Network Resource Personalized Recommendation System Based on Collaborative Filtering Algorithm

Gang Qiu[1,2](✉) and Jie Cheng[2]

[1] School of Software, Shandong University, Jinan 250101, China
[2] Changji University, Changji 831100, China

Abstract. Several existing personalized recommendation systems for network resources do not specifically classify user history information personalized classification criteria, resulting in a low degree of matching between system personalized recommendations and user interests. In order to improve the accuracy of personalized recommendations for network resources, we designed a collaborative filtering based Algorithm-based personalized recommendation system for network resources. In the hardware design, design the overall circuit module, configure the bus timing, and improve the operating speed of the system hardware. In software design, calculate the time weight function to construct the user's implicit scoring matrix; calculate the similarity between network resources and user browsing history items, design personalized classification criteria for user history information based on collaborative filtering algorithms; calculate predicted item scores and actual scores Establish a personalized recommendation model for network resources.In the experiment, the system is compared with several existing systems, and the average absolute error in different adjacent sets is tested. According to the data results, in the five data test sets, the average absolute error of the system is less than that of other systems, so the personalized recommendation system based on collaborative filtering algorithm has better recommendation accuracy.

Keywords: Collaborative filtering algorithm · Overall circuit module · Bus timing · Time weight function · Recommendation model

1 Introduction

Modern e-commerce has the main characteristics of personalization, which is not possessed by the traditional business model. E-commerce has an urgent demand in the dynamic provision of precise marketing "comparative shopping" services and real-time provision of optimized distribution services. With the emergence of e-commerce, it provides an unprecedented opportunity to achieve this demand, making it a new direction of e-commerce development, "personalized" business model came into being [1].

Literature [2] analyzes the user's historical behavior data based on the system dynamics algorithm, mines the user's interest preferences, classifies users according to different

S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 416, pp. 645–655, 2022.
https://doi.org/10.1007/978-3-030-94551-0_50

interest preferences, and recommends items with similar preferences. This kind of system requires a combination of qualitative and quantitative methods, and the calculation process is more complicated and the calculation speed is slow. Literature [3] designed a personalized recommendation system for college books and bibliographies in view of the poor effect and poor accuracy of the current college bibliographic personalized recommendation system combined with association rules. However, the hardware of this system is not up to the standard and cannot realize fast calculation and processing.In reference [4], a personalized recommendation system is designed based on tensorflow. The recurrent neural network module of the system can build a model for time series, fully tap the changing interests of users, and save training time to a great extent through multi super parameter adjustment, and can significantly reduce the error rate. But this kind of system needs a lot of parameters to build the model, otherwise the accuracy is difficult to reach the standard. In this paper, we design a new personalized recommendation system based on collaborative filtering algorithm.

2 Hardware Design of Network Resource Personalized Recommendation System

2.1 Circuit Module Design

The structure before the expansion of the network port of the ultrasonic flowmeter only includes three parts: TMS320F28335 as the control core; an analog front end composed of a transmitter circuit that transmits ultrasonic probe drive signals, a receiver circuit that receives ultrasonic probe signals, and a switch circuit that switches the transmit and receive states [5]; The ADC sampling module that samples the received signal.

DM9000 Ethernet control chip provides three kinds of bus connection modes: 8-bit, 16-bit and 32-bit. The 16-bit bus connection mode is chosen in this article. In the actual connection process, what needs to be connected includes 16-bit bus SD0~SD15, chip select signal AEN, processor read command signal IOR and processor write command signal IOW. It should be noted that the data bus and address bus are multiplexed in DM9000, and the actual access type of the bus is determined by the input signal of the *CMD* pin. When the input signal of *CMD* pin is a high level signal, the data bus is accessed; When the input signal of *CMD* pin is low level, the address bus is accessed. Therefore, in TMS320F28335, a signal is needed to control the *CMD* pin. The *CMD* pin in this paper is controlled by the SA2 pin in the address bus of TMS320F28335. In this way, TMS320F28335 can access dm9000 data bus and address bus by accessing different addresses.

2.2 Bus Timing Configuration

In the process of using xinf bus of TMS320F28335, the key point is to configure the timing of xinf [6]. In the internal clock signal of TMS320F28335, xtimclk and xclkout are used by xinf. The relationship between these two clock signals and CPU clock sysclkout is shown in Fig. 1.

