Innovative Design and Intelligent Manufacturing L.C. Jain et al. (Eds.) © 2024 The Authors. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/FAIA241098

# Design and Optimization of an Intelligent Recommendation System for Instant Retail Services Based on Machine Learning

Xingyao ZHOU<sup>a</sup>, Fujun WAN<sup>a,1</sup>, Yuchen ZHANG<sup>a</sup>, Shuohan QIU<sup>b</sup> <sup>a</sup>China National Institute of Standardization, Beijing 100080, China <sup>b</sup>China Electronics Standardization Institute, Beijing 100010, China

Abstract. In order to realize the desire of all parties involved in instant retail, it can better understand the behavior of users, discover the hidden interests of users and the behavior rules of group users, and an intelligent recommendation system for instant retail services based on machine learning is proposed. Analyze and process the relevant data of e-commerce websites, use the improved rough set attribute reduction algorithm and Apriori improved algorithm for data mining, use the two improved algorithms together as a scheme to apply to the recommendation system, and then compare the recommendation efficiency of the system through empirical analysis of relevant data. The experimental results show that the time to generate frequent item sets based on the improved Apriori algorithm is significantly less than that based on the Apriori algorithm when the minimum support is fixed. In terms of the efficiency of generating frequent item sets, the Apriori improved algorithm is more efficient. With the minimum support getting smaller and smaller, the comparison becomes more obvious. When the number of transactions is fixed, the time used by the recommendation system to apply Scheme II is obviously less than that used by Scheme I. Moreover, with the continuous increase of the number of transactions in the transaction database, the difference in the time used to apply the two schemes increases significantly. In terms of recommendation efficiency of instant retail intelligent recommendation system, Scheme II is better than Scheme I. Conclusion: The system effectively improves the recommendation efficiency of the recommendation system.

Keywords. rough set; Data mining; Apriori algorithm; Instant retail; Personalized recommendation

## 1. Introduction

With the rapid development of the Internet, e-commerce platform functions continue to improve, the mode of retail industry change is moving forward at an unprecedented speed. 2023 Central Document No. 1 pointed out that we should comprehensively promote the construction of the county commercial system, accelerate the improvement of the county, townships and villages of e-commerce and courier logistics and distribution system, and vigorously develop the common distribution, instant retailing and other new modes [1]. Instant retail mode has been applied from the first and second tier cities, and gradually penetrated into the county area. Due to the increasing demand of county residents and

<sup>&</sup>lt;sup>1</sup>Corresponding Author: Fujun Wan, China National Institute of Standardization, E-mail: wanfujun@126.com

rapid economic growth, the operation and management mode of instant retail has been well developed, and the scope of service is also expanding.

At present, with the surge in the number of customers, how to analyze the customer's purchasing behavior to identify their preferences for goods, and recommend satisfactory products for them, has become an important challenge and opportunity for online businesses, is considered to be an effective way to solve this problem of e-commerce recommender system came into being [2]. The current e-commerce recommender system design method can be roughly divided into two categories: one is based on the fixed price method, for a single customer of a single transaction to perform the recommendation task, its limitations are: the consideration of the factors is not comprehensive, often nonautomatic recommendation mode, accuracy is poor, does not have the intelligence. Another type is the use of content-based methods, mainly relying on the content and related features as the basis for recommending products to customers. For example, by analyzing the user's historical transaction records, mining their purchasing patterns and purchasing interests, and based on the discovered patterns of their services or commodities intelligent recommendation [3]. The limitation is that there is no modeling of the customer's purchasing behavior, and there is a lack of effective processing of the customer's preferences, so the accuracy of the recommendation is not high.

# 2. Literature review

In order to solve the problem of "information overload" caused by the massive amount of information, such as Google, Baidu and other search engines as a representative of the information retrieval system, the retrieval system will meet the user's needs for information submitted to the user, to a large extent, to help the user from the rich resources to obtain the required network information. However, when using a search engine, if the user enters the same keyword for information retrieval, then the information obtained is the same [4]. From another point of view, the user's needs for information are diverse and different, and the information retrieval system represented by the search engine does not meet the user's personalized needs, and therefore cannot effectively solve the problem of information overload.

