等业务；

三是由电子商务企业与金融机构或不同种类金融机构之间进行合作而产生的新兴业态，如余额宝(阿里与天弘基金的合作)等理财产品。



互联网金融常见的风险主要包括5种：

一是信用风险，由于交易双方基于虚拟的网络进行交易，容易被交易对手利用技术和法律漏洞谋取不当利益，信用风险较大；

二是信息科技风险，包括信息安全风险、技术选择风险和技术支持风险；

三是操作风险，由于交易主体对互联网金融业务的操作要求不太了解，或信息系统设计缺陷引起的操作风险；

四是声誉风险，互联网金融机构不能与客户建立良好的关系，从而导致其无法有序开展金融业务的风险；

五是法律风险，现有的法律法规都是为传统金融业务设置，不完全适合于互联网金融，如交易者身份认证、资金监管、市场准入等尚未有明确的监管规定，故在互联网金融的交易过程中容易出现由于交易主体间权利义务模糊而导致法律风险。

本文主要从信用风险角度展开，并结合江苏银行的实践对互联网金融的大数据风险管控应用实践进行分享。

和传统金融业相比，互联网金融的信用风险的防控具有以下两个关键点：

一是风险管理面临“免担保”模式的冲击，银行必须从更广阔的范围进行信用风险的防范，需要解决信息不对称的问题，并建立更加全面的客户信用风险量化体系。互联网的普及降低了信息成本，很大程度上解决了信息分散和不对称问题。互联网金融借鉴传统商业银行的信誉评估标准，整合电商、第三方支付等多平台数据，利用云计算、大数据等技术，深入挖掘企业信息，减少人力成本，提高风险预判。同时对企业财务经营状况、上下游关系等信息实时监测，方便违约后不良资产的及时处理变现，打造和谐信贷生态圈(链)。

二是在虚拟的网络进行交易，欺诈风险高，对客户信息的甄别更为重要，需要更先进的技术手段。除了传统数据反欺诈模型分析外，互联网时代必须引入移动数据的支持，如利用移动设备的位置信息，帮助商业银行验证贷款申请人居住地，分析贷款用户真实工作地点，识别出现在同一个经纬度的群体性恶意欺诈事件，降低恶意欺诈的风险。

二、大数据在风险管控领域的技术研究

大数据从内涵上看，其特征可归纳为三个方面：


一是数据类型方面，除了包括海量的结构化和半结构化的交易数据，还包括海量非结构化数据和交互数据；

二是技术方法方面，核心是从各种类型的数据中快速获取有价值信息的技术及其集成；

三是分析应用方面，重点是采用大数据技术对特定的数据集合进行分析，及时获得有价值的信息。

与其他行业相比，大数据对银行业更具潜在价值。一方面，大数据决策模式更符合银行发展需求。银行业发展模式转型、金融创新和管理升级等都需要充分利用大数据技术、践行大数据思维。另一方面，银行业数据特点是量大、类型多，不仅拥有账户信息和交易信息等结构化数据，还拥有客服音频、网点视频、网上银行记录、电子商城记录等非结构化数据。

在新的形势下，银行业需要加强大数据应用，如利用大数据图分析与流处理技术，快速统计历史数据、一段时间窗口的信息流和触发计算的事件，并匹配模型，在百毫秒级别内进行响应；处理非结构化数据，整合网页、文本、JSON、XML、图像和语音等非结构化数据，转化成结构化字段；通过引入和整合人民银行征信、税务、工商、公安、法院、电信



服务商、P2P平台等网络数据源，实现客户的云数据360度画像标签;探索基于并行数据处理技术环境下R语言的运用，实现客户担保圈关系的自动挖掘，自动标识预警担保圈的形成。

基于大数据的风险管控分为五个步骤。

1.全面风险视图的建立

通过建立数据交互渠道，获得税务、司法、环保、工商等在线信息，通过爬虫等技术手段获得舆情信息，并利用半结构化和非结构化数据加工分析技术，将上述数据转化成结构化数据，加工整合形成全面的客户征信视图。在此基础上，不断进行迭代设计，完善业务需求。

