

Analysis of Electronic Commerce Data Mining Technology Based on Cloud Computing

Kai Xiao

Wuhan Railway Vocational College of Technology, Wuhan, Hubei, China

wtzyxk@126.com

Keywords: Cloud Computing, Electronic Commerce, Data Mining

Abstract: EC(Electronic Commerce) is the engine of the future economy and the best way for enterprises to show their images and products to the world in an all-round way, find partners and expand sales scale. Effective organization and utilization of EC information enables customers to find truly valuable information and knowledge from a large amount of complex information. DM(data mining) technology originates from data warehouse technology. Nowadays, the information structure of EC is developing towards the direction of large amount of information and complex information structure. The huge amount of commodity information buries the available information, which leads to the lower utilization efficiency of commodity information by EC enterprises and makes it difficult to develop relationships with customers. In this paper, the massive data analysis solution based on cloud computing (CLO) platform is applied to EC. The super-large server cluster of CLO has super computing power, and cloud storage has powerful storage capacity, data analysis capacity and data management capacity, which together constitute a favorable foundation for the development and application of massive DM, improve the capacity of massive DM and business intelligence analysis, and achieve high-performance computing at a lower cost to meet the business needs of EC.

1. Introduction

In recent years, EC has been developing unprecedentedly. Dealers and customers have exchanged via the Internet, saving a lot of costs and time. However, EC is full of a large amount of data. How to dig out the really valuable information from these large amounts of data and help enterprise dealers develop better marketing strategies is an urgent problem for EC. EC is the engine of the future economy and the best way for enterprises to show their image and products to the world in an all-round way, find partners and expand their sales scale. How to effectively organize and utilize EC information so that customers can find valuable information and knowledge from a large number of complex information DM technology comes from data warehouse technology. It is used to process huge business data, find the underlying regularity, and model it to provide reference for decision-makers^[1]. Today, with the rapid growth of large-scale data sets, DM has become a frontier research topic in the current information field and database technology. It makes data processing technology enter a higher stage and is recognized as one of the most promising backbone technologies of future information processing. As a knowledge discovery process, the purpose of DM is to obtain "proven knowledge" from original data^[2]. It needs to extract useful, previously unknown but ultimately understandable knowledge from a large number of incomplete, noisy, fuzzy and random data.

After long-term operation, a large number of user visits and transaction records will be accumulated, which is a huge potential resource. If we can effectively mine it and find out the hidden laws, we can provide decision-makers with information of great commercial value, so that enterprises can occupy a more favorable position in the market competition^[3-4]. In order to manage mass commodity information and user information more efficiently, and improve DM and business intelligence analysis systems, this paper applies mass data analysis solutions based on CLO platforms to EC. CLO super large server clusters have super strong computing capabilities, and

cloud storage has strong storage capabilities, data analysis capabilities and data management capabilities, Together, they form a favorable foundation for the development and application of massive DM [5]. Improve the ability of massive DM and business intelligence analysis, and achieve high-performance computing at a lower cost to meet the needs of EC business.

2. Electronic commerce and data mining technology

2.1. Overview of Electronic Commerce

The English name of EC is EC, which can also be called Electronic Business. It refers to various commercial activities, trading activities, financial activities and related comprehensive service activities conducted and completed worldwide through the Internet network in the form of electronic data information circulation. The "modern information technology" included in "EC" should cover all kinds of communication methods based on electronic technology; "Commerce" refers to various matters arising from all commercial relationships, whether contractual or non contractual [6]. If "modern information technology" is regarded as a subset and "commerce" as another subset, the scope of EC should be the intersection of the two subsets, that is, the various uses of the Internet, intranet and electronic data exchange in trade that may be widely involved under the title of "EC" .

EC is a new business model. Enterprises in all industries will be connected together through the network, making it possible to cooperate with each other in real and virtual situations [7]. EC is beneficial to large enterprises and small and medium-sized enterprises. Because large and medium-sized enterprises need more trading activities.

