Personalized E-commerce Website Construction Based on Data Mining

Xin Chu¹, Dan Lv², Danni Zhao³

¹College of Management, Wuhan Donghu University, Wuhan, Hubei, China, 430212
²Correspondence author: Teaching and Research Section of International Trade and Economics of College of Economics, Wuhan Donghu University, Wuhan, Hubei, China, 430212
³Teaching and Research Section of Electronic Commerce of College of Management, Wuhan Donghu University, Wuhan, Hubei, China, 430212

Abstract. With the development of supply chain and logistics, e-commerce system can provide a growing variety of goods and quantity. The traditional website-centered retrieval business model has become a bottleneck restricting the value-added of e-commerce enterprises. Data mining technology enables people to finally understand the real value of data, that is, the information and knowledge contained in the data. For e-commerce enterprises, rich data sources and reliable data collected automatically make it easy to meet the necessary factors of data mining. Web server logs contain a large number of records of users accessing e-commerce systems, and data mining technology is used to process and analyze the data. Create a convenient, fast and personalized e-commerce website to provide a reliable analysis basis for the market positioning, strategy improvement, customer relationship improvement and service quality improvement, so as to improve the competition of enterprises in the market. Force to create better benefits. Through data mining, the analysis of user access behavior, frequency, content, etc., can extract the description characteristics of each user, so as to obtain knowledge about the behavior and manner of group user access.

1. Introduction

With the maturity of data mining technology, more and more business problems can be solved by this technology. At the same time, with the growing momentum of e-commerce development, data mining technology will have a very broad development space in the field of e-commerce [1]. In the traditional business era, people's individualized needs can not be met, and the emergence, development and popularization of the Internet is a liberation of individuals in traditional order-oriented economic organizations, so that personality and creativity play a more effective platform [2]. Highlight the advantages of personalized business information services. In order to ensure the accuracy and stability of customer types, it is necessary to first sample and count the customers with longer registration time, browsing and purchasing more products [3]. In application, it is more intuitive and more accurate. Through the big category to the small category, we can finally find the products we need, and can conveniently compare the same kind of products with enterprises and their products, so as to select the more suitable supplier of goods [4]. According to the log records classified by the user in the scan log file, the session identification of all users can be completed, and the session identification of the entire log file is completed [5]. Session recognition is the foundation and key of Web log mining. The system performance of the website has become a bottleneck affecting its quality service for customers. How to
improve the performance of the website system, optimize the structure of the website system and provide customers with better personalized services directly affects the efficiency of the business website [6].

E-commerce is the most appropriate application area of data mining technology, because e-commerce can easily meet the necessary factors of data mining: rich data sources, reliable data automatically collected, and the results of mining can be transformed into business activities, business investment can be timely evaluated [7-9]. Combining access information, reference log and site topology, we can list many paths for users to browse, if the requested page does not have a link relationship with the pages that users have browsed. Personalized e-commerce is the development of enterprise e-commerce to a certain stage. In order to attract customers to the greatest extent and maintain and enhance customer loyalty, the enterprise e-commerce website is developed and managed in depth in terms of information release and transaction mode, so that each customer can get "one-to-one" tailored services and information [10]. To meet the individual needs of consumers, but also to enhance the market competitiveness of enterprise products. At the same time, because of the long-term interaction with consumers, enterprises can timely understand the changes in market demand, targeted production, will not cause product backlog [11]. Make full use of the effectiveness of the Internet platform, no space span, no time interval and other characteristics, to provide opportunities for both parties with buying and selling needs, then the site has a lot of information, rich content, complex levels, diverse forms [12]. In the process of planning and design, the website must be analyzed for complete requirements, the purpose of website construction should be clarified, the access objects of the website and the service content and domain name that the website should provide, the icons, styles, related directories and structures of the design website should be determined. Content [13].

