Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2020.010202
ISSN 2790-0983 Vol. 1, Issue 2: 10-18

Collaborative Filtering Algorithm based on Hybrid Machine Learning Optimization

Shengwei Qiu*

Department of Information Engineering, Heilongjiang International University, Harbin 150025, China

qiushengwei@hiu.net.cn

**corresponding author*

Keywords: Hybrid Machine Learning, Optimization Algorithm, Collaborative Filtering Algorithm, Recommendation Algorithm

Abstract: Under the influence of the mobile Internet era, users' patience is increasingly limited. Future recommendation algorithms should quickly respond to users' urgent needs to save users' time. Under the background of big data, how to ensure the relatively low complexity and high accuracy of information push and display is a very valuable topic. Therefore, this paper discusses and analyzes the collaborative filtering algorithm (CFA) based on hybrid machine learning (ML) optimization. In this paper, the research background and significance of CFA are first described, including the development status of collaborative filtering recommendation algorithm, the research status of association rule recommendation algorithm and particle swarm optimization algorithm, and the problems in collaborative filtering recommendation algorithm. A new hybrid CFA model is proposed based on the hybrid ML optimization. This paper designs a simulation table of three recommendation algorithms to verify the proposed CFA. The experiment shows that the recall rate of the CFA proposed in this paper based on the hybrid ML optimization is not lower than other recommendation algorithms, which verifies the effectiveness of this algorithm.

1. Introduction

Collaborative filtering recommendation algorithm can effectively solve the problem of information overload. However, with the dramatic increase of data volume, the data of the relative user rating matrix becomes more and more sparse. At the same time, the research and development of collaborative filtering recommendation algorithm also encounter problems. The CFA based on hybrid ML optimization solves the problem of data sparsity in collaborative filtering recommendation algorithm, and can not only provide users with more personalized services.

Therefore, research on CFA of the hybrid ML optimization has great practical significance.

CFA based on hybrid ML optimization has been studied and analyzed by many scholars at home and abroad. In the model-based collaborative filter method, the neural network is applied to some related work of the recommendation system. Limiting Boltzmann machine is the first work to apply neural network model to recommendation system. However, RBM's goal score prediction, rather than ToP-N recommendation, only considers the observed score as its loss function [1]. It is a technical challenge to incorporate the negative sampling required for ToP-N recommendation into RBM training. Peng J proposed Auto Rec, a new automatic encoder framework for collaborative filtering. Auto Rec is a compact and efficient trainable model [2].

This chapter first describes the project based collaborative filtering recommendation algorithm in detail, explains its concept, principle and algorithm process, and proposes the evaluation indicators of the algorithm. It analyzes the cosine similarity of the project similarity, and obtains the hybrid recommendation algorithm to be studied in this paper. Then, using the data in the dataset, it conducts numerical simulation on the three algorithms, and compares each evaluation indicator of the algorithm. According to the comparison results, it is verified that the CFA based on hybrid ML optimization has good reliability and stability [3-4].

2. Analysis of Collaborative Filtering Recommendation Algorithm

2.1. Principle of Collaborative Filtering Recommendation Algorithm

Collaborative filtering recommendation algorithm can quickly find useful information for users, and it is a reliable intelligent retrieval tool for users. It fully analyzes users' past historical behavior data, and actively recommends products and news that meet users' needs and interests. In the whole recommended algorithm work process, users do not need to provide any explicit requirements. The algorithm predicts the targets that are likely to be interested in the future based on the user's historical behavior data, and then makes recommendations [5-6]. Neighborhood collaborative filtering, also known as memory collaborative filtering, is characterized by intuitiveness, easy implementation, and a process that does not require too long training, and has been fully applied and developed [7-8].

2.2. Three Ideas of CFA for hybrid ML Optimization

2.2.1. Integral Type

The overall idea is to integrate a variety of different recommendation strategies into one algorithm to achieve hybrid design. The implementation method is to adjust the algorithm internally, so that different types of data can be well used for input. It is usually described that specific data

The processing step is to first process the data into the format required by a specific algorithm. The hybrid recommendation is divided into two hybrid designs: feature combination and feature supplement. The integrated hybrid design is shown in Figure 1 below.

