https://doi.org/10.26599/IJCS.2022.9100032 International Journal of Crowd Science

A Correlation Analysis Based Risk Warning Service for Cross-Border Trading

Anting Zhang[1], Bin Wu[2], and Yinsheng Li[2] ✉

ABSTRACT

Obtaining a high-precision risk warning service, which can improve trading efficiency and legality, by reducing sampling proportion and customs clearance time dramatically is critical for cross-border trades. However, existing anti-fraudulent services are weak either in the precision or the mining capacity of discovering hidden risks. Among the reasons are incomplete data, untrustworthy resources, and old analysis models. On the basis of these observations, this article makes a combined technical solution for a risk warning service to address data resource, integrity, and mining capacity issues. The provided risk warning service is featured with a correlation analysis approach, which is advanced and efficient at addressing multisource and heterogeneous data to identify deep-seated risks with cross-border products, such as fake documents, price concealment, epidemic events, and ingredient pollution. To reveal the hidden correlation risks in cross-border clearance, a set of correlation-oriented data models and multi-attribute, multi-object, and multi-level methods are developed. The involved data sources and objects can be collected from inside businesses and public resources. Data are further structured to depict the whole portrait of a trade. The correlation analysis approach proves to be feasible and efficient in processing multisource and heterogeneous data to discover deep-seated risks with cross-border products. The risk warning service and the used correlation analysis approach have been studied and developed on the basis of a pilot project at an exit-and-entry port in Shanghai.

KEYWORDS

risk warning; correlation analysis; cross-border trading

The new entry-exit inspection and quarantine policy issued by the General Administration of Quality Supervision, Inspection and Quarantine calls for a significant reduction in sampling proportion and customs clearance time. Cross-border trades increase calls for scientific and intelligent supervision. A precise risk warning service is critical for checking cross-border products.

The potential risk identification methods currently used in customs clearance, such as risk rule based concealment, inspector experience, and enterprise credit system in customs, are all coarse-grained. Most research[1–4] in risk identification and prediction at ports uses machine learning or deep learning methods based on historical inspection data. However, this kind of method is inapplicable in the port scenario. Analyzing some critical risk types in ports, we find that identifying many risks is difficult due to limited information and features learned from inspection data. Risk warning services nowadays lack supporting data and efficient evaluation models.

Su et al.[5] pointed out that the impacts of interpersonal relationships (correlated risk analysis) are more reliable than those of individual behaviors (traditional credit risk analyses, such as credit score). Through correlation analysis on multisource and heterogeneous data collected from multiple links and enterprises, deep-seated risks in cross-border products can be identified. We define specific data models and identification methods to reveal multiple correlated risks. Hidden risks are identified by correlations between attributes, objects, and hidden relationships. Dynamic control strategies are designed with good interpretability for government departments according to risk types and levels.

As an unusual production scenario spanning many countries, enterprises and governments at ports cooperate but distrust each other. Government mediation dependency, information fragmentation, and inefficient collaboration among enterprises lead to incomplete risk data. Facing the issue of collecting sensitive data from multiple participants who lack trust, blockchain and smart contract technologies have brought opportunities to change production relationships, which make the development of a credible, autonomous system possible, depending on the characteristics of non-tampered data and trustworthy code execution. Risk data are collected from internal sources, including invoices, packing lists, and inspection and quarantine declaration data and results, and external sources, including epidemic announcements and consensus information.

The aim of the service is to provide a precise risk warning for cross-border products. In comparison with other techniques, the correlation analysis based service has the advantage of identifying hidden correlated risks with good interpretability and scalability. Port inspection is an important part of the cross-border trade process. The proposed risk warning service provides interfaces for the Trade Finance Open Platform launched by the Digital Currency Research Institute of the People's Bank of China.

The remainder of this paper is organized in the following manner. Section 1 cites previous related works. Section 2 presents the scenarios, objectives, and features of the risk warning approach. Section 3 describes the design of the risk warning service. Section 4 demonstrates the evaluation and provides the

1 Department of Automation, Tsinghua University, Beijing 100084, China
2 School of Computer Science, Fudan University, Shanghai 200433, China
Address correspondence to Yinsheng Li, liys@fudan.edu.cn

results and discussion. Section 5 concludes.

1 Literature

1.1 Literature on risk warning for cross-border clearance

Studies and explorations on product management modes have been carried out early in the inspection and quarantine system of China. At the government level, the development of risk warning methods is slow. Dai[6] showed that the government inspects every batch of products in the category list first. Yuan and Xu[7] pointed out that “classified management” first appeared in 1984, which includes three parts: examination and confirmation of qualified product types, sample inspection of export batches, and supervision of factory quality systems. Current methods mainly focus on classification and rule systems, which are coarse and inaccurate. The “Shanghai Free Trade Port Construction Plan” highlights that the management mode has changed from mainly managing products to enterprises. The General Administration of Quality Supervision, Inspection and Quarantine shows that the current evaluation of enterprises considers indicators such as credit status, management level, and business ability. This evaluation demonstrates the contradiction between the coarse-grained characteristics of existing models and the need for accurate risk warnings in the government.

