

评语：北京宜信致诚信用管理有限公司阿福平台通过接入各种金融数据并采用先进的数据处理技术，提升了信用判断的准确性，有效防范了多头负债等过去难以规避的风险。

金融科技在互联网金融行业性 风险防范领域的应用

胡鹏飞

北京宜信致诚信用管理有限公司, 北京 100022

摘要

互联网金融领域出现的越来越多的欺诈风险、多头负债等现象，已经严重影响了行业的健康发展。详细分析了互联网金融行业面临的挑战，采用分布式架构设计及API调用技术，设计了一个行业机构间的共享生态系统，并以知识图谱为底层架构，设计了更适用于金融场景的社交网络图谱，构建了反欺诈的风控体系，实现了互联网金融行业数据的互联与互通，有效防范了行业多头负债的发生，提升了行业反欺诈能力。

关键词

反欺诈；共享平台；知识图谱

中图分类号：TP393

文献标识码：A

doi: 10.11959/j.issn.2096-0271.2018012

Application of FinTech in internet financial industry risk prevention

HU Pengfei

Zhicheng Credit Service Co., Ltd., Beijing 100022, China

Abstract

In view of the increasing fraud risk and long-term liabilities appearing in the field of internet finance, the healthy development of the industry has been severely affected. The challenges facing the internet financial industry were analyzed in detail. Based on distributed architecture design and API calling technology, a shared ecosystem between industry organizations was designed. A social network knowledge graph for a financial scene was proposed. The anti-fraud risk control system was designed. The industry's anti fraud ability was promoted.

Key words

anti-fraud, share-platform, knowledge graph

1 引言

2006—2016年,中国的普惠金融已经发展成为中国金融体系的重要组成部分,随着行业监管细则的不断落地,行业发展愈加有章可循。但与此同时,借款人的信用风险(借款人多头负债和欺诈风险)严重影响了行业的健康发展。这些风险如果得不到很好的防控,很可能会演化成一个行业的系统性风险。在此背景下,如何做好风险管理与防范成为行业发展的重中之重。

2 互联网金融行业面临的挑战

随着互联网金融行业规模的不断扩大,借款需求持续增加,行业整体的风控水平亟待提升,主要面临的挑战包括以下3点。

- 行业数据“孤岛”现象严重。互联网金融机构每天都会接到大量借款申请,但由于缺乏数据共享机制,机构获知有关借款人的信用数据,特别是强金融属性数据的难度非常大。一方面导致机构无法有效判断客户的风险等级,另一方面也导致多头负债现象频生。这就需要通过技术手段实现信用信息的互联互通,并从海量数据中挖掘有效信息用以识别信用风险。

- 行业恶意欺诈现象越来越严重,欺诈手段不断翻新,且趋于科技化、专业化、规模化,也更具有隐蔽性,这已经严重制约了互联网金融行业的健康发展。

- 金融科技在风控制领域渗透度低,行业机构的风控水平参差不齐,很多机构仍采用线下人工的传统信用审计方式,缺

乏金融科技的应用与助力,在风控成本、风控质量、风控时效性上均无法满足实际的业务需求。

因此,首先应当解决行业数据割裂的瓶颈,形成了一个行业机构间的共享生态系统,实现互联网金融行业数据的互联与互通,有效防范行业多头负债的发生。其次,应拥有强大的信用数据分析与挖掘能力,推动数据在风控决策中的运用,帮助机构防范信用风险。最后,应最大化地将金融科技与业务融合,并通过量化评分、风险分级,对风险实现最优排序性和区分度,准确识别用户欺诈的可能性,精准反馈风险信息。

3 技术难点及应对思路

为了解决当前互联网金融行业面临的挑战,防范行业系统性风险的集中爆发,推动行业健康发展,北京宜信致诚信用管理有限公司(以下简称致诚信用)推出了致诚阿福风控平台(以下简称阿福平台),以共享为核心,基于强金融属性数据及先进风控经验、大数据分析应用及评分建模能力,采用分布式服务化的系统架构,通过RESTful应用程序编程接口(application programming interface, API)方式提供海量数据对接查询服务,帮助信贷机构防范在贷前调查、贷中授信和贷后管理中因信息不对称带来的潜在风险。除此之外,还将知识图谱、大数据分析等金融科技应用于信用审计风控领域,以提升机构风险管理的效率和效果,让金融科技成为金融行业发展的动力。这是专为网络借贷机构、消费信贷、小额信贷、银行信用卡中心等提供身份识别、反欺诈、信用评估等服务的一站式智能风控云平台。

