Papers recommendation with Item-level Collaborative Memory Network

First Author[1[0000-1111-2222-3333]] and Second Author[2[1111-2222-3333-4444]]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com

Abstract. The recommendation system can provide users with personalized and efficient recommendation information, which satisfies the user's demand for information in the information age, and has become a hot topic in the current era. In the recommendation system, users and items and the interaction of their own information has a crucial impact on the efficiency and accuracy of the recommendations. However, most of the existing recommendation systems usually design the systems as user-base only, considering the user's influence on the item in the recommendation, which to some extent blurs the interaction between items and users at the item level, unknown and potential connections between items and users are not well considered. In this paper, we are working on a collaborative memory network that can focus on the potential links between items and users, and consider the impact of items on user behavior. Experiments have shown that our improvement is better than the original method.

Keywords: Recommendation Systems, Memory Network, Collaborative Filtering.

1 INTRODUCTION

In the context of the era of big data, modern recommendation systems undoubtedly become an indispensable part of people's needs in their life. As an application that filters a large amount of information, the recommendation system can collect historical behavior data of the users, process the data through various processing methods or using relevant recommendation technologies, in order to obtain personalized recommendation for the users, recommend information or items such as movies, music, books, news, etc. that users may be interested in or like to users. The data of the recommendation system includes explicit data and implicit data, the explicit data can more accurately reflect the user's real preference on the items, such as user rating, etc., and the implicit data can reflect the user's preference after being analyzed and processed such as user clicks, watch time, etc. However, there are a large number of data and information that can be selected by the user. The effective recommendation system can provide the most relevant content for the user under the premise of large data volume, improve the efficiency and accuracy of the recommendation, and optimize the user experience.

The recommendation system based on collaborative filtering can analyze the dependence between users and items, and use the similarity information between users to recommend content that users are interested in [1,2.3]. The recommended personalization and automation are also high, at the same time, feedback can be used to speed up learning. Collaborative filtering technology is one of the most classic and mainstream technologies in recommendation systems. At the same time, applying matrix factorization (MF) to collaborative filtering technology is one of the most popular methods [4,5]. Many improvements are based on this. Singular value factorization (SVD) is also a matrix factorization method. But unlike feature factorization, SVD does not require that the factored matrix must be a square matrix [6]. It can represent a complex matrix with smaller and simpler sub-matrices, these sub-matrices describe the important characteristics of the original matrix, which has the advantage of simplifying the data, removing noise points, and improving the effect of the algorithm. However, in real life, the user's rating data is very rare, that is, explicit data is much less than implicit data, so how to add implicit data to the model becomes a problem. SVD++ integrates the user's implicit behavior into the SVD model, enhancing the user's implicit feedback [7].

Another recommended system algorithm is the K-nearest neighbor classification algorithm (KNN). The core idea of KNN is to find the majority of the k most neighboring samples in a feature space, if most of these samples belong to a certain category [8], then the sample also belongs to this category and has the characteristics of all samples in this category. The KNN method determines the category to which the sample to be classified belongs according to the category of the nearest one or several samples.When making the category decision, it is only related to a very small number of adjacent samples, and is more suitable for a set of samples to be classified that intersect or overlap in a category.

Recently, Tay et al. [9] proposed a flexible adaptive metric learning algorithm for collaborative filtering and ranking. By learning the adaptive relationship vector between user-item interaction, find an optimal translation vector between each user-item interaction pair,while introducing a relational vector to simulate the potential relationship between the user and the item. Hseih et al. [10] revealed the potential impact of internal products to simulate user-item relationships. Inspired by the memory network, Ebesu et al. [11] presented Collaborative Memory Network (CMN),they used the memory module as the nearest neighbor model of similar users, and used the attention mechanism based on specific users and items to learn the adaptive nonlinear weighting of the user neighborhood, while utilizing the adaptive neighborhood state and a non-linear interaction between the user and the item memory to generate recommendations.

