Self Contrastive Learning
for Session-Based Recommendation

Zhengxiang Shi[(✉)], Xi Wang, and Aldo Lipani

University College London, London, UK
`{zhengxiang.shi.19,xi-wang,aldo.lipani}@ucl.ac.uk`

Abstract. Session-based recommendation, which aims to predict the next item of users' interest as per an existing sequence interaction of items, has attracted growing applications of Contrastive Learning (CL) with improved user and item representations. However, these contrastive objectives: (1) serve a similar role as the cross-entropy loss while ignoring the item representation space optimisation; and (2) commonly require complicated modelling, including complex positive/negative sample constructions and extra data augmentation. In this work, we introduce Self-Contrastive Learning (SCL), which simplifies the application of CL and enhances the performance of state-of-the-art CL-based recommendation techniques. Specifically, SCL is formulated as an objective function that directly promotes a uniform distribution among item representations and efficiently replaces all the existing contrastive objective components of state-of-the-art models. Unlike previous works, SCL eliminates the need for any positive/negative sample construction or data augmentation, leading to enhanced interpretability of the item representation space and facilitating its extensibility to existing recommender systems. Through experiments on three benchmarks, we demonstrate that SCL consistently improves the performance of state-of-the-art models with statistical significance. Notably, our experiments show that SCL improves the performance of two best-performing models by 8.2% and 9.5% in P@10 (Precision) and 9.9% and 11.2% in MRR@10 (Mean Reciprocal Rank) on average across different benchmarks. Additionally, our analysis elucidates the improvement in terms of alignment and uniformity of representations, as well as the effectiveness of SCL with a low computational cost. Code is available at https://github.com/ZhengxiangShi/SelfContrastiveLearningRecSys.

Keywords: Recommendation System · Session-based Recommendation · Contrastive Learning

1 Introduction

Session-based recommendation [13,14,37,38] is a crucial aspect of modern recommender systems for various platforms such as e-commerce websites [9,12], music streaming services [3], and social media [29], with the goal of predicting a user's next interest by focusing on their current intent. Recently, Contrastive Learning (CL) [18] has been applied in session-based recommendation tasks to
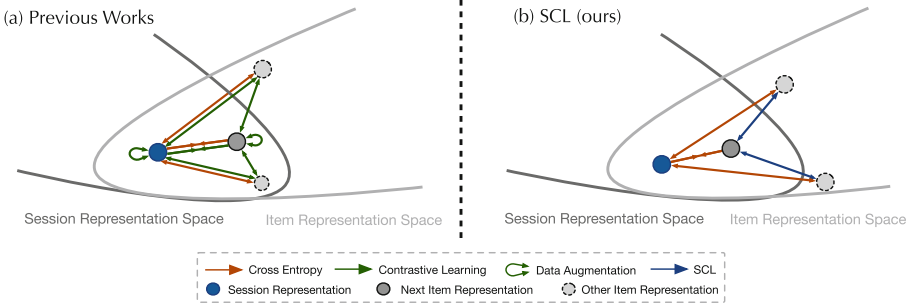
Fig. 1. The illustration of the framework of SCL. In previous works, contrastive learning (CL) objectives (depicted in green) typically involve complicated modelling, leading to a relatively lesser emphasis on optimising the item representation space. In contrast, SCL (depicted in blue) specifically addresses this issue and provides a better complement to the role of cross-entropy loss. (Color figure online)

enhance recommendation accuracy via improved representation quality, with the goal of aligning the session representation with the next item's representation, while also distinguishing it from other item representations. However, two key limitations exist within these methods.

Firstly, **the importance of optimising the item representation space *itself* by ensuring that the representations are uniformly distributed is not receiving adequate attention**. The CL objectives in previous research [17,19,35,37–39,47] serve a similar role as the cross-entropy loss while the optimisation of item representation space is not adequately addressed. As shown in Fig. 1(a), both cross-entropy loss and CL loss have the capacity to align the session representation with the representation of the next item and differentiate it from other item representations (§4). This has led to the marginalization of the role played by the uniformity of item representations. While some studies [17,19,35,39,47] have touched upon improving the uniformity of representations, these efforts generally contribute only a small fraction to the overall loss function.

Secondly, **current CL-based approaches often utilise complex techniques, including the sophisticated creation of positive and negative pairs and extra data augmentations, leading to limited adaptability across models**. Specifically, two state-of-the-art (SOTA) session-based recommendation models, S^2-DHCN [38] and COTREC [37] are the typical examples of complex CL-based applications. S^2-DHCN encompasses two encoder networks that generate varied session representations (positive) and compare them to corrupted session representations (negative) for noisy data-augmented CL. Similarly, COTREC requires two item representations to interact with the corresponding session representation in the CL objective, obtained through model-specific data augmentation techniques. These methods are heavily dependent on the model architecture and may not be compatible with various other models. Moreover, while recent studies have highlighted the importance of uniformity

in user/item representations for recommendation tasks, this has simultaneously triggered a rise in the use of extra data augmentation methods, such as applying noise perturbation [47] or dropout [39,50] to augment representations, as shown in Fig. 1(a).

In this work, we argue that the importance of the uniformity of item representations has been considerably undervalued and that intricate CL objectives could be streamlined. We propose a novel approach, *Self Contrastive Learning* (SCL), which directly enforces the representation of each item distinct from those of all other items through a new loss objective and thus promotes a uniform distribution within the item representation space. SCL can be easily integrated into SOTA models to effectively replace other CL objectives, eliminating the need for creating complex positive/negative samples or engaging in any form of data augmentation. Different from previous approaches in recommendation systems that utilise the CL [19,31,35,40,47], SCL represents the first attempt to simply enforce uniformity of item representation without resorting to other CL objectives. Through our research, we aim to address the following research questions:

RQ$_1$ To what extent does SCL improve the session-based recommendation tasks? (§5.2)

RQ$_2$ How does SCL improve the model performance in terms of the alignment and uniformity of representations? (§5.3)

RQ$_3$ Are those sophisticated CL objectives still necessary in the presence of SCL? (§5.4)

RQ$_4$ Can SCL maintain SOTA performance with a low computational cost? (§5.5)

To address **RQ$_1$**, we experiment on three datasets, TMALL, NOWPLAYING, and DIGINETICA (§5.2). Our results show that SCL consistently improves the performance of SOTA models across various evaluation metrics and datasets. In particular, our experiments on TMALL show that SCL improves the performance of S^2-DHCN from 28.65% to 35.14% in P@10 and from 15.94% to 20.39% in MRR@10, and it also boosts the performance of COTREC from 30.44% to 35.03% in P@10 and from 17.28% to 20.46% in MRR@10, outperforming all existing approaches by large margins.

