Hindawi
Advances in Multimedia
Volume 2023, Article ID 9840641, 1 page
https://doi.org/10.1155/2023/9840641

Retraction

Retracted: The Role of Cognitive Science and Big Data Technology in the Design of Business Information Management Systems

Advances in Multimedia

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Li, "The Role of Cognitive Science and Big Data Technology in the Design of Business Information Management Systems," *Advances in Multimedia*, vol. 2022, Article ID 2761661, 15 pages, 2022.

Hindawi
Advances in Multimedia
Volume 2022, Article ID 2761661, 15 pages
https://doi.org/10.1155/2022/2761661

Research Article

The Role of Cognitive Science and Big Data Technology in the Design of Business Information Management Systems

Yangfan Li

School of Art and Design, Wuchang Shouyi University, Wuhan 430064, China

Correspondence should be addressed to Yangfan Li; 2015225010015@stu.scu.edu.cn

Received 29 August 2022; Revised 19 September 2022; Accepted 28 September 2022; Published 17 November 2022

Academic Editor: L. T. T. Nguyen

With the surge in the amount of data in the internal and external environment, the collection, analysis, processing, and storage of the increasing data sources and data volume, as well as the problems of big data management, are the current situation and dilemmas of data management faced by enterprises today. Cognitive science and big data technology can provide good auxiliary support for enterprise management decision-making. This study takes the business information management of cold chain logistics enterprises as an example, aiming at the characteristics of business intelligence data in real applications, based on cognitive science and big data technology, from low-cost and high-performance storage, security management, and big data analysis. This paper is mainly through the research of big data processing theory and key technologies. Based on analyzing the logistics industry's data access rules and characteristics, this study proposes a hot data prediction model for multilayer hybrid storage systems. It is verified that the prediction model has good accuracy, robustness, and universality. For the application scenario of multitenant distributed data access, a data transparent security management model is proposed. Simulation experiments show that this method can realize data security management when the performance loss is controlled within an acceptable range. Based on the real-time computing technology of massive data, the label optimization scheme of collaborative filtering and reinforcement learning is used to realize the logistics distribution recommendation model and to solve the accuracy and real-time problems of logistics service distribution analysis.

1. Introduction

Informatization has become a significant technology to promote the development of commercial modernization [1, 2]. Massive information transactions and production data are generated through various informatization systems in the enterprise [3]. At present, the important production data in the enterprise information system can only be stored for one month. If it is to be stored effectively for a long time, it requires expensive storage costs. In addition, to deal with massive commercial data information, extracting useful data is very important for the production management of enterprise managers [4, 5]. However, traditional data management and data mining technologies are unable to store and analyze large amounts of data efficiently [6, 7]. Therefore, although companies have valuable data resources, they cannot be effectively utilized them [8, 9]. In the process of business information management, the effective storage and use of big data can effectively reduce the cost of enterprise information technology operations [10, 11]. In addition, managers can conduct an in-depth understanding of data, provide quantified decisions and technical support for enterprise management, improve production and sales efficiency, and ultimately improve economic efficiency.

With the in-depth integration and integration of the domestic e-commerce and logistics industries, the latest technology in the information field is gradually combined in the research of logistics informatization [12, 13]. A large number of crossresearches on domestic logistics information technology and cloud computing technology are carried out [14, 15]. The development of logistics enterprises depends on the development of e-commerce and promotes the development of e-commerce [16, 17]. On the basis of the original industrial system, logistics, commercial flow, and information flow, combined with the agency and distribution of goods, have developed a socialized logistics and distribution system [18, 19]. Due to the explosive increase of business information data, there is an urgent need to use big data processing technology in modern business information management to discover potentially exploitable information, which can not only provide data support for enterprise management when making decisions but also provide better data services for users of the enterprise [20, 21]. Big data technology has been used by many researchers in business information management [22, 23]. Machine learning can help management handle commercial data to make judgments and decisions on key issues in operations, optimize logistics information systems, optimize services, save space, and control inventory [24, 25]. In the logistics industry, the Internet of Things technology provides it with comprehensive information collection capabilities [26, 27]. RFID, GPS, infrared induction, sensors, and other technologies are used to collect commercial information anytime and anywhere to form a full range of spatiotemporal monitoring data. These massive amounts of data need to be processed urgently, and solving this problem requires the help of big data technology [28]. A complete cloud computing architecture is formed in the analysis system, which can help business managers to grasp the business situation more clearly, provide scientific data for the development trend of the enterprise, and provide decision suggestions for managers.

In business development, key technologies such as cloud computing and big data analytics can be used to better manage modern business information [29, 30]. However, researchers working in computer science in this multidisciplinary field of study lack knowledge about business confidence management. Therefore, the specific application of cognitive science and big data technology in business information management is difficult to achieve [31]. The research foundation of information technology in business information management is relatively weak, resulting in the slow development of existing cloud computing and big data analysis technologies in business information management [32, 33]. In the development process of modern logistics, the fourth-party logistics mode appears, to realize the sharing and unification of information flow, capital flow, and logistics of various logistics enterprises and users [34]. The fourth-party logistics information platform provides basic information processing and data processing services for various logistics enterprises and realizes the safe sharing of information [35]. Since the shared information flow can run through the existing logistics and capital flow, the shared economic model can easily develop the shared transportation mode of logistics on the current information platform [36]. Each logistics enterprise constructs an independent transportation network, and the loading rate and unloading rates of vehicles are heavily dependent on the supply of goods. However, a large number of vehicles travel on the same route during transportation, resulting in a huge waste of transportation resources. Due to the lack of an information platform for logistics data sharing, the current research on logistics sharing transportation is still in its infancy. However, due to different application scenarios, the existing theoretical research results of transportation route optimization and transportation sharing optimization in the field of intelligent transportation are difficult to be directly applied in logistics shared transportation, and the logistics industry needs to process large-scale data such as trajectories and packages in real time. In transportation scheduling, parallel computing solutions suitable for big data analysis and processing need to be considered, to meet users' needs for computing performance.

This study takes logistics information management as an example to illustrate the application of cognitive science and big data technology in business information management. Massive of transaction data, management data, and surveillance video data will be generated in the information management system of commercial logistics enterprises, and the analysis and real-time processing technology of big data will be studied. The research topic belongs to the field of big data processing and application, and the research results can provide theoretical and technical guidance for the informatization construction of local trade and logistics cities. The research content includes three contents: the first is the storage of commercial data, the next is the deal with commercial data, and the last one is the analysis and application of commercial data. In this thesis, we study and deeply analyze data storage, computation, and clues. The distributed storage of big data is the foundation, and the computation is the support for data analysis. The ultimate goal is to analyze the data and realize the application of big data in the decision-making of commercial logistics enterprises. This paper focuses on the design and implementation of the logistics big data processing basic system, as well as the application of big data analysis and calculation. Based on analyzing the laws and characteristics of data access in the logistics industry, a hotspot data prediction model for multilayer hybrid storage systems is proposed. Aiming at the application scenario of multitenant distributed data access, a data-transparent security management model is proposed. On the big data analysis platform, two parallel algorithms of collaborative filtering and label optimization of logistics distribution services are implemented, which solves the performance bottleneck of recommendation service computing.