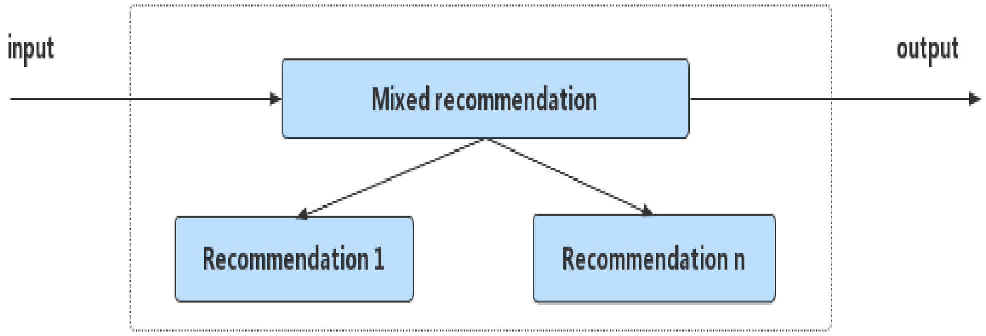

Figure 1. Integrated hybrid design

2.2.2. Flow Type

The pipelined hybrid design is a design method that divides the implementation process into different stages and knows the recommended results through the sequential effect of technology. The output of this stage can show the biggest difference between your pipelining scheme and other design schemes in the next stage [9]. That is to say, the design in the previous stage prepares for the preprocessing in the next stage. The constructed model can be used in the next stage, and the recommended list can also be given to optimize the calculation in the next stage [10].

There are two design ideas for the pipelining hybrid design scheme. The first one is serial mixing, where a group of recommended algorithms are connected in series. The former recommendation result is the input of the latter recommendation algorithm, and the latter recommendation algorithm optimizes the former recommendation result; The second is graded mixing [11-12]. The implementation steps of pipelining design are shown in Figure 2.

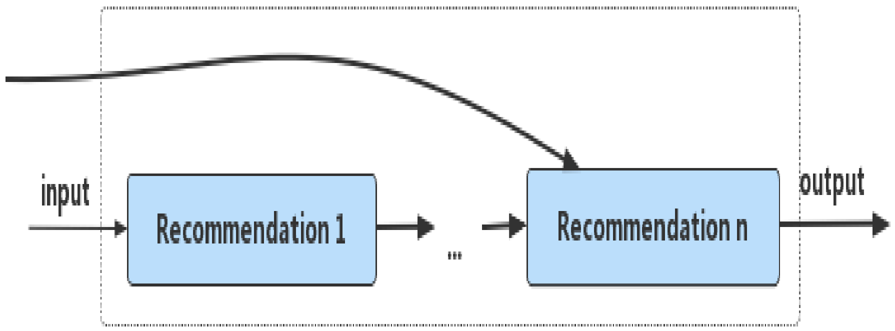

Figure 2. Flowing hybrid design

2.2.3. Parallel

Parallel hybrid design uses multiple different recommendation algorithms at the same time, and integrates the recommendation results formed by them with a specific hybrid mechanism. Its common strategies include weighting, crossing and switching. In this paper, we adopt the parallel strategy [13]. The implementation steps of parallel hybrid design are shown in Figure 3.