To understand how the model is improved (**RQ$_2$**), we investigate the transformations of the session and item representation in terms of alignment and uniformity (§5.3). Our study reveals that SCL learns item representations with a lower uniformity loss, leading to significant improvements in performance. Our findings suggest that SOTA approaches may have placed excessive emphasis on the alignment of session and item representations.

To answer **RQ$_3$**, we carry out an ablation study to evaluate the necessity of sophisticated CL objectives employed in prior works (§5.4). Our experiment reveals that SCL is capable of attaining the comparable model performance on its own, suggesting the advance of SCL and the redundant use of existing heavy and sophisticated CL objectives.

To understand the computational efficiency (**RQ$_4$**), we further study the impact of selecting the k-nearest item representations in SCL on the model performance (§5.5). Our results show that SCL generally benefits from contrasting to more item representations. However, it can still achieve SOTA performance even just using a value of k equal to 2, indicating that SCL can be implemented with a low computational cost.

2   Related Work

Session-Based Recommendation. The session-based recommendation aims to predict the next item by utilizing user behaviours within a short time period [10,14,33,36]. Early studies on session-based recommendation focused on utilising temporal information from session data through the use of Markov chain models [5,13,21,22,42,44,51]. Then neural networks are widely applied for session-based recommendation [6,10,14,33,36]. Recurrent neural networks [11] have been applied to session-based recommendation models to capture the sequential order between items [48]. GRU4Rec [10] was the first model to use gated recurrent units (GRUs) [4] to model the sequential relations of item interactions. NARM [13] extended GRU4Rec by incorporating the attention mechanism [2,23,25,26] STAMP [14] also replaced the recurrent structure with attention layers to capture a user's general and current interests. FGNN [20] rethinks the sequence order of items to exploit users' intrinsic intents using GNNs. GCE-GNN [34] aggregates item information from both the item-level and session-level through graph convolution and self-attention mechanism. S^2-DHCN [38] utilizes hyper-graph convolutional networks to capture high-order item relations within individual sessions. COTREC [37] integrated self-supervised learning into the graph training through sophisticated positive and negative constructions.

Contrastive Learning. CL has achieved great success in various research domains, such as computer vision [8] and natural language processing [6,7,24,27,28], with the goal of obtaining high-quality representations by pulling positive or similar instances closer in the representation space while simultaneously separating dissimilar, or negative instances. Recently, CL has recently been applied to sequential recommendation tasks, with several studies exploring its potential benefits in this area [43,45,46]. Bert4Rec [30] adapts the cloze objective from language modelling to a sequential recommendation by predicting random masked items in the sequence with surrounding contexts. S^3-Rec [49] utilizes intrinsic data correlations among attributes, items, subsequences, and sequences to generate contrastive signals and enhance data representations through pre-training. [16] proposed a seq2seq training strategy based on latent self-supervision and disentanglement of user intention behind behaviour sequences. CoSeRec [15] uses GNNs to capture more complex patterns than sequential patterns via CL objectives. CL4SRec [41] combines recommendation loss with CL loss of self-supervised tasks. DuoRec [19] retrieves the positive view of a given user sequence by identifying another user's sequence that shares the same next item through

its proposed supervised CL. Unlike DuoRec, our task does not have access to user information. CL has also been applied to other recommendation paradigms, such as general recommendation [43] and social recommendation [45,46]. Previous work [1] also proposes a framework as self contrastive learning approach where every sample sharing a ground-truth label with an anchor is treated as a positive pair. In contrast, our approach is different because we define the positive sample exclusively as the sample itself, not considering other samples with the same label.

3   Preliminaries

Task Definition. In the session-based recommendation, the full set of item candidates is represented as $I = \{i_1, \ldots, i_n\}$, where n is the total number of item candidates. A session s, consisting of m items, is represented as a sequence $S = [i_1^s, \ldots, i_m^s]$ ordered by timestamps, where $i_k^s \in I$ $(1 \leq k \leq m)$ represents the k-th item that has been interacted with by a user. The objective is to predict the next item, i_{m+1}^s, from a full set of item candidates I, based on the corresponding session sequence S. For a session s, the output of the session-based recommendation model is a ranked list of item candidates $R = [r_1^s, \ldots, r_n^s]$, where r_*^s is the corresponding predicted ranking or preference score of the i-th item. Afterwards, the top-k items $(1 \leq k \leq n)$ will be selected as recommendations.

Contrastive Learning. Contrastive learning aims to pull the representation of an anchor sample and the representations of its corresponding positive sample pairs closer while simultaneously pushing the representations of the negative sample pairs away [7]. INFONCE [18], where NCE stands for Noise Contrastive Estimation, is a type of contrastive loss function commonly used in recommendation systems [19,38,40,41]. Formally, let $\boldsymbol{a}$ denote an anchor representation and $\mathcal{X} \triangleq \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}, \boldsymbol{x}_n\}$ denote the set of negative representations $(1 \leq k \leq n-1)$ and one positive representation $(k = n)$ with respect to $\boldsymbol{a}$, the INFONCE loss is defined as:

$$\mathcal{L}_{\text{INFONCE}} = -\log \frac{f(\boldsymbol{a}, \boldsymbol{x}_n)}{\sum_{j=1}^{n} f(\boldsymbol{a}, \boldsymbol{x}_j)}, \tag{1}$$

where f can be approximated by a real-valued scoring function and typically a function of the cosine similarity is used.

In the field of CL, two key properties, known as alignment and uniformity, have been proposed by [32] as measures of the quality of representations. The uniformity of the embedding distribution is measured as follows:

$$\ell_{\text{uniform}} \triangleq \log \mathop{\mathbb{E}}_{\substack{x \sim p_{\text{data}}, \\ x' \sim p_{\text{data}}}} e^{-2\|f(x)-f(x')\|^2}, \tag{2}$$

where p_{data} denotes the data distribution. ℓ_{uniform} is lower when random samples are farther from each other. Therefore, the examination of item representation

uniformity ensures their semantic interpretability for a potential improvement in identifying the items of true interest. In contrast, instead of assessing the dispersion of item representations for uniformity, alignment gauges the expected distance between the embeddings of positively paired instances, assuming that representations are normalized, as expressed by the following equation:

$$\ell_{\text{align}} \triangleq \mathop{\mathbb{E}}_{\substack{x \sim p_{\text{data}}, \\ x' \sim p_{\text{pos}}(x)}} \|f(x) - f(x')\|^2, \tag{3}$$

where $p_{\text{pos}}(x)$ denotes the data distribution of samples that are positive to the instance x. ℓ_{align} is lower as all positive samples are closer to each other.