2. Data Security in Storage

At present, the commonly used data storage method is multilevel hybrid distributed storage, which is a complex storage system, and the multitenant storage outsourcing model has become the main application scenario. However, business survey results show that data security has gradually become a prominent obstacle to the development of distributed storage systems and cloud storage applications. Multilevel hybrid distributed storage is a complex storage system, and the multitenant storage outsourcing model has become the main application scenario. However, business survey results show that data security has gradually become a prominent obstacle to the development of distributed storage systems and cloud storage applications. At present, only 20% of users are willing to put private data in the cloud or distributed storage system, and 50% of users only want to store data backup and disaster recovery data in the cloud. Therefore, establishing an effective data security model for multilevel hybrid distributed storage systems has become a new

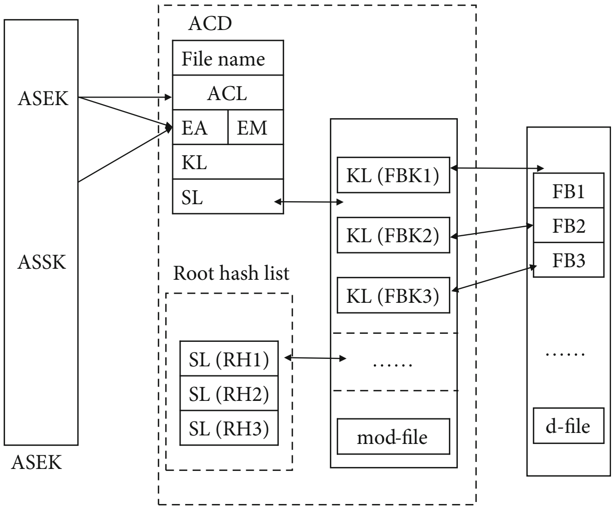


Figure 1: HMSS data logical structure and key management.

challenge. There are two fundamental problems in a multilevel hybrid distributed storage system: (1) the optimistic assumption of the trusted domain of distributed storage systems leads to ignoring attacks and threats from within the storage system; (2) the storage management mechanism of the multilevel hybrid storage system is more complex, such as copy mechanism, load balancing, data synchronization, and data migration. Complex data management requires a corresponding data security mechanism to work together. Aiming at these two fundamental problems, this section focuses on the data security model of multilevel distributed hybrid storage. Design the data security model from the perspective of a computer system to realize the effective integration of data security control and storage management in the hybrid storage system. The cooperative working mechanism of control and data storage plane is studied, and an authentication scheme of multilevel key control is proposed.

2.1. Key Technologies of the Data Security Model. Data security in HMSS is guaranteed by multiple levels of keys. This scheme not only enhances data security but also reduces the communication cost during key usage and maintenance. The multilayer key management and distribution usage scheme is shown in Figure 1. Data security in HMSS is guaranteed by multilevel keys. Data privacy controls address two key issues: (1) structural design of data files and (2) key management in a distributed key environment. For the structural design of the data file, the data file is logically divided into two parts in the data storage node: the metadata file and the data file. The metadata file (mod-file) stores attribute related to data security and storage management, such as access information, root hash linked list, data file popularity, and storage node location. The data file (d-file) stores the ciphertext data of the file. Secondly, the key management of HMSS adopts a hierarchical management scheme, which not only enhances data security but also reduces the communication cost in the process of key use and maintenance.

Data security model HMSS is to realize the effective integration of the storage management plane and the data security control plane; it is necessary to realize the cooperative management and data synchronization of the two planes. Collaborative management is implemented using event-triggered data synchronization. To realize collaborative management, two main problems need to be solved: the basic data structure and collaborative events to realize data synchronization in a distributed environment. The HMSS data synchronization mechanism is shown in Figure 2.

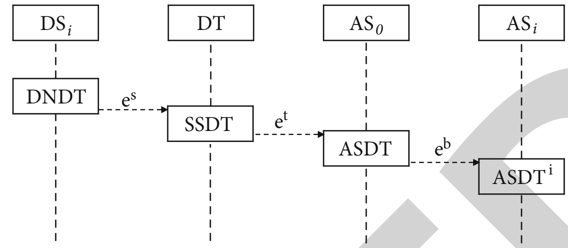


Figure 2: Schematic diagram of HMSS collaboration and data synchronization.

2.2. Testing and Analysis. This section tests the performance consumed by HMSS data security management, regardless of the network latency and system performance consumed by the client reading and writing data. By writing a test script to simulate virtual users to initiate data access requests, in the test script, starting from 100 virtual users, the initial step is 50 virtual users, and when more than 1000 virtual users are adjusted, the virtual user adjustment step is increased to 100 users. The experimental benchmark authentication model is a single authentication server working mode, where an authentication server deployed in the LAN is responsible for the authentication and data encryption management of 30 storage nodes. All virtual users have to initiate data authentication and encryption requests to this authentication server. The experimental observation points are the performance metrics of the first-level authentication server AS0: CPU utilization, memory utilization, network utilization, and disk utilization. As shown in Figures 3–6, the horizontal coordinates indicate the number of virtual users and the vertical coordinates indicate the utilization of various hardware metrics.

To improve the adaptability to complex storage management systems, this section first realizes the decoupling of storage management and data security management from the logical level, forming two independent spaces for storage management and data security. In the storage management layer, the scheme of data storage is allowed to be adjusted without affecting the design of data security. Experiments show that the method proposed in this study has better performance for file reading, and the performance loss caused by data security is within the acceptable range for users.

3. Recommended Models and Algorithms for Mass Logistics Data Distribution

This section is aimed at studying the recommendation model applicable to logistics distribution and the implementation of distributed parallel algorithms in business

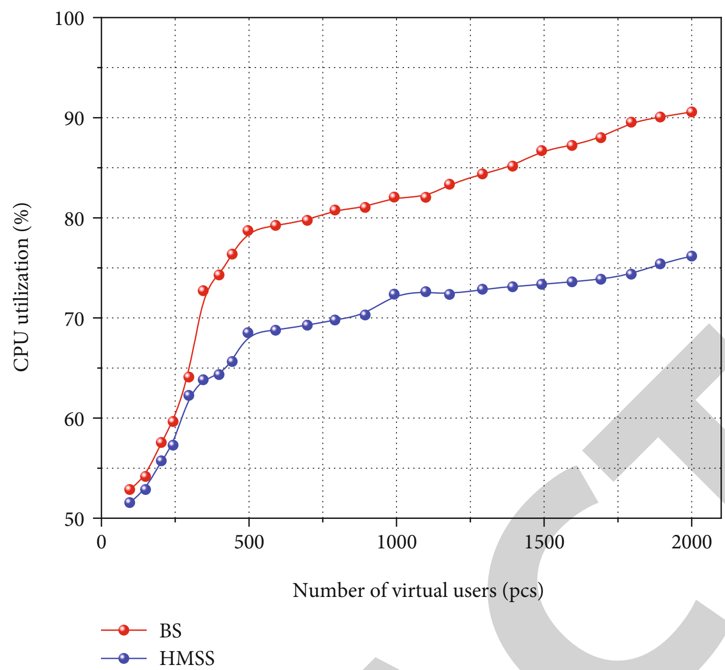


Figure 3: Comparison of CPU utilization between HMSS and benchmark authentication mode.

information management. The advantages of the algorithm are as follows: (1) the algorithm can collect data quickly in real time, although the data in the logistics operation environment is exceptionally complex. (2) The sparsity of high-dimensional data in this algorithm is more obvious in the context of big data. (3) In the fusion of multisource data, the mainstream data types of this algorithm are unstructured and semistructured stream data which are conducive to storage and processing. Data in the logistics industry is very complex, leading to serious problems such as noisy and redundant data. To address the above-mentioned problems, it is necessary to study the recommendation algorithm for logistics distribution services. In addition, due to data accumulation, algorithms suitable for distributed parallel processing are developed to meet the user's demand for processing massive data. This research can improve the user experience, increase the utilization of data resources in the logistics industry, and bring economic benefits to the business management of enterprises.

3.1. Logistics Distribution Recommendation Model. In the condition of massive business data, the high-dimensional sparsity of collaborative recommender systems is more obvious, and because the data types become more complex, the problem of redundant and noisy data is more serious. The processing of real-time streaming data also puts forward new requirements on the computational performance of recommender systems. To better deal with high-dimensional sparse data, reduce the sensitivity of redundant and noisy data, and reduce the computational complexity of the algorithm. In this section, the matrix decomposition method is used to design a logistics distribution recommendation model, and the recommendation is transformed into a matrix decomposition problem, and the sparse user rating matrix is mapped to a specific set of users and items, which can effectively reduce the sensitivity of high-dimensional data. On the other hand, the recommendation system can collect more contextual data in the big data analysis environment. This section focuses on the optimization of recommended labels and uses the optimized labels to build a more accurate logistics distribution recommendation model.

