International Core Journal of Engineering Volume 8 Issue 3, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

A Novel Deep Matrix Factorization Recommendation Model based on Attention Mechanism Fusion Side Information

Kunlin Miao[1,2], Changwei Zhao[1,2], Bin Song[1,2], and Zhiyong Zhang[1,2]

[1] School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

[2] Henan International Joint Laboratory of Cyberspace Security Applications, Henan University of Science and Technology, Luoyang 471023, China

Abstract

Recommendation systems have been extensively applied in various scenarios, such as e-commerce, personalized advertising, multimedia content push service, and so on. However, in the context of digital society and big data, a large amount of redundant interference information is mixed, which greatly affects the recommendation model and click rate prediction accuracy. This paper proposes a deep matrix factorization recommendation model and recommendation algorithm based on attention mechanism and side information. The attention mechanism is adopted to learn the weight between user-recommended items, eliminate the interference of redundant items on the model, learn the combined user-recommended item high-dimension and low-dimension data, perform operations at the output layer, and obtain the optimal recommendation list. Experimental comparative analysis based on real social network data sets shows that the proposed model is superior to the existing models in terms of precision, recall and comprehensive index F-measure, and has better superiority and usability.

Keywords

Recommendation System; Attention Mechanism; Side Information; Deep Matrix Factorization.

1. Introduction

With the advent of the era of big data, the scale of digital social data is growing exponentially. Information overload has become an important problem and challenge for users to obtain information[1]. The recommendation system is an effective method to solve the problem of information overload by analyzing the user's historical interaction data, mining the user's potential preferences, and providing users with personalized recommendation items (recommended items, content and other information services)[2]. At present, the recommendation system has become the basic functional configuration of the mobile Internet information infrastructure. Major social media platforms and portals have selected and deployed accurate recommendation systems.

The recommendation model is the core of the recommendation system. The traditional recommendation models take rating information or click information as input to predict user preferences. Restricted by factors such as privacy protection and users' reluctance to actively provide information, recommendation systems also face problems such as cold start and data sparseness. In the face of users with relative lack of interactive information, it is difficult for the system to achieve accurate prediction [3]. Although the matrix factorization algorithm has achieved great success in personalized recommendation, there are still some problems, such as sparse matrix elements and unstable calculation results. Because each recommended item is only considered separately, the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

correlation between the recommended item and the user is not considered. As a shallow model, the matrix factorization algorithm cannot effectively model the nonlinear relationship between the user and the recommended item, nor can it deeply understand the user. With the hidden features of recommended items, the recommendation accuracy is also very limited [4]. In addition, there are many factors affecting user interaction behavior in digital society, such as attribute information, image information, text information, label information and other types of side information, which contain rich user behaviors and needs [5]. Combining side information with scoring or click data streams becomes one of the most effective ways to solve the problem of data sparseness. Side information has complex characteristics such as multi-modality, data heterogeneity, large scale, sparse data and uneven distribution. Hybrid recommendation methods incorporating side information have become a research hotspot in the academic and industrial circles at home and abroad in recent years [6].

Recommendation systems incorporating side information can effectively utilize richer data to alleviate the cold start and data sparsity problems faced by traditional recommendation algorithms [7]. Sun[8] et al. used recommended item information (such as category, type, location, and brand, etc.) as side information to deeply analyze recommended item attributes and user preferences. However, considering that different interaction information has different effects on users' preferences, researchers have paid extensive attention to incorporating attention mechanism into recommendation system. He[9] et al. proposed an attention-based neural network recommendation item similarity model (NAIS), which uses the attention mechanism to distinguish which recommended items are most important for prediction. Wide & Deep[10] model aims to perform simultaneous interaction of low-dimension and high-dimension features on recommendation data by introducing a hybrid network structure, that is, the network structure contains wide layers and deep layers. He[11] et al. proposed a neural matrix factorization model, which adopts the automatic learning function method of multi-layer perceptron to learn the non-linear interaction between product ratings recommended by users. Kabbur[12] et al. proposed the Element Similarity Model (FISM) to improve the recommendation results by learning the similarity matrix between the recommended items, and associate two recommended items that were not purchased at the same time through a third recommended item. All of the above models use deep learning to acquire user characteristics, but there are still some limitations. For example, when users interact with multiple recommendations, these interactive features may reflect users' interest in different degrees and functions, so the lack of effective analysis and utilization of different features seriously affects the main performance of the recommendation model.

