DOI 10.15302/J-SSCAE-2020.06.015

基于大数据的智能风险防控平台设计与实现

章明，刘培

（中国银联股份有限公司，上海 200135）

摘要：金融安全是国家安全的重要组成部分，防范化解金融风险是金融工作的根本性任务。为帮助商业银行加快打造适应数字经济时代发展需要的风险防控平台，本文基于大数据应用的关键技术，提出了一种“五层两域”智能风险防控平台总体框架；纵向包含风险数据层、特征计算层、风险模型层、决策引擎层、业务接入层，各层之间松耦合、无状态、可扩展；横向则划分为生产部署域、业务运营域，可最大程度兼顾系统运行的稳定性与业务应用的灵活度。该设计有助于商业银行实现风险数据的统一治理和统一管理，在保证风险防控平台高效稳定运行的同时，又能在风险防控运营、数据分析、模型设计、规则调整等方面为风险防控业务人员提供充足的支撑。以某金融机构部署的智能风险防控平台为例，阐述了该平台的应用情况及实际成效，并对智能风险防控平台的应用发展提出建议。

关键词：风险防控；大数据；机器学习；实时计算；金融行业

中图分类号：TP309　　**文献标识码：**A

Design and Implementation of Intelligent Risk Control Platform Based on Big Data

Zhang Ming, Liu Pei

(China UnionPay Co., Ltd., Shanghai 200135, China)

Abstract: Since financial security is an important part of national security, controlling financial risks is the fundamental task for financial management. To help banks accelerate the establishment of risk control platforms in the era of digital economy, this study proposes an overall framework of an intelligent risk control platform with “five layers and two domains” based on the key technologies of big data. Specifically, the framework vertically consists of a risk data layer, a feature computing layer, a risk model layer, a decision engine layer, and a business access layer and all these layers are loosely coupled, stateless, and extensible. Horizontally, the framwork can be divided into a production deployment domain and a business operation domain, which considers both the stability of system operation and flexibility of business application. This design is helpful for commercial banks to realize the unified governance and management of risk data. While ensuring the efficient and stable operation of the risk control platform, it can also provide sufficient support for risk control experts in risk control operation, data analysis, model design, and rule adjustment. Finally, using the intelligent risk control platform deployed by a financial institution as an example, this study expounds the application situation and practical effect of the platform and provides some suggestions.

Keywords: risk control; big data; machine learning; real-time computation; financial industry

收稿日期：2020-09-15；**修回日期：**2020-10-12

通讯作者：章明，中国银联股份有限公司工程师，研究方向为网络安全、风险防控等；E-mail: zhangming@unionpay.com

资助项目：中国工程院咨询项目“网络空间安全保障战略研究”(2017-XY-45)

本刊网址：www.engineering.org.cn/ch/journal/sscae

一、前言

近年来，受宏观经济下行压力加大、监管要求趋严、市场竞争加剧与犯罪形态升级等多重因素影响，防控金融风险的重要性日益凸显。商业银行作为金融中介机构，其经营本质是对风险的承担和管理 [1]。伴随着金融体系复杂程度的提高以及全球金融一体化进程的加快，商业银行的经营环境日益复杂，面临风险进一步加大；在新形势下，智能风险防控能力已成为商业银行获取竞争优势的关键。基于大数据、人工智能（AI）、生物识别等新技术培育大数据风险防控能力，加快智能风险防控平台的应用落地，已成为金融领域专家及学者研究的热点。陈稀 [2] 结合大数据技术和 AI 技术，通过引入内置分析工具与监测模块，为商业银行审计部门设计并实现了以风险为导向的智能审计系统。丁世博 [3] 针对互联网企业在业务快速增长时所面临的业务安全问题，研究了基于面向服务的架构（SOA）框架的安全风险防控平台。张鲁男等 [4] 以风险防控系统的架构、规则引擎和阈值体系的设计为基础，详细介绍了基于规则引擎并利用 AI 算法的实时业务风险防控系统。郭锐 [5] 从大数据风险防控平台应用的概念特征及理论基础出发，论述了大数据风险防控平台对金融信贷发展的重要作用，并以某公司为案例，分析了大数据风险防控平台构建与运营发展过程中存在的问题并提出对策建议。

目前多数风险防控应用系统是针对特定交易场景或业务需求进行逻辑处理的，并没有建立实时、动态、可更新、可扩展的风险防控体系 [6]。本文以智能风险防控平台的设计框架和实现方法为研究对象，论述数字化转型背景下商业银行对智能风险防控平台的迫切需求；同时基于大量实践经验，从大数据智能平台的关键技术出发，提出一种高可用、高复用、易扩展、易伸缩的风险防控平台架构以及各功能模块的设计方法；以某金融机构部署的智能风险防控平台为例，从应用角度说明该方法的实际成效，据此对智能风险防控平台的应用发展提出建议。