在开发阿福平台的过程中,存在一些技术上的难点,主要包括以下几个问题。

(1) 第三方机构的系统稳定性问题

第三方机构系统规范程度不同,提供服务的能力参差不齐,如果千万级的访问量直接运行在第三方机构的系统上,绝大部分系统都难以承载。即使有缓存层作为防护,还是会存在缓存穿透的可能性,导致大量请求涌入第三方机构,引发机构的服务“雪崩”。因此,一方面要做好监控预警,另一方面也要充分挖掘第三方机构的系统承载潜力。因此,提出了第三方机构系统“健康度”的概念。

健康度是根据第三方机构历史吞吐量、响应时间、服务返回数据质量综合计算得出的。在分发请求时,会根据健康度智能决定查询数据的方式。如果健康度高,能够满足秒级响应,则实时反馈查询结果;如果健康度一般,则从缓存获取,随后异步请求机构,更新缓存内容,从而保证秒级响应;如果健康度差,则直接从缓存获取数据。同时,监控系统也会检测到机构的健康度问题,及时和第三方机构进行沟通处理。

(2) 并发访问性能问题

由于对并发访问性能有很高的要求,需要应对10亿级的被查历史数据量、千万级访问请求量、亿级的请求分发量、秒级的访问响应时间,因此通过优化缓存命中率和并发访问调度机制来解决。

首先,建立多级缓存机制,包含内存级缓存和固态硬盘(solid state drive, SSD)级缓存,内存级缓存过期时间为24 h, SSD级缓存过期时间为7天(可配置),在两级缓存的作用下,缓存命中率将达到80%以上,缓存集群采用Redis集群方式部署。缓存失效分为主动失效和被动失效,被动失效采用最近最少使用(least recently used, LRU)算法,主动失效为确保查询的信息

准确性,在系统容量充足且第三方机构健康度指标达标的情况下,主动发起信息更新行为。

其次,采用NIO+连接池的方式提高并发访问性能,减少分布式锁的使用,同时通过多级测试获得单个服务节点的服务能力上限,使线程数为最佳设置值,最大限度地利用系统性能。

(3) 异构数据处理问题

将知识图谱技术应用于信用审计风控领域,可提升机构风险管理的效率和效果。在构建知识图谱的过程中,需要处理大量的不同数据源的异构数据,如果针对每一个新增的数据源都写一个对应的适配器,会极大地影响开发效率,增加维护难度。因此,可通过引入MoonBox和Wormhole解决异构数据处理问题。

通过引入上述技术,可以实现支持自动适配不同数据源(包括MySQL、Oracle、HDFS、MongoDB等)、透明化异构数据系统异构交互方式、跨异构数据系统混算。Wormhole还可提供可视化的操作界面、极简的配置流程、基于SQL的业务开发方式,并屏蔽大数据处理底层技术细节,使知识图谱项目开发和管理变得更加可控可靠。

4 平台架构

阿福平台采用分布式架构设计,系统主要包括展示层、服务层、消息层、缓存层、数据层、监控层6部分,涵盖了从用户请求接收和分发、异构数据的接入和归一化处理、数据模型的构建、对外服务能力的输出以及系统整体健康度的检测和预警等内容,系统架构如图1所示。



图1 系统架构

(1) 展示层

通过使用Shrio和分布式会话管理技术,支持不同机构用户动态展示不同菜单功能项,优化了展示效果。同时通过使用单点登录(single sign on, SSO)技术,提供各子系统用户统一信息查询入口,包括通过用户界面(user interface, UI)直接查询和通过RESTful API批量查询,在确保用户体验的基础上,满足了用户多种查询模式的需求。

(2) 服务层

将Dubbo作为分布式服务框架,使用ZooKeeper提供服务的自动注册与发现,使用Hystrix提供服务熔断机制,使用UAVStack提供服务限流和降级、自动化应用/服务画像、无侵入调用链跟踪、一站式线程分析、秒级大规模服务图谱绘制、浏览器访问跟踪、多维可视化看板等功能。

