An Anchor-Semantics-Aware Neural Preference Propagation Model for Session-based Recommendation

Yaxuan Ye
College of Computer Science
Sichuan University
Chengdu, China
yyaxuan999@gmail.com

Ning Yang*
College of Computer Science
Sichuan University
Chengdu, China
yangning@scu.edu.cn

2023 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-8867-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IJCNN54540.2023.10191659

***Abstract*—Session-based recommendation aims to predict user's next action based on the anonymous sessions. Recent studies mainly apply Graph Neural Networks (GNN) to model the complex item-transitions and transfer the collaborative signals among sessions. However, most existing methods ignore the semantic inconsistency problem cause by item functional diversity and popular item during the process of cross-session collaborative signal capturing, which may lead to a preference negative transfer between irrelevant sessions. In addition, current works neglect to retain the semantic discriminability when learning the representations of sessions and items, and consequently degenerate the personalized recommendation. In this paper, we propose a novel model, called Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP), to simulate the transfer of collaborative signals with an anchor-semantics-aware recursive neural propagation over a designed graph. Specifically, we devise a Session Graph with Anchor Links (SGAL) based on all sessions, and then present an Anchor-Semantics Attention Network (ASAN) to transfer the semantically consistent cross-session collaborative signals and learn session-specific embeddings encoding different session semantics. Furthermore, we propose a Session Semantics Enhancement (SSE) module to improve the semantic discriminability of the learned representation via an elaborate self-supervised learning task. Extensive experiments on real-world datasets demonstrate the effectiveness and superiority of ASA-NPP over the state-of-the-art methods.**

I. Introduction

In recent years, session-based recommendation has been attracting increasing attention due to its important role in recommender systems that contain multitudes of anonymous user-item interaction data [1]–[3]. Session-based recommendation aims to predict next item that a user will interact with in her ongoing session, where a session is a list of user-item interactions that happen together in a certain period of time, e.g., a basket of products purchased in one transaction or a set of songs listened to in one hour [3].

Due to the highly practical value, a variety of approaches have been proposed for session-based recommendation, where the key challenge is how to infer a user's short-term preference carried in her current session by exploiting the observed historical anonymous sessions of all users. Most of early studies employ Markov chain based models [4], [5], which assume that a user's next action is only dependent on her previous one. Such an oversimplified assumption may impair the recommendation performance due to its inability to capture long-range dependences which commonly exist in most scenarios. Recently, inspired by the impressive power of deep learning, many neural models have been proposed for session-based recommendation, some of which apply Recurrent Neural Network (RNN) to capture the nonlinear dynamics of sequential interactions [6]–[8], while some others model the sequential dependence based on Convolution Neural Network (CNN) [9], [10] or attention mechanism [11], [12]. However, these abovementioned models only focus on the item-transitions within a single session, but fail to capture the cross-session transitions. To address this issue, researchers propose a large number of Graph Neural Network (GNN) based models [1], [2], [13]–[15], which model pair-wise relations of items embedded in session data with a graph and show a promising potential in capturing high-order collaborative signals among sessions. Although a significant progress has been made, these GNN-based methods still have following two challenges.

- **Semantic Consistency** In order to infer the preference of a target session, the existing GNN-based methods often transfer the collaborative signals from other sessions via *anchor items* (i.e., the items shared between sessions) and learn the preference representation of the target session by combining those of its neighbor sessions connected via the anchor items, which underlyingly assumes that the anchor items cause similar preferences of the sessions connected by them [2], [14], [16], [17]. This assumption, however, might become invalid in real world because it overlooks the problem of an anchor item's inconsistent semantics in different sessions. Intuitively, due to the functional diversity, an anchor item can play different roles in the sessions of different purposes, which implies that two sessions sharing items do not necessarily share similar preferences. For example, in the left part of Fig.

* Corresponding author.

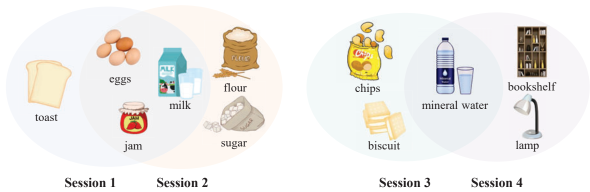


Fig. 1. Motivation example.

1, the milk, eggs, and jam are the anchor items shared between Session 1 and Session 2, but serve different purposes. In Session 1, they are parts of a breakfast, while in Session 2, they serve as the ingredients of making cakes. In such case, the existing methods will fail to infer the preference of Session 1 based on that of Session 2 and vice versa, since contrary to the assumption, the anchor item with inconsistent semantics in the two sessions carries no cross-session collaborative signals. In a worse case, the anchor item is a popular item that frequently appears in many sessions, resulting in false connections between irrelevant sessions. As shown in the right part of Fig. 1, the mineral water (anchor item) appears in Session 3 for snacks, while in Session 4 for buying furniture. Again due to the inconsistent semantics of the anchor item, the existing methods will conduct a negative transfer of preference between the two irrelevant sessions.