4 Methodology

In this section, we first discuss the potential limitations of SOTA session-based recommendation systems (§4). To address these issues, we then introduce a novel approach, *Self Contrastive Learning* (SCL), which aims to improve the uniformity in item representations by utilising a novel loss function (§4).

Motivation. In the field of session-based recommendation, existing works [34, 37,38], that utilise CL objectives, generally employ a framework in which the loss function is a combination of cross-entropy ($\mathcal{L}_{ce}$) and CL $\mathcal{L}_{cl}$ losses:

$$\mathcal{L} = \mathcal{L}_{ce} + \alpha\mathcal{L}_{cl}, \tag{4}$$

where $\mathcal{L}_{ce}$ aims to maximize the likelihood of selecting the correct next item, and $\mathcal{L}_{cl}$ aims to improve the learned representations, with the scalar coefficient α controlling the relative importance of these two objectives. Typically, the INFONCE loss, as in Eq. 1, is used as $\mathcal{L}_{cl}$. However, these two objectives are similar in nature, as shown in Fig. 1. Specifically, let s denote a learned session representation and $L = \{(\boldsymbol{x}_k, y_k)\}_{k=1}^n$ denote a set of n learned item representations and their corresponding ground-truth labels, where y_k is 1 if the k-th item is the user's next click item and 0 otherwise. The categorical cross-entropy of classifying the next item correctly is computed as follows:

$$\mathcal{L}_{ce} = -\sum_{\boldsymbol{x}, y \in L} y \log p(\boldsymbol{x}|\boldsymbol{s}), \tag{5}$$

where p measures the probability that the item represented by $\boldsymbol{x}$ is drawn from the full set of item candidates conditioned on the session representation $\boldsymbol{s}$. The probability measure p is typically normalized using a real-valued scoring function f (e.g., cosine similarity). Thus, we can rewrite the Eq. 5 as:

$$\mathcal{L}_{ce} = -\sum_{\boldsymbol{x}, y \in L} y \log \frac{f(\boldsymbol{s}, \boldsymbol{x})}{\sum_{j=1}^n f(\boldsymbol{s}, \boldsymbol{x}_j)} = -\log \frac{f(\boldsymbol{s}, \boldsymbol{x}^+)}{\sum_{j=1}^n f(\boldsymbol{s}, \boldsymbol{x}_j)}, \tag{6}$$

where $\boldsymbol{x}^+$ is the user's next clicked item. Therefore, $\mathcal{L}_{ce}$ can be considered as an alternative expression of $\mathcal{L}_{cl}$ ($\mathcal{L}_{\text{INFONCE}}$) when they use the same function f.

It is important to note that, while the loss functions $\mathcal{L}_{ce}$ and $\mathcal{L}_{cl}$ in Eq. 4 may have marginal variations, (*e.g.,* $\mathcal{L}_{cl}$ may use the extra temperature parameter τ in the function f in Eq. 6 and different positive and negative samples from data augmentation), their directions of optimising the representation spaces are the same: both objectives aim to push the session representation $\boldsymbol{s}$ closer to the next item representation $\boldsymbol{x}^+$ while pulling it away from other representations $\boldsymbol{x}_j$, thus improving session and item representations. Although using INFONCE as CL objectives in conjunction with cross-entropy loss may result in a marginal improvement in performance, we argue that it is not the most effective strategy. Firstly, this may place an overemphasis on the alignment of the session and item representations, as shown in Fig. 1(a) where the green lines and red lines both have an impact on alignment. Secondly, while prior work [39] attempted to improve the uniformity of the item representation space with some auxiliary losses, the importance of blue lines (see Fig. 1(b)) appears to be diluted by other CL objectives. A more straightforward regularization approach that specifically targets the item representation distributions and effectively complements the cross-entropy loss is necessary to improve the overall recommendation performance.

Self Contrastive Learning (SCL***).*** To address the aforementioned issues, we propose Self Contrastive Learning (SCL), a straightforward solution to improve the uniformity of the item representation space by introducing an additional loss objective, as shown in Fig. 1(b). This objective operates by directly penalizing the proximity of item representations based on our assumption that the representation of each item representation should be distant from those of all other items. Formally, given a set of n learned item representations $\mathcal{X}$, the objective of the SCL loss is calculated as follows:

$$\mathcal{L}_{\mathrm{SCL}} = -\sum_{i=1}^{n} \log \frac{g(\boldsymbol{x}_i, \boldsymbol{x}_i)}{\sum_{j=1}^{n} g(\boldsymbol{x}_i, \boldsymbol{x}_j)}, \tag{7}$$

where the function $g(\boldsymbol{x}, \boldsymbol{x}')$ is computed by $e^{\mathrm{sim}(\boldsymbol{x}, \boldsymbol{x}')/\tau}$, the exponential of the cosine similarity controlled by a temperature parameter τ. Using the cosine similarity, this loss pulls apart all items on the unit hypersphere. Next, we integrate $\mathcal{L}_{\mathrm{SCL}}$ into the existing session-based recommendation models. Given the loss objective $\mathcal{L}_{model}$ from the original model (**all other CL objectives are excluded**), the overall loss function is computed as follows:

$$\mathcal{L} = \mathcal{L}_{model} + \beta \mathcal{L}_{\mathrm{SCL}}, \tag{8}$$

where β is a hyperparameter that determines the relative importance of the two objectives. Complementary to the $\mathcal{L}_{model}$, which typical uses a $\mathcal{L}_{ce}$ to positively impact both ℓ_{align} and $\ell_{uniform}$, $\mathcal{L}_{\mathrm{SCL}}$ has a stronger positive effect on $\ell_{uniform}$.