This paper proposes a new deep matrix factorization recommendation model, Deep Matrix Factorization Recommendation Model (DMFRM), which combines attention mechanism and deep matrix factorization. Through deep matrix decomposition, high-dimension and low-dimension features are interacted, and the importance of different features is distinguished by the attention mechanism to construct user preferences, which further improves the prediction accuracy. The main chapters of this paper are arranged as follows: Section 2 proposes a new deep matrix factorization recommendation model; Section 3 combines the new model to describe the recommendation algorithm; Section 4 compares and analyzes the proposed model and related models; Section 5 is summary and prospect.

2. Proposed Recommendation Model

2.1 Selection and Representation of Side Information

In the recommendation system, side information refers to other main characteristic information of the recommending subject or object in addition to the basic information, such as the recommending user's gender, age, occupation, geographic location and registration time, the category label of the recommended item, the recommended item's Basic description, etc. There is a certain correlation between side information and user preferences. Rational use of side information can help us better

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

mine user preference. The model proposed in this paper mainly selects the characteristic information of the user subject and the recommended item object, which can accurately analyze the users' needs. Mining the potential relationship between users and recommended items to achieve more accurate personalized recommendation services.

In the context of digital social big data, there are many kinds of side information in the Internet information system, but most of the features cannot be used directly, so it needs to be preprocessed to provide preparation for the input model. The features of users and recommended items of side information can be represented by one-hot encoding, p_i and q_j are the input information of the user i and the recommended item j respectively. After fusing the side information, it is expressed as follows:

$$\begin{aligned} p_i &= r_i \oplus s_i \oplus \cdots \\ q_j &= r_j \oplus s_j \oplus \cdots \end{aligned} \tag{1}$$

In formula (1), r_i and r_j represent the vector representation of the user i and the recommended item j, s_i and s_j represent the side information vector representation of the user and the recommended item, $\oplus$ represents the connector, p_i and q_j represents the input data of users and recommended items after splicing, which can flexibly add various useful information.

2.2 DMFRM Model

A new deep matrix factorization recommendation model, DMFRM (Deep Matrix Factorization Recommendation Model, DMFRM), is proposed by integrating side information and attention mechanism into the deep matrix factorization recommendation model, and its structure is shown in Figure 1. The model consists of input layer, embedding layer, hidden layer and output layer, which integrates the interaction information between users and recommended items, and other side information.

(1) The input layer integrates the feature data of users and recommended items.

(2) The embedding layer processes the input feature data of users and recommended items, and converts the input high-dimensional sparse original feature data into low-dimensional dense embedding vectors. Converting raw data into embedded vectors requires two steps: first, the raw data is converted into consecutive integers by building a map, and then those integers are converted into embedded vectors.

(3) The main function of the hidden layer is to learn the hidden representation between users and recommended items through a deep neural network. The attention mechanism is added to the hidden layer to learn the influence of different feature information on the whole in parallel, and the weight of feature information with greater global impact is increased. The high-dimension and low-dimension features obtained can interact with each other. Distinguish the importance of different interactive features, and use neural network representation to learn the attention mechanism to learn the weights of different features in interactive items, eliminate the interference of redundant items, and obtain preferences that are helpful to users in interactive items.