二、构建智能风险防控平台的需求分析

（一）宏观需求分析

1. 国际环境震荡多变，风险形势复杂严峻

在世界经济陷入低迷、贸易摩擦不断升级、地缘政治持续紧张等诸多因素的影响下，我国经济转型发展阻力加剧。金融是经济的血脉，防范化解金融风险，促进经济健康高质量发展，是我国决胜全面建成小康社会、全面建成社会主义现代化国家的必然要求。“十九大”报告中把坚决打好防范化解重大风险列为三大攻坚战之首，其中防控金融风险是重中之重 [7]。十九届四中全会和中央经济工作会议提出，要打赢防范化解重大风险攻坚战，必须推进治理体系和治理能力现代化。

2. 监管要求持续从紧，风险打击治理从严

由于缺乏相应监管，支付行业经历一段时间的“野蛮”发展，造成支付市场乱象丛生，风险事件频频发生，网络赌博与电信诈骗风险尤为突出。针对该情况，中国人民银行及监管部门陆续出台了一系列规范与监管措施，严厉整顿支付市场乱象。2016 年，中国人民银行发布 261 号文件，提出加强支付结算管理防范电信网络新型违法犯罪的有关事项；2019 年的 85 号文件强调需进一步加强支付结算管理，防范电信网络新型违法犯罪的发生；2020 年的 155 号文件部署开展为跨境赌博、电信网络诈骗等违法违规活动提供支付结算服务的风险排查与整治工作。面对“严监管常态化”的政策环境，商业银行应严格落实监管政策要求，补齐风险防控短板，严防发生系统性风险。

3. 业态变革不断加速，风险特征升级演变

随着支付参与主体更加开放和多元，支付的内涵和外延发生全方位变革，新型支付方式不断推陈出新，扫码支付、手机闪付、无感支付等移动创新业务成为主流，在便利人们生活方式的同时，也对传统银行的风险防控能力提出挑战。犯罪团伙通过网络化渠道并借助程序多开、分身软件、短信嗅探等黑灰产工具对移动创新业务各环节实施精准化攻击，风险防控压力向注册、开户、交易、转账等全链条渗透。商业银行应与时俱进，提前布局新型支付产品的风险防控体系，针对犯罪分子攻击新业务

的手段和特征变化快的特点，升级风险防控技术能力，强化智能风险防控建设。

（二）技术需求分析

传统的风险防控体系以定性风险管理为主 [8]。然而，基于传统架构所设计和研发的风险防控系统已经不能满足业务快速发展的需要，突出表现在以下三方面。

风险防控系统与业务系统的紧耦合导致重复建设和数据孤岛 [3]。传统系统设计通常采用垂直应用架构，风险防控系统往往作为业务系统的一个子模块；在业务形态较为单一的早期，这种架构的问题并不突出，但随着业务创新的加快，这种架构将导致大量重复的功能建设。例如，某商业银行重复建设信用卡风险防控系统、手机银行风险防控系统、在线支付风险防控系统等多套类似功能的系统，造成系统维护和升级的高昂成本；这样的架构也不利于数据沉淀，各个风险防控系统彼此难以打通，数据视角只能局限在其对接的业务场景中，而无法建立全局风险防控策略。

单机存储与算力的限制导致风险防控特征计算范围的瓶颈。风险防控系统的核心是风险特征计算，即从卡片、商户、设备等不同维度计算一段时间窗口内的统计指标，从而刻画风险程度的高低，统计指标的时间窗口跨度、统计函数的复杂度直接决定了风险防控能力的强弱。然而，传统的以 AIX/DB2 为代表的小型机架构一般只能通过增加单机的中央处理器（CPU）、内存、磁盘等方式提高处理能力，代价高昂；随着数字互联时代的到来，在大规模高并发的交易行为处理方面显得力不从心。

规则模型迭代周期长导致无法应对层出不穷的新欺诈。当前的犯罪形态已经从个体化和作坊式向集团化、专业化、智能化和国际化转变，加之猫池、伪基站、自动化脚本、流量劫持等网络黑灰产已形成一个庞大的产业链，进一步降低了犯罪成本。然而，传统风险防控系统仍然大量依赖“事后分析”的专家规则，规则参数与模型变量迭代周期长，无法满足“事前甄别、事中干预”的新需求。此外，受制于底层的数据治理和模型训练环境，单纯依靠机器学习算法并不能解决所有的风险防控难题。

三、基于大数据的智能风险防控平台关键技术

构建一个能够有效支撑大数据应用的智能风险防控平台，涉及大数据处理、实时计算、机器学习算法等多项关键技术 [10]。

（一）大数据处理

大数据可以概括为“海量数据 + 复杂类型的数据”[9,10]。Hadoop 是典型的大数据批量处理架构，目前已发展成为以分布式文件系统（HDFS）、分布式计算框架（MapReduce）、分布式数据库（HBase）等功能模块为核心的完整生态系统 [11]，支持在大型集群服务器上对文件进行分布式处理。

Hadoop 主要采用“分而治之”的思想，先对大规模数据的计算任务进行分解，然后派发到众多计算节点分别完成。其中，HDFS 负责将大规模文件分布式存储在多台服务器中，适用于海量数据的存储和读取 [12]；MapReduce 实现任务分解和调度，负责协调计算任务在多台机器上并行运算 [13]；Hbase 是运行于 HDFS 文件系统上的分布式非关系型数据库，主要用来存储非结构化和半结构化的松散数据，支持数据的实时随机读写。

