Self Contrastive Learning for Session-based Recommendation

Zhengxiang Shi
University College London
London, United Kingdom
zhengxiang.shi.19@ucl.ac.uk

Xi Wang
University College London
London, United Kingdom
xi-wang@ucl.ac.uk

Aldo Lipani
University College London
London, United Kingdom
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 benchmark datasets, 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.

CCS CONCEPTS

• **Information systems → Recommender systems**; **Learning to rank**.

KEYWORDS

Recommender System, Contrastive Learning, Representation Learning

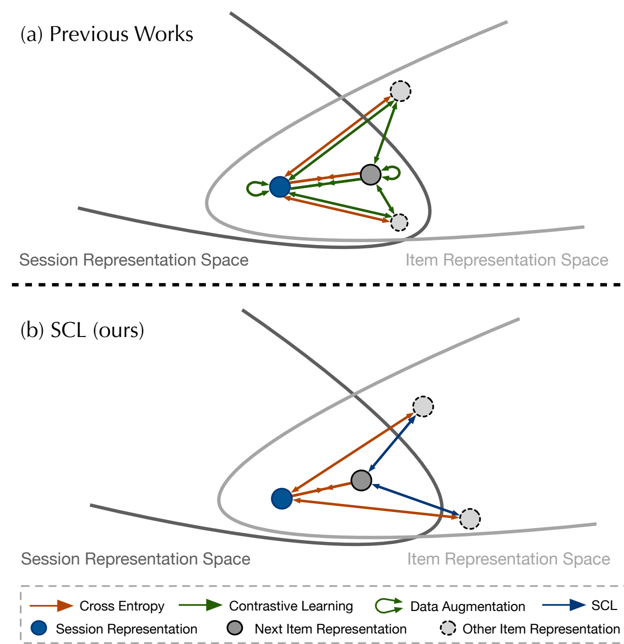

Figure 1: The illustration of the framework of SCL. In previous works, contrastive learning (CL) objectives (depicted in green) typically involve complicated modelling, such as extra data augmentation and complex creation of positive/negative samples. Additionally, the cross-entropy loss (depicted in red) and CL objectives (depicted in green) play a similar role. These designs have led 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.

1 INTRODUCTION

The session-based recommendation [15, 16, 47, 48] is a crucial aspect of modern recommender systems [19, 24–26, 40–43], as it aims to predict a user's next interest by focusing on their current intent. It has become an important area of research due to the growing amount of data generated by users through their interactions on various platforms such as e-commerce websites [10, 13], music streaming services [3], and social media [35]. Recently, contrastive learning (CL) [21] has been applied in session-based recommendation tasks, with the goal of aligning the session representation with the next item's representation, while also distinguishing it from other item representations. The intention behind these approaches is to enhance recommendation accuracy via improved representation quality. However, two key limitations exist within these methods.

Firstly, CL loss in previous research [20, 22, 45, 47–49, 58] serves a similar role as the cross-entropy loss while the optimisation of item representation space is not adequately addressed. As shown in Figure 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. We delve into their overlapping role further in §4.1. Furthermore, the uniformity of item representation typically plays a relatively minor supplementary role in contrast to other CL objectives, while earlier studies [20, 22, 45, 49, 58] have emphasized improving the uniformity of representations. For instance, it only contributes to a small portion of the overall loss [49]. We argue that the importance of optimising the item representation space *itself* by ensuring that the representations are uniformly distributed is not receiving adequate attention.

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. Indeed, two state-of-the-art session-based recommendation models, S^2-DHCN [48] and COTREC [47] 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 [58] or dropout [49, 61] to augment representations, as shown in Figure 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 state-of-the-art 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 [22, 37, 45, 50, 58], 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:

- $\mathbf{RQ}_1$ To what extent does SCL enhance performance in session-based recommendation tasks? (§5.2)
- $\mathbf{RQ}_2$ How does SCL improve the model performance in terms of the alignment and uniformity of representations? (§5.3)
- $\mathbf{RQ}_3$ Is the use of those sophisticated CL objectives still necessary in the presence of SCL? (§5.4)
- $\mathbf{RQ}_4$ Can SCL maintain state-of-the-art performance with a low computational cost? (§5.5)

In order to address $\mathbf{RQ}_1$, we conduct extensive experiments on three datasets, Tmall, Nowplaying, and Diginetica (§5.2). Our experimental results demonstrate that SCL consistently improves the performance of state-of-the-art 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. Additionally, SCL also brings notable improvement on Nowplaying and Diginetica, leading to new state-of-the-art performance.

To gain insight into how the model is improved ($\mathbf{RQ}_2$), we investigate the transformations of the session and item representation distributions 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, albeit with increased alignment loss. Our findings suggest that state-of-the-art approaches may have placed excessive emphasis on the alignment of session and item representations.

To answer $\mathbf{RQ}_3$, 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.

Given that the computational complexity of SCL is of the quadratic order with respect to the number of item candidates ($\mathbf{RQ}_4$), we further study the impact of selecting the k-nearest item representations on the model performance (§5.5). Our results show that SCL generally benefits from contrasting to more item representations. However, it can still achieve state-of-the-art 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

2.1 Session-based Recommendation

The session-based recommendation aims to predict the next item by utilizing user behaviours within a short time period. Early studies on session-based recommendation focused on utilising temporal information from session data through the use of markov chain models [5, 27, 28, 55, 62]. With the advent of neural networks [6, 30, 33, 34], recurrent neural networks (RNNs) [12] have been applied to session-based recommendation models to capture the sequential order between items [59]. GRU4Rec [11] was the first model to use gated recurrent units (GRUs) [4] to model the sequential relations of item interactions. NARM [15] extended GRU4Rec by incorporating the attention mechanism [2, 29, 31, 32] to extract the main intent in the current session while also modelling its sequential orders. STAMP [16] also replaced the recurrent structure with attention layers to capture a user's general and current interests.

