

大数据及人工智能技术在交易监察领域的应用

邢波

(东软集团股份有限公司, 辽宁 沈阳 110179)

摘要: 随着中国金融市场的繁荣和发展, 金融领域监管及风险防控力度不断加大, 已成为国家级的战略重点, 保障着金融体系的安全、高效、稳健运行。大数据和人工智能技术的引入为监管领域带来新的变革, 本文结合监察领域多年来的应用实践, 将大数据及人工智能技术引入交易监察平台的规划设计, 有效的解决日趋复杂的监管环境下, 交易监察信息系统建设面临的难题, 探索出新一代智能监察平台的建设之路。

关键词: 大数据; 人工智能; 交易监察

中图分类号: TP392 **文献标识码:** A

Application of Big Data and AI in the Supervision of Financial Transactions

XING Bo

(Neusoft Corporation, Shenyang 110179, China)

Abstract: With the prosperity and development of financial markets in China, the supervision and risk control of financial markets have been intensified and have become a national strategy for safeguarding the safe, efficient and stable operation of the whole financial system. The introduction of big data and AI brings reforms and innovations to the supervision sector. In combination with years of practical applications in the field of supervision, this paper introduces big data and AI into the design of the supervision platform for financial transactions to resolve efficiently the difficulties encountered during the construction of the financial transaction information system in the increasingly complicated supervising environment and to explore a way of building a new intelligent supervision platform.

Keywords: big data; artificial intelligence; supervision of financial transactions

1 引言(Introduction)

金融体系是我国国民经济的核心, 是市场经济发展的支柱产业, 但是金融本身的高风险性及金融危机的传导效应, 使得金融市场永远是一个机会与风险并存的领域, 因此保障金融体系的安全、稳健、有效运行是国家及各级金融机构的重要任务, 对国民经济的发展起到至关重要的作用。证券交易市场是金融体系的重要组成部分, 无论在监管层面还是一线的金融机构都将交易监察和风险防控作为市场稳定运行的重要关注点, 国家陆续颁布一系列政策法规, 例如: 《中国证券监督管理委员会证券市场操纵行为认定指引》《证券公司风险控制指标管理办法》《中国证券监督管理委员会证券市场上内幕交易行为认定指引》《证券投资基金管理公司监察稽核报告内容与格式指引》《证券期货投资者适当性管理办法》等, 要求证券领域金融机构建立健全一整套风险管理、监察稽核、合规管理的内控体系, 监管层也加大了监管力度, 进而保证金融市场平稳发展^[1]。

中共中央总书记习近平主持中共中央政治局第四十次集体学习时强调, “金融活, 经济活; 金融稳, 经济稳。”并提出金融风险防控的六大任务, 从宏观角度来看, 未来金融

监管力度会继续加大, 从长远来看, 合规与风险管理将有助于营造出更健康有序的金融环境, 脱虚向实, 把资金导向合规稳健发展的实体企业。因此加强金融监管风险防控是国家级的战略重点。

2 交易监察领域信息系统建设面临的问题和挑战 (Problems and challenges encountered during the construction of the financial transaction supervision system)

按照国际上监管体系模型的划分, 我国目前证券市场采用的是公共权力机关模型, 证券及期货市场经过20多年的发展, 逐步形成了以国务院证券监督管理机构、国务院证券监督管理机构的派出机构、证券交易所、保证金监控中心、行业协会“五位一体”的监管体系和自律管理体系。无论哪个层面其信息系统建设都是行使监管职责的重要支撑, 随着证券市场的发展, 投资品种、交易模式的不断创新, 高频交易行为不断涌现, 导致证券市场交易形态日趋复杂, 异常及违规交易行为呈现多样化、隐蔽化的发展态势, 给监管机构带来了巨大的挑战。

2.1 数据的非集中管理为监管带来困难

信息系统的设计和构建是随着业务的发展不断完善的,

但是随着证券市场的发展,创新业务、创新产品的不断涌现,监管规则和手段的不断优化,原有分散的系统将导致数据被割裂,大大影响数据整体分析统计的效率,以及数据管理的便利性,为监管带来困难,首先是操作权限和数据权限无法统一的问题,其次是关联的数据无法有效对接的问题,这些问题都为数据的使用者带来工作的不便利。

