Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2022 doi:10.20944/preprints202208.0353.v1

Multimodal Fusion Context-aware Semi-supervised Recommendation Method

Soren Pham, Arne Kamen, Jeffrey P Alarcón, Yehuda Hachulla,

Abstract—The advent of the era of big data will bring more convenience to people and greater development to society. But at the same time, it will also bring people the problem of 'information overload', i.e., when people are faced with huge information data, there are many redundant and worthless data. The redundant and worthless data information seriously interferes with the accurate selection of information data. Even though people can use Internet search engines to access information data, they cannot meet the individual needs of specific u sers i n s pecific co ntexts. Th e pe rsonalized ne eds of a pa rticular us er in a pa rticular co ntext. Th erefore, ho w to find useful and valuable information quickly has become one of the key issues in the development of big data. With the advent of the era of big data, recommendation systems, as an important technology to alleviate information overload, have been widely used in the field of e-commerce. Recommender systems suffer from a key problem: data sparsity. The sparsity of user history rating data causes insufficient training of collaborative filtering r ecommendation m odels, w hich l eads t o a s ignificant de crease in th e accuracy of re commendations. In fact, traditional recommendation systems tend to focus on scoring information and ignore the contextual context in which users interact. There are various contextual modal information in people's real life, which also plays an important role in the recommendation process. In this paper we achieve data degradation and feature extraction, solving the problem of sparse data in the recommendation process. An interaction context-aware sub-model is constructed based on a tensor decomposition model with interaction context information to model the specific i nfluence of in teraction co ntext in th e re commendation pr ocess. Th en an at tribute co ntext-aware su b-model is constructed based on the matrix decomposition model and using attribute context information to model the influence o f u ser a ttribute c ontexts and item attribute contexts on recommendations. In the process of building the model, the method not only utilizes the explicit feedback rating information of users in the original dataset, but also utilizes the interaction context and attribute context information of the implicit feedback and the unlabeled rating data. We evaluate our model by extensive experiments. The results illustrate the effectiveness of our recommender model.

Index Terms—Recommender; Multimodal; Context-aware

1 Introduction

In recent years with the rapid development of information technologies such as big data, cloud computing, and the size of data contained in the Internet has exploded. The scale of data is growing explosively. The advent of the era of big data will bring more convenience to people and greater development to society. But at the same time, it will also bring people the problem of 'information overload', i.e., when people are faced with huge information data, there are many redundant and worthless data. The redundant and worthless data information seriously interferes with the accurate selection of information data. Even though people can use Internet search engines to access information data, they cannot meet the individual needs of specific users in specific contexts. The personalized needs of a particular user in a particular context. Therefore, how to find useful and valuable information quickly has become one of the key issues in the development of big data. With the advent of the era of big data, recommendation systems, as an important technology to alleviate information overload, have been widely used in the field of e-commerce [1]. In order to provide better personalized recommendation services, accurate prediction of users' ratings of products is the key problem that recommendation systems need to solve. The current research directions in the field of recommendation systems are mainly divided into three types: content-based recommendation, collaborative filtering based recommendation [2, 3, 4] and hybrid recommendation [5, 6]. However, the traditional matrix decomposition algorithm [7]. It is difficult to adapt to the current complex environment. In recent years, deep learning techniques have developed rapidly and made great breakthroughs in areas such as images, and more and more scholars have applied deep learning techniques to recommender systems [8].

Recommender systems suffer from a key problem: data sparsity. The sparsity of user history rating data causes insufficient training of collaborative filtering recommendation models, which leads to a significant decrease in the accuracy of recommendations. In fact, traditional recommendation systems tend to focus on scoring information and ignore the contextual context in which users interact. There are various contextual modal information in people's real life, which also plays an important role in the recommendation process [9]. By incorporating these contextual modal information into the recommendation system, the effect of data sparsity can be appropriately mitigated to meet the needs of more personalized users. In the recommendation method based on deep learning and combined with textual information, the key is to obtain the context of textual