Access to XINTF is based on XTIMCLK. XTIMCLK can be configured as SYSCLKOUT or SYSCLKOUT/2 through XINTFCNF2 register, and its default value

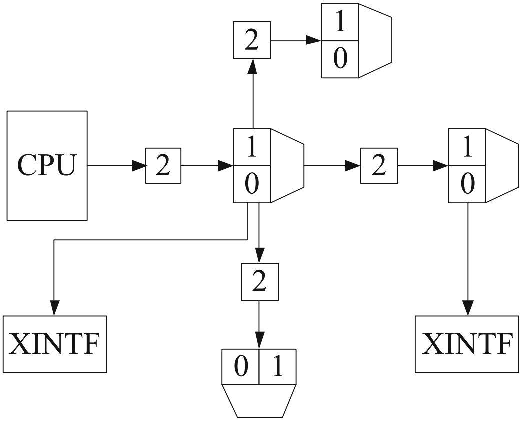


Fig. 1. Bus configuration relationship

is SYSCLKOUT/2. XCLKOUT is the output clock provided by XINTF module. In the process of DM9000 accessing Zone0 memory area of TMS320F28335 through XINTF bus, the process of accessing Zone0 includes three stages: establishment, activation and tracking.The timing of the three stages can be configured in the register XTIMING. In the setup phase, XZCSO is set to low level, and then the address is placed on the address bus. In the activation phase, if data is read from DM9000, the read signal is set to low level again; if data is written to DM9000, the write signal XWE is set to low level. In the tracking period, when the read and write signal becomes high, it will continue to be maintained for a period of time to ensure that the read and write process can be completed.The establishment, activation and tracking cycle of each XINTF can be configured through the XTIMING register. Through the design of the circuit module and the configuration of the bus timing, the operating speed of the network resource personalized recommendation system can be improved, so that it can search for network resources and sort them within the allowable range of the software [7].

3 Software Design of Network Resource Personalized Recommendation System Based on Collaborative Filtering Algorithm

3.1 Constructing the User Implicit Rating Matrix

The system's mining of user's implicit information is mainly through the log files of the website, where the main analysis and use of user behavior records, which mainly include user name, target IP, number of visits, visit duration, application type, etc., can be used for user interests [8]. It is expressed by the type of item visited, that is, the type of website visited. The interests of users are reflected by the websites they browse, and are classified according to the application types of the websites, and each application type includes at least one item. Since the log data has been classified according to application types, the realization of user interest classification is relatively simple. According to the record information of users visiting learning website, the user interest is mapped, and the user interest degree is transformed into implicit scoring of project resources by users, and the implicit scoring matrix of users is established. Considering the timeliness of the system recommendation, the user's latest browsing record is the most favorable for the recommendation effect, which makes the recommendation result very accurate, so the system selects the magazine file data of the last month as the analysis object [9]. Because this article analyzes and mines the user's implicit rating information, the rating information is indirectly reflected by the user's behavior when browsing the page. In order to intuitively describe the user's interest in the project, the interest is converted into an interest value, the data weight based on the user's visit time is introduced, and the user's interest value is generated in combination with the number of visits. Based on the time weight function, the calculation formula of $W_i(m, n)$ can be obtained as:

$$W_i(m, n) = (1 - \beta_x) + \beta \times \frac{D_{mn}}{T_m} \tag{1}$$

Where, $W_i(m, n)$ is the time weight function of a user accessing a network resource; β_x is the index to adjust the weight growth, that is, to change the speed of weight change; D_{mn} represents the access time of user m to resource n; T_m is the time interval for user m to access such resources. Considering that in a period of time, some users' interests may be more scattered or concentrated, which will have a certain impact on the recommendation effect to a certain extent, the appropriate correction should be made according to the number of types of users' interests in the quantitative scoring. The user scoring value formula is as follows:

