Digital Marketing Strategies for E-commerce Supply Chain Risk Management Based on Big Data Analysis and Predictive Modeling

Yifei Du[1] and Lin Zhu[2*]

[1,2]Department of Business, Sias University, Zhengzhou, 451150,China
[1]zhulin@sias.edu.cn,[2]duyifei@sias.edu.cn

Corresponding author: Lin Zhu, zhulin@sias.edu.cn

Abstract. In order to improve the risk prediction and management effect of e-commerce supply chain, this paper combines big data technology to carry out quantitative risk prediction and management of e-commerce supply chain, and proposes a bitemporal database middleware indexing algorithm. Moreover, this paper deals with e-commerce risk data, draws lessons from the 4R tree index method, and eliminates bitemporal variables through effective data transformation and query transformation. In addition, this paper analyzes and optimizes the two sub-processes of the original node insertion algorithm of the QR tree, and proposes the corresponding query transformation method after data transformation. Through simulation experiments, it can be seen that the risk management system of e-commerce supply chain based on big data technology proposed in this paper has good risk prediction and risk management effects.

Keywords: e-commerce; supply chain intelligence; big data; predictive modeling; quantitative risk; Digital Marketing Strategies
DOI: https://doi.org/10.14733/cadaps.2024.S4.1-18

1 INTRODUCTION

In the actual operation of the enterprise, the complex and changeable environment brings various risks to the operation of the supply chain, so the enterprise shows a certain degree of risk aversion attitude for the prevention of risk events.

Literature [2] studies a two-level supply chain composed of monopoly suppliers and retailers, in which both retailers and supply chains take the mean variance as the objective function to reflect risk aversion. The research shows that the optimal buyback strategy of suppliers is affected by the degree of risk aversion of suppliers and retailers, and rational suppliers can obtain more profits from the buyback strategy. Literature [3] considers the retailer as risk neutral and risk averse in its mean variance model. By comparing the retailer's optimal ordering and pricing strategies in two situations, it is found that when the demand scale is affected by price, the retailer will reduce the order quantity and increase the selling price; while when the demand distribution location is affected by the price, the retailer will Lower its selling price. Literature [18] used the mean variance analysis method to study the newsboy model and some standard multi-period inventory models. The customer waiting time and inventory levels in a multi-period inventory model were analyzed using mean variance. Literature [20] used mean variance to study a single supply chain and a two-level supply chain composed of multiple retailers. The risk aversion degree of retailers in the model obeys a continuous distribution function, and the supply chain coordination is discussed under the menu contract using optimal control theory. Literature [11] used mean variance to study a two-level supply chain model consisting of a single retailer and a supplier. It is found that the degree of risk aversion of the retailer and the supply chain will have an impact on the realization of coordination. When the degree of risk aversion of both the supply chain and the retailer is small, the supply chain is easy to achieve coordination, and when the degree of risk aversion of one party is large, The supply chain is difficult to achieve coordination. Literature [13] applied the revenue sharing contract to study the three-level supply chain model with risk-averse characteristics of supply chain members. This paper discusses the coordination mechanism of the revenue sharing contract under the risk neutrality of supply chain members. On this basis, the mean variance method is used to analyze the revenue sharing when the distributor or the retailer is risk averse, and when both are risk averse. The characteristics of the contract, and a numerical study was done. Literature [12] studies the risk-averse newsboy model with mean variance as the objective function. It is found that out-of-stock costs can have a significant impact on the optimal decision-making of newsboys, and it is worth noting that when there are out-of-stock costs, the order volume of risk-averse newsboys is not necessarily lower than that of risk-neutral newsboys.This article explores the intersection of digital marketing strategies, big data analysis, and predictive modeling in the context of supply chain risk management for e-commerce companies. We will delve into the key components of this approach, discuss the benefits it offers, and highlight some effective strategies that organizations can implement to enhance their supply chain resilience.

Literature [2] establishes a multi-product order risk decision model based on the analysis of investment portfolio and product portfolio, and draws on the CVaR method, and converts the model into a linear programming problem, which can be solved with the help of software tools. The effects of out-of-stock penalty and risk aversion degree on the optimal order quantity of risk-averse retailers are studied. Research shows that when there is a stock-out penalty, the relative size of the optimal order quantity in risk-averse and risk-neutral situations depends on the distribution of demand, the size of the unit out-of-stock penalty and the degree of risk aversion. Literature [15] studies the single-period newsboy model based on the CVaR risk measurement criterion, and points out that the CVaR risk measurement criterion makes the planning problem exhibit convexity, so that the problem can be solved, and is obtained in two different situations. Analytical solution to the problem. In the literature [8], under the two-level supply chain composed of dominant suppliers and following retailers, applying CVaR risk measurement criteria, two different objective function models for retailers are established, and according to the characteristics of the model two-level planning Convert the original model to solve. Literature [9] studied the risk-averse newsboy problem by using the CVaR risk metric under the condition that stochastic market demand depends on price. They discussed the optimal ordering and pricing decisions of risk-averse newsboys in the additive and multiplicative modes of demand, respectively. Literature [10] uses the CVaR criterion to measure the risk-return of retailers, studies the return policy of a two-level supply chain consisting of a single manufacturer and two risk-averse retailers, and discusses the risk aversion of retailers through numerical examples. and other parameters on the manufacturer's return strategy, profit, and retailer's decision-making.

Literature [4] discussed the repurchase strategy under the common sales channel, and he pointed out that it is not optimal for suppliers to provide full-price repurchase or no repurchase policy; when suppliers provide partial-price returns, not only can channel coordination be achieved, but also The coordination efficiency can be improved by Pareto optimization. Literature [6] extended the repurchase model and considered the effect of retail price on stochastic demand on his basis, based on the premise that the repurchase price is related to cost. Literature [17] studied the influence of wholesale price and buyback price on retailer's order quantity under the condition that demand is negatively correlated with price and is normally distributed. When the wholesale price exceeds a certain threshold, the buyback strategy can increase the expected profit of the supplier and the retailer, and the uncertainty of demand leads to the increase of the retail price. Literature [16] studies repurchase contracts in a two-stage production environment by considering two different production models. The first production model is cheaper but requires a longer lead time; the second model is more expensive but provides better responsiveness. The contract needs to determine the repurchase price and the different wholesale prices corresponding to the two models, pointing out that the contract can achieve supply chain coordination. Literature [7] studied the impact of a type of risk-free return policy on supply chain performance, and found that this return policy can improve the profits of suppliers and retailers compared with not allowing returns. Literature [19] discussed the influence of the degree of risk aversion on the repurchase strategy, and the study showed that ignoring the degree of risk aversion of supply chain decision makers will cause serious losses in supply chain performance. When demand is correlated with retail price, repurchase contracts cannot achieve supply chain coordination. With the rapid development of e-commerce, researchers began to study the repurchase policy model under the e-commerce environment. Literature [14] compared the performance of repurchase policy in traditional market and e-commerce market, and provided a Literature for decision makers whether to adopt e-commerce means. This paper analyzes the repurchase model in the e-commerce environment by using the "mean-method" method, and further studies the relationship between the repurchase policy and risk factors in the e-commerce market. Literature [5] studied a two-level supply chain problem with a loss-averse retailer, and pointed out that the repurchase contract can be used to coordinate the supply chain, so as to achieve a Pareto improvement of the profits of the supplier and the retailer, and the coordination can be achieved. Realize the arbitrary distribution of supply chain profits between the two. A two-level supply chain coordination problem composed of suppliers and retailers is studied, in which the supplier is risk-neutral and the retailer is risk-averse and the risk is measured by the CVaR criterion, and it is proved that the repurchase contract and the revenue sharing contract Supply chain coordination can be achieved, and when the retail price is externally given, the buy-back contract and the revenue-sharing contract are equivalent

