International Journal of Smart Business and Technology
Vol.7, No.2 (2019), pp.1-14
http://dx.doi.org/10.21742/ijsbt.2019.7.2.01

Research on Financial Risk Prevention and Control Methods Based on Big Data

Sami Mohammed

*Department of Computer Science, University of Victoria, 3800 Finnerty Road,
Victoria, British Columbia V8W 3P6, Canada
smohamme@uvic.ca*

Abstract

Facing the complex and severe risk prevention and control situation, the financial industry is stepping up the formulation of relevant technical standards. To effectively ensure the application of technical standards, this article takes the prominent problems of traditional risk prevention and control systems, such as data islands, computing power limitations, and long model iteration cycles, as the entry point. Based on emerging technologies such as machine learning, a new design scheme for a smart risk prevention and control platform based on big data is proposed. The platform adopts the design concept of "large and medium platform" and adopts the overall framework and implementation method of "five layers and two domains." The closed loop of risk prevention and control gives full play to the value of risk prevention and control of financial big data; secondly, the coupling between the layers is low, the dependence is small, and the applications in the layers use distributed architecture, which makes horizontal expansion convenient; at the same time, deployment from production, The two dimensions of business operation have carried out specific implementation and combined application of related functional modules, which maximizes the stability of system operation and the flexibility of business applications. The design scheme and implementation method proposed in this paper can better meet commercial banks' risk prevention and control needs, thereby supporting commercial banks' business transformation and high-quality development in the digital economy era.

Keywords: *Financial risk, Big data, Risk prevention, Commercial banks*

1. Introduction

In recent years, preventing and controlling financial risks has become increasingly prominent due to increasing downward pressure on the macroeconomy, stricter regulatory requirements, intensified market competition, and increased crime patterns. As a financial intermediary agency, commercial banks are essentially risk-taking and managing [1]. With the increase in the complexity of the financial system and the acceleration of global financial integration, the business environment of commercial banks is becoming increasingly complex, and risks facing further increase; in the new situation, intelligent risk prevention and control capabilities have become the key to commercial banks gaining competitive advantages. Cultivating big data risk prevention and control capabilities based on new technologies such as big data, artificial intelligence (AI), and biometrics and accelerating the application of

Article history:
Received (September 5, 2019), Review Result (October 9, 2019), Accepted (November 14, 2019)

Print: 2288-8969, eISSN: 2207-516X, IJSBT

intelligent risk prevention and control platforms has become a hot spot for experts and scholars in the financial field. Chen Xi [2] combined big data technology and AI technology by introducing built-in analysis tools and monitoring modules and designed and implemented a risk-oriented intelligent audit system for the audit department of commercial banks. Paper [3] given the business security problems faced by internet companies during the rapid business growth, a security risk prevention and control platform based on a service-oriented architecture (SOA) framework was studied. Paper [4] is based on the design of the risk prevention and control system's architecture, rule engine, and threshold system and introduces in detail a real-time business risk prevention and control system based on the rule engine and using AI algorithms. The paper [5] discusses the important role of the big data risk prevention and control platform in developing financial credit from the conceptual characteristics and theoretical basis of applying the significant data risk prevention and control platform. It analyzes big data risk prevention and control using a company as a case. Problems exist in platform construction and operation development, and countermeasures and suggestions have been put forward. Currently, most risk prevention and control application systems are logically processed for specific transaction scenarios or business needs, and there is no real-time, dynamic, updateable, and extensible risk prevention and control system [6]. This paper takes the design framework and implementation methods of the intelligent risk prevention and control platform as the research object, discussing the urgent needs of commercial banks for the intelligent risk prevention and control platform under the background of digital transformation; at the same time, based on a lot of practical experience, starting from the key technologies of the big data intelligent platform, Propose a high-availability, high-reuse, easy-scalable, and easy-scalable risk prevention and control platform architecture and design method of each functional module; taking the intelligent risk prevention and control platform deployed by a financial institution as an example, the method is explained from the perspective of application Based on the actual results, suggestions are made for the application development of the intelligent risk prevention and control platform.

2. Requirement analysis for building an intelligent risk prevention and control platform

The traditional risk prevention and control system is mainly based on qualitative risk management [8]. However, the risk prevention and control system designed and developed based on the traditional architecture can no longer meet the needs of rapid business development, which is prominent in the following three aspects.