*Soren Pham is the corresponding author; soren_pham@hotmail.com

- *Soren Pham, Arne Kamen, Jeffrey P Alarcón, and Yehuda Hachulla are with Technical University of Denmark, Denmark. (e-mail: .dk).*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2022 doi:10.20944/preprints202208.0353.v1

information. In terms of extracting contextual features, Kim et al. [10] proposed a context-aware convolutional matrix decomposition model, which integrates convolutional neural networks into a probabilistic matrix decomposition model to improve the accuracy of prediction scores, but there is a large lack in context-awareness. liu et al. [11] proposed a CA-RNN model based on recurrent neural networks (RNN), which introduces a context-aware input matrix and a context-aware transfer matrix and is able to perceive context better, but is not effective in linking bidirectional contextual information. Devlin et al. [12] proposed the BERT model, which can fuse bidirectional contextual information and further fuse semantic information in sentences to better extract feature representations containing contextual information; in terms of learning contextual features, Hochreiter et al. [13] proposed the LSTM , which can only obtain antecedent information related to the antecedent word and cannot obtain contextual related information. Zheng et al. [14] proposed the BiLSTM based on the LSTM, which was designed to obtain the relationship between the current word and the context by designing LSTMs in 2 directions before and after, respectively. Most of the available recommendation algorithms use explicit rating information to make recommendations [15], but most of the users on the platforms only generate implicit interaction information such as user browsing and clicking, which makes the traditional recommendation algorithms based on rating prediction unable to meet the needs of the relevant platforms. In recent years, recommendation algorithms based on users' implicit historical feedback information have received extensive academic attention, and research has found that implicit feedback can be an alternative to explicit feedback in an interactive environment, which provides the possibility of using the implicit feedback matrix as input for movie recommendation in this paper. Currently, most of the mainstream recommendation models incorporating implicit feedback are based on the Bayesian personalized ranking framework [16]. With the widespread use of deep learning, Wu et al. [17] proposed cooperative noise-reducing autoencoders (denoisingauto-encoders), which used autoencoder technology combined with implicit feedback to obtain better recommendation results [18].

In response to the fact that traditional methods of representing contextual information often ignore the specific impact of contextual information, by fusing contextual features, we achieve data degradation and feature extraction, solving the problem of sparse data in the recommendation process. An interaction context-aware sub-model is constructed based on a tensor decomposition model with interaction context information to model the specific influence of interaction context in the recommendation process. Then an attribute context-aware sub-model is constructed based on the matrix decomposition model and using attribute context information to model the influence of user attribute contexts and item attribute contexts on recommendations. In the process of building the model, the method not only utilizes the explicit feedback rating information of users in the original dataset, but also utilizes the interaction

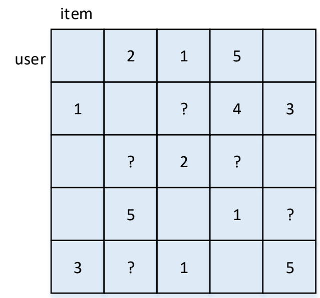

Fig. 1: Example of user-item rating matrix.

context and attribute context information of the implicit feedback and the unlabeled rating data. We evaluate our model by extensive experiments. The results illustrate the effectiveness of our recommender model.

2 Related Work

2.1 Recommender System

The purpose of a recommendation system is to tap into the user's preferences and interests to provide personalized recommendations [19, 20]. Collaborative filtering recommendation is a recommendation method that uses historical interaction behavior information to mine the views of other users with similar preferences to meet users' personalized needs [21, 22, 23]. The main advantages of collaborative filtering recommendation methods are the ability to handle comlicated item content, and the ability to discover more new user preferences, etc [24].

A classical collaborative filtering recommendation method was proposed by Resnick et al. [25] first constructed a social recommendation sub-model based on matrix decomposition, then constructed a social recommendation sub-model based on neighbors, and then fused the two sub-models into a collaborative filtering-based social recommendation master model. He et al. [26] proposed a new effective model learning algorithm and constructed an online recommendation model based on matrix decomposition technique using implicit feedback information. Yang et al. [27] proposed a social collaborative filtering recommendation model based on matrix decomposition technique using traditional user rating data of items and social credibility networks among similar users. As group-oriented recommendations become more and more popular in social networks, Ortega et al. [28] proposed a collaborative filtering recommendation model for group users using matrix decomposition techniques. The main drawback of collaborative filtering recommendation methods is that they are susceptible to the sparsity of users' historical rating data, which leads to inaccurate recommendations.