This paper combines big data technology to carry out quantitative risk prediction and management of e-commerce supply chain, and improves the risk management effect of e-commerce supply chain through intelligent methods.

2 E-COMMERCE SUPPLY CHAIN RISK DATA INDEXING ALGORITHM

In this paper, combining big data technology for risk prediction and management of e-commerce supply chain, a bitemporal database middleware indexing algorithm is proposed.

2.1 Overall Improvement Scheme for Qr Tree Indexing

Referring to the processing method of 4R tree index on temporal variables, this paper firstly divides bitemporal data into four basic types. Then, this paper performs data transformation for each type of data to eliminate bitemporal variables. After data transformation, the query conditions need to be transformed accordingly to ensure the semantic equivalence with the query before transformation. Specifically, it is to map the query conditions to the corresponding R-tree indexes, and then perform the query on the corresponding R-tree, and finally merge to obtain the query results.

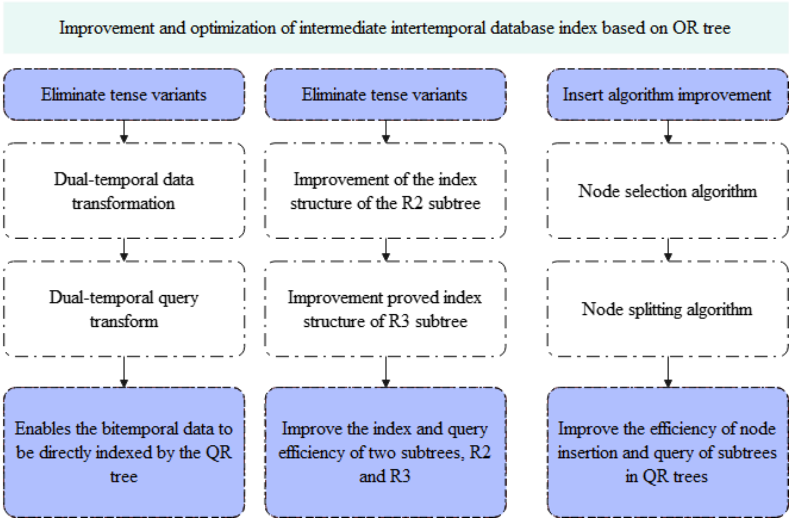


Figure 1: Optimization and improvement scheme of QR index tree.

In this paper, the subtree index structure of QR tree is improved at the same time, which effectively improves the index efficiency of large batch data. Finally, this chapter also optimizes the insertion algorithm of R tree, which can improve the efficiency of query. Figure 1 briefly summarizes the improvement and optimization scheme of the QR tree index in this paper.

2.2 Eliminate Tense Variables "Now" and "Uc"

The temporal query model TQue1 proposed by Snodgrass is used, that is, VT→, VT←, TT→ and TT← are used to indicate the start and end of valid time and the start and end of transaction time. Then, 4-tuple <TT→, TT←, VT→, VT←> can be used to represent temporal data.

Type	*VT→*	*VT←*	*TT→*	*TT←*
A	*vt1*	*Now*	*tt1*	*UC*
B	*vt1*	*Now*	*tt1*	*tt2*
C	*vt1*	*vt2*	*tt1*	*UC*
D	*vt1*	*vt2*	*tt1*	*tt2*

Table 1: Four bitemporal data types.

First, bitemporal data is divided into four types listed in Table 1 according to whether the termination value of valid time is "Now" and whether the termination value of transaction time is "UC". Among them, vf1, vt2, tt1 and tt2 are known and determined time points. The main technology to eliminate the two temporal variables "Now" and "UC" is to perform data transformation and query transformation on bitemporal data, which is divided into two sections.

The bitemporal regions corresponding to the four bitemporal data types A, B, C and D in Table 1 are respectively a growing staircase, a fixed staircase, a growing rectangle and a fixed rectangle, as shown in Figure 2.

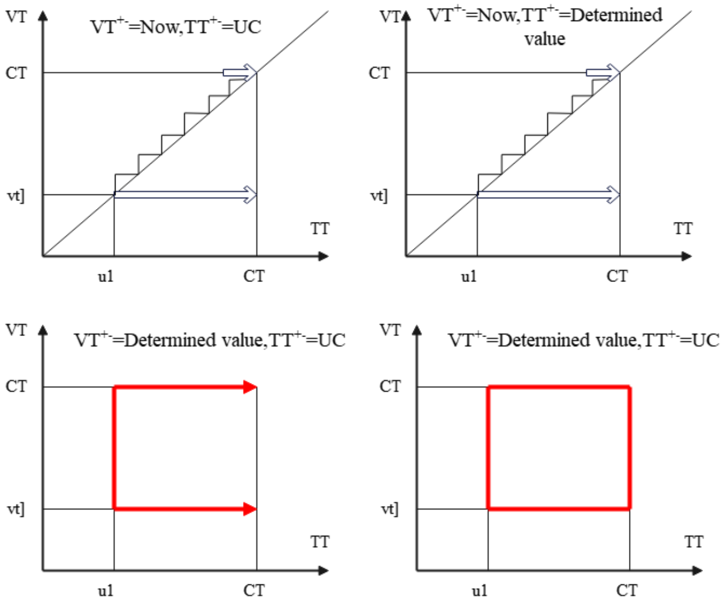


Figure 2: Four bitemporal data types.