At the research level, most studies use machine learning or deep learning based on historical customs data, including a dynamic K-means clustering algorithm on historical customs declaration data in Ref. [1], a deep learning method based on civil aviation and public security databases in Ref. [2], a classification analysis on historical customs declaration data in Ref. [3], outlier mining on import and export data in Ref. [4], and a blockchain-based production credit mechanism for manufacturing services in Ref. [8]. The shortcomings of the current approaches lie mainly in the single data source used and the lack of consideration of correlations among products. Identifying product risks on the basis of limited data is difficult.

The evaluation methods in the current clearance facing classified management and enterprise credit are coarse-grained and inaccurate. Plenty of works using machine learning or deep learning have been performed on the basis of historical customs data to improve the risk warning accuracy of the traditional information system. However, these methods can only solve the risk types that have existing labels, and the identification accuracy is not high due to the limited support features obtained from the data.

1.2 Literature on correlation algorithms

Correlation algorithms on risks have been widely used in fields such as credit and investment. Su et al.[5] argued that the impacts of interpersonal relationships (correlated risk analysis) are more reliable than those of individual behaviors (traditional credit risk analyses, such as credit score). Some studies have used correlations among data for risk identification. Kim et al.[9] presented a risk-based categorization for each country—commodity combination on the basis of the probabilities of quarantine pest interceptions and the uncertainties in that assessment. Zhou et al.[10] proposed a two-layer neural network model, which solves the problem that multi-attribute and low correlation data accumulation affects risk calculation. Qian[11] carried out a series of studies, including game theory, infectious epidemic model, and complex network theory, on the contagious effect of related credit risk. Based on these factors, an effective risk control strategy is put forward.

The use of correlation algorithms in port scenarios, which have multiple objects and levels, contributes to the identification of potential risks. However, these methods cannot access and use data from different sources or units and, therefore, cannot take advantage of correlations among multiple data sources. Such advantages include active access to external data and cooperative access to other units in the product supply chain.

1.3 Problem

Risks in ports are hidden in complex relationships among products and enterprises. The current supervision mode facing a single product is inaccurate due to limited aspects of identifying risks. The government can only obtain limited data from limited participants in the supply chain. The mediation dependency on custom brokers leads to incomplete and untrustworthy data. Thus, using sufficient risk evaluation methods and models becomes further difficult. The main issues of the current system are as follows:

Inaccurate risk evaluation. The current risk product evaluation mainly relies on some rules based on product category, inspector experience, and enterprise credit rating. Enterprise evaluation mainly focuses on enterprise macro situations, including credit status, management level, and business ability. Evaluation methods are coarse-grained and ignore the specific information of each batch of products. Many studies have used historical customs data to improve risk warning accuracy. However, these methods can only solve risk types that have existing labels, and the identification accuracy is not high due to the limited support features obtained from the data. Potential risks are difficult to be identified due to the separation of various business links, participants, and products.

Incomplete and untrustworthy risk data. An accurate risk evaluation is based on sufficient information. In the current working mode, multiple links and participants are involved in the supply chain of cross-border products. Participants include not only producers, consignees, and the government. Multiple intermediate service providers, such as logistics service providers, freight companies, tallying warehouses, agents, and customs declaration enterprises, are involved in the process. However, the government mainly obtains limited data from custom brokers. At present, the data uploaded to the system only include basic declaration information, such as harmonized system (HS) code, component, country of origin, and total product value. Critical information related to the process is uncollected and unused, such as traceability, epidemic, and consensus information.

The government data source is quite single. Intermediate service providers can deceive upstream or downstream participants. The low-trusted data transmission and inefficient process collaboration lead to untrustworthy data.

A platform that includes all services is usually good at solving data fragmentation. Traditionally, a centralized platform is always controlled by one institution that has more power than general participants. However, it is unsuitable in this case because enterprises involved in cross-border products often span at least two different countries. Neither institution is willing to give control of sensitive business data to other institutions.

2 Correlation Analysis Approach for Risk Warning

The risk warning in the current situation makes government supervision insufficient under the following scenarios. The proposed service is to solve issues and improve supervision accuracy and efficiency.

2.1 Scenario

We take some milk products as examples shown in Fig. 1 to illustrate risks with cross-border products. According to the domain knowledge of experts and analysis of historical events at ports, we list six critical risk types as follows:

- **Fake documents.** The so-called "wash the document" refers to the collusion between smugglers and illegal overseas forwards. When products are re-exported into a country, the forwards conceal the original documents and issue false documents in the name of the forwards to the smugglers for use in the domestic customs declaration. Correlations on documents cannot be used nowadays due to the lack of original documents.
- **Unqualified products in affiliated enterprises.** When milk powder A has problems such as unqualified quarantine, fresh milk A, which is from the same consignee A, must be quarantined in appropriate proportion. The product risk is strongly related to its enterprise. Different products from the same producer or consignee are supposed to have correlations.
- **Cooperative enterprise dishonesty.** Shipping company A has a long-term close cooperative relationship with customs broker B. When shipping company A is found to be dishonest from internal business processes, government announcements, or external consensuses, the credit on customs broker B should decrease, which affects the products handled by the company to some extent. Hidden relationships among enterprises are unused effectively.
- **Ingredient pollution.** The products of producer B have a pollution issue. The raw material of milk powder B is from producer B. A correlation exists among risks in the upstream and downstream of the production chain.
- **Epidemic events.** The epidemic disease has transmissibility. When animals from one pasture have epidemic diseases, the pastures from the same country are likely to be affected.
- **Price concealment.** The same products will be transported to different ports for regional reasons. Inherent product properties, such as price and raw materials, should be the same or similar. Correlations on data from different ports contribute to concealment identification.