First, we construct a network graph about the previous relationship between the three parts of user-resource-tag and then use a tag ranking algorithm to obtain the popularity of tags. At the same time, considering the characteristics of tags decaying over time, the tags whose popularity is ranked in the front are filtered out, and the garbage or redundant tags whose popularity is lower than a certain threshold are deleted. First, three network graphs are established with users, recommended resources, and labels as network nodes. There are directions and weights between nodes in the three-part network graph. The three-part network graph is formally expressed as $G = \{V, E, W\}$, where $U = \{u_1, u_2, \cdots, u_m\}$ represents the user set; $R = \{r_1, r_2, \cdots, r_n\}$ represents the set of resources to be recommended; $T = \{t_1, t_2, \cdots, t_l\}$ represents the set of tags; V is the set of all users, resources, and tag nodes; $V = \{u_1, u_2, \cdots, u_m, r_1, r_2, \cdots, r_n, t_1, t_2, \cdots, t_l\}$; E is the set of directed edges; $E = \{e_{uv}, e_{rs}, e_{tp}, e_{ur}, e_{ut}, e_{rt}\}$; e_{uv} denote the user edge from u to user v; e_{rs} denote the edge from resource r to resource s; e_{tp} denote the edge from label t to label p; e_{ur} denote the edge from user u to resource r; e_{ut} denote the edge from user u to label t; e_{rt} denote the resource, r denote the edge to label t; and W denote the set of edge weights.

The six methods for calculating the weights of directed edges between nodes in a tripartite graph network are as follows:

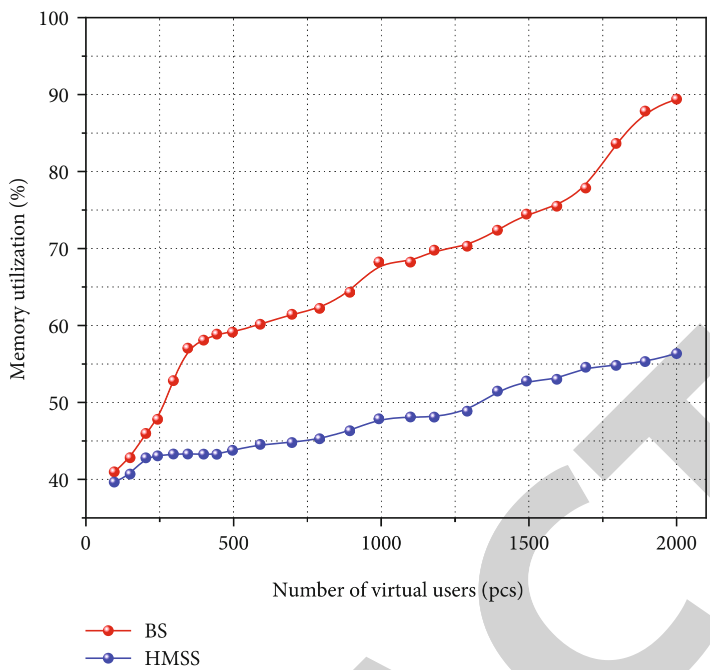


Figure 4: Comparison of memory utilization between HMSS and benchmark authentication mode.

(1) Weights between user and user nodes. Set user nodes u_1 and u_2; if u_1 and u_2 use the same label or label the same resource, then there is a bidirectional edge between u_1 and u_2, and its weight is

$$w(u_1, u_2) = \frac{1}{2}\left(\frac{\|T(u_1) \cap T(u_2)\|}{\|T(u_1) \cup T(u_2)\|} + \frac{\|R(u_1) \cap R(u_2)\|}{\|R(u_1) \cup R(u_2)\|}\right) \quad (1)$$

$T(u_1)$ and $T(u_2)$ represent the set of labels marked by u_1 and u_2, respectively; $R(u_1)$ and $R(u_2)$ represent the set of resources to be recommended marked by u_1 and u_2, respectively.

(2) Weights between resource and resource nodes. Assuming resource nodes r_1 and r_2, the same user or the same label has marked r_1 and r_2; then, there is a bidirectional edge between r_1 and r_2, and its weight is

$$w(r_1, r_2) = \frac{1}{2}\left(\frac{\|T(r_1) \cap T(r_2)\|}{\|T(r_1) \cup T(r_2)\|} + \frac{\|R(r_1) \cap R(r_2)\|}{\|R(r_1) \cup R(r_2)\|}\right) \quad (2)$$

$T(r_1)$ and $T(r_2)$ represent the set of labels labeled r_1 and r_2, respectively; $U(r_1)$ and $U(r_2)$ represent the set of users labeled r_1 and r_2, respectively.

(3) Weights between tag and tag nodes. Assuming label nodes t_1 and t_2, if the same user or the same resource has labeled t_1 and t_2, then there is a bidirectional edge between t_1 and t_2, and its weight is

$$w(t_1, t_2) = \frac{1}{2}\left(\frac{\|T(t_1) \cap T(t_2)\|}{\|T(t_1) \cup T(t_2)\|} + \frac{\|R(t_1) \cap R(t_2)\|}{\|R(t_1) \cup R(t_2)\|}\right) \quad (3)$$

$U(t_1)$ and $U(t_2)$ represent user sets marked by t_1 and t_2, respectively, and $R(t_1)$ and $R(t_2)$ represent resource sets marked by t_1 and t_2, respectively.

(4) Weights between user and resource nodes. If user u_i labels resource r_j, there is a directed edge from u_i to r_j, and the weight is set to 1

(5) Weights between user and tag nodes. If user u_i uses label t_j, there is a directed edge from u_i to t_j; then, the weight is

$$w(u_i, t_j) = \frac{\mathrm{Freq}(u_i, t_j)}{\sum_{t \in T(u_i)} \mathrm{Freq}(u_i, t)} \quad (4)$$

T [25] represents the set of tags used by u_i, and Freq(u_i, t_j) represents the number of times the user u_i has used the tag t_j.

(6) Weights between resource and tag nodes. If t_j labels resource r_i, a directed edge exists from r_i to t_j with the following weights:

$$w(r_i, t_j) = \frac{\mathrm{Freq}(r_i, t_j)}{\sum_{t \in T(r_i)} \mathrm{Freq}(r_i, t)} \quad (5)$$

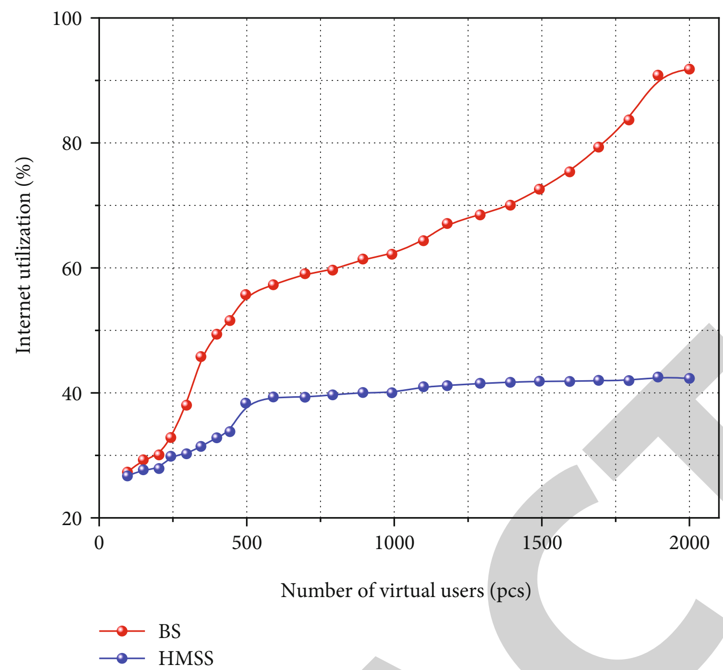


Figure 5: Comparison of network bandwidth utilization between HMSS and benchmark authentication mode.