The close connection between the risk prevention and control system and the business system leads to duplication of construction and data islands [3]. Traditional system design usually adopts a vertical application architecture, and the risk prevention and control system is often used as a sub-module of the business system; in the early days, when the business form is relatively single, the problem of this architecture is not prominent, but with the acceleration of business innovation, this architecture will lead to a lot of repeated function construction. For example, a commercial bank repeated the construction of multiple systems with similar functions, such as a credit card risk prevention and control system, a mobile banking risk prevention and control system, and an online payment risk prevention and control system, resulting in high costs for system maintenance and upgrading; such an architecture is also not conducive to data Precipitation, various risk prevention and control systems are difficult to connect, and the data perspective can only be limited to the business scenarios in which they are connected. It is impossible to establish a global risk prevention and control strategy.

The limitation of stand-alone storage and computing power restricts the calculation range of risk prevention and control features. The core of the risk prevention and control system is the calculation of risk characteristics, that is, the calculation of statistical indicators in a window of time from different dimensions, such as cards, merchants, and equipment, to describe the level of risk. The time window span of statistical indicators and the complexity of statistical functions determine the strength of risk prevention and control capabilities. However, the traditional minicomputer architecture represented by AIX/DB2 can generally only increase the processing capacity by increasing the single central processing unit (CPU), memory, disk, etc., which is costly. With the advent of the digital interconnection era, processing high-concurrency transaction behaviors seems inadequate.

Unable to cope with the endless new frauds. Crime has changed from individualization and workshop type to groupization, specialization, intelligence, and internationalization. Network black and gray products such as fake base stations, automated scripts, and traffic hijacking have formed a huge industrial chain, reducing crime costs. However, the traditional risk prevention and control system still relies heavily on the expert rules of "post-analysis," the rule parameters and model variables have a long iterative cycle, which cannot meet the new requirements of "pre-identification and mid-event." In addition, subject to the underlying data governance and model training environment, relying solely on machine learning algorithms cannot solve all risk prevention and control problems.

3. Key technologies of intelligent risk prevention and control platforms based on big data

Building an intelligent risk prevention and control platform that can effectively support big data applications involves several key technologies, such as processing, real-time computing, and machine learning algorithms [10].

3.1. Big data processing

Big data can be summarized as "massive data and complex types of data" [9][10]. Hadoop is a typical big data batch processing architecture. It has developed into a complete ecosystem with functional modules such as Hadoop Distributed File System (HDFS), MapReduce, and HBase as the core [11]. It supports distributed processing of files on large cluster servers.

The computing tasks of large-scale data are decomposed and then distributed to many computing nodes for completion. Among them, HDFS is responsible for the distributed storage of large-scale files in multiple servers, which is suitable for the storage and reading of massive data [12]; MapReduce implements task decomposition and scheduling and is responsible for coordinating computing tasks in parallel operations on multiple machines [13]; Hbase is a distributed non-relational database running on the HDFS file system, mainly used to store unstructured and semi-structured loose data, and supports real-time random read and write of data.

Spark is another well-known batch data processing platform system. Still, unlike MapReduce, which stores intermediate calculation results on disk, it stores intermediate calculation results in memory to reduce data landing during iteration, achieve efficient data sharing, and improve iterative calculations Efficiency [14].

3.2. Real-time computing

Hadoop and other methods of operating static data in batches make it difficult to meet application requirements when processing real-time requirements. The streaming calculation can directly process the continuous data stream in motion, calculate the data while receiving the data, and realize the second-level response. Storm and Flink are important representatives of streaming computing frameworks. The Storm is a distributed and streaming data processing system supported by Twitter. It adopts a master-slave architecture, including a master node and multiple slave nodes [15]; the master node manages system resources and coordinates tasks. The node is responsible for performing specific tasks. Flink is a distributed processing engine developed by the Apache Software Foundation to execute arbitrary stream data in a data-parallel and pipeline manner [16]. The outstanding feature is that all tasks are processed as streams; batch data can be used as a special case of stream data. Therefore, Flink supports processing batch data and stream data simultaneously and uses multi-threading to improve the efficiency of CPU usage significantly. It has the characteristics of high throughput, low latency, high reliability, and accurate calculation.

Real-time computing is also inseparable from the message system and memory database support. Kafka represents distributed publish and subscribes message components [17]. The Apache Software Foundation develops it and supports central stream data processing; the publisher publishes messages to the broker, and the subscriber subscribes to the message to process the stream data and integrate the message system. The combination of the storage and stream processing systems constitutes a flexible and scalable stream data processing platform. Redis is a data structure storage system that stores data in the form of key values and runs in memory. It can be used as a database, cache, and message middleware [18]. It is suitable for processing high concurrency and large data volumes and can overcome single-use relationships. The slow disk read/write speed and other serious performance drawbacks caused by the large-scale database to save data.

