Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 3, pp. 1765–1772, DOI 10.12694/scpe.v25i3.2682

E-COMMERCE DATA MINING ANALYSIS BASED ON USER PREFERENCES AND ASSICIATION RULES

ZHIYING FAN[*]

Abstract. Improving the sales of e-commerce platforms is the primary goal of this paper. This paper studies the data of e-commerce product recommendations from the perspective of user preference and association rules. The characteristics of positive and reverse association rules in data mining are analyzed. Then, a multi-dimension association rule calculation method is proposed. Create a data attribute unit set. By analyzing each attribute's weighted coefficient and similarity, the attribute confidence degree is obtained, and the data is preprocessed. An example is given to verify the effectiveness of the proposed method. The recommendation engine based on user preferences and association rules significantly improves the accuracy, recall rate and prediction coverage of e-commerce recommendation systems.

Key words: User preference; Association rules; Electronic commerce; Data mining; Recommendation system

1. Introduction. Nowadays, most e-commerce websites have adopted various ways to provide information services. A sound referral system can improve the chances of individual items being viewed. It can improve the time consumers spend in online stores, help consumers find products they are genuinely interested in, and improve their purchasing experience. This will increase site visits and product sales. Generally speaking, product recommendations on an e-commerce website can be based on the product's characteristics, for example, brand, category, applicable age group, etc. However, this requires a professional to review the product. Suggestions can also be made based on the user's browsing habits. Some scholars have proposed an automatic classification method based on association rules [1]. It can find the correlation law between commodities based on analyzing various transaction records in the transaction database, and then assist merchants in making corresponding business decisions—for example, purchase, sales and inventory management, shelf placement, etc. Compared with similar methods, this method has higher computational speed and accuracy. Meaningful associations can be found by analyzing the correlation of massive user behaviors and product metadata in the database to realize personalized customer recommendations based on the association rule algorithm. E-commerce extensive data processing recommendation systems can provide convenience to customers and bring more revenue to the company [2]. Therefore, this paper analyzes the big data of e-commerce systems based on user preference and association rule algorithm. The conclusion obtained in this paper has significant theoretical research value and practical guiding significance.

2. Research on data mining methods and user preferences based on multidimensional association rules.

2.1. Application of multidimensional association rule method in e-com merce data preprocessing. The preprocessing of e-commerce data is the premise of data mining for the same structure data [3]. Feature extraction is the first step to preprocessing e-commerce data. Aiming at the sparse problem of e-commerce data, an e-commerce data association method based on multidimensional association rules is proposed. This allows you to find features and properties in your e-business data. The set of characteristic elements of the e-business data can then be expressed as:

$$R = \{[r_1, C(r_1)], [r_2, C(r_2)], \cdots, [r_i, C(r_i)]\} \tag{2.1}$$

[*] Shanxi Institute of Mechanical and Electrical Technology, Changzhi, Shanxi, 046011, China (Corresponding author, F13191259977@163.com)

r_i stands for the characteristic unit of electronic commerce data. $C(r_i)$ represents the number of characteristic units in electronic commerce data. When the value of the multidimensional correlation function is significant, the parameter value is also considerable. In this case, the EPC model parameter estimation obtained is relatively accurate. In this way, the parameter estimation of e-commerce big data is transformed into a target-optimal problem with limitations [4]. The generated target functionality is described as follows:

$$h(x) = \max_x y(x) \tag{2.2}$$

The multidimensional association function $h(x)$ is a fitness function of e-commerce data set parameters. A matching index based on multidimensional association rules is proposed to measure the matching degree of e-commerce data features and attributes [5]. These formulas are:

$$W(a_i, r_j) = \sum_{i=1}^{n} C_i(r_j) \frac{U}{\sum\limits_{j=1}^{n} C_i(r_j)} \tag{2.3}$$