- **Semantic Discriminability** While the existing GNN-based models enjoy the benefit offered by the propagation of high-order collaborative signals across sessions, they likely suffer from the problem of over-smoothing incurred by the information propagation between high-order neighbors [18], which will hurt the semantic discriminability of the learned preference representations of sessions and items, and consequently degenerate the personalized recommendation. For example, as shown in Fig. 1, the anchor items account for the majority of both Session 1 and Session 2, but the two sessions have different intents that are mainly determined by their minority items, the toast in Session 1 and the flour and sugar in Session 2, respectively. Reasonably, their learned preference representations are supposed to be able to discriminate their different semantics. However, after the iterative information propagation of session to item and item to session, the preference transfer via the anchor items between Session 1 and Session 2 might cause their preference representations to improperly converge to indistinguishable vectors with little discriminant information of session preference.

To solve the aforementioned challenges, we propose a novel Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP) model for session-based recommendation. The **main idea** of ASA-NPP is to *simulate the transfer of the cross-session collaborative signals with an anchor-semantics-aware recursive neural propagation over a designed graph*. To address the **semantic consistency** challenge, we propose a Session Graph with Anchor Links (SGAL) to model the session data, where a node represents an appearance of an item in a specific session, an edge exists between two nodes if they are two item appearances in the same session or two different appearances of the same item (anchor item) in different sessions, and a session is a maximal clique consisting of the appearances of different items. With the help of SGAL, we can learn embeddings for items' appearances in sessions, not just items, due to which an anchor item appearing in multiple sessions can possess multiple session-specific embeddings encoding different session semantics. In particular, for the transfer of the semantics-consistent cross-session collaborative signals, we devise an Anchor-Semantics Attention Network (ASAN) to realize the recursive neural propagation over SGAL, which can learn an anchor item's session-specific embedding in a session with an attentional aggregation of its session-specific embeddings in other sessions with consistent semantics, while blocking those with inconsistent semantics. The neural propagation conducted on SGAL brings the benefit that the consistency of an anchor item's semantics in different sessions can be identified via its anchor links (the edges connecting two different appearances of the same item). Furthermore, to overcome the **semantic discriminability** challenge, we propose a Session Semantics Enhancement (SSE) module, by which ASA-NPP can improve a session representation by preserving the discriminative information carried by the item that could reflect personalized needs of the session user via an elaborate contrastive learning task during training. The main contributions of this work can be summarized as follows:

- We propose a novel Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP) model for session-based recommendation, which models the session data with an elaborately designed Session Graph with Anchor Links (SGAL), and can capture the cross-session collaborative signals with a recursive neural preference propagation over SGAL.
- We propose an Anchor-Semantics Attention Network (ASAN) for ASA-NPP to realize the neural propagation over SGAL. ASAN can identify the consistency of an anchor item's semantics in different sessions, which makes ASA-NPP to prefer the transfer of semantics-consistent collaborative signals across sessions while restrain the transfer of semantics-inconsistent ones.
- To improve the semantic discriminability of a session representation, we propose a Session Semantics Enhancement (SSE) module, which can preserve the discriminative information conveyed by the item that could reflect personalized needs of the session user via an elaborate contrastive learning task.
- The extensive experiments conducted on real-world datasets demonstrate the superiority of the proposed model over state-of-the-art models on session-based recommendation.

The rest of this paper is organized as follows. We introduce

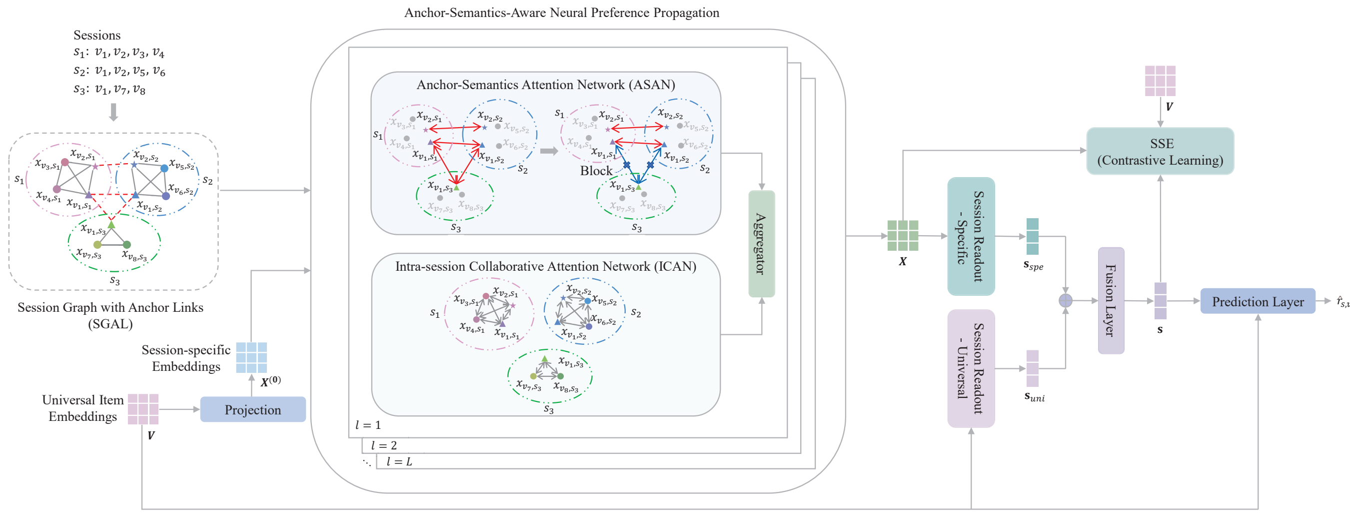


Fig. 2. The overview of the proposed model ASA-NPP.

the preliminaries in Section II. In Section III, we first give an overview of ASA-NPP and then describe its details. We empirically evaluate the performance of ASA-NPP over real-world datasets in Section IV. At last, we briefly review the related works in Section V and conclude in Section VI.