2.2 Data Sparsity Problem

In the recommender system, as the data for the recommendation becomes more, the rating matrix of the users and items becomes very sparse as an explicit feedback data. Therefore, the current data sparsity largely limits the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2022 doi:10.20944/preprints202208.0353.v1

accuracy of collaborative filtering a lgorithms, a nd some scholars are currently studying recommendation methods that can alleviate the data sparsity problem.

Zhang et al. [29] proposed to construct two types of matrix decomposition sub-models in the collaborative filtering a lgorithm o f r ecommendation s ystem a nd apply semi-supervised The two sub-models are mutually optimized using a semi-supervised co-training algorithm to alleviate the data sparsity problem. Qu et al. [30] in view of the data sparsity problem, proposed a semi-supervised collaborative training algorithm to optimize the three different views of text, image, and audio of the movie. Choi et al. [31] concluded that traditional collaborative filtering only treats the items with common user ratings equally when calculating user similarity, and is susceptible to data sparsity. Based on the principle of semi-supervised learning approach, the unrated elements are inferred and estimated using neighboring ratings. Jeong et al. [32] proposed an iterative collaborative filtering r ecommendation algorithm with semi-explicit ratings. The effect of data sparsity is mitigated. Wei et al. [33] reformulated the cold-start item representation learning from an information-theoretic standpoint to maximize the mutual dependencies between item content and collaborative signals.

2.3 Context-aware Recommender System

Context-aware recommender system(CARS) was first conceptualized by Adomavicius t al. [1]. A context-aware recommendation model based on multidimensional tensor decomposition is proposed by Karatzoglou et al. [34]. Rendle et al. [21] proposed a fast context-aware recommendation model based on factor decomposition technique, which expresses different contextual The model expresses different contextual information as corresponding contextual feature vectors, and fuses them directly with the user and item feature vectors at the same time to perform predictive scoring. Zhang et al [29] proposed a matrix decomposition-based attribute context-aware recommendation model that combines different types of model, in which different types of attribute context information are expressed as different attribute context bias variables in the matrix decomposition to thus performing predictive scoring. All the above recommendation models only consider the common impact of different contexts on users and items, but ignore the specific i mpact o f contexts on users and items respectively, i.e., the action-specific impact of contexts on users and the action-specific impact of contexts on items.

3 Methodology

In our recommender system, we use sets $U = \{u_1, u_2, \ldots, u_n\}$ and $V = \{v_1, v_2, \ldots, v_m\}$ to represent users and items respectively. The latent vectors of user u_i and item v_j are $u_i \in R^d$ and $v_j \in R^d$. The rating of item j by user i is denoted as $r_{i,j}$. Represents the value of multiple interaction contexts with $C_1, C_2, \ldots, C_n$. Define c a s the variable of context values C and use $(c_{1,k}, c_{2,k}, \ldots, c_{n,k})$ to denote the feature vector of the current interaction context k, where each context variable c can be represented by a d_c dimensional potential vector h. Thus the corresponding potential matrix for interaction context k can be expressed as:

$$H_k = [h_{1,k}, h_{2,k}, \ldots, h_{n,k}] \in R^{d_c \times n}, \quad (1)$$

In perceptual interaction contexts, different interaction contexts have similar or common contextual influences. We use the context operation tensor to represent this common contextual influence and the context potential vector to represent context-specific properties, thus defining the context operation matrix as the product of the corresponding context potential vector and the context operation tensor, which is used to represent the operational influence of the context on the user and the item, as follows:

$$M_{U,k} = a_k^t T_U^{[1:d]}, \quad (2)$$

$$M_{V,k} = a_k^t T_V^{[1:d]}, \quad (3)$$

The user's contextual operation matrix and the project's contextual operation matrix, constructed separately, can enables to perceive the context-specific contextual operation effects of interaction contexts on users and items, respectively. Where $M_{U,k}$ and $M_{V,k}$ denote the $d \times d$ dimensional contextual operation matrices of users and items in contextual situation k, respectively. $T_U^{[1:d]}$ and $T_V^{[1:d]}$ denote the $d \times d \times d$ dimensional context operation tensor, $[1 : d]$ denotes the tensor contains d pieces, and t denotes the transpose of the matrix. a_k denotes the contextual potential vector.

$$a_k = H_k W, \quad (4)$$

where H_k denotes the d-dimensional contextual potential matrix under contextual situation k, and W is an n-dimensional vector that represents denotes the weight of each context.