$C_i(r_j)$ represents the statistic of the e-commerce data attribute. U represents the statistical number of all units with characteristics in the e-commerce data [6]. If the $V = \{v_1, v_2, \cdots, v_n\}$set is used to define the relevant rules of EPC data, the correlation between e-commerce data can be expressed as:

$$L\{v_1, v_2, \cdots, v_n\} = \frac{1}{h_i} \sum_{i=1}^{n} \max v_i \tag{2.4}$$

$H = \{h_1, h_2, \cdots, h_n\}$ represents the weight vector for e-commerce data. n represents the range of sub-business data. Assuming that there are weights in the e-commerce data in the attribute set, the similarity between e-commerce data c_i and attribute a_i is expressed as:

$$Sam(h_i, a_i) = \frac{h_i}{||h_i||} \tag{2.5}$$

The attribute set X^a is labeled by the similarity of the data and its attributes [7]. The reliability analysis of the characteristics of e-commerce data is carried out. The calculation formula is:

$$Cor_{sam} = \sum_{sam_{\in} \in SamX^a} \frac{Sam_i}{|X^a|} \tag{2.6}$$

$SamX^a$ in formula (2.6) is the similarity group of e-commerce data and its attributes. Sam_i stands for characteristics similar to electronic commerce data. $|X^a|$ represents the number of attributes present in the attribute set X^a of the e-commerce data. This paper transforms the parameter estimation problem of e-commerce big data into a multi-objective optimal problem with constraints [8]. Then, the method of solving multidimensional association rules is proposed. E-commerce data is preprocessed by combining similarity and confidence.

2.2. Construction of e-commerce data model. Electronic commerce data includes a lot of network information. Website Q_i also contains a large amount of content and structure of electronic transaction data [9]. The structure of web pages is analyzed by constructing an e-commerce data model. The e-business data schema is represented as follows:

$$Q_i = (Z_i, Y_i, T_i) \tag{2.7}$$

Z_i represents electronic business data in the organized form of Web pages. Y_i represents the electronic transaction data target in the Web page, which is detected by the entity. T_i stands for electronic business data included on Web pages. E-commerce data priority value is calculated according to association rules and user preference algorithms.

$$W\left(\psi_i^j\right) = \frac{\frac{1}{\sigma_{ij}^k}\left(\sum_{k=1}^{m}\varphi_{ij}^k\right)^2}{\Omega\left(C_i^j\right)} \tag{2.8}$$

σ_{ij}^k and φ_{ij}^k in formula (2.8) represent the threshold for determining the e-commerce data block j in the adjacent window k. $\left(\psi_i^j\right)$stands for the j e-commerce data block in the web page window i. $\Omega\left(C_i^j\right)$ represents the amount of ψ_i^j contained in the page. m represents the number of electronic commerce data in adjacent Windows [10]. The data are arranged in order of importance. If the maximum sorting time of e-commerce data is set to $t_{\max}$, the sorting result of e-commerce data can be expressed as:

$$C^h = \frac{1}{t_{\max}}\sum_{t=1}^{n} Q_i \tag{2.9}$$

When classifying e-commerce information, the whole minimization principle is used to classify the data and determine its suitability [11]. The model of the e-commerce data object is established. The following formula is used to calculate fitness:

$$Fithess = \sum_{i=1}^{N}\left(x_i^2 - x_i\right)^2 \tag{2.10}$$

x_i represents the expected output. N represents the sample size of e-commerce data. The $\tilde{x}_i$ stands for the actual result. The e-commerce data thus constructed can be expressed as:

$$\phi = \frac{1}{A_i}\sum_{i=1}^{m}\lambda_i\frac{L_i}{\vartheta_i\cdot\kappa_i} + Cor_{sam} \tag{2.11}$$

L_i represents the E-Business Data item at bit i in E-Business Data object Q_i. ϑ_i stands for the name of the e-commerce data item i. κ_i represents the value of the e-commerce data item i. Cor_{sam} stands for the degree of trust of item i of e-commerce data. λ_i represents the weighting of e-commerce data item i. A complete target model of electronic commerce business is formed by calculating the above links.

2.3. E-commerce data mining algorithm flow. According to the target pattern of e-commerce data, the data segmentation scheme with the highest priority is found [12]. The objective function of e-commerce data optimization is defined.

$$\min\varepsilon = \frac{1}{h_i^{(a)}}\sqrt{\frac{q_i\Omega(Z(\theta) - Z(r))}{\phi_n(v)}} \tag{2.12}$$

$h_i^{(a)}$ of formula (2.12) represents the characteristics of e-commerce data. q_i represents the eigenvector of e-commerce data. $Z(\theta)$ stands for the amount of information contained in the e-commerce data. $\phi_n(v)$ represents the difference component in the e-commerce data of the two characteristics. Ω stands for the optimal classification threshold of e-commerce data. $Z(r)$ represents the number of e-commerce data characterized by r.

