International Journal of Scientific Engineering and Science
Volume 6, Issue 3, pp. 20-22, 2022. ISSN (Online): 2456-7361

Design of Product Recommendation System based on Restricted Boltzmann Machine

Linyuan Liang[1], Zhan Wen[1], Yihan Shi[1], Hantao Liu[2], Wenzao Li[1,2]
[1]Chengdu University of Information Technology, Chengdu, Sichuan, China, 610225
[2]Sichuan Education Informatization and Big Data Center, Chengdu, Sichuan, China, 610000

Abstract— *The traditional recommendation system is to use the evaluation of products by neighbors with high similarity to the target user to predict how much the target user likes the product, but its drawback is that the degree of individual user profiling is not enough, and the recommendation accuracy should be improved again. In this paper, we integrate the RBM model with clustering, which can optimize the shortcomings of traditional recommendation models. The restricted Boltzmann machine is a graphical model of binary random variables, based on a complete bipartite graph separating hidden and observed variables, and is a binary simulation of the factor analysis model. The RBM model can fully consider the connection between users and products, and the clustering algorithm focuses on the connection between users, and combining them both can improve the accuracy of the recommendation system.*

Keywords— *K-Means; Restricted Boltzmann Machine; recommendation system.*

I. Introduction

Recommendation system [1] provides a new way to solve information overload. It has the functions of satisfying users' needs, obtaining happiness, expanding horizons, etc. For websites, it has the value of retaining users and achieving business goals. For content providers, it also has the functions of obtaining long tail traffic, obtaining interaction, recognition and benefits. Therefore, it is very important to design and build an accurate recommendation system. At present, the mainstream recommendation system algorithms [2] includes content-based recommendation algorithm, collaborative filtering recommendation algorithm, hybrid recommendation algorithm, association rule recommendation algorithm, and recommendation algorithm based on network structure. Among them, collaborative filtering is a recommendation algorithm based on user behavior. In 1992, Xerox Palo Alto Research Center first proposed the word collaborative filtering, and developed Tapestry [3] system to solve the problem of mail overload and help users filter mail. Under the background of user behavior precipitation information, this model combines tag-based collaborative filtering algorithm [4] with deep learning [5] to obtain user personalized product recommendation system with high accuracy. However, when using the preferences to a group of similar interests and common experience to recommend information that users are interested in, the mining of personal interests is not thorough enough. If users do not feedback new information, the quality of new projects depends on the historical data set, and the recommendation accuracy is not guaranteed. Therefore, it is necessary to develop a commodity recommendation system that uses groups interest to predict personal interest and explores personal interest, giving users a better shopping experience and better promoting the development of e-commerce.

II. Model Design Idea

In the article of hot points and visual analysis of recommendation system research in China [6], Multi-Dimensional Scaling (MDS) analysis gets one of the hot topic domains of recommendation system research in China as the application of user-based personalized recommendation of e-commerce and other fields and the research of recommendation system ontology. As the core part of personalized recommendation system, the accuracy of recommendation algorithm directly affects the accuracy of recommendation system. The model proposed in this paper is designed for the application of recommendation systems in e-commerce, where K-Means focus on the use of group scores to obtain individual scores and RBM focus on the post-user profiling scores. First, obtain user behavior data, perform data pre-processing, and set up user labels. Secondly, the users are clustered to calculate the correlation between the tags belonging to a cluster of users and predict the personal commodity score. Third, this score and user behavior are introduced into the RBM neural network for personal interest exploration, and the final prediction results are obtained. Fourth, the development website provides user interfaces.

III. Design Scheme

According to the design idea, the design scheme for the product recommendation system can be obtained as shown in the figure.

As can be seen from the figure, the system is mainly divided into the steps of data collection, data processing, obtaining group ratings, constructing user RBM model, and building the website, which are described below.

International Journal of Scientific Engineering and Science
Volume 6, Issue 3, pp. 20-22, 2022. ISSN (Online): 2456-7361

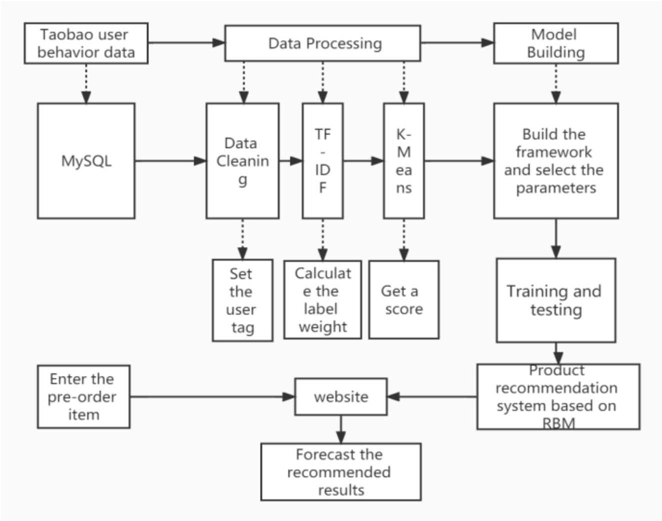

Fig. 1. Flow chart of the design scheme.

A. User Behavior Data from Taobao for Recommendation

This dataset contains all behaviors (behaviors include clicks, purchases, adds, likes) of approximately one million random users with behaviors between November 25, 2017 and December 3, 2017. The dataset is organized similarly to MovieLens-20M, i.e., each row of the dataset represents a user behavior, consisting of user ID, product ID, product category ID, behavior type, and timestamp, separated by commas. A detailed description of each column in the dataset is as follows.

