

金融大数据技术及基础支撑平台研究*

周峰 陈伟 李莉 林佳佳

(恒生电子股份有限公司, 浙江 杭州 310051)

摘要: 随着科技的发展, 国内金融业正迈入大数据阶段, 金融数据分析问题面临挑战, 智能化挖掘和处理金融大数据迫切需要得到实施。针对目前的数据分析模式效率低、投资风险高、应用范围窄等问题, 研究基于大数据技术的金融基础支撑平台, 推动大数据分析服务完善与进步, 促进金融行业智能化建设。

关键词: 大数据 金融业 智能分析

中图分类号: F830 **文献标识码:** A **文章编号:** 1003-9082 (2022) 04-0125-03

引言

随着数据化时代的到来, 中国金融业也逐渐受到大数据的影响, 经过长期的发展, 金融业已累积大量的数据, 尤其是非结构化的数据。目前我国金融数据量增长速度极快是由于金融行业在实际生活中具有巨大的优势去积累数据: 在业务开展过程中, 金融企业充分积累了包括客户身份信息、资产负债状况、资金往来交易等大量且具有高度价值的数 据, 在充分地利用专业技术分析和挖掘之后, 这些数据将在相应领域内产生巨大的商业价值^[1]。

目前, 大数据在金融领域的应用已取得初步成效, 得到了行业的认可。但是重大挑战依然存在, 尤其是在金融大数据场景落地方面, 巨大的数据量需要灵活可扩展的云计算技术进行存储与高效的查询处理, 也需要复杂的机器学习、数据挖掘等技巧来获得准确的结果, 这就要求金融企业在其现有平台基础上不断建设提高, 以满足更加实时、准确、可靠的金融大数据服务^[2]。

本文在广泛调研与深入分析现阶段我国金融应用领域对大数据应用需求的基础上, 参照国家《资本市场信息化建 设总体规划(2014—2020)》, 提炼金融行业大数据应用共 性需求, 对接金融行业相关技术规范、标准, 通过系统化 评测方法整合安全可靠软硬件环境, 结合申报单位多年的 行业积累, 研制“金融大数据基础支撑平台”, 聚焦包括银 行、证券、期货、基金、保险等在内的中国资本市场信息 化应用领域, 对接中国证监会等行业主管部门, 并实现解 决方案的产业化。目标平台主要涉及金融大数据存储、金 融流式大数据查询处理、金融大数据挖掘以及可靠性技术 四个方面, 虽然当前国内外已有较多的相关大数据产品在 上述各方面出现, 但尚缺乏一个统一的基础支撑平台以提

供面向金融行业的通用解决方案^[3]。

为此, 本文针对目前金融大数据研究开发方面的局限 性, 通过结合目前主流的数据挖掘、大数据采集及存储、 隐私保护、数据分析等高新技术, 提出研究开发金融大 数据基础支撑平台, 研究智能金融数据处理分析模式和体 系, 构建金融大数据管理系统^[4]。

一、金融大数据技术研究

如图1所示, 金融大数据基础支撑平台的主要技术研究 内容包括计算基础层、查询、应用分析层等三大部分。基 于上述的大数据技术, 为金融理财产品营销(精准营销、 客户挽留), 产品运营(客户画像、用户行为、智能客服), 风险管控(舆情分析打非平台、客户风险报告)、量化交易 等各类金融业务的创新提供强有力的技术基础^[5]。

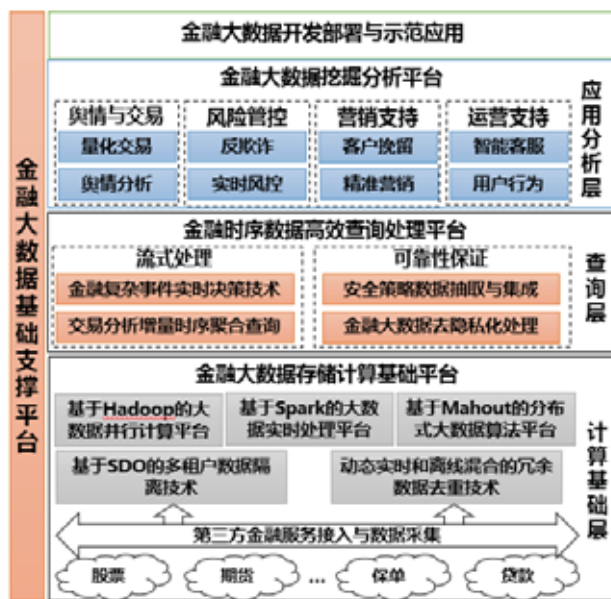


图1 研究内容结构图

* 基金项目: 2016 年浙江省重点企业研究院项目(2016Y0020)的阶段性和成果。项目编号: 2016Y0020。

二、金融大数据基础支撑平台研究

1. 金融大数据计算基础层

金融大数据分析与管理所需的数据来自多个不同的金融机构，如券商的证券交易数据、期货、基金的交易、用户信息、用户App浏览行为数据以及第三社交（微博QQ）和电信数据等。这些数据大多涉及个人隐私或公司的利益，因此如何设计一套完整的金融安全防护机制来保护这些敏感数据是整个基础平台的首要研究内容。在此之上，还需要针对海量的多源异构的数据，提供一个统一的平台来自动接入相关第三方金融数据并对其进行有效采集^[9]。