2.4. User preference degree model. A preference modeling method based on user behavior data is proposed. Because the display area of the page in the recommendation scenario is limited, the items ranked higher in the recommendation list will be more likely to attract users' attention. If the user's classification model can be used to classify products, it will have a practical guiding effect [13]. This method is of great significance to improve the system's accuracy and customer loyalty. A description of the user priority algorithm is shown in Figure 2.1.

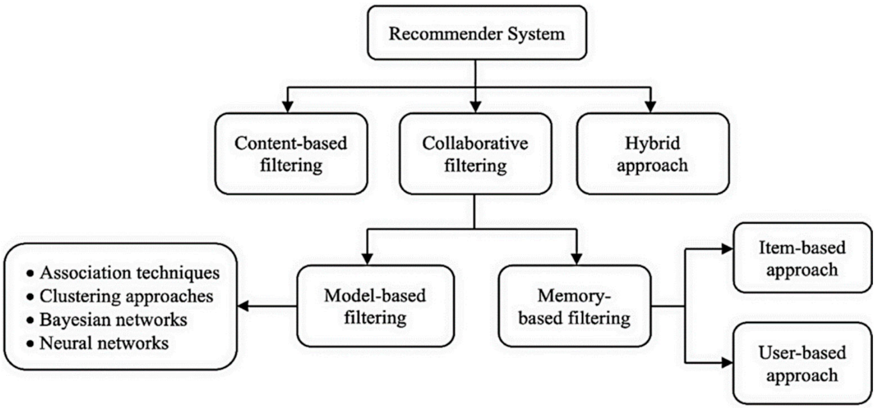

Fig. 2.1: User preference algorithm description.

3. Design of e-commerce data mining system.

3.1. System Structure. Establish an intelligent e-commerce site recommendation system. First, we must go through the data collection and extraction process. The log center of the mall is used to extract the information of users and items related to the algorithm. In addition, the matching process between the converted data and the loading module is carried out to improve the algorithm's effectiveness and the service's confidentiality. Before the establishment of the model, the corresponding countermeasures are put forward for the violation of ”brushing the list.” The aim is to improve the resistance of the rule association algorithm [14]. Minimize the direction of the specific combination of the modeled objects and the product, and store the modeled objects in the modeled database of the recommendation system. The intelligent recommendation system reads the algorithm input model required by the algorithm from the corresponding modeling database for algorithm calculation. Finally, the recommended results are transmitted to the relevant commercial system of the mall in the form of a protocol. Figure 3.1 shows the overall architecture design of the e-commerce recommendation system (the picture is quoted in Egyptian Informatics Journal, Volume 23, Issue 1, March 2022, Pages 33-45).

3.2. Functional architecture design. It is necessary to design and implement many functions in network information service reasonably to successfully implement good network information service. The differences in modular particles, coupling and cohesion among modules will directly affect the development efficiency and operation performance of the whole system [15]. This paper completes the essential system management, realizes the control of the running process of the system, and quickly configures and updates. The basic framework for intelligent recommendations in the marketplace is shown in Figure 3.2 (Frontiers in big Data, 2023, 6:1157899). The electronic commerce intelligent recommendation system consists of nine main functional modules.

Data Collection Module. The task of this module is to collect item metadata, user metadata, user usage habits and other data required by each storage and record center of the mall. According to the way and length of information storage in each storage center, the data collection module must extract relevant data from the journal center by SFTP and then save it to the distributed file system for use in the recommendation system. Given the characteristics of numerous product types and complex product levels in commercial networks, a model based on commercial relationships is proposed. Some data of each business department, such as user transaction data, are stored separately [16]. The data acquisition module is required to set up multiple data collection points. An effective data transmission method is proposed to ensure the system's data transmission speed and reliability.