2.2 穿透式监管要求多维数据处理问题

传统的证券市场监管,更多的数据是来源于交易的相关数据,但是随着违规行为日趋复杂,对于多维监管数据的获取是监管准确有效的基础,获取的渠道除了在一线的交易机构,还会扩展到公安、工商、社交媒体、银行征信等领域,为投资者和上市公司构建准确的投资者画像和上市公司画像,绘制交易关系图谱,这为监管信息系统的建设带来多类型数据存储和处理的挑战,其中包含传统的结构化数据,大量有价值的半结构化、非结构化数据,传统的数据存储和处理方式无法满足监察数据类型多样化的存储和处理要求。

2.3 实时交易数据处理的性能问题

在一线的监管机构,对于实时数据的密切监控是至关重要的,同时需要在线对数据进行大规模的计算和分析,但是随着市场的发展,交易量的大幅攀升,数据来源的不断丰富,系统推送的数据量大大增加,当数据量超过一定的量级后,导致传统监察系统的实时计算性能和大规模数据统计查询效率大幅下降,使得系统监察的实时性整体延迟,为实时违规行为的发现和处理工作带来滞后。

2.4 创新产品监管模型快速构建问题

证券市场近年来呈现快速发展的态势,不仅体现在交易量上,还体现在创新产品的不断涌现,而新产品的监管模型上线周期过长也会为监管的有效性带来问题,因此缩短新产品监管模型的构建、测试、验证、上线周期是亟需解决的关键问题。

监管数据的非集中管理、跨领域数据维度缺失、实时数据的处理性能瓶颈、创新产品的监管模型构建周期较长都为监管的及时性、精准性带来了一定困难,传统的金融监管模式和系统建设难以满足现有证券市场的数据多维性、监管时效性、分析智能化的要求。

3 大数据、云计算、人工智能技术的引入,为交易监察领域提供从理论到技术到平台的有利支撑 (Introduction of big data,cloud computing and artificial intelligence provides the supervision sector with the support of theory,technology and platform)

近年来,大数据、云计算、人工智能技术在金融体系内广泛应用为金融监管带来了深刻的改变,通过与前沿技术的融合,让金融监管体系发挥更大的效力,促进和稳定了我国金融市场的发展。

东软作为国内领先的IT解决方案供应商,为交易及监管机构提供IT解决方案已有十几年的历程,为多个大型的交易所提供了监察领域的技术咨询及系统建设服务,我们与中国证券市场共同成长起来。随着监管形态的发展和变化,我们发现在交易监察领域的系统设计和构建中对于海量数据处理、实时监管时效性、分析智能化的要求不断攀升,为监管带来了革命性的挑战。针对这一问题我们将自主研发的大数

据及人工智能共性技术平台与交易监管领域业务模式深度融合,推动大数据、云计算、人工智能技术与交易监管领域的实践验证及应用,不但解决了监察领域信息系统建设面临的关键技术难题,而且为新一代智能监察平台的规划和建设开启了创新之路。

3.1 采用业界成熟领先的开源大数据技术框架实现技术自主可控

金融领域的IT系统建设要求具有完全的自主可控,我们在监察平台技术框架设计中核心组件采用开源的Hadoop、Spark主流生态圈产品进行构建,避免技术繁杂而造成开发和运维的难度增加,这一应用在实践中得到充分的技术验证。平台的具体技术产品选型按数据采集、数据预处理、数据存储、数据计算框架、数据分析、开放服务分类,在数据采集环节选择适合多数据源收集的开源产品,并按照数据的实时性与结构分类选择适合的技术产品,使系统全面支持结构化、半结构化、非结构化、日志文件数据,以及关系型数据库数据的采集;在数据预处理环节对采集到的数据在入库前进行规则化操作,对原始数据进行清洗提升存储数据的质量;在数据存储环节选择支持结构化、半结构化、非结构化数据的存储分布式文件系统与分布式数据库,支持了多数据源的存储;在数据计算框架层面采用分布式计算的技术框架解决了传统关系型数据库计算技术无法实现的海量数据并行计算问题;在数据分析环节采用Hadoop生态圈的开源分析工具,支持包括即席查询、在线分析、文本挖掘、图谱挖掘、机器学习、全文搜索引擎等功能;在开放服务方面提供多种与业务系统松耦合的服务访问接口,包括Restful风格访问、SDK方式访问等,并提供多项可视化的基于Spring Cloud的微服务框架构建的Web应用服务供业务人员与数据分析模型设计人员使用^[2]。