Graph-based methods have been applied to session-based recommendation systems in order to learn item transitions over graphs. SR-GNN [46] models session sequences in session graphs and employs a gated Graph Neural Network (GNN) model to aggregate information between items into session representations. MGNN-SPred [39] builds a multi-relational item graph based on all session clicks to learn global item associations and uses a gated mechanism to predict the next item. GC-SAN [53] dynamically constructs session-induced graphs and uses self-attention networks on the graphs to capture item dependencies through graph information aggregation. FGNN [23] rethinks the sequence order of items to exploit users' intrinsic intents using GNNs. GCE-GNN [44] aggregates item information from both the item-level and session-level through graph convolution and self-attention mechanism. S^2-DHCN [48] utilizes hyper-graph convolutional networks to capture high-order item relations within individual sessions and uses self-supervised learning to enhance session representations. COTREC [47] integrated self-supervised learning into the graph training through sophisticated positive and negative constructions.

2.2 Contrastive Learning

CL has achieved great success in various research domains, such as computer vision [8, 9] and natural language processing [7], 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. Bert4Rec [36] 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 [60] utilizes intrinsic data correlations among attributes, items, subsequences, and sequences to generate contrastive signals and enhance data representations through pre-training. In addition, Xie et al. [50] proposed three data augmentation strategies to construct contrastive signals from original user behaviour sequences, in order to extract more meaningful user patterns and encode effective user representations. Ma et al. [18] proposed a sequence-to-sequence training strategy based on latent self-supervision and disentanglement of user intention behind behaviour sequences. CoSeRec [17] uses Graph Neural Networks (GNNs) to capture more complex patterns than sequential patterns through CL objectives. CL4SRec [51] combines recommendation loss with contrastive loss of self-supervised tasks to optimize the sequential recommendation model. DuoRec [22] 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. CL has also been applied to other recommendation paradigms, such as general recommendation [54] and social recommendation [56, 57]. In this work, we specifically focus on session-based recommendation tasks, where the most closely related works to our study are S^2-DHCN [48] and COTREC [47] While these two approaches have been acknowledged as state-of-the-art models with satisfactory performance, they suffer from two primary limitations, *i.e.,* the complex creation of positive/negative samples and modelling and the overlook of the importance of optimising the item representation space.

3 PRELIMINARIES

3.1 Task Definition

In the session-based recommendation task, 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 of a session-based recommendation task 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.

3.2 Contrastive Learning

Contrastive learning is introduced 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]. For instance, in the field of information retrieval, the anchor vector is commonly the representation of a query, while the positive and negative samples are the relevant and irrelevant documents to the query, respectively [1, 14, 52]. Here we introduce a specific CL method known as InfoNCE, which is commonly used in recommendation systems [22, 48, 50, 51], and two metrics (*i.e.,* alignment and uniformity) [38] to evaluate the quality of learned representations.

Noise Contrastive Estimation. InfoNCE [21], where NCE stands for Noise Contrastive Estimation, is a type of contrastive loss function. 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.

Alignment and Uniformity. In the field of CL, two key properties, known as alignment and uniformity, have been proposed by Wang and Isola [38] as measures of the quality of representations. The uniformity of the embeddings 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:

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

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.

These two measures align well with the CL objective, which is to keep positive instances close and scatter embeddings for random instances on the unit hypersphere. Hence, we aim to leverage the alignment and uniformity of item representations to gain deeper and additional insights into the inner workings of our novel CL-based approach.

4 METHODOLOGY

In this section, we first discuss the potential limitations of state-of-the-art session-based recommendation systems (§4.1). To address these issues, we then introduce a novel approach, referred to as *Self Contrastive Learning* (SCL), which aims to improve the uniformity in item representations by utilising a novel loss function (§4.2).

4.1 Motivation

In the field of session-based recommendation, existing works [44, 47, 48], 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, as follows:

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

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 Figure 1. Specifically, let $\boldsymbol{s}$ 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:

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

$$= -\log \frac{f(\boldsymbol{s}, \boldsymbol{x}^{+})}{\sum_{j=1}^{n} f(\boldsymbol{s}, \boldsymbol{x}_j)}, \tag{7}$$

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. 7 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 Figure 1(a) where the green lines and red lines are overlapped. Secondly, while prior work [49] attempted to improve the uniformity of the item representation space with some auxiliary losses, the importance of blue lines (see Figure 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.

4.2 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 Figure 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:

$$\mathcal{L}_{\text{SCL}} = -\sum_{i=1}^{n} \log \frac{g(\boldsymbol{x}_i, \boldsymbol{x}_i)}{\sum_{j=1}^{n} g(\boldsymbol{x}_i, \boldsymbol{x}_j)}, \tag{8}$$

where the function $g(\boldsymbol{x}, \boldsymbol{x}')$ is computed by $e^{\text{sim}(\boldsymbol{x}, \boldsymbol{x}')/\tau}$, the exponential of the cosine similarity controlled by a temperature parameter τ. Using the cosine similarity, this loss pulls apart items on the unit hypersphere, which is what ℓ_{uniform} measures.

Next, we integrate $\mathcal{L}_{\text{SCL}}$ 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:

$$\mathcal{L} = \mathcal{L}_{model} + \beta \mathcal{L}_{\text{SCL}}, \tag{9}$$

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 $\ell_{uniform}$, $\mathcal{L}_{\text{SCL}}$ 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 [58] or dropout [49]. 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.

Table 1: Statistics of datasets.

Dataset	Tmall	Nowplaying	Diginetica
Train Size	351,268	825,304	719,470
Test Size	25,898	89,824	60,858
Items Size	40,728	60,417	43,097
Average Length	6.69	7.42	5.12

5 SESSION-BASED RECOMMENDATION

In this section, we evaluate our proposed SCL method in three session-based recommendation 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 §1.

5.1 Experimental Setup

Datasets. Adhering to previous works [44, 47, 48], we evaluate our proposed SCL method using three benchmark datasets. The statistics for these datasets are presented in Table 1. The datasets include:

- **Tmall**[1]: The Tmall dataset is sourced from the IJCAI-15 competition and includes anonymized shopping logs from users on the Tmall online shopping platform.
- **Nowplaying**[2]: The Nowplaying dataset describes the music-listening behaviour of users.
- **Diginetica**[3]: The Diginetica dataset, from CIKM Cup 2016, comprises of typical transaction data.