Data preprocessing module. This module is mainly divided into data ETL and data cleaning modules. The two are combined to complete the preprocessing of business data. Finally, standard and efficient transaction

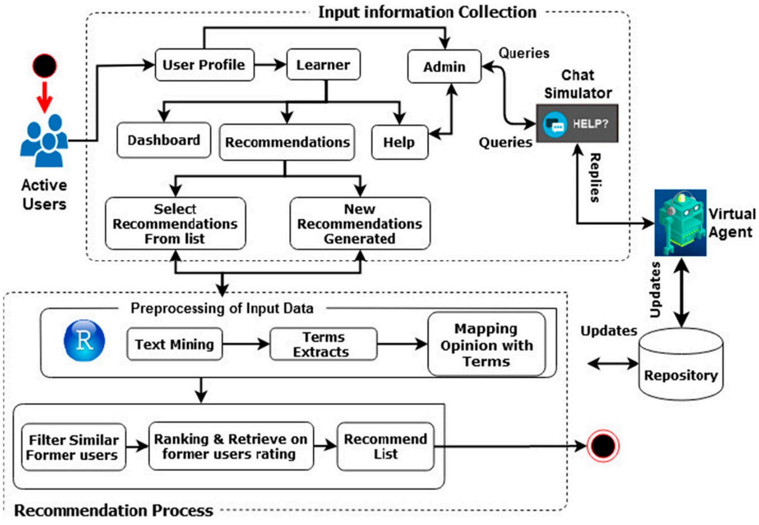
Input information Collection
User Profile
Learner
Admin
Queries
Chat Simulator
HELP?
Queries
Active Users
Dashboard
Recommendations
Help
Replies
Select Recommendations From list
New Recommendations Generated
Virtual Agent
Updates
Preprocessing of Input Data
R
Text Mining
Terms Extracts
Mapping Opinion with Terms
Updates
Repository
Filter Similar Former users
Ranking & Retrieve on former users rating
Recommend List
Recommendation Process

Fig. 3.1: Overall architecture design of e-commerce recommendation system.

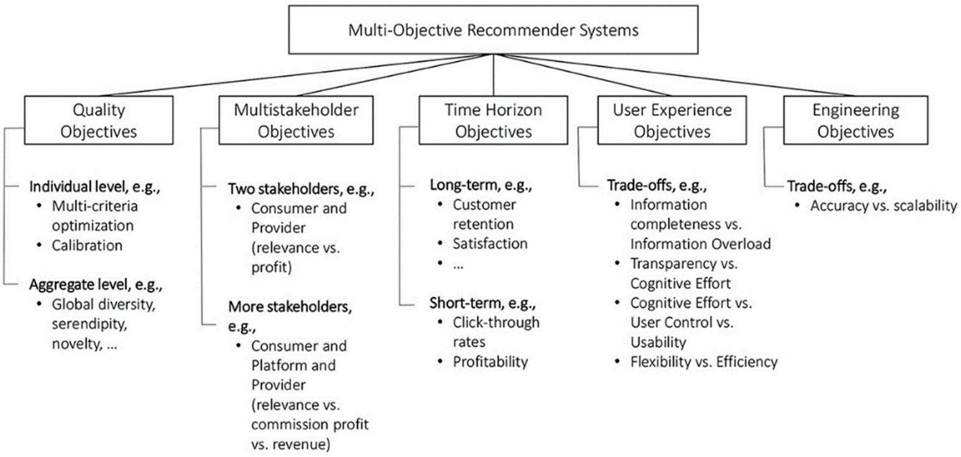
Multi-Objective Recommender Systems
Quality Objectives
Multistakeholder Objectives
Time Horizon Objectives
User Experience Objectives
Engineering Objectives
Individual level, e.g.,
• Multi-criteria optimization
• Calibration
Aggregate level, e.g.,
• Global diversity, serendipity, novelty, ...
Two stakeholders, e.g.,
• Consumer and Provider (relevance vs. profit)
More stakeholders, e.g.,
• Consumer and Platform and Provider (relevance vs. commission profit vs. revenue)
Long-term, e.g.,
• Customer retention
• Satisfaction
• ...
Short-term, e.g.,
• Click-through rates
• Profitability
Trade-offs, e.g.,
• Information completeness vs. Information Overload
• Transparency vs. Cognitive Effort
• Cognitive Effort vs. User Control vs. Usability
• Flexibility vs. Efficiency
Trade-offs, e.g.,
• Accuracy vs. scalability