(4) The feature vector learned by the hidden layer is sent to the output layer to complete the click-through rate prediction. In this model, the recommendation prediction first extracts the target user and recommended item information from the user and recommended item information pool, and then generates latent factors from the user and recommended item feature conversion functions, and sends them to the fully connected layer to obtain the final predicted value $\widehat{R}_{ij}$.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

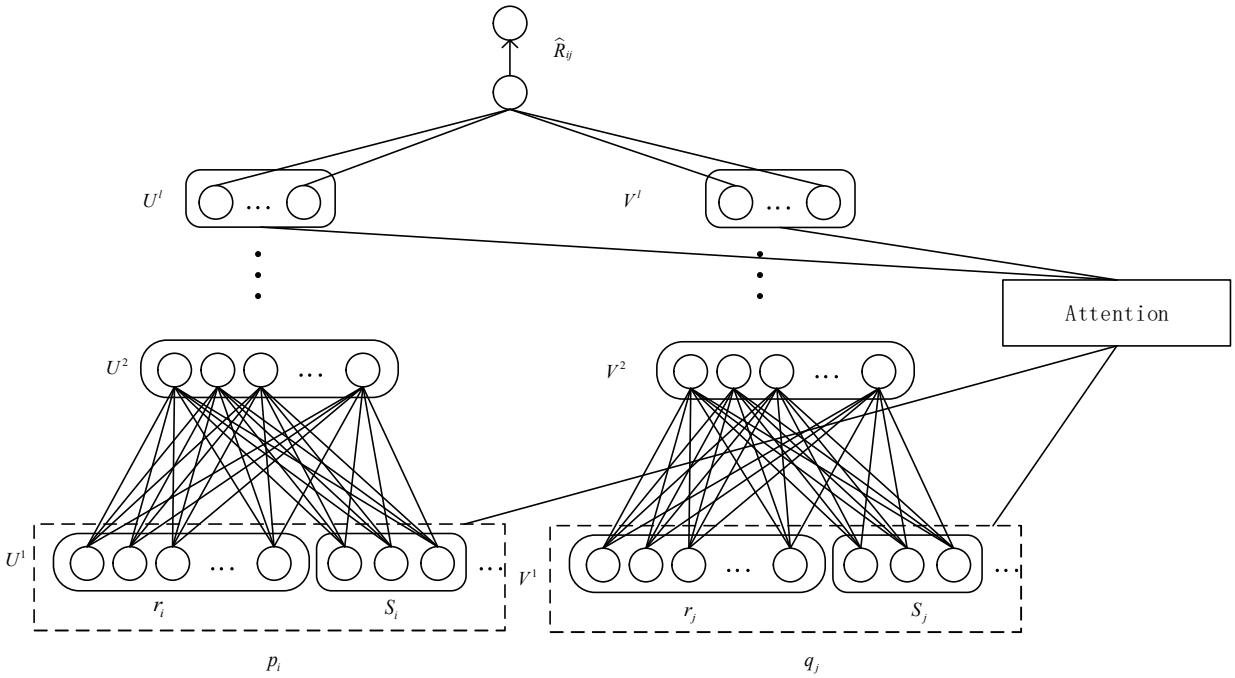

Figure 1. DMFRM model structure

2.3 Deep Matrix Factorization

The deep neural network part uses a multilayer perceptron to obtain a deep representation of the user. Input the concatenated data p_i, W_u represents the user weight, b_u represents the user's bias, and the active function uses Relu,

$$\begin{aligned} U^1 &= p_i \\ U^2 &= \mathrm{Re}\,lu(W_n^2 U^1 + b_u^2) \\ &\cdots \\ U^l &= \mathrm{Re}\,lu(W_u^l U^{l-1} + b_u^l) \end{aligned} \tag{2}$$

Similarly, the deep model factor representation of the recommended item can be obtained, w_v and b_v represent the weight and bias of the recommended item:

$$V^l = \mathrm{Re}\,lu(W_v^l U^{l-1} + b_v^l) \tag{3}$$

The multi-layer neural network structure for users and items can be constructed respectively, where input the characteristics of users and items into r_i and r_j respectively, and add side information s_i and s_j at the same time. After the combination, the latent factor vectors of users and items are established, which are sent to the multilayer perceptron, and two k-dimensional vectors are trained as the deep nonlinear feature representation of users and items.

