Dual Contrastive Network for Sequential Recommendation

Guanyu Lin
Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology, Tsinghua University, China

Chen Gao†
Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology, Tsinghua University, China

Yinfeng Li
Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology, Tsinghua University, China

Yu Zheng
Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology, Tsinghua University, China

Zhiheng Li
Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University, China

Depeng Jin
Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology, Tsinghua University, China

Yong Li
Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology, Tsinghua University, China

ABSTRACT

Widely applied in today's recommender systems, sequential recommendation predicts the next interacted item for a given user via his/her historical item sequence. However, sequential recommendation suffers data sparsity issue like most recommenders. To extract auxiliary signals from the data, some recent works exploit self-supervised learning to generate augmented data via dropout strategy, which, however, leads to sparser sequential data and obscure signals. In this paper, we propose **D**ual **C**ontrastive **N**etwork (DCN) to boost sequential recommendation, from a new perspective of integrating auxiliary user-sequence for items. Specifically, we propose two kinds of contrastive learning. The first one is the dual representation contrastive learning that minimizes the distances between embeddings and sequence-representations of users/items. The second one is the dual interest contrastive learning which aims to self-supervise the static interest with the dynamic interest of next item prediction via auxiliary training. We also incorporate the auxiliary task of predicting next user for a given item's historical user sequence, which can capture the trends of items preferred by certain types of users. Experiments on benchmark datasets verify the effectiveness of our proposed method. Further ablation study also illustrates the boosting effect of the proposed components upon different sequential models.

†Chen Gao is the corresponding author (chgao96@gmail.com).

SIGIR '22, July 11–15, 2022, Madrid, Spain.

ACM ISBN 978-1-4503-8732-3/22/07.
https://doi.org/10.1145/3477495.3531918

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Computing methodologies** → *Neural networks.*

KEYWORDS

Sequential recommendation, Self-Supervised Learning, Contrastive Learning

ACM Reference Format:
Guanyu Lin, Chen Gao†, Yinfeng Li, Yu Zheng, Zhiheng Li, Depeng Jin, and Yong Li. 2022. Dual Contrastive Network for Sequential Recommendation. In *Proceedings of the 45th Int'l ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3477495.3531918

1 INTRODUCTION

Widespread in today's big data applications, recommender systems that vastly facilitate user's soaking or purchasing behavior are vital for online platform's business success. Sequential recommendation (SR) [19] that predicts a user's next interacted item is popular in both industrial applications and academic researches. Despite the achievements of existing SR models [7, 10, 17, 24], they suffer from data sparsity issue [20, 21]. To improve SR models, some recent works [16, 29] have attempted to exploit self-supervised learning [8]. However, existing self-supervised SR still encounters two bottlenecks as follows.

- **Sparse sequential behaviors**. The dropout strategy of existing works is not suitable for sparse sequential behaviors, leading to extremely sparse data, generating obscure self-supervised signals.
- **Obscure self-supervised signals**. Existing works that generate augmented data via dropout strategy will lead to sparser sequential data and unreliable signals.

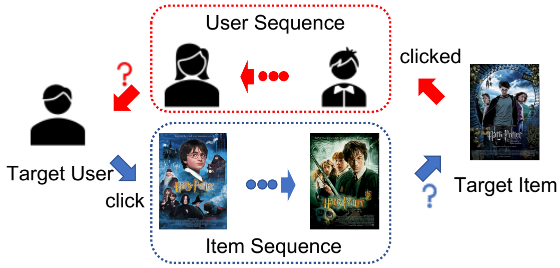

Figure 1: Illustration of next user and item prediction. (1) Traditional sequential recommendation based on item sequence (next item prediction) aims to predict the probability of the target item that the target user will click. (2) Our proposed next user prediction aims to predict the probability of the target user that the target item will be clicked by.

To address these two bottlenecks, we propose a novel solution named DCN (short for **D**ual **C**ontrastive **N**etwork for Sequential Recommendation) based on auxiliary user sequence. In contrast to item sequence that reflects the dynamic interest of the given user, user sequence reflects the dynamic trend of given item preferred by certain types of users. As illustrated in Figure 1, user sequence serves as a complementary role to traditional sequential recommendation based on item sequence. Specifically, we first encode the user/item sequence for a given target item/user based on user/item encoder, respectively. Then we treat the output of the user/item encoder as the representation of the given target item/user and perform contrastive learning to minimize the distance between it and the target item/user embedding, extracting self-supervised signals for representation learning. Finally, to capture the dynamic interest and static interest simultaneously, we perform contrastive learning on the next item prediction with the target user-item interaction prediction, as well as the next user prediction.