2.2 Objective

For a precise risk warning, the service needs to meet the following requirements:

(1) **Correlated risk identification.** Compared with the current research on discovering intuitive risks, the service needs to discover multiple types of hidden risks with cross-border products.

(2) **Interpretability.** Government service has its particularity. At present, machine or deep learning based methods are often poorly interpretable and cannot convince the government staff. The risk warning service must provide reasons for inspection or quarantine.

(3) **Scalability.** The government and law-breakers are continually competing and growing. Risk types are changing day by day. The risk warning service should thus adapt to new risk types quickly, conveniently, and accurately.

2.3 Principles and features

The correlation analysis based risk warning service is proposed to solve these issues. The service can identify hidden risks on the basis of the correlations in multiple aspects. Sufficient information and features are extracted from multisource and heterogeneous data closely related to products. Data covering multiple links and enterprises are collected with the help of changed production relationships by blockchain. The risk warning service effectively improves the accuracy of early warning for cross-border products. System features are as follows:

Multisource and heterogeneous data

Multisource and heterogeneous data are collected as the fundamental of the risk warning service. To identify risks accurately, rich information is extracted from data collected from the whole participants and links related to cross-border products. Benefiting from blockchain technology, we can change the production relationship and obtain private and sensitive business data from enterprises.

Correlation analysis

Identifying product risks on the basis of their own data is too limited. Applying the correlation analysis on multisource and heterogeneous data, deep-seated risks in cross-border products can be identified. The correlation analysis is applied to multiple

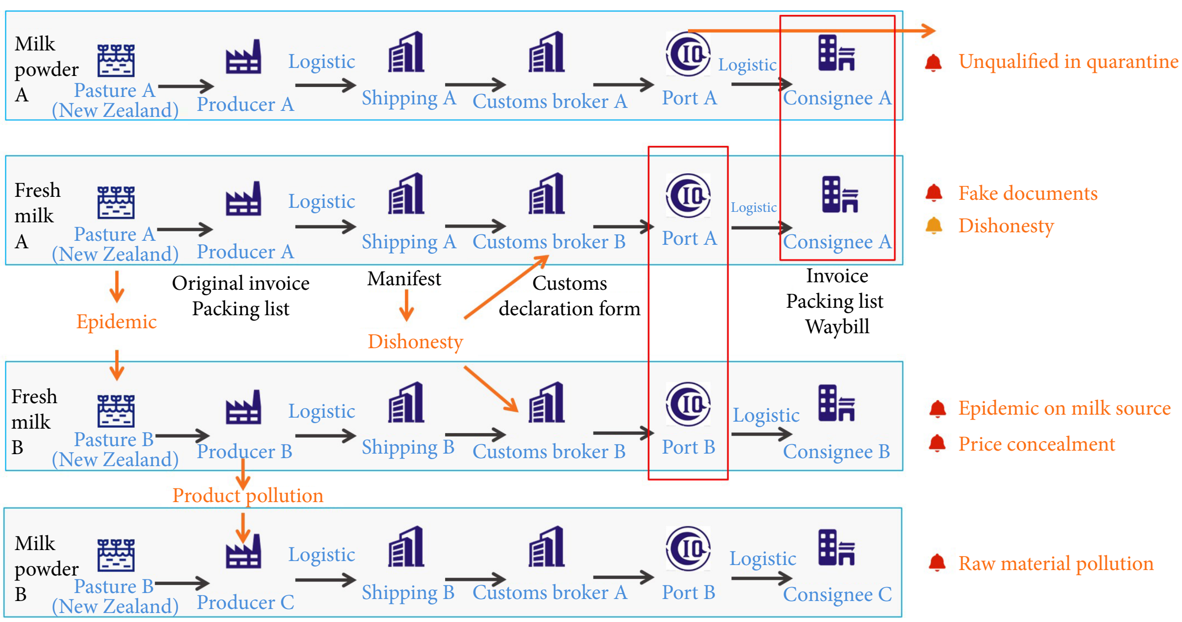

Fig. 1 Correlation analysis on a specific example. Use of correlation analysis in port scenarios with some milk products as examples.

aspects, including documents, products, enterprises, ingredients, epidemics, and ports.

Dynamic control

Benefiting from the interpretability of the correlation analysis, risk evaluation results provide not only risk probabilities or degrees but also specific reasons to prompt the government. Giving risk scores or tips on products is not the end of the risk warning service. To achieve effective supervision, designing different control strategies in each government department for different risk types and levels is necessary.

3 Correlation Analysis Based Risk Warning Service

We present the architecture of the risk warning service in Fig. 2. Heterogeneous data are acquired by different methods, including business clients and image processing methods, such as optical character recognition (OCR), web crawler, and application programming interface (API). The business process starts from the producers, through several intermediate service providers, such as warehousing and customs declaration enterprises, to the inspection and quarantine department, and finally reaches the consignees, with the logistics service providers involved in the process. The batch is used as the tracing unit, and essential data and files, such as qualification certificates, inspection reports, and waybills, are associated with the tracing unit based on business logic. Products are traced by batch through the close association of data carriers in each link.

Information related to each product link is stored in the blockchain. The risk warning service obtains the required data from the blockchain. It sends out risk warnings to relevant inspection and quarantine departments when risks are detected.