Small and medium-sized enterprises can obtain business information equivalent to that of large enterprises through the operation process of EC, so as to improve their own transaction and management efficiency without increasing operating costs. Enterprises that have greatly improved the competitiveness of SMEs can also provide better services to customers in a new way, which is not the idea that large enterprises can achieve. The Internet provides a new development opportunity for enterprises. Any enterprise may establish a second tier relationship with suppliers or customers worldwide.

2.2. The role of Data mining in Electronic commerce

The reason why DM technology can serve EC is that it can mine the potential information in the process of activities to guide EC activities. Personalized service is divided into two stages: first, user clustering and classification, Web page clustering, frequent access path discovery, etc. Are carried out according to users' historical access records and users' personal information; Then use appropriate recommendation methods to combine user access mode requirements with request pages by mining user personalized information, and conduct Web page access prediction and product recommendation through user group similarity [8]. Generally speaking, classical DM algorithms can be directly used for Web usage mining. However, in order to improve the mining quality, researchers have made efforts to extend algorithms, including compound association rule algorithm, improved sequence discovery algorithm, etc. The EC classification diagram of Web mining is shown in Figure 1.

With the popularization of broadband Internet, EC has grown into a new economic growth point. The rich information resources in the EC field also provide a broad application platform for the application of DM, which can greatly improve the ability of enterprises to extract information [9]. The process of DM in EC generally includes three main stages: data preparation, DM, result interpretation and evaluation.

① Data preparation

Through the analysis of user needs, search and determine the target data consistent with it from the database.

② In the stage of DM, the first step is to determine the goal of DM and the type of knowledge to be mined. After the mining task is determined, the appropriate mining algorithm is selected

according to the type of mining knowledge. Finally, the DM operation is implemented, and the selected mining algorithm is used to extract the required knowledge from the database.

③ Interpretation and evaluation of results

The knowledge found in the DM stage may have redundant or irrelevant knowledge after evaluation. At this time, it needs to be eliminated, or the knowledge may not meet the needs of users, and the above mining process needs to be repeated for re-mining.

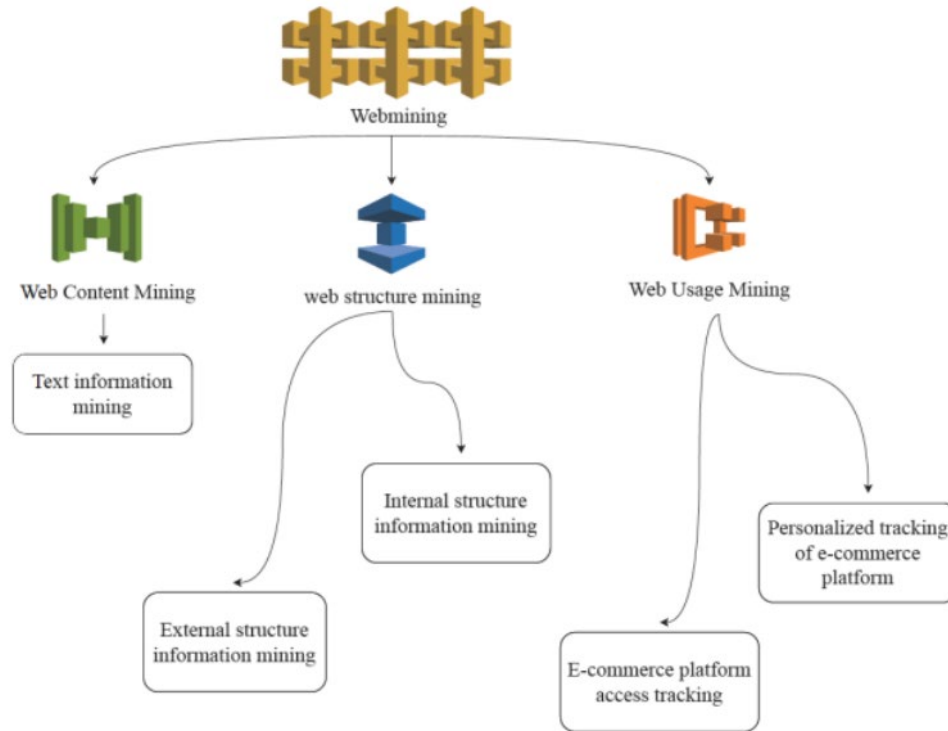


Figure 1 E-Commerce Classification Diagram of Web Mining.