为将收集到的数据进行存储以及为下一阶段的分析计算提供支撑，提出对当前主流开源大数据技术的集成应用研究，包括使用Hadoop大数据存储计算平台来满足基本查询需求、基于Spark内存计算的大数据处理平台对实时金融决策提供快速响应。由于这些开源平台在业界已经被广泛应用，所以其在实际应用方面具有高效性、可靠性以及安全性。同时，为了保证数据独立性与节省空间的问题，本文设计对数据隔离与压缩去重技术进行研究，这样可以大大降低基础平台的建设和运维成本^[7-8]。

2. 金融大数据查询层

金融领域中的数据大多是连续多维的时序数据，而现有的金融数据平台仍然采用传统的查询方法，导致难以满足较高的查询效率。本文针对金融大数据这个特定场景进行查询优化，考虑采取增量处理来避免多余的计算，提高时序数据的查询效率。除了查询以外，提出搭建一套面向金融大数据的大规模复杂流事件处理系统，其能够侦听在线数据并对特定事件做出实时决策。为精准营销、量化交易、用户行为分析的强实时性提供强有力的技术支持^[9]。

2.1 金融复杂事件实时决策技术。金融数据在交易过程中具备非常高的实时性特点，需要一种能确保实时对数据模式进行匹配并产生结果的决策系统。大规模复杂流事件处理（CEP）系统正好满足这一需求，因此这部分决定采用CEP进行实现。

2.2 交易分析增量时序聚合查询。物化视图、概要表等索引优化方案以及树形数据结构的结合方面可以为金融时序数据的查询效率提供进一步的优化；同时，引入特殊树形结构的特征可以提供对数据增量操作等功能。

2.3 安全策略数据抽取与集成。拟开发安全策略驱动完备的数据抽取与集成平台，设立安全策略驱动步骤。为了最终达到异构业务数据的有机整合，现需要对各类金融数据进行采集、提取、分析、转换、存储等步骤整合集成。

2.4 金融大数据去隐私化处理。差分隐私保护是一种采用数据失真的隐私高级专业技术，在同时保持某些数据属性不变的情况下，采用添加噪声的方法使关键数据被隐藏，这样可以保证处理后的数据仍然具有某些统计性质，便于开展后期的数据挖掘和分析。

3. 金融大数据应用分析层

大数据的核心主要是对海量数据进行分类、聚类或关联分析来挖掘出数据之间的隐含关系或有意义的结论。在金融大数据中，希望在数据中能发现频繁出现的模式或行为，以此作为参考来指导一些金融活动。同时，在监管层面，需要在金融领域这个欺诈高发区进行有效监测。通过相应的数据挖掘手段，来有效发现个体或群体的欺诈行为以及对应的关系，并且能自动适应动态变化的欺诈行为。因此，基于冗余去重的舆情分析系统、基于CEP的量化交易系统、频繁模式的挖掘、金融用户行为分析的实时聚合查询技术以及自适应的欺诈检测将是本文的一个研究重点。

3.1 量化交易。为了避免因市场极度激进或悲观的情形下作出不理智的投资决策，采用量化交易的决策方式，所谓量化交易是指利用计算机技术将高级的数学模型代替人为主观的判断方法，通过从大量的历史数据中挖掘出能带来巨大收益的事件以制定智慧策略，极大地避免投资者因心绪波动而对最终决定产生影响。

3.2 舆情分析。有效控制和管理网络、促进和谐社会发展是舆情技术有效应用所带来的效益。主要的网络舆情技术包含话题查找与检测、自动摘取、智能分类、形势分析、舆情控制与预测等。

3.3 实时风控。建立公司的数据基础平台，完成公司要求的数据输入输出，将网状的数据关系优化为星状。达到管控数据的自动化，能够对ETL和数据效果进行有效且全面的监控。

3.4 反欺诈。利用大数据手段，将欺诈者遗留的痕迹进行挖掘和分析，形成诚信体系，避免欺诈行为发生。其具有较高的现实意义，在欺诈行为发生前利用反欺诈手段进行监测、制止，降低欺诈行为发生，提高欺诈成本，进而促进社会诚信体系建设。