3.3. Machine learning

Traditional data analysis techniques are based on specific tasks and use present methods to analyze data hiding laws. Machine learning automatically discovers laws in historical data and uses the laws to classify or predict unknown data. Standard machine learning algorithms include supervised, unsupervised, semi-supervised, and graph algorithms.

Because there is no need to label the data set, the corresponding training cost is low, but the training effect is complex to quantify

The supervised algorithm uses the identified data as the training set to build a function model. Then, it uses the model to predict unknown samples, such as logistic regression, random forest, etc. The training effect of the model with prior knowledge as input is relatively good, but the data needs to be manually labeled, So the training cost is relatively high. The unsupervised algorithm acquires the data's internal patterns and statistical laws by learning the unidentified sample set data, such as K-means clustering, principal component factor analysis, etc. The corresponding training cost is low because there is no need to label the data set, but the training effect is difficult to quantify. The semi-supervised algorithm is a combination of supervised and unsupervised. In the training process, a small part of the labeled and most unlabeled data, such as a label propagation algorithm, are used for training and learning. The graph algorithm uses the relationship network to establish a global relationship diagram through information such as the behavior of individuals and then finds groups with a certain behavior pattern on the worldwide relationship diagram.

4. The overall framework design of an intelligent risk prevention and control platform based on big data

4.1. Design objectives

To comply with the banking industry's digital transformation, from the perspective of platform positioning, the risk prevention and control system should not be regarded as a subsidiary subsystem of the business system. Still, they should be considered an essential part of the "big and middle platform." All business data, everything "data business," has gradually become an industry consensus. The risk prevention and control platform needs to have complete data access capabilities. Through the flexible message structure design, it can actively or passively collect data from various business systems in real-time, complete risk assessment, and perform risk prevention and control actions.

Real-time decision-making has gradually become the standard configuration of risk prevention and control systems. However, the investment of hardware resources also far exceeds that of quasi-real-time and batch systems. Maximizing the use of computing resources is an unavoidable problem. We should fully use the advantages of big data platforms to process massive amounts of data and balance computing resources from two aspects. On the one hand, indicator features should be calculated into online and offline categories, and the big data platform should calculate and complete multiple features in advance. On the one hand, the model is divided into online and offline models. The online supervised tree model improves model calculation efficiency for risk prevention and control scenarios, prioritizing timeliness. For risk prevention and control scenarios that prioritize analysis breadth, offline Unsupervised and complex network models maximize potential risks.

Risk prevention and control is a hot area of AI technology application. Various software packages and modeling tools are becoming more and more perfect. Risk prevention and control models based on massive samples and mathematical statistics gradually replace risk prevention and control rules based on small samples and expert experience. However, the risk prevention and control model also faces the practical difficulties of a long and difficult iteration cycle. Therefore, in the design of the risk prevention and control platform, the integration of the model training environment and the operating environment should be considered. The model iteration configuration should be a lightweight update that business personnel can independently operate so that there is no need to rely on the update of the entire risk prevention and control platform. Under data desensitization, try to ensure the data consistency between the modeling and operating environments as much as possible to avoid the difference in the model's offline performance from causing model iteration failure.

4.2. Frame composition

The overall framework proposed in this paper is shown in Figure 1, which covers five functional layers: risk data layer, feature calculation layer, risk model layer, decision engine layer, and business access layer. The risk data layer includes the underlying data mart, various dimensions of data tags, and good data management functions; the feature calculation layer also supports online calculation of real-time data and offline calculation of batch data, and realizes the feature calculation function through a unified feature calculation scheduling module The risk model layer is mainly used to deploy supervised, unsupervised, semi-supervised, complex network and other machine learning algorithms, and has the entire life cycle management functions of model training, verification, and deployment; the decision engine layer completes each For the configuration and management of risk prevention and control rules, the calculation results of the risk model can be called when the rules are executed, and the final decision is to block, suspend or warn the transaction; the business access layer completes the docking with each business system, according to the unified The filtering standards of risk prevention and control elements complete data collection and matching of various black and gray lists. In addition, it is horizontally divided into production deployment modules and business operation modules. The core functional modules related to online transaction processing belong to the production deployment module. In contrast, the supporting parameters, rules, models, features, data, and other management modules belong to the business operation module. The production deployment module should focus on the stability of the platform operation, and the business operation module needs to have a good human-computer interaction interface to ensure that business personnel can flexibly configure related content according to their needs.