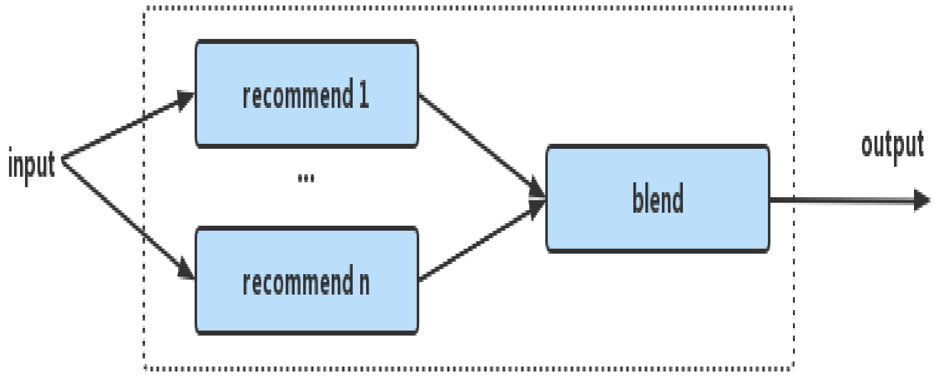

Figure 3. Parallel hybrid design

The cross strategy is to combine the recommendation results from different recommendation algorithms on the user interface interaction layer and recommend them to users. The switching hybrid recommendation strategy is that an authority decides which recommendation strategy to use under which conditions according to the recommendation quality of user history [14-15]. The weighted hybrid recommendation strategy is to weight different recommendation results, and the sum of the weighted values is 1.

3. CFA based on Hybrid ML Optimization

3.1. Improve Collaborative Filtering Recommendation Algorithm

According to the item matrix we have solved, we use calculation formulas to solve the distance of vectors, such as cosine distance calculation, Pearson distance calculation, Euclidean distance calculation and likelihood distance calculation to calculate the similarity between each item vector and item vector, and obtain the item similarity matrix [16].

The user matrix can be calculated according to the extracted raw data. We reorganize all the scored items of a single user, and then according to the above method, all users will get their corresponding user vectors, which will be combined into a user matrix. Find out K items similar to the products evaluated by a user, and form a recommended item matrix for users [17-18].

The algorithm first uses the scoring matrix to build the model, and then calculates the similarity between users or items to obtain the most similar neighbors. Finally, the recommendation engine will recommend the corresponding strategies according to the similar neighbors. Calculation method of logarithmic likelihood similarity:

3.1.1. Euclid Distance

Euclidean distance, also known as Euclidean distance, takes the object (or user) as the coordinate axis, and each user (or each item evaluated) who has participated in the evaluation of the project is regarded as a point in the N-dimensional space, and then the linear distance between the user and the object is calculated as their similarity. The Euclidean distance calculation formula between points in space is shown in formula (1):

$$d(x,y)=\sqrt{(\sum(x_i-y_i)^2)} \tag{1}$$

Wherein, user xi's rating of article i is described, and user yi's rating of article i is described. The calculated Euclidean distance is a positive number, and then it is normalized. At the same time, the result is controlled between 0 and 1, as shown in Formula (2):

$$sim(x,y)=\frac{1}{1+d(x,y)} \tag{2}$$

It can be seen from Formula (2) that as long as there is at least one item that has been scored jointly between users or two items that have been scored by the same user, the European distance formula can be used. If there is no common term, the Euclidean distance cannot be calculated.

3.1.2. Cosine Similarity

The cosine similarity calculation method of cosine similarity is to use the cosine value of the angle between two vectors in space to measure whether two vector individuals are similar. Compared with Euclidean distance method, cosine similarity pays more attention to the difference between their two vectors in the direction of each other than the difference in length or distance. The formula for calculating the similarity is shown in Formula (3):

$$\sin(x,y)=\cos\theta=\frac{\vec{x}\cdot\vec{y}}{\|x\|\cdot\|y\|} \tag{3}$$

Similar to the European distance, the preselected similarity algorithm also uses the user's preference for the item as a spatial point in the N-dimensional space, and then connects the point with the original point in the coordinate system to form a straight line. Because these spatial points connecting the user's score must intersect the origin, they must form an included angle between each other.

3.2. CFA Model based on Hybrid ML Optimization

We will integrate the improved CFA to get the final algorithm implementation process. It can be seen from Figure 4 that the basis of our hybrid recommendation algorithm is two single recommendation algorithms. The two single recommendation algorithms are hybrid ML optimization algorithm, improved cosine similarity recommendation algorithm for collaborative filtering. The final recommendation results can be obtained after the weighted average of the recommendation results of the two recommendation algorithms. Therefore, the first step in using the hybrid model is to use a single recommendation algorithm to predict the scores, and then use the scores of the two recommendation algorithms as the input data of the personalized hybrid model.