$$H_{mn} = \frac{\beta_m}{C_i} \times K_{mn} \tag{2}$$

In the formula, H_{mn} represents the value of user m's interest in network resource n; β_m represents the adjustment parameter of network personalized information classification, generally taking a value of 1; C_i represents the standard threshold of user i for the item; K_{mn} represents the user's interest in a certain network resource The value of the score.The behavior records of a user in the network are obtained from the log file, and then these behaviors are classified as $C_i = \{C_1, C_2, \cdots, C_n\}$ according to the user's interest, and then the implicit scoring matrix of the user is output. The length and times of user I's access to item J are extracted from the log file and recorded as D_{mn} and T_m respectively. The time weight $W_i(m, n)$ is obtained by formula (1), and then multiplied by the number of visits to obtain the user interest value. According to formula (2), the user's rating value is obtained, denoted as H_{mn}, and then the user's rating value of the item is divided into 6 standards from 0 to 5, and a threshold value is set for each standard, denoted as $CL_i = \{CL_1, CL_2, \cdots, CL_n\}$, and then the relevant threshold value is passed Evaluate the corresponding interest value, repeat the above steps, calculate the user's score for all access items and the score for all application types, and finally form the user's implicit score matrix.

3.2 Design User History Information Personalized Classification Standard Based on Collaborative Filtering Algorithm

Firstly, preprocess the historical query information on the web, and use a vector composed of multiple search keywords to represent the vector model of the similarity of historical query information on the web to improve the accuracy of retrieval. On this basis, the temporal semantic classification of the information in the network historical query information database is carried out, and the temporal semantic feature vector is extracted. At this point, a vector model should be established first, so that the vector can represent network history query information in this feature space, and all the temporal semantic sentences related to the network history query information in the database are extracted to establish this vector model. Every feature in the vector model has its eigenvalue, that is the weight of the feature. Each temporal semantic keyword is one of the feature dimensions. Assuming that there are I temporal semantic keywords in the database, the feature dimension in the database is also I. For a web history query information H_i, we can use the frequency of temporal semantic keyword H_j in the database to calculate the weight of temporal semantic keywords, and take the similarity as its continuous measurement index. The formula for calculating the frequency of temporal semantic keywords is as follows:

$$H_{ij} = \frac{\mu_{ij}}{\gamma_j} \tag{3}$$

Among them, H_{ij} represents the frequency of temporal semantic keywords, μ_{ij} represents the frequency of temporal semantic keywords appearing in historical online query information, and γ_j is the most frequent occurrence of temporal semantic keywords in historical online query information. The number of times. This formula is mainly used to calculate the relative frequency of a temporal semantic keyword in the online historical query information. The temporal semantic keywords appearing in the network history query information are transformed into a vector according to the frequency order. The content of the network history query information is extracted, the attributes and attribute values are extracted, and the temporal semantic vector describing the content of the network history query information is characterized. Each web history query information has different temporal semantic features, so when calculating the similarity of web history query information, we need to first consider the concept attribute.Assuming that the concept G has an instance g, the instance g can be represented as $g_y = W_\eta[C_L]$, where $D = (D_1, D_2, \cdots, D_m)$. At this point, the similarity of instance G and Q can be calculated. First, the attribute vector of instance G and Q is a common attribute vector through the above method, and then the similarity of the historical query information of the network is calculated according to the attribute value, and the comparison of the two instances [10]. The formula for attribute value and similarity is as follows.

$$\alpha_G(m, n) = \sum_{t=1}^{n} \frac{\mu_{ij} + \gamma_j}{2} \cdot \alpha_G(m_1, n_1) \tag{4}$$

Among them, μ_{ij} and γ_j are the weight coefficients of attributes m_1 and n_1 in each vector, which are preset parameters. They are usually the statistical values obtained after preprocessing the network history query information. The final weight value of the network history query information is determined by this statistical value, and its value range is [0.1]. By substituting the above similarity into the semantics of the network history query information, the similarity between the network history query information and the standard semantics can be calculated, and then the personalized classification standard of the user's network history query information can be obtained [11].