3.5 客户挽留。

对于有价值的客户，需要采取相应的措施延长客户消费周期，这也是保持市场份额、平衡运营效益的关键手段，可以采用科学的方法对将要流失的客户进行挽留。

3.6 精准营销。采用现代高新的大数据技术建立金融行业与客户之间高效的沟通体系，采取精准营销实现精准的

产品推广,在低成本的情况下实现高收益的回报,促进金融业的快速发展。

3.7 智能客服。利用计算机技术设计智能客服,实现细粒度管理客户需求,建立海量用户与金融产品之间的联系,实现精准推广,同时,利用智能客服为企业制定合理高效的信息管理体系,在行业应用中充分利用知识处理手段挖掘用户的消费心理,有针对性地为用户推荐合适的金融产品。

3.8 用户行为。通过挖掘客户行为规律,将用户进行分类,有针对性地拓展客户价值。发掘潜在客户,对现有客户进行日常维护,客户挽留以及价值深挖;建立长效机制,对预警客户及时采取措施。

伴随着经济的进步和互联网行业的快速发展,大数据已经充斥在我们的日常,每个人都是产生数据的个体,当这些数据被我们收集并被人们分析时,我们的生活需求和市场需求被充分反映。首先,金融大数据的应用,避免了信息不对称的局面的产生。通过对大数据的分析,利用网络共享以及云存储,在金融产品的提供者与客户之间架起沟通的桥梁,及时地了解需求、完善需求。对于消费者而言,避开传统的产品中介,可以更直接地了解金融产品信息,掌握动态,规避风险,提高经济效益。

金融大数据降低了金融产品的系统性风险。由于缺乏相应的手段对金融产品提供者的信用进行分析评判,金融产品提供者的信用对传统金融产品领域,具有较大的影响,乃至造成严重的风险。大数据技术的广泛使用,可以在交易前对供应商的信用行为、信用价值进行查询,做到事前规避,而不是仅仅依靠财务分析或其他财产担保,而不仅仅只能事后进行追溯评估。同时,利用数据挖掘手段在交易过程中进行实时监管,这也将成为风险规避的常规措施。

金融大数据提升了金融服务效率。大数据技术基于智能处理和在线服务,其对于数据的理性分析和大量统计,使得供应商能够在正确的时间以正确的方式,有针对性地为用户提供恰当的金融服务。大大地提高了金融服务的效率。

结语

本文介绍了金融大数据技术以及基础服务平台建设,该平台主要由计算基础层、查询、应用分析层等三大部分组成,构建智能金融大数据管理,提高金融服务的效率以及降低金融投资风险。构建了独立数据处理分析体系,为金融行业提供了一个可复制推广的社会化智慧数据分析模式。而金融风险预警是金融数据挖掘中的一个重要研究方

向,由于金融数据具有类型多样、关系复杂、数据动态性、数据量大等一般特征,此外还有高噪音、非正态等特征,因此,金融风险预警更有挑战性。运用数据挖掘技术能够从海量的金融数据中发现隐藏在其背后的规律,有效地降低金融机构的运营风险,在金融风险预警有着广阔的应用价值和市场前景,因此,未来会围绕金融风险预警方向深入研究,提高金融大数据管理技术能力。

参考文献

- [1]王玲琳.大数据技术在金融信用卡领域的应用研究[J].商展经济,2021(22):75-78.
- [2]吴大胜.浅谈大数据技术在金融风控领域研究[J].中国新通信,2021,23(18):99-100.
- [3]张旭.大数据技术在金融统计分析中的应用初探[J].当代经济,2021(07):26-29.
- [4]魏思思,邱继源.基于大数据技术的金融发展研究[J].商业经济,2021(06):181-183.
- [5]王昌宜.大数据技术与金融行业的深度融合[J].商业文化.2021(16):108-109.
- [6]刘昊来,王姝霖.大数据技术背景下金融风险管理研究[J].中国集体经济,2021(04):97-98.
- [7]栗锶.大数据技术在金融行业的应用及未来展望[J].财富时代,2020(11):18-19.
- [8]张波.大数据技术及其在互联网金融风险监测领域的应用[J].金融科技时代,2020(05):43-49.
- [9]陈永坚.大数据技术与金融行业的深度融合研究[J].中国商论,2020(04):68-69.

作者简介:

周峰(1963.01—),男,汉族,浙江宁波,中级职称,硕士研究生,研究方向金融大数据、人工智能、风险管理,恒生电子股份有限公司。

陈伟(1970.07—),男,汉族,浙江杭州,本科,研究方向:金融大数据、人工智能、风险管理等。

李莉(1976.08—),女,汉族,山东昌邑,硕士研究生,研究方向:金融大数据、人工智能、风险管理等,恒生电子股份有限公司。

林佳佳(1982.01—),女,汉族,浙江温州,中级职称,本科,研究方向:金融大数据、人工智能、风险管理、项目管理等。