According to the architecture, blockchain system participants are limited to specific institutions. We choose the permissioned chain, which is limited to alliance member participants. Blockchain permissions are formulated according to alliance rules. Different system participants are in equal positions. Nobody controls the system by themselves. Furthermore, we apply permission control on the basis of their identities so that they can control their data.

3.1 Data processing

Data sources

To collect sufficient data, data sources of the risk warning service include internal and external sources. Internal data are mainly from the business process. The blockchain provides an equal environment that eliminates concerns about sharing sensitive and private business data. Data sources should cover the whole product supply chain. According to critical supply chain processes, we sort out the main participants.

- **Producer.** The producer is the provider of products and the starting point of the product flow, which is usually in a different country from the consignee. Data from the producer mainly contain producer subject information, such as name, country, and address, and order information, including product name, brand, composition, HS code, and price.
- **Consignee.** The consignee is the receiver of products and the endpoint of the product flow. Data from the consignee mainly contain the consignee's subject information, such as name, country, and address, and receiving information, such as quantity.
- **Intermediate service provider (ISP).** ISPs connect producers and consignees, which play critical roles in the product flow. ISPs include logistics service providers, freight forwarders, and customs brokers. The data contain information about the transport enterprise, loading time, and quantity information in transport; the name and address of the forwarding company; and the category, code, and price information of customs declaration.
- **Government departments.** There are several departments responsible for different types of products and laboratories responsible for quarantine. The data include airport arrival time, inspection department information, inspection and quarantine certificate, and sample inspection result.

External data are mainly from websites and other government institutions, including the following sources.

- **Government announcements.** The websites of customs or the inspection and quarantine departments of domestic and

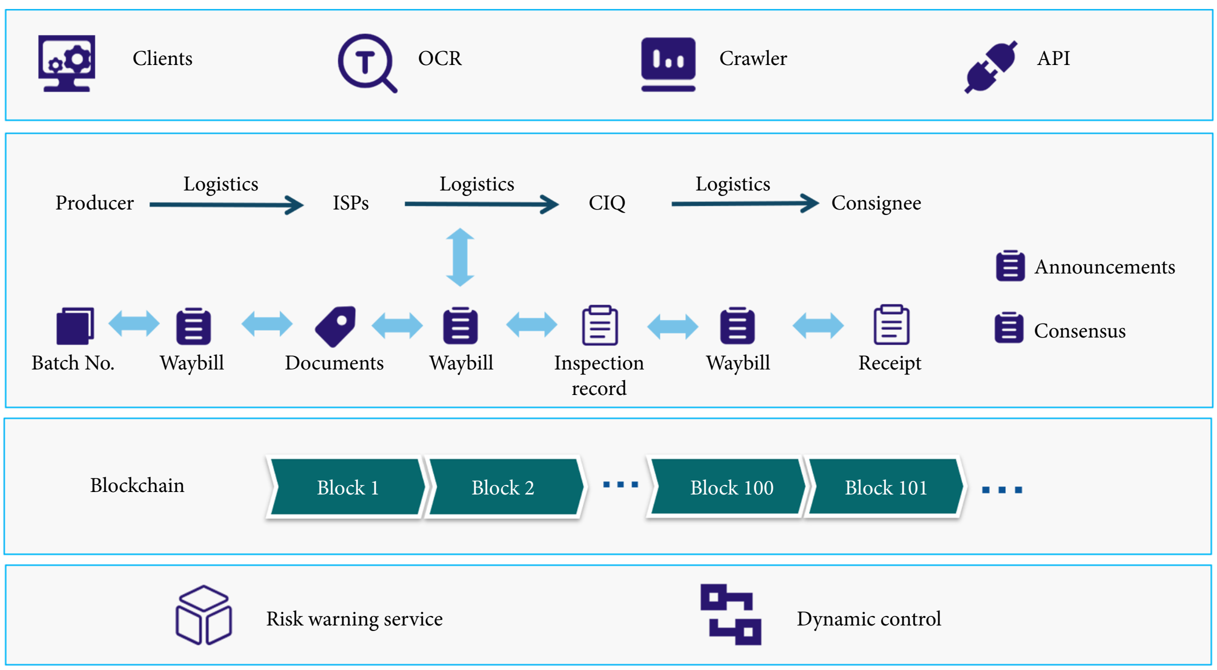

Fig. 2 Architecture of risk warning service, mainly including four parts: data acquisition, data model, data storage, and risk warning services with dynamic control.

foreign countries publish epidemic situations and other announcements. This part contains the inspection and quarantine information issued by the General Administration of Customs; foreign information, such as warnings issued by the US Consumer Product Safety Commission and the US Food and Drug Administration; and recall information issued by the EU Rapid Alert System for Food and Feed and the Rapid Alert System for Non-Food Consumer Products. Credit China and other websites show the dishonest behaviors of enterprises. This part contains enterprise reputation data, illegal records, and business conditions. Obtaining and processing such information in time is the key to early warning.