At present, the database technology is developing rapidly, and the database system is widely used in information management. The data information in the database is also increasing rapidly [14]. Then, how to organize and utilize the massive information on the database effectively. The mining results are presented and analyzed in a form that is easy to understand by users. After evaluation, some redundant or irrelevant patterns are eliminated. Some useful patterns may not meet the actual needs of users [15-17]. E-commerce websites involve the complexity and diversity of subject content. At the beginning of website creation, when doing website system analysis, we should consider the overall content of the website comprehensively [18]. On the premise of interactive websites and databases as supporting tools, enterprises can integrate and use personalized websites, personalized E-mail, personalized webpages, customer interest tracking and related tools, which can not only establish close and friendly contact with customers, but also create personalized marketing information [19]. Each type of commodity has its own specific user group, and different groups have different consumption habits, preferences, and spending power. In the increasingly fierce market competition, personalization, specialization and specialization occupy an increasingly important position in customer service [20]. Through the analysis and processing of the log data information, the browsing mode of the user visiting the site can be found, and the decision information used by the merchant for targeted marketing to a specific consumer group or individual can be obtained, and the relationship and timing relationship can be revealed [21]. Therefore, designing a good e-commerce website is the key to the success of e-commerce activities. By using data mining technology to reasonably mine the sales information of the website, the law of its formation and existence, as well as the rules of user access can be found [22].

2. Materials and Methods
Mining Web log files and customer transaction data to discover meaningful user access patterns and related potential user groups. Its main feature is to extract, convert, analyze and model user information data. From which key data to assist business decisions are extracted. In order to make it easier for some potential users to find our website on search engines, thus showing the advantages of products and attracting users to buy. Data will play an increasingly important role in our life, and the related data mining technology will become more practical. However, e-commerce personalized recommendation service also faces many challenges. When users visit the website, the management and maintenance of the website will follow. In fact, the management and maintenance of the website has been running
through the construction of the website. As long as the website is running, it must be managed and maintained. Correctly arrange a large number of enterprise requirements for advertising, and reasonably structure the position of enterprise advertisements on the front page, so that the advertisement position does not conflict with the business service content of the front page, and is suitable for browsing habits of viewers. In the data warehouse, a multi-dimensional data model with the themes of customers, sales and finance is established, and the data of the data warehouse are continuously refreshed on the basis of retaining the original data. On the one hand, the server extracts the customer's historical record from the storage system (database, knowledge base, information base), and on the other hand records the customer's latest access path and residence time. The server passes all the collected information to the data mining module, and the corresponding operations are performed by data mining. Analyze how many different users are accessing. The general approach is to use heuristic rules to uniquely identify users with user IP and proxy. That is, the user IP address and the agent are the same at the same time. User access modes are shown in Table 1 and Figure 1.

| Table 1. User access mode parameters |
|-------------------------------------|
| **Access**                          | **Pre-mathematics** |
| Frequent path                       | 19.71               | 16.05               |
| Maximum path                        | 17.36               | 8.33                |

Figure 1. User access mode parameters

In the aspect of website optimization design, the object of mining is the content of the website providing services, which can optimize the website structure and mine the content that users pay more attention to. In customer relationship management, data mining can be based on the analysis of Web log data. The website information is simulated with data, data is set, and part of customer information is selected from the background database of the website, including the number of customers purchased, the type of goods purchased, the browsing depth of users in the website when purchasing goods, and the return visit rate of customers. Session refers to the activities carried out by a user during a visit to a website. The task of session identification is to identify the same visit request belonging to the same user. Completing the path is to complete the access path in the user session so as to better reflect the user's access process. It is mainly used to solve the problem that the access log does not completely record the user's access behavior due to cache and other reasons. For e-commerce enterprises, it is easy to meet the necessary factors for data mining: rich data sources, reliable data collected automatically, and can translate the results of mining into personalized business practices such as personalized network marketing, business investment can be directly evaluated. Enhance the practicability of the website and meet the buyer's need for full comparison of products. This method can maximize the click rate, extend the browsing time of netizens, and increase the repeated clicks of the browsing crowd.
Due to different factors such as visitor identity, access intention, and length of access time, the management and maintenance of the website is the most complicated step. It mainly includes security management, performance management and content management. Use various channels to aggregate, organize and classify resources, and then provide personalized recommendation information to users to meet the individual needs of users. The structure of the personalized marketing service e-commerce system is shown in Table 2 and Figure 2.