Baselines. Our proposed SCL method is compared with the following representative methods:

- **FPMC** [27] is a method for a sequential recommendation that uses Markov Chain. To apply it to the session-based recommendation, user latent representations are not taken into account when calculating recommendation scores.
- **GRU4REC** [11] is a method for modelling user sequences in the session-based recommendation. It employs a parallel training process for mini-batches of sessions and uses ranking-based loss functions to optimize the model.
- **NARM** [15] is a RNN-based method for session-based recommendation. It uses an attention mechanism [2] to understand the main purpose of the user and combines this with their sequential behaviour to generate recommendations.
- **STAMP** [16] is a session-based recommendation model that uses attention layers instead of RNN encoders. It employs the self-attention mechanism to improve its performance.
- **SR-GNN** [46] is a session-based recommendation model that utilizes a gated graph convolutional layer to generate item embeddings and a soft-attention mechanism to compute session embeddings.
- **GCE-GNN** [44] is a state-of-the-art session-based recommendation model that creates two types of session-induced graphs to capture both local and global information at different levels.
- **S^2-DHCN** [48] is a state-of-the-art session-based recommendation model that creates two types of hypergraphs to capture both inter- and intra-session information. It also employs self-supervised learning to improve its performance.
- **COTREC** [47] is a state-of-the-art session-based recommendation model that utilizes two separate graph encoders to generate additional self-supervised signals via session-based data augmentation. The model employs a self-supervised objective to enhance performance.

Evaluation Metrics. Following the protocol in previous works [46–48], 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. P@k is a commonly used measure of predictive accuracy, which reflects the proportion of correctly recommended items among the top-k items. MRR@k is a measure that takes into account the order of the recommended items and calculates the average of the reciprocal ranks of the correctly recommended items. A large MRR@k value indicates that correct recommendations are placed at the top of the ranking list.

Implementation Details. We conduct experiments with the proposed SCL method using three state-of-the-art models, GCE-GNN[4], S^2-DHCN[5], and COTREC[6]. We first reproduce the experimental results of these models by following the settings and protocols specified 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 dataset as we were unable to replicate the results.

It is noteworthy that, to demonstrate the effectiveness of our proposed SCL method, we adhere to the settings of hyperparameters suggested in original papers and do not perform any additional hyperparameter optimization during our implementation process. In other words, we do not make any changes to the existing settings, except for incorporating our proposed SCL method. Further exploration of the hyperparameter space may lead to additional performance enhancements introduced by the proposed SCL method.

5.2 Main results (RQ1)

Table 2 presents the performance of all comparison methods, where the proposed SCL is applied to three state-of-the-art models, GCE-GNN, COTREC, and S^2-DHCN. Our experimental results demonstrate 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 state-of-the-art performance (highlighted in blue). The significance tests further corroborate the effectiveness of SCL. Below we present the experimental results on each dataset in more detail.

Particularly remarkable is that SCL achieves a notable improvement compared to the state-of-the-art models on the Tmall dataset.

[1]https://tianchi.aliyun.com/dataset/dataDetail?dataId=42
[2]http://dbis-nowplaying.uibk.ac.at/#nowplaying
[3]https://competitions.codalab.org/competitions/11161
[4]https://github.com/CCIIPLab/GCE-GNN
[5]https://github.com/xiaxin1998/DHCN
[6]https://github.com/xiaxin1998/COTREC

Table 2: Performances of all comparison methods on the development set on three datasets (RQ_1). Results marked with † are taken from the original paper, while those marked with ‡ are our own reproductions. Triangles in colours indicate an improvement in performance compared to our reproduced results. The highest results in each column are highlighted in bold font, and new state-of-the-art performances are indicated in blue. The asterisks denote the level of statistical significance of the improvement: *** indicates a p-value < 1e-20, ** 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%)

Specifically, SCL improves the performance of GCE-GNN by more than 2.3% in terms of MRR@10 and MRR@20. Similarly, the 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, our proposed method, S^2-DHCN + SCL achieves a new state-of-the-art 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%. COTREC + SCL and S^2-DHCN + SCL have their own advantages over different evaluation metrics, exceeding the previous state-of-the-art performance with substantial improvements. This demonstrates the effectiveness of our proposed SCL method.

On the Nowplaying dataset, our results indicate that the proposed method SCL consistently improves the performance of the state-of-the-art models, GCE-GNN and S^2-DHCN. Specifically, the GCE-GNN + SCL method results in an increase of 1.5% and 1.7% in HIT@10 and HIT@20 respectively. Additionally, the S^2-DHCN + SCL method demonstrates a marked improvement, with 2.2% and 3.2% increases in HIT@10 and HIT@20 respectively, compared to its own performance without SCL.

On the Diginetica dataset, the proposed SCL method consistently improves the performance of all models, as observed in the Tmall and Nowplaying datasets. In the case of GCE-GNN, SCL provides a 2.3% and 1.8% increase in HIT@10 and MRR@10, respectively. Similarly, for the COTREC model, SCL leads to an increase of 1.3% and 1.4% in HIT@10 and MRR@10, respectively. The performance improvement is also observed for the S^2-DHCN model, where SCL brings a 3.5% and 2.8% increase in terms of HIT@10 and MRR@10, respectively. Overall, the GCE-GNN model with the proposed SCL method attains a new state-of-the-art performance, as shown in Table 2.

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**). In this section, we explore this question from the perspective of alignment of session and item representations and uniformity of item representations. 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 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. Below we delve deeper into these two findings and discuss them in more detail.