2.4 Fusion Attention Mechanism

The user selects an item because the item contains some important features, and the attention mechanism is used to mine these hidden features. Because different features have different contributions to user preferences, different attention weights can be assigned to different features. To

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

relate the attention weight a_{ij} to the user and item embedding vectors p_i, q_j, parameterize a_{ij} as a function of the embedding vector as input:

$$a_{ij} = f\left(p_i, q_j\right) \tag{4}$$

The advantage of this is that even if the user has not interacted with the item, as long as it has been learned from the data, it can be used to estimate the attention weights. In this paper, a fully connected neural network is used as the attention network to learn different feature weights. The active function uses Relu, and a_{ij} calculation formula is as follows:

$$a_{ij} = W^T \operatorname{Re}lu(w(p_i \odot q_j^T) + b) \tag{5}$$

where W^T denotes the attention weight from the hidden layer to the input layer, w and b represents the weight and bias from the input layer to the hidden layer, respectively. Using the Softmax function can correctly standardize the attention and has good interpretability. The input of the attention network is the processed k-dimensional vector, and the output is the attention weight corresponding to the feature. After normalizing the obtained attention score by Softmax, the attention weight is converted into the form of probability distribution, and the formula is as follows:

$$a_{ij} = \frac{\exp(f(p_i, q_j))}{\sum_{j \in R+} \exp(f(p_i, q_j))} \tag{6}$$

The attention mechanism eliminates the influence of redundant features on the model by dynamically weighting the features, and the attention mechanism is added in this part to learn the user's attention weight for the recommended items.

2.5 Prediction and Optimization

Input the two interactions obtained into a fully connected layer to obtain the final prediction $\widehat{R}_{ij}$. Formally, the predictive model is as follows:

$$\widehat{R}_{ij} = p_i^T \left(\sum_{j \in R+} a_{ij} q_j\right) \tag{7}$$

The loss function is used to evaluate the inconsistency between the prediction and the reality of the model. The smaller the loss function, the better the model's performance. Observed user-item interactions as positive samples and negative samples from the remaining unobserved interactions. where R+ and R- represent the positive sampling set and negative sampling set, σ represents the active function Sigmoid, $\widehat{R}_{ij}$ denotes the prediction, represents the possibility that user i will rate item j, λ represents the strength of regularization, $\theta = \{\{p_i\}, \{q_j\}, w, b, h\}$.

$$Loss = -\frac{1}{N}\left(\sum_{(i,j \in R+)} \log \sigma(\widehat{R}_{ij}) + \sum_{(i,j \in R-)} \log(1 - \sigma(\widehat{R}_{ij}))\right) + \lambda \left\|\theta\right\|^2 \tag{8}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

In order to obtain the optimal value, the model optimizer used Adam, the optimization algorithm used the Mini-batch Gradient Descent (MBGD) algorithm. The batch size selected for each training is 512, and the parameters are updated iteratively to reduce the error and minimize the objective function. The experimental data set is randomly divided into training set (60%), test set (40%). The test set is used for performance evaluation.