II. PRELIMINARY

A. Notations and Definitions

Let $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ denote the set of N unique items involved in all sessions. $s = \{v_{1,s}, v_{2,s}, \ldots, v_{|s|,s}\}$ represents an anonymous session s which is a subset of $\mathcal{V}$, $v_{i,s} \in \mathcal{V}$ represents an interactive item v_i of the user within the session s, and the length of session s is $|s|$. The whole session set is defined as $\mathcal{S} = \{s_1, s_2, \ldots, s_M\}$, M is the number of sessions in the system. Each item v and each session s can be described with an embedding vector $\mathbf{v} \in \mathbb{R}^d$ and $\mathbf{s} \in \mathbb{R}^d$, respectively, and the representation of the whole item set is denoted as $\mathbf{V} \in \mathbb{R}^{N\times d}$.

B. Problem Statement

Given a session s, the aim of our model is to predict the next interactive item $v_{|s|+1,s}$ of the current session s based on the known sessions $\mathcal{S}$. This problem can be formulated as a top-K recommendation problem that the item v with a top-K probability $\hat{r}_{s,v}$ will be selected to the candidate set and recommended to the user of the session s.

III. PROPOSED MODEL

A. Overview

Fig. 2 shows the overview of ASA-NPP, which mainly comprises four components: 1) Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP) layer. It learns session-specific embeddings for items' appearances by employing an Anchor-Semantics Attention Network (ASAN) and an Intra-session Collaborative Attention Network (ICAN) to capture the collaborative signals across and within sessions correspondingly based on the Session Graph with Anchor Links (SGAL). 2) session representation generation. It leverages the attention mechanism to generate the session embeddings based on universal item embeddings and session-specific embeddings, respectively, and then fusing them to obtain the final session representation. 3) Session Semantics Enhancement (SSE) module. It improves the semantic discriminability of session representation by maximizing the mutual information between representations of session and the most unique item in the session. 4) prediction layer. It outputs the predicted probability of items for recommendation.

For a given anonymous session s, the specific workflow is as follows: Firstly, each item $v_{i,s}$ contained in the session s is mapped to a node $x_{v,s}$ which represents the item's appearance specific to that session, then an Anchor-Semantics-Aware Neural Preference Propagation is carried out on the Session Graph with Anchor Links (SGAL) to learn the session-specific embedding $\mathbf{x}_{v,s}$. After that, attention mechanisms and a linear transformation are exploited to obtain session representation $\mathbf{s}$ based on both universal item embeddings and session-specific embeddings. Next, the mutual information between representations of session and the most unique item in the session is maximized so that the discrimination of the final session representation $\mathbf{s}$ is improved. Finally, the recommendation scores of items are generated.

B. Converting Sessions to Session Graph with Anchor Links

In order to model the diverse semantics of items across sessions and preserve the personalization needs of session user, we develop a Session Graph with Anchor Links (SGAL) to depict the items' appearances in sessions. Inspired by [19], we construct a Session Graph with Anchor Links (SGAL) $\mathcal{G} = (\mathcal{V}_x, \mathcal{E}_x)$ based on all sessions, where $\mathcal{V}_x$ is the node set consisting of all items' appearances in sessions, and two nodes are connected if they are two item appearances in the same session or two different appearances of the same item (anchor item) in different sessions. Each session is modeled as a maximal clique consisting of the appearances of different items. Formally, we denote an appearance of item v in a specific session s as $x_{v,s}$, where $v \in \mathcal{V}$, $s \in \mathcal{S}$. $\mathcal{G} = (\mathcal{V}_x, \mathcal{E}_x)$, where $\mathcal{V}_x = \{x_{v,s} \mid v \in \mathcal{V} \ and \ s \in \mathcal{S}\}$ and $\mathcal{E}_x = \{(x_{v,s}, x_{v',s'}) \mid v = v' \ or \ s = s'\}$. As illustrated in the left part of Fig. 2, the constructed Session Graph with Anchor Links (SGAL) contains two types of edges, the solid lines depict the co-occurrence of items inside a session, while the dashed lines, dubbed anchor links, present the cross-session transitions. The anchor link can be considered as the channel for the exchange of cross-session collaborative signals.