Table 2. Individualized Marketing Service Electronic Commerce System Structure Reference

| Index analysis | Multidimensional analysis |
|----------------|--------------------------|
| Risk early warning | 6.08 | 7.31 |
| Management suggestion | 5.92 | 6.25 |
| Personalized Service | 5.73 | 5.19 |
| Information consultation | 6.81 | 3.92 |

Figure 2 Individualized Marketing Service Electronic Commerce System Structure Reference

E-commerce operation mode is one of many e-commerce modes. It has obvious personalized service characteristics and broad application prospects. However, it is limited by the application of data mining technology, and it can not accurately locate personalized consumer characteristics. In order to further develop rapidly, it can be combined with the specific project application practice of the model mall. Typical browsing behaviors reflecting users'interests are summarized as saving pages, printing pages, adding pages to favorites, copying page content, visiting the same page and browsing time on the page. Registered domain names, site configuration and shading, and web design and production can be carried out simultaneously. Web site testing can only be carried out after the completion of the other three items. After all the content is built, we can publish the website, and users can access it through the domain name of the website on the Internet. Left and right sidebar advertisements refer to the advertisements on the most left and right side of the web page. Generally, they use the blank areas on both sides of the web page to launch advertisements. The effect is obvious and the audience is high. Instead of focusing on technical tasks (such as coding), focus on the problem itself. Provides a variety of graphical technology, assistant to solve the key links between data, guide users to find the most convenient way to find the final solution to the problem. Data preprocessing extracts a part of the attributes that have a significant impact on the target output from a large number of data attributes, that is, reduces the dimensionality of the original data, thereby improving the quality of the data in the original information stream and improving the speed of data mining. The data mining process is shown in Figure 3. When there is a promotion or discount in the online store, you can inform the customer through the contact information left by the customer, and keep the customer's loyalty as much as possible. The depth of browsing for customers at the time of purchase is relatively large, which means that the product layout of the website is not very reasonable, and can be optimized according to the browsing records of the customers on the website.
Information generated by data collection includes user access records and information extracted from storage system, which comes from database, information base and knowledge base. The characteristics of storage system determine that it has the characteristics of small amount of data and high degree of integration. Generally speaking, pretreatment is not required. After data mining and analysis of all products on the website, cluster analysis is carried out on the sales situation of commodities in each category, so that website owners can clearly understand the main purchase sources of commodities in each category and the data such as which commodities are better sold on the website. Scanning all MFPs of each user session one by one, iteratively generating candidate sub-paths with length len, and generating frequent access paths by calculating the frequent access support of the candidate paths. It can automatically realize the whole set of data mining operations including data selection, data conversion, data mining and result presentation. If necessary, this process can be repeated for the result data set until a satisfactory result is obtained. On the front page of the e-commerce website, a special shopping guide page is regularly launched. The form of this special page can be set flexibly, but the timeliness and short-term nature of the page must be maintained. We need to seriously study users' habit of searching for keywords. For example, since each of us has different search habits, in order to balance most people, when keywords are selected, keywords need to be combined. By grasping the information of websites and products browsed by online customers, judging the product information that customers are concerned with, and quickly establishing service files for customers. In addition, when the customer exchanges new information with the customer service, the customer service personnel can store the data communicated with the customer in the database of the computer.