The advantages of SCL can be summarized in three main aspects: (1) **Improved representation spaces.** By incorporating SCL as an additional loss objective, we achieve improved uniformity in the item representation space, leading to better model performance; (2) **Simplified modelling process.** By

leveraging the SCL objective, we avoid the need for complex creation of positive/negative sample pairs and data augmentation techniques, such as noise perturbation [47] or dropout [39]. In SCL, each item representation serves as the sole positive sample, and all other item representations are considered negative samples without further modifications. This greatly simplifies the construction of recommendation systems, making them more efficient and easier to implement; and (3) **Seamless integration into existing systems.** SCL can be seamlessly integrated into existing session-based recommendation systems that utilise session and item representations, without any additional modification to the architecture of the model. This high level of compatibility makes SCL widely applicable and adaptable to various settings and scenarios. Overall, these advantages make SCL a valuable solution for enhancing recommendation systems, offering improved uniformity, simplified training, and easy integration into existing models.

5   Experiments and Results

In this section, we evaluate our SCL in three benchmarks. We first describe the experimental setup, including the used datasets, baselines, evaluation metrics, and implementation details. Then we present our experimental results with respect to the four research questions introduced in Section §1.

5.1   Experimental Setup

Datasets and Baselines. **Tmall** is sourced from the IJCAI-15 competition and includes anonymized shopping logs from users on the Tmall online shopping platform, with train size 351,268, test size 25,898, items Size 40,728 and an average length 6.69. **Nowplaying** describes the music-listening behaviour of users, with train size 825,304, test size 89,824, items Size 60,417 and an average length 7.42; and **Diginetica** , from CIKM Cup 2016, comprises typical transaction data, with train size 719,470, test size 60,858, items Size 43,097 and an average length 5.12. Our proposed SCL method is compared with the following representative methods: **FPMC** [21], **GRU4REC** [10], **NARM** [13], **STAMP** [14], **SR-GNN** [36], **GCE-GNN** [34], S^2-**DHCN** [38], and **COTREC** [37].

Implementation Details. We evaluate the performance of our proposed SCL method using the metrics of P@k (Precision) and MRR@k (Mean Reciprocal Rank), where the cutoff k is set to 5, 10 or 20. We conduct experiments with the proposed SCL method using three SOTA models, GCE-GNN, S^2-DHCN, and COTREC. We first reproduce the experimental results of these models by following the settings in their original papers. Then, we apply the SCL to these three models. For the hyperparameters used in the SCL, the temperature parameter, denoted by τ, is set to 0.1, and the loss weight parameter, denoted by β, is varied within a range of 0.1 to 100. We have omitted the evaluation of COTREC on the Nowplaying as we were unable to replicate the results.

Table 1. Performances of all comparison methods on the dev set (RQ_1). † from the original paper. ‡ from our own reproductions. Triangles in colours indicate an improvement over our reproduced results. The highest results in each column are highlighted in bold font. SOTA performances are indicated in blue. *** indicates a p-value $< 1e-20$ for the improvement, ** indicates a p-value $< 1e-5$, and * indicates a p-value $< 1e-2$.

Method	Tmall				Nowplaying				Diginetica			
	P@10	MRR@10	P@20	MRR@20	P@10	MRR@10	P@20	MRR@20	P@10	MRR@10	P@20	MRR@20
FPMC	13.10	7.12	16.06	7.32	5.28	2.68	7.36	2.82	15.43	6.20	26.53	6.95
GRU4REC	9.47	5.78	10.93	5.89	6.74	4.40	7.92	4.48	17.93	7.33	29.45	8.33
NARM	19.17	10.42	23.30	10.70	13.6	6.62	18.59	6.93	35.44	15.13	49.70	16.17
STAMP	22.63	13.12	26.47	13.36	13.22	6.57	17.66	6.88	33.98	14.26	45.64	14.32
SR-GNN	23.41	13.45	27.57	13.72	14.17	7.15	18.87	7.47	36.86	15.52	50.73	17.59
GCE-GNN†	28.01	15.08	33.42	15.42	16.94	8.03	22.37	8.40	41.16	18.15	54.22	19.04
GCE-GNN‡	27.48	14.85	32.52	15.20	17.19	8.09	22.42	8.45	40.98	18.12	54.23	19.04
w/ SCL	28.67**	15.20*	33.65**	15.55*	17.44*	8.10	22.81*	8.47	41.93**	18.45*	54.93*	19.38*
Δ (%)	(▲4.3%)	(▲2.4%)	(▲3.5%)	(▲2.3%)	(▲1.5%)	(▲0.1%)	(▲1.7%)	(▲0.2%)	(▲2.3%)	(▲1.8%)	(▲1.3%)	(▲1.8%)
COTREC†	30.62	17.65	36.35	18.04	-	-	-	-	41.88	18.16	54.18	19.07
COTREC‡	30.44	17.28	36.09	17.67	-	-	-	-	40.26	17.75	53.75	18.69
w/ SCL	35.03***	20.46***	39.29***	20.76***	-	-	-	-	40.78*	18.00*	53.78	18.90*
Δ (%)	(▲15.1%)	(▲18.4%)	(▲8%)	(▲17.5%)	-	-	-	-	(▲1.3%)	(▲1.4%)	(▲0.1%)	(▲1.1%)
S^2-DHCN†	26.22	14.60	31.42	15.05	17.35	7.87	23.50	8.18	39.87	17.53	53.18	18.44
S^2-DHCN‡	28.65	15.94	34.54	16.35	17.23	7.70	23.00	8.10	39.54	17.31	52.76	18.22
w/ SCL	35.14***	20.39***	39.13***	20.67***	17.61*	7.92*	23.74**	8.32*	40.91**	17.79**	53.91*	18.69*
Δ (%)	(▲22.7%)	(▲27.9%)	(▲13.3%)	(▲26.4%)	(▲2.2%)	(▲2.9%)	(▲3.2%)	(▲1.7%)	(▲3.5%)	(▲2.8%)	(▲2.2%)	(▲2.6%)

5.2 Main Results (RQ1)

Table 1 presents the performance of all comparison methods, where our SCL is applied to three SOTA models, GCE-GNN, COTREC, and S^2-DHCN. Results show that SCL consistently improves the model performance in terms of P@k and MRR@k across three datasets, Tmall, Nowplaying, and Diginetica, achieving the new SOTA performance (highlighted in blue). The significance tests further corroborate the effectiveness of SCL. Particularly remarkable is that SCL achieves a notable improvement compared to the SOTA models on the Tmall dataset. COTREC model with the proposed SCL method also shows significant improvement, with an increase of 18.4% and 17.5% in terms of MRR@10 and MRR@20, respectively. Additionally, S^2-DHCN + SCL achieves a new SOTA performance on the Tmall dataset. It records a 27.9% increase of the MRR@10 from 15.94% to 20.39% and a 26.4% increase of MRR@20 from 16.35% to 20.67%. Similar improvement in performance can also be observed on the Nowplaying and Diginetica datasets.