Spark 是另一个知名的批量数据处理平台系统，但与 Mapreduce 将计算中间结果保存在磁盘上不同，将中间计算结果存放在内存来减少迭代过程中的数据落地，能够实现数据高效共享，提高迭代运算效率 [14]。

（二）实时计算

Hadoop 等批量操作静态数据的方式在处理实时性要求较高的业务时，难以满足应用需求。流式计算可以直接处理运动中的连续数据流，在接收数据的同时计算数据，实现秒级响应。

Storm、Flink 是流式计算框架的重要代表。Storm 是 Twitter 公司支持开发的分布式、处理流式数据的系统，采用主从式体系结构，包含一个主节点和多个从节点 [15]；主节点负责系统资源的管理和任务的协调，从节点负责执行具体的任务。Flink 是 Apache 软件基金会开发的，以数据并行和流水线方式执行任意流数据的分布式处理引擎 [16]，突出特点是将所有任务当成流来处理；批数据可作为

流数据的一个极限特例，因此 Flink 同时支持批数据和流数据的处理，采用多线程的方式来极大提高 CPU 的使用效率，具有高吞吐、低延迟、高可靠、精确计算等特性。

实时计算同样离不开消息系统和内存数据库的支撑。Kafka 是分布式发布订阅消息组件的代表[17]，由 Apache 软件基金会开发，支持中央式流数据处理；由发布者向代理发布消息，订阅者订阅消息的方式处理流式数据，将消息系统、存储系统、流处理系统组合构成了灵活伸缩的流式数据处理平台。Redis 是以 key-value 形式存储数据、运行在内存中的数据结构存储系统，可以用作数据库、缓存和消息中间件[18]，适用于高并发、大数据量的处理，能克服单一使用关系型数据库来保存数据所导致的磁盘读 / 写速度较慢等严重性能弊端。

（三）机器学习

传统数据分析技术是基于特定任务，使用预先设定的方法分析数据隐藏规律。机器学习则是在历史数据中自动发现规律并利用规律对未知数据进行分类或预测，常见的机器学习算法有监督算法、无监督算法、半监督算法、图算法[19]。

有监督算法利用已标识数据作为训练集来建立函数模型，再用模型来预测未知样本，例如逻辑回归、随机森林等；以先验知识作为输入模型训练效果相对较好，但因需要人工标注数据，所以训练成本相对较高。无监督算法通过对无标识样本集数据的学习来获取数据的内在模式及统计规律，如 K 均值聚类、主成分因子分析等；由于不需要对数据集进行标记，相应训练成本较低但训练效果难以量化。半监督算法是有监督与无监督的结合，在训练过程中利用小部分的标记数据、大部分的非标记数据进行训练学习，如标签传播算法等。图算法借助关系网络，通过个体之间的行为等信息建立全局的关系图，进而在全局关系图上发现具有一定行为模式的团体。

四、基于大数据的智能风险防控平台总体框架设计

（一）设计目标

1. 打通数据壁垒

为顺应银行业的数字化转型，从平台定位的角度看，风险防控系统不应作为业务系统的附属子系统，而应视为“大中台”的重要组成部分，“一切业务数据化、一切数据业务化”也已逐步成为行业共识。风险防控平台需要具备完备的数据接入能力，通过灵活的报文结构设计，能主动或被动地从各个业务系统实时采集数据、完成风险评估、执行风险防控动作。

2. 平衡计算资源

实时决策已逐渐成为风险防控系统的标配，但是硬件资源的投入同样远超准实时和批量系统，如何最大化利用计算资源的投入是不可回避的问题。应充分借助大数据平台处理海量数据的优势，将特征计算与模型计算解耦，从两个方面平衡计算资源：一方面，对于指标特征的计算应区分为在线与离线两类，对于多日特征提前交由大数据平台计算完成，*T*+1 日加载至内存中；另一方面，将模型区分为在线模型、离线模型，对于时效性优先的风险防控场景（如交易反欺诈、申请反欺诈场景），采用在线的有监督树类模型来提高模型计算效率，而对于分析广度优先的风险防控场景（如洗钱团伙场景），采用离线的无监督与复杂网络模型来最大程度地挖掘潜在风险网络。

3. 具备迭代能力

风险防控是 AI 技术应用的热点领域，各类软件包、建模工具日趋完善，基于海量样本和数学统计学的风险防控模型也正在逐渐取代基于小样本和专家经验的风险防控规则，但风险防控模型同样面临迭代周期长、迭代难度大的现实困难。因此，在风险防控平台设计中应考虑模型训练环境与运行环境的一体化，将模型迭代配置为一种业务人员可自主操作的轻量级更新，从而无需依赖整个风险防控平台的更新；在数据脱敏的前提下，尽可能保证建模环境与运行环境的数据一致性，避免模型在离线表现的差异性导致模型迭代失败。

（二）框架组成

本文提出一种“五层两域”总体框架（见图 1），纵向涵盖风险数据层、特征计算层、风险模型层、决策引擎层、业务接入层等 5 个功能层，各层之间松耦合、无状态、可扩展。风险数据层包含底层的数据集市、各维度的数据标签以及良好的数据管理功能；特征计算层同时支持实时数据的在线计算、

