

Research on Algorithm of Personalized Recommendation System Based on Deep Learning

Feng Liu^{1, a} and Weiwei Guo^{1, b}

¹ Heilongjiang University of Technology, JiXi, 158100, China

^aliufeng8038@163.com, ^bgwwguoweimei@163.com

Keywords: Recommendation System; Deep learning; Collaborative Filtering

Abstract. Deep learning is an important direction in the field of artificial intelligence. Integrating deep learning technology into recommendation system to improve the performance and recommendation accuracy of recommendation system has become a research hotspot. This paper first analyzes the recommendation ideas and recommendations of traditional recommendation algorithms. The advantages of the algorithm are then combined with the deep learning framework to establish a personalized recommendation algorithm model. Finally, the simulation experiment proves that the personalized algorithm proposed in this paper can effectively improve the performance and recommendation accuracy of the recommendation system.

Introduction

In recent years, with the rapid development of new technologies such as grid computing, cloud computing, and big data, the widespread use of the Internet has led to a large-scale increase in data. According to the relevant report of the International Data Group, by 2020, the total amount of data generated by Internet applications will be about 24 times that of 2010. In this huge data space, there are many valuable data resources that will make decisions for the development of various fields in society. However, a large number of data resources provide value and lead to information overload. The problem, how to quickly obtain valuable data resources in a large amount of messy data has become a bottleneck in the development of big data, and the recommendation system is the most effective way to solve the information overload problem. The research and application of the recommendation system has become a hot issue in research and attention in various fields, and certain results have been formed in the research process. The recommendation system can discover information of interest to the user, such as product information, news information, movies, etc., from a large number of data resources according to the user's interests, needs, characteristics, and the like through a certain recommendation algorithm, and then the found result information is performed. Personalized recommendations to users, thereby increasing the efficiency of users to find valuable data. At present, the recommendation system has been widely used in various fields such as e-commerce, social networking, information inquiry, entertainment, and information push.

The recommendation algorithms used in the recommendation system mainly include a content-based recommendation algorithm, a collaborative filtering recommendation algorithm, and a hybrid recommendation algorithm. The application is widely used in the collaborative filtering recommendation algorithm, and has achieved certain results. At the same time, there are problems such as cold start and data sparsity; the content-based recommendation algorithm needs to search for relevant items through the user's selection of project attributes for recommendation. The feature extraction process mainly relies on artificially designed feature extraction, and its scalability is not good. The hybrid recommendation algorithm combines the advantages of both, effectively alleviating the cold start and data sparsity problems, but the diversity and distribution of data types are uneven. Other features make the hybrid recommendation still face certain challenges.

Deep learning is a kind of artificial intelligence research. It is a deep nonlinear network structure and has achieved certain results in image processing, artificial intelligence and natural language processing. Deep learning simulates the human brain for analysis, which analyzes and understands the data by simulating the human brain. Deep learning is a kind of unsupervised learning. The training

center contains multi-layered perceptual information, and the lower-level features are combined to form a more abstract high-level representation attribute category. Or feature to discover the distributed characteristics of the data. Research on personalized recommendation system based on deep learning has become a hot issue in the field of recommendation system [1].

Research Work

The purpose of the recommendation system research is to actively mine the data of interest to the user from a large amount of data, and filter the data information that the user is not interested in, so as to accurately recommend the user. At present, the main recommended system algorithms are mainly content-based recommendation algorithms, collaborative filtering recommendation algorithms and hybrid recommendation algorithms [2].

Content-Based Recommendation Algorithm. The content-based recommendation algorithm is based on the selected user, by analyzing the user historical data, searching for the feature attributes of the historical data, and recommending data similar to the feature attributes to the user. The content-based recommendation algorithm is mainly based on the user's historical data and the feature attribute information of the data. This recommendation algorithm can ignore the data sparseness of the user's historical data score. The key point of recommendation is to establish a mathematical model based on the user's and user's preference for the item, and then calculate the similarity between the items through the algorithm. It is to find high-similar items with the items that the user likes, and then recommend [3].