5.3 Alignment and Uniformity (RQ2)

The substantial improvement in performance achieved by the SCL raises the research question of where these improvements come from (RQ_2). Figure 2 depicts the impact of the proposed SCL method on the alignment and uniformity on Tmall and Diginetica datasets. In general, we find that (1) SCL has improved the uniformity of item representations, leading to an improvement in
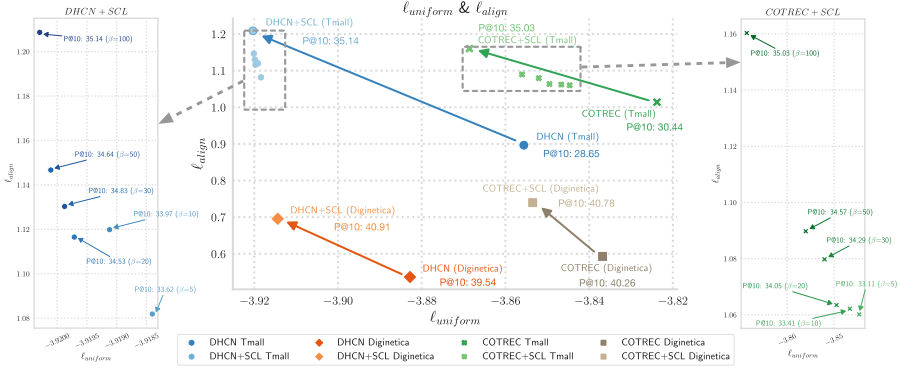
Fig. 2. The analysis of alignment loss ℓ_{align} and uniformity loss ℓ_{uniform}, where P@10 is reported as model performance. While it is generally acknowledged that a decrease in the alignment or uniformity loss leads to improved model performance, an excessive emphasis on alignment and insufficient attention to uniformity can result in sub-optimal performance.

model performance; and (2) a higher loss in alignment ℓ_{align} does not necessarily result in worse performance if the uniformity loss ℓ_{uniform} is improved.

Better Uniformity of Item Representations Brings Substantial Improvement in Performance. The sub-figure in the centre of Fig. 2 illustrates how SCL improves the uniformity of item representations of S^2-DHCN and COTREC on Tmall and Diginetica. This is indicated by a lower uniformity loss when SCL is applied. The uniformity loss measures the dissimilarity between the item representations themselves and a lower uniformity loss indicates that the item representations are becoming more discriminative and less correlated with each other. Specifically, the use of SCL results in a reduction of the uniformity loss of S^2-DHCN from -3.86 to -3.92 on the Tmall dataset, and this improvement is accompanied by an increase in P@10 from 28.65% to 35.14%.

The Trade-Off Between Alignment and Uniformity. We also observe that the proposed SCL method leads to an increase in the alignment loss. This indicates that the next item representations are not only becoming more discriminative to other item representations but also less correlated with the session representations. We conduct additional studies on the Tmall dataset by adjusting the alignment and uniformity loss through controlling the SCL loss weight β. The results are depicted in the two sub-figures of Fig. 2. Specifically, for the S^2-DHCN model, when the uniformity loss gradually decreases from -3.86 to -3.92 and the alignment loss increases from 1.08 to around 1.20, the model performance in P@10 is generally improved from 33.62% to 35.14%. Similar results are observed in the experiments using the COTREC model. This suggests that an excessive focus on alignment and inadequate attention to uniformity could result in sub-optimal model performance.
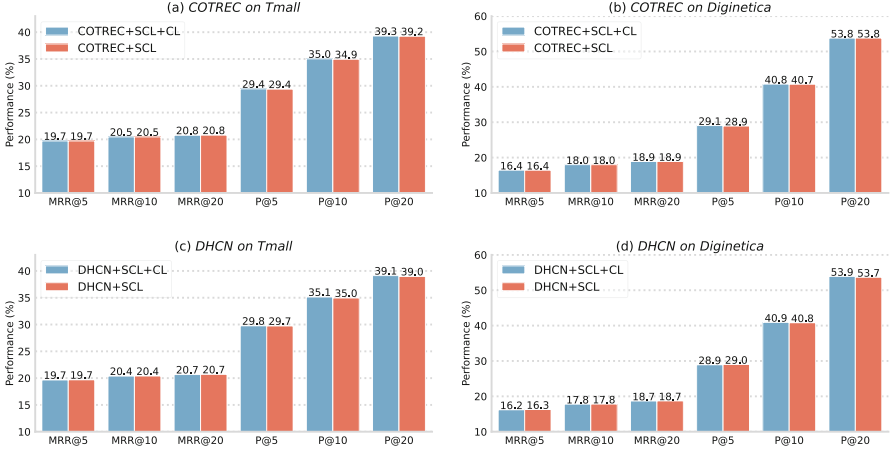
Fig. 3. Ablation studies (**RQ$_3$**) on other CL objectives. Blue indicates the model performance using SCL and other CL objectives. Red represents the model performance using SCL only.

5.4   Sophisticated CL Objectives Are Unnecessary (RQ3)

Given the complexity of CL objectives used in the SOTA models, we investigate the necessity of these complex contrastive objectives with our proposed SCL approach (**RQ$_3$**). We conduct experiments for COTREC and S^2-DHCN on two datasets, TMALL and DIGINETICA with two different settings as follows: (1) **Model + SCL + CL** refers to the model performance with our SCL and all CL objectives in the original model; (1) **Model + SCL** refers to the model performance with our SCL only.

Results. Figure 3 depicts the performance of the models. It can be observed that MODEL + SCL + CL and MODEL + SCL achieve very similar performance results, which implies that the utilization of other sophisticated CL objectives may not be necessary and that the proposed SCL is able to effectively improve the model performance on its own. Below we delve deeper into these findings and discuss them in more detail. Specially, for the TMALL dataset, when using the COTREC model as the backbone, the MODEL + SCL + CL and the MODEL + SCL achieve P@10 scores of 35.0% and 34.9%, respectively, and the same M@10 scores of 20.5%. Similar performance can be observed for the S^2-DHCN model and DIGINETICA dataset, implying that the complexity of these objectives may not be necessary when using SCL.