Definition 1: If T represents the time domain, and ID represents the temporal data tuple tag, then D^B represents the bitemporal data tuple domain that contains arguments, and D^s represents the bitemporal data tuple domain that does not contain arguments:

$$D^B = \left\{ < TT_r^{-}, TT_r^{\leftarrow}, VT_r^{\rightarrow}, VT_r^{\leftarrow}, id_r > \in T \times T \cup \{UC\} \times T \times T \cup \{Now\} \right. \\ \left. \times ID \,|\, \left(TT_r^{\leftarrow} = UC \vee TT_r^{\rightarrow} \le TT_r^{\leftarrow} \right) \wedge \left(VT_r^{\leftarrow} = Now \vee VT_r^{\leftrightarrow} \le VT_r^{\leftarrow} \right) \right\} \tag{1}$$

$$D^S = \left\{ < TT_r^{\leftrightarrow}, TT_r^{\leftarrow}, VT_r^{\rightarrow}, VT_r^{\leftarrow}, id > \right\} \in T \times T \times T \times T \times ID \,| \\ TT_r^{\leftrightarrow} \le TT_r^{\leftarrow} \wedge VT_r^{\rightarrow} \le VT_r^{\leftarrow} \Big\} \tag{2}$$

Among them, the subscript r represents the data rectangle, which is convenient for distinguishing the query rectangle represented by the subscript q later.

By observing Definition 1, it is not difficult to find that D^B contains the set of four bitemporal data types in Figure 2, while D^S represents the bitemporal data of the fixed rectangle in Figure 2. On the basis of Definition 1, data transformation is to transform those uncertain bitemporal data tuple fields that contain variables into fixed bitemporal data tuple fields that do not contain variables, so as to facilitate representation and storage in R-tree. The basic idea is to extend the bitemporal data tuple that does not contain arguments, and add a Type field representing the transformation type to the tuple attribute.

Definition 2: If $R \subseteq D^B$, and the tuple type is denoted as $Type=\{1,2,3,4\}$, the data transformation $\Gamma_D : 2^{D^D} \to 2^{D^S \times T\ fype}$ that eliminates temporal variables is defined as follows:

$$\Gamma_D(R)=\Big\{\Gamma_r\big(\langle TT_r^{\leftrightarrow},TT_r^{\leftarrow},VT_r^{\leftrightarrow},VT_r^{\leftarrow},id_r>)kTT_r^{\rightarrow},TT_r^{+},$$

$$VT_r^{\rightarrow},VT_r^{\leftarrow},id_r>\in R\Big\}\, among,$$

$$\Gamma_r\Big(\big\langle TT_r^{+},TT_r^{\leftarrow},VT_r^{\rightarrow},VT_r^{+},id_r\big\rangle\Big)=$$

$$\begin{cases} <TT_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{+},VT_r^{\leftarrow},id_r,1> \ if \ \left(TT_r^{\leftarrow}=UC \wedge VT_r^{+}=Now\right) \\ <TT_r^{\rightarrow},TT_r^{+},VT_r^{+},VT_r^{+},id_r,2> \ if \ \left(TT_r^{\leftarrow}\neq UC \wedge VT_r^{+}=Now\right) \\ <TT_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{+},VT_r^{+},id_r,3> \ if \ \left(TT_r^{\leftarrow}=UC \wedge VT_r^{+}\neq Now\right) \\ <TT_r^{\rightarrow},TT_r^{+},VT_r^{+},VT_r^{+},id_r,4> \ if \ \left(TT_r^{\leftarrow}\neq UC \wedge VT_r^{\leftarrow}\neq Now\right) \end{cases} \quad (3)$$

From formula (3), it can be seen that the four types of bitemporal data tuple fields A, B, C and D have more Type sub-attributes after transformation than before, and their values can be 1, 2, 3 or 4. When establishing an index, they will be distinguished according to the value of the Type field, and then different indexes will be established, so the index needs 4 R trees. They are named R1, R2, R3 and R4, respectively. Figure 3 shows the comparison between the actual data area and the transformed data area of the four types of bitemporal data tuples A, B, C and D.

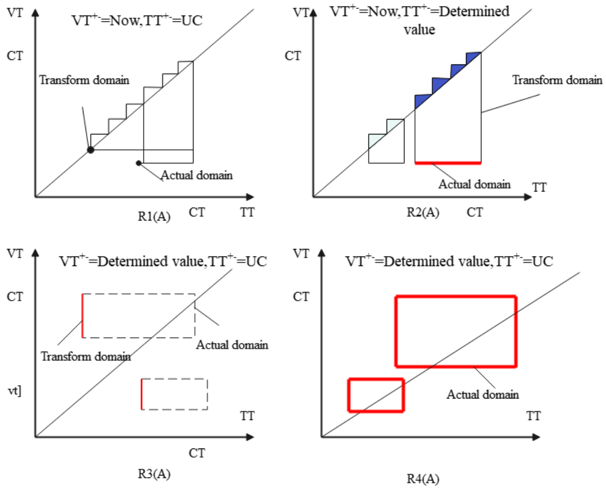


Figure 3: Comparison of the four bitemporal data regions before and after transformation.

The validity time of type A data and the termination of transaction time are both uncertain times. The original data area extends forward to the current time CT in the direction of the TT axis, and extends upward to the current time CT in the direction of the VT axis, showing a staircase shape that grows along the straight line VT=TT. After the transformation, the tuple structure of this type of data is expressed as $\left\langle TT_r^{\rightarrow}, TT_r^{\leftarrow}, VT_r^{\rightarrow}, VT_r^{\leftarrow}, id_r, 1 \right\rangle$. Among them, the type field 1 indicates that the data is converted from bitemporal data of type A.

The termination of the valid time of type B data is not a definite value, while the termination of transaction time is known. The original data area extends up to the current time CT in the direction of the VT axis, showing a fixed staircase shape along the straight line VT=TT. After the transformation, the tuple structure of this type of data is represented as $\left\langle TT_r^{\rightarrow}, TT_r^{\leftarrow}, VT_r^{\rightarrow}, VT_r^{\leftarrow}, id_r, 2 \right\rangle$. Among them, the type field 2 indicates that the data is converted from the bitemporal data of type B.

The termination of the C-type data transaction time is not a definite value, while the termination of the valid time is known. The original data area extends in the direction of the TT axis to the current time CT, presenting as a growing rectangle along the TT axis. After the transformation, the tuple structure of this type of data is represented as $\left\langle TT_r^{\rightarrow}, TT_r^{\leftarrow}, VT_r^{\rightarrow}, VT_r^{\leftarrow}, id_r, 3 \right\rangle$. Among them, the type field 3 indicates that the data is converted from bitemporal data of type C.

The effective time and transaction time of D-type data are known, and the original data area is a fixed rectangle, and no data transformation is required. This is just to be consistent with the previous three types of data tuple structure A, B and C, so add the type field 4, which is expressed as $\left\langle TT_r^{\rightarrow}, TT_r^{\leftarrow}, VT_r^{\rightarrow}, VT_r^{\leftarrow}, id_r, 4 \right\rangle$.

On the basis of the R-tree index, the query is performed in the rectangular domain, also known as the rectangular intersection query. $< TT_q^{\rightarrow}, TT_q^{\leftarrow}, VT_q^{\rightarrow}, VT_q^{\leftarrow} > \in T \times T \times T \times T$ represents the rectangular area of the intersecting query, and T is the time domain.