Fig. 3.2: Functional architecture of e-commerce recommendation system.

data is generated. Among them, data ETL is mainly based on the data accumulated by the data acquisition module, the format requirements of the algorithm input for the data, and the processing of large-scale data expansion data extraction, transformation and loading. The data cleaning module mainly deletes or performs other operations on invalid data, missing field data and other illegal data in the process of data ETL to avoid the impact of illegal data on the recommendation quality of the algorithm.

Data mode and parsing module. This module aims to realize the analysis and modeling of ETL and cleaned data. Attribute analysis and canonical modeling of such data are carried out according to input data format and data type requirements. In the rule association algorithm, it is necessary to normalize each transaction in the transaction database [17]. How to define consumer behavior is an important question. This is because there are many scenarios where users purchase fewer items at a time. If every purchase is treated as a transaction, the links between items become small and complex, quantitative and efficient. In terms of user/commodity metadata information, it can be seen that the algorithm will analyze the user's preference degree based on the commodity's level, so the commodity's level attribute and user preference should be part of the algorithm

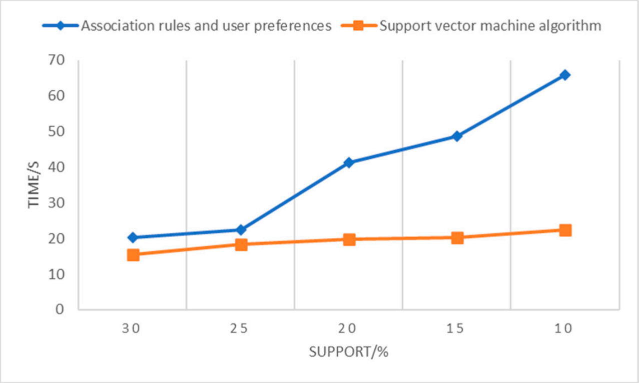

Fig. 3.3: Comparison of support vector machine algorithm and association rules algorithm in execution time.

modeling. In addition, for the recommendation system, the input model of the algorithm can be correlated with and sorted out from multiple data tables and generate specific shopping or browsing records. Finally, according to the required characteristics or other requirements, the corresponding data content and the corresponding association format are generated, and the information is saved to the unassociated database MongoDB cluster to achieve access to the recommendation algorithm.

Recommended Algorithm module. An e-commerce information recommendation method is proposed based on association rules and user preferences. The association rules algorithm presented in Chapter 2 is used in this module [18]. The method is divided into two stages: offline operation and online operation. In practical application, the offline operation mainly aims at the problem of high complexity and extended time in the solving process. The online calculation is based on the initial recommendation value and combined with the corresponding optimization algorithm to get the final recommendation value.

Distributed Recommendation Engine Module. This paper presents the implementation method of Hadoop based on the Jar package. It is done according to a particular order and needs. Select the corresponding recommendation list for the specific recommendation environment and target. The corresponding suggestions are provided to users by using the HTTP protocol.

Optimized Modules are recommended. The task of this module is to optimize the recommendation results generated by various recommendation algorithms to generate the final user recommendation results. The system mainly realizes the following three optimization aspects: initial recommendation filtering, ranking, and interpretation.

Recommended Data Interaction Interface Modules. This module aims to realize the adaptation of commodity recommendation and trading interface [19]. The data collection problem of the mall record center involves data type, data date, data channel and so on. The second is an interactive recommendation. The commercial system of the mall must respond to the recommendation system according to different recommendation scenarios, people and products. The recommendation system recommends the corresponding list to the user based on the data.

Recommended Evaluation Module. Three experimental methods evaluate the recommendation system. It includes an AIB test that evaluates the index offline, surveys users, and online trials [20]. In the practical implementation, the offline evaluation of the method is emphasized. Among them, there are mainly precision, recall rate and other indicators.

Basic Management Function Modules. A recommendation algorithm based on a data warehouse is proposed, and the algorithm is analyzed in detail. It includes two parts of: data storage and processing. The system management module includes interface management, log management, data maintenance, parameter configuration and rule intervention.