3. Application of electronic commerce data mining technology based on CLO

3.1. CLO-based Electronic Commerce Data Mining Service Architecture

The DM service pattern provides EC services through composite applications, that is, applications built by data mining functions and application collections. CLO has built a network environment to realize the sharing of computer equipment, storage devices, server clusters, etc. The dynamic and scalable computing power based on CLO brings the possibility of efficient massive DM; The swarm intelligence with public participation in CLO environment provides an environment for the research of new DM methods with swarm intelligence [10]. The emergence of EC has changed the traditional business model and the relationship between sellers and customers. According to the behavior and requirements of DM, combined with the architecture and architecture of CLO, an EC service architecture based on CLO is designed, as shown in Figure 2.

The expansion of CLO based customers' choice makes them pay more attention to the value of goods, rather than considering brand and geographical factors first. Therefore, for sellers through CLO, they should try their best to understand their customers' needs, carry out targeted EC business activities, and provide customers with the goods and services they need efficiently and conveniently, so as to remain invincible in the competition. Massive DM requires high network environment and application conditions, and the application of CLO effectively improves the efficiency of EC DM.

The EC resource management system based on CLO can store a large amount of data, and combine the requirements of DM to reasonably distribute network resources, improve the extensibility of DM algorithms, and has a certain degree of fault tolerance, improving the reliability and security of DM. The data resource service provides a remote hosting database service, enabling

users to use remote EC data resources just like local databases, and provides demand driven database and data warehouse technology.

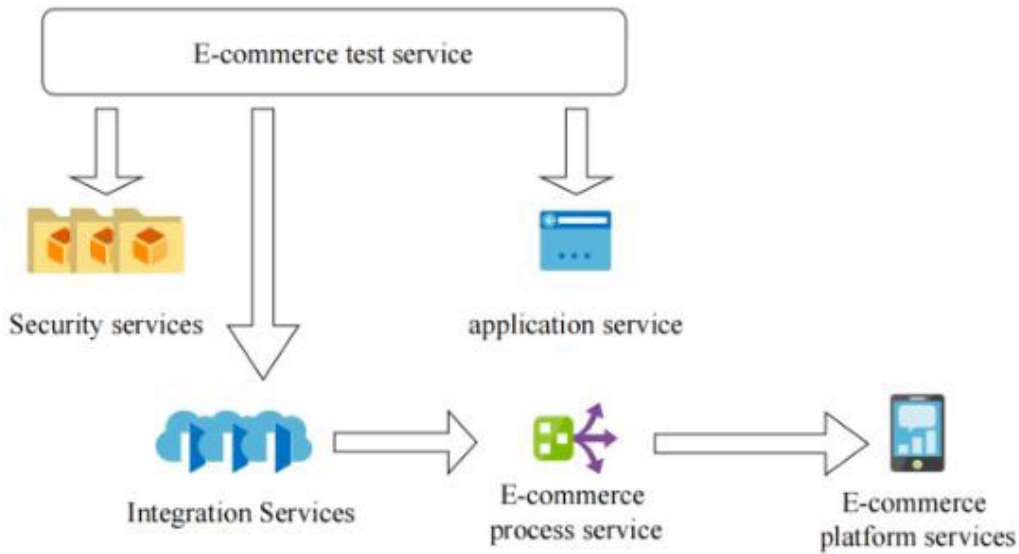


Figure 2 EC DM service architecture based on CLO.

3.2. Electronic Commerce Data Mining Service Process Based on CLO

DM is a service process from the perspective of management; From a technical perspective. Combined with the life cycle of software, the service process of CLO EC DM is discussed in the process of CLO EC distributed development. The DM service process based on CLO is defined as four stages: analysis , design, development, maintenance and decay. After defining the purpose of DM, first extract the data set of EC DM from large databases and data warehouse targets, and describe the data object with a set of basic attributes, which mainly covers the background information when customers log on EC websites and past purchase and click stream information. First, prepare the algorithm data, convert the file format of the link structure to make it more targeted, and assign values to each node. Generally, the initial value is 1-d, and the specific value of d is 0.85, as the input data for calculation. Data preprocessing is composed of massive information web nodes, and the entire node also needs to be parallelized. The file format of the link structure is shown in Table 1.