With the rapid development of recommendation systems, in addition to analyzing and modeling users' historical behavior records, people also try to use a variety of hybrid models and recommendation methods to solve the cold start problem and data sparse problem of the recommendation system [5-8]. Recommendation systems have gradually been widely accepted and applied, and more and more scholars have begun to focus on the robustness and security of recommendation algorithms, as well as the effectiveness evaluation methods of recommendation systems [9-10]. Low, M. P., et al. used 4932579 pieces of borrowing data from readers in their experiment. Relatively speaking, they used a large-scale dataset, more than 10 times the largest scale of the previous experiment. Although it still does not fully represent the big data environment, it is relatively close to the current real big data application scenario [11].

As a scientific research of intelligent computing, rough set theory has made great progress in both theoretical research and practical application, showing its broad application prospects. The improved attribute reduction algorithm related to rough set theory is applied to association rule data mining, and the Apriori algorithm is improved, which can optimize the data mining scheme and effectively improve the recommendation efficiency of e-commerce intelligent recommendation system based on rough set. This paper first briefly describes e-commerce personalized service recommendation system, then proposes a data mining model of association rules based on rough set, and then comprehensively uses the improved algorithm of attribute reduction based on SQL and the improved algorithm of Apriori as a data mining solution, and then applies this solution to e-commerce intelligent recommendation system, Use relevant data to conduct experiments to analyze the recommendation efficiency of the system.

#### 3. The research methodology

## 3.1 Instant Retail Personalized Service Recommender System

The instant retail personalized service recommendation system is a system applied to the field of instant retail, aiming at achieving the win-win goal of enterprise income increase and customer satisfaction. Its advantage lies in improving customers' satisfaction with instant retail websites through good recommendation functions, helping them establish and maintain good customer relationships, which is a virtuous circle for enterprises and customers to achieve win-win results. Web mining can be divided into three categories. The system mainly uses Web content mining and Web usage mining, and uses the methods based on database and server-side Web access logs to analyze and summarize customer access rules and purchase patterns, classify according to different rules and patterns, and then according to online customer registration information Purchase history, evaluation information and other data will be classified into one of the above categories, and the recommendation results will be displayed to customers in combination with realtime commodity database information; In addition, the system can also track the user's access behavior in different ways, such as tracking each purchase behavior of the user, each evaluation of the purchased goods, and the marking of some goods (such as putting into the shopping cart). Through the operation of related algorithms, the system can achieve real-time and dynamic adjustment of the recommendation set, thus further improving the accuracy of recommendations [12].

As shown in Fig. 1, the system mainly has the following functional modules: First, collect the data related to customers and commodities to provide data sources for the next data mining, which is the basis of all related work. Second, the collected data are preprocessed with continuous attribute discretization, data completion, attribute simplification, etc. The result of data mining depends on the effect of pre-processing to a large extent. Third, the pre-processed data are stored to form a user transaction database. Fourth, the transaction database using relevant data mining techniques (such as association rules, sequence patterns, etc.) for analysis and research, the formation of valuable statistical results and rules, found that the user's purchase mode, and in an appropriate way to explain, determine the recommended mode used to guide the user's actual online trading behavior [13]. Fifth, the recommendation engine is formed on the basis of the previous modules, it is the whole recommendation system of the "big manager", taking into account the user's purchase history, evaluation information, marking information, current browsing behavior, the current product database and other information, the use of the fourth part of the mining out of the rules and other results of the realization of the customer's recommendations to the user's interface and friendly visualization effect of the display to the user. The visualization effect is displayed to the user to achieve the purpose of personalized recommendation service. At the same time, the recommendation results and the user's current shopping behavior are sent to the website management center and stored in the corresponding database, so as to optimize the structure of the website and dynamically change the recommendation results in order to realize the real-time accuracy of the recommendation.



Figure 1. Overview diagram of the instant retail personalized service recommendation system

# 3.2 Rough set based association rule data mining model

Generally speaking, a complete commercial data mining process based on rough set theory includes the following three conceptual layers: data preparation layer; data preprocessing layer; mining evaluation layer. In this paper, the association rule data mining model based on rough set is shown in Figure 2, and the whole mining process is divided into three steps: data preprocessing, attribute simplification, and association rule mining.



Figure 2. Rough set based association rule data mining model.