3. Result Analysis and Discussion

Since most e-commerce websites have a background database and the front pages are mostly generated dynamically, we can write a web crawler to simulate users' browsing and capture the final page content that users see. Through the analysis of hyphenation and page content, the purpose is to collect and analyze data clearly and turn it into useful information. From the early market research to the later follow-up service and the final processing of the whole set of operating products. Data analysis will inevitably be used in every link, and we can finally obtain a set of keywords in all documents. Focus on analyzing their characteristics, formulate more friendly promotion information according to their characteristics, promote their own products, and guide these users to consume, so as to improve the dependence of these users on products. The function of the commodity display area is to quickly capture commodity information from the database service information in a timely and dynamic manner. For example, the picture, name, price, seller, latest product, recommended product and other information of the product, especially the display of special products and the display of classified products, should be more obvious, convenient and fast. The information extracted from the repository is generally represented as a tree structure in which nodes store a summary of page information and traces left by customer visits (e.g., dwell time). Extracted from the knowledge base is the priority algorithm for data mining (for different customers, the mining algorithm has different priorities. It should not be expressed in Flash animation form, otherwise it will cause viewers to dazzle and carry content in visual effects.
Problems such as unclear performance affect the viewer's evaluation of the entire web page. Page
genericity, customer affiliation, and frequent access to paths and pages, in order to optimize the topology
of the Web site, more effectively identify market goals, improve decision-making provide help.

In Internet e-commerce, the customer's browsing information is automatically collected by the Web
server and saved in the log file. In the Internet age, the number of users and the number of commodities
are both increasing exponentially, which results in high dimensions of input information. There are still
many users who provide incomplete data information and many new projects have not been scored by
any users since they were just launched, which results in sparse input information and cold start of new
projects. The information extracted from the database is a two-dimensional table composed of attributes
and fields. The information extracted from the information base is the customer's previous access
information, including content information and structure information. Through the analysis of customer
attribute characteristics, transaction behavior and capital capability, various knowledge related to the
static and dynamic characteristics of customer transactions are extracted to make necessary
segmentation of customers. The commodity inquiry area, where you can quickly inquire about the
commodities the customer wants according to the customer's selection, classification and screening type.
The design of the shopping cart, when the customer service selects a certain commodity, immediately
starts the shopping cart function, quickly and conveniently adds to the shopping cart, and quickly
completes the order. When the customer deletes a product in the shopping cart, the order cancellation
function is completed in time. To act as a viewer, you will know what kind of website will make viewers
more like. With this kind of thinking as a guide to optimize the website, we can effectively improve the
loyalty of the viewer. Increase your understanding of the product, learn about the performance of the
product, and the virtual experience of online transactions. These visitors will behave in a behavioral
mode that likes to browse product-related information, such as other people's comments on the product,
and other customers' perceptions of the product.

Because the products of e-commerce websites will remain basically unchanged for a period of time,
this part of the workload in addition to the initial keyword extraction and content mining, the rest of the
time does not have a lot of workload. Every time new products are added, the original log files are
pretreated and transformed into data forms suitable for mining. Then the data mining algorithm in pattern
recognition is used to mine, and finally the mining results are summarized and applied to practice. It
also calculates the different weights of the newly added product page according to the pre-calculated
keyword list, and constructs the weight vector of the newly added product for the keyword list. More
accurately find the nearest neighbor set, but also alleviate the problem of cold start of new projects.
Compared with the traditional collaborative filtering recommendation architecture, it alleviates the
problems of low real-time, sparse data and cold start of new projects through user-based mining, project-
based mining optimization and two similarities. Through a variety of communication channels to obtain
a series of services to participate in customer surveys, direct contact with customers, give attention and
guidance at every step of the consumer shopping process, strengthen the relationship between the system
and corporate customers, and improve consumer satisfaction. And loyalty. Divide and classify
customers to analyze the common characteristics of customers in the group, and provide high-quality
personalized services for the corresponding customers, so that customers can be targeted when browsing
information, saving time cost, spiritual cost, physical cost, etc., and finally making the total customer
The cost is reduced. Dynamically detect patterns that were not previously discovered. Using data mining
technology, the e-commerce system is able to scan all data and identify hidden patterns. For example,
by analyzing retail data to identify products that appear to be unconnected.