(3) 消息层

使用Kafka和RocketMQ提供消息服务,其中Kafka应用于日志信息同步,

RocketMQ应用于订单等要求高可用的使用场景。构建统一的消息中心,支持多种消息中间件,同时抽象出消息的发送和接收、消息限流、消息去重等功能。

(4) 缓存层

使用Redis、Memcache等建立CacheManager,提供统一缓存服务接口、分布式锁服务,同时针对对象存储提供透明的序列化/反序列化服务,支持无缝扩展NoSQL数据源。

(5) 数据层

采用MySQL+HBase集群的方式,其中MySQL采用Sharding-JDBC进行分库分表,同时为提高读取性能,采用了读写分离技术,一主多从。使用DBus进行流式同步,解决不同业务系统的数据源的同步效率问题,使数据同步效率从原来的 $T+1$ 提高到准实时。

(6) 监控层

建立多级监控体系,引入ELK+UAVMonitor,能够对虚拟机、Docker、

物理机的基础性能指标、Java虚拟机、线程状态、服务整体生命周期、服务调用栈、统一日志、数据库连接池等进行全面监控和预警。

5 知识图谱在平台的应用

平台将知识图谱技术应用在反欺诈领域，为传统风控带来了极具竞争力的革新。目前利用致诚信用11年来的海量数据，构建了一个包含企业实体超过1亿、人的实体超过4 000万、关系量超过50亿的知识图谱。

知识图谱对数据的存储和组合有巨大优势，但是不利于进一步的分析，因此结合金融行业的特点，形成了一个更适用于金融场景的社交网络图谱。不同于微信、微博等一般社交网络，金融场景下的社交网络更为稀疏，且节点含有丰富的金融属性，状态变化频繁。社交网络的属性描述如下。

- 节点：基于贷前贷后数据进行用户画像，如进件次数、被拒次数、最近申请时

间等。

- 边：把实体关系细分为 own、contact、call等，并通过组合得到人与人之间的不同关系。

构建知识图谱基于的数据来源于两大块数据源：一是软件开发工具包 (software development kit, SDK) 传递的用户设备数据，如设备基本信息、IP地址信息、全球定位系统 (global positioning system, GPS) 信息等；二是致诚信用历史用户授权信息。针对不同的数据源，采用不同的处理方式。对于实时数据，基于 Spark Streaming 开发了数据导入工具，对其进行数据清洗、格式转换、自然语言处理等相关操作，可以在秒级时延内将数据导入系统。对于一些历史数据，使用 Spark 和 MapReduce 进行处理。另外，对于已经存储的社交网络数据，可以通过 RESTful API 方式进行查询，也可以将数据导入 Hive 中，供数据分析人员进行数据分析。知识图谱架构如图2所示。