After perceiving the interaction context, the constructed contextual operation matrix represents the operational impact on the entity properties of the user and the item in the context. Thus the model can use the context operation matrix to manipulate the potential vectors dealing with users and items as follows:

$$u_{i,k} = M_{U,k} u_i, \quad (5)$$

$$v_{j,k} = M_{V,k} v_j, \quad (6)$$

where u_i and v_j are the original potential vectors of users and items.

$$u_{i,k} = (H_k W)^t T_U^{[1:d]} u_i, \quad (7)$$

$$v_{j,k} = (H_k W)^t T_V^{[1:d]} v_j, \quad (8)$$

where $u_{i,k}$ and $v_{j,k}$ denote the d-dimensional potential vectors of users and items, respectively, after being affected by the contextual context k operation.

After the potential vectors of users and items are affected by the contextual operation matrix operations, the prediction function of the model can be expressed as:

$$\hat{r}_{i,j} = w_0, + w_i + w_j + \sum_{m=1}^{n} w_{m,k} + u_{i.k}^t v_{j,k}, \quad (9)$$

where the prediction score $\hat{r}$ consists of six components: the global average bias w_0, the bias w_i of user i , the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2022 doi:10.20944/preprints202208.0353.v1

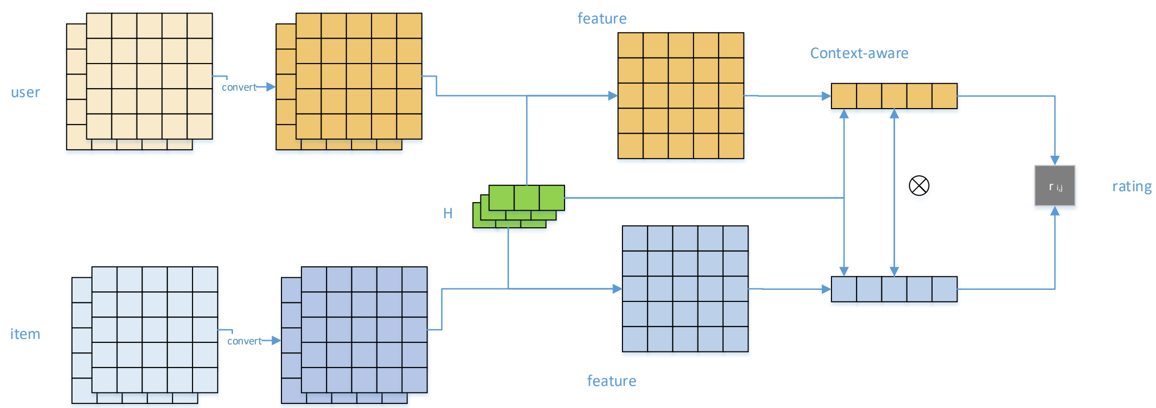

Fig. 2: Model structure diagram.

TABLE 1: Statistical information of datasets.

dataset	Movielens 100k	Movielens 1M
users	943	6040
items	1682	3706
rates	100000	1000209

bias w_j of item j , the bias $w_{m,k}$ of contextual values, and the d-dimensional potential vectors $u_{i.k}$ and $v_{j,k}$ of users and items after being affected by the contextual context k operations.

Based on these, the interaction context-aware score prediction model $h_1(i, j)$ is constructed as follows:

$$h_1(i,j) = \hat{r}_{i,j} = w_0 + w_i + w_j + \sum_{m=1}^{n} w_{m,k} \\ +[(H_k W)^t T_U^{[1:d]} u_i]^t (H_k W)^t T_V^{[1:d]} v_j, \tag{10}$$

To further optimize the model parameters, the following objective function of the model is defined:

$$minJ_1 = \sum_{r_{i,j} \in \omega} (r_{i,j} - \hat{r}_{i,j})^2 + \frac{2}{\lambda}(||U||^2 + ||V||^2 \\ +||H||^2 + ||T||^2 + ||W||^2), \tag{11}$$

where ω denotes the training set, λ denotes the regularization factor, and all parameters can be optimized by the following corresponding SGD formula.

$$\theta = \theta - \frac{\partial J_1(\theta)}{\partial \theta}, \tag{12}$$

where θ represents each parameters.