Better uniformity of item representations brings substantial improvement in performance. The sub-figure in the centre of Figure 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%. Similarly, the uniformity

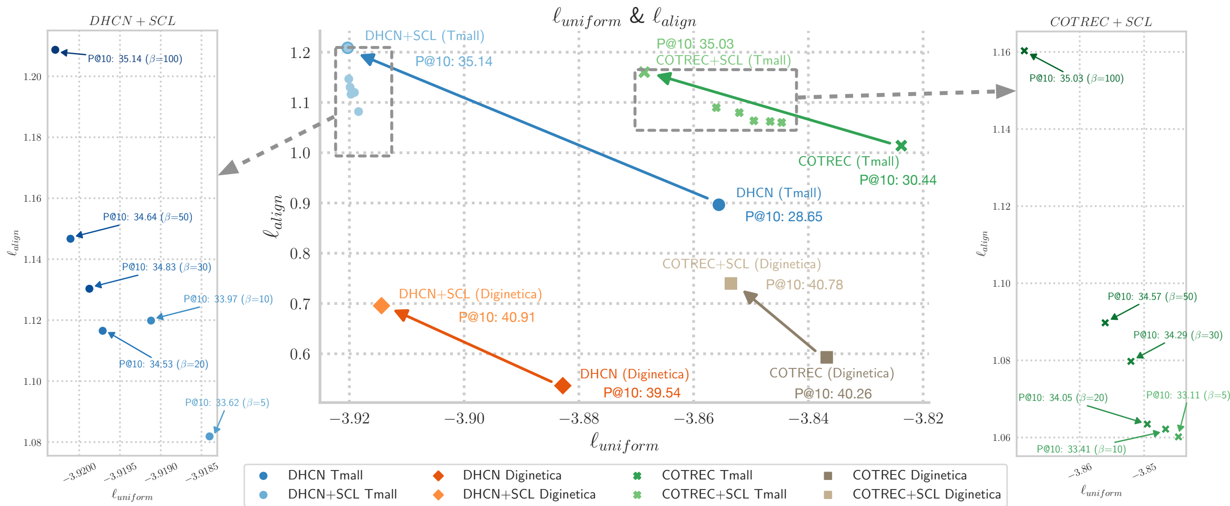

Figure 2: The analysis of alignment loss ℓ_{align} and uniformity loss ℓ_{uniform} on the Tmall and Diginetica datasets, where P@10 is reported as the representative of model performance. The central sub-figure illustrates the impact of the proposed SCL method on the S^2-DHCN and COTREC model in terms of alignment and uniformity loss. The sub-figure on either side is the close-up of a portion of the central sub-figure, where the trade-off between alignment and uniformity loss is controlled by the SCL loss weight β. 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.

loss of COTREC is reduced from -3.82 to -3.87 by applying the SCL on the Tmall dataset, with an improvement in P@10 from 30.44% to 35.03%. These results suggest that the proposed SCL method is effective in encouraging the item representations to be more distinct from one another, which leads to improved performance.

The trade-off between alignment and uniformity. In addition to the reduction in uniformity loss that results in improved model performance, 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. To further understand the trade-off between these two factors, we conducted additional studies on the Tmall dataset by adjusting the alignment and uniformity loss through controlling the SCL loss weight β.

The results of these studies are depicted in the two sub-figures (on two sides) of Figure 2, which provide a closer look at the effect of different combinations of alignment and uniformity loss on the model performance. Specifically, as we increase the SCL loss weight β using the S^2-DHCN model, the uniformity loss gradually decreases from -3.86 to -3.92, while the alignment loss increases from 1.08 to around 1.20. During the process, 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.

5.4 Sophisticated CL objectives are unnecessary (RQ3)

Given the complexity of CL objectives used in the state-of-the-art models, S^2-DHCN and COTREC, we investigate the necessity of these objectives in the presence of our proposed SCL approach. Specifically, we aim to investigate the effect of these contrastive objectives in the presence of the proposed SCL (**RQ**$_3$).

Setup. To evaluate the effectiveness of the CL objectives used in COTREC and S^2-DHCN, we conduct experiments with two different settings as follows:

- **Model + SCL + CL** refers to the model performance with the proposed SCL method and all CL objectives in the original model;
- **Model + SCL** refers to the model performance with the proposed SCL method only.

The experiments are conducted on two datasets, Tmall and Diginetica. It is worth mentioning that we use the default and same hyperparameter for each model, and no additional hyperparameter tuning is performed for different settings.

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.

For the Tmall dataset, when using the COTREC model as the backbone, the Model + SCL + CL method and the Model + SCL

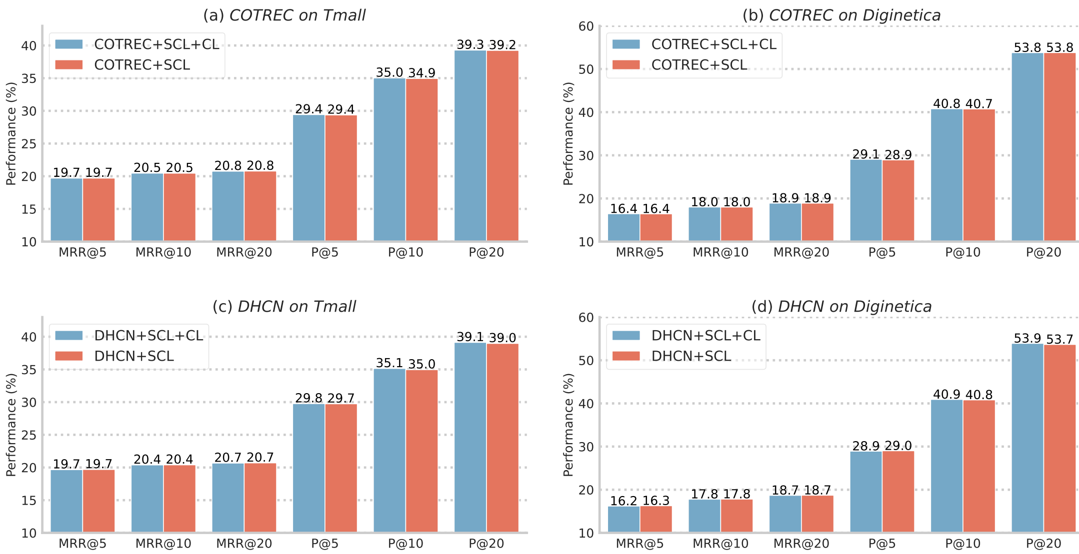

Figure 3: Ablation studies (RQ_3) for other CL objectives in the S^2-DHCN or COTREC model on Tmall and Diginetica datasets. Blue indicates the model performance of the proposed SCL method together with other CL objectives. Red represents the model performance of the proposed SCL method.