Definition 3:We set the data domain $r1 :< TT_{r1}^{\rightarrow}, TT_{r1}^{\leftarrow}, VT_{r1}^{\rightarrow}, VT_{r1}^{\leftarrow} >$,

data domain $r2 :< TT_{r2}^{\rightarrow}, TT_{r2}^{\leftarrow}, VT_{r2}^{\rightarrow}, VT_{r2}^{\leftarrow} >$

If the condition $\left(TT_{r1}^{\leftrightarrow} \le TT_{r2}^{\leftarrow}\right) \wedge \left(TT_{r1}^{\leftarrow} \ge TT_{r2}^{\leftrightarrow}\right) \wedge \left(VT_{r1}^{\rightarrow} \le VT_{r2}^{\leftarrow}\right) \wedge \left(VT_{r1}^{\leftarrow} \ge VT_{r2}^{\rightarrow}\right)$ is true, then r1 and r2 are said to intersect, denoted by $< TT_{r1}^{\rightarrow}, TT_{r1}^{+}, VT_{r1}^{\rightarrow}, VT_{r1}^{\leftarrow} > \cap \left\langle TT_{r2}^{+}, TT_{r2}^{+}, VT_{r2}^{\rightarrow}, VT_{r2}^{\leftarrow} \right\rangle$. If the query rectangle is $q = TT_q^{\rightarrow}, TT_q^{\leftarrow}, VT_q^{\rightarrow}, VT_q^{\leftarrow} > \rangle$, the current time domain is CT, and $R \subset D^B$, then the intersection query $Inter\,sect^B$ on R is defined as follows:

$$Inter\,sec\,t^{B}[\,q,CT\,](\,R\,)=\{i_r\,|\,\alpha\wedge(\,\beta\vee\gamma\vee\delta\vee\omega\,)\}$$

Among them,

$$\begin{aligned}
&\alpha=\left(<TT_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{\rightarrow},VT_r^{\leftarrow},id_r>\in R\right),\\
&\beta=\left(TT_r^{\leftarrow}=UC\wedge VT_r^{\leftarrow}=Now\wedge TT_q^{\rightarrow}\geq VT_q^{\rightarrow}\wedge q\cap\right.\\
&\left.<TT_r^{\rightarrow},CT,VT_r^{\rightarrow},CT>\right),\\
&\gamma=\left(TT_r^{+}=UC\wedge VT_r^{\leftarrow}\neq Now\wedge q\bigcap\langle TT_r^{\rightarrow},CT,VT_r^{\rightarrow},VT_r^{\leftarrow}>\right),\\
&\delta=\left(TT_r^{\leftarrow}\neq UC\wedge VT_r^{\leftarrow}=Now\wedge TT_q^{\leftarrow}\geq VT_q^{\rightarrow}\wedge q\cap.\right.\\
&\left.<TT_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{\rightarrow},CT>\right),\\
&\omega=\left(TT_r^{\leftarrow}\neq UC\wedge VT_r^{\leftarrow}\neq Now\wedge q\cap\langle TT_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{\leftrightarrow},VT_r^{\leftarrow}>\right)
\end{aligned}\tag{4}$$

Definition 4: If $q=\left\langle TT_q^{+},TT_q^{\leftarrow},VT_q^{\rightarrow},VT_q^{+}\right\rangle,R\in D^{S}$, the intersection query on R with query rectangle q as parameter is defined as follows:

$$\begin{aligned}
&Inter\,sec\,t^{s}[\,q\,](\,R\,)=\left\{id_r\,|\,TT_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{\rightarrow},VT_r^{\leftarrow},id_r>\in\right\rangle\\
&R\wedge q\cap<TT_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{\rightarrow},VT_r^{\leftarrow}>\}
\end{aligned}\tag{5}$$

On the basis of definition 3 and definition 4, the query transformation Γ_q of 4R-tree index is defined. For four types of bitemporal data tuples, the query transformation maps an intersecting query condition to multiple transformations of the original data.

Definition 5: First, the following premise definitions are introduced:

$$\begin{aligned}
&R\subseteq D^{B},S=\Gamma_D(\,R\,)S_i=\{\langle T_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{\rightarrow},VT_r^{\leftarrow},id>/\\
&<TT_r^{\leftrightarrow},TT_r^{\leftarrow},VT_r^{\leftrightarrow},VT_r^{\leftarrow},id_r,i>\in S\}\;among\;i=1,2,3,4,\\
&q=TT_q^{\rightarrow},TT_q^{\leftarrow},VT_q^{\rightarrow},VT_q^{\leftarrow}>\rangle\\
&q_1=\left\langle 0,TT_q^{\leftarrow},0,VT_q^{\leftarrow}\right\rangle\\
&q_2=\left\langle max\left(TT_q^{\leftrightarrow},VT_q^{\leftarrow}\right),TT_q^{\leftarrow},0,VT_q^{\leftarrow}\right\rangle\\
&q_3=\left\langle 0,TT_q^{\leftarrow},VT_q^{\leftrightarrow},VT_q^{\leftarrow}\right\rangle\\
&q_4=TT_q^{\rightarrow},TT_q^{\leftarrow},VT_q^{\leftrightarrow},VT_q^{\rightarrow}>\rangle
\end{aligned}\tag{6}$$

Then, the query transformation $\Gamma_q = \left[2^{D^B} \to 2^{ID} \right] \to \left[2^{D^S \times Type} \to 2^{ID} \right]$ is defined as follows:

$$\Gamma_q = \left(Inter\,sec\,t^B[q,CT] \right)(S) = \begin{cases} \left\{ \bigcup_{i=1,2,3,4} Inter\,sec\,t^s[q_i](S_i) \text{ if } \left(TT_q^{\leftarrow} \geq VT_q^{\leftrightarrow} \right) \right. \\ \left\{ \bigcup_{1=2.4} Inter\,sec\,t^s[q_1](S_i) \text{ if } \left(TT_q^{\leftarrow} < VT_q^{\rightarrow} \right) \right. \end{cases} \tag{7}$$

It can be seen from formula (7) that each query on the $4R$ tree needs to go through 2 subtrees or 4 subtrees. The query conditions corresponding to each sub-tree are also different. The equivalence of the data set before and after the transformation between the data transformation of the $4R$-tree and the query transformation is analyzed below.

Theorem 1: For any $q = \left\langle TT_q^{\leftrightarrow}, TT_q^{\leftarrow}, VT_q^{\leftrightarrow}, VT_q^{+} \right\rangle$ and any data set $R \subseteq D^B$, the following formula holds:

$$Inter\ sec\,t^B[q,CT](R) = \Gamma_q\left(Inter\ sec\,t^B[q,CT] \right)\left(\Gamma_D(R) \right) \tag{8}$$

To fully implement the QR tree index, the system needs to perform some unification and transformation operations on the corresponding index subtrees of the four types of bitemporal data. Among them, the most important function includes how to determine which subtree to perform before inserting, querying and deleting operations for a given data object.