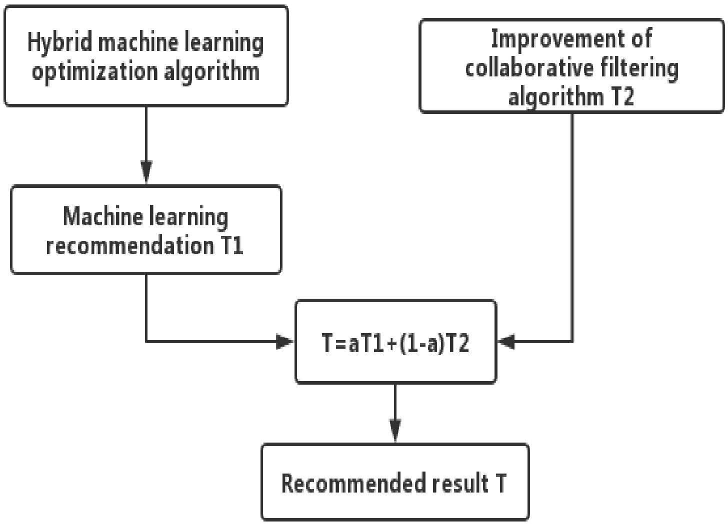

Figure 4. Flow chart of hybrid algorithm

Next, we give a mathematical definition of the parallel hybrid model. Suppose that user u scores item i as Rui when using the collaborative filtering recommendation algorithm, and the user scores item i as R'ui when using the association rule recommendation algorithm. No matter from a theoretical or practical point of view, we can know that the user's score must be between the CFA and the strong association rule recommendation algorithm, so we can regard the score obtained by using the model as a mixture of two scoring behaviors.

In order to simplify the model, we assume that the user's score is a linear combination of the two scores when establishing the hybrid model in this paper. In the final score of user u on item i, the proportion of collaborative filtering recommendation algorithm score is aui, and the proportion of association rule recommendation algorithm score is β Ui, so the user's final score is:

$$P_{ui} = a_{ui}R_{ui} + \beta_{ui}R_{ui} + e \tag{4}$$

Since the final score is obtained after the weighted average of collaborative filtering and association rule scores, the coefficients aui and β Ui satisfaction relation aui+ β ui=1 The experimental results show that the optimal parameters solved by the simplified model will be more accurate. In this paper, when solving au and u in the model, we did not use the solution method of linear equations. Instead, we used the optimized mean square error method, because: (i) we considered noise e in the model. The scoring matrix is a data sparse matrix. The objective function we use is:

$$\min_{a_{ui}\beta_{ui}} = \sum_{R\in S}(R_{u,j} - \overline{R}_u)^2 \tag{5}$$

4. Experimental Test Analysis

In many cases, when collecting and obtaining data, we will intentionally or unintentionally ignore some relevant information, but we need to ensure the integrity of key information in the data, such as the collected data should include behavioral objects or all user behaviors. Different datasets are usually used to describe different user behaviors. At present, there are four influential experimental datasets. Data set is a data set from a foreign film recommendation system website. At present, this data has been widely used in the research of various recommendation algorithms.

In order to verify the effectiveness of the CFA based on hybrid ML optimization studied in this paper, the improved CFA, ML algorithm and the recall rate of the algorithm in this paper are compared and analyzed; The following is the simulation analysis, and the comparison simulation results of the recall rates of the three algorithms are shown in Table 1 and Figure 5.

