DOI 10.15302/J-SSCAE-2020.06.015

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

Zhang Ming, Liu Pei

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

Abstract: Financial security is a key aspect of national security and controlling financial risk is a fundamental task within financial management. To help banks accelerate the establishment of risk control platforms in the era of the digital economy, this study proposes an overall framework for an intelligent risk control platform with "five layers and two domains," based on the key technologies of big data. Specifically, the framework consists of a risk data layer, a feature computing layer, a risk model layer, a decision engine layer, and a business access layer, all of which are loosely coupled, stateless, and extensible. Horizontally, the framework can be divided into a production deployment domain and a business operation domain, which considers both the stability of system operation and the flexibility of business applications. This design is helpful for commercial banks to realize 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 to experts in risk control operations, 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 on the application situation and practical impact of the platform and provides some suggestions.
Keywords: risk control; big data; machine learning; real-time computation; financial industry

1 Introduction

In recent years, the importance of preventing and controlling financial risk has become increasingly important due to multiple factors, such as increasing downward pressure on the macroeconomy, stricter regulatory requirements, intensified market competition and escalation of crime patterns. As a financial intermediary institution, the primary operation of a commercial bank is to bear and manage risk [1]. With the increase in the complexity of the financial system and the acceleration of the global financial integration process, the business environment of banks is becoming more complicated and the degree of risk has increased. Under the new situation, intelligent risk control capabilities have become a competitive advantage for commercial banks. The key issue lies in identifying ways to cultivate big data risk control capabilities based on new technologies, such as big data, artificial intelligence (AI), and biometrics, and accelerate the application of intelligent risk control platforms. This area has become a hotspot for experts and scholars in the financial field. Chen [2] combined big data technology and AI technology to design and implement a risk-oriented intelligent auditing system for the audit department of commercial banks by introducing built-in analysis tools and monitoring modules. Ding [3] studied the risk control platform based on the service-oriented architecture to solve the business security problems faced by Internet companies during rapid business growth. Zhang et al. [4] introduced, in detail, a real-time business risk control system with a rule engine and AI algorithms, based on the design of risk control system architecture, a rule engine, and a threshold system. Guo [5] first described some conceptual features and presented a theoretical basis for the application of a big data risk control platform; they then discussed the key role played by big data risk control platforms in the development of financial credit. In addition, taking a company as the entry point, they analyzed the existing problems associated

Received date: September 15, 2020; **Revised date:** October 12, 2020
Corresponding author: Zhang Ming, engineer of China UnionPay Co., Ltd. Major research fields include network security and risk control. E-mail: zhangming@unionpay.com
Funding program: CAE Advisory Project "Strategic Research on Cyberspace Security Assurance"（2017-XY-45）
Chinese version: Strategic Study of CAE 2020, 22 (6): 111–120
Cited item: Zhang Ming et al. Design and Implementation of Intelligent Risk Control Platform Based on Big Data. *Strategic Study of CAE*, https://doi.org/10.15302/J-SSCAE-2020.06.015

with the construction, operation and development of a big data risk control platforms and put forward countermeasures and suggestions.

At present, most risk control application systems perform logical processing for specific transaction scenarios or business needs, and no such system has yet established a real-time, dynamic, updateable, and scalable mechanism [6]. This study takes the design framework and implementation methods of an intelligent risk control platform as the research object and discusses the urgent needs of commercial banks for intelligent risk control platforms in the context of digital transformation. At the same time, starting from the key technology of the big data intelligent platform, this study proposes a high-availability, high-reusable and easily scalable platform architecture and design method for each functional module through practical experience. Taking an intelligent risk control platform, deployed by a financial institution as an example, the actual applications from a practical perspective are illustrated. Finally, this study puts forward relevant suggestions for the development of an intelligent risk control platform.

2 Demand analysis for building an intelligent risk control platform

2.1 Macro demand analysis

2.1.1 The volatile international environment and severe risk situation

Economic transition and the development of China encounters numerous obstacles, such as a sluggish world economy, escalation of trade friction, and continued geopolitical tension. Finance is the bloodline of an economy. Preventing and dissolving financial risks and promoting healthy economic development is an inevitable requirement for China to build a moderately prosperous society and a modern socialist country. In the report of the 19th National Congress of the Communist Party of China, General Secretary Xi resolutely mentioned that prevention and resolution of major risks is among the top three challenges and that risk control in finance is a priority. [7]. The Fourth Plenary Session of the 19th Central Committee and the Central Economic Work Conference proposed that we must advance the modernization of governance systems and capabilities to prevent and resolve major risks.

2.1.2 Tighter regulatory measures and stricter risk management

Due to the lack of corresponding supervision, the payment industry has experienced a series of barbaric developments, which had led to chaos in the payment market and frequent risk events. Among these, the risks of network gambling and telecom fraud are particularly prominent. In response to this situation, the People's Bank of China and other regulatory authorities have successively introduced multiple norms and measures to rectify chaos in the payment market. In 2016, the People's Bank of China issued Document No. 261, proposing to strengthen the management of payments and settlement to prevent new types of illegal crimes through telecommunications and networks, and in 2019 Document No. 85 emphasized this further. In 2020, the People's Bank of China issued Document No. 155 to lay out the risk investigation and rectification work of providing payment and settlement services for illegal activities, including cross-border gambling and telecommunications network fraud. Faced with strict supervision, commercial banks should strictly implement the requirements of regulatory policies, make up for the shortcomings in risk control, and strictly prevent the occurrence of systemic risk.