批量数据的离线计算，通过统一的特征计算调度模块实现对特征计算函数和计算周期的灵活配置；风险模型层主要用于部署有监督、无监督、半监督、复杂网络等机器学习算法，具备模型训练、验证、部署的模型全生命周期管理功能；决策引擎层完成各类风险防控规则的配置和管理，规则执行时可以调用风险模型的计算结果，最终决策当笔交易的阻断、挂起或预警等；业务接入层完成与各个业务系统的对接，按照统一的风险防控要素过滤标准完成数据采集和各类黑灰名单的匹配。此外，横向划分为生产部署域、业务运营域，与联机交易处理相关的核心功能模块属于生产部署域，而配套的参数、规则、模型、特征、数据等各类管理模块则属于业务运营域；生产部署域应重点考虑平台运行的稳定性，业务运营域则需要有良好的人机交互界面，保证业务人员根据需要灵活配置相关内容。

五、基于大数据的智能风险防控平台功能模块实现

（一）风险数据层

风险数据层位于平台的最底层，主要包括数据集市、数据标签两个功能模块以及数据查询、数据管理两个数据管理模块（见图 2）。

1. 数据集市

通过搭建 Hadoop 生态，实现对海量业务数据、风险数据、外部数据的存储和治理。业务数据的关键是通过唯一主键（如账户号、用户身份标识号）将同一客户从不同渠道发起的交易关联勾兑在一起，形成一个完整的历史交易序列，从源头上解决数据孤岛的问题。风险数据是在日常风险防控运营过程中积累的黑灰名单，如发生盗刷的卡号、涉嫌电信诈骗的商户、恶意拒付的持卡人等。外部数据是通过行业联防联控获得的外部补充数据，如运营商提供的手机基站定位数据，公安部门提供的实名认证数据等。

2. 数据标签

基于数据集市，通过数据的抽取、转换和加载，形成人、卡片、商户、设备等多维度的数据标签，并按照主要的风险类型（欺诈风险、信用风险、合规风险）形成标签分类。数据标签不是静态的，而是随着基础数据的更新不断延展和扩充。例如，根据风险数据中的欺诈卡片，关联历史数据得到交叉度最高的商户，从而对该商户增加一个疑似信息泄漏点的数据标签。

3. 数据管理

通过对接 Hive 或 Impala 等大数据查询分析引擎，业务人员可以方便地查询数据集市中存储的数据，开展日常的风险排查和建模准备，也可以对数据标签库进行加工和维护。

（二）特征计算层

特征计算层负责在线特征和离线特征的计算，

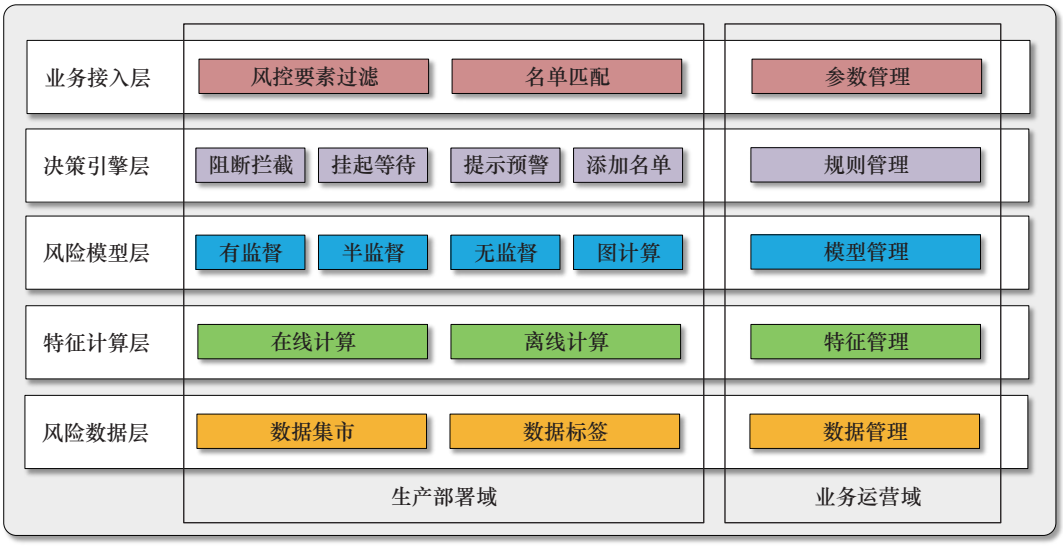

图 1　基于大数据的智能风险防控平台总体框架图

将计算完成的特征统一加载至分布式内存中，被上层的模型和规则所调用（见图 3）。

1. 在线计算

在线计算模块通常负责计算当日内的指标特征，用于刻画当前的行为变化轨迹，如卡片 30 min 的失败交易笔数、商户 90 min 内失败交易的卡片数。由于当日的交易数据尚未在数据集市中落库，因此计算的数据源来自于实时采集的数据，通过 Kafka 等消息队列组件源源不断地进入流式计算引擎，随后流式计算引擎通过滑动窗口的方式完成实时计算，并将计算结果实时加载至分布式缓存中。在线特征管理模块通过结构化查询语言（SQL）语句和可视化界面，可以对在线特征的计算对象、计算函数、窗口大小、匹配条件等进行灵活配置，配置完成后支持实时生效。