2.客户线上信息识别

通过人脸识别、反欺诈侦测技术核实客户身份的真实性，判断申请者是否存在欺诈行为。一般来说，人脸识别系统包括图像摄取、人脸定位、图像预处理以及人脸识别(身份确认或者身份查找)。系统输入的一般为一张或多张含有未确定身份的人脸图像、人脸数据库中的若干已知身份的人脸图像或者相应的编码，输出的则是一系列相似度得分，表明待识别的人脸的身份。

在线反欺诈是互联网金融必不可少的一部分，常见的反欺诈系统由用户行为风险识别引擎、征信系统、黑名单系统等组成，包含五个部分：数据采集、欺诈侦测、系统管理、报表、数据库。为了进一步提升反欺诈能力，设备指纹技术、代理检测技术、生物探针技术被应用到反欺诈系统中，实现从多维度降低风险。

3.信用评分模型建设以及与之匹配的业务策略设计

智能模型是一种欺诈风险量化的模型，它利用可观察到的交易特征变量，计算出一个分值来衡量该笔交易的欺诈风险，并进一步将欺诈风险分为不同等级。智能模型会在客户交易的第一个行为开始进行分析，为客户每一个动作赋予相对应的风险分数，为智能型反交易欺诈授权决策提供科学依据，对欺诈风险高的交易可以拒绝授权和展开调查。

银行业欺诈风险主要集中在注册、登录、借款、提现、支付、修改信息6个业务场景。如注册场景中的虚假注册、垃圾注册;登录场景中的撞库登录、暴力破解等;借款场景中的多头借贷、信用恶化;提现和支付场景中的欺骗行为等。

4.实时风控技术框架

针对个人线上消费贷款的风控需求，反欺诈系统需具备稳定、快速、准确的特点，以平衡业务拓展、客户体验和风险控制三方的关系。通过引入反欺诈风险规则引擎，可以将不断变化的业务规则剥离出来，进行动态管理和多规则多重组合，从而使系统变得更加灵活，适用范围更加广泛。在交易过程中，通过实时计算当前交易和历史交易特征的偏离值，如平均交易金额、常用的交易类型等，计算该笔交易发生欺诈的概率。

5.智能决策与业务应用流程结合

基于行内和行外数据，将开发和设计出的智能模型及业务策略统一部署在企业级决策平台之后，需要将这些智能决策服务嵌入到现有的作业流程中，从而改进传统的作业模式，实现客户服务模式、体验模式及管理模式的创新。

在整个技术实现框架中，数据是基础，智能模型和业务策略是核心和灵魂，与各种业务流程及渠道结合是应用成果的外在展现。

三、江苏银行在互联网金融业务中应用大数据开展风险管控的探索与实践

从2014年起，江苏银行对互联网金融进行了多方面探索和实践。

为解决产品设计中的风险管控问题，江苏银行基于Hadoop开放式的大数据分布式处理技术平台，整合了内外部海量

数据，开发了风险数据集市、资产负债管理集市、监管报送集市等多个内部数据集市，打造了面向全行的开放共享大数据平台。

针对外部数据，通过统一的外部数据平台采集、转换和存取包括人民银行、银监会、工商、税务、法院、环保、海关、电信等20余个数据源，多达几千项字段的外部数据，并运用网络爬虫技术和命名实体识别技术，抓取公共网络媒体舆情信息；

内部数据包括行内核心系统、信贷管理系统、网银系统、个贷系统等几十个业务系统中的交易数据、账户数据和客户基础数据，以及通过线上行为分析系统收集的设备信息和客户行为数据。在丰富的数据基础上，通过流计算引擎和风险规则库，实时计量互联网金融中的信用风险。

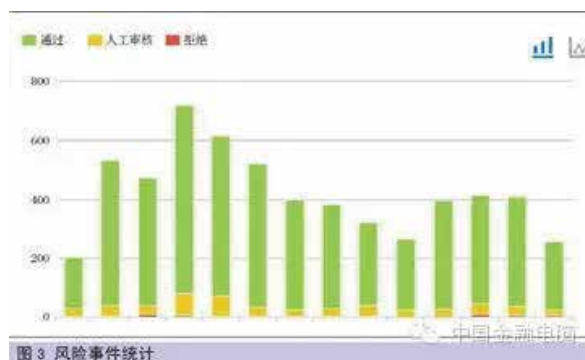
基于大数据平台，江苏银行推出了独具特色的两款互联网金融产品——联网线上个人消费贷款产品“享e融”和小微金融贷款产品“税e融”。这两款产品均具有“纯线上”、“高效率”的特点。基于大数据技术的决策模型实现系统自动审批，替代传统人工审批，业务办理效率大大提高。从客户发起申请提交到后台数据搜集、分析、审批，仅需要5~7秒左右。