C. Anchor-Semantics-Aware Neural Preference Propagation

After transforming the all sessions to Session Graph with Anchor Links (SGAL), we innovatively devise a new method to simulate the transfer of collaborative signals, called Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP), to learn the session-specific embeddings for items' appearances based on the constructed graph. As presented in Fig. 2, the Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP) layer consists of three modules: an Anchor-Semantics Attention Network (ASAN), an Intra-session Collaborative Attention Network (ICAN), and an aggregator. The ASAN is responsible for capturing the cross-session collaborative signals, while the ICAN takes charge of intra-session collaborative signals collection. The aggregator fuses the two types of collaborative signals and refines the session-specific embeddings. Technically, the item embeddings $\mathbf{V}$, which is universal for all sessions, is mapped into session-specific embeddings for items' appearances by a projection matrix $\mathbf{P} \in \{0,1\}^{N' \times N}$, where N' is the number of all items' appearances and N is the item number. The element of the projection matrix $\mathbf{P}$ represents the relationship between the item and item's appearance that if they are two different appearances of the same item (anchor item) in different sessions, then the element is set to one, otherwise zero. We obtain the initial session-specific embeddings $\mathbf{X}^{(0)} \in \mathbb{R}^{N' \times d}$ after the following projection:

$$\mathbf{X}^{(0)} = \mathbf{P}\mathbf{V} \tag{1}$$

With the Session Graph with Anchor Links (SGAL), the Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP) learns the session-specific embeddings by incorporating cross-session collaborative signals with consistent semantics, as well as intra-session collaborative signals. Note that when capturing the collaborative signals, we take into account the information of the item's appearance itself, just like adding the self-loop, which is generally used in the graph-based model to prevent the loss of self-information.

1) Cross-session Collaborative Signal Capturing: To overcome the semantic inconsistency problem during the process of capturing collaborative signals across sessions, we propose a novel Anchor-Semantics Attention Network (ASAN), which is capable of aggregating semantics-consistent collaborative

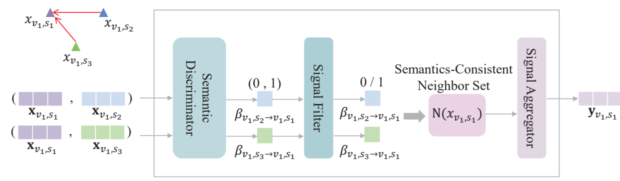


Fig. 3. Anchor-Semantics Attention Network (ASAN).

signals to learn more accurate session-specific embeddings. As mentioned before, two sessions sharing items do not necessarily share similar preferences, and the semantics of anchor item in different sessions are not always consistent that they may be irrelevant or even opposite, which may lead to a negative transfer of preference. To tackle this problem, we particularly focus on the anchor links since the cross-session collaborative signals are delivered through the anchor links. Instead of directly aggregating collaborative signals from the sessions sharing items, we design a semantic discriminator and a signal filter to select the truly relevant collaborative signals as shown in Fig. 3. The semantic discriminator computes the semantic consistency score between anchor items through Euclidean distance. For a pair of anchor items $(x_{v,s}, x_{v,s'})$, their semantic consistency score $\beta^{(l)}_{v,s' \to v,s}$ at l-th layer is computed by:

$$\beta^{(l)}_{v,s' \to v,s} = \exp(-\mathrm{d_E}(x^{(l-1)}_{v,s}, x^{(l-1)}_{v,s'})) \tag{2}$$

where $\mathrm{d_E} = (\cdot)$ represents the Euclidean distance between $x_{v,s}$ and $x_{v,s'}$. Then, we adopt a threshold a to filter the semantics-inconsistent collaborative signals and preserve the really relevant one. For example, the semantic consistency score $\beta^{(l)}_{v,s' \to v,s}$ is set to 0 when $\beta^{(l)}_{v,s' \to v,s} < 0.5$, otherwise one. $\beta^{(l)}_{v,s' \to v,s} = 0$ indicates that the anchor link is broken so as the semantics-inconsistent collaborative signals cannot be transmitted. Formally, the signal filter can be defined as:

$$\beta^{(l)}_{v,s' \to v,s} = \begin{cases} 0 & \text{if } \beta^{(l)}_{v,s' \to v,s} < a \\ 1 & \text{if } \beta^{(l)}_{v,s' \to v,s} \geq a \end{cases} \tag{3}$$

After that, we select the anchor items whose corresponding anchor links are retained to form the semantics-consistent neighbor set of $x_{v,s}$ denoted as $\mathcal{N}^{(l)}(x_{v,s}) = \left\{x_{v,s'} \mid \beta^{(l)}_{v,s' \to v,s} > 0\right\}$. Subsequently, we utilize self-attention mechanism to further distinguish the importance of the cross-session collaborative signals from each semantics-consistent neighbor and aggregate them to obtain embedding $\mathbf{y}^{(l)}_{v,s}$. Inspired by [20], we do not use the feature transformation since each session-specific embedding here only specifies to one item appearance and we empirically find that including the feature transformation not only increases the difficulty of training but also degrades the recommendation performance. Thus, we implement the signal aggregator as follows:

$$\mathbf{y}_{v,s}^{(l)} = \sum_{x_{v,s'} \in \mathcal{N}^{(l)}(x_{v,s})} \alpha_{v,s' \to v,s}^{(l)} \mathbf{x}_{v,s'}^{(l-1)}$$
$$\alpha_{v,s' \to v,s}^{(l)} = \mathrm{softmax}(\frac{(\mathbf{x}_{v,s'}^{(l-1)})^{\mathrm{T}} \mathbf{x}_{v,s}^{(l-1)}}{\sqrt{d}}) \tag{4}$$

where d is the dimension size. Different from directly using the attention mechanism, we filter out the irrelevant collaborative signals rather than just softly weaken their adverse effects to make the learned representation more accurate. It is necessary to remove the semantics-inconsistent collaborative signals for learning high-quality embeddings, especially when most of the collaborative signals transmitted are semantically irrelevant that including them may be misleading.