通过构建的知识图谱，可以进行图特征提取，计算人脉信用得分和反欺诈得分，进行触黑关联查询等，其应用如图3所示。

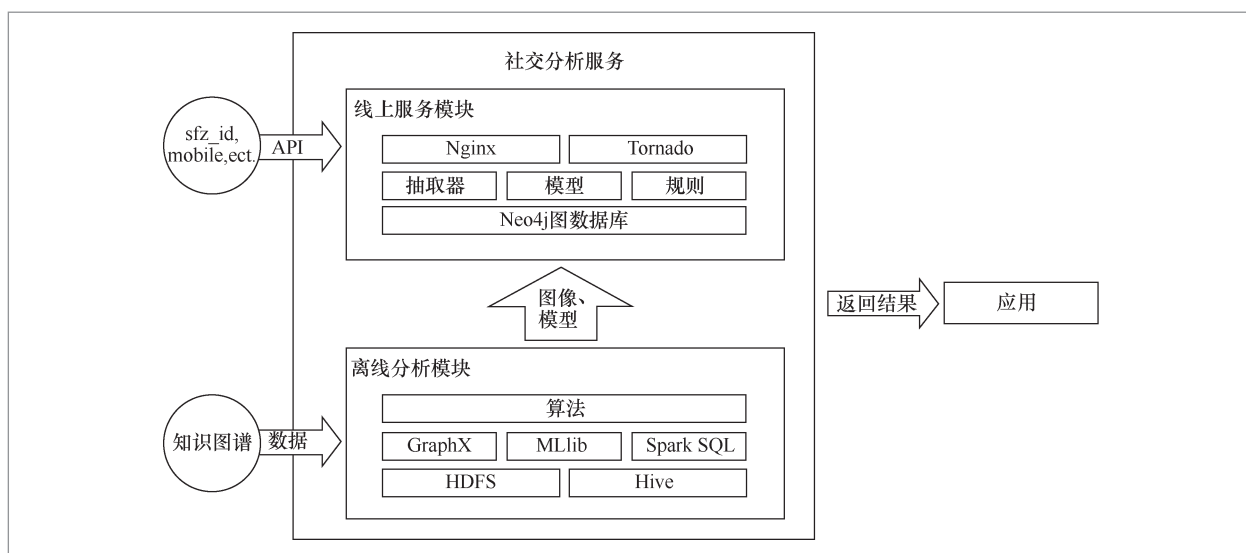


图2 知识图谱架构

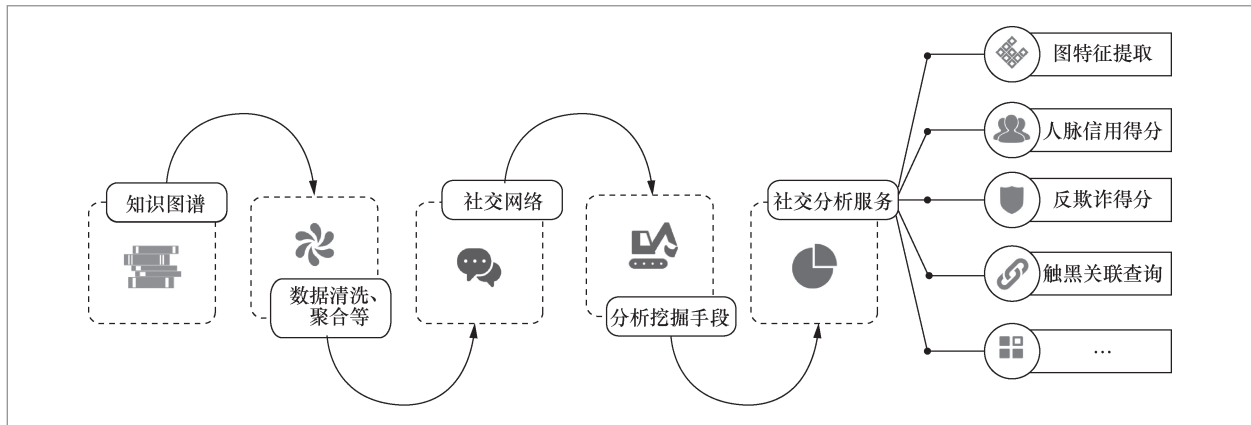


图3 知识图谱的应用

6 应用效果

(1) 大幅提升信审时效

通过强大的信用数据分析与挖掘能力，解决了行业数据割裂的瓶颈，在得到客户授权的前提下，以共享为基础，通过分布式架构设计及API调用技术，形成了一个行业机构间的共享生态系统，实现了互联网金融行业数据的互联与互通，有效防范了行业多头负债的发生。截至2017年10月，阿福平台已经为行业累计预警多头借贷1 733万次。其中，在2家及以上机构申请借款的总人数达443万人，在5家及以上机构申请借款的总人数达137.8万人，同一借款人最多向30家机构申请了借款。此外，还帮助行业机构将传统线下审批模式升级为线上智能决策的模式，并成功解决了千万级并发访问性能问题，实现了查询结果秒级反馈，大大降低了风控成本，提升了信用审计效率。

(2) 提升行业反欺诈能力

以知识图谱为底层架构建立了反欺诈的风控体系，创新性地研发了福网，以经过校验核实的多条业务线的风险数据、社交数

据以及其他数据为基础进行整合，在提升数据纯度的基础上，有效释放了数据价值，通过自然语言处理、机器学习、聚类算法等方法构建了模型，创建了多维度数据画像。通过量化评分、风险分级，对风险实现最优排序性和区分度，准确识别用户欺诈的可能性，并为机构反馈精准且区隔度高的风险信息。在欺诈客户的造假手段识别，构建客户知识图谱、社交关系网络，从更多维度识别隐蔽性欺诈、团体欺诈预警等方面，都取得了非常好的防范效果。截至2017年10月，阿福平台有效识别欺诈借款人31 784人，拦截欺诈申请109 367次。