2. 离线计算

离线计算模块通常负责 *T*-1 日以前的历史特征计算，用于刻画长期行为特点，如 7 d、30 d、6 个月的特征，典型的有商户 30 d 内日交易量的标准差、卡片 6 个月内发生交易的主要城市等。由于计算的跨度周期长、数据量大，需要调度大数据分布式集群来完成特征计算，计算完成后生成特征文件通过每日定时加载的方式加载至分布式内存中。离线特征管理模块主要用于管理大数据计算任务，包括计算逻辑和计算周期；由于离线特征的计算资源开销大，因此还需要有相应的监测机制，如果发现某个任务的计算时长出现超时，则应尽快完成计算资源的弹性扩容。

（三）风险模型层

风险模型层不仅承担机器学习模型计算的任务，还需负责模型的全生命周期管理，主要包括模型训练、模型测试、模型运行功能模块（见图 4）。

1. 模型训练

模型训练时首先需要从风险数据层中抽取建模所需的数据，并在抽取的过程中通过哈希算法完成卡号、手机号、身份证号等个人敏感信息的脱敏；其次进行数据清洗和特征筛选；最后是模型搭建，完成算法选择和模型验证。此外，为提高建模效率，

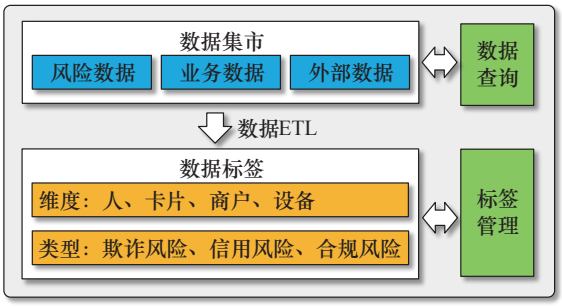

图 2　智能风险防控平台的风险数据层

注：ETL 即数据抽取、转换、装载的过程。

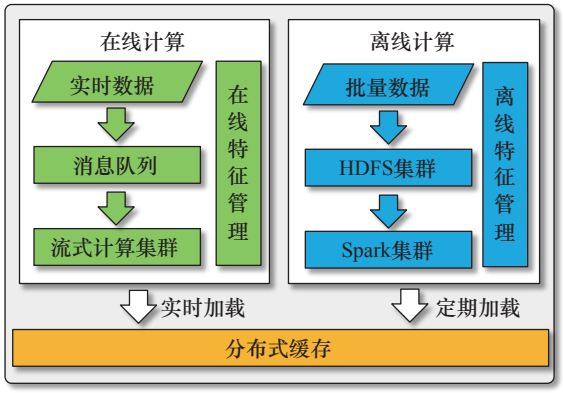

图 3　智能风险防控平台的特征计算层

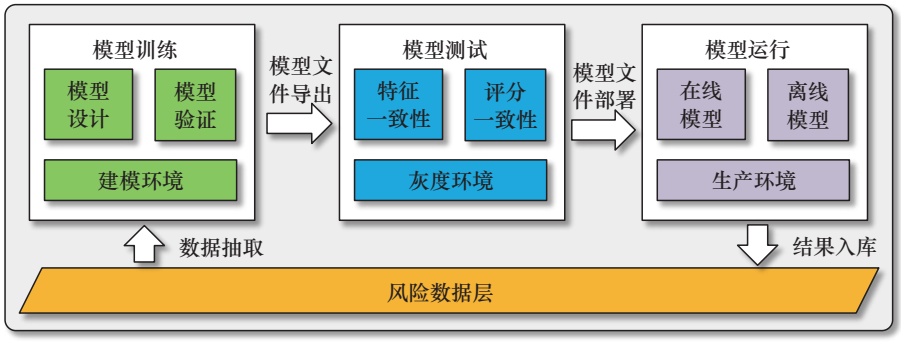

图 4　智能风险防控平台的风险模型层

建模环境支持 Python、SAS、R 等常用的建模工具和算法包，在模型指标（如 ROC 取现、KS 值等）达到预期目标，即模型定稿后生成标准格式的模型文件。

2. 模型测试

为降低由于训练样本的不充分、在离线特征计算的差异等原因所导致的模型效果欠佳，在模型正式部署前，需要通过模型测试。模型测试是在灰度发布环境中进行，全部模拟生产环境的运行情况；其中离线特征同样由离线计算模块完成，在线特征则通过实时数据的并行分流，由在线计算模块完成。最后将模型测试结果与建模环境中的训练结果进行比对，包括特征计算结果的一致性、评分分布结果的一致性，前者旨在发现特征筛选的缺陷，后者用于发现算法选择的缺陷。