3. Recommendation Algorithms

3.1 Recommendation Algorithm Process

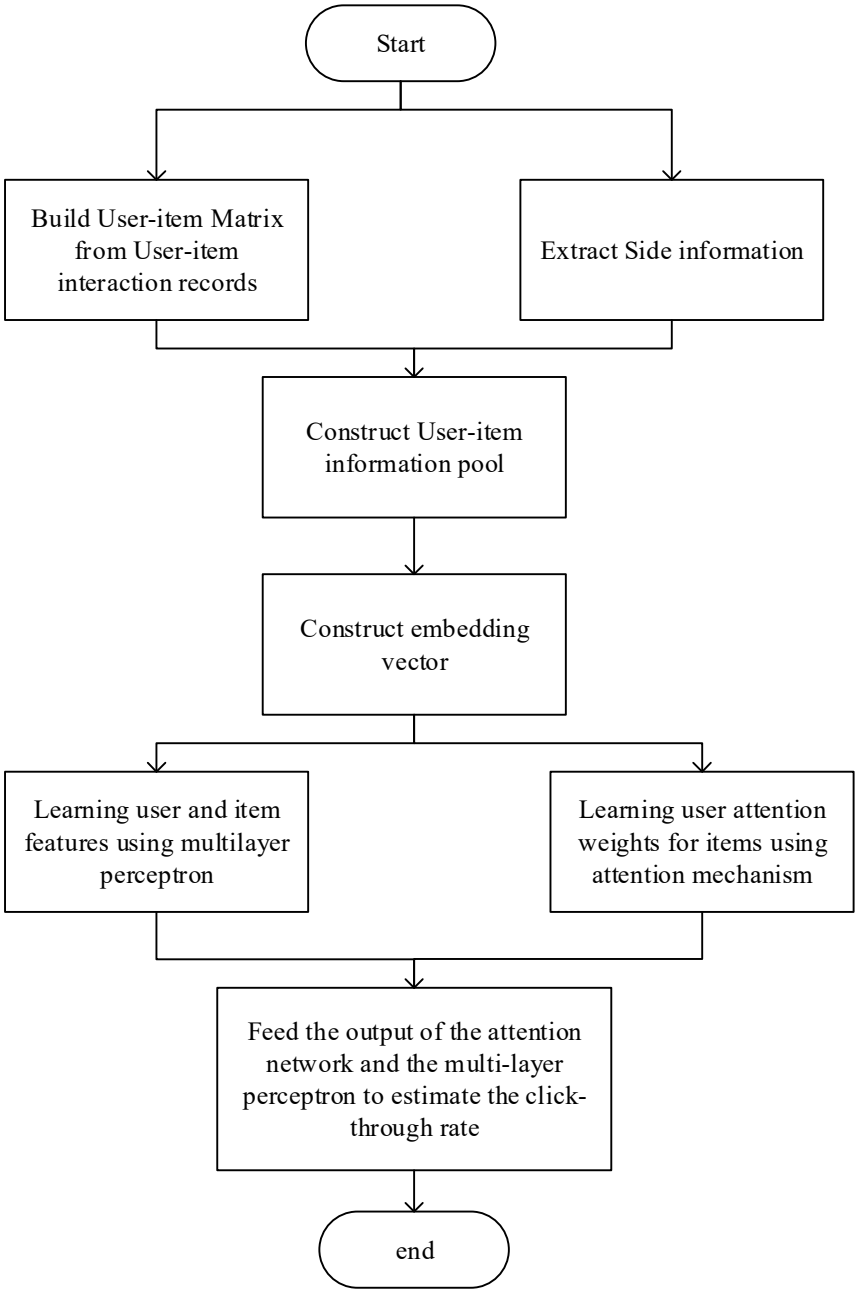

Figure 2. Flowchart of deep matrix factorization based on attention mechanism and side information.

Based on the new deep matrix factorization recommendation model proposed in this paper, the recommendation process is shown in Figure 2, and the specific steps are described as follows:

Step 1: Collect the user's browsing records, and construct a user-item interaction matrix. Extract features of side information for users and items.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

Step 2: Concatenate user features and item features with side information features respectively, and convert sparse features into dense vectors (embedding vectors) in vector space through embedding technology.

Step 3: the embedding vector is sent to the multi-layer perceptron, and the network is constructed through the full connection method, which is used for implicit high-dimension feature extraction to learn the nonlinear relationship between users and items. At the same time, the attention network is used to learn the combined feature weights, and the softmax function is used to normalize the obtained attention weights.

Step 4: The output of the attention network and the multi-layer perceptron is passed to the output layer to estimate the click rate, and the error is back-propagated through the loss function until convergence, and the model training is completed.