参考文献:

- [1] ABREU D D, FLORES A, PALMA G, et al. Choosing between graph databases and RDF engines for consuming and mining linked data[C]//The 4th International Conference on Consuming Linked Data, July 13, 2013, Sydney, Australia. Aachen: CEUR-WS.org, 2013: 37-49.
- [2] BORDES A, USUNIER N, WESTON J, et al. Translating embeddings for modeling multi-relational data[C]//The 26th International Conference on Neural Information Processing Systems,

December 5–10, 2013, Lake Tahoe, USA.
New York: Curran Associates Inc., 2013:
2787–2795.

- [3] JENATTON R, ROUX N L, BORDES A, et al.
A latent factor model for highly multi-

relational data[C]//The 25th International
Conference on Neural Information
Processing Systems, December 3–6,
2012, Lake Tahoe, USA. New York:
Curran Associates Inc., 2012: 3167–3175.

作者简介



胡鹏飞(1983-), 男, 北京宜信致诚信用管理有限公司技术总经理, 主要从事分布式平台、高并发、高可用系统和大数据的应用和研究工作。

收稿日期: 2017-12-15

2016年《大数据》高被引论文Top10

排名	论文信息
1	杨丽彬, 李海林, 张飞波. 大数据环境下的管理信息系统发展研究[J]. 大数据, 2016, 2(1): 86-98. YANG L B, LI H L, ZHANG F B. Research on development of management information system in big data environment[J]. Big Data Research, 2016, 2(1): 86-98.
2	方滨兴, 贾焰, 李爱平, 等. 大数据隐私保护技术综述[J]. 大数据, 2016, 2(1): 1-18. FANG B X, JIA Y, LI A P, et al. Privacy preservation in big data: a survey[J]. Big Data Research, 2016, 2(1): 1-18.
3	潘柱廷, 程学旗, 袁晓如, 等. CCF大专委2016年大数据发展趋势预测[J]. 大数据, 2016, 2(1): 105-113. PAN Z T, CHENG X Q, YUAN X R, et al. Developing trend forecasting of big data in 2016 from CCF TFBD: interpretation and proposals[J]. Big Data Research, 2016, 2(1): 105-113.
4	冯仕政. 大数据时代的社会治理与社会研究: 现状、问题与前景[J]. 大数据, 2016, 2(2): 3-16. FENG S Z. Big data and social studies in China's governance: status quo, problems, and prospects[J]. Big Data Research, 2016, 2(2): 3-16.
5	朱扬勇, 熊贲. 大数据时代的数据科学家培养[J]. 大数据, 2016, 2(3): 106-112. ZHU Y Y, XIONG Y. Training data scientists in the era of big data[J]. Big Data Research, 2016, 2(3): 106-112.
6	王小兵. 农业信息化与大数据[J]. 大数据, 2016, 2(1): 21-27. WANG X B. Agricultural informatization and big data[J]. Big Data Research, 2016, 2(1): 21-27.
7	洪学海, 范灵俊, 洪筱楠, 等. 智慧城市建设中政府大数据开放与市场化利用[J]. 大数据, 2016, 2(3): 17-26. HONG X H, FAN L J, HONG X N, et al. Government big data opening and market utilization for smart city construction[J]. Big Data Research, 2016, 2(3): 17-26.
8	李海英. 大数据的法律挑战和建议[J]. 大数据, 2016, 2(2): 100-107. LI H Y. Legal challenges and recommendations for big data[J]. Big Data Research, 2016, 2(2): 100-107.
9	高丰. 共治共创视角下的开放数据发展: 趋势、挑战和反思[J]. 大数据, 2016, 2(2): 38-45. GAO F. Open data development in the era of co-governance and co-creation: trends, challenges and reflections[J]. Big Data Research, 2016, 2(2): 38-45.
10	王文生, 郭雷风. 关于我国农业大数据中心建设的设想[J]. 大数据, 2016, 2(1): 28-34. WANG W S, GUO L F. Envisagement of the construction of national agricultural big data center[J]. Big Data Research, 2016, 2(1): 28-34.

数据来源: 中国知网; 统计日期: 2018-01-04