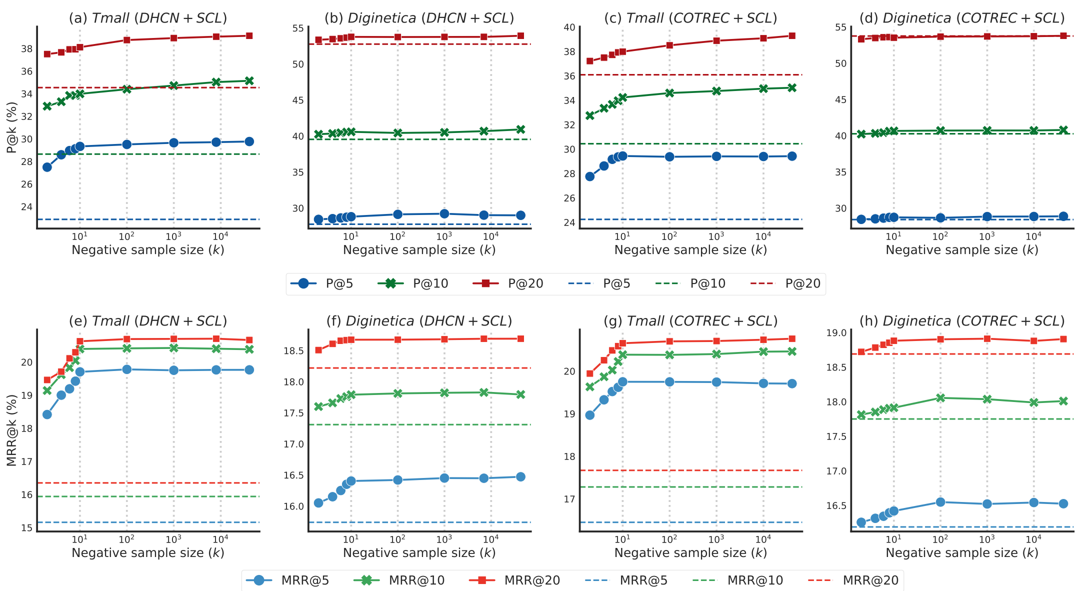

Figure 4: The impact of negative sample sizes with S^2-DHCN + SCL and COTREC + SCL on the Tmall and Diginetica datasets (RQ_4). The original model performance without using SCL is represented by the corresponding dash line with the same colour. SCL could achieve state-of-the-art performance even when the negative sample size k is equal to 2.

method achieve P@10 scores of 35.0% and 34.9%, respectively, and the same M@10 scores of 20.5%. Similarly, when using the S^2-DHCN model as the backbone, the Model + SCL + CL and Model + SCL methods achieve comparable performance, with P@10 scores of 35.1% and 35.0%, respectively, and the same M@10 scores of 20.4%.

For the Diginetica dataset, when using the COTREC model as the backbone, the Model + SCL + CL method and the Model + SCL method yield P@10 scores of 40.8% and 40.7%, respectively, and M@10 scores of 18.0%. Similarly, when using the S^2-DHCN model as the backbone, the Model + SCL + CL and Model + SCL methods achieve comparable performance, with P@10 scores of 40.9% and 40.8%, respectively, and the same M@10 scores of 17.8%.

We also observe that Model + SCL can achieve even better performance than Model + SCL + CL. One such example is on the Diginetica dataset, where the Model + SCL attains a P@5 score of 29.0% while the Model + SCL + CL achieves a P@5 score of 28.9% when utilizing the S^2-DHCN model as the backbone. Our experimental results reveal that the proposed SCL method can achieve similar performance without relying on more complex CL objectives, implying that the complexity of these objectives may not be necessary when using SCL.

5.5 The trade-off between model performance and computational cost (RQ4)

The size of negative samples plays a critical role in the model performance, particularly in the context of CL. However, in addition to the potential benefits, it is crucial to also consider the potential drawbacks, including increased computational resources and model complexity, which may make the proposed method impractical for certain applications or settings. In this section, we carefully consider the trade-off between model performance and computational cost when training session-based recommendation models with our proposed SCL method (**RQ**$_4$).

Setup. The proposed SCL method has a time complexity of $O(n^2 * d)$, where n is the number of item representations used in Eq. 8 and d is the dimension of item representations. To reduce the time complexity and computational cost, we update the objective function of SCL by encompassing a k-Nearest Neighbour (kNN) component that boosts the efficiency of SCL with a fast dense embedding retrieval method. As a result, with pre-computed item representations, the kNN-variant of SCL reduces the time complexity from $O(n^2 * d)$ to $O(n * k * d)$, where $k \ll n$. The updated objective is calculated as follows:

$$\mathcal{L}_{\text{SCL}}^{\text{knn}} = -\sum_{i=1}^{n} \log \frac{f(\boldsymbol{x}_i, \boldsymbol{x}_i)}{\sum\limits_{\boldsymbol{x}' \in \mathcal{K}_i} f(\boldsymbol{x}_i, \boldsymbol{x}')}, \tag{10}$$

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. To optimise the value of k from the full set of candidates, we conduct further 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 state-of-the-art 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 of the model tends to level off and become less noticeable. 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 Figure 4. Our experimental results show that using a small value for k can produce comparable results to using values greater than 10 000, thus demonstrating that the SCL can be implemented with a reasonable computational cost.

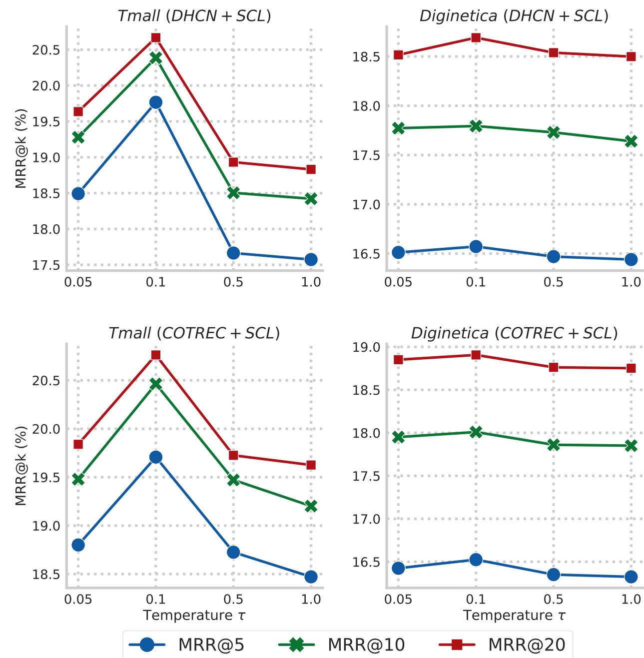

Figure 5: The effect of the temperature τ using the S^2-DHCN + SCL and COTREC + SCL model on the Tmall and Diginetica datasets, where MRR@k is reported as the representative of model performance.