2) Intra-session Collaborative Signal Capturing: To incorporate the intra-session collaborative signals, we devise an Intra-session Collaborative Attention Network (ICAN). Since the strengths of the collaborative signals between co-occurring items in the session might be different, we exploit the attention mechanism to pay various attention on different items and highlight the informative collaborative signals. Similarly, we do not utilize the feature transformation here. For each item's appearance $x_{v,s}$, we aggregate the session-specific embeddings of co-occurring items within session s to generate the embedding $\mathbf{z}_{v,s}^{(l)}$ as follows:

$$\mathbf{z}_{v,s}^{(l)} = \sum_{x_{v',s} \in s} \gamma_{v',s \to v,s}^{(l)} \mathbf{x}_{v',s}^{(l-1)}$$
$$\gamma_{v',s \to v,s}^{(l)} = \mathrm{softmax}(\frac{(\mathbf{x}_{v',s}^{(l-1)})^{\mathrm{T}} \mathbf{x}_{v,s}^{(l-1)}}{\sqrt{d}}) \tag{5}$$

where d is the dimension size, and $\gamma_{v',s \to v,s}^{(l)}$ indicates the signal strength at l-th layer.

3) Session-specific Embedding Generation: Once both kinds of collaborative signals are propagated, we update the representation of each item appearance by integrating them. The aggregator is implemented as follows:

$$\mathbf{x}_{v,s}^{(l)} = \mathbf{W}_x (\mathbf{y}_{v,s}^{(l)} \parallel \mathbf{z}_{v,s}^{(l)}) \tag{6}$$

where $\mathbf{W}_x \in \mathbb{R}^{d \times 2d}$ is the learnable transformation matrix, and $\parallel$ denotes concatenation.

While a single Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP) layer is capable of capturing collaborative signals that propagated via direct link, we stack multiple layers to obtain multi-hop high-order collaborative signals. After the anchor-semantics-aware recursive neural propagation of all layers (L layers) is finished, we generate the final session-specific embeddings by averaging the session-specific embeddings obtained at each layer, which is defined as $\mathbf{X} = \frac{1}{L+1} \sum_{l=0}^{L} \mathbf{X}^{(l)}$.

D. Session Representation Generation

As mentioned earlier, in order to better describe the items and their dynamically changing semantics, we introduce two kinds of embeddings for items in ASA-NPP model including universal item embeddings and session-specific embeddings. The universal item embeddings are commonly used by all sessions, while the session-specific embeddings are tailored for items' appearances that are capable of depicting various item semantics in different sessions. While the semantics of item is changing, there are some characteristics of an item may be retained and shared by all sessions [21], such as the low-fat characteristic of milk, the less-sugar characteristic of jam, etc. Hence, we take both universal item embeddings and session-specific embeddings into consideration during session representation generation. Specifically, we adopt attention mechanism, which is able to automatically emphasize the informative members' embeddings in the current session, to generate two kinds of session embeddings corresponding to universal item embeddings and session-specific embeddings, respectively, and then fusing them to obtain the final session representation. For a session s, we aggregate the universal item embeddings of items contained in that session to obtain the session embedding $\mathbf{s}_{uni}$ as follows:

$$\mathbf{s}_{uni} = \sum_{v \in s} \eta_{vs} \mathbf{W}_{uni}^{v} \mathbf{v}$$
$$\eta_{vs} = \mathrm{softmax}(\frac{(\mathbf{W}_{uni}^{k} \mathbf{v})^{\mathrm{T}} \mathbf{W}_{uni}^{q}}{\sqrt{d}}) \tag{7}$$

where $\mathbf{W}_{uni}^{q} \in \mathbb{R}^{d}$ represents a trainable query vector, $\mathbf{W}_{uni}^{k} \in \mathbb{R}^{d \times d}$ and $\mathbf{W}_{uni}^{v} \in \mathbb{R}^{d \times d}$ are the trainable transformation matrices used to generate key vector and value vector, correspondingly. Meanwhile, we get the session-specific embedding based session embedding $\mathbf{s}_{spe}$ with:

$$\mathbf{s}_{spe} = \sum_{x_{v,s} \in s} \mu_{xs} \mathbf{x}_{v,s}$$
$$\mu_{xs} = \mathrm{softmax}(\frac{(\mathbf{x}_{v,s})^{\mathrm{T}} \mathbf{W}_{spe}^{q}}{\sqrt{d}}) \tag{8}$$

where $\mathbf{W}_{spe}^{q} \in \mathbb{R}^{d}$ is the trainable query vector. The feature transformation of session-specific embeddings is omitted with the same reason as before. According to observations in previous studies [1], [2], [22], user's behaviors in the short-term sequence may not have strict chronological order and the order of the interactive items is less likely to be related to the user's interests. Therefore, we do not take account of the order information in session-based recommendation scenario, and omit the position embedding while utilizing attention mechanism. Afterwards, we generate the final session representation $\mathbf{s}$ by taking linear transformation over the concatenation of $\mathbf{s}_{uni}$ and $\mathbf{s}_{spe}$:

$$\mathbf{s} = \mathbf{W}_s (\mathbf{s}_{uni} \parallel \mathbf{s}_{spe}) \tag{9}$$

where $\mathbf{W}_s \in \mathbb{R}^{d \times 2d}$ is a learnable matrix parameter.