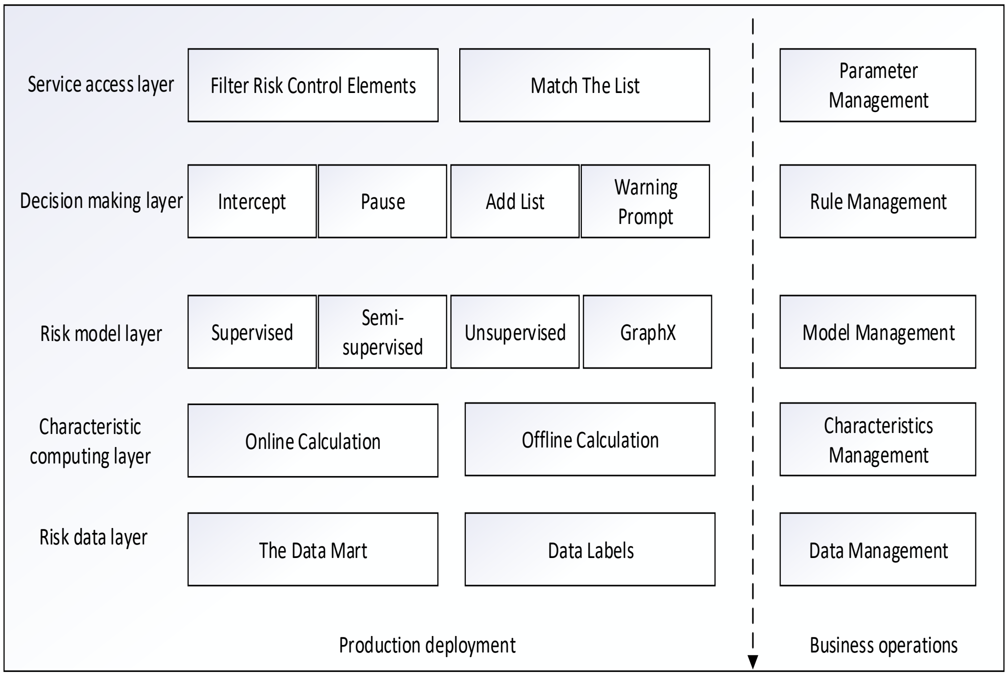

Figure 1. The overall framework

5. Implementation of function module of intelligent risk prevention and control platform based on big data

5.1. Risk data layer

The risk data layer is located at the bottom of the platform, which mainly includes two functional modules, data mart and data label, and two data management modules, data query and data management, as shown in Figure 2.

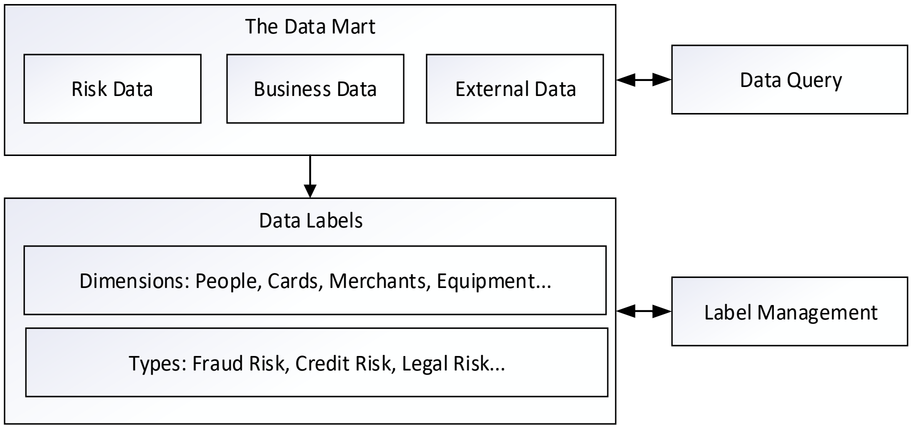
The Data Mart
Risk Data
Business Data
External Data
Data Query
Data Labels
Dimensions: People, Cards, Merchants, Equipment...
Types: Fraud Risk, Credit Risk, Legal Risk...
Label Management

Figure 2. Risk data layer structure

5.1.1. Datamart

By building a Hadoop ecosystem, the storage and governance of massive business data, risk data, and external data are realized. The key to business data is to associate transactions initiated by the same customer from different channels through a unique primary key (such as account number or user identification number) to form a complete historical transaction sequence and solve the problem of data islands from the source. Risk data is a black-and-gray list accumulated in the daily risk prevention and control operations, such as card numbers that have been stolen, merchants suspected of telecom fraud, cardholders who maliciously refuse to pay, etc. External data is supplementary data obtained through industry joint prevention and control, such as mobile phone base station positioning data provided by operators, real-name authentication data provided by public security departments, etc.