5.5   Computational Cost (RQ4)

The size of negative samples plays a critical role in the model performance for CL. However, it also has some potential drawbacks, including increased computational resources and model complexity, which may make the proposed method
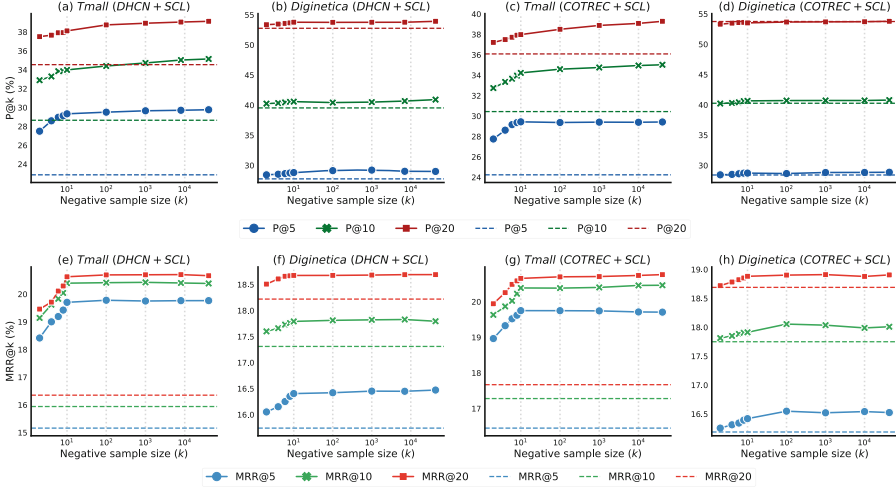
Fig. 4. The impact of negative sample sizes with S^2-DHCN + SCL and COTREC + SCL on the TMALL and DIGINETICA datasets (**RQ**$_4$). The original model performance without using SCL is represented by the corresponding dash line with the same colour. SCL could achieve SOTA performance even when the negative sample size k is equal to 2.

impractical for certain applications or settings. The proposed SCL method has a time complexity of $O(n^2 * d)$, where n is the number of item representations used in Eq. 7 and d is the dimension of item representations. To reduce the computational cost, we simplify the objective function of SCL by encompassing a k-Nearest Neighbour (kNN) with a fast dense embedding retrieval method, which reduces the time complexity from $O(n^2 * d)$ to $O(n * k * d)$, where $k \ll n$. The updated objective is as follows:

$$\mathcal{L}_{\text{SCL}}^{\text{knn}} = -\sum_{i=1}^{n} \log \frac{f(\boldsymbol{x}_i, \boldsymbol{x}_i)}{\sum_{\boldsymbol{x}' \in \mathcal{K}_i} f(\boldsymbol{x}_i, \boldsymbol{x}')}, \tag{9}$$

where $\mathcal{K}_i$ is a set of k nearest item representations in the distance measured by the cosine similarity for the i-th item representation, including its own representation. We conduct experiments to evaluate the impact of negative sample size k, with the values of 2, 4, 6, 8, 10, 100, 1 000, 10 000 and the full set.

Results. Figure 4 presents the model performance in P@k and MRR@k with respect to various values of the negative sample size k on the TMALL and DIGINETICA datasets, where S^2-DHCN + SCL and COTREC + SCL are evaluated. Overall, our experimental results indicate SCL could improve the performance of SOTA models even when the negative sample size k is equal to 2, and that the performance of the models generally improves as the size of negative samples increases. As the negative sample size continues to increase, the improvement
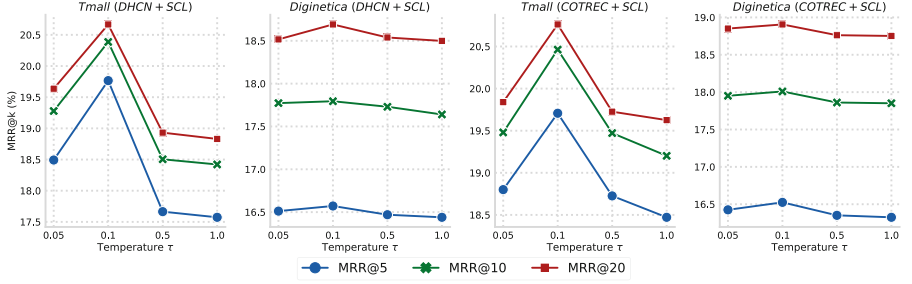
Fig. 5. The effect of the temperature τ using the S^2-DHCN + SCL and COTREC + SCL model on the TMALL and DIGINETICA datasets, where MRR@k is reported as the representative of model performance.

of the model tends to level off and become less noticeable, using a small value for k can produce comparable results to using values greater than $10\,000$. For example, the performance of P@k and MRR@k for the S^2-DHCN + SCL model tends to become stable once the negative sample size reaches 10 on the TMALL dataset, as shown in sub-figure (a) and (e) of Fig. 4.

5.6 Hyperparameter Sensitivity

We conduct an additional study to investigate the effect of varying the hyperparameter temperature τ on model performance. In the experiment, 4 distinct values of τ (namely 0.05, 0.1, 0.5, and 1.0) are evaluated with the S^2-DHCN + SCL and COTREC + SCL on the TMALL and DIGINETICA datasets. The experimental results are presented in Fig. 5, indicating that the model achieves optimal performance when the temperature τ is set to 0.1.

6 Conclusion

In this work, we propose *Self Contrastive Learning* (SCL), which improves the performance of SOTA models with statistical significance across three datasets. SCL targets the optimization of item representation uniformity in SOTA session-based recommendation systems. SCL complements the use of cross-entropy loss, eliminating the need for sophisticated CL objectives. This simplicity makes SCL highly adaptable across a variety of models. Moreover, we shed light on how SCL enhances representation spaces from the alignment and uniformity viewpoints, thus emphasizing the importance of uniformity in item representations. Our analysis also points out that achieving an optimal balance between alignment and uniformity loss is a crucial aspect of designing recommendation systems. Lastly, we demonstrate that the implementation of SCL is efficient and entails low computational costs.