Despite the success of CMN, it ignored some content. Each item has a connection with a group of users, and similar items do not necessarily have the same appeal to the same user [12]. When considering the relevance of a item to a user, the CMN only considers users with similarities to construct a neighborhood representation. However, in general, the problem of collaborative ranking has many-to-many nature, the item will also have a certain degree of impact on the user's choices and preferences, which may result in changes in the user's preferences.CMN ignores the potential relationships between items and users, especially those related to large data sets.

To solve this problem, we propose a network called Item-level Collaborative Memory Network (ICMN) for ranking and recommendation, which can fully consider the distribution of potential relationships between items and users, while strengthen the relationship between specific item and neighborhoods,it can take into account the impact of users and items on each other and their neighborhoods. Specifically, in order to be able to take into account the potential relationship between the item and the user, we adopt a higher-order method to obtain the weighted sum of the relationship between the specific item and the user and the fields they forms. In addition, we have separately considered the individual results of focusing on specific items and specific users. At the same time, we also considered the combination of focusing on specific items and specific users. Undoubtedly, our work has been inspired by the latest developments in NLP and recommendation systems [13,14,15]. The experimental results show that our method is better than CMN and can capture the hidden semantics between implicit interactions.

Our contributions are summarized as follows:

- We propose ICMN, a new type of network that takes into account the potential relationship between items and users, used to generate rankings for recommendations. For the first time, we have fully considered the distribution of potential relationships between items and users on CMN, and strengthened the relationship between specific items and neighborhoods.
- We evaluated our proposed ICMN on a public dataset. Experiments have shown that our proposed ICMN is not only superior to CMN, but also superior to many other powerful baselines with improved performance.
- We also consider the impact of items and users on each other and their neighborhood, and combine two modules that focus on specific items and specific users to demonstrate the importance of the item's impact on users. At the same time, neurological attention provides greater insight and interpretability for this model.

2 RELATED WORK

2.1 Collaborative Filtering

The rapid development of the Internet era, the construction of a large number of information systems, resources, and websites has enabled people to enjoy the rapid way of obtaining information, but they also have created the confusion of how to obtain the information they need efficiently and accurately. At the same time, users have also created more personalized needs, and recommendation systems have emerged. The recommendation system can solve the need for users to quickly and efficiently obtain the required information while improving the quality of service, and provides users with more personalized services. The collaborative filtering recommendation algorithm is the most classic and currently the most widely used algorithm in the recommendation system. It can also be divided into explicit collaborative filtering and im-

plicit collaborative filtering. The former learns the user's explicit information such as rating, score, etc., while the latter involves the user implicit interaction data.
Usually, in implicit collaborative filtering, if there is an interaction between the user and the item, it is recorded as 1, otherwise 0 [16].However, this does not indicate that the feedback between the user and the item that generated the interaction is positive or negative.

In a traditional data-based approach, all unobserved interactions are treated as negative samples and have the same weight, but unobserved interactions and data may be positive or missing data [17], and also have different weights. To this end, some recent efforts [18,19] have focused on weighting schemes and have considered whether unobserved samples are indeed negative. For example, certain non-uniform weighting schemes on negative samples, such as user-oriented and item popularity orientation, have been proposed and proven to be more effective than uniform weighting schemes [20].However, one major limitation of the non-uniform weighting method is that the weighting scheme is defined based on the hypothesis proposed by the authors, which may not be correct in the actual data. He et al. [21] weighted missing data according to item popularity. Yuan et al. [22] used the richness of content information in social media to find potential content for user-item pair with missing information to improve the accuracy of One Class Collaborative Filtering (OCCF).

2.2 Attention Mechanism

The attention mechanism has a great improvement effect on the sequence learning task. The attention model can perform data weighted transformation on the data sequence. Xu et al. [23] proposed two attention modes, one is hard attention and the other is soft attention, hard attention only pays attention to a certain position of the model sequence at each moment t, and soft attention takes care of all the positions at a time, and the weight of each position is different.Luong et al. [24]proposed two improved attentions, namely global attention and local attention. Among them, local attention can be regarded as a mixture of hard attention and soft attention. Unlike hard attention, local attention is almost everywhere. Easy and training, only focus on a small part of the source position at a time, and global attention needs to scan all source hidden states each time.

