

Java Data Structure Optimization of Talent Training Performance Evaluation System Based on Distributed Hierarchical Data Clustering Algorithm

Chen Jinju
Hangzhou Polytechnic
Hangzhou, 311402, China

Abstract: In the implementation scheme, the data source uses the message middleware Kafka, and the intermediate results are stored in the memory database Redis through serialization technology. Multi-faceted experimental results and analysis show that social environmental factors, social value orientation utilitarian. Based on this, the following main countermeasures are put forward to enhance the endogenous motivation of applied talents to cultivate college students: to stimulate their own endogenous motivation from the student level, to improve their own subject consciousness, to adjust their own need's structure, and therefore to teach mixed data structure based on Java. Demonstration system for optimized design. First, we should optimize the design of the framework of the hybrid data structure teaching demonstration system, and then optimize the design of the teaching demonstration database, that is, optimize the design of the database tables.

Keywords: Java Data Structure Optimization, Talent Training, Performance Evaluation System, Data Clustering Algorithm

1. INTRODUCTION

In 2003, the famous financial writer Michael Lewis wrote "Moneyball: The Wisdom of Winning in Adversity" (Moneyball), describing the way Billy Beane, the general manager of the Oakland Athletics baseball team, used big data to gain a huge advantage. Story [1]. Recommendation engine technology was first proposed to discover user preferences through data mining algorithms, and to predict the products that users may be interested in by comparing them with product attribute information. This idea has been widely used in various e-commerce websites [2].

Special intelligence research and judgment is an important module in the Jinguan Phase II project, which refers to the process in which intelligence personnel use the model provided by the information system to analyze and judge the research and judgment tasks. The comprehensive research and judgment model is based on business experience [3]. After induction and refinement, it is divided into two major contents: enterprise theme research and judgment and commodity theme research and judgment. Conscientiously fulfill the responsibility of employee development, thoroughly implement the policy that talents are the first resource, fully implement the strategy of cultivating talents to strengthen the enterprise [4], strengthen the construction of the workforce, cultivate leadership and backup members, establish, and improve democratic management and democratic supervision of employees, and enhance team cohesion, Realize the personal value of employees [5],

Promote the majestic development of enterprises. Data mining, also known as knowledge discovery, is described as extracting implicit, potentially useful, human-understandable patterns from data [6]. Data mining improves the deep understanding, understanding and application of large amounts of original data by data owners by discovering useful new laws and new concepts. Not being able to apply what they have learned is a headache for the entire society and employers at present. Under the background of the new era development trend of "mass entrepreneurship and innovation" [7]. At the same time, the sustainable development of applied

talents with social responsibility, practical ability and innovative spirit is the most important part to support my country's economic development, and it is also the part with the least talents [8].

However, big data can only generate correct and valuable information through reasonable data mining. Before the 2016 U.S. election, Microsoft's PredictWise [9], The Upshot of The New York Times, and Princeton's Sam Wang all predicted that the probability of Trump winning on the day of the election was only around 10%, but in the end, it was Trump with 306.: 232 votes result a big win. Take Microsoft as an example [10]. In the real-time news recommendation service, user interest modeling and content mining and extraction of news texts can be done periodically offline. When users use the recommendation service, they need to process the user's interest matching with the latest news in real time. The processing speed is slow, and the recommended result delay is long [11].

Commodity information is a part of customs declaration information, which is mainly described by fields such as commodity tax number, commodity name, specification and model, and country of origin [12]. Among them, only the product tax number and the country of origin are normalized, and the product name and specification model are all unchecked strings. Performance evaluation is the most important part of the performance management system [13]. The performance evaluation work is the beginning of the entire performance management process. The performance management personnel should evaluate the evaluation object based on the business objectives and strategies of the enterprise, and then correspond to the evaluation results [14].

Applying data mining technology to e-commerce and mining these data can find out this valuable "knowledge". Based on this "knowledge", corporate users can grasp customer trends, track market changes, and make correct and targeted decisions, such as improving websites [15]. From the perspective of the current demand for talents in my country's economic and social development, undergraduate-level higher education is the priority The task is to cultivate a large

number of applied talents [16]. With the popularization of higher education, university education has achieved a leapfrog extension development, but at the same time, the connotative development has not been able to maintain a consistent pace. This is indeed the case. The current problem faced by big data is not the amount of data, but the quality of the data, big data is about the relevance of the data [17].

The greater the correlation, the higher the data quality, and the more accurate the applications based on this will be. How to establish a strong correlation of data in the data ocean and continuously optimize the algorithm and mathematical model can be described as the current big data application field key. The recommendation result has a long delay, which will seriously affect the user experience. The matching process between massive news texts and users is the main reason for performance loss: the information is redundant, and the distribution is disordered in the case of unprocessed news text data.