To summarize, the contributions of this paper are as follows.

- To the best of our knowledge, we take the first step to well exploit user sequence with self-supervised learning, *i.e.*, modeling user sequence and item sequence simultaneously, for the sequential recommendation.
- We perform contrastive learning from both representation-aspect and interest-aspect, respectively, to refine the user-item representation learning, and capture both the static and dynamic interest.
- We conduct experiments on two benchmark datasets where the results outperform the state-of-the-art models. Further experimental results of the ablation study well confirm the effectiveness of our designed self-supervision.

2 PROBLEM FORMULATION

We first introduce the symbols and then formulate the problem of sequential recommendation. We use $\mathcal{U}_i$ and $\mathcal{I}_u$ to denote the sequential sets of the given user u and given item i, respectively. Precisely, supposing $u_t \in \mathcal{U}_{i_{t+1}}$ and $i_t \in \mathcal{I}_{u_{t+1}}$ are the t-th user and item that a given target item i_{t+1} and target user i_{t+1} have interacted with. Then t-length historical sequences of users/items can be represented as $\mathcal{U}_{i_{t+1}} = (u_1, u_2, \ldots, u_t)$ and $\mathcal{I}_{u_{t+1}} = (i_1, i_2, \ldots, i_t)$, respectively. The sequential recommendation aims to better predict the probability that the target user *i.e.*, u_{t+1} will interact with the target item *i.e.*, i_{t+1}. With the above symbols, we can formulate the problem as follows.

Input: User sequence $\mathcal{U}_{i_{t+1}} = (u_1, u_2, \ldots, u_t)$ and item sequence $\mathcal{I}_{u_{t+1}} = (i_1, i_2, \ldots, i_t)$ for a given target item i_{t+1} and user u_{t+1}.
Output: The recommender which estimates the probability that target user u_{t+1} clicks item i_{t+1}.

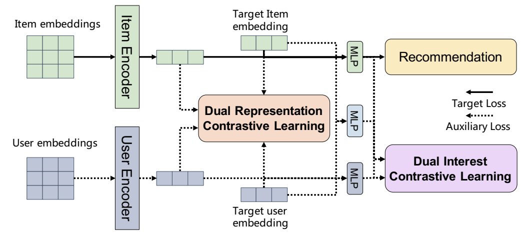

Figure 2: Illustration of our DCN. (1) Item embeddings and user embeddings are looked up based on the item sequence and user sequence; (2) Dual Representation Contrastive Learning self-supervises the output of item encoder with the target user embedding and that of user encoder with the target item embedding; (3) Dual Interest Contrastive Learning self-supervises the next item prediction with the target user-item interaction prediction and the next user prediction.

3 METHODOLOGY

Our DCN model is illustrated in Figure 2, enhancing the sequential recommendation with contrastive learning from representation aspect and interest aspect based on the novel auxiliary user sequence.

- **Sequential Encoding Layer**. We build a user encoder and item encoder to capture the user-to-user transition pattern and item-to-item transition pattern for the given target item and user.
- **Dual Representation Contrastive Layer**. We treat the outputs of the user encoder and item encoder as the representations for the target item and target user, respectively, and propose dual representation contrastive learning to self-supervise them with their embeddings.
- **Mixed Prediction Layer**. For the dynamic interest, we predict the next user (next item) by feeding the target user (item) embedding with the output of the user encoder (item encoder) into the prediction layer. For the static interest, we predict the interaction between the target user and the target item by feeding their embeddings into the prediction layer.
- **Dual Interest Contrastive Layer**. We propose dual interest contrastive learning to self-supervise the dynamic interest for next-item prediction with the static interest. Besides, we also self-supervise the next user score with the next item score.