5.1.2. Data label.

Based on the data mart, through data extraction, conversion, and loading, multi-dimensional data labels such as persons, cards, merchants, and equipment are formed, and label classifications are formed according to the main risk types. The data label is not static but continuously expanded as the basic data is updated. For example, according to the fraud card in the risk data, the merchant with the highest degree of crossover is obtained by associating the historical data, thereby adding a data label of the suspected information leakage point to the merchant.

5.1.3. Data management

Data management can be achieved by docking with big data queries and analysis engines such as Hive or Impala. Business personnel can easily query the data stored in the data mart for daily risk investigation and modeling preparation. It can also process and maintain the data tag library.

5.2. Feature calculation layer

The feature calculation layer is responsible for calculating online and offline features. The calculated features are uniformly loaded into the distributed memory and called by the upper model and rules, as shown in Figure 3.

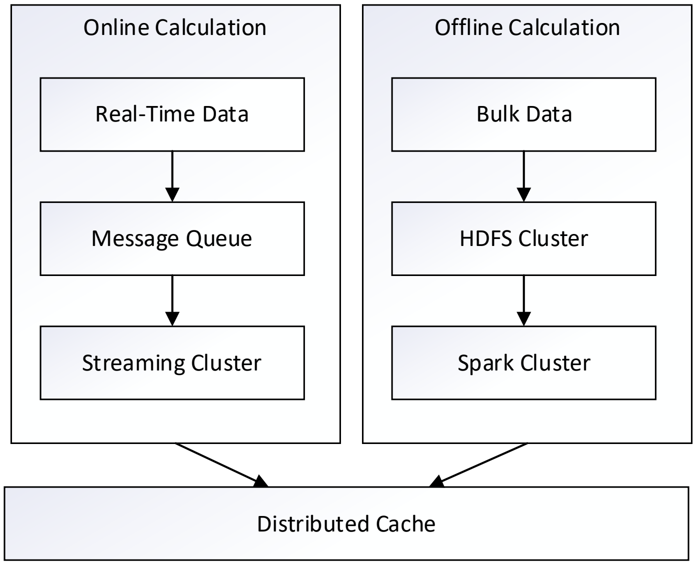

Figure 3. Feature calculation layer structure

5.2.1. Online calculation

The online calculation module usually calculates the indicator characteristics within the day. It describes the current behavior change track, such as the number of failed transactions of a card within 30 minutes and the number of failed transactions of a merchant within 90 minutes. Since the transaction data of the day has not been released in the data mart, the data source for calculation comes from the data collected in real time. It continuously enters the streaming computing engine through Kafka and other message queue components, and then the streaming computing engine passes through the sliding window. The real-time calculation is completed this way, and the results are loaded into the distributed cache in real time. The online feature management module can flexibly configure the calculation objects, functions, window sizes, matching conditions, etc., of online features through structured query language (SQL) statements and visual interfaces and support real-time effects after the configuration is completed.

5.2.2. Offline calculation

The offline calculation module is usually responsible for calculating historical characteristics before T-1 day and is used to describe long-term behavior characteristics, such as 7-day, 30-day, and 6-month characteristics. Typical examples include the standard deviation of the daily transaction volume of the merchant within 30 days and the card in 6 major cities where transactions occurred within a month, etc. Due to the long calculation and extensive data, scheduling a big data distributed cluster is necessary to complete the feature calculation. After the calculation, a feature file is generated and loaded into the distributed memory through daily timing loading. The offline feature management module is mainly used to manage big data computing tasks, including computing logic and computing cycles; due to the high cost of computing resources for offline features, a corresponding monitoring mechanism is also required. If a certain task's computing time is over, then the elastic expansion of computing resources should be completed as soon as possible.

5.3. Risk model layer

The risk model layer not only undertakes the task of machine learning model calculation but also needs to be responsible for the entire life cycle management of the model. It mainly includes model training, model testing, and model running function modules, as shown in Figure 4.