The traditional encoder-decoder structure relies on the limitation of an internal fixed-length vector in encoding and decoding [25], and shows very effective results in various machine learning tasks.

2.3 Deep Learning

Deep learning which stems from the research of artificial neural networks, is one of the latest trends in artificial intelligence and machine learning research, among the most popular scientific research trends today. It combines low-level features to form more abstract high-level representation attributes. Category or feature is used to represent the distributed characteristics of the data [26].The deep learning approach has brought historic progress to machine learning.In the study of machine learning, the motivation of deep learning lies in the establishment and simulation of the neural network for the analysis and learning of the human brain. It can mimic the mechanism of the human brain to interpret data such as images, sounds and texts.

In the recommendation system, the general idea of common algorithms is to convert high-dimensional discrete features into fixed-length continuous features through the Embedding layer, then pass through multiple fully-coupled layers, and finally convert to a 0-1 value through a sigmoid function, representing clicks probability [27].This method can fit high-order nonlinear relationships through neural networks while reducing the workload of artificial features.Recently, Zhou et al. [28] proposed a Deep Interest Evolution Network for Click-Through Rate Prediction, and proposed a Deep Interest Network (DIN) to adaptively derive historical behavior from an advertisement by designing a local activation unit. Learning the expression of user interest to address this challenge.

2.4 Collaborative Memory Network(CMN)

CMN is a recent deep architecture that demonstrates excellent and competitive performance in several baseline tests. In this network, memory components and neural attention mechanisms are merged into neighborhood components to learn the user's For a particular neighborhood, the ranking score is generated by the neighborhood and user memory [10].At the same time, the network can stack multiple memory modules together to create a more complex architecture for capturing deep and complex user-item interactions.

However, CMN is not without defects. CMN focuses its attention on the user's information, which leads to the neglect of information about the item.In the process of user interaction with the item, in addition to the explicit relationship, there is an implicit relationship. At the same time, in addition to the user's influence on the item, the item also has a non-negligible connection with the user. It must be equally attractive to the same user, and items that have not interacted with the user will affect the user's choice to some extent.

In real life, the channels through which users can accept item information are diverse, which means that different information will make the characteristics of the item have different effects on users.At the same time, it also shows that CMN cannot flexibly recommend in the case of huge data volume and dense data set.

3 OUR MODEL

In this section, we will introduce CMN and our network model ICMN that can focus on the item's attractiveness and potential relationships to users.The overall architecture and construction process of our network model,as shown in Figure 1, which includes the part of the item.Finally, we will introduce the relevant content that combines the user and the item's impact on each other and their neighborhoods to show the item's own influence, as shown in Figure 3.

3.1 Collaborative Memory Network

The CMN combines the memory component and the neural attention mechanism to form a neighborhood component that learns the user's specific neighborhood [10]. The stacking of the two memories enables them to obtain a deeper relationship between the user and the item.In a single-layer memory, given a user u, a preference vector q_{ui} for a particular user is formed, and then a similarity q_{uiv} of a particular user to a user in the user domain is derived, which is represented by $q_{uiv} = m_u^T m_v + e_i^T m_v$.When constructing the final neighborhood representation, the CMN constructs a neighborhood representation p_{uiv} for a particular user for subsequent operations.We focus on item information and features, embedding users and items, and using different methods to get the final neighborhood s_{iu} for a particular item. On this basis, we get the final ranking score.