- **Consensus.** Contents about products or enterprises released by news media or online communities reflect the potential risks and their influence degree levels from the side. This part mainly includes feedback information involving product quality and safety and epidemic released through micro-blogging and other social media. Consumers provide product quality and safety, epidemic, and other risk information through telephone calls, networks, and letters.
- **Other government departments or third-party institutions.** Through cooperation with other government departments, such as industry and commerce, quality supervision, and public security or third-party institutions, such as inspection certification agencies, the basis for risk identification has been greatly increased. This part includes risk information related to inspection and quarantine from customs, quality supervision, commerce, industry and commerce, health, and other ministries; conformity assessment data from inspection and testing certification organizations; and product injury accident information from hospitals, fire, and forensic institutions.

Data acquisition

According to multiple data sources, data are heterogeneous with different formats and media. Internal data include invoices, packing lists uploaded by producers, manifests uploaded by transportation companies, waybills uploaded by consignees, declaration forms uploaded by customs brokers, and inspection and quarantine results. External data include epidemic announcements provided by domestic and foreign government websites, consensus information in news media, and conformity assessment data provided by inspection certification agencies.

The acquisition method includes acquiring data uploaded by users through clients, acquiring image-based declaration data through intelligent image processing methods, collecting consensus and announcement data through network crawlers, and acquiring data of relevant institutions through business database APIs.

Risks in cross-border trades are closely related to product types. Different product types are involved in different risks. An example is the 44 669 reported records of apparel products and 67 102 jewelry products in 2017. For apparel types, the current inspection of unqualified items is mainly based on composition and clothing style. For jewelry types, the inspection of unqualified items focuses on metal content that exceeds the standard. For food types, unqualified quarantine and other phenomena are involved. The correlation analysis of risks must consider the various levels of correlation and be flexible enough to accommodate the different risk types that exist in different product types.

Data model

The correlation analysis of risk data is oriented to different objects and levels. We define five objects in the application scenario at different levels. We provide the specific meanings of these objects according to the level from low to high.

(1) Batch (B). The batch is the object we decide whether to inspect and quarantine.

(2) Product (P). A product contains many batches. The producer and consignee of a product are often fixed, but the ISPs are not. Considering that the batch is too specific, we often adjust the inspection and quarantine proportion at the product level.

(3) Enterprise (E). Enterprises are those responsible for handling every batch of products, such as shipping companies A and B and customs brokers A and B in Fig. 1.

(4) Link (L). The production chain includes many links, from production, transportation, customs declaration, inspection, and quarantine to receiving. Each link contains many enterprises.

(5) Port (T). The port refers to government departments in different regions. Different ports have their own set of the above objects.

We also define Risk (R) as the collection of objects identified to have risks. For example, $R(B)$ means the collection of batches identified to have risks. For the correlation analysis, we define a relationship between two objects: "$O_1 \circ O_2$". When the level of O_1 is lower than that of O_2, it means that O_1 belongs to O_2. When the level of O_1 is higher than that of O_2, it means that O_1 contains O_2. For example, "$B_1 \circ P_1 \circ E_1 \circ L_1 \circ E_2 \circ P_2 \circ B_2$" means that B_2 has a correlation with B_1 in link L_1.

3.2 Correlation analysis

In Section 2, we take some milk products as examples to illustrate potential risks in ports. We provide specific methods for these scenarios as follows:

- **Document correlation.** Data items such as name, package quantity, unit price, currency value, quantity, packing type, import customs, the trading country on original invoices, packing lists uploaded by producers, manifest uploaded by transportation companies, and waybills uploaded by consignees are cross-validated to check whether a risk of false documents exists. A batch can be represented as follows:

$$B = (E_1, E_2, \ldots, E_n).$$

Data cross-validation is performed between E_i.property and E_j.property on this batch, where $1 \leqslant i, j \leqslant n$.

- **Product correlation.** According to the product risks found during on-site inspection or laboratory quarantine, it carries out risk warnings for other enterprise products. The products needed to be warned are

$$\{P_h | B \circ P \circ E \circ P_h, B \in R(B)\}.$$

- **Enterprise correlation.** When a risk exists in an enterprise, such as public opinion (e.g., dishonesty), a risk warning is given to the enterprise closely cooperating with another enterprise. Association algorithms, such as the Apriori algorithm or clustering and community detection algorithms, should be used on historical batches to find enterprises close to the enterprise with risk. The target enterprises are

$$\{E_h | E \circ E_h, d(E, E_h) < \xi, E \in R(E)\},$$

where d is the distance variable representing the close degree of enterprise cooperation, and ξ is a constant.

- **Ingredient correlation.** When the raw materials of products have problems such as pollution, a risk warning should be given to the products using the same raw materials or from the same production place. The target products are

$$\{P_h | B \circ P \circ E \circ P_h, E.\text{link} = \text{Pasture}, B \in R(B)\}.$$

- **Epidemic correlation.** In the case of an epidemic situation in a country or a region, a risk warning must be carried out for products from the relevant production place. The target products are

$$\{P_h | E_1 \circ E_2 \circ P_h, E_1.\text{link} = E_2.\text{link} = \text{Pasture}, E_1.\text{country} = E_2.\text{country}, E_1 \in R(E)\}.$$

- **Port correlation.** Comparing the price differences of products declared at different ports, a risk warning should be given to these concealed products. Data cross-validation is performed between P_i.property and P_j.property, where P_i and P_j refer to the same product in different ports.

Analyzing the six listed critical correlation types, we conclude correlation analysis types as follows:

(1) Based on the correlation among self-owned attributes. The document correlation is in the scope of this correlation. Through the correlation analysis of the data extracted from multisource documents that should be the same or similar, the risk warning service finds concealment in products.