3.3 Establish a Personalized Recommendation Model for Network Resources

When building the recommendation model, the set of neighboring users of the target item is obtained through the above formula, and then the target user's score of unknown items is predicted according to the score of neighboring users, and the items with higher prediction score are recommended to users [12]. The algorithm for predicting, scoring and recommending unknown items is shown in Fig. 2:

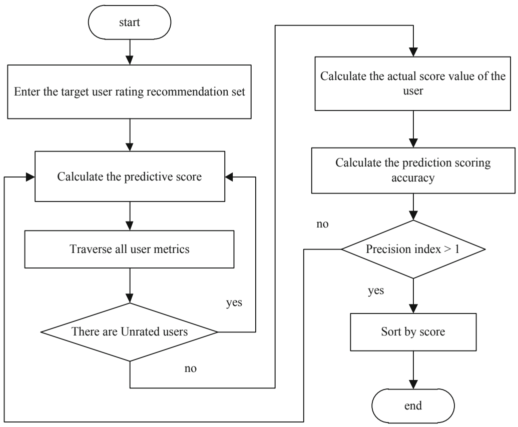


Fig. 2. Algorithm flow

As shown in Fig. 2, it is first necessary to input the nearest neighbor sets of various types of target users in the algorithm, such as $C_i = \{C_1, C_2, \cdots, C_n\}$ above, and then use formulas to calculate the target user's predicted scoring value for unknown items and the ranking of recommended items. Assuming that the project user is C_i, the unrated projects of the network resource X_n can be roughly summarized in a set of projects. Then use the evaluation formula to calculate the predicted accuracy value:

$$G_{m,n} = \frac{\sum_{i=1}^{m} sim(m, G_i) \times H_M}{\sum_{j=1}^{n} \left| sim(n, G_j) \right|} \tag{5}$$

Where, $G_{m,n}$ is the precision index between the predicted value and the actual value of user m in network resource item n; $sim(m, G_i)$ represents the predicted value of a user relative to a network resource; $sim(n, G_j)$ the actual Recommendation Index of a user relative to a network resource; H_M represents the user's own rating of the item.If the accuracy does not meet the standard, repeat the calculation steps of the user's recommended predicted value for the network resource until the accuracy of the scoring items of all users reaches the standard value. Finally, the score values of the unknown items in all types of the target user are calculated, and the predicted score values are sorted from high to low, and the top X items are selected as recommended items.The algorithm can recommend the users who first visit the network resources, classify the interests according to the types of network resources accessed by users, mine the implicit information of users in the system described in this paper, and calculate the similarity to produce the nearest neighbor, and finally recommend the similar items. With the increase of access to network resources and the extension of time, the number of recommended projects will be more and more, and the accuracy is also higher and higher.

4 Experimental Research

4.1 Experiment Preparation

The evaluation of personalized recommendation system is also an important content. If the results of recommendation can not meet the personalized needs of users, it will reduce the user's trust in personalized recommendation service, and recommendation is meaningless. On the contrary, if the personalized recommendation results are consistent with the user's personalized needs, it will enhance the user's sense of trust in the system. More importantly, it can help users solve the problem of "information overload" and lack of personalized services, so that users can reduce the time cost of finding the resources they need.Therefore, it is very necessary to evaluate the recommendation effect of the algorithm. Use the SPM method to calculate the similarity between the predicted score and the user's real score in the recommendation results of the personalized recommendation system to measure the recommendation accuracy of the recommendation system, calculate the average absolute error (MAE), and detect the items recommended by the personalized recommendation system Whether it is appropriate, statistics recommend the frequency of correct or incorrect. The MAE value is used to calculate the error value between the predicted score of the recommendation algorithm and the actual score, so MAE only calculates the resource items that the user has scored in the test data set.The smaller the MAE value, the higher the recommended accuracy of the proposed algorithm. If the score set of I users in n network resources can be expressed as n_i, the user score predicted by the system can be expressed as $ky_{i,j}$ and the real score of the user can be expressed as $kz_{i,j}$, then the calculation of the average absolute error can be expressed as:

$$\eta_i = \frac{\sum_{j-1}^{n_i} \left| ky_{i,j} - kz_{i,j} \right|}{n_i} \tag{6}$$

In the formula, η_i represents the absolute error of the resource recommendation accuracy of the network resource personalization system. Take the average of the MAE values of the selected i users to get the overall average absolute deviation of the algorithm:

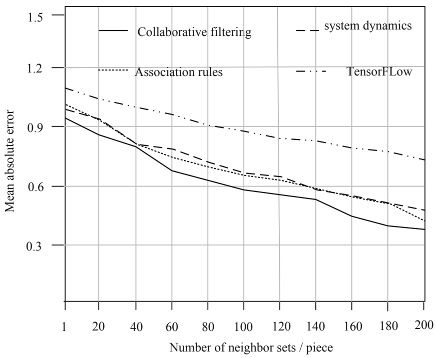
$$\overline{\eta}_i = \frac{\sum_{i=1}^{n} \eta_i}{i} \tag{7}$$