3.2 User and Item Embedding

The input user and item information are embedded in the memory component of the user and the item respectively. The user's memory component is M, and the item's memory component is E, where $M \in R^{P \times d}$, $E \in R^{Q \times d}$, P, and Q are the number of users and items, d is the dimension of the memory. The user's preferences are stored in the user memory component slice m_u, and the specific features of the item are stored in e_i of the item memory slice to get the similarities a_{iut} and q_{uiv} for a particular item and a particular user in their given neighborhood:

$$a_{iut} = m_u^T e_t + e_i^T e_t \quad \forall t \in N(u) \tag{1}$$

$$q_{uiv} = m_u^T m_v + e_i^T m_v \quad \forall v \in N(i) \tag{2}$$

Where v is a specific user in the user's neighborhood with implicit feedback, and t is a specific item in the item neighborhood with implicit feedback.

3.3 ICMN-User and Item Neighborhood Attention

In order to focus on the influential set of items and users in the field, the attention mechanism can learn the adaptive weighting function, which is different from the traditional neighborhood method used in Pearson correlation or cosine similarity [29], adaptive weighting. The function does not need to set the number of users or items. At the same time, we do not need to pre-define or set the weighting function based on past experience in the learning process of the whole attention mechanism. In this way, we can better learn the unique contribution of the item. And the unique contributions of users:

$$x_{iut} = \frac{\exp(a_{iut})}{\sum_{g \in N(u)} \exp(a_{iut})} \quad \forall t \in N(u) \tag{3}$$

$$p_{uiv} = \frac{\exp(q_{uiv})}{\sum_{k \in N(i)} \exp(q_{uik})} \quad \forall v \in N(k) \tag{4}$$

Among them, the generated neighborhood distribution can be dynamically assigned to specific users and items.

The weights are then manipulated with the neighborhood by using the weights formed by the attention mechanism to get the final neighborhood representation.

$$\mathrm{s_{iu}} = \sum_{\mathrm{t \in N(u)}} \mathrm{x_{iut}} \times \mathrm{f_t} \tag{5}$$

$$\mathrm{o_{ui}} = \sum_{\mathrm{v \in N(i)}} \mathrm{p_{uiv}} \cdot \mathrm{c_v} \tag{6}$$

Where f_t is the vector of another embedded matrix of the item, c_v is the vector in the user's other embedded matrix, $\times$ is the operation at the matrix level, and $\cdot$ is the operation at the vector level, focusing on the respective characteristics of the user and the item, the attention mechanism The resulting weights are selectively weighted in different ways to the user and the item and its neighborhood to obtain the final weighted neighborhood.

This not only consider the impact of the item on the user, but also strengthens the relationship between the specific item and the neighborhood, taking into account the distribution of the potential relationship between the item and the user [30], as shown in Figure 1.

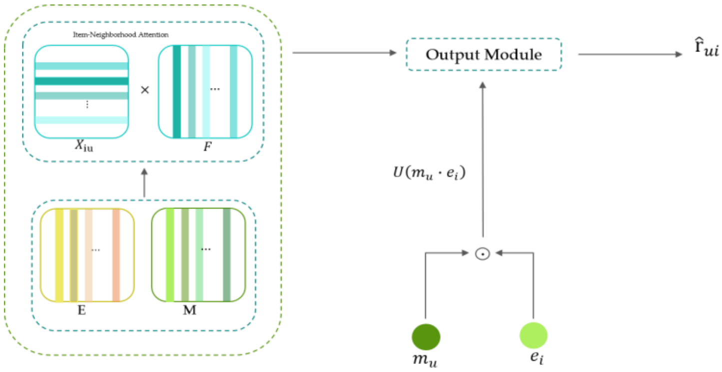

Fig. 1. A layer of memory structure about the item. Our model can focus on the item's attractiveness and potential relationships to users, and enhance the item's appeal to users.