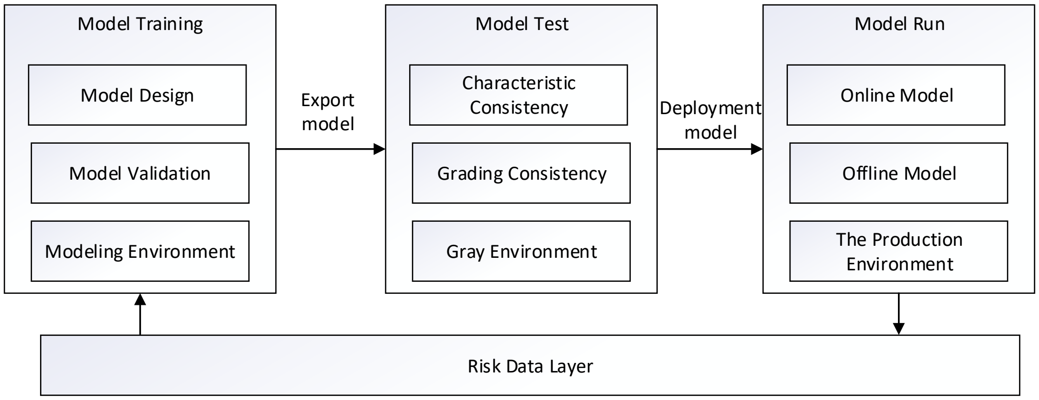

Figure 4. Risk model layer structure

5.3.1 Model training

Model training first needs to extract the data required for modeling from the risk data layer and complete the desensitization of sensitive personal information such as card number, mobile phone number, and ID number through the hash algorithm during the extraction process; secondly, perform data cleaning and Feature screening; the final step is model building, which completes algorithm selection and model verification. In addition, to improve modeling efficiency, the modeling environment supports commonly used modeling tools and algorithm packages such as Python, SAS, R, etc., to generate model files in a standard format after the model indicators reach the expected target, that is, the model is finalized.

5.3.2. Model test

To reduce the model's poor performance due to insufficient training samples and differences in offline feature calculations. Before the model is officially deployed, it must pass the model test. Model testing is carried out in a gray-scale release environment, all simulating the operation of the production environment. The offline computing module also completes the offline features, and the online computing module completes the online features through the parallel distribution of real-time data. Finally, the model test results are compared with the training results in the modeling environment, including the consistency of the feature calculation results and the consistency of the score distribution results. The former is used to find defects in feature selection, and the latter is used in algorithm selection.

5.3.3. model running

After passing the model test, the model file is officially deployed in the production environment. During deployment, different operating modes must be configured according to the model's adaptation scenarios, including online and offline models. The former mainly applies to real-time anti-fraud scenarios requiring real-time evaluation of each transaction, such as anti-telecom fraud scenarios. The latter suits non-real-time risk prevention and control scenarios, such as money laundering networks and gang cash-out detection. On the one hand, the result of the model operation is called the decision engine layer. On the other hand, it is also synchronously recorded in the risk data layer as a training sample for the continuous iteration of the model.

5.4. Decision engine layer

The decision engine layer mainly completes matching risk prevention and control rules and the execution of decision actions by calling the feature calculation layer and the risk model layer based on real-time events. The relevant functional modules are shown in Figure 5.

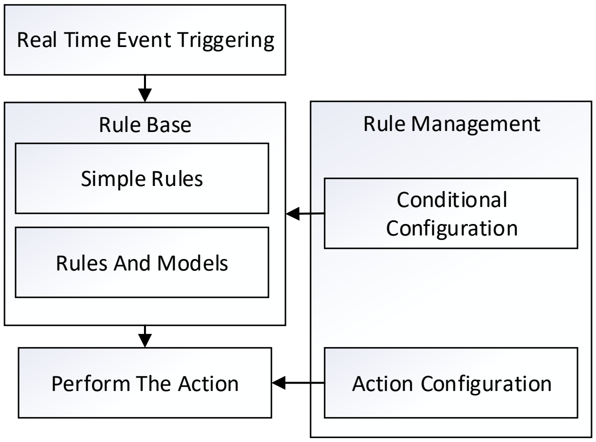

Figure 5. Decision engine layer structure

5.4.1. Rule base

The rule base is composed of specific rule conditions, and a single condition can be expressed as a left operand, a right operand, and a relational operator. Operands can be abstracted into different types, such as current transaction elements, previous transaction elements, online features, offline features, and collections. Relational operators include greater than, equal to, belongs to, does not belong to, contains, does not contain, and regular expression matching. The conditions are generally combined with AND/OR logical relations. If the conditions do not refer to the calculation result of the model, it is a simple rule. Otherwise, it is a complex rule.