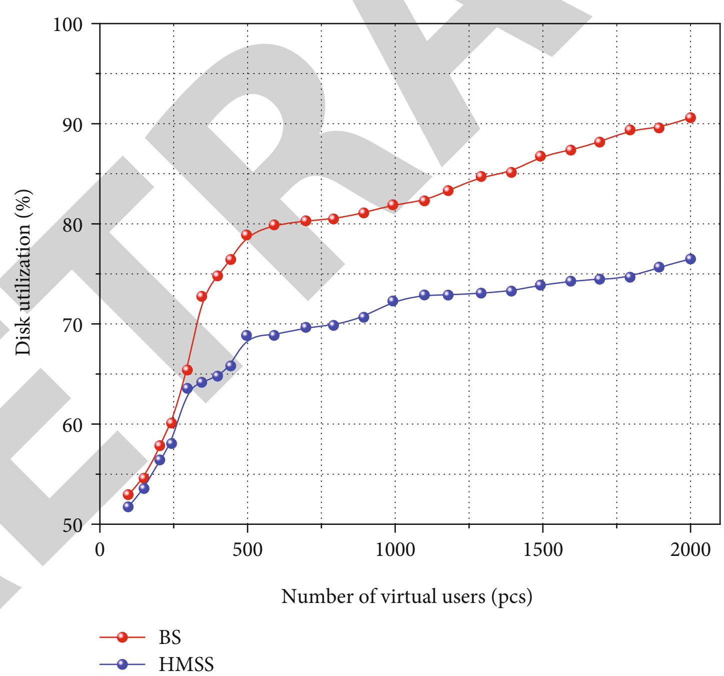


Figure 6: Disk utilization comparison between HMSS and benchmark certification mode.

$T(r_i)$ represents the set of labels that label r_i, and Freq(r_i, t_j) represents the number of times t_j is used to label r_i.

The tag optimization strategy is based on three assumptions: (1) high-quality users who use high-quality tags to tag resources have higher value; (2) users who use high-quality tags to tag high-value resources also have higher quality; and (3) the labels used by high-quality users to label high-value resources are of better quality; the above assumptions constitute a mutually reinforcing learning relationship. This mutually reinforcing learning relationship is introduced based on the search engine's scoring algorithm (Hyperlink-Induced Topic Search, HITS) to obtain the popularity of users, resources, and recommended tags. Sorting according to the popularity of the tags, the tags with low popularity are determined as spam tags and deleted.

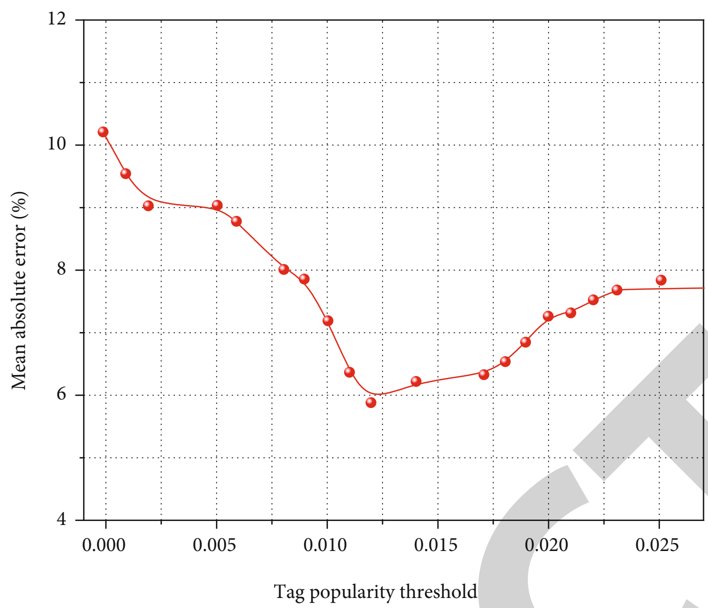


Figure 7: The effect of tag popularity threshold ε on MAE.

The HITS algorithm that introduces mutual reinforcement learning is as follows: Any node $g_k(\text{auth}, \text{hub})$ in the network graph G is represented by two attribute values, and the initial position of each node is set to (1,1). At the beginning of the algorithm, all nodes in the network graph G are input, and the following updates are performed for each node g_k:

$$\begin{aligned} \text{auth}(g_k) &= \sum_{i=1 \text{and } i\neq k}^{n} \text{hub}(g_i) \times w(g_k, g_i), \\ \text{hub}(g_k) &= \sum_{i=1 \text{and } i\neq k}^{n} \text{auth}(g_i) \times w(g_k, g_i), \end{aligned} \tag{6}$$

where the sum of nodes in G is denoted by n and $w(g_{\text{k}}, g_i)$ represents the weight of nodes g_i to g_k. In each node iterative update, the attribute values of each node should be renormalized to prevent numerical overflow. After K iterations converge, return the auth values of all label nodes. To reduce data size, delete all tags with auth value that is lower than ε. In the recommendation process of social network tags, there is another problem that needs to be solved: the recommended tags reduce the recommendation effect over time. In order to comprehensively consider the characteristic that the quality of labels decreases over time, a temporal attention function is introduced here to realize the decay of labels over time:

$$f(t_i^{\text{cur}}) = 1/(1 + b|t_i^{\text{cur}} - t_i|). \tag{7}$$

In Equation (7), $f\,(t_i^{\text{cur}})$ is the label time attention function, parameter b controls the decay speed, and t_i^{cur} and t_i are the current time and the time when the label was created, respectively.

3.2. Logistics Distribution Recommendation Algorithm. Based on user preferences and resource characteristics, this section recommends a business information management model characterized by tag vectors. Although these tags are widely used and popular with the public, these tags lack personalization characteristics. For example, users mark "big data" tags in network resources, and such popular tags cannot reflect the characteristic of the recommended resources. Hence, judging and limiting the weight of these labels are necessary.

First, a preference function for user tags needs to be constructed. Based on how often users use tags, the preference function can be determined, that is, when a user uses a tag frequently, which means the user prefers this tag. The function for user tags is

$$p(u_i, t_j) = tf(u_i, t_j) \cdot idf_u(t_j), \tag{8}$$

where $tf(u_i, t_j)$ is the number of resources marked by the user u_i, as shown in the following equation:

$$\begin{aligned} tf(u_i, t_j) &= w(u_i, t_j), \\ idf_u(t_j) &= \log \frac{M}{m_j}, \end{aligned} \tag{9}$$

where $idfu(tj)$ represents the ratio between the number of users and the number of tag t_j. Therefore, u_i can be defined as

$$p(u_i, t) = (p(u_i, t_1), p(u_i, t_2) \cdots p(u_i, t_j)). \tag{10}$$

Second, establish a resource tag preference function. The tag preference function that defines the resource is as

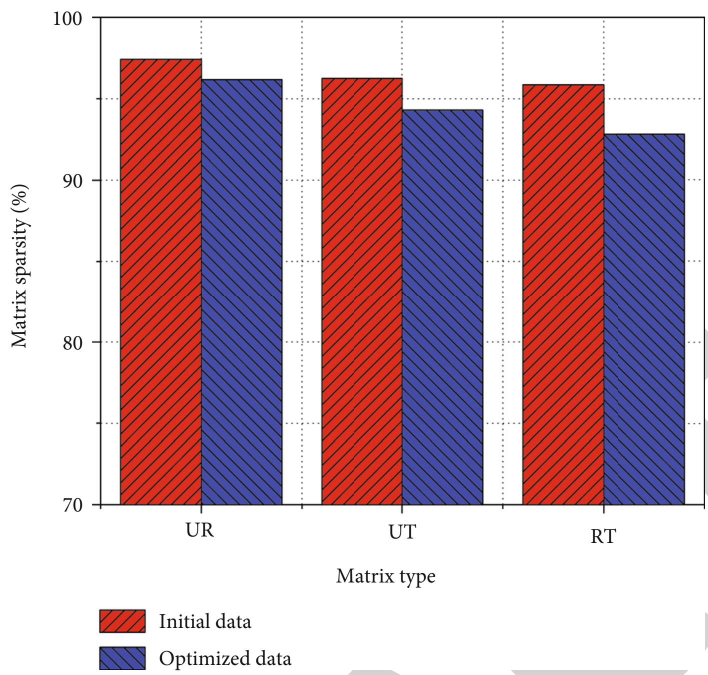


FIGURE 8: Sparsity comparison of matrices.