3.1 Sequential Encoding Layer

3.1.1 ***Embedding Layer***. We create a user embedding matrix $\mathbf{M}^u \in \mathbb{R}^{n \times D}$ and an item embedding matrix $\mathbf{M}^i \in \mathbb{R}^{m \times D}$ where n and m denote the numbers of users and items, respectively. Here D denotes the latent dimensionality. Given user sequence $\mathcal{U}_{i_{t+1}} = (u_1, u_2, \ldots, u_t)$ and item sequence $\mathcal{I}_{u_{t+1}} = (i_1, i_2, \ldots, i_t)$, we lookup the input embeddings as:

$$\mathbf{E}^u = [\mathbf{M}^u_{u_1}, \mathbf{M}^u_{u_2}, \ldots, \mathbf{M}^u_{u_t}], \mathbf{E}^i = [\mathbf{M}^i_{i_1}, \mathbf{M}^i_{i_2}, \ldots, \mathbf{M}^i_{i_t}] \tag{1}$$

where $\mathbf{E}^u$ and $\mathbf{E}^i \in \mathbb{R}^{T \times D}$ denote the user and item embeddings for the model inputs, respectively.

3.1.2 ***User Encoder and Item Encoder.*** As we have obtained $\mathbf{E}^u$ and $\mathbf{E}^i$ from above embedding layers, we then exploit two single sequential models as user encoder and item encoder to learn the sequential patterns for user sequence and item sequence, respectively, as follows,

$$\boldsymbol{h}^u_{i_{t+1}} = \mathbf{Encoder}_u(\mathbf{E}^u), \boldsymbol{h}^i_{u_{t+1}} = \mathbf{Encoder}_i(\mathbf{E}^i) \tag{2}$$

where $\boldsymbol{h}^u_{i_{t+1}}$ and $\boldsymbol{h}^i_{u_{t+1}}$ are the hidden states for the user sequence of target item i_{t+1} and the item sequence of target user u_{t+1}, respectively. Here $\mathbf{Encoder}_u$ and $\mathbf{Encoder}_i$ are the sequential encoders for user sequence and item sequence, respectively, where **SLi-Rec** [24] is exploited as the main encoder in our experiments. Note that the encoder can also be replaced by other sequential models, such as GRU4REC [7] and SASRec [10].

3.2 Dual Representation Contrastive Layer

In this section, we tend to perform a contrastive learning layer from the representation aspect. After obtaining the sequential embeddings $\boldsymbol{h}^u_{i_{t+1}}$ and $\boldsymbol{h}^i_{u_{t+1}}$ under user sequence $\mathcal{U}_{i_{t+1}} = (u_1, u_2, \ldots, u_t)$ and item sequence $\mathcal{I}_{u_{t+1}} = (i_1, i_2, \ldots, i_t)$, we treat them as the representations for target item i_{t+1} and target user u_{t+1}, respectively. Then we further perform contrastive learning on them with the target item embedding $\mathbf{M}^i_{i_{t+1}}$ and target user embedding $\mathbf{M}^u_{u_{t+1}}$, respectively, as:

$$\mathcal{L}^e = \lambda_e \left(\left(\boldsymbol{h}^u_{i_{t+1}} - \mathbf{M}^i_{i_{t+1}} \right)^2 + \left(\boldsymbol{h}^i_{u_{t+1}} - \mathbf{M}^u_{u_{t+1}} \right)^2 \right), \tag{3}$$

where λ_e is the L2 regularization parameter for the representation contrastive learning.

3.3 Mixed Prediction Layer

With the outputs from the user encoder and item encoder, we concatenate them together and feed them into the MLP-based prediction layer [27, 28], which can be formulated as follows,

$$P(u_{t+1}|u_1, u_2, \ldots, u_t) = \mathbf{MLP}\left(\boldsymbol{h}^u_{i_{t+1}} \| \mathbf{M}^u_{u_{t+1}} \right), \tag{4}$$

$$P(i_{t+1}|i_1, i_2, \ldots, i_t) = \mathbf{MLP}\left(\boldsymbol{h}^i_{u_{t+1}} \| \mathbf{M}^i_{i_{t+1}} \right), \tag{5}$$

$$P(u_{t+1}, i_{t+1}) = \mathbf{MLP}\left(\boldsymbol{h}^u_{i_{t+1}} \| \boldsymbol{h}^i_{u_{t+1}} \right), \tag{6}$$

where $P(u_{t+1}|u_1, u_2, \ldots, u_t)$, $P(i_{t+1}|i_1, i_2, \ldots, i_t)$ and $P(u_{t+1}, i_{t+1})$ are the predicted scores (probabilities) of next user prediction, next item prediction and static interest.