(2) Based on the correlation among multiple objects. Product, ingredient, epidemic, and port correlations are in the scope of this correlation. Through the correlation analysis on different objects or levels and through the risk transmission path, the hidden risks in products are identified.

(3) Based on the correlation between hidden relationships. The enterprise correlation is in the scope of this correlation. The correlation between products and their enterprises is easy to catch. However, many hidden correlations between enterprise and product levels exist in the dataset, such as the previously mentioned close cooperation enterprises.

3.3 Dynamic control

Risk result

For every correlation risk, the result is stored in a triplet {Tag, Score, Msg}.

- The Tag stands for the sign that products must be inspected when it equals 1. In some cases, for example, when a product is declared for the first time, this batch must be inspected. We set this Tag to prevent indicator weights from influencing the results in some particular cases.
- The Score stands for the risk degree. When the Tag equals 0, the Score defines the risk degree of the batch. It is similar to most models.
- The Msg stands for the message that is shown to inspectors. Most evaluation models only provide a binary classification. To not lose correlation analysis information, the Msg defines the risk content description used to assist inspectors in their judgments.

The final triple of the risk warning service is

$$\{\sum \text{Tag} > 0 ? 1 : 0, \sum W \cdot \text{Score}, \text{Msg}[]\}.$$

where W represents the weights of different identified correlation risks. W can be obtained on the basis of prior knowledge or training data. The final Msg of the service contains all messages given by every risk result.

Department dynamic control

Dynamic control is oriented to departments within the government. In the current administration of inspection and quarantine, the government mainly includes four departments:

- **First-class warehouse.** Products are first transmitted here. It is the starting point of products in the government. Products are sorted by type here.
- **Customs declaration.** It is the department to which enterprises submit customs declaration documents. It is also the department where our risk warning service is involved and decides whether to inspect products.
- **Inspection station.** It is the place for on-site inspection. Unqualified products are allowed for rectifications. Qualified products after rectifications will be in the rapid release. Meanwhile, unqualified products will be destroyed or returned.
- **Laboratory.** It is the place for quarantine. The process here is similar to that in the inspection station.

The specific dynamic control flow is illustrated in Fig. 3. The risk warning service gives out risk notifications to other government institutions through interfaces, risk tips on first-class warehouses, and direct interventions on the inspection and quarantine proportion in the inspection station and laboratory. The risk warning service also works with random and immediate controls. Random control is applied to prevent enterprises with high credits and products with low risks from being inspected for a long time. Immediate control is triggered by on-site staff in

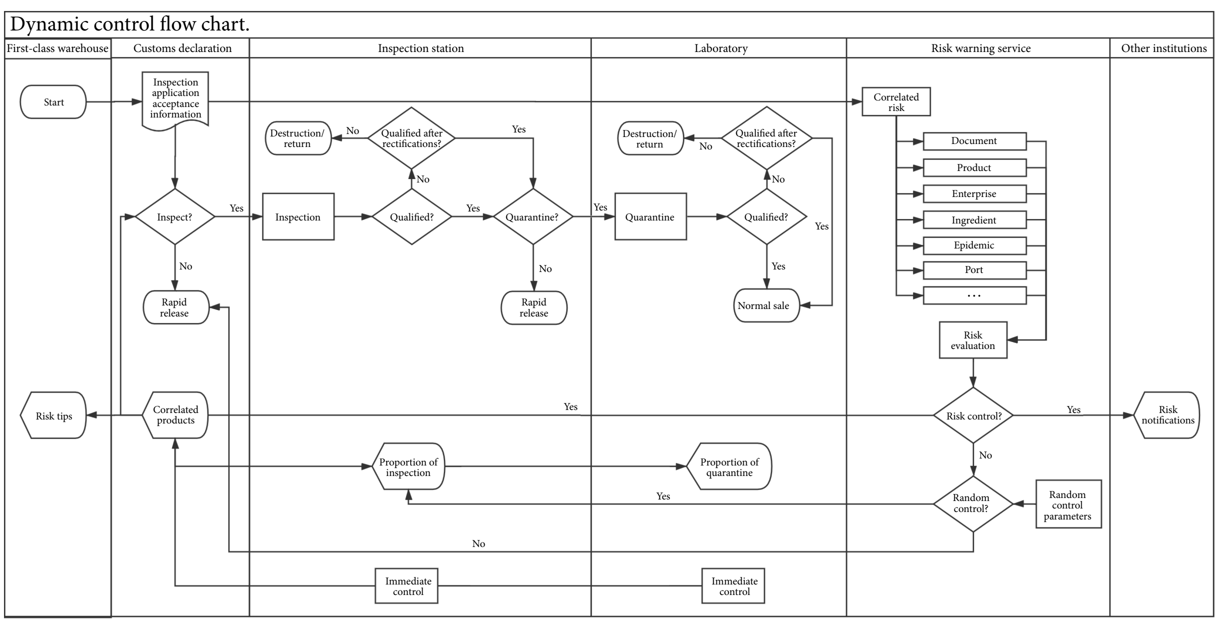

Fig. 3 Dynamic control strategies for different departments in the government.

some particular cases.