1 Insert operation. When inserting data objects into data, the database should be able to determine which of the four types of bitemporal data the new data belongs to. Here, it is only necessary to simply judge whether the termination value of the valid time and the termination value of the transaction time are the current time. The definition of insert operation is:

$$IntertQR\left(TT^{\rightarrow}, TT^{\leftarrow}, VT^{\rightarrow}, VT^{\leftarrow}, Type \right) =$$

$$\begin{cases} QR1\,Insert\left(TT^{\rightarrow}, VT^{\rightarrow} \right),\ if\ (Type = 1) \\ QR2\,Insert\left(TT^{\rightarrow}, TT^{\leftarrow}, VT^{\rightarrow} \right),\ if\ (Type = 2) \\ QR3\,Insert\left(TT^{\leftrightarrow}, VT^{\leftrightarrow}, VT^{\leftarrow} \right),\ if\ (Type = 3) \\ QR4\,Insert\left(TT^{\rightarrow}, TT^{\leftarrow}, VT^{\rightarrow}, VT^{\leftarrow} \right),\ if\ (Type = 4) \end{cases} \tag{9}$$

2 Query operation. The query operation needs to divide the query into the corresponding subtree according to the query conditions. It should be noted that it is entirely possible for a query to become a query on multiple subtrees after division and transformation, which requires the query results to be merged at the end. The query transformation in the previous article has been clearly summarized, so it will not be repeated here. The definition of the query operation is:

$$SearchQR\left(TT^{\rightarrow},TT^{\leftarrow},VT^{\rightarrow},VT^{\leftarrow}\right)=$$

$$\begin{cases} Search\ QR1\left(0,TT^{\rightarrow},0,VT^{\leftarrow}\right)\cup \\ Search\,QR2\left(max\left(TT^{\rightarrow},VT^{\rightarrow}\right),TT^{\rightarrow},0,VT^{\leftarrow}\right)\bigcup \\ Search\,QR3\left(0,TT^{\leftarrow},VT^{\rightarrow},VT^{\leftarrow}\right)\bigcup \\ Search\,QR4\left(TT^{\rightarrow},TT^{\leftarrow},VT^{\rightarrow},VT^{\leftarrow}\right) \\ \ if\ \left(TT^{\rightarrow}\neq CT\wedge TT^{\leftarrow}\geq VT^{\rightarrow}\right) \\ Search\,QR3\left(0,TT^{\leftarrow},VT^{\rightarrow},VT^{\leftarrow}\right)\cup \\ SearchQR4\ \left(TT^{\rightarrow},TT^{\leftarrow},VT^{\rightarrow},VT^{\leftarrow}\right) \\ \ if\ \left(TT^{\rightarrow}\neq CT\wedge TT^{\leftarrow}<VT^{\rightarrow}\right) \\ Search\,QR1\left(0,TT^{\leftarrow},0,VT^{\leftarrow}\right)\bigcup \\ SearchQR\,3\left(0,TT^{\leftarrow},VT^{\rightarrow},VT^{\leftarrow}\right) \\ if\ \left(TT^{\rightarrow}=CT\wedge\left[VT^{\rightarrow},VT^{\leftarrow}\right]\neq\left[T_{min}T_{max}\right]\wedge TT^{\leftarrow}\geq VT^{\rightarrow}\right) \\ SearchQR\,3\left(0,TT^{4},VT^{-},VT^{+}\right) \\ if\left(TT^{\rightarrow}=CT\wedge\left[VT^{\rightarrow},VT^{\leftarrow}\right]\neq\left[T_{min},T_{max}\right]\right) \\ QR1\bigcup QR2 \\ if\left(TT^{\rightarrow}=CT\wedge\left[VT^{\rightarrow},VT^{\leftarrow}\right]=\left[T_{min},T_{max}\right]\right) \end{cases} \quad (10)$$

3 Delete operation. When deleting a temporal data object, you must first know the index subtree where the data object is located. Similar to the insertion operation, the corresponding subtree must be judged by the value of the Type field of the data. The definition of the delete operation is:

$$\begin{cases} QR1\ Delete\ \left(TT^{\rightarrow},VT^{\rightarrow}\right),\ if\left(\ Type\ =1\right); \\ QR2\ Delete\ \left(TT^{\rightarrow},VT^{\rightarrow},VT^{\leftarrow}\right),\ if\left(\ Type\ =2\right); \\ QR3\ Delete\ \left(TT^{\rightarrow},TT^{\leftarrow},VT^{\rightarrow}\right),\ if\left(\ Type\ =3\right); \\ QR4\ Delete\ \left(TT^{\rightarrow},TT^{\leftarrow},VT^{\rightarrow},VT^{\leftarrow}\right),\ if\left(Type=4\right) \end{cases} \quad (11)$$

4 Update operation. The update operation can be understood as the delete operation first, and then the insert operation. It should be noted that for a bitemporal database, the update can only be performed on the current version of the data, but not on the historical version of the data. The data of the historical version only provides data snapshots for querying.

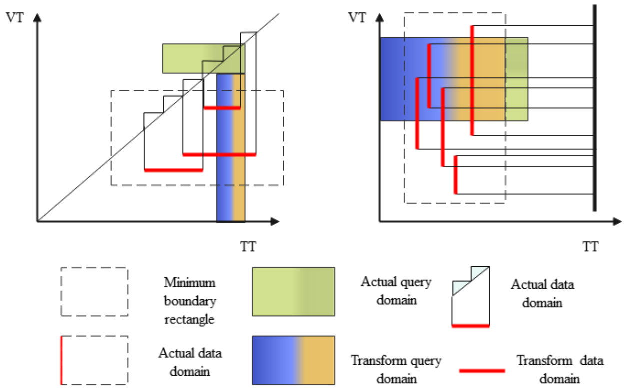


Figure 4: Shows the redundant query space caused by the extension of the query area.

For the R4 tree, the indexed data itself is a fixed area and does not need to be transformed. It can be directly indexed by the R tree, and its index efficiency and query efficiency are high, so there is less room for improvement and optimization.

2.3 Optimization Of Qr Tree Node Insertion Algorithm

Most of the current R-tree node splits are implemented by the optimal bound search method of the objective function, which generally requires the minimum MBR area or perimeter value after splitting. However, in some cases, such an objective function does not achieve optimal splitting. Figure 5 shows two different splitting methods for the same data object, and the dotted box represents the outer rectangle after splitting. The area and perimeter of the outer rectangle of split A are smaller than those of split B, but obviously split B is more reasonable. This shows that using the objective function alone does not guarantee optimal node splitting.