Where $\overline{\eta}_i$ is the average absolute error of the recommendation accuracy of the network resource personalization system; i is the number of users who scored. Through the previous article, we know that the selection of similar user groups has a great impact on the final recommendation results. In the case of selecting different neighbor sets, the stability of recommendation accuracy is a big index to measure the efficiency of collaborative filtering recommendation algorithm. Therefore, this experiment will design experiments to test the algorithm.Taking movie scoring as an example, a data set on network resources is established, and the Movie Lens data set provided by the Group Lens project research team of the University of Minnesota is used as the data set of this experiment. This includes nearly 200,000 rating data on 2,249 movies from 10,484 users, and each Web user has rated at least 19 of them. In the experiment of this article, the training set and the test set are divided into 5 cross-validation, and the data set is randomly divided into 5 groups. Each test set occupies about 20% of the total data. Therefore, it can be known that the total data of the data set The amount is about 80% to form 5 sets of test sets and training sets.

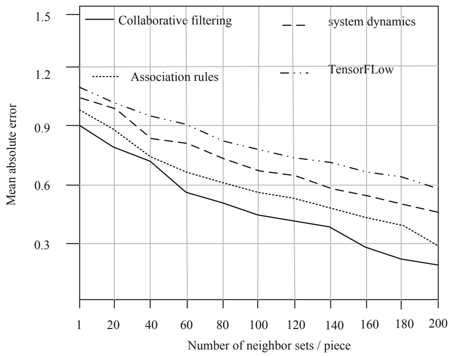
4.2 System Personalized Recommendation Test

Under the same experimental conditions and experimental environment, five rounds of cyclic experiments were carried out on the experimental group and the control group with different training sets and test sets, and their recommended results were compared with the test set data, and five sets of experiments were calculated. In, set up the experimental group and the control group respectively. The experimental group is the recommendation system based on the collaborative filtering algorithm designed in the article, and the control group is the personalized recommendation system based on system dynamics proposed in the literature [2], literature [3] The personalized recommendation system based on association rules mentioned in [4] and the personalized recommendation system based on TensorFlow proposed in [4]. In the same test set, different sizes of neighbor sets are selected to test and compare the traditional collaborative filtering recommendation algorithm with the optimized recommendation algorithm. In this experiment, we increase the number of neighbors from 1 to 200, and the test results are shown in Fig. 3.

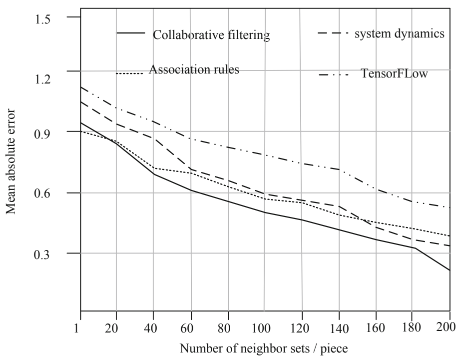
According to the experimental data of the five test sets in Fig. 3, the size of the neighbor set has a significant impact on the average relative error value, that is, the accuracy of the system recommendation. With the increase of the neighbor set, the average relative error value will gradually decrease, and the accuracy of the recommendation will be improved. In the comparative test of the four systems, the network resource personalized recommendation system based on collaborative filtering algorithm is smaller than other systems, and has higher recommendation accuracy in the five test sets. From the experimental data in Fig. 4, we can know that the personalized recommendation system based on collaborative filtering algorithm designed in this paper has better recommendation accuracy than several existing systems.