Compared with the nearest neighbor query efficiency based on user partition data mining and the
best clustering number, the nearest neighbor query efficiency compared with the collaborative filtering
algorithm based on user partitioning data and the collaborative filtering algorithm based on user mining
is shown in Figure 4.
Figure 4. Nearest Neighbor Query Efficiency of User-based Partition Mining Improvement Scheme

When the data mining tool runs on a high-performance parallel processing system, it can analyze a very large database in a few minutes. This faster processing speed means that users have more opportunities to analyze data, making the analysis results more accurate, reliable and easy to understand. Mining pattern can be used for customer clustering. Since it recommends resources according to similar customers, i.e. predicts current customers' interests according to the most similar neighbors, it is possible to recommend new interesting contents for potential customers. Dynamic category information management, customer service category management and content management; Member management, order management and email notification of new orders; The functions of mailing list, sending and template management; Management and classification of feedback information, comments on commodities, stock-out registration of commodities, etc. Detailed access statistics. It plays an important role in carrying out accurate user information push service of websites and individual specific targeted services, so that users can feel good user experience with different personality situations and different treatment, thus effectively improving the recognition among users. If a page has a high weight in the overall user of the website, that is, the "word of mouth" in the user group of the website is quite good, even if the performance of the page in the user mining is not good enough, we should consider recommending it. Therefore, when recommending the page, two factors should be considered together. In the aspect of collaborative mining based on user mining, user-item scoring matrix and user-project type preference matrix are used as the common reference basis. Based on collaborative mining of project mining, data sparsity is well compensated by data preprocessing. The lack of, and improvement based on the similarity of the project.

4. Conclusion
This paper studies the construction of personalized e-commerce websites based on data mining. The selection of initial clustering centers based on user mining is optimized, and the distance similarity formula is optimized, so that users with similar preferences are divided into the same class cluster, thus greatly reducing the scope and improving the retrieval speed. Through data mining and analysis of user's access behavior, frequency and content, the descriptive characteristics of each user can be extracted and knowledge about group user's access behavior and mode can be obtained. Through understanding and analysis of these user characteristics, targeted e-commerce activities can be carried out. Visits refer to the classification of websites visited by customers, and the accuracy rate of comprehensive visits is lower than that of single visits, which is due to too few visits by customers in system tests. The calculation shows that the accuracy rate increases with the number of valid customer visits. When we use these different indicators to measure, we can see from a multi-angle perspective whether the positioning, content, functions, etc. of the website can cause customers' interest points, thus reflecting the shortcomings of the website. The information in the original database system placed on the Web server as a common resource can be shared with multiple enterprises without any influence on each other. By discussing data mining tools and their implementation models, combined with the
analysis of the realization goals of personalized marketing, in order to enable e-commerce enterprises to make full use of the rich underlying data sources, and further excavate on the basis of mining the largest forward path. The user's frequent access path, find out the user's shopping characteristics and habits, achieve the purpose of recommending products to users, and truly integrate data mining into business decisions.

Acknowledgements

This work was supported by the grants from Youth Foundaion WuHan Donghu University (2016) and the grants from Hubei Provincial Collaborative Innovation Centre of Agricultural E-Commerce(Wuhan Donghu University Research [2015] No. 11 Document).

References

[1] Hong H,Tsangaratos P,Ilia I,et al. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China[J]. Science of The Total Environment, 2018, 625:575-588.

[2] Elsabahy M,Wooley K L . Data Mining as a Guide for the Construction of Cross-Linked Nanoparticles with Low Immunotoxicity via Control of Polymer Chemistry and Supramolecular Assembly[J]. Accounts of Chemical Research, 2015, 48(6):1620-1630.