2.1.3 The gradually accelerating business form and escalating risk characteristics

As payment participants become more diversified, the connotations and extensions of payments have undergone several changes. Further, new payment methods are constantly emerging. Mobile innovative services, such as QR codes, Mobile QuickPass, and contactless payments have become popular. Although they make our lives easier, they also challenge the ability of traditional banks to control risk. Criminal gangs carry out precise attacks through network channels on the links of innovative mobile payments with the help of illegal technical tools, such as executive multiprogramming, split software, and SMS sniffing. Consequently, the pressure of risk control has penetrated the entire chain comprising registration, account opening, transactions, and transfers. Commercial banks should keep pace with the times, deploy a risk control system for new payment products in advance, upgrade their risk control technology capabilities, and strengthen the construction of intelligent risk control platforms in response to various types of criminals' attacks, which are characterized by rapid changes.

2.2 Technical demand analysis

Traditional risk control systems are mainly based on qualitative risk management [8]. However, a risk control system that is designed and developed based on a traditional architecture is incapable of adapting to the rapid developments, which are prominent in the following three aspects.

DOI 10.15302/J-SSCAE-2020.06.015

The tight coupling between the risk control system and the business system leads to repeated construction and data islands [3]. Traditional system designs typically adopt a vertical application architecture. Risk control systems are often used as sub-modules in business systems. In the early days, when the form of business was relatively simple, the problems associated with the aforementioned type of architecture were not significant. However, the acceleration in business innovation has led to repeated functional constructions; for example, a commercial bank repeatedly builds systems for credit card risk control, mobile banking risk control, online payment risk control, and other systems with similar functions, resulting in high system maintenance and upgrade costs. In addition, such an architecture is not conducive to data accumulation owing to the difficulty for various risk control systems to be integrated. As the data perspective can only be limited to the business scenarios in which it is connected, a global risk control strategy cannot be established.

The limitations of single-machine storage and computing power lead to a reduction in the ability to calculate the risk characteristics. The core of risk control systems lies in the calculation of risk characteristics, that is, calculating the statistical indicators within a period of time using different dimensions, such as cards, merchants, and devices, to describe the degree of risk. The risk control ability is directly determined by the time span and the complexity of the functions of statistical indicators. However, the traditional minicomputer architecture, represented by AIX/DB2, can only improve the processing capacity by increasing the CPU, memory, and disk of a single machine by incurring high costs. Therefore, in the current era of digital interconnection, it is difficult to handle large-scale and highly concurrent transactions using traditional machines.

The long iteration cycle of the rule model makes it impossible to deal with the endless new frauds. The current form of crime has shifted from individualization and workshops to groupization, specialization, intelligence, and internationalization. At the same time, illegal technical tools, such as modem pools, pseudo base stations, automated scripts, and website traffic hijacking, have formed a huge industrial chain, further reducing the cost of crime. However, traditional risk control systems still rely heavily on the expert rules of post-analysis; the rule parameters and model variables have a long iterative cycle, which cannot meet the new demands of pre-identification and intervention during the event. In addition, given the underlying data governance and model training environment, relying solely on machine learning algorithms cannot solve all the problems related to risk.

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

An intelligent risk control platform can be developed to effectively support big data applications involving several key technologies, such as big data processing, real-time computing, and machine learning algorithms [10].

3.1 Big data processing

Big data can be simply summarized as massive data + complex types of data [9,10]. Hadoop is a typical architecture used to handle bulk data. It has now developed into a complete ecosystem centered on functional modules, including the Hadoop distributed file system (HDFS), distributed computing framework (MapReduce), and distributed database (HBase) [11]. The system supports the distributed processing of files on large integrated servers.

Hadoop essentially adopts the idea of divide and conquer, wherein large-scale computing tasks are first decomposed and then dispatched to several computing nodes for completing each task separately. The HDFS is responsible for the distributed storage of large-scale files on multiple servers, which is suitable for the storage and reading of massive data [12]. MapReduce realizes task decomposition and scheduling; it is responsible for coordinating computing tasks executed on multiple machines in parallel operations [13]. HBase is a distributed nonrelational database running on an HDFS file system. It is mainly used to store unstructured and semi-structured loose data, and supports real-time reading and writing of data.

Spark is another well-known bulk data processing system. However, unlike MapReduce, which stores the intermediate calculation results on a disk, Spark stores the results in memory, reducing data landing during the iteration process, enabling efficient data sharing, and improving the efficiency of iterative operations [14].

3.2 Real-time calculation

Hadoop, and other methods, that operate on static data in batches, face difficulty in meeting the needs of businesses requiring high real-time performance. Stream computing can directly deal with a continuous data stream in motion, calculate the data while receiving the data, and realize a second-level response.