3.4 Dual Interest Contrastive Layer

In this section, we tend to perform contrastive layer at interest aspect. After the mixed prediction layer, we obtain the predicted scores of next user prediction, next item prediction and static interest. We then tend to further mix them together for the same optimization objective.

Table 1: Data statistics for Taobao dataset and Amazon Toys dataset filtered by 10-core setting.

Dataset	#Users	#Items	#Records	Avg. records per user	Avg. records per item
Taobao	37,271	64764	1,497,029	40.17	23.12
Amazon Toys	6,919	28,695	120,334	17.39	4.19

3.4.1 ***Dual Interest Contrastive Learning.*** Apart from the representation contrastive learning, we further propose dual interest contrastive learning on the dynamic interest for next item prediction with the static interest as follows.

$$\begin{aligned} \mathcal{L}^p = &\lambda_p ((P(i_{t+1}|i_1, i_2, \ldots, i_t) - P(u_{t+1}, i_{t+1}))^2 + \\ &(P(i_{t+1}|i_1, i_2, \ldots, i_t) - P(u_{t+1}|u_1, u_2, \ldots, u_t))^2) \end{aligned} \tag{7}$$

where λ_p is the L2 regularization parameter for the interest contrastive learning. Here next user score is also self-supervised.

3.4.2 ***Model Optimization.*** Following the popular next item prediction, we then utilize the widely-used *LogLoss* function [27, 28] as the main loss function, formulated as follows,

$$\begin{aligned} \mathcal{L}^i = -\frac{1}{|\mathcal{R}|} \sum_{(u_{t+1}, i_{t+1}) \in \mathcal{R}} &(y_{u_{t+1}, i_{t+1}} \log P(i_{t+1}|i_1, i_2, \ldots, i_t) + \\ &(1 - y_{u_{t+1}, i_{t+1}}) \log (1 - P(i_{t+1}|i_1, i_2, \ldots, i_t))), \end{aligned} \tag{8}$$

where $\mathcal{R}$ is the training set. Here $y_{u_{t+1}, i_{t+1}} = 1$ and $y_{u_{t+1}, i_{t+1}} = 0$ denote the positive/negative training samples, respectively.

To jointly optimize the next item prediction with the next user prediction and static interest, our final loss function combine $\mathcal{L}^i$ with $\mathcal{L}^e$ and $\mathcal{L}^p$ in Eqn.(3) and Eqn.(7). It can be formulated as follows,

$$\mathcal{L} = \mathcal{L}^i + \mathcal{L}^e + \mathcal{L}^p + \lambda \|\Theta\|_2, \tag{9}$$

where Θ denote learnable parameters with hyper-parameter λ for the regularization penalty.

Discussion. Our early attempt has discovered that predicting the next user based on the historical user sequence of a given target item is poorer than next item prediction, which may result from the bias of collected sequential behaviors recommended by next item prediction models. Therefore, mere user sequence is not suitable to achieve sequential recommendation tasks, which may be the reason why existing works do not attempt the next user prediction.

4 EXPERIMENTS

4.1 Experimental Setup

4.1.1 Datasets. We conduct evaluations of recommendation performance on two benchmark datasets. The statistics of the filtered datasets with 10-core setting are shown in Table 1.

- **Taobao**[1]. This dataset is collected from the largest e-commerce platform in China. We preserve the click data from November 25 to December 3, 2017. We split the first 7 days as the training set, the 8-th day as the validation set, and the last day as the test set, following the existing work [2].
- **Amazon Toys**[2]. This high sparse dataset is collected from the largest e-commerce platform around the world. We split the

[1]https://tianchi.aliyun.com/dataset/dataDetail?dataId=649
[2]https://www.amazon.com

Table 2: Performance comparisons for DCN on Taobao and Amazon Toys.