1. “享e融”个人纯线上贷款产品

通过从网贷平台调用反欺诈风控系统接口，对申请网贷业务的客户进行反欺诈甄别，并拒绝触犯反欺诈规则的申请客户。反欺诈引擎的交互如图2所示。



该产品上线两周即收到近6000笔申请，反欺诈风控系统能快速帮助江苏银行识别出风险事件(调用结果反馈约200ms/次)。其中高风险贷款事件占比约1%，需人工审核确认的贷款事件占比约9%，风险事件统计如图3所示。





在发现的贷款风险事件中，失信风险事件约占8.9%，异常借款事件约占91.1%。决策引擎发现了各类风险情况，如一天内同一设备或账户借款次数过多(一天37次)、不在手机归属地借款事件、3个月内同一身份证在多个平台进行多头借贷(同一身份证向17个不同平台申请借款)，有效地帮助风控部门识别出潜在欺诈风险，提高决策效率和准确性。

大数据风控系统除了具备贷前阶段的风险事件识别功能，还具备贷后监控功能，针对已放款的借款用户进行监控，通过批量计算客户风险分值的方式(当借贷人在贷款期间发生逾期记录、重复借贷或者卷入经济法律纠纷等时，其风险评分将恶化)，及时预警客户风险，通知客户经理或后台管理部门，采取有效防范风险措施。

2. “税e融”小微金融贷款

在“税e融”产品中，江苏银行内部系统首先获取国税数据中客户的缴税记录、纳税额、纳税评级等各类数据，并将其解析成客户模型分析所需的参数，判断授信目标为诚信纳税且纳税额较高的客户；

然后数据采集系统通过爬虫等技术围绕该客户进行一系列的外部资料搜索和获取，如工商信息中的集团关系、担保信息、关联人信息，法院、公安信息、征信、银监信息、舆论信息等；

最后通过数据整合及完整模型分析，生成客户肖像概况，并计算出违约概率，进而自动对客户进行风险等级评定，确定可授信金额。

对客户而言，只需要提供营业执照号、组织机构代码号、纳税识别号等信息，即可实现1分钟内贷款审批完成、入账。“税e融”为小微企业打开了一条便捷融资的“绿色通道”。自2015年6月正式推广，至10月底，已收到贷款申请40301笔，为4239户小微企业发放39.8亿元信用贷款，赢得了广大小微企业客户的一致赞誉。

此外，江苏银行大数据风险防控系统已走出江苏，通过为甘肃银行提供甘肃地区“税e融”云服务，实现了大数据技术和业务模式的输出。

互联网金融已经大举进军金融业，并不断冲击着传统金融业务。在以网络化和数据化为特征的新经济时代，金融与大数据交叉融合。

大数据有助于提升金融市场的透明度，通过从海量的数据中快速获取有价值的信息以支持商业决策，进一步推动金融业发展；大数据促进互联网金融企业实现精准营销、规避风险、优化经营绩效、提高运营效率，增强企业融资的便捷性和经济性；同时，利用大数据技术可以逐步解决金融格局演变、信用评估、风险管控、信息安全等的一系列难题。

除了协助传统商业银行建立风险管控体系外，大数据还可以在征信共享服务、寻找新业务和客户价值、实现多渠道协同精准营销、精细化资本管理、优化产品、改善客户体验及提高决策科学性等方面开展更多的应用。

传统银行业如何借力移动互联网寻求突围，如何在大数据时代持续保持稳健发展的竞争力，降低“互联网+”时代带来的各种金融风险，将是互联网金融近几年来需要不断探索和实践的课题。

另外在数据共享的同时，也需要保护个人隐私，唯有如此，数据才能可持续地释放正能量。

星环信息科技（上海）有限公司

🏠 地址：上海市徐汇区桂平路481号18幢3层301室（漕河泾新兴技术开发区）

✉ 邮编：200233 ☎ 电话：4008 079 976

🌐 网址：www.transwarp.io

TRANSWARP
DATA HUB