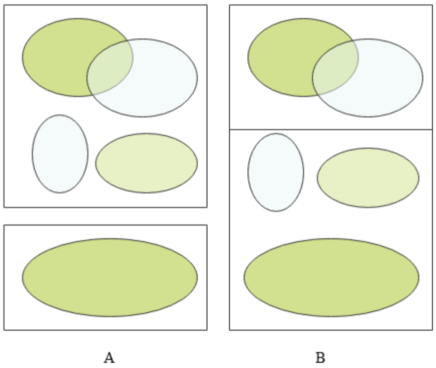


Figure 5: Inadequacy of optimal split objective function.

Under the constraints of the objective function, further constraints need to be added to ensure the rationality of QR tree node splitting. The rationality of node splitting has an important relationship with the two enclosing rectangles after splitting, instead of only looking at the comprehensive parameters accumulated by the two enclosing rectangles. Therefore, this paper introduces the following node splitting constraint factor λ :

$$\lambda = \left|S(R_1) - S(R_2)\right| / \left(S(R_1) + S(R_2)\right), \; among \lambda \in [0,1] \tag{12}$$

We use λ to measure the rationality of the child node MBR after node splitting, and restricting the upper bound value of λ can ensure the rationality of node splitting. Obviously, if λ is too small, unreasonable splitting is likely to occur, and if it is too large, the situation in Figure 5 cannot be avoided. Through experiments, for data with normal distribution, the more appropriate λ should be around 0.7, and for excessively distorted data, the λ factor may take a larger value. For example, it can be taken as above 0.9.

The main steps of the improved algorithm are to firstly traverse each node in the leaf node layer in order, and find the node that can cover the object to be inserted in a minimum and complete manner. The meaning of minimum complete coverage here is that under the premise that node P completely covers the object obj to be inserted, the area difference between P and obj is the smallest. Doing so ensures that the object obj is inserted in the optimal position. If a node covering the object to be inserted is found in the previous step, the algorithm stops the search, and then starts to search for the position to be inserted top-down until the leaf node. If the node that can cover the object to be inserted is not found in the previous step, it means that a new node needs to be created to contain and point to the object to be inserted, and the search continues to the upper level until a non-leaf node that can contain the node to be inserted is found.

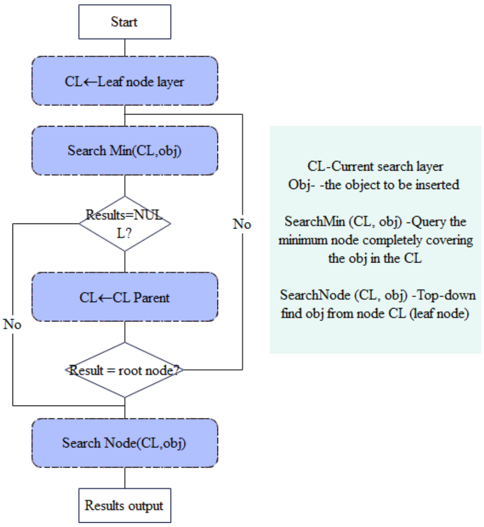


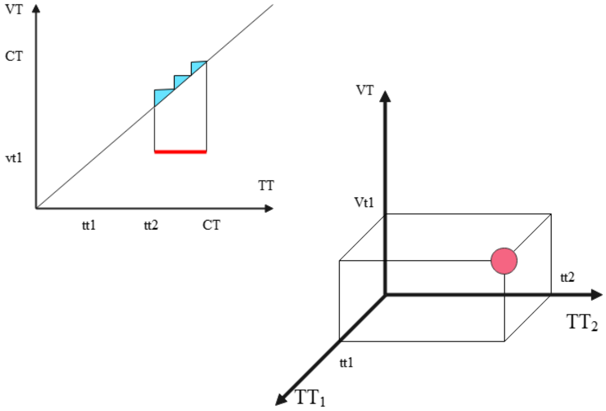
Figure 6: Flowchart of the improved node query algorithm.

The improved algorithm flow is shown in Figure 6, and the functions and symbols are marked in the box on the left.

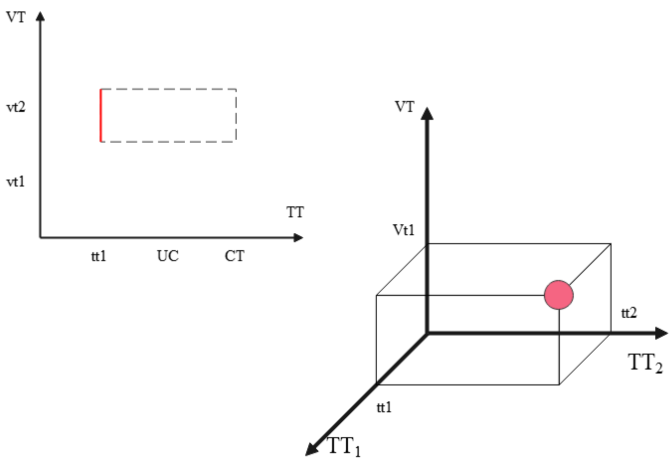
2.4 Improvements To the Subtree Structure Of Qr Trees

On the R2 tree, the original data tuple after data transformation is represented as $\langle tt1, vt1, vt2, 2\rangle$. The start and end values of the transaction time are equal to t1. This representation is to save storage space. The three dimensions after the improvement are the transaction time start dimension, the effective time start dimension and the effective time end dimension, denoted as TT_1, TT_2 and VT. Figure 7(a) compares the corresponding temporal data areas in the R2 subtree before and after the improvement. On the R3 tree, the original data tuple after data transformation is expressed as $< tt1, tt2, vt1, 3 >$. Among them, the start and end values of the valid time are equal, both are vt1. Such a representation can save index storage space. After the improvement, the three dimensions are the transaction time start dimension, the transaction time end dimension and the effective time dimension, denoted as TT, VT_1 and VT_2. Figure 7(b) compares the corresponding temporal data areas in the R3 subtree before and after the improvement.

Comparing the MBRs of the two types of temporal data objects before and after improvement, we can find that the minimum bounding rectangle of the data item $< OI_1, MBR_1 >$ in the original node in the two subtrees is denoted as $MBR_i = (I_1, I_2)$. Among them, I_1 represents the transaction time dimension and I_2 represents the effective time dimension. Since the common feature of temporal data on $R2$ and $R3$ subtrees is that they are represented as a line segment on the original two-dimensional plane, the structure of $< TT^{\rightarrow}, VT^{\rightarrow}, VT^{\leftarrow} >$ or $\langle TT^{\leftarrow}, TT^{\leftarrow}, VT^{\rightarrow}\rangle$ can be used to represent and store the corresponding MBR_r.



(a) Comparison of temporal data regions of R2 subtrees before and after improvement



(b) Comparison of temporal data regions of R3 subtrees before and after improvement

Figure 7: Comparison of temporal data regions before and after improvement.