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 Figure 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 state-of-the-art models with statistical significance across three datasets. SCL targets the optimization of item representation uniformity in state-of-the-art session-based recommendation systems. SCL serves as a valuable supplement to the use of cross-entropy loss, eliminating the need for sophisticated CL objectives, which usually require extra positive/negative creation and training processes. This simplicity makes SCL highly adaptable across a variety of models. Moreover, we delve into the workings of SCL, shedding light on how it 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] Abhijit Anand, Jurek Leonhardt, Koustav Rudra, and Avishek Anand. 2022. Supervised Contrastive Learning Approach for Contextual Ranking. In *Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval* (Madrid, Spain) *(ICTIR '22)*. Association for Computing Machinery, New York, NY, USA, 61–71. https://doi.org/10.1145/3539813.3545139

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *3rd International Conference on Learning Representations, ICLR 2015*, Yoshua Bengio and Yann LeCun (Eds.). ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. http://arxiv.org/abs/1409.0473

[3] Brian Brost, Rishabh Mehrotra, and Tristan Jehan. 2019. The Music Streaming Sessions Dataset. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia (Eds.). ACM, usa, 2594–2600. https://doi.org/10.1145/3308558.3313641

[4] Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *CoRR* abs/1412.3555 (2014). arXiv:1412.3555 http://arxiv.org/abs/1412.3555

[5] Xiao Fu and Aldo Lipani. 2023. Priming and Actions: An Analysis in Conversational Search Systems *(SIGIR '23)*. Association for Computing Machinery. https://doi.org/10.1145/3539618.3592041

[6] Xiao Fu, Emine Yilmaz, and Aldo Lipani. 2022. Evaluating the Cranfield Paradigm for Conversational Search Systems. In *Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval* (Madrid, Spain) *(ICTIR '22)*. Association for Computing Machinery, New York, NY, USA, 275–280. https://doi.org/10.1145/3539813.3545126

[7] Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 6894–6910. https://doi.org/10.18653/v1/2021.emnlp-main.552

[8] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 9726–9735. https://doi.org/10.1109/CVPR42600.2020.00975

[9] Mariya Hendriksen, Maurits Bleeker, Svitlana Vakulenko, Nanne van Noord, Ernst Kuiper, and Maarten de Rijke. 2022. Extending CLIP for Category-to-Image Retrieval in E-Commerce. In *Advances in Information Retrieval: 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10–14, 2022, Proceedings, Part I* (Stavanger, Norway). Springer-Verlag, Berlin, Heidelberg, 289–303. https://doi.org/10.1007/978-3-030-99736-6_20

[10] Mariya Hendriksen, Ernst Kuiper, Pim Nauts, Sebastian Schelter, and Maarten de Rijke. 2020. Analyzing and predicting purchase intent in e-commerce: anonymous vs. identified customers. *arXiv preprint arXiv:2012.08777* (2020).

[11] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1511.06939

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

[13] Dietmar Jannach, Malte Ludewig, and Lukas Lerche. 2017. Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts. *User Modeling and User-Adapted Interaction* 27, 3 (2017), 351–392.

[14] Omar Khattab and Matei Zaharia. 2020. ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, Jimmy Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu (Eds.). ACM, 39–48. https://doi.org/10.1145/3397271.3401075

[15] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural Attentive Session-based Recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, Ee-Peng Lim, Marianne Winslett, Mark Sanderson, Ada Wai-Chee Fu, Jimeng Sun, J. Shane Culpepper, Eric Lo, Joyce C. Ho, Debora Donato, Rakesh Agrawal, Yu Zheng, Carlos Castillo, Aixin Sun, Vincent S. Tseng, and Chenliang Li (Eds.). ACM, Singapore, 1419–1428. https://doi.org/10.1145/3132847.3132926

[16] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. 2018. STAMP: Short-Term Attention/Memory Priority Model for Session-based Recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, Yike Guo and Faisal Farooq (Eds.). ACM, UK, 1831–1839. https://doi.org/10.1145/3219819.3219950

[17] Zhiwei Liu, Yongjun Chen, Jia Li, Philip S Yu, Julian McAuley, and Caiming Xiong. 2021. Contrastive self-supervised sequential recommendation with robust augmentation. *ArXiv preprint* abs/2108.06479 (2021). https://arxiv.org/abs/2108.06479

[18] Jianxin Ma, Chang Zhou, Hongxia Yang, Peng Cui, Xin Wang, and Wenwu Zhu. 2020. Disentangled Self-Supervision in Sequential Recommenders. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash (Eds.). ACM, 483–491. https://dl.acm.org/doi/10.1145/3394486.3403091

[19] Zaiqiao Meng, Richard McCreadie, Craig Macdonald, Iadh Ounis, Siwei Liu, Yaxiong Wu, Xi Wang, Shangsong Liang, Yucheng Liang, Guangtao Zeng, Junhua Liang, and Qiang Zhang. 2020. BETA-Rec: Build, Evaluate and Tune Automated Recommender Systems. In *Proceedings of the 14th ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 588–590. https://doi.org/10.1145/3383313.3411524

[20] Ping Nie, Yujie Lu, Shengyu Zhang, Ming Zhao, Ruobing Xie, William Yang Wang, and Yi Ren. 2022. MIC: Model-Agnostic Integrated Cross-Channel Recommender. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management* (Atlanta, GA, USA) *(CIKM '22)*. Association for Computing Machinery, New York, NY, USA, 3400–3409. https://doi.org/10.1145/3511808.3557081