References

1. Bae, S., Kim, S., Ko, J., Lee, G., Noh, S., Yun, S.Y.: Self-contrastive learning: single-viewed supervised contrastive framework using sub-network. Proc. AAAI Conf. Artif. Intell. **37**(1), 197–205 (2023). https://doi.org/10.1609/aaai.v37i1.25091. https://ojs.aaai.org/index.php/AAAI/article/view/25091

2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015 (2015). http://arxiv.org/abs/1409.0473

3. Brost, B., Mehrotra, R., Jehan, T.: The music streaming sessions dataset. In: Liu, L., et al. (eds.) The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, 13–17 May 2019, pp. 2594–2600. ACM, USA (2019). https://doi.org/10.1145/3308558.3313641

4. Chung, J., Gülçehre, Ç., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. CoRR abs/1412.3555 (2014). http://arxiv.org/abs/1412.3555

5. Fu, X., Lipani, A.: Priming and actions: an analysis in conversational search systems. In: Association for Computing Machinery, SIGIR 2023, July 2023. https://doi.org/10.1145/3539618.3592041

6. Fu, X., Yilmaz, E., Lipani, A.: Evaluating the Cranfield paradigm for conversational search systems. In: Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2022, pp. 275–280. Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3539813.3545126

7. Gao, T., Yao, X., Chen, D.: SimCSE: simple contrastive learning of sentence embeddings. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online and Punta Cana, Dominican Republic, pp. 6894–6910. Association for Computational Linguistics (2021). https://doi.org/10.18653/v1/2021.emnlp-main.552. https://aclanthology.org/2021.emnlp-main.552

8. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020, pp. 9726–9735 (2020). https://doi.org/10.1109/CVPR42600.2020.00975

9. Hendriksen, M., Kuiper, E., Nauts, P., Schelter, S., de Rijke, M.: Analyzing and predicting purchase intent in e-commerce: anonymous vs. identified customers. arXiv preprint arXiv:2012.08777 (2020)

10. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: Session-based recommendations with recurrent neural networks. In: Bengio, Y., LeCun, Y. (eds.) 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, Conference Track Proceedings, 2–4 May 2016 (2016). http://arxiv.org/abs/1511.06939

11. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

12. Jannach, D., Ludewig, M., Lerche, L.: Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts. User Model. User-Adap. Inter. **27**(3), 351–392 (2017)

13. Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., Ma, J.: Neural attentive session-based recommendation. In: Lim, E., et al. (eds.) Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, 06–10

November 2017, pp. 1419–1428. ACM, Singapore (2017). https://doi.org/10.1145/3132847.3132926

14. Liu, Q., Zeng, Y., Mokhosi, R., Zhang, H.: STAMP: short-term attention/memory priority model for session-based recommendation. In: Guo, Y., Farooq, F. (eds.) Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, 19–23 August 2018, pp. 1831–1839. ACM, UK (2018). https://doi.org/10.1145/3219819.3219950

15. Liu, Z., Chen, Y., Li, J., Yu, P.S., McAuley, J., Xiong, C.: Contrastive self-supervised sequential recommendation with robust augmentation. arXiv preprint arXiv:2108.06479 (2021)

16. Ma, J., Zhou, C., Yang, H., Cui, P., Wang, X., Zhu, W.: Disentangled self-supervision in sequential recommenders. In: Gupta, R., Liu, Y., Tang, J., Prakash, B.A. (eds.) The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2020, Virtual Event, CA, USA, 23–27 August 2020, pp. 483–491. ACM (2020). https://dl.acm.org/doi/10.1145/3394486.3403091

17. Nie, P., et al.: MIC: model-agnostic integrated cross-channel recommender. In: Proceedings of the 31st ACM International Conference on Information and Knowledge Management, CIKM 2022, pp. 3400–3409. Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3511808.3557081

18. van den Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)

19. Qiu, R., Huang, Z., Yin, H., Wang, Z.: Contrastive learning for representation degeneration problem in sequential recommendation. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, WSDM 2022, pp. 813–823. Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3488560.3498433

20. Qiu, R., Li, J., Huang, Z., Yin, H.: Rethinking the item order in session-based recommendation with graph neural networks. In: Zhu, W., et al. (eds.) Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, 3–7 November 2019, pp. 579–588. ACM, Beijing (2019). https://doi.org/10.1145/3357384.3358010

21. Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.: Factorizing personalized Markov chains for next-basket recommendation. In: Rappa, M., Jones, P., Freire, J., Chakrabarti, S. (eds.) Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, 26–30 April 2010, pp. 811–820. ACM, USA (2010). https://doi.org/10.1145/1772690.1772773

22. Shani, G., Heckerman, D., Brafman, R.I.: An MDP-based recommender system. J. Mach. Learn. Res. **6**, 1265–1295 (2005)

23. Shi, Z., Feng, Y., Lipani, A.: Learning to execute actions or ask clarification questions. In: Findings of the Association for Computational Linguistics, NAACL 2022, Seattle, United States, pp. 2060–2070. Association for Computational Linguistics (2022). https://doi.org/10.18653/v1/2022.findings-naacl.158. https://aclanthology.org/2022.findings-naacl.158

24. Shi, Z., Lipani, A.: Don't stop pretraining? Make prompt-based fine-tuning powerful learner. In: Thirty-seventh Conference on Neural Information Processing Systems. NeurIPS (2023). https://openreview.net/forum?id=s7xWeJQACI

25. Shi, Z., Ni, P., Wang, M., Kim, T.E., Lipani, A.: Attention-based ingredient parser. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Bruges, Belgium (2022). https://doi.org/10.14428/esann/2022.ES2022-10

26. Shi, Z., Ramos, J., Kim, T.E., Wang, X., Rahmani, H.A., Lipani, A.: When and what to ask through world states and text instructions: IGLU NLP challenge solution. In: Advances in Neural Information Processing Systems (NeurIPS), IGLU Workshop (2023). https://nips.cc/virtual/2022/66405

27. Shi, Z., Tonolini, F., Aletras, N., Yilmaz, E., Kazai, G., Jiao, Y.: Rethinking semi-supervised learning with language models. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) Findings of the Association for Computational Linguistics, ACL 2023, Toronto, Canada, July 2023, pp. 5614–5634. Association for Computational Linguistics (2023). https://doi.org/10.18653/v1/2023.findings-acl.347. https://aclanthology.org/2023.findings-acl.347