3.2 Description of the Recommendation Algorithm

The new deep matrix factorization algorithm uses the interaction matrix and side information as input to reduce the dimensionality of the sparse combined data and obtain the low-dimensional embedding vector. Feed into multilayer perceptrons and attention networks to learn nonlinear features and attention weights of users and items. The output result is sent to the output layer, and the inner product is used to obtain the final prediction result. Based on the DMFRM model and recommendation process, the corresponding recommendation algorithm is described as follows:

Algorithm 1: Deep Matrix Factorization Recommendation Algorithm Based on Attention Mechanism and Side Information

Input: User-item interaction matrix: $R_{ij,}$User side information set:S_i,Item side information set:S_j

Output: multimedia content to be recommended MediaID

Begin

Initialize parameter randomly

Link $R_{ij,}$ with S_i and S_j

For

For each i and j

Compute the U^l, V^l of user and item via equation (2), (3)

Compute the a_{ij} of user and item via equation (4), (5), (6)

Compute the $\widehat{R}_{ij}$ via equation (7)

End for

Update *Loss* via equation (8)

until convergence

End for

End

4. Experiments

4.1 Experimental Settings

In this paper, a real-world technology community platform (CyVOD) dataset is used to experimentally verify the proposed model [16]. The platform obtains user click behavior through data

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

embedding, collects user click data on the server side, and quantifies user behavior. This experiment selects 1000 users and 100,352 click data as the data set, which contains side information including the user's gender, profession, equipment used, comment information, geographic location, etc. The side information of the recommended project content mainly includes the description of the video, type of video, etc.

This paper adopts three important basic indicators used in the recommendation system, that is, precision, recall and F-measure. The recommended N recommended items for user u are recorded as R(u), and the set of recommended items that user u likes on the test set is recorded as T(u). Then, the accuracy of the recommendation algorithm can be evaluated by the precision rate and recall rate. The mathematical expression of the three indicators is as follows:

$$Precision = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u R(u)} \tag{9}$$

$$Recall = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u T(u)} \tag{10}$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

4.2 Comparison and Analysis

In order to prove the superiority and effectiveness of the model in this paper, the following related recommendation system models are selected for comparative analysis.

FM [14]: Factorization Machine, which has the advantages of SVM and factorization models.

SLIM [15]: A sparse linear recommendation algorithm, which mainly uses a sparse matrix as a weight to complete matrix scoring.

NCF [9]: Neural matrix factorization model, which uses implicit feedback as input and learns user-item interaction information through neural networks.

SituRecommender[16]: A recommendation algorithm based on social context analysis and collaborative filtering, which predicts the target user's rating based on the target user's nearest neighbors and historical behavior.

DMFRM: An attention mechanism recommendation model incorporating side information, which integrates side information for recommendation on the basis of attention mechanism and matrix decomposition.