[21] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *ArXiv preprint* abs/1807.03748 (2018). https://arxiv.org/abs/1807.03748

[22] Ruihong Qiu, Zi Huang, Hongzhi Yin, and Zijian Wang. 2022. Contrastive Learning for Representation Degeneration Problem in Sequential Recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining* (Virtual Event, AZ, USA) *(WSDM '22)*. Association for Computing Machinery, New York, NY, USA, 813–823. https://doi.org/10.1145/3488560.3498433

[23] Ruihong Qiu, Jingjing Li, Zi Huang, and Hongzhi Yin. 2019. Rethinking the Item Order in Session-based Recommendation with Graph Neural Networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, Wenwu Zhu, Dacheng Tao, Xueqi Cheng, Peng Cui, Elke A. Rundensteiner, David Carmel, Qi He, and Jeffrey Xu Yu (Eds.). ACM, Beijing, 579–588. https://doi.org/10.1145/3357384.3358010

[24] Hossein A. Rahmani, Mohammad Aliannejadi, Mitra Baratchi, and Fabio Crestani. 2020. Joint Geographical and Temporal Modeling based on Matrix Factorization for Point-of-Interest Recommendation. In *European Conference on Information Retrieval (ECIR)*. Springer, 205–219.

[25] Hossein A. Rahmani, Mohammad Aliannejadi, Rasoul Mirzaei Zadeh, Mitra Baratchi, Mohsen Afsharchi, and Fabio Crestani. 2019. Category-Aware Location Embedding for Point-of-Interest Recommendation. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval* (Santa Clara, CA, USA) *(ICTIR '19)*. Association for Computing Machinery, New York, NY, USA, 173–176. https://doi.org/10.1145/3341981.3344240

[26] Hossein A. Rahmani, Mohammadmehdi Naghiaei, Mahdi Dehghan, and Mohammad Aliannejadi. 2022. Experiments on Generalizability of User-Oriented Fairness in Recommender Systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 2755–2764. https://doi.org/10.1145/3477495.3531718

[27] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized Markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, Michael Rappa, Paul Jones, Juliana Freire, and Soumen Chakrabarti (Eds.). ACM, usa, 811–820. https://doi.org/10.1145/1772690.1772773

[28] Guy Shani, David Heckerman, and Ronen I Brafman. 2005. An MDP-based recommender system. *Journal of Machine Learning Research* 6, Sep (2005), 1265–1295.

[29] Zhengxiang Shi, Yue Feng, and Aldo Lipani. 2022. Learning to Execute Actions or Ask Clarification Questions. In *Findings of the Association for Computational Linguistics: NAACL 2022*. Association for Computational Linguistics, Seattle, United States, 2060–2070. https://doi.org/10.18653/v1/2022.findings-naacl.158

[30] Zhengxiang Shi and Aldo Lipani. 2023. Don't Stop Pretraining? Make Prompt-based Fine-tuning Powerful Learner. *arXiv preprint arXiv:2305.01711* (2023). https://arxiv.org/pdf/2305.01711.pdf

[31] Zhengxiang Shi, Pin Ni, Meihui Wang, To Eun Kim, and Aldo Lipani. 2022. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*. Bruges, Belgium.

[32] Zhengxiang Shi, Jerome Ramos, To Eun Kim, Xi Wang, Hossein A Rahmani, and Aldo Lipani. 2023. When and What to Ask Through World States and Text Instructions: IGLU NLP Challenge Solution. *arXiv preprint arXiv:2305.05754* (2023).

[33] Zhengxiang Shi, Francesco Tonolini, Nikolaos Aletras, Emine Yilmaz, Gabriella Kazai, and Yunlong Jiao. 2023. Rethinking Semi-supervised Learning with Language Models. In *Findings of ACL 2023*. Association for Computational Linguistics,

Toronto, Canada.

[34] Zhengxiang Shi, Qiang Zhang, and Aldo Lipani. 2022. StepGame: A New Benchmark for Robust Multi-Hop Spatial Reasoning in Texts. In *Association for the Advancement of Artificial Intelligence*. AAAI Press.

[35] Weiping Song, Zhiping Xiao, Yifan Wang, Laurent Charlin, Ming Zhang, and Jian Tang. 2019. Session-Based Social Recommendation via Dynamic Graph Attention Networks. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*, J. Shane Culpepper, Alistair Moffat, Paul N. Bennett, and Kristina Lerman (Eds.). ACM, 555–563. https://doi.org/10.1145/3289600.3290989

[36] 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, CIKM 2019, Beijing, China, November 3-7, 2019*, Wenwu Zhu, Dacheng Tao, Xueqi Cheng, Peng Cui, Elke A. Rundensteiner, David Carmel, Qi He, and Jeffrey Xu Yu (Eds.). ACM, Beijing, 1441–1450. https://doi.org/10.1145/3357384.3357895

[37] Lei Wang, Ee-Peng Lim, Zhiwei Liu, and Tianxiang Zhao. 2022. Explanation Guided Contrastive Learning for Sequential Recommendation. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management* (Atlanta, GA, USA) *(CIKM '22)*. Association for Computing Machinery, New York, NY, USA, 2017–2027. https://doi.org/10.1145/3511808.3557317

[38] Tongzhou Wang and Phillip Isola. 2020. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 9929–9939. http://proceedings.mlr.press/v119/wang20k.html

[39] Wen Wang, Wei Zhang, Shukai Liu, Qi Liu, Bo Zhang, Leyu Lin, and Hongyuan Zha. 2020. Beyond Clicks: Modeling Multi-Relational Item Graph for Session-Based Target Behavior Prediction. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen (Eds.). ACM / IW3C2, Taiwan, 3056–3062. https://doi.org/10.1145/3366423.3380077

[40] Xi Wang, Iadh Ounis, and Craig Macdonald. 2019. Comparison of Sentiment Analysis and User Ratings in Venue Recommendation. In *Advances in Information Retrieval*, Leif Azzopardi, Benno Stein, Norbert Fuhr, Philipp Mayr, Claudia Hauff, and Djoerd Hiemstra (Eds.). Springer International Publishing, Cham, 215–228.