4 EXPERIMENTS

4.1 Dataset

To verify the validity of the model in this paper, Movielens 100k and Movielens 1M movie rating datasets were used to test and validate the performance of the proposed model and other models used for comparison. After reading the data, the scoring matrix is populated with the scoring data, with the user users as rows and items as columns, forming Movielens 100k and Movielens 1M matrices.

TABLE 2: Model MAE loss on two datasets.

	Movielens 100k	Movielens 1M
SVD	0.0748	0.0713
PMF	0.0752	0.0747
Ours	0.0472	0.0469

TABLE 3: Model MSE loss on two datasets.

	Movielens 100k	Movielens 1M
SVD	0.0493	0.0364
PMF	0.0465	0.0357
Ours	0.0402	0.0282

4.2 Metrics

The mean square error (MSE) and mean absolute error (MAE) are used to evaluate the predictive performance of the model, as:

$$MSE = \frac{1}{n} \sum_{u_i=1}^{n} (r_{u,i} - \hat{r}_{u,i})^2, \tag{13}$$

$$MAE = \frac{1}{n} \sum_{u_i=1}^{n} |r_{u,i} - \hat{r}_{u,i}|, \tag{14}$$

where n is the total number of predicted movies, $r_{u,i}$ denotes predicted rating of user u on movie i, and $\hat{r}_{u,i}$ denotes real rating of user u on movie i.

4.3 Result analysis

Because SVD [23] and PMF [35] are traditional algorithms that implement recommendations according to the matrix decomposition principle as well as the limitation of experimental conditions, these algorithms are selected for experimental comparison with the model in this paper, and MAE is used as the evaluation index. The MAE losses of different models on different data sets are shown in Table 2.

As can be seen from Table 2, the results of the model in this paper are significantly better than the compared traditional algorithms on 2 different data sets. In this paper's model, the process of training the model improves the efficiency of matrix decomposition and the prediction of the model, which results in lower loss values. However, the MAE values of individual models do not differ significantly on the Movielens 1M dataset and the Movielens 100k dataset. The reason may be that the MAE evaluation metric is more robust to outliers in the dataset, and the gradient of updates is always the same when training the model using a fixed learning rate, which is not conducive to the convergence of MAE values and the learning of the model.

As can be seen from Table 3, the loss values of the model in this paper are significantly lower than those of the other comparison models when the experiments are performed on the Movielens 1M dataset. And when the experiments are conducted on Movielens 100k, the loss values obtained are lower than the loss values of other comparison models. Analyzing the reasons, it can be seen that this paper's model uses deep learning for recommendation, which can better capture the deep semantic information, and therefore achieves better results compared with other models. Also observe Table 3, it is found that the loss of this paper's model on Movielens 100k increases compared to that on Movielens 1M, and there is a big difference. Analyzing the reason, it is clear that Movielens 100k has less data volume compared to Movielens 1M, so it may be more difficult to learn the influence of semantic information on user interest migration, which leads to higher loss values of the same model on different data sets.

5 Conclusion

Recommender systems suffer from a key problem: data sparsity. The sparsity of user history rating data causes insufficient training of collaborative filtering recommendation models, which leads to a significant decrease in the accuracy of recommendations. In fact, traditional recommendation systems tend to focus on scoring information and ignore the contextual context in which users interact. There are various contextual modal information in people's real life, which also plays an important role in the recommendation process. By incorporating these contextual modal information into the recommendation system, the effect of data sparsity can be appropriately mitigated to meet the needs of more personalized users. In the recommendation method based on deep learning and combined with textual information, the key is to obtain the context of textual information. In terms of extracting contextual features, Kim et al. proposed a context-aware convolutional matrix decomposition model, which integrates convolutional neural networks into a probabilistic matrix decomposition model to improve the accuracy of prediction scores, but there is a large lack in context-awareness. liu et al. proposed a CA-RNN model based on recurrent neural networks (RNN), which introduces a context-aware input matrix and a context-aware transfer matrix and is able to perceive context better, but is not effective in linking bidirectional contextual information. Devlin et al. proposed the BERT model, which can fuse bidirectional contextual information and further fuse semantic information in sentences to better extract feature representations containing contextual information; in terms of learning contextual features, Hochreiter et al. proposed the LSTM , which can only obtain antecedent information related to the antecedent word and cannot obtain contextual related information. Zheng et al. proposed the BiLSTM based on the LSTM, which was designed to obtain the relationship between the current word and the context by designing LSTMs in 2 directions before and after, respectively. Most of the available recommendation algorithms use explicit rating information to make recommendations, but most of the users on the platforms only generate implicit interaction information such as user browsing and clicking, which makes the traditional recommendation algorithms based on rating prediction unable to meet the needs of the relevant platforms. In recent years, recommendation algorithms based on users' implicit historical feedback information have received extensive academic attention, and research has found that implicit feedback can be an alternative to explicit feedback in an interactive environment, which provides the possibility of using the implicit feedback matrix as input for movie recommendation in this paper. Currently, most of the mainstream recommendation models incorporating implicit feedback are based on the Bayesian personalized ranking framework. With the widespread use of deep learning, Wu et al. proposed cooperative noise-reducing autoencoders (denoisingauto-encoders), which used autoencoder technology combined with implicit feedback to obtain better recommendation results.