Dataset	Taobao				Amazon Toys			
Metric	**AUC**	**MRR**	**NDCG**	**WAUC**	**AUC**	**MRR**	**NDCG**	**WAUC**
DIN	0.7556	0.6841	0.7608	0.8511	0.6483	0.4625	0.5887	0.6813
Caser	0.8204	0.6690	0.7491	0.8415	0.7113	0.4840	0.6063	0.7151
GRU4REC	0.8449	0.6939	0.7680	0.8540	0.7304	0.4926	0.6132	0.7250
DIEN	0.7884	0.6816	0.7590	0.8519	0.6761	0.4676	0.5932	0.6956
SASRec	0.8031	0.6427	0.7292	0.8303	0.6890	0.4879	0.6091	0.7133
SLi-Rec	0.8591	0.7073	0.7784	0.8624	0.7356	0.4965	0.6161	0.7266
DCN	**0.8627**	**0.7137**	**0.7832**	**0.8658**	**0.7442**	**0.5112**	**0.6271**	**0.7311**

behaviors before the last year as the training set, the first half of the last year as the validation set, and the second half of the last year as the test set.

4.1.2 Baseline Models and Metrics.

- **DIN** [28]: DIN represents user by aggregating his/her weighted historical item embeddings based on attention weights.
- **Caser** [17]: Caser performs convolution on user's item sequence embeddings to extract the dynamic interest of given user.
- **GRU4REC** [7]: GRU4REC proposes GRU [3] to predict the next interacted items in given session by the final state of unit output.
- **DIEN** [27]: DIEN proposes two GRU layers for interest extraction and evolution, respectively, to encode the item sequence.
- **SASRec** [10]: SASRec encodes the item sequence via the hierarchical self-attention network.
- **SLi-Rec** [24]: SLi-Rec proposes an attention network to model the long-term interest and time-aware LSTM to model the short-term interest.

We use AUC and GAUC (two widely-used accuracy metrics [6]), and MRR and NDCG@10 (two widely-used ranking metrics [2]) for evaluation.

4.1.3 Hyper-parameter Settings. All models are optimized by Adam with a learning rate 0.001 [11], whose parameters are initialized by Xavier initialization [5]. We search the regularization parameters within $[1e^{-7}, 1e^{-5}, 1e^{-3}]$. Both datasets are fed with batch size 200. The embedding sizes are fixed as 32 for all models. The prediction layers of all models are two-layer MLPs with layer sizes 100 and 64. The maximum sequence length of the Taobao is 50, while that of the Amazon Toys is 20.

4.2 Overall Performance

We show the performance comparisons in Table 2, from which we have the following observations.

- **Our DCN achieves the best performance**. Our model achieves the best performance compared with these six baselines for four metrics. Specifically, MRR is improved by 0.9% on the Taobao and by 2.96% on the Amazon Toys when comparing DCN with other baselines. The improvement is more obvious on the Amazon Toys with sparser data, which means our approach can well tackle the sparse data issue by integrating user sequence and extracting dynamic item trends for auxiliary training based on self-supervised learning.

Table 3: Ablation study on contrastive learning.

Model		Taobao		Amazon Toys	
		NDCG	**WAUC**	**NDCG**	**WAUC**
Dual Representation	w/o URC	0.7811	0.8650	0.5874	0.6999
	w/o IRC	0.7820	0.8640	0.6206	0.7294
Dual Interest	w/o SIC	0.7806	0.8630	0.6073	0.7188
	w/o UPC	0.7803	0.8645	0.6173	0.7253
Ours	w all	**0.7832**	**0.8658**	**0.6271**	**0.7311**

- **Sequential recommendation suffers from sparse data**. The observation of model performance is consistent between two datasets, i.e., SLi-Rec > GRU4REC > Caser > SASrec > DIEN > DIN, where the former five ones are sequential models, which illustrates that it is necessary for us to extract the dynamic interests of users. Besides, SLi-Rec outperforms other models more sharply under Taobao dataset with a longer sequence, which illustrates the ability of SLi-Rec's long-term modeling to model the long sequence. Most importantly, SLi-Rec also outperforms other models more sharply under Taobao dataset with denser data. Though SLi-Rec is effective with long and short-term modelings, it still encounters the sparse data bottleneck. Therefore, we tend to enhance the sequential learning by incorporating user sequence for auxiliary training instead of generating sparser augment data by dropout strategy based on self-supervised learning.