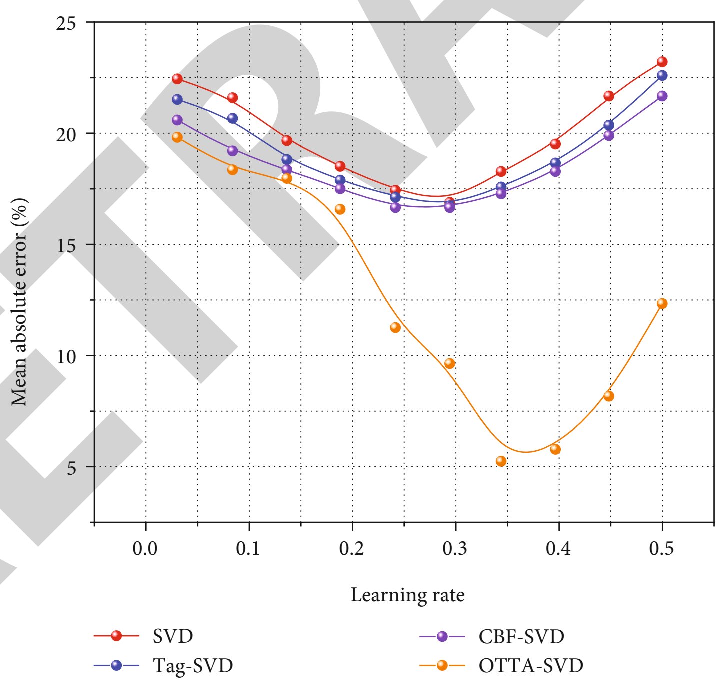


FIGURE 9: Effect of learning rate α on MAE.

follows:

$$p(r_i, t_j) = tf(r_i, t_j) \cdot idf_r(t_j), \tag{11}$$

where$tf(r_i, t_j)$represents the number of users who haver_i label resource with labelt_j, which is

$$\begin{aligned} tf(r_i, t_j) &= w(r_i, t_j), \\ idf_r(t_j) &= \log \frac{N}{n_j}, \end{aligned} \tag{12}$$

where $idfr(tj)$ represents the ratio of the number of

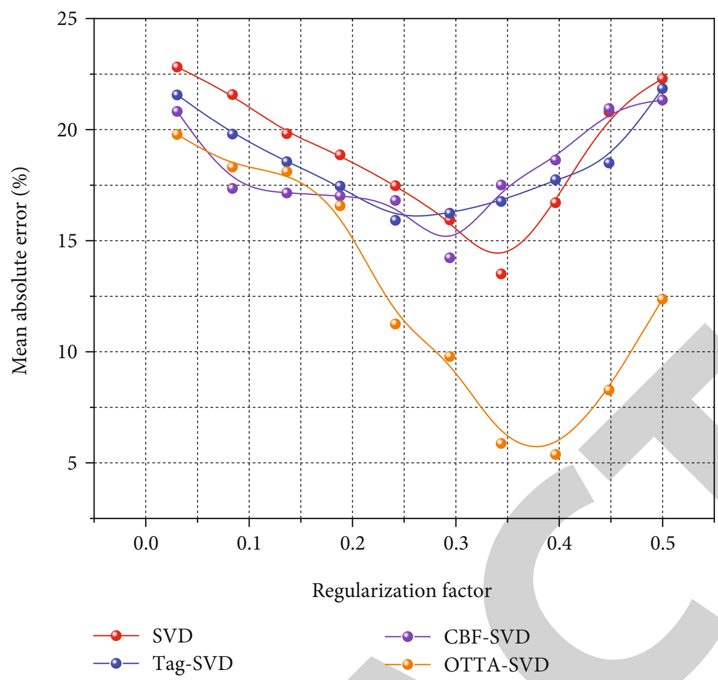


Figure 10: Influence of regularization coefficient β on MAE.

resources to the number of resources labeled with tag t_i. So the resource r_i can be expressed by

$$p(r_i, t) = (p(r_i, t_1), p(r_i, t_2) \cdots p(r_i, t_j)) \quad (13)$$

Then, the user's score for the resource can be calculated by

$$\mathrm{Pre}(u, r) = \sum_{t \in T(u) \cap T(r)} p(u, t) \times p(r, t). \quad (14)$$

The recommendation model needs to input various vector data of users and resources, and then, it will automatically output the information of recommended resources. The basic principle is a matrix-based recommendation model that decomposes the user's preference matrix for a resource into a low-dimensional feature matrix, a process shown in the following equation:

$$R_{m\times n} = D_{m\times F} V_{n\times F}^T, \quad (15)$$

where $D_{m\times F}$ represents the user feature matrix, $V_{n\times F}$ represents the resource feature matrix, and F denotes the amount of latent features. From this, the feature matrix of user and the feature matrix of resource can be applied to estimate the user u_i's score for the resource r_j:

$$r_{i,j} = \sum_{f=1}^{F} d_{\mathrm{if}} \times v_{jf}^T. \quad (16)$$

The function of objective loss is

$$E = \sum_{i\in U \text{ and } j\in R} \left(r_{i,j} - \sum_{f=1}^{F} d_{\mathrm{if}} \times v_{jf}^T \right)^2 + \beta\left(\|d_i\|^2 + \|v_j\|^2 \right). \quad (17)$$

The recursive formula of the iterative process is as follows:

$$\begin{aligned} d_{\mathrm{if}} &\longleftarrow d_{\mathrm{if}} + \alpha \times (e_{i,j} \times v_{jf} - \beta d_{\mathrm{if}}), \\ v_{jf} &\longleftarrow v_{jf} + \alpha \times (e_{i,j} \times d_{\mathrm{if}} - \beta v_{jf}), \end{aligned} \quad (18)$$

where α is the learning rate and $e_{i,j}$ is the deviation between the score known and the value predicted. The above process is repeated n times unless the objective function E is less than the threshold predefined, and the final feature vectors D and V were netted.

4. Experiment Verification and Improvement

4.1. Simulation Experiment Verification. The experimental data set is the data set on Delicious, with a total of 10,000 users, 9,007 resources, and 4,438 tags. Since the labels used by users in the data set have timestamps, the records of each user in the data set are sorted by time in the experiment, the first 20% of the records are used as test data, and the last 80% of records are used as training data. In the experiment, the mean absolute error (MAE) is used as the evaluation metric to evaluate the degree of deviation between the resource score predicted by the recommendation algorithm and the user's actual resource score. The calculation formula is as

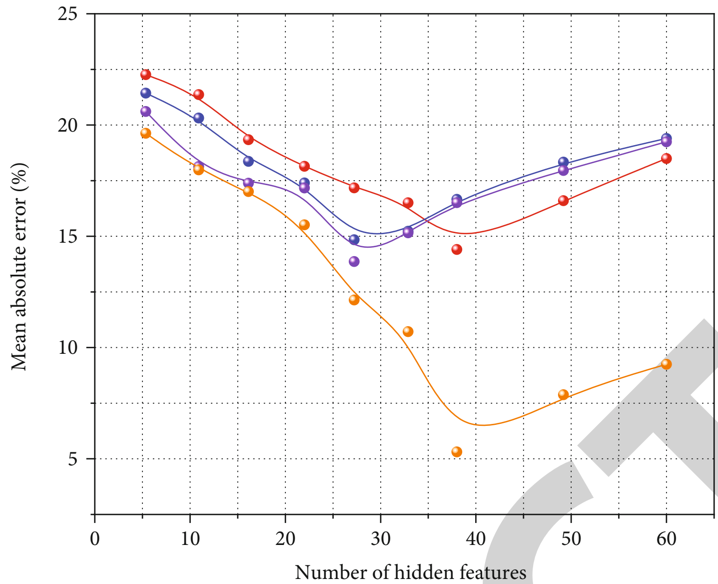


FIGURE 11: Influence of the number of latent features F on MAE.