2. THE PROPOSED METHODOLOGY

2.1 The Distributed Hierarchical Data Clustering Algorithm

The two core parts of the FP-Stream algorithm are the global frequent pattern tree Pattern-Tree shown in Figure 3.1 and the skewed time windows embedded in the pattern tree. The frequent pattern tree reflects the frequent patterns in the current time window, and its structure is like FP-tree, which is conducive to subsequent reuse of the FP-growth algorithm. The system can use Storm-based distribution based on user behavior records and expert outfit data. The parallelized stream association algorithm DPFP-Stream algorithm mines the recommendation results of related products; All kinds of data involved in the system are reasonably designed and stored in the database.

Involving millions of products, tens of millions of users, and purchase data of 100 million, it is necessary to achieve low latency, load balancing, and accurate and efficient prediction. Mechanical word segmentation is also called word segmentation based on string matching, and its principle is to match text with existing dictionaries one by one. The advantages of this type of algorithm are low time complexity and fast word segmentation. The disadvantage is that the dependence on the dictionary is stronger, and the effect is not good when dealing with ambiguous words. Haiyan Power Supply Department consists of General Office, Finance Department, Electricity Section, Emergency Repair Section, Power Distribution Section, Engineering Section, Metrology Section, Low Voltage Section, Business Hall, Marketing and Scheduling and two power supply offices. The total number of employees has reached people, including regular employees, employed employees, and rural power workers. The clothing shopping recommendation system needs to realize the following four main parts: the realization of the clothing shopping website, providing users with a platform for purchasing goods.

The implementation of the data source generation module is the preparation for generating the recommended results. To adapt to the application scenarios of high throughput and high concurrency, the system uses Kafka to implement this module. The previous text clustering algorithms are usually only based on text feature for clustering. Clustering based on text content, that is, clustering based on literal similarity, usually introduces more noise, and cannot identify synonyms,

resulting in low similarity of text objects and too many clustering results.

2.2 The Talent Training Performance Evaluation System

Dendrogram itself is also a tree, and the mathematical meaning of its branch nodes is when two nodes are merged at layer d of Dendrogram. Openness principle: The assessment may be based on different periods, and its focus will be different. The manager should explain the criteria, procedures, methods, time, and other matters of the assessment to the assessed and inform the assessed about the focus of the assessment as soon as possible.

College students need to have correct self-awareness and self-cognition ability, which is the premise for college students to form the subject consciousness of applied talents training in colleges and universities, so that college students can understand their own subject status, ability, and value in the process of college application-oriented talents training and even in the future development. Compared with the traditional teaching demonstration system, this teaching demonstration system has more stable system performance, faster teaching demonstration speed, more comprehensive database, and more excellent overall performance. The system's clothing shopping website is developed in MyEclipse using Java language, the server uses Tomcat, and the client uses the Web terminal to display.

The clothing website mainly implements five modules: registration, login, viewing products, adding to shopping cart, and purchasing. In the application of personalized news recommendation engine, the news clustering system belongs to the background processing process of the system and does not need to interact with users., just interact with the data in the background database. The location of the clustering system in the news recommendation service architecture is shown in the figure. For hierarchical clustering, whether it is a single-connection, fully connected or average-connected hierarchical clustering, when each layer of clustering is calculated, it is necessary to measure the distance between all categories to find the categories that meet the merge requirements. Any division of data points will cause the loss of distance information between categories. Use interviews and questionnaires to collect information and data and use brainstorming and fishbone diagrams to analyze the information flow and data flow in the work. Sort out and find problems in existing performance evaluation programs and their causes.

2.3 The Java Data Structure Optimization of Talent Training Performance Evaluation System

It is mainly used to store data about individuals and groups, business departments and enterprises. The optimization design of this hybrid data structure teaching demonstration system mainly involves the optimization of the teaching demonstration database. Database optimization is a hierarchical clustering constructed for the system application environment, whether it is a single connection, a complete connection, or an average connection. When each layer of clustering is calculated together, it is necessary to measure the distance between all categories and find the category that meets the merge requirements. For any division of data points, the distance information between categories will be lost.

Participate in the process of cultivating applied talents with a proactive and subjective attitude. Of course, it is necessary for college students to actively enrich their own knowledge and experience and expand and supplement their own cognitive scope. For the item_id of the product to be predicted, first check whether there is an item_id of the product to be predicted from the item Recommend in Redis. The first three items are returned; otherwise, the similarity between the word segmentation result terms of item_id and the terms of the key in item Recommend is compared. If the word segmentation results are the same, it reaches more than 80%. The functional test environment of the system is shown in the figure, and the experimental database is deployed on the data mining server. On the other hand, if the news clustering system department is on a single station in the same local area network, the data exchange between the two, including the original news data and the generated clustering results.