#### 3.2.1 Data pre-processing

The initial information of the data source is processed by data cleaning, continuous attribute discretization and data completeness, followed by data selection and transformation, clarifying the set of conditional attributes and decision-making attributes, forming a complete transaction one by one, and depositing it into the database. This thesis adopts the test data of Nellie's Mall (an online shopping platform completed during the internship) as the data source.

## 3.2.2 Attribute Minimization

For the relevant data in 3.2.1, use the attribute reduction algorithm based on discernibility matrix to reduce the condition attributes, delete redundant attributes, and generate the reduced attribute set on this basis; In addition, the improved algorithm of attribute reduction based on SQL mentioned is used to reduce the conditional attribute set and obtain the reduced attribute set. Compare the efficiency of the two attribute reduction algorithms and their impact on the efficiency of association rule mining. This paper proposes an improved algorithm for attribute reduction based on SQL [14]. The key step in this algorithm is to calculate the information entropy H (P). Its SQL language implementation algorithm is: SELECT SUM (P (X i log (P (X i)) FROM S GROUP BY P, where P (X i is calculated according to formula (1)

 $P(X_i) = COUNT(*)/|U|$ (1)

## 4. Analysis of results

Association rule mining can be divided into two steps: (1) According to the minimum support, search for frequent itemsets in the transaction database with a large number of transactions; (2) According to the minimum confidence, association rules are generated based on the found frequent item sets. The first step is the key and more complex step. The superiority of the algorithm directly determines the complexity and efficiency of this step. The second step is relatively simple. On the basis of frequent item sets generated in the first step, association rules can be generated through subset generation.

According to the data reduction results in 3.2.2, this paper determines the support threshold and confidence threshold, executes Apriori algorithm and improved Apriori algorithm on the reduction results, and compares the efficiency of the two algorithms in discovering frequent item sets. The verification results are shown in Figure 3.

It can be seen from Figure 3 that under the condition of a certain minimum support, the time to generate frequent item sets based on the improved Apriori algorithm is obviously less than the time to generate frequent item sets based on the Apriori algorithm [15]. This shows that the Apriori improved algorithm is more efficient in generating frequent item sets, and it can be seen from the figure that the contrast becomes more obvious as the minimum support becomes smaller and smaller. In addition, it can be seen from the two steps of association rule mining that the operation of the second step depends on the first step, so in the process of generating association rules in the second



Figure 3. Variation of the efficiency of the two algorithms in generating frequent itemsets with respect to the minimum support

step, the Apriori improved algorithm is still better than the Apriori algorithm. In conclusion, in the whole process of association rule mining, the efficiency of Apriori improved algorithm is higher than that of Apriori algorithm.

The above analysis compares the efficiency of Apriori's improved algorithm in the process of mining association rules such as generating frequent item sets. Next, in order to study and analyze the impact of the algorithm on the recommendation efficiency of the recommendation system, we apply the improved algorithm of attribute reduction based on SQL and the improved algorithm of Apriori to the instant retail intelligent recommendation system, The following two different schemes are adopted respectively: the combination of attribute reduction algorithm based on discernibility matrix and Apriori algorithm is used as Scheme 1, and the combination of improved attribute reduction algorithm based on SQL and Apriori algorithm is used as Scheme 2. These two schemes are applied to personalized recommendation service of instant retail intelligent recommendation system, and the original data is imported into the system, Compare the running time of the recommended system [16]. The verification results are shown in Figure 4: when the number of transactions is fixed, the time used by the recommendation system to apply Scheme 2 is obviously less than that used by Scheme 1, and with the continuous increase of the number of transactions in the transaction database, the time difference between the two schemes increases significantly. This shows that Scheme 2 is better than Scheme 1 in terms of recommendation efficiency of instant retail intelligent recommendation system.