3.4 Output Module

Capture the local structure of the item and the user's neighborhood and the global interaction between the item and the user, the item and the user have established a deeper and wider relationship, and then through a non-linear approach, can more thoroughly reflect the item A potential relationship with the user to generate a ranking score that focuses on the impact of the item:

$$\mathrm{\hat{r}_{iu}} = \mathrm{v^T} \emptyset \left(\mathrm{U(m_u \cdot e_i)} + \mathrm{W^2}_{\mathrm{s_{iu}^2}} + \mathrm{b} \right) \tag{7}$$

Where $\cdot$ is the operation of the element level. After obtaining the product of the item and the user, we operate the neighborhood representation and parameters obtained by learning in a linear projection manner. We use the nonlinear activation function

$\emptyset(x) = max(0, x)$, which can improve the expressive ability of the model [31,32]. In addition, when user information is added to consider the impact of the item and the user on the other party and its neighbors, the ranking score becomes:

$$\hat{r}_{iu}=v^{T}\emptyset\left(U(m_u \cdot e_i)+W^2{}_{s_{iu}^2}+W^2{}_{o_{ui}^2}+b\right) \tag{8}$$

3.5 Parameter Estimation

We use the Bayesian Personalized Sorting (BPR) optimization standard [33] as our loss function:

$$L=\sum_{(u,i^+,i^-)} \log \sigma\left(\hat{r}_{i^+u}-\hat{r}_{i^-u}\right) \tag{9}$$

Where $\sigma(x) = 1/(1 + exp(-x))$ is a logical sigmoid function.

3.6 Memory Stack

In the concrete model construction, we use a stack of two memoriese, the stacked memory can capture a deeper and wider information between the item and the user, improve the efficiency and performance of the entire model, as shown in Figure 2. Through the stacking of memories, you can dynamically assign weights to items, and capture the appeal of the item to users more comprehensively and in detail.

3.7 Integrated User and Item

Throughout the construction of the model, we not only focus on the item, consider the unique information of the item itself and the impact of the item on the user's appeal. At the same time, we also combine the user and the item's influence on each other and their neighborhoods, such a way to clearly show the non-negligible influence of the item itself., as shown in Figure 3.

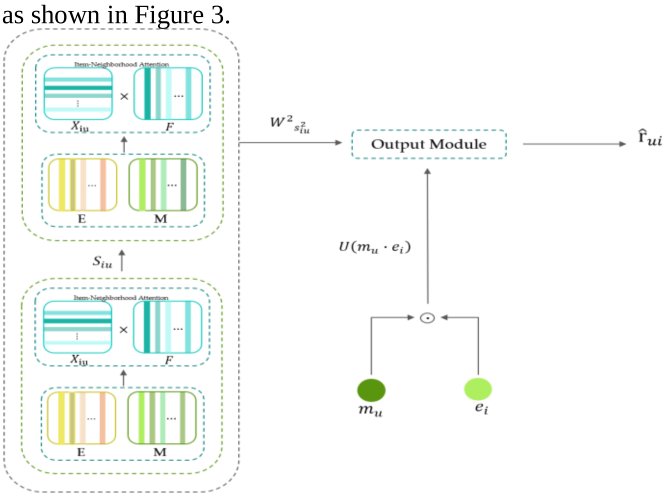

Fig. 2. Two memory stack. The stacking of the two memory components allows us to dynamically obtain the weight information generated by the attention mechanism.

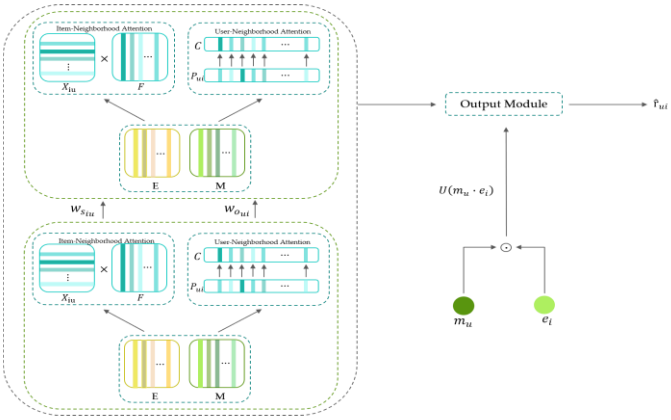

Fig. 3. Join the model of user information. Comparing with the original model after adding user information can clearly reflect the importance of item information and the non-negligible impact on users.