Definition 6: The query condition is $q=\left(\left(VT_q^{\rightarrow},VT_q^{\leftarrow}\right),\left(TT_{1q}^{\leftrightarrow},TT_{1q}^{\leftarrow}\right),\left(TT_{2q}^{\leftrightarrow},TT_{2q}^{\leftarrow}\right)\right)$,

The data rectangle is $r=\left(\left(VT_r^{\rightarrow},VT_r^{\leftarrow}\right),\left(TT_{1r}^{\rightarrow},TT_{1r}^{+}\right),\left(TT_{2r}^{\rightarrow},TT_{2r}^{\leftarrow}\right)\right)$.

If the condition $\left(VT_q^{\rightarrow}\le VT_r^{\leftarrow}\right)\wedge\left(VT_q^{\leftarrow}\ge VT_r^{\rightarrow}\right)\wedge\left(TT_{1q}^{\rightarrow}\le TT_{1r}^{\leftarrow}\right)\wedge \left(TT_{1q}^{\leftarrow}\ge TT_{1r}^{\rightarrow}\right)\wedge\left(TT_{2q}^{\rightarrow}\le TT_{2r}^{\leftarrow}\right)\wedge\left(TT_{2q}^{\leftarrow}\ge TT_{2r}^{\rightarrow}\right)$ is true, then q and r are said to intersect, denoted by $q\cap r$. The current time is represented by CT, then in the new R2 index space, the intersection query of bitemporal data is expressed as:

$$\begin{aligned}&Intersect^{B}[q,CT](R)=\left\{id_r \mid TT_r^{\rightarrow},TT_r^{\leftarrow},VT_r^{\rightarrow},VT_r^{\leftarrow}>\right\rangle\\&\in R\wedge\left(VT_r^{\leftarrow}=Now\wedge TT_r^{\leftarrow}\ne UC\wedge q\cap\right.\\&\left.<\left(0,TT_r^{\leftarrow}\right),\left(TT_r^{\rightarrow},TT_r^{\leftarrow}\right),\left(VT_r^{+},CT\right)>\right\}\end{aligned}\tag{13}$$

In order to ensure the equivalence of the original two-dimensional index and the improved three-dimensional index query, the original query condition also needs to be converted, the original two-dimensional query condition $q=\left(max\left(TT_q^{\leftrightarrow},VT_q^{\rightarrow}\right),TT_q^{\leftarrow},0,VT_q^{\leftarrow}\right)$ should be converted to the three-dimensional query condition $q'=\left(\left(0,VT_q^{+}\right),\left(TT_q^{\rightarrow},max(T)\right),\left(0,TT_q^{+}\right)\right)$. Among them, $max(T)$ represents the maximum time value on the time domain T, $\left(0,VT_q^{+}\right)$ represents the start and end interval on the effective time axis, $\left(0,TT_q^+\right)$ represents the start and end interval on the transaction time start axis, and $\left(TT_q^{\rightarrow},max(T)\right)$ represents the start and end interval on the transaction time end axis. The corresponding bitemporal data area is shown in Figure 8.

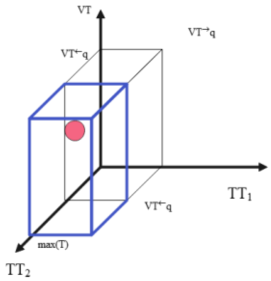
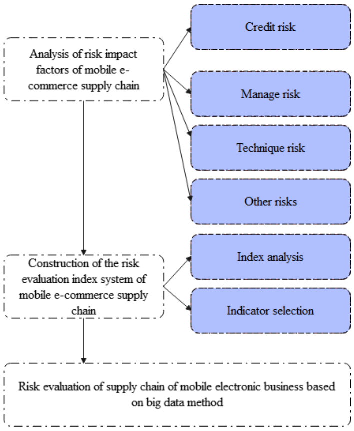


Figure 8: Query domain in 3D space after R2 subtree improvement.

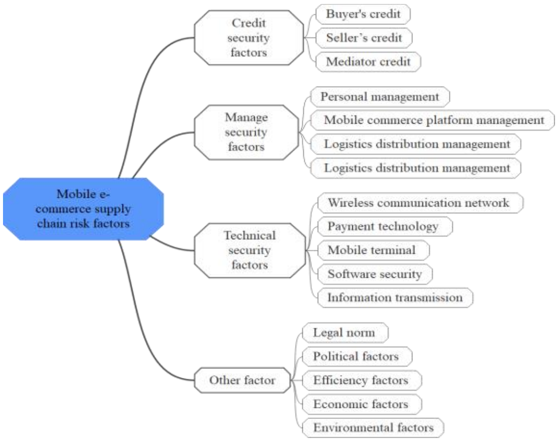
The data on the R2 subtree is transformed into a spatial point object, and the points satisfying the query conditions should be included in the solid line cuboid in Figure 8. When querying, it is only necessary to determine whether the point data object is in the cuboid (query rectangle).

3 RISK PREDICTION AND MANAGEMENT OF E-COMMERCE SUPPLY CHAIN BASED ON BIG DATA TECHNOLOGY

Figure 9(a) presents the evaluation idea of constructing the risk factors of mobile e-commerce supply chain. After fully considering the basic functions that distinguish traditional businesses, this paper gives the following evaluation index system. These factors directly or indirectly affect the security of the mobile e-commerce supply chain and omit intermediate sub-goals, and the evaluation index system formed is shown in Figure 9(b).



(a) Ideas for risk factor analysis of e-commerce supply chain



(b) E-commerce supply chain risk factor evaluation index system

Figure 9: E-commerce supply chain risk management system based on big data technology.

The effect of the e-commerce supply chain risk management system based on big data technology proposed in this paper is verified, and the risk prediction and risk management effects of the system are calculated. This paper obtains multiple sets of basic test data from the network, and combines the simulation platform to verify the effect, and finally obtains the simulation test results shown in Figure 10.

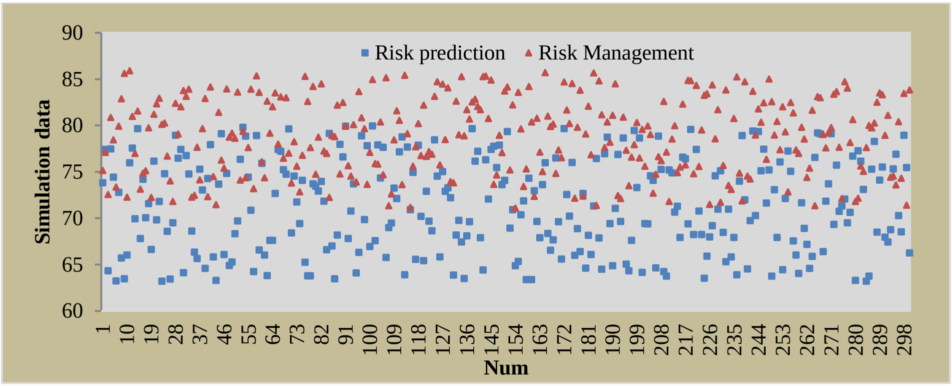


Figure 10: Experimental results of system simulation.