Table 1 File format of link structure

Start Node	Initial value	Target	Node
1	1-d	3	4
2	1-d	4	6
3	1-d	5	
4	1-d	3	1

The output link sequence is saved in the platform, which is an important content in the data preparation stage of the algorithm. The process of EC DM based on CLO technology is as follows: users send out DM instructions in the input module and transmit them to the EC system server. The server automatically calls out data from the database according to the user's mining instructions, and then reprocesses the data, including checking the integrity and consistency of data, dealing with noise and errors, dealing with unknown attribute values, discretization and fuzzification of numerical attributes, dealing with continuity, grouping of symbolic attribute values, attribute selection and sequencing, attribute construction and transformation, etc. At the same time, the optimal mining algorithm is called out from the algorithm library, and after preprocessing the EC data, it is transmitted to the operation module for in-depth DM.

4. Conclusions

CLO technology has the advantages of practicality, efficiency, low cost, etc., so this paper

applies CLO technology in DM. In view of the deficiencies in the research and application of existing DM services, this paper proposes DM services under CLO environment and constructs the architecture of DM services according to the information resource service model of CLO, its service-oriented architecture and CLO platform architecture , The process of creating DM service is designed, the architecture of DM service model is given, and the process of DM service is defined. In the future research, we need to build the service model of DM and implement the corresponding DM services under the CLO platform, Combined with the application of DM, the proposed model is verified. By integrating other CLO middleware, we can expand the functions of the platform, promote the development of EC applications, and further apply it to the Internet of Things to provide a high-performance, dynamically expanding recommended basic platform for EC under the three networks integration. This paper analyzes the advantages of using CLO technology to realize EC DM, investigates the current research status of EC DM technology based on CLO, summarizes the problems and challenges faced by EC DM based on CLO, and puts forward some solutions and measures. With the further development of CLO technology, it is bound to bring new breakthroughs and better technical support to EC DM.

References

- [1] Yang, ZH, Pan, et al. Research on Electronic Commerce Information Management System based on Data Mining[J]. *Agro Food Ind Hi Tec*, 2017, 28(11):17-27.
- [2] Xie C, Xiao X, Hassan D K. Data mining and application of social e-commerce users based on big data of internet of things[J]. *Journal of intelligent & fuzzy systems: Applications in Engineering and Technology*, 2020, 22(39-4 Pt.1):17-33.
- [3] Qing H, Zhang J, Fu D. Data Mining Technology in Business Data Analysis[J]. *Journal of Physics Conference Series*, 2021, 1852(2):022045-022075.
- [4] Qing H, Zhang J, Fu D. Data of E-Commerce Users Based on Data Mining Technology[J]. *Journal of Physics: Conference Series*, 2021, 1852(2):022038-022058.
- [5] Deng J, Qing H. Practice Teaching Method of Electronic Commerce Major Based on Data Mining Algorithm[J]. *Journal of Physics: Conference Series*, 2021, 1852(2):022039-022069.
- [6] Shi H, Deng H. Teaching and Experimental Schema of Data Mining Technology Combined with the Cloud Computing[J]. *Taiwan Academic Network Management Committee*, 2021, 65(1):37-74.
- [7] Shen X. Research on the Internet Data Mining Technology Based on the Cloud Computing[J]. *Basic & clinical pharmacology & toxicology*. 2019, 18(124-S1):22-26. .
- [8] Xu H, Fan G. Application of Big Data Mining Technology in Intelligent Safe Production on Cloud Computing Platform[J]. *Springer, Cham*, 2019, 36(18):26-47.
- [9] Wang Y. Research on the big data feature mining technology based on the cloud computing[J]. *International English Education Research: English*, 2019, 36(3):3-11.
- [10] Bruin B D, Floridi L. The Ethics of Cloud Computing[J]. *Science and Engineering Ethics*, 2017, 25(11): 12-32.