E. Session Semantics Enhancement Module

In order to improve the semantic discriminability and learn more discriminative session representation, we present a contrastive learning task to encourage the learned session representation to be mindful of the item that could reflect personalized needs of the session user by utilizing the mutual information maximization (MIM) principle. For a session s, we treat the item with the lowest popularity in the session as the positive sample and make the session representation $\mathbf{s}$ and the session-specific embedding of positive sample $\mathbf{x}^+$ closer. On the other hand, we randomly select a popular item (the number of sessions appeared ranked in the top 20%) that the session user has not interacted with as a negative sample and denoted its corresponding universal item embedding as $\mathbf{v}^-$. The objective function of SSE module can be defined as a standard binary cross-entropy (BCE) loss:

$$\mathcal{L}_{MI} = -\sum_{s\in\mathcal{S}_{train}} \left(\log\sigma(D(\mathbf{s},\mathbf{x}^+) + \log\sigma(1 - D(\mathbf{s},\mathbf{v}^-)\right) \tag{10}$$

where $D(\cdot) : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ is the discriminator function that takes two representations as input and then scores the agreement between them. Here, we simply implement the discriminator as dot product.

F. Prediction and Optimization

With the learned final session representation, we are able to make recommendations by computing the preference scores of session s on all of N items in the system. For each item $v \in \mathcal{V}$, we first compute the score by inner product:

$$\tilde{r}_{s,v} = \mathbf{s}^{\mathrm{T}}\mathbf{v} \tag{11}$$

Then, the predicted probability of the next item being item v, $\hat{r}_{s,v}$, can be defined as:

$$\hat{r}_{s,v} = \frac{\exp(\tilde{r}_{s,v})}{\sum_{v'\in\mathcal{V}}\exp(\tilde{r}_{s,v'})} \tag{12}$$

We select the items with top-k probabilities as candidate items and recommend to the user of session s.

Let $r_{s,v}$ denotes the ground truth, we formulate the learning objective of recommendation task as cross entropy loss function:

$$\mathcal{L}_{Rec} = -\sum_{s\in\mathcal{S}_{train}}\sum_{v=1}^{N} r_{s,v}\log\hat{r}_{s,v} \tag{13}$$

Finally, we combine the recommendation task and the contrastive learning task and jointly optimize them in an end-to-end fashion. The overall objective function of ASA-NPP model is defined as follows:

$$\mathcal{L} = \mathcal{L}_{Rec} + \lambda_1\mathcal{L}_{MI} + \lambda_2\left\|\theta\right\|_2^2 \tag{14}$$

where λ_1 is a hyper-parameter used to control the magnitude of contrastive learning task, and λ_2 is the coefficient of the L2 regularization on θ which is a set of the trainable parameters of the proposed model.

TABLE I
STATISTICS OF THE USED DATASETS.

Dataset	# of clicks	# of sessions	# of items	average length
Diginetica	874,627	189,734	37,242	4.61
Nowplaying	1,699,693	369,889	59,647	4.60

IV. EXPERIMENTS

A. Experimental Setup

1) Datasets: We conduct experiments on real-word datasets: Diginetica[1] and Nowplaying[2]. Diginetica is a dataset that comes from CIKM Cup 2016, which includes typical transaction data extracted from e-commerce search engine logs. Nowplaying is a dataset, which consists of music listening event of users.

Similar to the previous works [6], [22], [23], we first perform data preprocessing on the datasets. Specifically, for Nowplaying, we treat the user's behaviors in a single day as a session. For the both datasets, we remove sessions with more than 20 or less than 2 items and filter out items appearing less than 5 sessions. Furthermore, we sort the sessions in chronological order and divide them into three parts, training data, validation data and test data that account for 80%, 10% and 10% respectively. Then, we filter out the items that do not appear in the training set, and take the last item of the session as the label for that session. The statistics of the datasets after preprocessing are presented in Table I.

2) Baselines: To evaluate the performance of the proposed ASA-NPP model, we compare it with traditional methods as well as state-of-the-art deep learning methods.

- **POP** is a frequency based method, which recommends top-K frequent items of the training set.
- **Item-KNN** [24] is a neighborhood based method that computes the cosine similarity between items and recommends items similar to the previous items.
- **FPMC** [5] is a sequential method that combines the matrix factorization with Markov Chain.
- **GRU4Rec** [6] is an RNN-based model that utilizes the Gated Recurrent Unit (GRU) to model user sequences.
- **NARM** [8] exploits GRU with attention mechanism to capture user's sequential behavior and main purpose.
- **STAMP** [25] employs attention memory network to capture user's general interest and current interest.
- **SR-GNN** [1] transforms sessions into directed unweighted graphs and applies gated graph neural networks to learn item transitions.
- **GCE-GNN** [14] devises session graph and global graph based on session data and employs session-aware attention mechanism and GNN to learn session-level item embeddings and global-level item embeddings respectively.