4.3 Ablation Study

As for the further evaluations, we study the performance of our model without dual-representation and dual-interest contrastive learning, and the results are shown in Table 3. It can be observed that these two contrastive learning approaches are generally more effective on Amazon Toys with sparser data, which demonstrates the advantage of our self-supervised approaches with genuine facts instead of augmentation data generated by dropout. Firstly, we compare the model without User Representation Contrastive Learning (URC) and without Item Representation Contrastive Learning (IRC) in Dual Representation Contrastive Learning. URC is more effective than IRC on Amazon Toys while comparable with it on Taobao, which means user representation learning is a decent complement for sequential recommendation with item sequence only. Besides, we compare the model without Static Interest Contrastive Learning (SIC) and without Next User Prediction Contrastive Learning

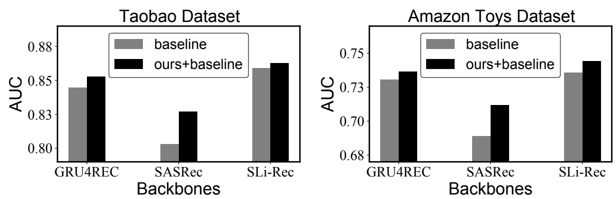

Figure 3: Performance comparisons of our proposed approaches with different backbones as encoders.

(UPC) in Dual Interest Contrastive Learning, from which we can observe that SIC is more effective than UPC on Amazon Toys while is comparable with it under Taobao. This may be because the item sequence has already captured the sequential pattern to some extent, and static preference is more complementary for sequential recommendation with item sequence only.

4.4 Study of Encoder Backbone

In this section, we tend to verify the generalization of our proposed approaches to other models such as GRU4REC and SASRec. The results of our approaches exploiting different backbones as encoders are as shown in Figure 3, where we can observe that (1) our self-supervised approaches can boost the performance of these three backbones; (2) the improvement is most obvious under SASRec with vast parameters and prone to be overfitting under sparse data, which verifies our approaches are more effective for sequential models requiring a lot of data for training; (3) the improvement is minor under SLi-Rec, which may because SLi-Rec's long and short term modelings have provided it with sufficient prior knowledge and thus keep it from data sparsity issue to some extent, as well as long term interest can be treated as static interest [24].

5 RELATED WORK

Sequential Recommendation. One related work of our solution is Sequential Recommendation [19], which aims to predict the probability of the next interacted item based on the historical item sequence of the given target user. Markov chain is the most initial sequential model applied in the sequential recommendation, namely FPMC [15]. Subsequently, some sprouted deep learning models such as recurrent neural network [3, 9], convolution neural network [13] and attention network [18] are then applied in recommendation [7, 10, 17, 27, 28] to achieve better generalization. To further capture the short-term and long-term interest simultaneously, matrix factorization [12] is also adopted to achieve fancy performance [24, 25]. RIMs [14] connects personalized item recommendation and marketing user recommendation by dual decomposition, to trade-off between recommendation relevance and item-diversity. However, existing works of sequential recommendation only consider the item sequence side without further considering the user sequence side. Indeed, item sequence reflects the dynamic interest of given user preferring items while user sequence reflects the dynamic trend of given item preferred by users.

Self-supervised Learning in Sequential Recommendation. Recently, researchers have also integrated self-supervised learning (SSL) [8] into the sequential recommendation. Indeed, self-supervised learning has already achieved decent performance in the fields of natural language processing [4] and computer vision [1], based on the random dropout upon raw data information such as masking the sentence or rotating/clipping the image. In the field of recommendation, more similar to natural language processing, Bert4Rec [16] predicts the random masked items in the sequence by dropout strategy. Although such self-supervised methods have achieved decent performance, they are not suitable for the sequential recommendation because the random dropout strategy would generate sparser data and obscure self-supervision signals for sequential learning. Hence S^2-DHCN [22] conducts contrastive learning between representations of different hyper-graphs without random dropout, but its fixed ground truths restrict the improvements. CSLR [26] generates self-supervised signal via long-term and short-term interest, but failing to further consider more effective signals. Besides, two works about social recommendation [23] and session-based recommendation [22] address this problem by the self-supervised co-training framework.

Different from them, our work takes the first step to reflect sequential recommendation from both user-sequence and item-sequence perspectives, based on the self-supervised learning.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we approached sequential recommendation from a new perspective that integrated user sequence with self-supervised learning instead of solely item sequential learning. Such exploration not only supplemented the existing sequential recommendation with user sequence but also addressed the challenge of current self-supervised learning with sparse augmentation data by dropout strategy. Specifically, we proposed DCN (**D**ual **C**ontrastive **N**etwork for Sequential Recommendation) with two kinds of contrastive learning. The first one was the dual representation contrastive learning that contained user representation contrastive learning and item representation contrastive learning. This self-supervised learning approach tended to refine the target user and item embeddings via the encoded sequential representations for the target user and target item, respectively. The second one was the dual interest contrastive learning between the dynamic interest for next item prediction and the static interest, which aimed to incorporate the static modeling into the item sequence modeling via auxiliary training. Besides, the predicted score of the next user was also self-supervised here with the next item prediction.