5.4.2. Action execution

When all the rule conditions are met, subsequent risk prevention and control actions must be executed. According to the different risk levels of the hit rules, corresponding actions will be performed, including blocking interception, suspending waiting, and prompting warning. Blocking the interception will directly cause the current transaction to fail, which is the most stringent intervention measure; the pending transaction is to give a second confirmation opportunity, prompting early warning does not affect the authority of the current transaction. In addition, the action of adding the list can be superimposed, and the user will automatically add a certain element of the current transaction to the black/white/grey list.

5.4.3. Rule management

The operator configures the rules' conditions and tasks through the interactive page. To improve the efficiency of rule editing, the reuse of rule templates should generally be supported.

5.5. Business access layer

The business access layer is responsible for collecting real-time data according to the needs of risk prevention and control scenarios, mainly including element filtering and list matching function modules, as shown in Figure 6.

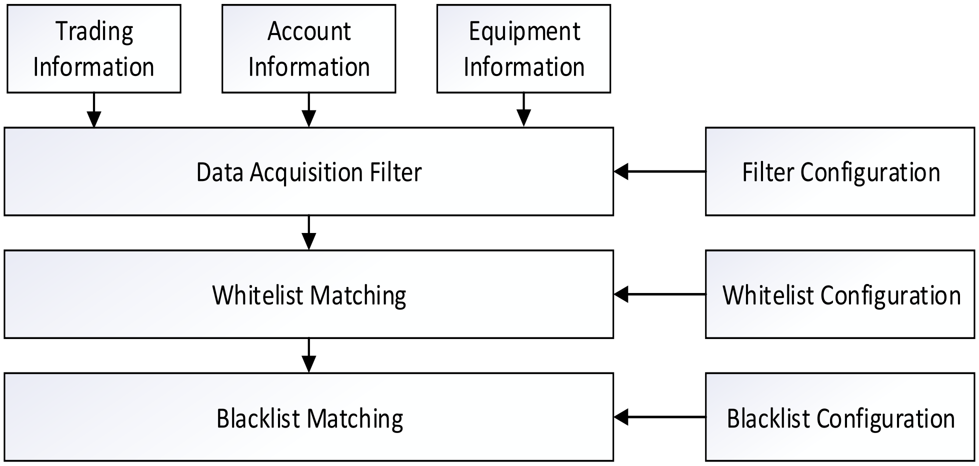

Figure 6. Business access layer structure

5.5.1 Filtering risk prevention and control elements

First, Risk prevention and control must identify the current risk prevention and control scenarios, such as login, transfer, and account opening. Complete the specific risk prevention and control elements filtering according to the filter configuration corresponding to the scene. Risk prevention and control elements mainly include transaction information, account information, and equipment information, including main account number, mobile phone number, transaction time, transaction amount, transaction area, etc. The account information includes account opening time, available quota, etc., and device information includes GPS location, device fingerprints, etc. In addition, while filtering the elements, it is necessary to check the legality of the fields of each component to avoid interference with the judgment of the risk prevention and control platform through malicious tampering of messages.

5.5.2 matching list

List matching includes whitelist matching and blacklist matching. If the transaction is matched to the whitelist, the transaction will be released directly, and the blacklist will be blocked directly. The list is updated in three aspects: automatic addition based on related actions triggered by rules, active addition by business personnel based on survey feedback, and manual addition based on industry sharing.

6. Application cases

Based on the design framework and implementation method of the intelligent risk prevention and control platform proposed in this paper, a financial institution reconstructed and upgraded the existing system to create an intelligent anti-fraud risk prevention and control platform. At the bottom of the platform, a data mart is built to realize the collection, storage and management of multi-source heterogeneous data generated in various businesses such as account opening, payment, transfer, and cash withdrawal; above the data mart is a portrait tag library, which is integrated Various AI technologies such as pattern recognition, natural language processing, and complex networks analyze and mine customer behavior characteristics from multiple dimensions such as people, cards, equipment, merchants, etc., and deposit them into portrait tags. The tag scale reaches 2 billion levels and tens of millions per month—the speed of the stage increases.

The exchange processing center will forward the transaction message to the real-time measurement risk scoring model engine for each transaction. The scoring model is based on current transaction elements and historical portrait features. Through the real-time streaming calculation engine, the feature calculation and the quantitative evaluation of the degree of risk of the current transaction are completed in milliseconds. The evaluation results are output as points to the acquiring transaction monitoring engine and the issuing transaction monitoring engine. Subsequently, the two engines complete the decision based on the quantitative score from the dimensions of the card and the merchant and return the decision result to the risk-scoring model engine. The scoring model completes the calculation of the comprehensive decision result. Then, the exchange processing center implements actions such as interception, suspension, secondary verification, or release of the current transaction based on the result of the comprehensive decision, thereby completing the risk prevention and control decision of the entire real-time transaction. To reduce the impact on the success rate of the transaction, the entire decision-making process mentioned above will be completed within 50 milliseconds. In addition, risk prevention and control operators can investigate and process cards and merchants that are at risk and constantly adjust and optimize features, models, and rule configurations based on the investigation situation to ensure that the platform is always running in the best state.