3.2 底层大数据与智能分析平台设计解决多元异构数据存储与海量数据计算问题

在监察领域中数据是业务开展的基础,包括证券行业体系内及体系外的多维数据,比如体系内交易数据、账户数据、监管数据、产品数据、同业数据、资金数据等,也包括体系外的投资者户籍信息、社会关系信息、社交媒体信息、征信信息、互联网日志信息等,其中大量的数据类型及结构差异巨大,包含传统的结构化数据,大量有价值的半结构化、非结构化数据,传统的数据存储方式无法满足监察数据类型多样化的存储。在监管平台建设架构设计中我们通过底层大数据基础平台与智能分析平台的设计解决了这一问题,通过Hadoop、Spark主流生态圈产品,进行底层的大数据基础平台的建设,解决了多元异构数据存储这一难题,为监察业务开展奠定了多维数据的基础。

大数据智能分析平台的设计建设解决了传统关系型数据库无法实现的海量数据计算问题。提供计算分析功能的智能分析平台设计时主要由分析引擎与模型管理两大部分组成,分析引擎负责承载分析模型底层运算功能,主要包含多维分析、即席查询、文本挖掘、机器学习、全文搜索、图谱挖掘等大数据分析算法底层执行运算引擎,大数据平台计算引擎部分采用Hadoop与Spark高性能的并行计算框架,实现对于海量数据进行高效的并行分析和计算,该设计不但可以处理多种类型的数据,并且能够支持对TB和PB级别的超大规模数

据处理，满足了海量数据计算的需求^[3]。

3.3 高频实时交易数据采集计算与集中化的基础引擎设计满足监察高并发与实时性要求

一线监管部门对于实时交易数据的分析是监控的重点，需要在业务运转中高效实时的获取实时流数据。实时流数据是一组顺序、大量、快速、连续到达的数据序列，被视为一个随时间延续而无限增长的动态数据集，实时数据的采集，对时效性要求比较高，平台通过分布式实时采集组件，以多进程/线程并发方式获取数据流，以满足对高频实时流数据的采集要求。分布式实时采集组件的设计，具备高吞吐量特点，且吞吐能力可随着采集组件节点的增加而增加，使得数据源端不产生数据积压，降低数据源端资源使用压力，提高了数据可用的周期，同时采用业界成熟的基于内存计算技术的Spark Streaming实时流计算框架构建实时流计算引擎，对实时采集到的交易数据，根据风险预警触发规则实时计算交易的风险预警状态，从而提高风控监察预警时效性，缩短了风控监察预警发生时事件处理流程周期^[4]。

此外，在大数据智能分析平台中我们集成了用于流程管理与规程管理的流程引擎与规则引擎，这种集中化的集成多种基础引擎组件的设计大大加强了平台对第三方技术组件的集成扩展能力，这种能力通过平台的统一服务接口对业务系统开发，避免了业务系统各自集成第三方基础技术组件造成的开发资源浪费，实现基础技术组件开发的标准化。

3.4 通过原子力度抽取及智能灵活组合配置方式提升监察模型定义效能

在监察业务运转过程中，对于创新业务、创新产品的监察模型快速构建是监察实效性的重要验证，原子力度的模型抽取及智能配置是系统设计中的一大亮点，可以有效的解决这一问题。我们在底层大数据智能分析平台的设计中融入了模型管理，包括模型资源库、模型组合智能配置、运行任务调度等核心模块。模型资源库用于存储监察及风控分析模型，包括相关的评级类模型、盈亏类模型、交易监管类模型、预测类模型等都预置入模型资源库中。模型组合智能配置用于支持自由选择模型库中的模型并配置合适的分析引擎，进而组合成为实现某一监管分析业务的新模型，模型库支持扩展并通过机器学习和模拟数据演练不断的进行优化，这一设计可以有效的复用已有公用模型来快速构建满足创新产品、创新业务监管需求的监管模型，同时在算法模型升级维护过程中采用统一升级的方式，避免了不同业务应用各自维护造成的升级不同步问题。运行任务调度的设计满足算法模型后台定时任务自动调度的需求，通过配置不同的时间间隔进行算法分析模型的自动运行。平台采用集中式的算法模型、计算引擎与硬件计算资源管理，保证了金融领域的公共算法模型在不同细分行业可被大量复用，避免了各个细分领域系统重复开发造成大量的人力成本浪费，最大化的发挥算法设计师与工程师的效能。