[3] Cao J,Yu X,Zhang Z . Integrating OWA and data mining for analyzing customers churn in E-commerce[J]. Journal of Systems Science and Complexity, 2015, 28(2):381-392.

[4] Elwakil E,Zayed T . Construction knowledge discovery system using fuzzy approach[J]. Canadian Journal of Civil Engineering, 2015, 42(1):22-32.

[5] Heredia D,Amaya Y,Barrionto E . Student Dropout Predictive Model Using Data Mining Techniques[J]. IEEE Latin America Transactions, 2015, 13(9):3127-3134.

[6] Wang J H,Zhao L F,Lin P,et al. GenCLiP 2.0: a web server for functional clustering of genes and construction of molecular networks based on free terms[J]. Bioinformatics, 2014, 30(17):2534-2536.

[7] Williams T P,Gong J . Predicting construction cost overruns using text mining, numerical data and ensemble classifiers[J]. Automation in Construction, 2014, 43:23-29.

[8] Li Z,Xu W,Zhang L,et al. An ontology-based Web mining method for unemployment rate prediction[J]. Decision Support Systems, 2014, 66:114-122.

[9] Kaplan S E,Nischwietz R J . A Web Assurance Services Model of Trust for B2C E-Commerce[J]. International Journal of Accounting Information Systems, 2003, 4(2):95-114.

[10] Sharma G,Lijuan W . The effects of online service quality of e-commerce Websites on user satisfaction[J]. The Electronic Library, 2015, 33(3):468-485.

[11] Rizoiu M A,velcin J,Lallich, Sté phane. Unsupervised feature construction for improving data representation and semantics[J]. Journal of Intelligent Information Systems, 2013, 40(3):501-527.

[12] Rubiano S M M,Garcia J A D . Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance[J]. IEEE Latin America Transactions, 2016, 14(6):2783-2788.

[13] Zhang L,Luo P,Tang L,et al. Occupancy-Based Frequent Pattern Mining *[J]. Acm Transactions on Knowledge Discovery from Data, 2015, 10(2):1-33.

[14] Agnello A,Kelly B C,Treu T,et al. Data mining for gravitationally lensed quasars[J]. Monthly Notices of the Royal Astronomical Society, 2015, 448(2):1446-1462.

[15] Karabadji N E I,Seridi H,Bousetouane F,et al. An evolutionary scheme for decision tree construction[J]. Knowledge-Based Systems, 2017, 119:166-177.

[16] Wang W,Fan Y,Li R . Optimal binary codes and binary construction of quantum codes[J]. Frontiers of Computer Science, 2014, 8(6):1024-1031.

[17] Xu F . Conception and Exploration of Using Data as a Service in Tunnel Construction with the NATM[J]. Engineering, 2018, 4(1):123-130.
[18] Zhang R, Xiong S, Chen Z. Construction method of concept lattice based on improved variable precision rough set[J]. Neurocomputing, 2016, 188: 326-338.

[19] Boryczka U, Kozak J. Enhancing the effectiveness of Ant Colony Decision Tree algorithms by co-learning[J]. Applied Soft Computing, 2015, 30: 166-178.

[20] Aljukhadar M, Senecal S. The user multifaceted expertise: Divergent effects of the website versus e-commerce expertise.[J]. International Journal of Information Management, 2016, 36(3): 322-332.

[21] Huang X, Dai X, Liang W. BulaPay: a novel web service based third-party payment system for e-commerce[J]. Electronic Commerce Research, 2014, 14(4): 611-633.

[22] Ramsey E, Fang Y, Qureshi I, et al. Trust, satisfaction, and online repurchase intention: The moderating role of perceived effectiveness of e-commerce institutional mechanisms[J]. MIS Quarterly, 2014, 38(2): 407-427.