3.4 System implementation

We use the software architecture shown in Fig. 4 to realize the risk warning service for cross-border products. The main modules include the data storage layer, data interaction layer, business service layer, and application interaction layer, including privilege management and logging. The main functions of each module are as follows:

- **Data storage layer.** Due to the differences in system functional requirements and database characteristics, data are stored in two different databases. Oracle is a relational database that stores data items related to various businesses. The blockchain stores important data items and provides functions, such as distributed shared ledger maintenance, state database maintenance, and smart contract full life cycle management.
- **Data interaction layer.** It is responsible for database access and operation, specifically providing data services for the business service layer. The blockchain provides the interface for the upper layer through the restful API.
- **Business service layer.** It is the back-end service of the application, which is responsible for processing business requests from the front end. It is realized through Spring Framework. Key functions, including risk evaluation and data cross-validation, are realized through smart contracts.
- **Application interaction layer.** It provides user interaction interface operations, including user and business operation functions.
- **Privilege management.** It provides functions such as member registration, member cancellation, and authority control for different blockchain system participants. It is realized through JSON web token (JWT) and the digital certificate in the public key infrastructure (PKI) system.
- **Logging.** It records user operations, system operation statuses, and error data.

4 Evaluation

In this section, we evaluate the correlation analysis based risk warning service by comparing it with other risk warning methods for cross-border products.

4.1 Evaluation design

For a precise risk warning, related indicators are proposed to evaluate the effect of the risk warning service. Given that the service uses data not accumulated at ports, traditional classification indicators are unsuitable.

(1) **Correlated risk identification.** The impacts of interpersonal relationships are more reliable than individual behaviors. Risk warning methods are supposed to find deep-seated risks in cross-border products.

(2) **Interpretability.** The early risk warning service is intended for government personnel in inspection and quarantine departments. Only with good interpretability can the inspection basis be clarified, making it easy to be accepted and judged by government personnel.

(3) **Scalability.** The risk warning model must be able to expand to new risk types quickly.

We choose some risk warning methods based on machine learning or deep learning for comparison, including Xiao[1] who used a dynamic K-means clustering algorithm, Feng and Huang[2] who employed a feature deep learning method, Zhou and Zhang[3] who conducted a classification analysis, and Yu[4] who performed outlier mining. We choose the correlation-based risk warning method proposed by Kim et al.[9] A pilot project of Tianjin-Port shown in Ref. [12], using blockchain technology to link source data and conduct cross-validation, is also selected.

4.2 Result and analysis

Due to the limited attributes of machine learning and deep learning technologies, the risk warning based on these technologies needs further data accumulation. The current applications of these methods for cross-border products are mainly based on historical data kept in customs. In view of the special and complex cross-border scene at the port, supporting the precise risk warning only by using the data from the original information system is difficult. Correlated risks are ignored in these methods. Kim et al.[9] found a correlation between country

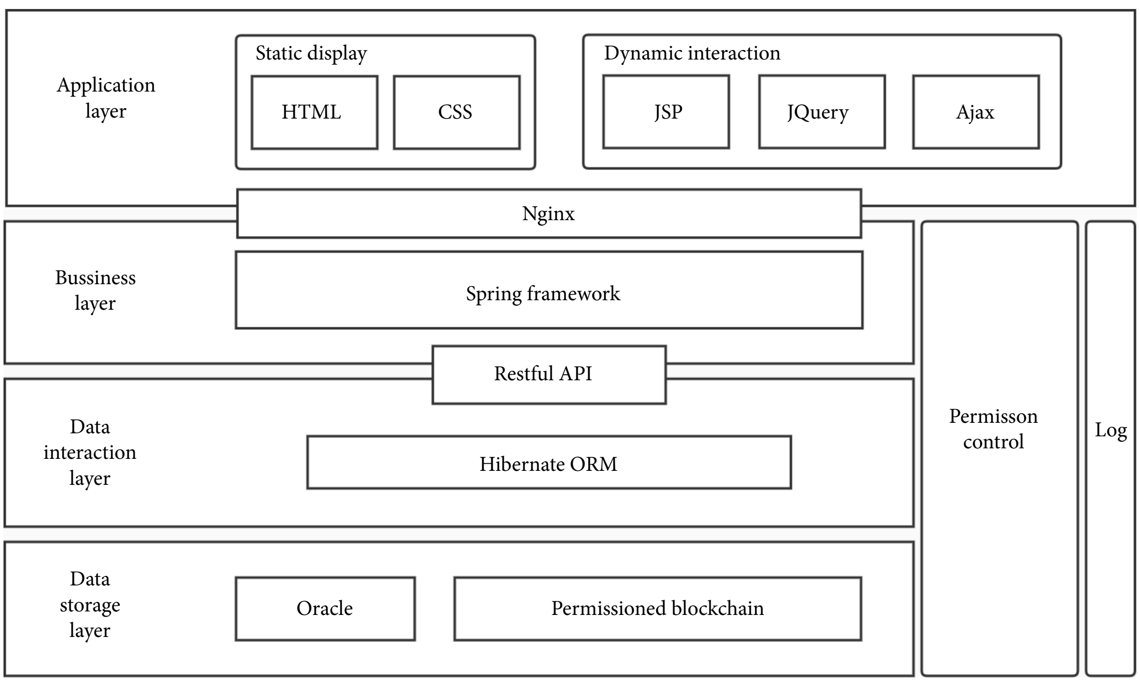

Fig. 4 System implementation of the risk warning service.

and commodity by constructing pairs. The pilot project in Tianjin Port checks eight data items in documents from different sources to ensure data authenticity. Our method can identify correlated risks among object attributes, among objects, and by using hidden relationships, as mentioned before.