[41] Xi Wang, Iadh Ounis, and Craig Macdonald. 2020. Negative Confidence-Aware Weakly Supervised Binary Classification for Effective Review Helpfulness Classification. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 1565–1574. https://doi.org/10.1145/3340531.3411978

[42] Xi Wang, Iadh Ounis, and Craig Macdonald. 2021. Leveraging Review Properties for Effective Recommendation. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 2209–2219. https://doi.org/10.1145/3442381.3450038

[43] Xi Wang, Iadh Ounis, and Craig Macdonald. 2022. BanditProp: Bandit Selection of Review Properties for Effective Recommendation. *ACM Trans. Web* 16, 4, Article 20 (nov 2022), 19 pages. https://doi.org/10.1145/3532859

[44] Ziyang Wang, Wei Wei, Gao Cong, Xiao-Li Li, Xianling Mao, and Minghui Qiu. 2020. Global Context Enhanced Graph Neural Networks for Session-based Recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, Jimmy Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu (Eds.). ACM, Virtual, 169–178. https://doi.org/10.1145/3397271.3401142

[45] Yinwei Wei, Xiang Wang, Qi Li, Liqiang Nie, Yan Li, Xuanping Li, and Tat-Seng Chua. 2021. Contrastive Learning for Cold-Start Recommendation. In *Proceedings of the 29th ACM International Conference on Multimedia* (Virtual Event, China) *(MM '21)*. Association for Computing Machinery, New York, NY, USA, 5382–5390. https://doi.org/10.1145/3474085.3475665

[46] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. 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, January 27 - February 1, 2019*. AAAI Press, 346–353. https://doi.org/10.1609/aaai.v33i01.3301346

[47] 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 and Knowledge Management*. ACM, Virtual Event Queensland Australia, 2180–2190. https://doi.org/10.1145/3459637.3482388

[48] Xin Xia, Hongzhi Yin, Junliang Yu, Qinyong Wang, Lizhen Cui, and Xiangliang Zhang. 2021. Self-Supervised Hypergraph Convolutional Networks for Session-based Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. AAAI, Virtual, 4503–4511. https://doi.org/10.1609/aaai.v35i5.16578

[49] Ruobing Xie, Zhijie Qiu, Bo Zhang, and Leyu Lin. 2022. Multi-granularity Item-based Contrastive Recommendation. *arXiv preprint arXiv:2207.01387* (2022).

[50] Xu Xie, Fei Sun, Zhaoyang Liu, Jinyang Gao, Bolin Ding, and Bin Cui. 2020. Contrastive Pre-training for Sequential Recommendation. *ArXiv preprint* abs/2010.14395 (2020). https://arxiv.org/abs/2010.14395

[51] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive learning for sequential recommendation. In *2022 IEEE 38th international conference on data engineering (ICDE)*. IEEE, IEEE, virtual, 1259–1273. https://ieeexplore.ieee.org/abstract/document/9835621

[52] Lee Xiong, Chenyan Xiong, Ye Li, Kwok-Fung Tang, Jialin Liu, Paul N. Bennett, Junaid Ahmed, and Arnold Overwijk. 2021. Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval. In *9th International Conference on Learning Representations, ICLR 2021, May 3-7, 2021*. OpenReview.net, Virtual Event, Austria. https://openreview.net/forum?id=zeFrfgyZln

[53] Chengfeng Xu, Pengpeng Zhao, Yanchi Liu, Victor S. Sheng, Jiajie Xu, Fuzhen Zhuang, Junhua Fang, and Xiaofang Zhou. 2019. Graph Contextualized Self-Attention Network for Session-based Recommendation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, Sarit Kraus (Ed.). ijcai.org, 3940–3946. https://doi.org/10.24963/ijcai.2019/547

[54] Tiansheng Yao, Xinyang Yi, Derek Zhiyuan Cheng, Felix Yu, Ting Chen, Aditya Menon, Lichan Hong, Ed H Chi, Steve Tjoa, Jieqi Kang, et al. 2020. Self-supervised learning for deep models in recommendations. *arXiv e-prints* (2020), arXiv–2007.

[55] Hongzhi Yin and Bin Cui. 2016. *Spatio-temporal recommendation in social media*. Springer.

[56] Junliang Yu, Hongzhi Yin, Min Gao, Xin Xia, Xiangliang Zhang, and Nguyen Quoc Viet Hung. 2021. Socially-Aware Self-Supervised Tri-Training for Recommendation. *ArXiv preprint* abs/2106.03569 (2021). https://arxiv.org/abs/2106.03569

[57] Junliang Yu, Hongzhi Yin, Jundong Li, Qinyong Wang, Nguyen Quoc Viet Hung, and Xiangliang Zhang. 2021. Self-Supervised Multi-Channel Hypergraph Convolutional Network for Social Recommendation. *ArXiv preprint* abs/2101.06448 (2021). https://arxiv.org/abs/2101.06448

[58] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. 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* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 1294–1303. https://doi.org/10.1145/3477495.3531937

[59] Yuyu Zhang, Hanjun Dai, Chang Xu, Jun Feng, Taifeng Wang, Jiang Bian, Bin Wang, and Tie-Yan Liu. 2014. Sequential Click Prediction for Sponsored Search with Recurrent Neural Networks. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, Carla E. Brodley and Peter Stone (Eds.). AAAI Press, 1369–1375. http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8529

[60] 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 *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, October 19-23, 2020*, Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux (Eds.). ACM, Virtual Event, Ireland, 1893–1902. https://doi.org/10.1145/3340531.3411954

[61] Xin Zhou, Aixin Sun, Yong Liu, Jie Zhang, and Chunyan Miao. 2023. SelfCF: A Simple Framework for Self-Supervised Collaborative Filtering. *ACM Trans. Recomm. Syst.* (apr 2023). https://doi.org/10.1145/3591469 Just Accepted.

[62] Andrew Zimdars, David Maxwell Chickering, and Christopher Meek. 2013. Using temporal data for making recommendations. *arXiv preprint arXiv:1301.2320* (2013).