Storm and Flink are important representatives of the stream computing framework. Storm is a distributed system

for processing stream data developed by Twitter. Storm adopts a master–slave architecture, including one master node and several slave nodes [15]. The master node is responsible for system resource management and task coordination, and the slave 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 parallel and pipeline manner [16]. The biggest feature of Flink is that it treats all tasks as streams. Bulk data can be regarded as a special case of stream data. It supports the processing of both bulk data and stream data. Owing to the multi-thread method, it can greatly improve CPU efficiency with the features of high throughput, low latency, high reliability, and accurate calculation.

Real-time computing is also dependent on the support of the message and memory databases. Kafka, developed by the Apache Software Foundation, is a representative of distributed publish–subscribe message components [17], and supports central stream data processing. In Kafka, publishers publish messages to agents, and subscribers subscribe to messages to deal with stream data. Kafka combines message systems, storage systems, and stream processing systems to form a flexible and scalable platform. Redis is a storage system of a data structure that stores data in the form of key-value and runs in memory. It can be used as a database, cache, and message middleware [18]. It is suitable for processing large amounts of data with high concurrency and can overcome the serious disadvantages of slow disk read/write speed caused by the use of only relational databases in order to save data.

3.3 Machine learning

Traditional data analysis techniques are based on specific tasks using preset methods to analyze hiding data laws. Machine learning automatically discovers laws from historical data and then uses the laws to classify or predict unknown data. Common machine learning algorithms include supervised, unsupervised, semi-supervised, and graph algorithms [19].

A supervised learning algorithm, such as logistic regression or random forest, uses identified data as the training set to establish a function model to predict unknown samples. The supervised algorithm uses prior knowledge as input to achieve a better training effect, but the training cost is relatively high owing to the requirement of manual annotation of data. An unsupervised algorithm, such as K-means clustering, principal components, or factor analysis, acquires the internal patterns and statistical laws of data by learning from an unidentified sample dataset. Because there is no need to label the dataset, the training cost of the unsupervised algorithm is lower, but the training effect is difficult to quantify. A semi-supervised algorithm, such as label propagation algorithm, is a combination of supervised and unsupervised algorithms. In the training process, a small part of the identified data and most of the unidentified data are used for training and learning. A graph algorithm uses the relationship network to establish a global relationship diagram through information, such as individual behavior, and then discovers groups with a certain behavior pattern on a global relationship diagram.

4 Design of an overall framework for an intelligent risk control platform based on big data

4.1 Design goals

4.1.1 Break through data barriers

To comply with the digital transformation of the banking industry, from the perspective of platform positioning, the risk control system should not be regarded as a subsystem but as a key component of the business system. The industry consensus has gradually become “all business can be data-oriented and all data can be business-oriented.” The risk control platform must therefore have good capability for data access. Through flexible message structure design, it can actively or passively collect data from various business systems in real time, and can complete risk assessments, perform risk control actions, etc.

4.1.2 Balance computing resources

Real-time decision-making has gradually become the standard configuration of risk control systems, but the investment of hardware resources far exceeds that of quasi-real-time systems and bulk data processing systems. Maximizing the use of computing resources is an unavoidable problem. Therefore, we should make full use of the advantages of big data platforms to process the mass of data to decouple feature calculations from model calculations, and then achieve a balance of computing resources from the following aspects. On one hand, the calculation of indicator features should be divided into two categories: online and offline. For multi-day features, these should be calculated by a big data platform in advance and loaded into memory on day T+1. In addition, the model should also

DOI 10.15302/J-SSCAE-2020.06.015

be divided into online and offline models. For scenarios with high timeliness requirements, such as transaction anti-fraud and application anti-fraud, online supervised tree models can be used to ensure the high computational efficiency of the model. For scenarios with high analysis breadth requirements, such as money laundering gangs, offline unsupervised and complex network models can be used to discover potential network relations.

4.1.3 Iterative ability

Risk control is a hot topic in the domain of AI technology. The increasingly rich types of software packages, modeling tools, and risk control models based on massive samples and mathematical statistics are gradually replacing the rules based on small samples and expert experience. However, risk control models also face difficulties, such as long iteration cycles and high iteration difficulty. Therefore, the integration of the model training environment and operating environment should be considered in the design of risk control platforms. The model iteration should be designed as a lightweight update that business personnel can operate independently, so that there is no need to rely on the update of the entire risk control platform. Under the premise of data masking, data consistency between the modeling environment and the operating environment should be ensured to avoid a difference in performance between online and offline models, leading to the failure of model iteration.