28. Shi, Z., Zhang, Q., Lipani, A.: StepGame: a new benchmark for robust multi-hop spatial reasoning in texts. In: Proceedings of the AAAI Conference on Artificial Intelligence, June 2022, vol. 36, pp. 11321–11329 (2022). https://doi.org/10.1609/aaai.v36i10.21383. https://ojs.aaai.org/index.php/AAAI/article/view/21383

29. Song, W., Xiao, Z., Wang, Y., Charlin, L., Zhang, M., Tang, J.: Session-based social recommendation via dynamic graph attention networks. In: Culpepper, J.S., Moffat, A., Bennett, P.N., Lerman, K. (eds.) Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, 11–15 February 2019, pp. 555–563. ACM (2019). https://doi.org/10.1145/3289600.3290989

30. Sun, F., et al.: BERT4Rec: sequential recommendation with bidirectional encoder representations from transformer. In: Zhu, W., et al. (eds.) Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, 3–7 November 2019, pp. 1441–1450. ACM, Beijing (2019). https://doi.org/10.1145/3357384.3357895

31. Wang, L., Lim, E.P., Liu, Z., Zhao, T.: Explanation guided contrastive learning for sequential recommendation. In: Proceedings of the 31st ACM International Conference on Information and Knowledge Management, CIKM 2022, pp. 2017–2027. Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3511808.3557317

32. Wang, T., Isola, P.: Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In: Proceedings of the 37th International Conference on Machine Learning, ICML 2020, Virtual Event, 13–18 July 2020, vol. 119, pp. 9929–9939. Proceedings of Machine Learning Research (PMLR) (2020). http://proceedings.mlr.press/v119/wang20k.html

33. Wang, W., et al.: Beyond clicks: modeling multi-relational item graph for session-based target behavior prediction. In: Huang, Y., King, I., Liu, T., van Steen, M. (eds.) The Web Conference 2020, WWW 2020, Taipei, Taiwan, 20–24 April 2020, pp. 3056–3062. ACM/IW3C2, Taiwan (2020). https://doi.org/10.1145/3366423.3380077

34. Wang, Z., Wei, W., Cong, G., Li, X., Mao, X., Qiu, M.: Global context enhanced graph neural networks for session-based recommendation. In: Huang, J., et al. (eds.) Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2020, Virtual Event, China, 25–30 July 2020, pp. 169–178. ACM (2020). https://doi.org/10.1145/3397271.3401142

35. Wei, Y., et al.: Contrastive learning for cold-start recommendation. In: Proceedings of the 29th ACM International Conference on Multimedia, MM 2021, pp. 5382–5390. Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3474085.3475665

36. Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., Tan, T.: Session-based recommendation with graph neural networks. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, 27 January–1 February 2019, pp. 346–353. AAAI Press (2019). https://doi.org/10.1609/aaai.v33i01.3301346

37. Xia, X., Yin, H., Yu, J., Shao, Y., Cui, L.: Self-supervised graph co-training for session-based recommendation. In: Proceedings of the 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, pp. 2180–2190. ACM (2021). https://doi.org/10.1145/3459637.3482388. https://dl.acm.org/doi/10.1145/3459637.3482388

38. Xia, X., Yin, H., Yu, J., Wang, Q., Cui, L., Zhang, X.: Self-supervised hypergraph convolutional networks for session-based recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, vol. 35, pp. 4503–4511. AAAI (2021). https://doi.org/10.1609/aaai.v35i5.16578. https://ojs.aaai.org/index.php/AAAI/article/view/16578

39. Xie, R., Qiu, Z., Zhang, B., Lin, L.: Multi-granularity item-based contrastive recommendation. arXiv preprint arXiv:2207.01387 (2022)

40. Xie, X., Sun, F., Liu, Z., Gao, J., Ding, B., Cui, B.: Contrastive pre-training for sequential recommendation. arXiv preprint arXiv:2010.14395 (2020)

41. Xie, X., et al.: Contrastive learning for sequential recommendation. In: 2022 IEEE 38th International Conference on Data Engineering (ICDE), virtual, pp. 1259–1273. IEEE (2022). https://ieeexplore.ieee.org/abstract/document/9835621

42. Xu, C., et al.: Graph contextualized self-attention network for session-based recommendation. In: Kraus, S. (ed.) Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, 10–16 August 2019, pp. 3940–3946. ijcai.org (2019). https://doi.org/10.24963/ijcai.2019/547

43. Yao, T., et al.: Self-supervised learning for deep models in recommendations. arXiv e-prints pp. arXiv-2007 (2020)

44. Yin, H., Cui, B.: Spatio-Temporal Recommendation in Social Media. SCS, Springer, Singapore (2016). https://doi.org/10.1007/978-981-10-0748-4

45. Yu, J., Yin, H., Gao, M., Xia, X., Zhang, X., Hung, N.Q.V.: Socially-aware self-supervised tri-training for recommendation. arXiv preprint arXiv:2106.03569 (2021)

46. Yu, J., Yin, H., Li, J., Wang, Q., Hung, N.Q.V., Zhang, X.: Self-supervised multi-channel hypergraph convolutional network for social recommendation. arXiv preprint arXiv:2101.06448 (2021)

47. Yu, J., et al.: Are graph augmentations necessary? Simple graph contrastive learning for recommendation. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2022, pp. 1294–1303. Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3477495.3531937

48. Zhang, Y., et al.: Sequential click prediction for sponsored search with recurrent neural networks. In: Brodley, C.E., Stone, P. (eds.) Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 27–31 July 2014, Québec City, Québec, Canada, pp. 1369–1375. AAAI Press (2014). http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8529

49. Zhou, K., et al.: S3-Rec: self-supervised learning for sequential recommendation with mutual information maximization. In: d'Aquin, M., Dietze, S., Hauff, C., Curry, E., Cudré-Mauroux, P. (eds.) The 29th ACM International Conference on Information and Knowledge Management, CIKM 2020, Virtual Event, Ireland, 19–23 October 2020, pp. 1893–1902. ACM (2020). https://doi.org/10.1145/3340531.3411954
50. Zhou, X., Sun, A., Liu, Y., Zhang, J., Miao, C.: SelfCF: a simple framework for self-supervised collaborative filtering. ACM Trans. Recomm. Syst. **1**, 1–25 (2023). https://doi.org/10.1145/3591469
51. Zimdars, A., Chickering, D.M., Meek, C.: Using temporal data for making recommendations. arXiv preprint arXiv:1301.2320 (2013)