In response to the fact that traditional methods of representing contextual information often ignore the specific impact of contextual information, by fusing contextual features, we achieve data degradation and feature extraction, solving the problem of sparse data in the recommendation process. An interaction context-aware sub-model is constructed based on a tensor decomposition model with interaction context information to model the specific influence of interaction context in the recommendation process. Then an attribute context-aware sub-model is constructed based on the matrix decomposition model and using attribute context information to model the influence of user attribute contexts and item attribute contexts on recommendations. In the process of building the model, the method not only utilizes the explicit feedback rating information of users in the original dataset, but also utilizes the interaction context and attribute context information of the implicit feedback and the unlabeled rating data. We evaluate our model by extensive experiments. The loss values of the model in this paper are significantly lower than those of the other comparison models when the experiments are performed on the two datasets. The results illustrate the effectiveness of our recommender model.

References

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE transactions on knowledge and data engineering*, vol. 17, no. 6, pp. 734–749, 2005.

[2] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5–53, 2004.

[3] M. D. Ekstrand, J. T. Riedl, J. A. Konstan *et al.*, "Collaborative filtering recommender systems," *Foundations and Trends® in Human–Computer Interaction*, vol. 4, no. 2, pp. 81–173, 2011.

[4] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, vol. 7, no. 1, pp. 76–80, 2003.

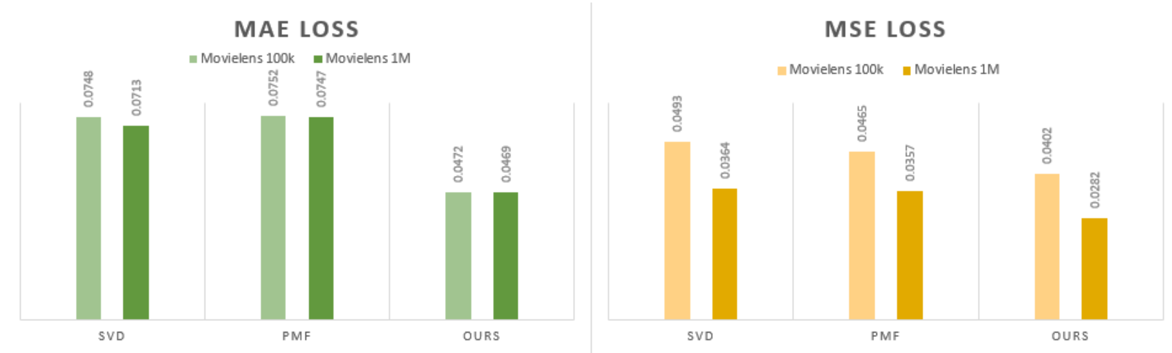

Fig. 3: Experiment result comparison on two datasets.

[5] R. Burke, "Hybrid recommender systems: Survey and experiments," *User modeling and user-adapted interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[6] M. Zanker and M. Jessenitschnig, "Case-studies on exploiting explicit customer requirements in recommender systems," *User Modeling and User-Adapted Interaction*, vol. 19, no. 1, pp. 133–166, 2009.

[7] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.

[8] Y. Deldjoo, M. Schedl, B. Hidasi, Y. Wei, and X. He, "Multimedia recommender systems: Algorithms and challenges," in *Recommender systems handbook*. Springer, 2022, pp. 973–1014.