Content-based recommendation algorithms are generally divided into three steps:

1. Item representation: The recommendation algorithm extracts some features or content from the user's historical items to represent the item and establish corresponding matrix information;
2. Feature learning: The feature data information matrix is established through the user's preference data, and the feature data is classified through the calculation link, one is the feature data that the user likes the item, and the other is the feature data that the user does not like, and then learns the user's Feature hobby
3. Generate a recommendation list: According to the user's characteristic hobbies, and the user hobbies found in the previous step, the candidate items are compared with the user's preferences, thereby generating a recommendation list of the user items.

The content-based recommendation method considers the similarity between items in the calculation process, and ignores some historical data of the user, so it can effectively link the cold start and data sparseness in the collaborative filtering recommendation process. Therefore, the content-based recommendation algorithm has the advantages of short response time and fast speed. At the same time, there are some shortcomings in this recommendation algorithm. For example, the items recommended to the user are highly similar to the user's historical preferences, so it may lead to items that the user may like but not enough to recommend to the user. The accuracy of the recommendation; secondly, in the recommendation process, the user's feature data needs to be processed in advance, and the amount of data processed is very large, which is difficult to implement [4].

Collaborative Filtering Recommendation Algorithm. Collaborative Filter (CF) algorithm is one of the most recommended recommendation system algorithms, and it is the most widely used personalized recommendation algorithm [5]. The basic idea of collaborative filtering recommendation algorithm can be understood as “classification by object” and “classification by people”, that is to say, the user's future preferences are consistent with historical preferences, then the user can be recommended to the historical preference. Or recommend items to users with the same historical preferences, so that personalized recommendations can be implemented. The flow of the collaborative filtering recommendation algorithm is shown in Figure 1.

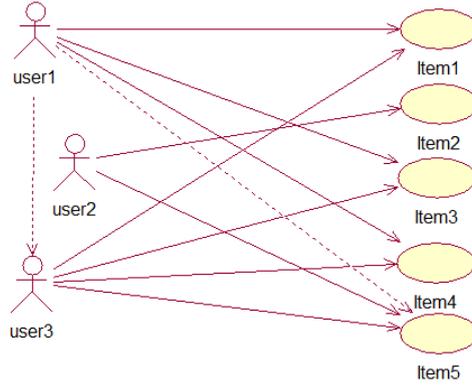


Figure. 1 collaborative filtering recommendation process

The collaborative filtering recommendation algorithm can be further divided into user-based collaborative filtering recommendation and project-based collaborative filtering recommendation. The advantage of the collaborative filtering algorithm is that the user and the item's own data information can be used, and only the user and the item can be mined through the scoring matrix according to the user's scoring matrix. Thereby achieving personalized recommendations. In the collaborative filtering process, the association rules between similar items or similar users can be fully utilized to discover the potential preferences of the users, thereby providing a higher quality personalized recommendation service. So building a scoring matrix for users and projects is critical. The scoring matrix of users and projects is constructed in Table 1 below.

Table 1 User and project scoring matrix

	Item1	Item 2	Item 3	...	Item n
user1	R_{11}	R_{12}	R_{13}	...	R_{1n}
user2	R_{21}	R_{22}	R_{23}	...	R_{2n}
user3	R_{31}	R_{32}	R_{33}	...	R_{3n}
...
user m	R_{m1}	R_{m2}	R_{m3}	...	R_{mn}

In the collaborative filtering recommendation algorithm, user-based collaborative filtering recommendation is the main content. In this process, finding the nearest user set is the most critical step, that is, calculating the similarity between users. At present, the methods for calculating user similarity mainly include Pearson correlation coefficient and cosine similarity [6].

1. Pearson correlation coefficient

If the scoring matrix $R(m,n)$ of the user and the project has been established, the Pearson correlation coefficient between the two different users a and b can be expressed by Equation 1.

$$\text{sim}(a, b) = \frac{\sum(R_{ai} - \bar{R}_a)(R_{bi} - \bar{R}_b)}{\sqrt{\sum(R_{ai} - \bar{R}_a)^2 \sum(R_{bi} - \bar{R}_b)^2}} \quad (1)$$

In Formula 1, R_{ai} and R_{bi} denote user a and user b 's score on item i , \bar{R}_a and \bar{R}_b denote the average value of user a and user b 's score on all items, respectively.