For interpretability, machine learning and deep learning methods only give binary classification or probability value, which makes it difficult for inspectors to acquire the specific reasons for the high risk of this batch and to provide targeted control measures. The country-commodity combination reveals the relationship between country and commodity, which can be explained at the statistical level. For Tianjin Port, the data cross-validation used among project attributes is easy to understand. Our work provides the risk result in a triple with tag, risk degree, and detailed basis for risk determination.

For scalability, machine learning and deep learning methods usually have poor performances when small amounts of negative data are used. Thus, a model must be retrained to identify new risk types. The country-commodity combination focusing on the combination of direct objects also has a difficult time dealing with objects with complex relationships. Data validation cannot expand to other risk types. Our work defines objects at different levels in the port scenario. Multiple correlation risk types have been identified using our method. The new risk types are suitable for the correlation analysis based risk warning approach.

The evaluation results are presented in Table 1.

5 Conclusion

The study analyses some shortcomings of the current supervision mode of inspection and quarantine for cross-border products. For government departments, incomplete and untrustworthy risk data and inaccurate risk warning services lead to insufficient supervision of enterprises and products. For this reason, we propose a correlation analysis based risk warning service for cross-border trading products.

Correlation analysis is performed on multisource and heterogeneous data to identify deep-seated risks in cross-border products. For collecting sufficient and trusted risk data covering the whole process, specific participants and data models are designed. Analyzing critical correlation risks in scenarios, a specific data model and multi-attribute, multi-object, and multi-level correlation analysis methods are proposed. A permissioned blockchain is developed to change production relationships and act as a trusted data infrastructure. Multiple participants, including producers, consignees, intermediate service providers, and government departments, are involved. The risk warning service comprehensively considers historical and current data from internal and external sources.

The permissioned blockchain is a new third-party service that provides a risk warning service for the Trade Finance Open Platform. Such a service effectively improves risk warning accuracy for cross-border products.

Acknowledgment

The work was supported by the National Key Research and Development Program of China (No. 2019YFB1404904).

Dates

Received: 26 June 2022; Revised: 18 September 2022; Accepted: 19 September 2022

References

[1] H. Xiao, Ensemble learning models and algorithms for risk decision-making problems, PhD dissertation, School of Economics and Business Administration, Chongqing University, Chongqing, China, 2017.

[2] W. Feng and J. Huang, Early warning for civil aviation security checks based on deep learning, *Data Analysis and Knowledge Discovery*, vol. 2, no. 10, pp. 46–53, 2018.

[3] X. Zhou and C. Zhang, Customs risk classification and forecasting model based on data mining, *Journal of Customs and Trade*, vol. 38, no. 2, pp. 22–31, 2017.

[4] Y. Yu, Mining and analysising of Chongqing customs' import and export data, Master dissertation, School of Economics and Business Administration, Chongqing University, Chongqing, China, 2008.

[5] M. Su, L. Bai, and T. Zhou, *Personalization: The Future of Business*, (in Chinese). Beijing, China: Mechanical Industry Press, 2012.

[6] Y. Dai, A research on a planning on entry-exit inspection and quarantine conformity conditions screening and inspection & risk pre-warning, PhD dissertation, School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China, 2009.

[7] C. Yuan and Y. Xu, Research on implementing classified inspection supervision for export mechanical and electrical products, (in Chinese), *Modern Commodity Inspection Science and Technique*, no. 5, pp. 7–9, 17, 2001.

[8] J. Liu, P. Jiang, and J. Leng, A framework of credit assurance mechanism for manufacturing services under social manufacturing context, in *Proc. 2017 13th IEEE Conference on Automation Science and Engineering* (*CASE*), Xi'an, China, 2017, pp. 36–40.

[9] B. J. Kim, S. C. Hong, D. Egger, C. S. Katsar, and R. L. Griffin, Predictive modeling and categorizing likelihoods of quarantine pest introduction of imported propagative commodities from different countries, *Risk Analysis*, vol. 39, no. 6, pp. 1382–1396, 2019.

[10] L. Zhou, L. Cai, L. Jiang, and L. Chen, Power grid enterprise intelligent risk identification model considering multi-attribute and low correlation data, *IEEE Access*, vol. 7, pp. 111324–111331, 2019.

[11] Q. Qian, Research on contagion effect and control strategy of associated credit risk, PhD dissertation, School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, China, 2018.

[12] Tianjin-Port, Tianjin port launches trials of blockchain verification technology, http://www.chinabankingnews.com/2019/04/22/tianjin-port-launches-trials-of-blockchain-verification-technology/, 2019.

Table 1 Evaluation results.

Method	Data	Correlation risk			Interpretability	Scalability
		Between attributes	Between objects	Hidden relationship		
Xiao[1], Feng and Huang[2], Zhou and Zhang[3], Yu[4]	Historical data in the customs	No	No	No	Bad	Bad
Kim et al.[9]	Country-commodity combination	No	Yes (country-commodity)	No	Middle	Bad
Tianjin-Port[12]	Multi-source data	Yes	No	No	Good	Bad
Ours	Multi-source and heterogeneous data	Yes	Yes	Yes	Good	Good