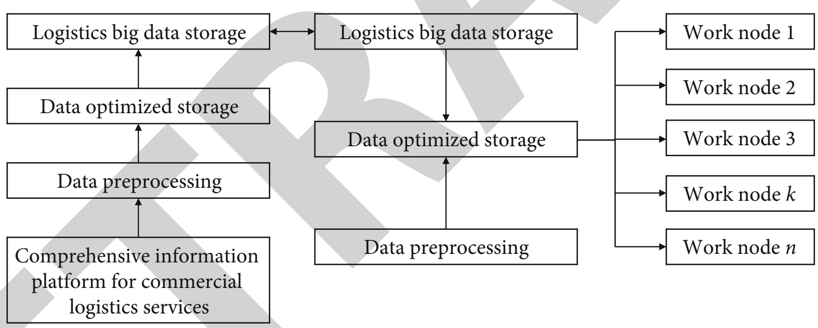


FIGURE 12: Logistics big data real-time processing system.

follows:

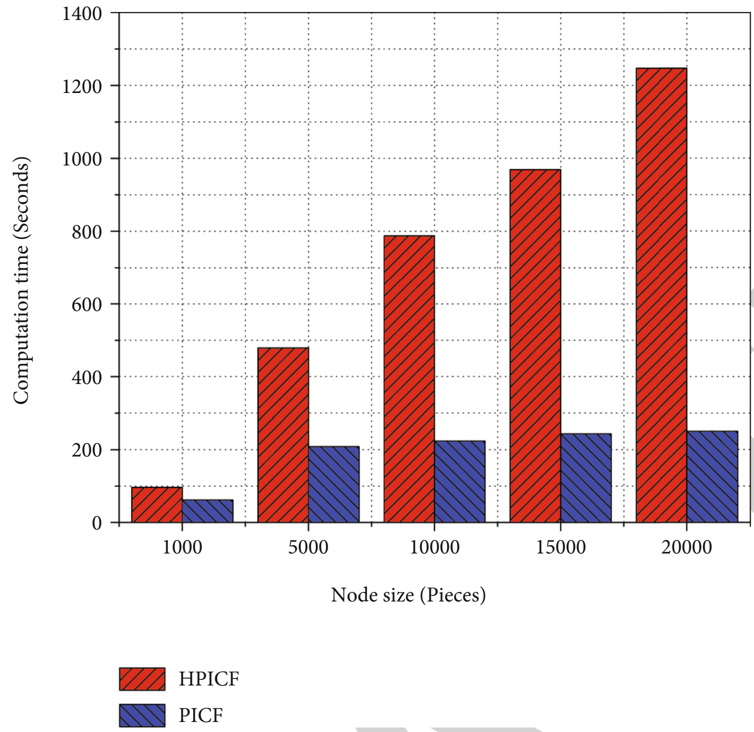
$$\mathrm{MAE} = \frac{\sum_{u \in U} \sum_{i \in R_{\mathrm{test}}(u)} \left| \mathrm{pre}(u, i) - r_{ui} \right|}{\sum_{u \in U} \left| R_{\mathrm{test}}(u) \right|}. \tag{19}$$

For the value of this parameter, an experimental method is used to explore the optimal value. As shown in Figure 7, the abscissa is the popularity threshold, and the ordinate is the average error of recommendation. The prediction error of different parameter values is verified by continuously increasing the value of the popular threshold.

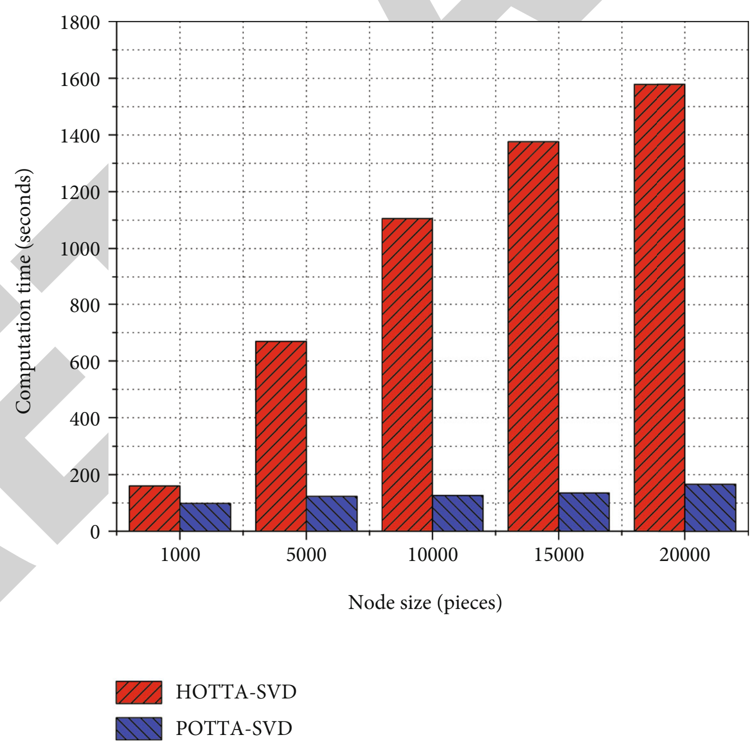
As shown in Figure 8, the user-resource matrix [35], user-label matrix (UT), and resource-label matrix (RT) in the original data set are compared with the sparsity of the optimized label matrix (the matrix sparsity is the null value in the matrix. The ratio of the number of elements to the total number of elements in the matrix). The comparison found that after label optimization, the sparsity of the three matrices was reduced, especially the user-label matrix and the resource-label matrix. This shows that label optimization can play a role in solving data sparsity from another perspective.

Figure 9 lists the influence of the learning rate parameter α in various recommendation algorithms on the recommendation accuracy. When the value of α is small, the MAE of the four algorithms changes quickly and is relatively low because the gradient descent method will quickly fall into the local optimal value in the iteration. Gradually increasing, the algorithm OTTA-SVD always maintains a better effect on different α values, and compared with tag-SVD, it can be seen that the recommendation accuracy is improved after optimizing the tags.

Figure 10 lists the effect of regularization coefficient β on recommendation accuracy in various algorithms. With the increase of β, the MAE values of the four recommendation algorithms all show a trend of first decreasing and then increasing, and OTTA-SVD has a better recommendation effect than the other three algorithms.