2. Cosine similarity

When cosine similarity is used for similarity calculation, the score of a user for all items is regarded as a vector in n -dimensional space. If the user does not score on one of the items, the value of the score can be considered to be zero. In some specific environments, many users do not rate the project, so these scores can be set to zero at this time, so this method has higher data sparsity when calculating the cosine similarity. In the scoring process, because each user has certain differences in the scoring standards of the project, the difference brings the error to the whole calculation. Therefore, in the calculation method of cosine similarity, the common practice is to reduce The mean value of the current user score is dropped, and the corrected similarity range is -1 to 1 , so the cosine similarity of user a and user b can be expressed by Equation 2 [7].

$$\text{sim}(a, b) = \frac{\sum_p (R_{a,p} - \bar{R}_a)(R_{b,p} - \bar{R}_b)}{\sqrt{\sum_p (R_{a,p} - \bar{R}_a)^2} \sqrt{\sum_p (R_{b,p} - \bar{R}_b)^2}} \quad (2)$$

Where \bar{R}_b represents the average score of user b for all items, and \bar{R}_a represents the average score of user a for all items. Equation 2 and Equation 1 are relatively similar.

Project-based collaborative filtering recommendation is similar to user-based recommendation process. The idea is that if users have certain preferences for a certain type of project, there will be certain preferences for similar or similar projects. The starting point is Project-based [8].

Although the collaborative filtering algorithm can get the item information similar to its own preferences, there are certain defects in the collaborative filtering algorithm. For example, the collaborative filtering algorithm can easily lead to the sparseness of the data, the cold start of the user and the cold start of the project.

Personalized Recommendation Algorithm

Deep learning is a set of methodology developed for the purpose of computer learning human behavior. The purpose is to mine rules and knowledge from a large amount of data. The algorithm used by the recommendation system (model-based recommendation) can use the deep learning method, which essentially acquires features from the training set and predicts that the recommendation system is a business application, and the deep learning is a basic algorithm, which is in an upstream and downstream relationship with the recommendation algorithm [9]]. Deep learning has become the core algorithm of recommendation system. The traditional collaborative filtering method is gradually eliminated, and it is generally only applied to hot and cold start, downgrade strategy, and preliminary screening strategy. Content recommendation needs to solve the problem of refined personalized recommendation and heat penetration of long tail content, and deep learning method has more advantages [10].

Deep Learning Framework. Combine deep learning with personalized recommendations, first of all to build a deep learning framework, as shown in Figure 2.

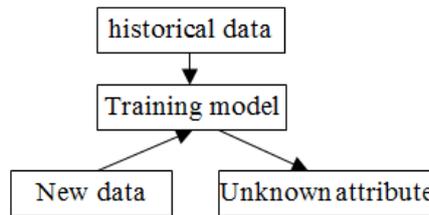


Figure. 2 deep learning framework

Firstly, the historical data is used for repeated training in the training model. In the process of learning and training, new data can be continuously introduced, and the training parameters should be constantly adjusted throughout the training process. Finally, the unknown properties are exported through the training model.

Personalized Algorithm Model Design. The model design of the algorithm firstly screens out the recall set according to the given user historical data information, and establishes the user's favorite list through the parameter adjustment of the deep learning training model. Under normal circumstances, if a user has two in a short time at the same time Piece of historical data, then the probability of obtaining user preferences will increase; secondly, by calculating the remaining string similarity between the user and the project, the recall set of the product to be recommended can be obtained, and after the recall set is obtained, filtering is performed by the matching category. Finally, the parameter values are repeatedly modified, the feature attributes are obtained through the deep learning model, the user's preferences are re-created, and a list of information to be recommended is established.

1. Establish the objective function

According to the collaborative filtering algorithm, the objective function is established as shown in Equation 3:

$$\min_{x^{(1)}, \dots, x^{(n_m)}} \frac{1}{2} \sum_{i=1}^{n_m} \sum_{j:r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)})^2 \quad (3)$$

It can be known from Equation 3 that the user parameter vector θ is first initialized, and then the item feature value x and the user parameter vector θ are obtained by an iterative method, which is the basic idea of the initial collaborative filtering method.