3. 模型运行

在通过模型测试后，将模型文件正式部署在生产环境中，部署时需要根据模型的适配场景，配置不同的运行方式，包括在线模型和离线模型两种：前者主要适用于实时反欺诈场景，需要对每一笔交易进行实时评估，如反电信诈骗等场景；后者适用于非实时的风险防控场景，如洗钱网络侦测、团伙套现侦测等场景。模型运算的结果一方面被决策引擎层调用，另一方面也将同步记录在风险数据层，作为模型持续迭代的训练样本。

（四）决策引擎层

决策引擎层主要根据实时事件的触发，通过调用特征计算层和风险模型层，完成风险防控规则的匹配和决策动作的执行，相关功能模块如图 5 所示。

1. 规则库

规则库由具体的规则条件组成，单个条件可以表示为左操作数、右操作数、关系运算符，操作数可以抽象为当笔交易要素、上笔交易要素、在线特征、离线特征、集合等不同类型，关系运算符包括大于、等于、属于、不属于、包含、不包含以及正则表达式匹配等。条件之间一般通过与（AND）、或（OR）的逻辑关系进行组合，如果条件没有引用模型的计算结果则为简单规则，否则为复杂规则。

2. 动作执行

当一个规则的所有条件全部满足时，需要执行

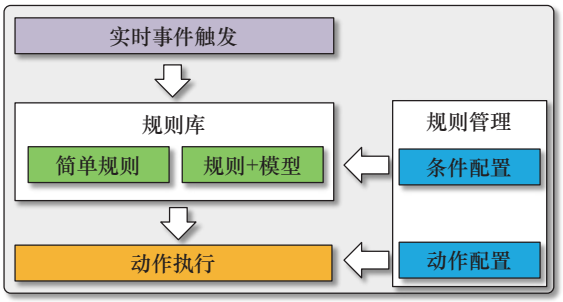

图 5　智能风险防控平台的决策引擎层

后续的风险防控动作。根据命中规则的不同风险等级，将执行相应的动作，主要包括阻断拦截、挂起等待、提示预警。阻断拦截将直接使得当笔交易失败，是最严格的干预措施；挂起等待则是给予二次确认的机会，提示预警不影响当笔交易的授权。此外，还可以叠加添加名单的动作，用户将当笔交易的某个要素自动添加至黑白灰名单中。

3. 规则管理

业务人员通过交互页面对规则的条件和工作进行配置。为提高规则编辑的效率，一般应支持规则模版的复用。

（五）业务接入层

业务接入层负责根据风险防控场景的需要开展实时数据的采集，主要包括要素过滤、名单匹配功能模块（见图 6）。

1. 风险防控要素过滤

风险防控首先需要识别当前的风险防控场景，如登陆、转账、开户等，并根据场景对应的过滤器配置，完成具体风险防控要素的过滤。风险防控要素主要包括交易信息、账户信息和设备信息等，其中交易信息有主账号、手机号、交易时间、交易金额、交易地区等，账户信息有账户开立时间、可用额度等，设备信息有全球定位系统位置、设备指纹等。此外，在要素过滤的同时还需要检查各要素的字段合规性，避免通过恶意篡改报文对风险防控平台的判断造成干扰。

2. 名单匹配

名单匹配包括白名单匹配、黑名单匹配，命中白名单则直接当笔交易直接放行通过，命中黑名单则直接阻断。名单的更新来源于三方面：根据规

则触发的关联动作自动添加、根据调查反馈由业务人员主动添加、根据行业共享人工添加。

六、应用案例

以某金融机构为例，为满足不同时期业务发展需要，该机构曾建设了多套风险防控系统，但不同渠道业务数据单独存储，各风险防控系统之间无法有效协同整合；为满足数字时代的发展需要，基于本文提出的“五层两域”智能风险防控平台的设计框架和实现方法，对现有系统进行重构和升级，打造了行业级的智能反欺诈风险防控平台（见图 7）。

该平台底层搭建了数据集市，实现对开户、支付、转账、取现等各类业务中产生的多源异构数据采集、存储和管理；在数据集市之上的是画像标签库，综合运用模式识别、自然语言处理、复杂网络等各类 AI 技术，从人、卡、设备、商户等多维度分析挖掘客户行为特征并沉淀为画像标签，标签规模达到 20 亿级，且每月以千万级的速度递增。

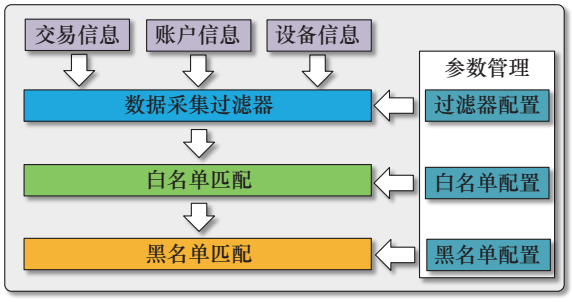

图 6　智能风险防控平台的业务接入层

对于每一笔交易，交换处理中心均会将交易报文实时转发至计量风险评分模型引擎，评分模型基于当笔交易要素、历史画像特征，通过实时流式计算引擎，在毫秒级时间内完成特征计算及当笔交易风险程度的量化评价，评价结果以分值的形式输出至收单交易监控引擎和发卡交易监控引擎中；随后，两个引擎分别从卡片和商户的维度完成基于量化评分的决策，并将决策结果返回至风险评分模型引擎，由评分模型完成综合决策结果的计算；然后，交换处理中心根据综合决策结果对当笔交易实施拦截、挂起、二次验证或者放行等动作，从而完成整个实时交易的风险防控决策。为了降低对交易成功率的影响，上述整个决策过程将在 50 ms 内完成。此外，风险防控运营人员可以对存在风险的卡片、商户开展调查处理，并根据调查情况不断调整优化特征、模型、规则配置，确保平台始终运行在最佳状态。