As for future work, we plan to improve this paper by 1) conducting more experiments on more datasets and metrics, 2) exploring more user-sequence-aware self-supervision signals, and 3) applying the model to real-world online scenarios to further evaluate its performance.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China under U1936217, 61971267, 61972223, and the Science and Technology Innovation Committee of Shenzhen under CJGJZD20200617102801005.

REFERENCES

[1] Philip Bachman, R Devon Hjelm, and William Buchwalter. 2019. Learning representations by maximizing mutual information across views. *Advances in neural information processing systems* 32 (2019).

[2] Jianxin Chang, Chen Gao, Yu Zheng, Yiqun Hui, Yanan Niu, Yang Song, Depeng Jin, and Yong Li. 2021. Sequential Recommendation with Graph Neural Networks. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 378–387.

[3] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[5] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*. 249–256.

[6] Asela Gunawardana and Guy Shani. 2015. Evaluating Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, 265–308. https://doi.org/10.1007/978-1-4899-7637-6_8

[7] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based recommendations with recurrent neural networks. In *ICLR*.

[8] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. 2018. Learning deep representations by mutual information estimation and maximization. *arXiv preprint arXiv:1808.06670* (2018).

[9] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[10] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 197–206.

[11] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

[12] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009).

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. 25 (2012), 1097–1105.

[14] Yifei Ma, Ge Liu, and Anoop Deoras. 2021. Bridging Recommendation and Marketing via Recurrent Intensity Modeling. In *International Conference on Learning Representations*.

[15] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized markov chains for next-basket recommendation. In *WWW*. 811–820.

[16] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1441–1450.

[17] Jiaxi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *WWW*. 565–573.

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NeurIPS*. 5998–6008.

[19] Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z. Sheng, and Mehmet Orgun. 2019. Sequential recommender systems: challenges, progress and prospects. (2019).

[20] Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z. Sheng, and Mehmet Orgun. 2019. Sequential recommender systems: challenges, progress and prospects. (2019).

[21] Zhenlei Wang, Jingsen Zhang, Hongteng Xu, Xu Chen, Yongfeng Zhang, Wayne Xin Zhao, and Ji-Rong Wen. 2021. Counterfactual data-augmented sequential recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 347–356.

[22] Xin Xia, Hongzhi Yin, Junliang Yu, Yingxia Shao, and Lizhen Cui. 2021. Self-Supervised Graph Co-Training for Session-based Recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2180–2190.

[23] Junliang Yu, Hongzhi Yin, Min Gao, Xin Xia, Xiangliang Zhang, and Nguyen Quoc Viet Hung. 2021. Socially-aware self-supervised tri-training for recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2084–2092.

[24] Zeping Yu, Jianxun Lian, Ahmad Mahmoody, Gongshen Liu, and Xing Xie. 2019. Adaptive User Modeling with Long and Short-Term Preferences for Personalized Recommendation. In *IJCAI*. 4213–4219.

[25] Wei Zhao, Benyou Wang, Jianbo Ye, Yongqiang Gao, Min Yang, and Xiaojun Chen. 2018. PLASTIC: Prioritize Long and Short-term Information in Top-n Recommendation using Adversarial Training. In *IJCAI*. 3676–3682.

[26] Yu Zheng, Chen Gao, Jianxin Chang, Yanan Niu, Yang Song, Depeng Jin, and Yong Li. 2022. Disentangling Long and Short-Term Interests for Recommendation. In *TheWebConf*.

[27] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep interest evolution network for click-through rate prediction. In *AAAI*. 5941–5948.

[28] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *KDD*. 1059–1068.

[29] Kun Zhou, Hui Wang, Wayne Xin Zhao, Yutao Zhu, Sirui Wang, Fuzheng Zhang, Zhongyuan Wang, and Ji-Rong Wen. 2020. S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 1893–1902.