7. Conclusion

There is no end to the development of big data intelligent risk prevention and control. As the fifth generation of mobile communications, the Internet of Things, blockchain, AI, and other new-generation information technologies become more mature, finance and technology will be further integrated, and intelligent risk prevention and control platforms will be gaining ground. There is a broader space for development in the entire chain of businesses, such as customer service, credit extension, anti-fraud, and marketing. Based on the current problems and status quo, this article puts forward the following suggestions for applying the intelligent risk prevention and control platform.

The rapid development of big data technology is also accompanied by hidden dangers of data abuse and information leakage that cannot be ignored. Enterprises must regard legal development as a red line for survival and to develop businesses in compliance with laws, regulations, and regulatory requirements. In the process of data collection, data storage, and data sharing, the requirements for supervision, privacy protection, and security should be met to ensure that customer data acquisition is legal.

Technology-driven is the foundation of the big data intelligent risk prevention and control industry, giving full play to the advantages of cutting-edge technologies such as machine learning, complex networks, blockchain, and cloud computing. Strengthen the research and application of intelligent risk prevention and control.

Business development is the fundamental goal of an enterprise, and risk control is a means to achieve and guarantee business development. The construction of an intelligent risk prevention and control platform must be based on current business needs, unify with the company's development goals, and achieve the coordination of business and risk control to ensure long-term development.

Reference

[1] V. Srinivasan, "The intelligent enterprise in the era of big data," John Wiley and0 Sons, **(2016)**

[2] I. Osband, "Risk versus uncertainty in deep learning: Bayes, bootstrap and the dangers of dropout," NIPS Workshop on Bayesian Deep Learning, pp.192, **(2016)**

[3] J. Sun and Y. Chen, "Intelligent enterprise information security architecture based on service-oriented architecture," 2008 International Seminar on Future Information Technology and Management Engineering, IEEE, pp.196-200, **(2008)**

[4] L. N. Zhang, B. G. Chang, and L. Mei, "Real-time business risk control system based on rule engine and intelligent threshold," Communication Technology, vol.52, no.11, pp.2720-2724, **(2019)**

[5] T. M. Choi, S. W. Wallace, and Y. Wang, "Big data analytics in operations management," Production and Operations Management, vol.27, no.10, pp.1868-1883, **(2018)**

[6] N. Siddiqi, "Credit risk scorecards: Developing and implementing intelligent credit scoring," John Wiley and Sons, **(2012)**

[7] P. T. Susca, "Using processes to prevent and predict risk," Professional Safety, vol.63, no.8, pp.18-21, **(2018)**

[8] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of big data - evolution, challenges, and research agenda," International Journal of Information Management, vol.48, pp.63-71, **(2019)**

[9] K. Noh and D. Lee, "Bigdata platform design and implementation model," Indian Journal of Science and Technology, vol.8, no.18, pp.1, **(2015)**

[10] C. Jiang, Z. Ding, and J. Wang, "Big data resource service platform for the internet financial industry," Chinese Science Bulletin, vol.59, no.35, pp.5051–5058, **(2014)**

[11] Research review of big data technology

[12] D. Borthakur, "HDFS architecture guide," Hadoop Apache Project, vol.53, no.1-13, pp.2, **(2008)**

[13] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," Communications of the ACM, vol.51, no.1, pp.107-113, **(2008)**

[14] J. Yang and X. Li, "MapReduce based method for big data semantic clustering," 2013 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, pp.2814-281, **(2013)**

[15] D. W. Sun, G. Y. Zhang, and W. M. Zheng, "Big data streaming calculation: Key technologies and system examples," Journal of Software, no.4, pp.153-176, **(2014)**

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

[17] S. Parhizi, H. Lotfi, and A. Khodaei A, "State of the art in research on microgrids: A review," Ieee Access, vol.3, pp.890-925, **(2015)**

[18] A. Fahad, N. Alshatri, and Z. Tari, "A survey of clustering algorithms for big data: Taxonomy and empirical analysis," IEEE Transactions on Emerging Topics in Computing, vol.2, no.3, pp.267-279, **(2014)**

This page is empty by intentions.