(A) Test set 1

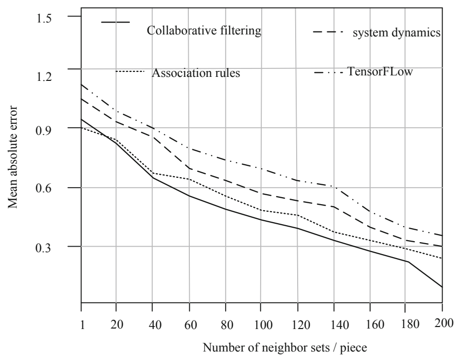


(B) Test set 2

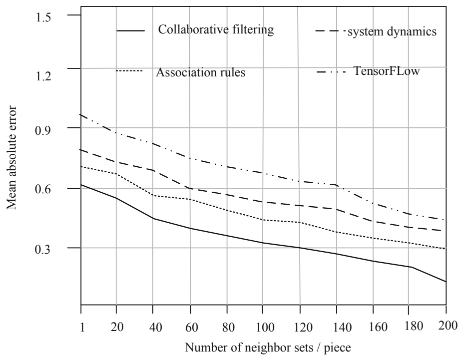


(C) Test set 3

Fig. 3. System performance comparison test



(D) Test set 4



(E) Test set 5

Fig. 3. continued

5 Conclusion and Prospect

The collaborative filtering recommendation system was born in the field of e-commerce and has been widely and successfully applied. Traditional websites can only recommend unified network resources to customers, but are unable to meet the individual needs of customers. In order to provide them with personalized network resource recommendation, we learn from the successful application of collaborative filtering recommendation technology in the field of e-commerce, and try to introduce this technology into the network environment design in this research.Based on the three systems of literature [2], literature [3] and literature [4], this paper optimizes them. Then, based on the optimized collaborative filtering recommendation algorithm, a personalized network resource recommendation model is proposed. Finally, a personalized network resource recommendation system is designed and implemented to meet the personalized needs of customers and improve the efficiency of the network.

Although this method seems to have achieved some results, there are still many deficiencies in the research. In the future, research will be conducted on the category and amount of network data to improve the effectiveness of recommended resources.

Fund Projects. The work was supported by the 2021 Autonomous Region Innovation Environment (Talents, Bases) Construction Special-Natural Science Program (Natural Science Foundation) Joint Fund Project (2021D01C004) and the 2019 Xinjiang Uygur Autonomous Region Higher Education Scientific Research Project (XJEDU2019Y057, XJEDU2019Y049).

References

1. Zhang, Y-J., Dong, Z., Meng, X,-W.: Pesearch on personalized advertising recommendation systems and their applications. Chin. J. Comput. **44**(3), 531–563 (2021)
2. Wang, Z., Si, L., Liu, B., et al.: Study on information personalized recommendation based on system dynamics. J. Hebei Univ. Sci. Technol. **42**(2), 170–179 (2021)
3. Xu, B.: Design of university library bibliography personalization recommendation system based on association rules. Modern Electron. Tech. **44**(8), 68–72 (2021)
4. Yang, H.: Design of personalized recommendation system based on TensorFLow. Adhesion **41**(2), 166–169 (2020)
5. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. Neurocomputing (2020). https://doi.org/10.1016/j.neucom.2019.12.143
6. Li, X., Liang, H., Feng, J., et al.: Design of personalized learning resource recommendation system for online education platform. Comput. Technol. Dev. **31**(2), 143–149 (2021)
7. Liu, S., Li, Z., Zhang, Y., Cheng, X.: Introduction of key problems in long-distance learning and training. Mob. Networks Appl. **24**(1), 1–4 (2018). https://doi.org/10.1007/s11036-018-1136-6
8. Chen, Y., Geng, X.: Recommendation of personalized product-service system scheme based on improved collaborative filtering. Comput. Integr. Manuf. Syst. **27**(1), 240–248 (2021)
9. Ou, D.: Research and implementation of personalized travel recommendation system based on data mining[J]. Hubei Agric. Sci. **60**(9), 123–126 (2021)
10. Gao, P., Li, J., Liu, S.: An introduction to key technology in artificial intelligence and big data driven e-learning and e-education. Mob. Networks Appl. (2021). doi: https://doi.org/10.1007/s11036-021-01777-7
11. Niu, Q., Yang, R., Tian, H.: Optimization scheduling simulation of network high coverage resource based on AGV communication. Comput. Simul. **38**(6), 410–414 (2021)
12. Chin, H.H., Varbanov, P.S., Kleme, J.J., et al.: Enhanced cascade table analysis to target and design multi-constraint resource conservation networks. Comput. Chem. Eng. **148**(3), 107262 (2021)