The risk analysis model in the special research and judgment of customs intelligence is a data model for grouping statistics of customs declaration data for enterprises, commodities, personnel, and other special topics. Aiming at the problems existing in these data models and the status quo of customs commodity data, this paper designs a short text clustering system based on the MapReduce framework. Carry out safeguard measures such as performance feedback and appeals to promote the smooth operation of the optimized performance evaluation system. Or provide preferential policies to retain high-churn customer groups, etc. The data required for data mining of e-commerce system are mainly seven contents, seven structures, dead usage records, background information of customers, transaction data, query information and so on. These data have the characteristics of distribution, heterogeneity, sparsity, high dimensionality, and mass

3. CONCLUSIONS

For the intermediate results generated by the DFPF-Stream algorithm, a Redis-oriented serialization algorithm is designed, which not only ensures data security, but also reduces network transmission overhead, and its efficient read and write efficiency is verified through experiments. Many large enterprises have begun to deploy their applications on the Java platform and introduced service-oriented software architecture into the construction of data mining systems. At the same time, the key technology of distributed data mining method in e-commerce is studied and Java also has a very broad development prospect in the computer field. The optimized design of the mixed data structure teaching demonstration system based on Java can fully develop the functions of the teaching demonstration system and fully demonstrate its advantages.

4. ACKNOWLEDGEMENT

Guangxi Educational Science "14th Five-Year" Plan 2022 Educational Evaluation Reform Special Key Project "Research on the Construction of Higher Vocational Colleges' Internal Education and Teaching Quality Assurance Evaluation System", project number: 2022ZJY681

5. REFERENCES

- [1] Bao Jiayang. Network View Optimization of SDN Distributed Controller Based on Hierarchical Clustering Algorithm [D]. Jilin University.
- [2] Zhang Yang, Mao Guojun. Hybrid Clustering Algorithm for Distributed Data Streams [J]. Microcomputer Information, 2011, 27(11):3.
- [3] Lv Lin. Hierarchical clustering algorithm based on ant colony optimization and its application in network forensics [D]. Shandong Normal University.
- [4] Ma Jing, Li Li. Distributed K-means Algorithm for Extended Index Layer Optimization on RDD [J]. Computer Engineering and Applications, 2019, 055(001):161-167,270.
- [5] Chen Xu. A log-based distributed incremental clustering algorithm [D]. University of Electronic Science and Technology of China.
- [6] Zhang Zhenya, Cheng Hongmei, Wang Jin, et al. Research on Matrix Storage Data Structure for Agglomerative Hierarchical Clustering Algorithm [J]. Computer Science, 2006, 33(1):4.
- [7] Ding Xuejun, Yang Yan, Yang Kejian, et al. Research on the application of attribute-based clustering algorithm in doctor's medical quality evaluation system [J]. Computer Application Research, 2005, 22(3):4.
- [8] Guo Wenjuan. K-means Clustering Algorithm Based on Optimizing Initial Clustering Centers [J]. Science and Technology Wind, 2022(4):63-65.
- [9] Meng Lili. Research on efficient clustering algorithm for large-scale data based on partition method [D]. Yanshan University.
- [10] Xu Yong. Design of Performance Evaluation System Based on Improved Fuzzy Algorithm [J]. Computer Programming Skills and Maintenance, 2016(21):3.
- [11] Zheng Wenjuan. Research and application of improved fuzzy C-means clustering algorithm based on Weka platform [D]. Taiyuan University of Technology.
- [12] Wang Lei, Sun Jinhai, Li Tuojuan. Research on the Optimization of Fitness Crowd Classifiers Based on Support Vector Machines and Random Forests [C]// The 11th National Sports Science Conference Abstracts Collection. 2019.
- [13] Yan Zhipeng. Particle swarm clustering optimization algorithm based on multi-task collaboration [J]. Modern Computer (Professional Edition), 2021, 000(019):32-40.
- [14] Li Sa. Data Stream Soft Clustering Theory and Its Application in Gas Disaster Early Warning [D]. Liaoning University of Engineering and Technology, 2014.
- [15] Mao Guojun, Cao Yongcun. Distributed data stream clustering model and algorithm based on data summary description [J]. Computer Science, 2013, 40(6):6.
- [16] Yang Liu, Zhong Cheng, Lu Xiangyan. Research and Analysis of Distributed Clustering Algorithm Based on P2P Network [J]. Microelectronics and Computer, 2009(8):83-85.