4 PERFORMANCE EVALUATION

In this section, we will introduce the data set and baseline settings used in the experiment, and experiment with the graphical data of the form.

4.1 Dataste

The dataset we use is the citeulike-a dataset [34], an online service collected from CiteULike, which allows users to create their own collection of articles. Each article has a summary, title and label, and the user provides a digital catalog to save and share academic papers, the dataset has 204,987 ratings and 5,551 users and 16,980 items.

4.2 Baselines

In this section, we will present the key baselines for comparison with our proposed ICMN.

SVD ++: SVD ++ [7] is a hybrid model that combines neighborhood-based similarity and latent factor models.

GMF: The Generalized Matrix Factorization (GMF) [35] model is a nonlinearly extended potential factor model.

KNN: KNN [8] is a domain-based method for calculating cosine item-item similarity.

BPR: Bayesian Personalized Sorting (BPR) [33] is a matrix factorization model of implicit feedback.

NeuMF: Neuromolecular Factorization (NeuMF) [32] is a matrix factorization model for item ordering through a multilayer perceptron model.

CDAE: Collaborative Denoising Autoencoder (CDAE) [36] is a item-based deep learning model.

FISM: Factority Similarity Model (FISM) [37] is a neighborhood-based model that can decompose the similarity matrix of item-item pairs and optimize the loss function.

CMN: Collaborative Memory Networking [11] is a model that fuses memory components and attention mechanisms and is the source of our ideas.

In addition, we set the number of negative samples as 4, and initialize in pre-training. In pre-training, l2's weight decay is set to 0.001. During training, l2's weight decay is set to 0.1, memory embedding. The size d is set to 50.

Table 1. Citeulike-a dataset experiment results.

	HR@5	HR@10	NDCG@5	NDCG@10
SVD++	0.6952	0.8199	0.5244	0.5649
GMF	0.7271	0.8326	0.5689	0.6034
KNN	0.6990	0.7348	0.5789	0.5909
BPR	0.6547	0.8083	0.4858	0.5357
NeuMF	0.7629	0.8647	0.5985	0.6316
CDAE	0.6799	0.8103	0.5106	0.5532
FISM	0.6727	0.8072	0.5106	0.5545
CMN	0.7959	0.8921	0.6185	0.6500
ICMN	**0.8031**	**0.8932**	**0.6372**	**0.6666**

4.3 Baseline Comparison

The data in Table 1 shows the experimental results of our model and each baseline model. We considered HR@5, HR@10, NDCG@5 and NDCG@10 [38] were used as the basis for the evaluation of experimental results.It can be seen that our results have good results and performance on these four indicators.

In Table 2, we made changes and comparisons of our own models. ICMN-u adds the user's influence on the item, considering the impact of the item and the user on each other and the neighborhood. At the same time, we also considered more memory components, for which ICMN-4 is a model containing four layers of memory. The performance improvement of ICMN-4 not only indicates the optimization of the performance brought by the memory component, but also shows that the processing of the weight by the attention component enables the model to obtain information more dynamically. Although the result of ICMN-u is better than the CMN model, but it is slightly insufficient compared with the model that focuses on item information. The comparison between ICMN and ICMN-u shows that the information and characteristics of the item itself have different degrees of non-negligible influence on the user and the potential connection between the item and the user. The experimental results show that our model is superior to all baselines.

Table 2. ICMN adjustment results.

	HR@5	HR@10	NDCG@5	NDCG@10
ICMN	0.8031	0.8932	0.6372	0.6666
ICMN-u	0.7977	0.8879	0.6308	0.6603
ICMN-4	**0.8044**	**0.8934**	**0.6434**	**0.6726**

5 CONCLUSION

In this paper, we propose an item-level collaborative memory network (ICMN) that can focus on the impact of the item level, which not only takes into account the attractiveness of the item to the user, but also fully considers the distribution of the potential relationship between the item and the user. To strengthen the relationship between specific items and neighborhoods, in addition, we also consider the impact of users and items on each other and the neighborhood.This enables our network to perform well on larger data sets.The attention mechanism also enables more rational, dynamic, and personalized processing of item features to obtain more comprehensive user preference information.We compared the other eight baselines on the citeulike-a dataset.