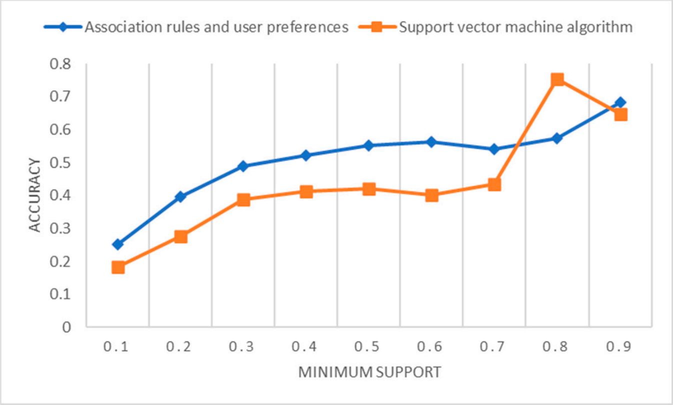

Fig. 4.1: Comparison of product recommendation accuracy between the SVM and association rules algorithms..

4. Application of user preference and association rule algorithm in product recommendation. The method uses a rotating database. When dealing with K-entry candidate sets, the search times of the original database can be reduced effectively, and the efficiency of data mining can be improved. Figure 4.1 compares the association rule algorithm and the support vector machine algorithm regarding execution time. As you can see in Figure 4.1, the relationship between user priority and the implementation time of the relevant rules is almost horizontal [21]. The support vector machine (SVM) 's running speed shows a significantly monotonically increasing trend. Compared with the traditional support vector machine method, the user preference and association rules method can obtain the initial candidate set only through a single scan of the initial transaction database. Subsequent processing, such as generating candidate sets, calculating support numbers, etc., no longer requires re-accessing the original database. This can significantly reduce the execution time and increase the operation's efficiency. The proposed method has obvious advantages over the SVM method in computation time.

If we want to discover all the e-commerce data, we need to do a lot of operations. Usually, the data in the database is divided into daily necessities, food, clothing, sporting goods, etc. There are many subdivisions under each category. This paper proposes a method based on association rules. This is also consistent with the general rules of what consumers buy. For example, when a customer searches for a digital camera, the site will display the digital camera the customer is searching for. Want to buy an SD card too? When consumers want red women's clothing, the site will suggest red leather bags and socks. Because it is only for product recommendation, only two frequent sets must be mined when data mining. In general, the number of recommendations that each project can provide is limited. When many items match the suggestions, you can select the most popular items from them [22]. The correctness of the method is tested by statistical analysis of the purchase records of different types of products. In this way, the precision comparison curve of the product can be obtained (Figure 4.2). From Figure 4.2, we can see that the accuracy of product recommendation of user preference and association rule algorithm is higher than that of support vector machine algorithm in terms of corresponding support degree.

5. Conclusion. The product recommendation function on the e-commerce platform is becoming increasingly prominent. A data mining method based on association rules and user preferences is presented. The algorithm can recommend products that users may be interested in more accurately based on sales history data. The results show that the method utilizes a rotating database and bit operation. The computational efficiency of this method is much higher than that of the conventional SVM method. Product recommendations are more accurate after introducing rule mining. But this approach has its limits. For example, if there is too much traffic on the site, the demand for storage will increase. This is the next step that needs to be addressed.

REFERENCES

[1] Zhang, H. N., & Dwivedi, A. D. (2022). Precise marketing data mining method of E-commerce platform based on association rules. Mobile Networks and Applications, 27(6), 2400-2408.

[2] Zhang, Y. (2021). Sales forecasting of promotion activities based on the cross-industry standard process for data mining of E-commerce promotional information and support vector regression. Journal of Computers, 32(1), 212-225.

[3] Loukili, M., Messaoudi, F., & El Ghazi, M. (2023). Machine learning based recommender system for e-commerce. IAES International Journal of Artificial Intelligence, 12(4), 1803-1811.

[4] Xie, C., Xiao, X., & Hassan, D. K. (2020). Data mining and application of social e-commerce users based on big data of internet of things. Journal of Intelligent & Fuzzy Systems, 39(4), 5171-5181.