The simulation test shows that the risk management system of e-commerce supply chain based on big data technology proposed in this paper has good risk prediction and risk management effects.

4 CONCLUSION

The realization of the maximum benefit of the supply chain requires each supply chain member to cooperate and cooperate with each other, and these members are independent individuals pursuing the maximization of their own interests. Supply chain members often ignore the optimal performance of the overall supply chain for the sake of maximizing their own interests, resulting in sub-optimal decision-making in the supply chain. Based on this, the supply chain contract coordinates the supply chain to ensure that members of the supply chain can achieve the optimal overall benefit without reducing their own benefits. However, in supply chain management research, the risk issue has been ignored. This paper combines big data technology to carry out quantitative risk prediction and management of e-commerce supply chain, and improves the risk management effect of e-commerce supply chain through intelligent methods. Through the simulation test, it can be seen that the risk management system of e-commerce supply chain based on big data technology proposed in this paper has good risk prediction and risk management effects. the integration of digital marketing strategies with big data analysis and predictive modeling holds immense potential for e-commerce companies in effectively managing supply chain risks. By leveraging the wealth of data available through various digital channels, companies can gain valuable insights into their supply chain operations, identify vulnerabilities, and make informed decisions to mitigate risks proactively.

Yifei Du, https://orcid.org/0000-0002-8236-841X
Lin Zhu, https://orcid.org/0000-0001-7022-3304

ACKNOWLEDGEMENT

This research work was MOE (Ministry of Education in China) Project of Humanities and Social Sciences “Research on the Organization Mode and Governance Mechanism of Shared Recycling Recreation” (20YJC630241).

REFERENCES

[1] Anitha, J.; Kalaiarasu, M.: Optimized machine learning based collaborative filtering recommendation system in e-commerce, Journal of Ambient Intelligence and Humanized Computing, 12(6), 2021, 6387-6398. https://doi.org/10.1007/s12652-020-02234-1

[2] Chang, D.; Gui, H. Y.; Fan, R.; Fan, Z. Z.; Tian, J.: Application of improved collaborative filtering in the recommendation of e-commerce commodities, International Journal of Computers Communications & Control, 14(4), 2019, 489-502. https://doi.org/10.15837/ijccc.2019.4.3594

[3] Chen, H.: Personalized recommendation system of e-commerce based on big data analysis, Journal of Interdisciplinary Mathematics, 21(5), 2018, 1243-1247. https://doi.org/10.1080/09720502.2018.1495599

[4] Chen, J.: The Application of Commodity Recommendation in Cross-border E-commerce: Current Situation and Prospect, Frontiers in Economics and Management, 2(1), 2021, 266-274.

[5] Fedirko, O.; Zatonatska, T.; Wolowiec, T.; Skowron, S.: Data Science and Marketing in E-Commerce Amid COVID-19 Pandemic, European Research Studies Journal, 24(2), 2021, 3-16. https://doi.org/10.35808/ersj/2187

[6] He, G.: Enterprise E-Commerce Marketing System Based on Big Data Methods of Maintaining Social Relations in the Process of E-Commerce Environmental Commodity, Journal of Organizational and End User Computing, 33(6), 2021, 1-16. https://doi.org/10.4018/JOEUC.20211101.oa16

[7] Hosseini, F.; Sadighi, H.; Mortazavi, S. A.; Farhadian, H.: An E-Commerce SWOT Analysis for Export of Agricultural Commodities in Iran, Journal of Agricultural Science and Technology, 21(7), 2019, 1641-1656.

[8] Huiliang, Z.; Yang, Q.; Liu, Z.: Impact of online customer reviews and deep learning on product innovation empirical study on mobile applications, Business Process Management Journal, 27(6), 2021,1912-25. https://doi.org/10.1108/BPMJ-12-2020-0542
[9] Liang, L.; Qin, X.: Research on consumers online shopping decision-making and recommendation of commodity based on social media network, Cluster Computing, 22(3), 2019, 6529-6539. https://doi.org/10.1007/s10586-018-2296-7
[10] Liu, D.; Huo, C.; Yan, H.: Research of commodity recommendation workflow based on LSH algorithm, Multimedia Tools and Applications, 78(4), 2019, 4327-4345. https://doi.org/10.1007/s11042-018-5716-z
[11] Pan, H.; Zhang, Z.: Research on context-awareness mobile tourism e-commerce personalized recommendation model, Journal of Signal Processing Systems, 93(2), 2021, 147-154. https://doi.org/10.1007/s11265-019-01504-2
[12] Shen, J.; Zhou, T.; Chen, L.: Collaborative filtering-based recommendation system for big data, International Journal of Computational Science and Engineering, 21(2), 219-225. https://doi.org/10.1504/IJCSE.2020.105727
[13] Subramaniyaswamy, V.; Logesh, R.; Chandrashekhar, M.; Challa, A.; Vijayakumar, V.: A personalised movie recommendation system based on collaborative filtering, International Journal of High Performance Computing and Networking, 10(1-2), 2017, 54-63. https://doi.org/10.1504/IJHPCN.2017.083199
[14] Sukrat, S.; Papasratorn, B.: An architectural framework for developing a recommendation system to enhance vendors' capability in C2C social commerce, Social Network Analysis and Mining, 8(1), 2018, 1-13. https://doi.org/10.1007/s13278-018-0500-7
[15] Wang, Z.; Wan, M.; Cui, X.; Liu, L.; Liu, Z.; Xu, W.; He, L.: Personalized recommendation algorithm based on product reviews, Journal of Electronic Commerce in Organizations,16(3), 2018, 22-38. https://doi.org/10.4018/JECO.2018070103
[16] Wei, C.; Niu, J.; Guo, Y.: DLGNN: a double-layer graph neural network model incorporating shopping sequence information for commodity recommendation, Sensors and Materials, 32(12), 2020, 4379-4392. https://doi.org/10.18494/SAM.2020.3056
[17] Xu, J.; Hu, Z.; Zou, J.: Personalized product recommendation method for analyzing user behavior using DeepFM, Journal of Information Processing Systems, 17(2), 2021, 369-384.
[18] Yang, F.: A hybrid recommendation algorithm–based intelligent business recommendation system, Journal of Discrete Mathematical Sciences and Cryptography, 21(6), 2018, 1317-1322. https://doi.org/10.1080/09720529.2018.1526408
[19] Yu, T.: A case study of B2C cross-border e-commerce challenges in China from Customs to consumers, World Customs Journal, 12(2), 2018, 121-132.
[20] Zhou, L.: Product advertising recommendation in e-commerce based on deep learning and distributed expression, Electronic Commerce Research, 20(2), 2020, 321-342. https://doi.org/10.1007/s10660-020-09411-6