该平台每年为行业挽回资金损失超过 1 亿元，在提升产业环境与输出、赋能合作机构方面发挥了重要价值。

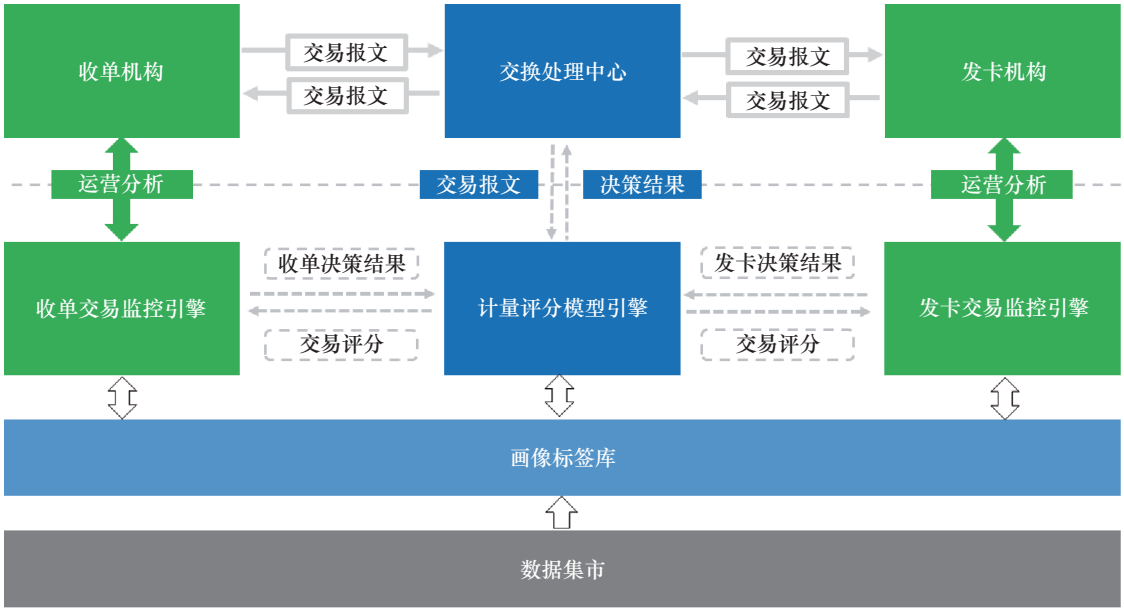

图 7　应用案例：行业级智能反欺诈风险防控平台

七、结语

面对复杂严峻的风险防控形势，金融行业正在加紧制定相关的技术标准。为切实保障技术标准的落地应用，本文以传统风险防控系统面临的数据孤岛、算力瓶颈、模型迭代周期长等突出问题为切入点，在充分理解、吸收和借鉴大数据处理、实时计算、机器学习等新兴技术的基础上，提出了一种全新的基于大数据的智能风险防控平台的设计方案。该平台运用了“大中台”的设计理念，通过“五层两域”的总体框架和实现方法，首先从风险数据、特征计算、风险模型、决策引擎、业务接入五方面构建了完整的风险防控闭环，发挥了金融大数据的风险防控价值；其次，各层之间的耦合度低、依赖性小且层内应用多采用分布式架构，使得横向扩展方便；同时从生产部署、业务运营两个维度对相关功能模块进行了具体实现和组合应用，最大程度地兼顾系统运行的稳定性与业务应用的灵活度。本文提出的设计方案与实现方法可以较好地满足商业银行的风险防控需求，从而支撑商业银行在数字经济时代的业务转型与高质量发展。

大数据智能风险防控的发展没有止境，随着第五代移动通信、物联网、区块链、AI 等新一代信息技术日趋成熟，金融与科技将进一步深度融合，智能风险防控平台在获客、授信、反欺诈、营销等全链条业务中拥有更加广阔的发展空间。对于智能风险防控平台的应用，基于当前存在的问题及现状，本文提出以下建议。

第一，坚守合规底线。大数据技术的快速发展，也伴生了不可忽视的数据滥用和信息泄露隐患。企业必须将合规发展作为生存红线，遵照法律法规和监管要求开拓业务。在数据搜集、数据存储、数据共享过程中，应满足监管、隐私保护、安全等方面的要求，确保获取客户数据合理、合规、合法。

第二，坚持技术驱动。技术驱动是大数据智能风险防控行业的根本，充分发挥机器学习、复杂网络、区块链、云计算等前沿技术的优势，强化智能风险防控的研究应用，根据经济、社会、场景、用户的变化，对算法、模型持续迭代升级，更好地为金融行业赋能。

第三，坚定业务导向。业务发展是企业的根本目标，风险控制是实现和保障业务发展的手段。智能风险防控平台的建设需要立足于当前的业务需求，统一于企业的发展目标，实现业务与风险控制协调并进，据此保障长远发展。

参考文献

[1] 刘刚. 大数据时代智能风控体系建设实践 [J]. 中国金融电脑, 2018, 349(8): 17–20.
Liu G. Practice of intelligent risk control system construction in the era of big data [J]. China Financial Computer, 2018, 349(8): 17–20.