4.2 Frame composition

To achieve the abovementioned objectives, this study proposes an overall framework comprising five layers and two domains, as shown in Fig. 1. The framework includes five functional layers: risk data, feature calculation, risk model, decision engine, and business access layers. All these layers are loosely coupled, stateless, and scalable. The risk data layer includes the underlying data mart, data labels of various dimensions, and good data management functions. The feature calculation layer is capable of supporting the online calculation of real-time data and offline calculation of bulk data and can realize flexible configuration of functions and cycles through a unified feature calculation scheduling module. The risk model layer is mainly used to deploy machine learning algorithms, such as supervised, unsupervised, semi-supervised, and complex networks. It has the full lifecycle management function of model training, model verification, and model deployment. The decision engine layer completes the configuration and management of various rules. The result of the risk model can be invoked when the rule is executed and the final decision is whether to block, suspend, or alert the transaction. The business access layer mainly completes the docking with various business systems and carries out data collection and matching of various black and gray lists in accordance with the filtering standards for unified risk control elements. In addition, the framework is divided horizontally into production deployment and business operation domains. The core functional modules, related to online transaction processing, belong to the production deployment domain, whereas the modules supporting the parameters, rules, models, features, data, etc., belong to the business operation domain. The production deployment domain should focus on the stability of the platform, and the business operation domain must have a good human–computer interaction interface to ensure that business personnel can flexibly configure related content as needed.

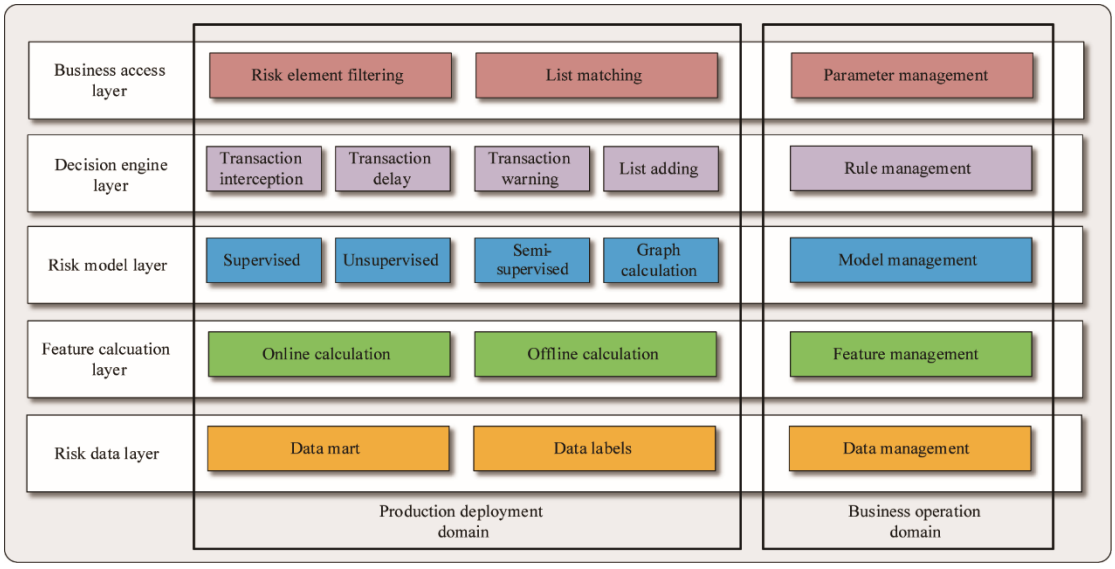

Fig. 1. Overall framework of the intelligent risk control platform based on big data.

5 The realization of function modules for the intelligent risk control platform based on big data

5.1 Risk data layer

The risk data layer is located at the bottom of the platform, and it mainly includes two functional modules, that is, data mart and data label, and two data management modules: data query and data management. The related functional modules are shown in Fig. 2.

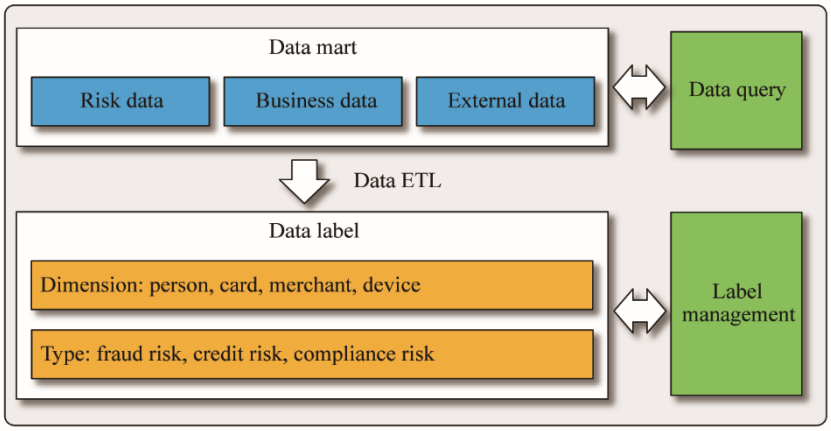

Fig. 2. Risk data layer of the intelligent risk control platform.

Note: ETL is the process of data extraction, transforming, and loading.

5.1.1 Data mart

A data mart can realize the storage and governance of business data, risk data, and external data by building, for example, Hadoop ecosystems. The point of business data is to associate transactions initiated by the same customer from different channels through a unique primary key (e.g., an account number or user ID) to form a complete historical transaction sequence and to solve the problem of data islands from the source. Risk data is the black and gray list accumulated in the daily risk control operation, such as stolen card numbers, merchants suspected of telecom fraud, and cardholders who maliciously refuse to pay. External data is supplementary data obtained through industry sharing in risk control practices, such as positioning data of mobile phone base stations provided by telecom operators, and identity authentication provided by the Ministry of Public Security.