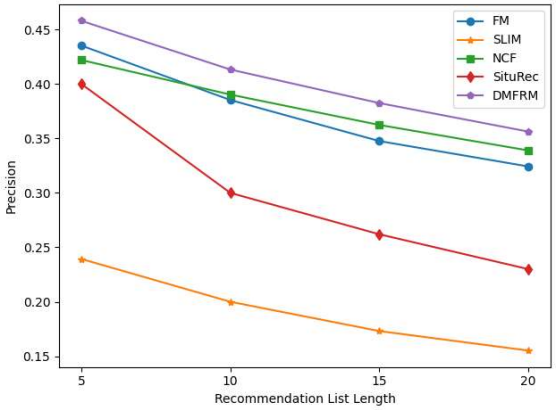

Figure 3. Precision comparison of five recommendation models

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

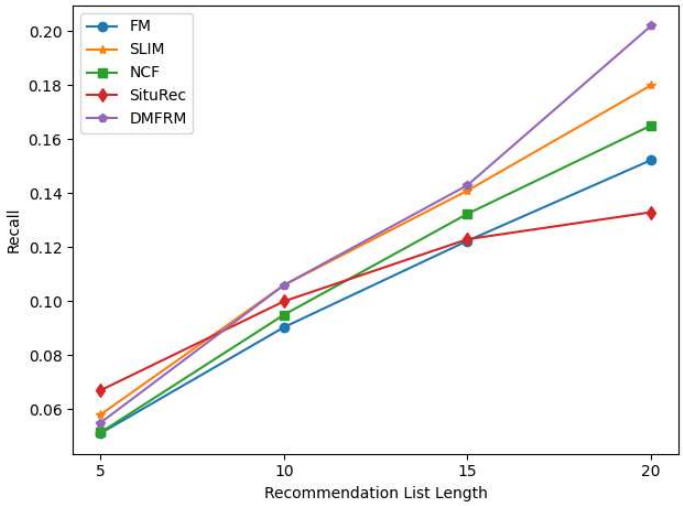

Figure 4. Recall comparison of five recommendation models

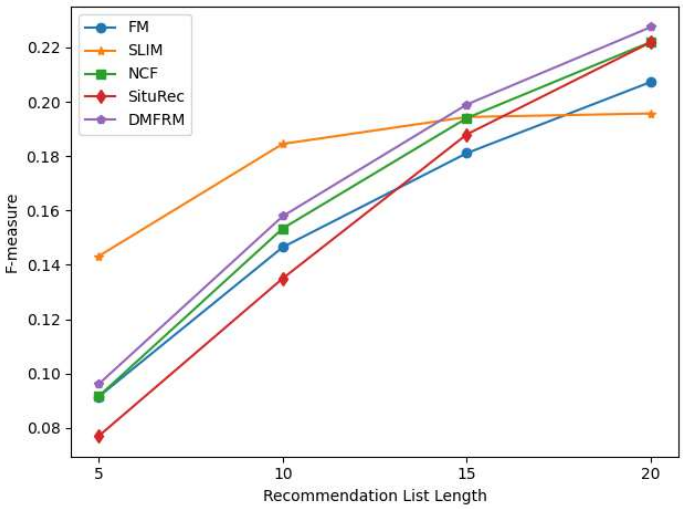

Figure 5. F-measure comparison of five recommendation models

Based on the CyVOD platform experiment, the Precision, Recall, and comprehensive evaluation index F-measure of five algorithms(FM, SLIM, NCF, Situ Recommender and DMFRM) on recommendation lists of four lengths are tested. The comparison results are shown in the table 1,Table 2, Table 3, Figure 3, Figure 4 and Figure 5.

As shown in Figure 3, the precision will decrease with the increase of the length of the recommendation list. When the dataset and the length of recommendation list are the same, the precision of the DMFRM model in the recommendation list of four lengths is significantly higher than that of the other four recommendation models. As can be seen from Figure 4, the recall increases with the length of the recommendation list, and the DMFRM model begins to perform better when the list length exceeds 10. As can be seen in Figure 5, the F-measure increases with the list length, and the DMFRM model outperforms the other four models when the recommended list length is 15, 20. As can be seen from Figure 3, Figure 4, and Figure 5, when the length of the recommendation list exceeds 10, the recommendation model proposed in this paper is superior to the two classical models and the proposed models in recent years in terms of accuracy, recall and F-measure.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

Table 1. Precision comparison of five recommended models

	FM	SLIM	NCF	SituRe-commend	DMFRM
5	0.435	0.235	0.422	0.437	0.458
10	0.385	0.203	0.390	0.403	0.413
15	0.347	0.173	0.362	0.400	0.382
20	0.324	0.155	0.339	0.332	0.356

Table 2. Recall comparison of five recommended models

	FM	SLIM	NCF	SituRe-commend	DMFRM
5	0.051	0.058	0.051	0.044	0.055
10	0.090	0.106	0.095	0.100	0.106
15	0.122	0.141	0.132	0.133	0.143
20	0.152	0.180	0.165	0.172	0.202

Table 3. F-measure comparison of five recommended models

	FM	SLIM	NCF	SituRe-commend	DMFRM
5	0.092	0.143	0.092	0.077	0.097
10	0.147	0.184	0.153	0.135	0.158
15	0.181	0.192	0.194	0.188	0.205
20	0.207	0.194	0.222	0.221	0.288

As can be seen from Table 1, Table 2 and Table 3, the precision of the model proposed in this paper is increased by 7% on average, the recall is increased by 1.5% on average, and the comprehensive evaluation index F-measure is increased by 2.8% on average. All three evaluation metrics are better than the compare models.