[2] 陈稀. 基于深度学习的智能风控系统 [D]. 北京: 北京交通大学(硕士学位论文), 2019.
Chen X. Intelligent risk control system based on deep learning [D]. Beijing: Beijing Jiaotong University(Master’s thesis), 2019.

[3] 丁世博. 基于SOA的安全风控平台研究与设计 [D]. 西安: 西安电子科技大学(硕士学位论文), 2018.
Ding S B. Research and design of security risk control platform based on SOA [D]. Xi’an: Xidian University(Master’s thesis), 2018.

[4] 张鲁男, 常宝岗, 梅利. 基于规则引擎及智能阈值的实时业务风控系统 [J]. 通信技术, 2019, 52(11): 2720–2724.
Zhang L N, Chang B G, Mei L. Real-time business risk control system based on rule engine and intelligent threshold [J]. Communication Technology, 2019, 52(11): 2720–2724.

[5] 郭锐. T公司大数据风险防控平台的研究 [D]. 南京: 南京大学(硕士学位论文), 2016.
Guo R. Research on T Company’s big data risk control platform [D]. Nanjing: Nanjing University(Master`s thesis), 2016.

[6] 王欣. 基于人工智能的移动金融风险防控体系建设中的实践 [J]. 信息安全研究, 2017, 3(11): 1000–1005.
Wang X. Practice of mobile financial risk control system construction based on artificial intelligence [J]. Information Security Research, 2017, 3(11): 1000–1005.

[7] 坚决打好防范化解重大风险攻坚战 [J]. 实践, 2018 (6): 18–19.
Fight resolutely to prevent and defuse major risks [J]. Practice, 2018 (6): 18–19.

[8] 刘瑞霞. 打造基于大数据的智能化风险防控体系 [J]. 金融电子化, 2018 (8): 57–58.
Liu R X. Building an intelligent risk control system based on big data [J]. Electronic Finance, 2018 (8): 57–58.

[9] 宫夏屹, 李伯虎, 柴旭东, 等. 大数据平台技术综述 [J]. 系统仿真学报, 2014, 26(3): 489–496.
Gong X Y, Li B H, Chai X D, et al. Overview of big data platform technology [J]. Journal of System Simulation, 2014, 26(3): 489–496.

[10] Jiang C, Ding Z , Wang J, et al. Big data resource service platform for the internet financial industry [J]. Chinese Science Bulletin, 2014, 59(35): 5051–5058.

[11] 刘智慧, 张泉灵. 大数据技术研究综述 [J]. 浙江大学学报: 工学版, 2014, 48(6): 957–972.
Liu Z H, Zhang Q L. Research review of big data technology [J]. Journal of Zhejiang University: Engineering, 2014, 48(6): 957–

972.

[12] Borthakur D. HDFS architecture guide [J]. Hadoop Apache Project, 2008, 53(1–13): 2.

[13] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters [J]. Communications of the ACM, 2008, 51(1): 107–113.

[14] 宋杰, 孙宗哲, 毛克明, 等. MapReduce大数据处理平台与算法研究进展 [J]. 软件学报, 2017, 28 (3): 514–543.
Song J, Sun Z Z, Mao K M, et al. Research advance on MapReduce based big data processing platforms and algorithms [J]. Journal of software, 2017, 28(3): 514–543.

[15] 孙大为, 张广艳, 郑纬民. 大数据流式计算: 关键技术及系统实例 [J]. 软件学报, 2014 (4): 153–176.
Sun D W, Zhang G Y, Zheng W M. Big data streaming calculation: Key technologies and system examples [J]. Journal of software, 2014 (4): 153–176.

[16] Katsifodimos A, Schelter S. Apache flink: Stream analytics at scale [C]. Berlin: IEEE International Conference on Cloud Engineering Workshop (IC2EW), 2016: 193.

[17] 吴璨, 王小宁, 肖海力, 等. 分布式消息系统研究综述 [J]. 计算机科学, 2019 (S1): 1–5.
Wu C, Wang X N, Xiao H L, et al. Research review of distributed message system [J]. Computer Science, 2019 (S1): 1–5.

[18] 黄健宏. Redis设计与实现 [M]. 北京: 机械工业出版社, 2014.
Huang J H. Design and implementation of Redis [M]. Beijing: China Machine Press, 2014.

[19] 何清, 李宁, 罗文娟, 等. 大数据下的机器学习算法综述 [J]. 模式识别与人工智能, 2014 (4): 327–336.
He Q, Li N, Luo W J, et al. A survey of machine learning algorithms for big data [J]. Pattern Recognition and Artificial Intelligence, 2014 (4): 327–336.