5.1.2 Data labels

Based on the data mart, multidimensional data labels, including person, card, merchant, and device labels, are obtained through data extraction, conversion, and loading. Label classifications are formed according to the main risk types, including fraud, credit, and compliance. Data labels are not static but are constantly extended and expanded as the underlying data are updated. For example, historical data can be linked to mark a merchant with a high degree of crossover as a suspected information leakage merchant, according to the fraud cards of the risk data.

5.1.3 Data management

By connecting to big data query and analysis engines, such as Hive or Impala, business personnel can conveniently query the data stored in the data mart, carry out daily risk investigation and modeling preparation, and can also maintain the data tag library.

5.2 Feature calculation layer

The feature calculation layer is responsible for calculating the online and offline features. The calculated results are uniformly loaded into distributed memory, which is used by the upper model and rules. The relevant functional modules are shown in Fig. 3.

DOI 10.15302/J-SSCAE-2020.06.015

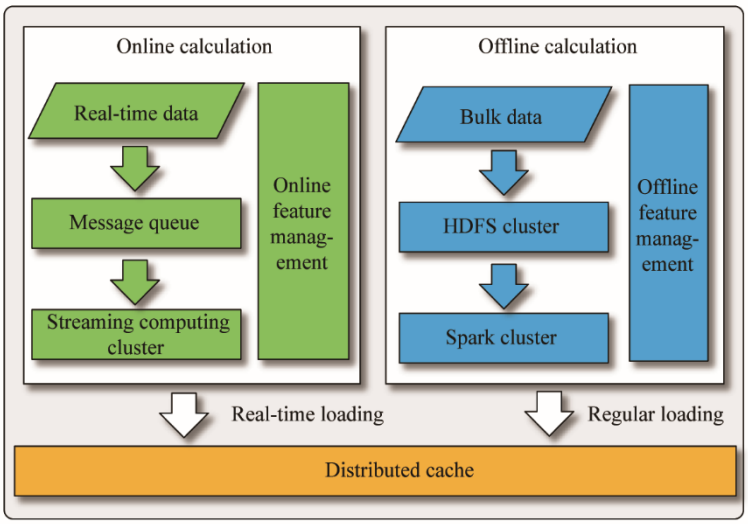

Fig. 3. Feature calculation layer of the intelligent risk control platform.

5.2.1 Online calculation

The online calculation module is usually responsible for calculating the indicator characteristics of the day and is used to describe the current behavior trajectory, such as the number of failed transactions for a card within 30 min, and the number of failed transactions by a merchant within 90 min. Since the transaction data of the day have not been released in the data mart, the source of online calculation comes from the data collected in real time. Through message queue components, such as Kafka, real-time data are continuously fed into the streaming engine, which completes the computation using a sliding window, and subsequently loads the calculation results into the distributed cache. The online feature management module can flexibly configure the calculation objects, calculation functions, time window span, matching conditions, etc., through SQL statements and visual interfaces. After the configuration is completed, it can take effect in real time.

5.2.2 Offline calculation

The offline calculation module is usually responsible for the historical feature calculation before day $T-1$, which is used to characterize long-term behavior, such as the characteristics of 7 days / 30 days / 6 months, the standard deviation of the daily transaction volume by the merchant within 30 days, and the main city where the card transaction occurred within 6 months. Owing to the long period of calculation and the large amount of data, it is necessary to schedule the big data distributed cluster to complete the feature calculation. After the calculation is completed a feature file, that can be loaded into distributed memory by daily timed loading, is generated. The offline feature management module is mainly used to manage big data computing tasks, including computing logic and computing cycles. Because offline features require costly resources, a corresponding monitoring mechanism is required. If the computing time of a certain task is found to be exceeded, 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 calculation task of the machine learning model, but also manages the full lifecycle of the model. Therefore, it mainly includes three main functional modules: model training, model test, and model run. The relevant functional modules are shown in Fig. 4.

5.3.1 Model training

First, model training needs to extract the required data from the risk data layer. In the process of extraction, the card number, mobile phone number, ID number, and other personal sensitive information are masked using the Hash algorithm. Second, data cleaning and feature screening should be performed. Finally, the model is built to complete the algorithm selection and model verification. In addition, to improve modeling efficiency, the modeling environment should support commonly used modeling tools and algorithm packages, such as Python, SAS, and R. After the model indicators (e.g., ROC curve and KS value) reach the expected goal, that is, after the model is finalized, a standard format model file can be generated.