2. Training model

On the basis of deep learning, the objective function can be further optimized. First, $x^{(1)}, \dots, x^{(n_m)}, \theta^{(1)}, \dots, \theta^{(n_u)}$ random initialization is performed in a small range, and then the gradient descent algorithm is applied to iteratively optimize the objective function. The iteration formula is shown in Equation 4 and Equation 5:

$$x_k^{(i)} = x_k^{(i)} - \alpha \left(\sum_j ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) \right) \quad (4)$$

$$\theta_k^{(j)} = \theta_k^{(j)} - \alpha \left(\sum_i ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) \right) \quad (5)$$

The feature matrix x obtained through deep learning in the recommendation method contains important data information of the project. Sometimes the information implies some attributes and relationships that are difficult to be understood, but it is still necessary to use the feature matrix x as an important project recommendation. in accordance with. For example, if a user is interested in the project $x^{(i)}$, the recommendation model can find a project similar to $x^{(i)}$ $x^{(j)}$ according to the given similarity measure. The user. In the collaborative filtering recommendation method, there is no need to set an additional feature offset term, so there is $x, \theta \in R^n$.

Simulation

In order to verify the personalized recommendation algorithm proposed in this paper, we compare it with the traditional method. The standard of verification is usually measured by statistical accuracy. The statistical accuracy measurement method mainly calculates the deviation between the score value of the item predicted by the recommendation algorithm and the actual score value of the item, and measures the recommended quality of the recommendation system recommendation algorithm by the calculated recommendation deviation. Through the Internet platform, some data sets of the Jingdong website shopping in 2014 were obtained. The data collection contains different product information and user characteristic information purchased by users at various levels.

Through the simulation platform, the simulation data to be tested is imported, and the neighboring numbers are selected as 10, 50, 100, 150, 200 respectively, and the personalized algorithm is compared with the traditional collaborative filtering algorithm and the hybrid algorithm to obtain the contrast deviation value as shown in Table 2.

Table 2 Comparison bias values

	Personalized recommendation algorithm	Collaborative filtering recommendation	Mixed recommendation
10	0.6725	0.6826	0.6995
50	0.6738	0.68	0.6893
100	0.6782	0.6812	0.6899
150	0.675	0.6855	0.6901
200	0.6793	0.6801	0.6943

Summary

It can be clearly seen from Table 2 that the deep recommendation-based personalized recommendation system algorithm proposed by the article establishes the objective function and the training model, and has low deviation values through simulation experiments, which has high recommendation accuracy. Effectively improve the user's personalized recommendations.

Acknowledgements

This work was supported by the project of Nature Scientific Foundation of Heilongjiang Province (F2016038).

References

- [1] L.W. Huang, et al. A review of research on recommendation systems based on deep learning[J]. Chinese Journal of Computers, 2018, v.41; No.427(07):191-219.
- [2] W. Guo. Research on Multipath data transmission protocol based on wireless Sensor Networks[J]. International Journal of Future Generation Communication and Networking, 2016, 9(2):49-58.
- [3] Jennings B, Stadler R. Resource management in clouds: Survey and research challenges. Journal of Network and Systems Management, 2015, 23(3): 567-619.
- [4] K.Wang, Y.Wang, X. Hu, Y. Sun, D.J. Deng, A. Vinel and Y. Zhang Wireless big data computing in smart grid. IEEE Wireless Communications, 2017, 24(2): 58-64.
- [5] F.Liu and W.W. Guo. Study on grid scheduling model based on hierarchical scheduling model. International Journal of Grid Distribution Computing, 2015, 8(3): 1-10.
- [6] Feng L and Wei-Wei G. Recommendation Algorithm Based on Tag Time Weighting [C]// 2018 International Conference on Smart City and Systems Engineering (ICSCSE). IEEE Computer Society, 2018.
- [7] Wei-Wei G and Feng L .Application Research of Hadoop's Weibo Recommendation System Prototype Based on customer Dynamic Behavior [C]// 2018 International Conference on Smart City and Systems Engineering (ICSCSE). IEEE Computer Society, 2018.
- [8] W. Guo and F. Liu, Research on Parallel Algorithm Based On Hadoop Distributed Computing Platform, International Journal of Grid & Distributed Computing, (2015).
- [9] Z. H. Zhan, X.F. Liu, Y. J.Gong, J. Zhang, H. S. H. Chung and Y. Li. Cloud computing resource scheduling and a survey of its evolutionary approaches. ACM Computing Surveys (CSUR), 2015, 47(4): 63.
- [10] Garg R. Singh A. K. Adaptive workflow scheduling in grid computing based on dynamic resource availability. Engineering Science and Technology, an International Journal, 2015, 18(2): 256-269.