[1]https://competitions.codalab.org/competitions/11161
[2]https://dbis.uibk.ac.at/node/263#nowplaying

- **SHARE** [22] converts each session to a hypergraph and applies hypergraph attention networks to learn the session-wise item embeddings.
- **DHCN** [2] constructs hypergraph and line graph based on all sessions and exploits hypergraph convolutional network and graph convolutional network to capture the intra- and inter-session information.
- **FMLP-Rec** [26] is a recently proposed all-MLP model with learnable filters which can adaptively attenuate the noise information contained in user behavior data.

3) Evaluation Metrics: Following previous studies [1], [2], [14], [22], we adopt the commonly used Hit Ratio (Hit@K) and Mean Reciprocal Rank (MRR@K) as evaluation metrics.

4) Parameter Settings: Similar to [25], we apply extensive grid search to find the optimal hyper-parameters by using the validation set. The ranges of hyper-parameters for grid search are the following: $\{16, 32, 64, 96, 128, 136, 144\}$ for embedding dimension size d, $\{0, 0.2, 0.4, 0.5, 0.6, 0.8, 1.0\}$ for the threshold of semantics-inconsistent collaborative signal filtering a, $\{0, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.5\}$ for the magnitude of contrastive learning task λ_1, and $\{0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1\}$ for learning rate. Furthermore, the batch size and the L2 penalty is set to 64 and 10^{-5} respectively. For fair comparison, the hyper-parameters of the baselines are set to the optimal values reported in the original papers and then fine-tuned on validation sets to ensure the best performance of baseline models. Note that the ASA-NPP is a graph-based transductive learning model. We need to input the entire graph including unlabeled validation data and test data, hence the complexity of ASA-NPP is positively related to the total number of user-item interactions contained in the dataset.

B. Performance Comparison

The experimental results of all methods are reported in Table II, from which we can draw the following observations.

The performance of traditional methods (i.e., POP, ItemKNN and FPMC) are not competitive in most cases, since they make the recommendation solely based on the frequency, similarity or adjacent items' transition, without considering the complex transitions across and within sessions.

Most of the neural network based methods achieve better performance compared with traditional methods, proving the superiority of deep neural network in modeling complex dependencies embedded in session data. Generally, the RNN-based methods (i.e., GRU4Rec and NARM) less competitive than GNN-based methods, because they only consider the sequential transitions of adjacent items within a session and ignore that over all sessions. STAMP yields competitive performance as it applies the attention mechanism and is more sensitive to user's true interests. FMLP-Rec, an all-MLP model, also achieves impressive results, since it applies learnable filters which can attenuate the effects of noise information contained in session. Both STAMP and FMLP-Rec demonstrate the importance of emphasizing the informative items in a session.

GNN-based methods (i.e., SR-GNN, GCE-GNN, SHARE and DHCN) show promising results and generally outperform other methods, especially on Nowplaying. This can be attributed to the great ability of GNN to model the complex item transitions. On the whole, the DHCN is the strongest baseline, which indicates the necessity of capturing both intra- and inter-session information. Although GCE-GNN exploits the item transitions over all sessions, it does not yield competitive performance. One possible reason is that GCE-GNN only models a part of item's global-level item transitions. In some cases, SHARE achieves competitive performance, despite it ignores the fertile cross-session collaborative signals, suggesting the importance of considering item's dynamic semantics.

The proposed ASA-NPP consistently outperforms all baselines under each of metrics on both datasets, which demonstrates the effectiveness and superiority of ASA-NPP in improving session-based recommendation. The ASA-NPP takes semantics consistency into account instead of accepting all collaborative signals in generalities, while capturing cross-session collaborative signals. Additionally, it employs contrastive learning to further enhance the semantic discriminability and personalization of the learned session representation.

C. Ablation Studies

To verify the effectiveness of each module in ASA-NPP, we conduct an ablation test with four variants of ASA-NPP on Diginetica and Nowplaying. ASA-NPP-w/o-suni and ASA-NPP-w/o-spe represent the version without using the universal item embeddings and session-specific embeddings, respectively. ASA-NPP-w/o-filtering directly aggregates all collaborative signals from the overlapping sessions without filtering. ASA-NPP-w/o-sse removes Session Semantics Enhancement (SSE) module. Table III shows the performance of ASA-NPP and its different variants. Due to the space limitation, we only report the Hit@10 and MRR@10 results. From Table III, we can observe that ASA-NPP outperforms all of its variants, proving the effectiveness of each component of ASA-NPP.

Comparing with ASA-NPP-w/o-suni and ASA-NPP-w/o-spe, ASA-NPP takes account of both universal item embeddings and session-specific embeddings while generating session representation. This indicates the necessity of carrying out the learning of both kinds of embeddings. Additionally, we find that the contributions of these two parts are vary according to the dataset. Although removing any of these two parts causes performance degradation, the Nowplying is more sensitive to session-specific embedding which results in a larger performance drop without considering it.