References

1. Hongyi Wen,Longqi Yang,Michael Sobolev,and Deborah Estrin.Exploring recommendations under user-controlled data filtering.In Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 18, 72-76. Vancouver, British Columbia, Canada,(2018).
2. Tiago Cunha, Carlos Soares, André C. P. L. F. de Carvalho. CF4CF: recommending collaborative filtering algorithms using collaborative filtering. In Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 18, 357-361. Vancouver, British Columbia, Canada,(2018).
3. Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John Riedl.Item-based collaborative filtering recommendation algorithms.In Proceedings of the 10th International World Wide Web Conference, WWW 10, 285-295. Hong Kong, China,(2001).
4. Panagiotis Symeonidis,.Matrix and Tensor Factorization in Recommender Systems. In Proceedings of the 10th ACM Conference on Recommender Systems, RecSys 16, 429-430. Boston, Massachusetts, USA,(2016).
5. Jorge Silva,Lawrence Carin. Active learning for online bayesian matrix factorization. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining.KDD 12,325-333.Beijing,China(2012).
6. Xilun Chen,K.Selcuk Candan. LWI-SVD: low-rank, windowed, incremental singular value factorizations on time-evolving data sets. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.KDD 14,987-996. New York, USA (2014).
7. Yehuda Koren.Factorization meets the neighborhood: a multifaceted collaborative filtering model.In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining.KDD 08.426-434. Las Vegas, Nevada, USA(2008).

8. Francesco Ricci,Lior Rokach,and Bracha Shapira. Introduction to recommender systems handbook. Recommender Systems Handbook.1-35. Springer(2011).
9. Yi Tay,Luu Anh Tuan,Siu Cheung Hui. Latent Relational Metric Learning via Memory-based Attention for Collaborative Ranking. In Proceedings of the 2018 World Wide Web Conference.WWW 18.729-739. Lyon, France(2018).
10. Cheng-Kang Hsieh,Longqi Yang,Yin Cui,Tsung-Yi Lin,Serge J. Belongie,and Deborah Estrin.Collaborative Metric Learning.In Proceedings of the 26th International Conference on World Wide Web, WWW 17, 193–201.Perth, Australia(2017).
11. Travis Ebesu,Bin Shen,Yi Fang. Collaborative Memory Network for Recommendation Systems. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval.SIGIR 18.515-524. Ann Arbor, MI, USA(2018).
12. Jingyuan Chan,Hanwang Zhang,Xiangnan He and Liqiang Nie.Attentive Collaborative Filtering:Multimedia Recommendation with Item- and Component-Level Attention. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 17.335-344. Shinjuku, Tokyo, Japan,(2017).
13. Alessio Ferrari. Natural language requirements processing: from research to practice. In Proceedings of the 40th International Conference on Software.ICSE 18.536-537. Gothenburg, Sweden(2018).
14. Xavier Amatriain, Justin Basilico. Past, Present, and Future of Recommender Systems: An Industry Perspective. In Proceedings of the 10th ACM Conference on Recommender Systems.RecSys 16.211-214. Boston, Massachusetts(2016).
15. Alexandros Karatzoglou, Balázs Hidasi, Deep Learning for Recommender Systems. In Proceedings of the Eleventh ACM Conference on Recommender Systems.RecSys 17.396-397. Como, Italy(2017).
16. Nathan N. Liu, Evan W. Xiang, Min Zhao, Qiang Yang.Unifying explicit and implicit feedback for collaborative filtering. In Proceedings of the 19th ACM international conference on Information and knowledge management.CIKM 10.1445-1448. Toronto, ON, Canada(2010).
17. 传统的基于全数据的方法
18. Jun Wang,Lantao Yu,Weinan Zhang,Yu Gong,Yinghui Xu,Benyou Wang,Peng Zhang,and Dell Zhang. IRGAN:A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models.In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval.SIGIR 17.515-524. Shinjuku, Tokyo, Japan (2017).
19. Dzmitry Bahdanau,Kyunghyun Cho,and Yoshua Bengio.Neural machine translation by jointly learning to align and translate. In Proceedings of the 2015 International Conference on Learning Representations.ICLR 15. San Diego (2015).
20. Qinyong Wang, Hongzhi Yin, Zhiting Hu, Defu Lian, Hao Wang, Zi Huang. Neural Memory Streaming Recommender Networks with Adversarial Training. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.KDD 18.2467-2475. London, United Kingdom(2018).
21. Xiangnan He,Hanwang Zhang,Min-Yen Kan,and Tat-Seng Chua.Fast Matrix Factorization for Online Recommendation with Implicit Feedback. In SiGIR.549–558(2016).
22. YUAN
23. Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention.In Proceedings of the 32nd International Conference on Machine Learning, ICML 15.Lille, France(2015).