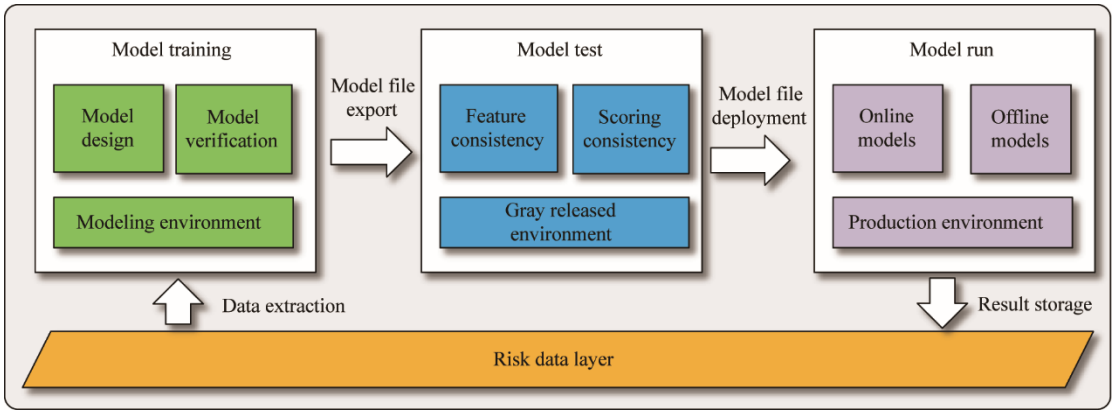

Fig. 4. Risk model layer of the intelligent risk control platform.

5.3.2 Model test

To avoid poor performance of the model, due to factors such as insufficient training samples and differences in offline feature calculations, it is necessary to carry out a model test before the model is officially deployed. The model test is carried out in a gray-release environment, fully simulating the operation of the production environment. The offline feature is also completed by the offline computing module, and the online feature is calculated by the online computing module through parallel shunting 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 involves finding defects in feature screening, and the latter involves discovering defects in the algorithm selection.

5.3.3 Model run

After passing the model test, the model file is deployed in the production environment. During deployment, different operating modes must be configured according to the adaptation scenarios of the model, including online and offline models. Among these, the online model is suitable for real-time anti-fraud scenarios. Each transaction needs to be evaluated in real time, such as anti-telecom fraud scenarios; the offline model is suitable for non-real-time risk control scenarios, such as money laundering network detection and gang cash-out detection. Finally, the results of the model calculation are used by the decision engine layer, which is also recorded synchronously in the risk data layer as a training sample for the continuous iteration of the model.

5.4 Decision engine layer

Based on the triggering of real-time events, the decision engine layer mainly matches the risk control rules and executes decision actions by invoking the feature calculation and risk model layers. The related functional modules are shown in Fig. 5.

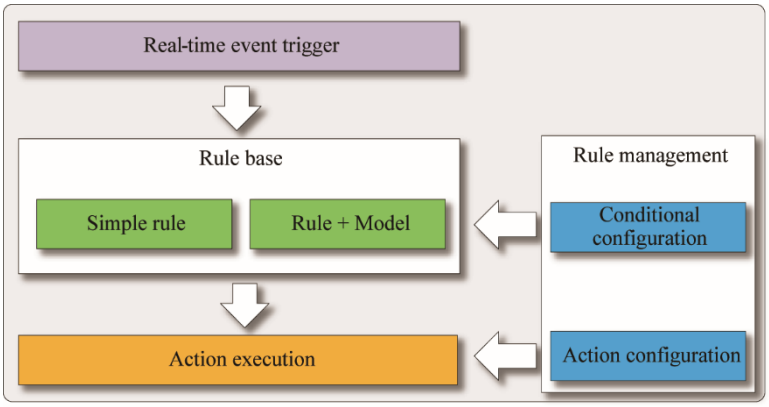

Fig. 5. Decision engine layer of the intelligent risk control platform.

5.4.1 Rule base

The rule base is composed of specific rule conditions, which can be expressed as left operands, right operands, and relational operators. Operands can be abstracted into current transaction elements, previous transaction elements,

DOI 10.15302/J-SSCAE-2020.06.015

online features, offline features, and collections. Relational operators include greater than, equal to, belonging to, not belonging to, containing, not containing, and regular expression matching. The conditions are generally combined through the logical relationship of AND and OR. If the conditions do not refer to the calculation results of the model, then it is a simple rule; otherwise, it is a complex rule.

5.4.2 Action execution

When all the conditions of a rule are met, subsequent risk control actions must be executed. According to the different risk levels of the rules, there are corresponding actions such as transaction interception and delaying and warning of transactions. Interception directly causes the current transaction to fail and is the most stringent intervention. Delaying provides a second confirmation opportunity, and warning does not affect the authority of the current transaction. In addition, a list can be added, namely, the user can automatically add a certain element of the current transaction to the black, white, or gray list.

5.4.3 Rule management

Business personnel can configure the conditions of the rules via the interactive page. The reuse of rule templates should generally be supported to improve the efficiency of rule editing.