3.5 基于负载实时侦测的弹性扩展设计解决监察业务高并发时计算资源智能调配问题

为了及时的防控市场风险，在交易监察业务中对于实时监察分析的性能要求极高，业务高并发时计算资源的智能调配是解决这一问题的有效途径。平台选择Hadoop生态圈下的

分布式计算与分布式存储技术，在技术架构层面上提供成熟稳定的集群化的存储与计算框架支撑，保证数据的存储安全与计算安全，同时提供数据可靠的备份机制。另外，基于分布式计算技术的大数据智能分析平台在架构层面支持线性扩展计算能力，通过平台提供的虚拟化管理模块实时侦测平台的运行状态，在发现存储与计算资源不足时可动态的调配或增加存储与计算资源，保证了在存储和计算资源不足时，智能的增加集群节点数量，快速提高平台的存储空间和计算能力，解决监察业务高并发时计算资源智能调配问题。此外，硬件计算资源的统一管理，避免了各个业务系统分别采购计算资源造成的资源使用率不均衡与维护成本高的问题，可以有效合理的分配硬件计算资源的使用，科学的利用计算资源对海量数据进行机器学习、人工智能、文本挖掘等高级分析，从数据中挖掘出潜藏的价值，推动业务的创新^[5]。

3.6 集中化的统一安全框架和多维安全控制模型保障监察数据的安全

数据的安全性和保密性是监察平台建设的重要关注点，平台设计中提供集中化的数据存储与访问管理，保证数据的安全。通过集中的数据权限管控，分别对不同安全等级的数据分配不同的访问权限，并基于统一的安全技术框架与多维度的安全控制模型，对用户授权及数据访问权限进行校验，保证金融机构核心数据的保密性，在设计中多维度安全模型包含组织单元、岗位、角色、用户、资源等维度，支持功能级、操作级、数据级权限控制，支持岗位、角色互斥，支持组织单元、角色、资源及权限的分级管理。

4 结论(Conclusion)

前沿技术的快速发展推动了金融创新与监管水平的日新月异，我们在金融领域应用与实践中将前沿研究成果与行业应用深度融合，为国内交易监察领域构建了从底层大数据平台到顶层智能监察系统全面的解决方案，通过大数据、云计算、人工智能技术的引入解决监管业务中面临的数据海量、监管时效化、分析智能化等问题。新一代的信息技术引领金融监管领域的巨大变革，通过多渠道、跨市场的大数据分析实现穿透式监管；通过机器学习和数据演练实现对监管模型、市场监测的精准修正；通过统一的平台建设实现资源的复用，降低市场监管的成本是新一代智能化监管的必由之路。

参考文献(References)

- [1] 洪艳蓉. 证券监管的独立性与监管治理研究[J]. 金融监管研究, 2017, 12: 5-20.
- [2] 陈俊宇. 大数据技术的发展及其研究综述[J]. 中国管理信息化, 2016, 19(20): 146-147.
- [3] 程学旗, 靳小龙, 杨婧. 大数据技术进展与发展趋势[J]. 科技导报, 2016, 34(14): 49-59.
- [4] 陈军成, 丁志明, 高需. 大数据热点技术综述[J]. 北京工业大学学报, 2017, 43(3): 358-367.
- [5] 郝庭毅, 吴恒, 吴国全. 面向微服务架构的容器级弹性资源供给方法[J]. 计算机研究与发展, 2017, 54(3): 597-608.

作者简介:

邢波(1971-), 男, 硕士, 教授. 研究领域: 软件规划设计与开发.