5. Conclusion

Recommendation system has become an important part of the mobile Internet and big data platforms, and a lot of related research work has been conducted at home and abroad. How to identify the user and recommended item data features effectively and improve the accuracy of recommendations has become the focus of current click-through rate prediction models. The innovative deep matrix factorization model proposed in this paper can extract high-dimension features and low-dimension features at the same time. Furthermore, the model can automatically learn the importance of different features, reduce the impact of redundant features effectively. The experimental results showed that the model in this paper is preferable to the current mainstream CTR prediction model. In the future, more valuable interactive information, such as knowledge maps, will be integrated to improve the accuracy and efficiency of Internet information content services.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0013

Acknowledgments

The work was supported by Project of Leading Talents in Science and Technology Innovation for Thousands of People Plan in Henan Province Grant No.204200510021, and Program for Henan Province Key Science and Technology No.222102210177, No.212102210383.We show gratitude to the reviewers and editor for their valuable comments, questions and suggestions.

References

[1] Hui L, Haining L, Shu Z, et al. Intelligent learning system based on personalized recommendation technology[J]. Neural Computing and Applications, 2018, 31:4455-4462.

[2] Xu C. A novel recommendation method based on social network using matrix factorization technique[J]. INFORMATION PROCESSING AND MANAGEMENT, 2018, 54(3):463-474.

[3] Huang L, Jiang B, Shou-Ye L, et al. Survey on Deep Learning Based Recommender Systems. [J]. Chinese Journal of Computers, 2018,41(07): 1619-1647.

[4] Fan J, Cheng J. Matrix completion by deep matrix factorization[J]. Neural Networks, 2017:34-41.

[5] Zhang F, Yuan N J, Lian D, et al. Collaborative knowledge base embedding for recommender systems[C]// Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016:353-362.

[6] Peng Y, Zhu W, Zhao Y, et al. Cross-media analysis and reasoning: advances and directions[J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18(1):44-57.

[7] Xue H, Dai X, Zhang J, et al. Deep matrix factorization models for recommender systems[J]. IJCAI International Joint Conference on Artificial Intelligence, 2017: 3203–3209.

[8] Sun Z, Yang J, Zhang J, et al. Recurrent knowledge graph embedding for effective recommendation[C]// Proceedings of the 12th ACM Conference on Recommender Systems,2018:297-305.

[9] He X,He Z,Song J,et al. NAIS:Neural attentive item similarity model for recommendation[J]. IEEE Transactions on Knowledge and Data Engineering,2018,30(12):2354-2366.

[10] Cheng H T, Koc L,Harmsen J,et al. Wide&deep learning for recommender systems[C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. ACM,2016:7-10.

[11] He X, Liao L, Zhang H, et al. Neural collaborative filtering[C]. Proceedings of the 26th international conference on world wide web,2017:173-182.

[12] Kabbur S,Ning X,Karypis G. FISM:Factored item similarity models for top- n recommender systems[C]// Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013:659-667.

[13] Zhang Z, Sun R, Zhao C, et al. CyVOD: A Novel Trinity Multimedia Social Network Scheme[J]. Multimedia Tools and Applications, 2017, 76(18): 18513-18529.

[14] Rendle S. Factorization Machines[C]// ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010. IEEE, 2010.

[15] Ning X,Karypis G. SLIM:Sparse linear methods for top-n recommender systems[C]//Proceedings of 2011 IEEE 11th International Conference on Data Mining,2011:497-506.

[16] Zhang Z, Sun R, Choo K, et al. A Novel Social Situation Analytics-Based Recommendation Algorithm for Multimedia Social Networks[J]. IEEE Access, 2019.