5.5 Business access layer

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

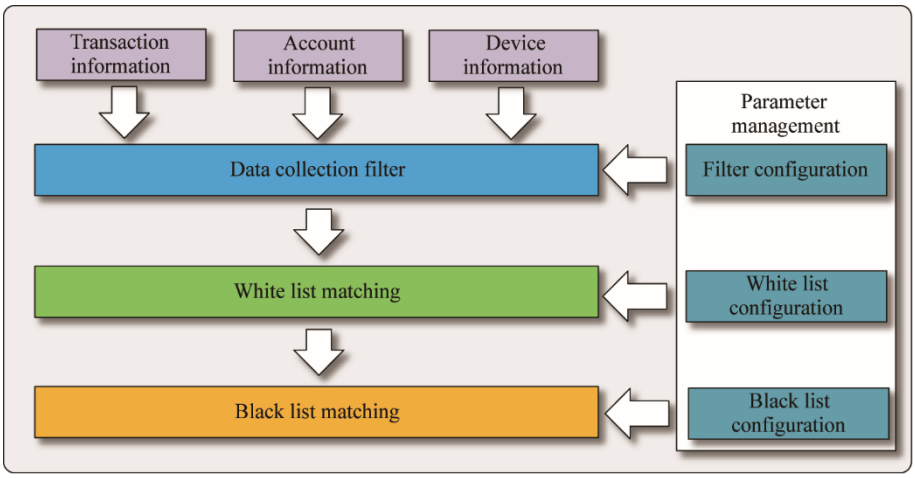

Fig. 6. Business access layer of the intelligent risk control platform.

5.5.1 Risk control element filtering

First, the current risk control scenarios should be clarified, such as login, transfer, and account opening. Next, the filtering of specific risk control elements is completed according to the filter configuration corresponding to the scenario. The risk control elements mainly include the transaction, account, and device information. Among these, the transaction information includes the primary account number, mobile phone number, transaction time, transaction amount, and transaction area. Account information includes account opening time and available quota. Device information includes GPS location and device fingerprint. In addition, while filtering the elements, it is necessary to check the compliance of the field for each element to avoid interference with the judgment of the risk control platform through malicious tampering of messages.

5.5.2 List matching

List matching includes white and black list matching. Hitting the white list, the current transaction is directly released, whereas the situation is reversed when hitting the black list. The list is updated from three aspects: first, it is added automatically based on related actions triggered by rules; second, it is added proactively by business personnel based on investigation feedback; and third, it is added manually based on industry sharing.

6 Application

Let us consider a financial institution that has built several risk control systems to meet the needs of business

development in different periods but the business data of different channels are stored separately; in other words, the risk control systems are not effectively coordinated and integrated. To adapt to the advancements of the current digital era, this financial institution reconstructs and upgrades the existing system to create an industry-level intelligent anti-fraud risk control platform based on the design framework and implementation method of "five layers and two domains" proposed in this study, as shown in Fig. 7.

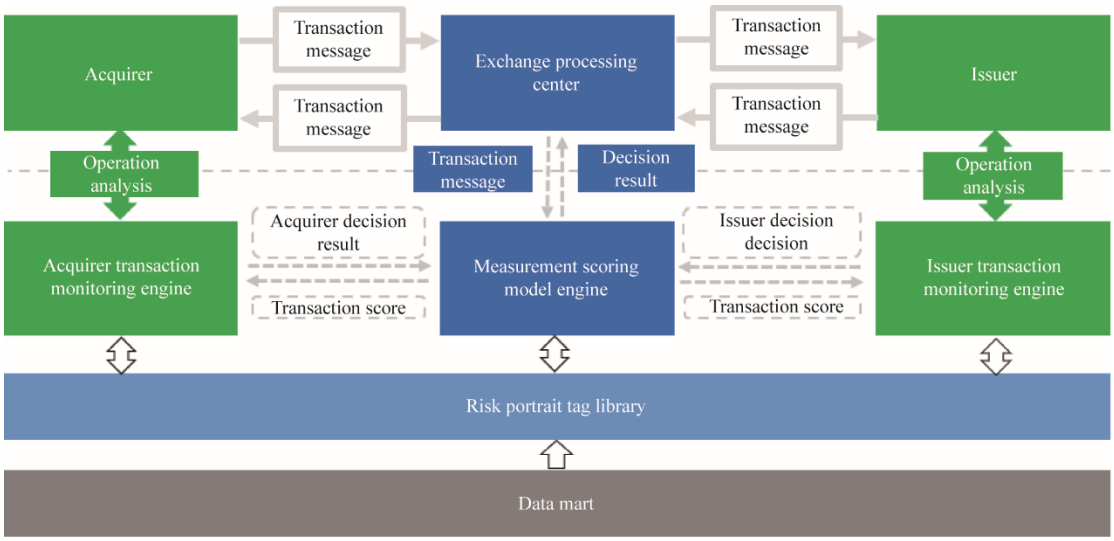

Fig. 7. Application case: industry-level intelligent anti-fraud risk control platform.

In the bottom layer of the platform, a data mart is built for the collection, storage, and management of multi-source heterogeneous data generated from various types of businesses such as account openings, payments, transfers, and cash withdrawals. This data mart is the library comprising risk portrait tags, which are estimated by extracting customer behavior characteristics from their personality dimensions, cards, devices, merchants, etc., through various AI technologies, such as pattern recognition, natural language processing, and complex networks. At present, there are two billion tags, which are increasing at the rate of tens of millions every month.