[5] Massaro, A., Mustich, A., & Galiano, A. (2020). Decision support system for multistore online sales based on priority rules and data mining. Computer Science and Information Technology, 8(1), 1-12.

[6] Sjarif, N. N. A., Azmi, N. F. M., Yuhaniz, S. S., & Wong, D. H. T. (2021). A review of market basket analysis on business intelligence and data mining. International Journal of Business Intelligence and Data Mining, 18(3), 383-394.

[7] Bakar, W. A. W. A., Zuhairi, M. A., Man, M. U. S. T. A. F. A., Jusoh, J. A., & Triana, Y. S. (2022). a Critical Review of Deep Learning Algorithm in Association Rule Mining. J. Theor. Appl. Inf. Technol, 100(5), 1487-1494.

[8] Xu, B., Huang, D., & Mi, B. (2020). Research on E-commerce transaction payment system basedf on C4. 5 decision tree data mining algorithm. Computer Systems Science and Engineering, 35(2), 113-121.

[9] Ünvan, Y. A. (2021). Market basket analysis with association rules. Communications in Statistics-Theory and Methods, 50(7), 1615-1628.

[10] Tran, D. T., & Huh, J. H. (2022). Building a model to exploit association rules and analyze purchasing behavior based on rough set theory. The Journal of Supercomputing, 78(8), 11051-11091.

[11] Zong, K., Yuan, Y., Montenegro-Marin, C. E., & Kadry, S. N. (2021). Or-based intelligent decision support system for e-commerce. Journal of Theoretical and Applied Electronic Commerce Research, 16(4), 1150-1164.

[12] Wu, Z., Li, C., Cao, J., & Ge, Y. (2020). On scalability of association-rule-based recommendation: A unified distributed-computing framework. ACM Transactions on the Web (TWEB), 14(3), 1-21.

[13] Murthy, T. S., Roy, M. S., & Varma, M. K. (2020). Improving the performance of association rules hiding using hybrid optimization algorithm. Journal of Applied Security Research, 15(3), 423-437.

[14] Putra, A. A. C., Haryanto, H., & Dolphina, E. (2021). Implementasi Metode Association Rule Mining Dengan Algoritma Apriori Untuk Rekomendasi Promo Barang. CSRID (Computer Science Research and Its Development Journal), 10(2), 93-103.

[15] Zhao, Z., Jian, Z., Gaba, G. S., Alroobaea, R., Masud, M., & Rubaiee, S. (2021). An improved association rule mining algorithm for large data. Journal of Intelligent Systems, 30(1), 750-762.

[16] Han, Q. Y. (2020). The study of personalized recommendation algorithm in e-commerce system. International Journal of Education and Economics, 3(2), 1-6.

[17] Zhang, Y. (2021). The application of e-commerce recommendation system in smart cities based on big data and cloud computing. Computer Science and Information Systems, 18(4), 1359-1378.

[18] Perumal, S. P., Sannasi, G., & Arputharaj, K. (2020). REFERS: refined and effective fuzzy e-commerce recommendation system. International Journal of Business Intelligence and Data Mining, 17(1), 117-137.

[19] Tingting, W. (2020). Research on user access pattern mining based on web log. Asia-pacific Journal of Convergent Research Interchange (APJCRI), 6(8), 135-148.

[20] Dogan, O., Kem, F. C., & Oztaysi, B. (2022). Fuzzy association rule mining approach to identify e-commerce product association considering sales amount. Complex & Intelligent Systems, 8(2), 1551-1560.

[21] Urbancokova, V., Kompan, M., Trebulova, Z., & Bielikova, M. (2020). Behavior-based customer demography prediction in E-commerce. Journal of Electronic Commerce Research, 21(2), 96-112.

[22] Rani, L. N., Defit, S., & Muhammad, L. J. (2021). Determination of student subjects in higher education using hybrid data mining method with the k-means algorithm and fp growth. International Journal of Artificial Intelligence Research, 5(1), 91-101.

Edited by: Zhigao Zheng
Special issue on: Graph Powered Big Aerospace Data Processing
Received: Oct 9, 2023
Accepted: Oct 23, 2023