TABLE I. Definition and description of the data

S. No.	Column	Description	Comment
1	user_id	Identity of users	Int,Sampled&desensitized
2	item_id	Identity of items	Int,Desensitized
3	behavior_type	The user behavior type	Int,Desensitized
4	item_category	The category id of the item	strings,Including pv, buy, cart, fav
5	time	The time of the behavior	Int,To the nearest hours

TABLE II. User behavior type

S. No.	Behavior_type	Comment
1	pv	click
2	buy	payment
3	cart	add-to-cart
4	fav	collect

TABLE III. Description of the size of the dataset

S. No.	Dimension	Quantity
1	number of users	987,994
2	number of commodity	4,162,024
3	Number of commodity categories	9,439
4	Number of all acts	100,150,807

B. Data Processing

Because the data set is too large, only 20% of the data are randomly selected. Firstly, we clean the acquired data, including missing value processing, consistency processing, outlier processing, date and time period processing, and making user tabs. Secondly, we record the browsing active time period and purchase active time period of the last 30 days, the last 7 days and the last behavior of the user, and pay attention to whether the user browses, adds an order without placing an order and the number of days between the last two purchases. Third, defines labels to users according to their behavior and preferences to determine whether they are repurchase users, whether they have high access activity, whether they have high purchase activity, and whether they purchase a single type of good, and group users by value. Finally, using the Term Frequency-Inverse Document Frequency (TF-IDF), the label weights are calculated.

C. Get a group score

Based on whether the training samples contain label information, machine learning can be divided into supervised and unsupervised learning. Clustering algorithm [7] is a typical unsupervised learning in which the value of the training samples contains the features of the samples and does not contain the label information of the samples. In the clustering algorithm, the samples with similar attributes are divided into uniform categories by using the characteristics of the samples. A user is regarded as a sample, clustered, and the similarity relationship between labels belonging to users in a cluster is calculated using cosine similarity, and the prediction score is calculated to be equal to the behavior weight multiplied by the label correlation for subsequent model building.

D. Building user RBM model

The RBM model [8] is built to set the appropriate number of neurons in the visual layer and the hidden layer. The neurons in the visual layer use softmax neurons, which is a vector of length K, indicating that the score of the project is 1 to K. Only one component of the vector is 1, and the other components are 0. Items that are not scored by the user are used as a vector of length K with all components equal to 0, and the connected weights and bias terms are not updated. During the training process, a separate RBM model is trained for each user, and the previous prediction scores and behavioral information are passed into the model, and the training weights are constantly trained. As a result, the user's interests can be further explored, and more accurate user-recommended products can be obtained.

E. Building a Website

The website is provided for use. When entering the name of the product to be queried, through the user's RBM model, the product information and model prediction and recommendation results are obtained, and the results are visualized on the web page to help users filter products.

Once completed, this system can be widely used in e-commerce to help users, websites and content providers reach their respective goals and promote economic and technological development.

IV. CONCLUSION

In this paper, we propose a product recommendation system combining K-Means algorithm and deep learning RBM model. This system can better filter information and

Volume 6, Issue 3, pp. 20-22, 2022.
ISSN (Online): 2456-7361

solve the problem of information overload, users can have a better shopping experience, and the website can better realize the business value. This model has not been proposed and applied in the field of e-commerce.

To alleviate data sparsity and scalability in collaborative filtering, the K-Means clustering model is applied to a product recommendation system. RBM is the first neural network model that has been applied to learn the hidden representation of a user by reconstructing the user's rating data. The system was proposed based on the application of K-Means in e-commerce and the study of RBM recommendation system. These models have different prediction methods, but each has its own advantages and they complement each other, so combining them can obtain more accurate recommendation results.

Acknowledgment

This work is supported by Student Innovation and Entrepreneurship Project of Chengdu University of Information and Technology (No. 202110621142), and it is also supported by education application research of intelligent technology based on 5G, General Education Research Fund Project of Sichuan Province (NO. 2019514) and Network and Data Security Key Laboratory of Sichuan Province, UESTC(NO.NDS2021-7).

References

[1] Beheshti, A. , Yakhchi, S. , Mousaeirad, S. , Ghafari, S. M , & Edrisi, M. A. , “Towards cognitive recommender systems”, Algorithms, vol. 13, issue 8 : 176, 2020.
[2] Georgia Koutrika, “Modern Recommender Systems: from Computing Matrices to Thinking with Neurons,” in Proceedings of the 2018 International Conference on Management of Data, pp. 1651-1654, 2018.
[3] Rahman, Deman . "Comparison of Java web application frameworks." , M.S. thesis, Enterprise Software Engineering, University of Greenwich, London, England, 2019.
[4] Caiyun Guo, Huaijin Wang, “Improved label-based collaborative filtering algorithm”, Computer Engineering and Applications, vol. 52, no. 8, pp. 56-61, 2016. (in Chinese)
[5] Golovko V. A., Kroshchanka A. A., Mikhno E. V. , “Deep Neural Networks: Selected Aspects of Learning and Application”, Pattern Recognition and Image Analysis, vol. 31, issue 1, pp. 132-143, 2021.
[6] Zhang Jun, Gu Chong, “Research hotspots and visual analysis of recommendation systems in China”, Modern commerce industry, vol. 39, no. 18, pp. 7-10, 2018. (in Chinese)
[7] Tao, Y. , Yang, F. , Liu, Y. , & Dai, B, “Research and optimization of K-means clustering algorithm”, Computer Technology and Development, vol. 28, no. 6, pp. 90-92, 2018. (in Chinese)
[8] Huang L.W., Jiang B.T., Lv Shouye, Liu Yanbo, Li D.Y.. “A review of deep learning-based recommendation system research”, Journal of Computer Science, vol. 41, no. 7, pp. 1620-1646, 2018. (in Chinese)