For each transaction, the exchange processing center sends messages to the measurement scoring model engine in real time. Based on the current transaction elements and historical portrait features, the scoring model completes feature calculation and quantitative risk evaluation of the current transactions within milliseconds through a real-time stream computing engine. The evaluation results are treated as inputs for the acquirer transaction monitoring engine and the issuer transaction monitoring engine. These two engines then complete the decision based on quantitative scoring calculated based on the dimensions of the card and merchant, and then return the decision result to the measurement scoring model engine, which yields the comprehensive result. Next, the exchange processing center implements actions, such as interception, suspension, secondary verification, or release of current transactions, based on the comprehensive decision result, thereby completing the risk control decision of the entire real-time transaction. To reduce the impact on the success rate of the transaction, all decision-making processes are completed within 50 ms. In addition, risk control operators can investigate and process the cards and merchants that are at risk and continuously adjust and optimize features, models, and rule configurations based on the investigation feedback, to ensure that the platform is always running in the best condition.

The platform recovers hundreds of millions of capital losses for the industry each year. It therefore plays a key role in improving the industrial environment and in exporting and enabling cooperative institutions.

7 Conclusion

Faced with a complex and severe risk situation, the financial industry is stepping up the formulation of relevant technical standards. To effectively ensure the application of technical standards, this study started from the three outstanding issues of data islands, computing power bottlenecks, and longer model iteration cycles, which exist in traditional risk control systems. Based on the complete understanding, absorption, and derivation from emerging technologies, such as big data processing, real-time computing, and machine learning, a new design of an intelligent risk control platform was proposed. The proposed platform is a "five layers and two domains" framework based on

DOI 10.15302/J-SSCAE-2020.06.015

"middle platform" ideas. First, an entire closed loop of risk control was constructed considering the following five aspects: risk data, feature calculation, risk model, decision engine, and business access, giving play to the risk control value of financial big data. Second, the layers were made to have a low coupling degree, small dependence, and the applications in the layers mostly adopt a distributed architecture, which is convenient for horizontal expansion. At the same time, related functional modules were specifically implemented and combined from the two dimensions of production deployment and business operation, to maximize the stability of system operation and the flexibility of business applications. Therefore, the design and implementation methods proposed in this study can better meet the risk control needs, thus supporting the business transformation and high-quality development of commercial banks in the era of digital economy.

The potential for the development of intelligent risk control for big data is endless. As next-generation technologies, such as 5G, Internet of Things, blockchain, and AI mature, it is expected that finance and technology will integrate further. The intelligent risk control platform has a broader development space in the entire chain of customer acquisition, credit granting, anti-fraud, and marketing. For the application of the intelligent risk control platform, this study puts forward the following suggestions based on the existing problems and current situation.

(1) Emphasize on the bottom line of compliance. The rapid development of big data technology has given rise to hidden dangers such as data abuse and information leakage, which cannot be ignored. Companies must regard compliance development as the red line for survival and develop businesses in compliance with laws, regulations, and regulatory requirements. In the process of data collection, storage, and sharing, the requirements of supervision, privacy protection, and security should be considered to ensure that the acquisition of customer data is reasonable, compliant, and legal.

(2) Adhere to technology. Technology forms the foundation of the industry of intelligent risk control based on big data. It is therefore necessary to strengthen the research and application of intelligent risk control, giving full play to the advantages of machine learning, complex networks, blockchain, cloud computing, and other cutting-edge technologies. Companies should iteratively upgrade algorithms and models to better empower the financial industry in accordance with the changes in the economy, society, scenarios, and users.

(3) Remain focused on business orientation. Business development is the fundamental goal of an enterprise. Risk control is a means of achieving and guaranteeing business development. Therefore, the construction of an intelligent risk control platform should be based on current business needs and be unified with the development goals of the enterprise to achieve the coordination of business and risk control as well as long-term development.

References

[1] Liu G. Practice of intelligent risk control system construction in the era of big data [J]. China Financial Computer, 2018, 349(8): 17–20. Chinese.

[2] Chen X. Intelligent risk control system based on deep learning [D]. Beijing: Beijing Jiaotong University(Master's thesis), 2019. Chinese.

[3] Ding S B. Research and design of security risk control platform based on SOA [D]. Xi'an: Xidian University(Master's thesis), 2018. Chinese.

[4] 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. Chinese.

[5] Guo R. Research on T Company's big data risk control platform [D]. Nanjing: Nanjing University(Master`s thesis), 2016. Chinese.

[6] Wang X. Practice of mobile financial risk control system construction based on artificial intelligence [J]. Information Security Research, 2017, 3(11): 1000–1005. Chinese.

[7] Fight resolutely to prevent and defuse major risks [J]. Practice, 2018 (6): 18–19. Chinese.

[8] Liu R X. Building an intelligent risk control system based on big data [J]. Electronic Finance, 2018 (8): 57–58. Chinese.

[9] 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. Chinese.

[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] 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. Chinese.

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

[14] 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. Chinese.

[15] 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. Chinese.

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

[17] Wu C, Wang X N, Xiao H L, et al. Research review of distributed message system [J]. Computer Science, 2019 (S1): 1–5. Chinese.

[18] Huang J H. Design and implementation of Redis [M]. Beijing: China Machine Press, 2014. Chinese.

[19] 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. Chinese.