24. Minh-Thang Luong, Hieu Pham, Christopher D. Manning Effective Approaches to Attention-based Neural Machine Translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing,EMNLP 15. 1412–1421.Lisbon, Portugal(2015).
25. Bing-Jie Sun, Huawei Shen, Jinhua Gao, Wentao Ouyang, Xueqi Cheng. A Non-negative Symmetric Encoder-Decoder Approach for Community Detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.597-606. Singapore, Singapore(2017).
26. Ian Goodfellow,Yoshua Bengio,and Aaron Courville.Deep learning.MIT Press.(2016)
27. Travis Ebesu and Yi Fang.Neural Semantic Personalized Ranking for item cold-start recommendation. Information Retrieval Journal 20,2,109–131(2017).
28. Guorui Zhou,Xiaoqiang Zhu,Chenru Song,Ying Fan. Deep Interest Evolution Network for Click-Through Rate Prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.KDD 18.1059-1068. London, United Kingdom(2018).
29. Francesco Ricci,Lior Rokach,and Bracha Shapira.Introduction to recommender systems handbook. Springer.(2011)
30. Homanga Bharadhwaj, Homin Park, Brian Y. Lim. RecGAN: recurrent generative adversarial networks for recommendation systems.In Proceedings of the 12th ACM Conference on Recommender Systems.RecSys 18.372-376. Vancouver, British Columbia, Canada(2018).
31. Alex Krizhevsky,Ilya Sutskever,Geoffrey E.Hinton. ImageNet classification with deep convolutional neural networks.Communications of the ACM.60.84-90(2017).
32. Giuseppe Amato,Fabio Carrara,Fabrizio Falchi, Claudio Gennaro.Efficient Indexing of Regional Maximum Activations of Convolutions using Full-Text Search Engines.In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval.ICMR 17.420-423. Bucharest, Romania(2017).
33. Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural Collaborative Filtering.In Proceedings of the 26th International Conference on World Wide Web.WWW 17.173-182. Perth, Australia(2017).
34. Chong Wang and David M Blei.2011. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.KDD 11.448-456. San Diego, California, USA (2017).
35. Steffen Rendle,Christoph Freudenthaler,Zeno Gantner,and Lars Schmidt-thieme. BPR:Bayesian Personalized Ranking from Implicit Feedback. UAI (2009).
36. Yao Wu,Christopher DuBois,Alice X.Zheng,and Martin Ester. Collaborative Denoising Auto-Encoders for Top-N Recommender Systems.In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining.WSDM 16.153-162. San Francisco, California, USA (2016).
37. Santosh Kabbur,Xia Ning,and George Karypis. Fism:factored item similarity models for top-n recommender systems.In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.KDD 13.659-667. Chicago, Illinois, USA (2013).
38. Francesco Ricci,Lior Rokach,and Bracha Shapira.Introduction to recommender systems handbook. Springer(2011).