Table 1. Comparison of recall rates under different algorithms

	Improved CFA	ML algorithm	Hybrid algorithm
10	0.37	0.48	0.51
20	0.33	0.44	0.49
30	0.32	0.41	0.47
40	0.34	0.42	0.48
50	0.35	0.45	0.50
60	0.36	0.49	0.52

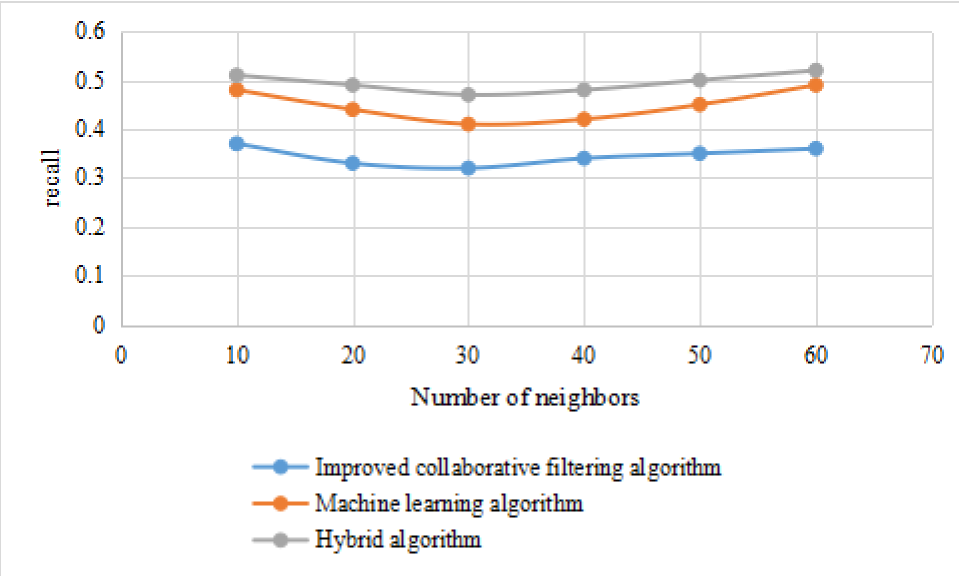

Figure 5. Comparison simulation trend chart of recall rate

The simulation results shown in the above figure show that the recall rates of the three algorithms do not change significantly with the increase of the number of neighbors. In general, people always hope that the accuracy rate and recall rate will increase with the number of neighbors, but this is an ideal state. Recall and accuracy often do not increase at the same time. Therefore, when we conduct experiments, we often make one increase while ensuring that the other does not decrease, which makes the hybrid recommendation algorithm have a high recall rate.

5. Conclusion

Most of the current collaborative filtering recommendation algorithms focus on the accuracy of the recommendation, but they often ignore that with the rapid development of mobile Internet, CFAs can provide services with high accuracy but are difficult to be recognized by users. The user's recommendation set formed by most collaborative filtering recommendation algorithms is the item set that users have not used or are interested in before. However, because of the sparse scoring data matrix, our understanding of users' preferences can only be part of their real interests. Therefore, this paper also needs to perfect the user's feedback and make full use of the various historical data left by users in the system, so as to analyze the historical access behavior of these users and get accurate prediction scores. The CFA based on hybrid ML optimization needs further research.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Peng J, Liu M, Zhang X, et al. Hybrid heuristic algorithm for multi-objective scheduling problem. Journal of Systems Engineering and Electronics, 2019, 30(02):109-124. https://doi.org/10.21629/JSEE.2019.02.12

[2] Dinarvand S, Nejad A M. Off-centered stagnation point flow of an experimental-based hybrid nanofluid impinging to a spinning disk with low to high non-alignments. International Journal of Numerical Methods for Heat & Fluid Flow, 2020, 32(8):2799-2818.

[3] Yousefi A, Pishvaee M S. A hybrid machine learning-optimization approach to pricing and train formation problem under demand uncertainty. RAIRO - Operations Research, 2020, 56(3):1429-1451.