[9] K. Verbert, N. Manouselis, X. Ochoa, M. Wolpers, H. Drachsler, I. Bosnic, and E. Duval, "Context-aware recommender systems for learning: a survey and future challenges," *IEEE transactions on learning technologies*, vol. 5, no. 4, pp. 318–335, 2012.

[10] D. Kim, C. Park, J. Oh, S. Lee, and H. Yu, "Convolutional matrix factorization for document context-aware recommendation," in *Proceedings of the 10th ACM conference on recommender systems*, 2016, pp. 233–240.

[11] Q. Liu, S. Wu, D. Wang, Z. Li, and L. Wang, "Context-aware sequential recommendation," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 2016, pp. 1053–1058.

[12] J. D. M. C. K. Lee and K. Toutanova, "Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] Z. Xiao and P. Liang, "Chinese sentiment analysis using bidirectional lstm with word embedding," in *International Conference on Cloud Computing and Security*. Springer, 2016, pp. 601–610.

[15] Y. Wei, X. Wang, L. Nie, X. He, and T.-S. Chua, "Graph-refined convolutional network for multimedia recommendation with implicit feedback," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 3541–3549.

[16] J. Pitkow *et al.*, "Personalized search: A content computer approach may prove a breakthrough in personalized search efficiency," *Communications of the ACM*, vol. 45, no. 9, 2002.

[17] Y. Wu, C. DuBois, A. X. Zheng, and M. Ester, "Collaborative denoising auto-encoders for top-n recommender systems," in *Proceedings of the ninth ACM international conference on web search and data mining*, 2016, pp. 153–162.

[18] A. Stuhlsatz, J. Lippel, and T. Zielke, "Feature extraction with deep neural networks by a generalized discriminant analysis," *IEEE transactions on neural networks and learning systems*, vol. 23, no. 4, pp. 596–608, 2012.

[19] Y. Wei, X. Wang, X. He, L. Nie, Y. Rui, and T.-S. Chua, "Hierarchical user intent graph network for multimedia recommendation," *IEEE Transactions on Multimedia*, 2021.

[20] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-based systems*, vol. 46, pp. 109–132, 2013.

[21] S. Rendle, Z. Gantner, C. Freudenthaler, and L. Schmidt-Thieme, "Fast context-aware recommendations with factorization machines," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 2011, pp. 635–644.

[22] Y. 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*, 2008, pp. 426–434.

[23] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[24] X. Yu, T. Gan, Y. Wei, Z. Cheng, and L. Nie, "Personalized item recommendation for second-hand trading platform," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 3478–3486.

[25] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 1994, pp. 175–186.

[26] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, "Fast matrix factorization for online recommendation with implicit feedback," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 549–558.

[27] B. Yang, Y. Lei, J. Liu, and W. Li, "Social collaborative filtering by trust," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1633–1647, 2016.

[28] F. Ortega, A. Hernando, J. Bobadilla, and J. H. Kang, "Recommending items to group of users using matrix factorization based collaborative filtering," *Information Sciences*, vol. 345, pp. 313–324, 2016.

[29] M. Zhang, J. Tang, X. Zhang, and X. Xue, "Addressing cold start in recommender systems: A semi-supervised co-training algorithm," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 73–82.

[30] W. Qu, K.-S. Song, Y.-F. Zhang, S. Feng, D.-L. Wang, and G. Yu, "A novel approach based on multi-view content analysis and semi-supervised enrichment for movie recommendation," *Journal of Computer Science and Technology*, vol. 28, no. 5, pp. 776–787, 2013.

[31] K. Choi and Y. Suh, "A new similarity function for selecting neighbors for each target item in collaborative filtering," *Knowledge-Based Systems*, vol. 37, pp. 146–153, 2013.

[32] B. Jeong, J. Lee, and H. Cho, "An iterative semi-explicit rating method for building collaborative recommender systems," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6181–6186, 2009.

[33] Y. Wei, X. Wang, Q. Li, L. Nie, Y. Li, X. Li, and T.-S. Chua, "Contrastive learning for cold-start recommendation," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 5382–5390.

[34] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, "Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 79–86.

[35] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," *Advances in neural information processing systems*, vol. 20, 2007.