Figure 4. Variation of the execution efficiency of the recommend system with the number of transactions under the two scenarios

## 5. Conclusion

Rough set theory has made great progress in theoretical knowledge and practical application. Applying the improved attribute reduction algorithm and Apriori algorithm related to rough set theory to the recommendation system based on association rule data mining can effectively improve the recommendation efficiency of the recommendation system. This paper first introduces the instant retail personalized service recommendation system, briefly describes a running process of the recommendation system, then proposes a rough set based association rule data mining model, introduces the components of the mining model and the data source of this paper, and then uses SQL to improve the attribute reduction algorithm based on rough set on the basis of reducing condition attribute set, On this basis, it is combined with Apriori improved algorithm to become a scheme, which is applied to the recommendation system to improve the recommendation efficiency of instant retail intelligent recommendation services. The experiment shows that the scheme composed of improved attribute reduction algorithm based on SQL and improved Apriori algorithm, and the scheme composed of attribute reduction algorithm based on discernibility matrix and Apriori algorithm are applied to the real-time retail intelligent recommendation system based on rough set. The recommendation efficiency of the former is obviously higher than that of the latter, and the effect is more obvious as the amount of data increases and the number of transactions increases.

## Funding

This research was funded by The Dean's Fund Project of China National Institute of Standardization " Research on the Design and Optimization of Key Elements in Instant Retail Services ", Project Number: 602024Y-11441.

# References

- Lee, D., Gopal, A., & Park, S. H. (2020). Different but equal? A field experiment on the impact of recommendation systems on mobile and personal computer channels in retail. Information Systems Research, 31(3), 892-912.
- [2] Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A survey of recommendation systems: recommendation models, techniques, and application fields. Electronics, 11(1), 141.
- [3] Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation systems: Algorithms, challenges, metrics, and business opportunities. applied sciences, 10(21), 7748.
- [4] Ferreira, D., Silva, S., Abelha, A., & Machado, J. (2020). Recommendation system using autoencoders. Applied Sciences, 10(16), 5510.
- [5] Onile, A. E., Machlev, R., Petlenkov, E., Levron, Y., & Belikov, J. (2021). Uses of the digital twins concept for energy services, intelligent recommendation systems, and demand side management: A review. Energy Reports, 7, 997-1015.
- [6] Alamdari, P. M., Navimipour, N. J., Hosseinzadeh, M., Safaei, A. A., & Darwesh, A. (2020). A systematic study on the recommender systems in the E-commerce. Ieee Access, 8, 115694-115716.
- [7] Karn, A. L., Karna, R. K., Kondamudi, B. R., Bagale, G., Pustokhin, D. A., Pustokhina, I. V., & Sengan, S. (2023). Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis. Electronic Commerce Research, 23(1), 279-314.
- [8] Pereira, A. M., Moura, J. A. B., Costa, E. D. B., Vieira, T., Landim, A. R., Bazaki, E., & Wanick, V. (2022). Customer models for artificial intelligence-based decision support in fashion online retail supply chains. Decision Support Systems, 158, 113795.
- [9] Bandyopadhyay, S., Thakur, S. S., & Mandal, J. K. (2021). Product recommendation for e-commerce business by applying principal component analysis (PCA) and K-means clustering: benefit for the society. Innovations in Systems and Software Engineering, 17(1), 45-52.
- [10] Karthik, R. V., & Ganapathy, S. (2021). A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. Applied Soft Computing, 108, 107396.
- [11] Low, M. P., Cham, T. H., Chang, Y. S., & Lim, X. J. (2023). Advancing on weighted PLS-SEM in examining the trust-based recommendation system in pioneering product promotion effectiveness. Quality & Quantity, 57(Suppl 4), 607-636.
- [12] Lee, D., & Hosanagar, K. (2021). How do product attributes and reviews moderate the impact of recommender systems through purchase stages?. Management Science, 67(1), 524-546.
- [13] Beheshti, A., Yakhchi, S., Mousaeirad, S., Ghafari, S. M., Goluguri, S. R., & Edrisi, M. A. (2020). Towards cognitive recommender systems. Algorithms, 13(8), 176.
- [14] Patro, S. G. K., Mishra, B. K., Panda, S. K., Kumar, R., Long, H. V., Taniar, D., & Priyadarshini, I. (2020). A hybrid action-related K-nearest neighbour (HAR-KNN) approach for recommendation systems. IEEE Access, 8, 90978-90991.
- [15] Himeur, Y., Sohail, S. S., Bensaali, F., Amira, A., & Alazab, M. (2022). Latest trends of security and privacy in recommender systems: a comprehensive review and future perspectives. Computers & Security, 118, 102746.
- [16] Mashal, I., Alsaryrah, O., Chung, T. Y., & Yuan, F. C. (2020). A multi-criteria analysis for an internet of things application recommendation system. Technology in Society, 60, 101216.