[4] Mohammad Zare, Manfred Koch:Hybrid signal processing/machine learning and PSO optimization model for conjunctive management of surface-groundwater resources. Neural Comput. Appl. 33(13): 8067-8088 (2020) https://doi.org/10.1007/s00521-020-05553-8

[5] Wei Xie, Jie-Sheng Wang, Cheng Xing, Sha-Sha Guo, Meng-wei Guo, Ling-Feng Zhu:Adaptive Hybrid Soft-Sensor Model of Grinding Process Based on Regularized Extreme Learning Machine and Least Squares Support Vector Machine Optimized by Golden Sine Harris Hawk Optimization Algorithm. Complex. 2020: 6457517:1-6457517:26 (2020) https://doi.org/10.1155/2020/6457517

[6] Mohamed Riad Youcefi, Ahmed Hadjadj, Abdelhak Bentriou, Farouk Said Boukredera:Rate of penetration modeling using hybridization extreme learning machine and whale optimization algorithm. Earth Sci. Informatics 13(4): 1351-1368 (2020) https://doi.org/10.1007/s12145-020-00524-y

[7] Sizhou Sun, Jingqi Fu, Feng Zhu, Dajun Du:A hybrid structure of an extreme learning machine combined with feature selection, signal decomposition and parameter optimization for short-term wind speed forecasting. Trans. Inst. Meas. Control 42(1): 3-21 (2020)

https://doi.org/10.1177/0142331218771141

[8] Wei He, Changyin Sun, Donald C. Wunsch, Richard Yi Da Xu:Guest Editorial Special Issue on Intelligent Control Through Neural Learning and Optimization for Human-Machine Hybrid Systems. IEEE Trans. Neural Networks Learn. Syst. 30(12): 3530-3533 (2019) https://doi.org/10.1109/TNNLS.2019.2952699

[9] Alluvenkateswara Rao, Chanamallu Srinivasa Rao, Dharma Raj Cheruku:An enhanced copy-move forgery detection using machine learning based hybrid optimization model. Multim. Tools Appl. 81(18): 25383-25403 (2020)

[10] Atiye Yousefi, Mir Saman Pishvaee:A hybrid machine learning-optimization approach to pricing and train formation problem under demand uncertainty. RAIRO Oper. Res. 56(3): 1429-1451 (2020)

[11] Lenin Kanagasabai:Real Power loss reduction by hybrid pan troglodytes optimization: extreme learning machine based augmented sine: cosine algorithms. Int. J. Syst. Assur. Eng. Manag. 13(3): 1102-1120 (2020)

[12] Dun Li, Cui Wang, Lun Li, Zhiyun Zheng:Collaborative filtering algorithm with social information and dynamic time windows. Appl. Intell. 52(5): 5261-5272 (2020)

[13] Tao Long, Shunli Wang, Wen Cao, Pu Ren, Mingfang He, Carlos Fernandez:Collaborative state estimation of lithium-ion battery based on multi-time scale low-pass filter forgetting factor recursive least squares - double extended Kalman filtering algorithm. Int. J. Circuit Theory Appl. 50(6): 2108-2127 (2020) https://doi.org/10.1002/cta.3250

[14] Liu Na, Ming-Xia Li, Qiu Hai-yang, Hao-Long Su:A hybrid user-based collaborative filtering algorithm with topic model. Appl. Intell. 51(11): 7946-7959 (2020)

[15] Pei Tian:Collaborative filtering recommendation algorithm in cloud computing environment. Comput. Sci. Inf. Syst. 18(2): 517-534 (2020) https://doi.org/10.2298/CSIS200119008T

[16] Ajaegbu Chigozirim:An optimized item-based collaborative filtering algorithm. J. Ambient Intell. Humaniz. Comput. 12(12): 10629-10636 (2020) https://doi.org/10.1007/s12652-020-02876-1

[17] Neetu Kushwaha, Millie Pant:Fuzzy electromagnetic optimisation clustering algorithm for collaborative filtering. J. Exp. Theor. Artif. Intell. 33(4): 601-616 (2020) https://doi.org/10.1080/0952813X.2019.1647557

[18] Hansaim Lim, Lei Xie:A New Weighted Imputed Neighborhood-Regularized Tri-Factorization One-Class Collaborative Filtering Algorithm: Application to Target Gene Prediction of Transcription Factors. IEEE ACM Trans. Comput. Biol. Bioinform. 18(1): 126-137 (2020)https://doi.org/10.1109/TCBB.2020.2968442