(a) Two parallel versions of ICF

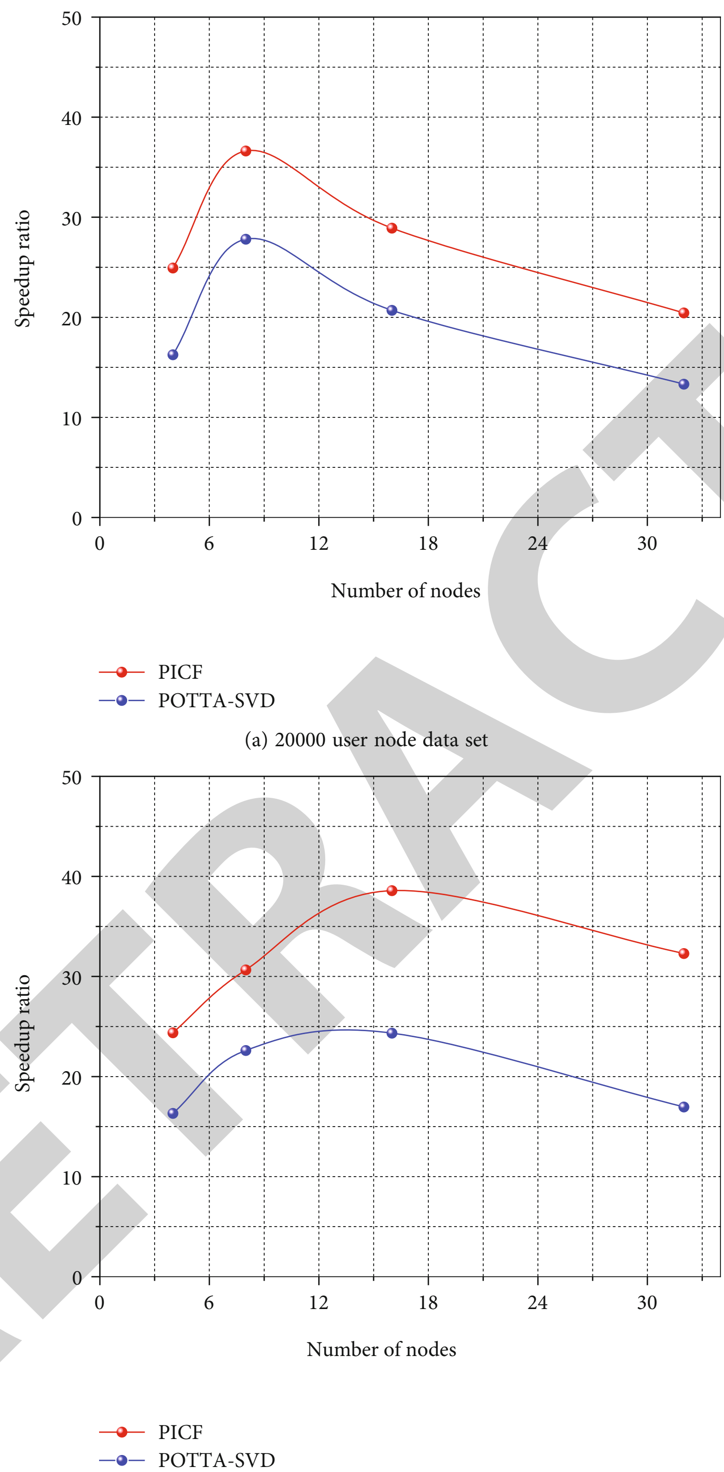


(b) Two parallel versions of OTTA-SVD

Figure 13: Comparison of different parallel performances of recommendation algorithms in different scale data sets.

Figure 11 lists the influence of the number of latent features F on the recommendation accuracy in various recommendation algorithms. Simulation experiments show that the recommendation effect of the OTTA-SVD algorithm is the best.

In conclusion, the OTTA-SVD algorithm can not only optimize the label data but also reduce the data sparsity and the overfitting problem of the recommendation model, and the results on the real data set show that OTTA-SVD has better recommendation accuracy. The



(a) 20000 user node data set

(b) 40000 user node data set

Figure 14: Performance comparison of two Spark parallel algorithms on different node clusters.

research of this chapter can improve the efficiency of distribution in logistics enterprises. In addition, collecting customer feedback information can improve the service quality of the distribution company, enhance the brand effect of the company, and accelerate the development of the company.

4.2. Method Improvements. To promote the computing performance of the recommendation model, this section studies the implementation of the parallel algorithm based on the existing matrix factorization recommendation algorithm and improves the performance by implementing the parallel algorithm of the memory iterative computing mode in the distributed cluster. The system consists of three parts: data collection, data storage, calculation, and data analysis, as shown in Figure 12.

To further improve the efficiency, the parallel transplantation of OTTA-SVD to the Spark system is carried out to improve the performance of the recommendation algorithm in logistics big data applications. First, use elastic data sets to realize parallel segmentation of massive data, and then, use parallel operation functions in Spark to describe complex parallel operation logic. Encapsulate different user data and user's scoring data on logistics services into elastic data set RDD, and then, implement parallel operations such as Map, Filter, and Union on different RDDs.

Figure 13 provides the performance of the Hadoop and Spark versions of the recommendation algorithms ICF and OTTA-SVD on various data scales. The abscissa is the number of users, reflecting the user size in the recommendation algorithm; the ordinate is the calculation time of the recommendation task. Since the calculation time is shorter, the calculation performance is better. It can be seen that the Spark parallel algorithm has obvious advantages.

To verify the effect of scalability on speedup in the Spark cluster, select the Spark version of the two recommended algorithms, OTTA-SVD and ICF, and use 20,000 and 40,000 user data nodes, respectively, to implement the experiment and calculate the parallel speedup ratio of 4, 8, 16, and 32 nodes to analyze the cluster scalability. The result is shown in Figure 14.

In the process of operation, logistics information platform generates huge amount of business data. The logistics distribution recommendation service is obtained based on the calculation of massive data, and the core algorithm recommended by this study is used to help enterprises quickly select the distribution solution suitable for user needs. This section analyzes the advantages and disadvantages of logistics service recommendation models and parallel distributed algorithms and discusses the related technologies for business data analysis, including big data frameworks, key data storage systems, and Spark and Hadoop. It is found that Hadoop and Spark big data computing frameworks are good for analyzing massive amounts of data. However, the recommendation algorithm involves a large number of iterative operations, and the way the intermediate data are stored affects the performance of the parallel algorithm. Spark parallel processing algorithms achieve better parallel acceleration performance at large data set sizes, which is about 10 times more efficient than Hadoop. In addition, the recommendation algorithm is very sensitive to the basic configuration parameters of data nodes. Coping with different data sizes requires constant experimentation to determine the configuration parameters. This is a shortcoming of the current study and will be investigated in the future.

5. Conclusion

With the development of commercial business, the calculation and analysis of big data have become a popular application. In the application field of business information management, corresponding theoretical models and algorithms are urgently needed to deal with the increasing data information. Combined with actual logistics projects, this paper studies three aspects of big data storage, calculation, and analysis. To optimize the storage performance and data security of logistics big data, theoretical models and key technologies are studied. Based on data storage, effective theoretical models, strategies, and key algorithms are proposed for the logistics distribution calculation and shared transportation problems. To further improve the application value of the logistics data analysis algorithm, a distributed parallel algorithm is implemented based on the original algorithm, the performance of the algorithm is deeply analyzed, and the key issues such as parameters are studied and adjusted. The above research can not only be applied to the data centre construction of large and medium-sized logistics enterprises but can also realize low-cost storage of logistics trajectory big data information and video surveillance data and guide logistics distribution and transportation schedule.

Firstly, based on the multilevel hybrid storage system with SSD as the cache layer, the optimization scheme to improve data access performance is studied. Then, from the perspective of data security, the multilevel hybrid storage model is studied, and a data security storage model suitable for a multilevel hybrid storage system is proposed. Finally, to optimize the distribution service in the logistics industry, based on the traditional matrix decomposition recommendation idea, a logistics distribution recommendation model is proposed. After optimization, it is incorporated into social networks and has a recommendation effect. The research results have achieved a good simulation recommendation effect. Based on this research achievement, the big data analysis research was carried out, the big data distributed computing platform based on the open-source software Spark was established, the typical recommendation algorithm was transplanted into the big data environment, and the key technical problems were solved. The parallel implementation scheme of the big data recommendation algorithm has certain universality and can be applied to various typical machine learning and data analysis algorithms. The business management system of this study can not only be used in logistics and transportation but can also provide a reference for other enterprises' business information management.

Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] Y. Yang and M. Chen, "Intelligent informatization early warning analysis of agricultural economy based on support vector sequential regression model," *Journal of Mathematics*, vol. 2021, Article ID 6334444, 9 pages, 2021.

[2] X. Wang and H. Wang, “A study on sustaining corporate innovation with E-commerce in China,” *Sustainability*, vol. 11, no. 23, p. 6604, 2019.

[3] Z. Li, P. Li, X. Zhao, and Z. Tu, “Business strategy and environmental information disclosure quality: empirical evidence from Chinese heavy pollution listed firms,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 14, p. 8325, 2022.

[4] M. Wu and Z. Lv, “Optimization of E-commerce supply chain management process based on internet of things technology,” *Complexity*, vol. 2021, Article ID 5569386, 12 pages, 2021.

[5] Y. Shi and D. Liu, “Relationship between urban new business indexes and the business environment of Chinese cities: a study based on entropy-TOPSIS and a Gaussian process regression model,” *Sustainability*, vol. 12, no. 24, article 10422, 2020.

[6] X. Li, Y. Zhang, Y. Li, K. Yu, and Y. Du, “Study of E-business applications based on big data analysis in modern hospital health management,” *Information Systems and e-Business Management*, vol. 19, no. 2, pp. 621–640, 2021.

[7] L. Wei and P. Wang, “Risk management of data flow under cross-cultural English language understanding,” *International Journal of System Assurance Engineering and Management*, vol. 28, 2022.

[8] S. Suoniemi, L. Meyer-Waarden, A. Munzel, A. R. Zablah, and D. Straub, “Big data and firm performance: the roles of market-directed capabilities and business strategy,” *Information & Management*, vol. 57, no. 7, article 103365, 2020.

[9] S. Weiwei and N. Jan, “Research on the integration of preschool language education resources based on metadata storage,” *Journal of Mathematics*, vol. 2022, Article ID 4802381, 8 pages, 2022.

[10] H. Wang, X. Ji, Z. A. Page, and J. L. Sessler, “Fluorescent materials-based information storage,” *Materials Chemistry Frontiers*, vol. 4, no. 4, pp. 1024–1039, 2020.

[11] Y. Liu, K. Zhao, Y. Ren et al., “Highly plasticized lanthanide luminescence for information storage and encryption applications,” *Advanced Science*, vol. 9, no. 7, article e2105108, 2022.

[12] X. Li, W. Zhang, X. Zhao, W. Pu, P. Chen, and F. Liu, “Wartime industrial logistics information integration: framework and application in optimizing deployment and formation of military logistics platforms,” *Journal of Industrial Information Integration*, vol. 22, article 100201, 2021.

[13] Z. You, C. Wu, L. Zheng, and L. Feng, “An Informatization scheme for construction and demolition waste supervision and management in China,” *Sustainability*, vol. 12, no. 4, p. 1672, 2020.

[14] Y. Song, G. Yeung, D. Zhu, L. Zhang, Y. Xu, and L. Zhang, “Efficiency of logistics land use: the case of Yangtze River Economic Belt in China, 2000-2017,” *Journal of Transport Geography*, vol. 88, article 102851, 2020.

[15] M. Hehua and G. Shi, “Application of passive wireless RFID asset management in warehousing of cross- border E-commerce enterprises,” *Journal of Sensors*, vol. 2021, Article ID 6438057, 12 pages, 2021.

[16] Y. Zhang, F. Rong, and Z. Wang, “Research on cold chain logistic service pricing—based on tripartite Stackelberg game,” *Neural Computing and Applications*, vol. 32, no. 1, pp. 213–222, 2020.

[17] K. Zheng, Z. Zhang, and B. Song, “Retracted: E-commerce logistics distribution mode in big-data context: a case analysis of JD.COM,” *Industrial Marketing Management*, vol. 86, pp. 154–162, 2020.

[18] O.-P. Brunila, V. Kunnaala-Hyrkki, and T. Inkinen, “Hindrances in port digitalization? Identifying problems in adoption and implementation,” *European Transport Research Review*, vol. 13, no. 1, 2021.

[19] V. Paulauskas, L. Filina-Dawidowicz, and D. Paulauskas, “Ports digitalization level evaluation,” *Sensors*, vol. 21, no. 18, p. 6134, 2021.

[20] P. B. Condon and W. G. Pothier, “Advancing data literacy: mapping business data literacy competencies to the ACRL framework,” *Journal of Business & Finance Librarianship*, vol. 27, no. 2, pp. 104–126, 2022.

[21] S. J. Miah, I. Solomonides, and J. G. Gammack, “A design-based research approach for developing data-focussed business curricula,” *Education and Information Technologies*, vol. 25, no. 1, pp. 553–581, 2020.

[22] C.-C. Qi, “Big data management in the mining industry,” *International Journal of Minerals, Metallurgy and Materials*, vol. 27, no. 2, pp. 131–139, 2020.

[23] M. Du, “Application of information communication network security management and control based on big data technology,” *International Journal of Communication Systems*, vol. 35, no. 5, 2022.

[24] Z. Cong, X. Luo, J. Pei, F. Zhu, and Y. Zhang, “Data pricing in machine learning pipelines,” *Knowledge and Information Systems*, vol. 64, no. 6, pp. 1417–1455, 2022.

[25] C. Hongsong, Z. Yongpeng, C. Yongrui, and B. Bhargava, “Security threats and defensive approaches in machine learning system under big data environment,” *Wireless Personal Communications*, vol. 117, no. 4, pp. 3505–3525, 2021.

[26] P. Zhang, “Image data security mechanism based on the Internet of Things cardiac catheterization laboratory information management system research and design,” *Journal of Healthcare Engineering*, vol. 2021, Article ID 5592185, 14 pages, 2021.

[27] T. Ou, X. Cai, M. Wang, F. Guo, and B. Wu, “A novel method of clinical nursing under the medical internet of things technology,” *Journal of Healthcare Engineering*, vol. 2021, Article ID 2234457, 10 pages, 2021.

[28] J. Yang, Y. Li, Q. Liu et al., “Brief introduction of medical database and data mining technology in big data era,” *Journal of Evidence-Based Medicine*, vol. 13, no. 1, pp. 57–69, 2020.

[29] X. Zhao, “Analysis and strategies of Internet + taxation risk management of listed companies in the big data era from the organizational psychology perspective,” *Frontiers in Psychology*, vol. 13, article 904417, 2022.

[30] H. S. Munawar, F. Ullah, S. Qayyum, and D. Shahzad, “Big data in construction: current applications and future opportunities,” *Big Data and Cognitive Computing*, vol. 6, no. 1, p. 18, 2022.

[31] A. Persaud, “Key competencies for big data analytics professions: a multimethod study,” *Information Technology & People*, vol. 34, no. 1, pp. 178–203, 2020.

[32] R. Busulwa, M. Pickering, and I. Mao, “Digital transformation and hospitality management competencies: toward an integrative framework,” *International Journal of Hospitality Management*, vol. 102, article 103132, 2022.

[33] A. L. Zolkin, R. V. Faizullin, and V. V. Dragulenko, “Application of the modern information technologies for design and monitoring of business processes of transport and logistics

system," *Journal of Physics: Conference Series*, vol. 1679, no. 3, article 032083, 2020.

[34] Q. Li, G. Wu, and S.-B. Tsai, "ERP system in the logistics information management system of supply chain enterprises," *Mobile Information Systems*, vol. 2021, Article ID 7423717, 11 pages, 2021.

[35] C. F. Durach, T. Blesik, M. Düring, and M. Bick, "Blockchain applications in supply chain transactions," *Journal of Business Logistics*, vol. 42, no. 1, pp. 7–24, 2021.

[36] A. Vafadarnikjoo, H. Badri Ahmadi, J. J. H. Liou, T. Botelho, and K. Chalvatzis, "Analyzing blockchain adoption barriers in manufacturing supply chains by the neutrosophic analytic hierarchy process," *Annals of Operations Research*, vol. 54, 2021.