When removing the semantic discriminator and signal filter, the ASA-NPP-w/o-filtering directly aggregates all cross-session collaborative signals without judging their semantics similarity, which degrades the performance. This verifies the validity of doing filtering and illustrates the importance of incorporating semantics-consistent collaborative signals.

The ASA-NPP-w/o-sse presents worse performance compared with ASA-NPP, which shows the effectiveness of contrastive learning in session semantics enhancement.

TABLE II

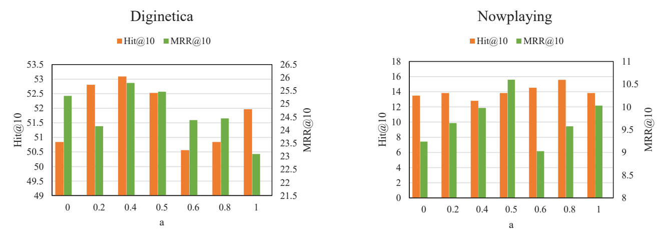
PERFORMANCE OF DIFFERENT MODELS ON REAL-WORLD DATASETS WITH % OMITTED. THE BEST AND THE SECOND BEST RESULTS ARE HIGHLIGHTED IN BOLDFACE AND UNDERLINE RESPECTIVELY. * DENOTES THE IMPROVEMENTS OVER THE STRONGEST BASELINE ARE STATISTICALLY SIGNIFICANT (P < 0.05) WITH PAIRED T-TESTS.

Method	Diginetica						Nowplaying					
	Hit@5	Hit@10	Hit@20	MRR@5	MRR@10	MRR@20	Hit@5	Hit@10	Hit@20	MRR@5	MRR@10	MRR@20
POP	1.1948	2.2760	3.7263	0.5325	0.6870	0.7843	2.9508	5.4182	8.2322	2.1302	2.4455	2.6293
ItemKNN	5.3901	6.4935	7.8211	5.2794	5.4237	5.5162	1.9340	2.4784	3.0765	1.7986	1.8638	1.9225
FPMC	18.4017	27.8155	36.8908	10.2859	11.5451	12.1916	8.2460	9.6231	11.4392	5.6861	5.8619	5.9850
GRU4Rec	9.7548	15.3578	22.1929	5.0507	5.7931	6.2566	5.7029	7.2593	9.1907	3.9614	4.1470	4.2876
NARM	11.1373	13.5376	16.8837	7.0420	7.3577	7.5898	7.5281	9.1240	9.9624	6.1278	6.3376	6.3903
STAMP	21.5917	31.3495	40.2390	11.8890	13.1951	13.7927	5.7595	7.2158	8.8632	4.2430	4.4385	4.5508
SR-GNN	16.1767	20.6023	24.8750	9.7326	10.3414	10.6275	8.2357	10.0686	11.5934	6.2777	6.5072	6.6129
GCE-GNN	14.7722	18.1910	21.7436	7.4163	7.8586	8.1001	9.6147	11.4772	13.2665	6.0156	6.2557	6.3774
SHARE	20.1835	23.7611	29.4085	12.4385	12.9062	13.2905	9.8622	11.4856	13.5843	7.4420	7.6476	7.8026
DHCN	26.9968	33.9009	39.8499	16.3270	17.2688	17.6852	10.1210	12.1791	15.5601	6.4557	6.7290	6.9674
FMLP-Rec	16.0286	20.9033	26.2454	9.5011	10.1584	10.5375	6.4516	8.0245	9.9505	4.3640	4.5606	4.6969
ASA-NPP	**29.9482***	**38.4893***	**48.1133***	**17.5490***	**18.7031***	**19.3640***	**11.8340***	**14.4133***	**17.0365***	**8.1057***	**8.4426***	**8.6131***
Improv. (%)	10.9324	13.5347	19.5688	7.4845	8.3057	9.4927	16.9252	18.3445	9.4884	8.9183	10.3954	10.3876

TABLE III

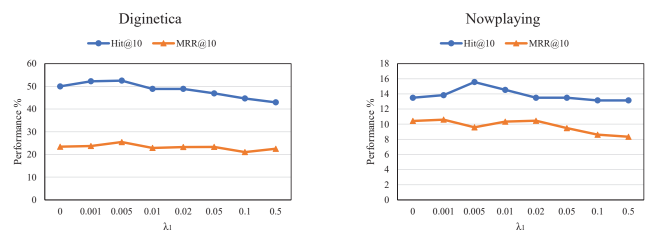
ABLATION TEST RESULTS WITH % OMITTED. THE BEST RESULT IN EACH COLUMN IS BOLDFACED.

Method	Diginetica		Nowplaying	
	Hit@10	MRR@10	Hit@10	MRR@10
ASA-NPP	**38.4893**	**18.7031**	**14.4133**	**8.4426**
ASA-NPP-w/o-suni	36.1618	16.9904	11.1779	6.8231
ASA-NPP-w/o-spe	37.5272	16.9877	12.3776	7.4539
ASA-NPP-w/o-filtering	36.9915	17.1460	12.9415	7.5467
ASA-NPP-w/o-sse	37.2897	17.6010	13.0809	7.5666



(a) (b)

Fig. 4. Performance of ASA-NPP with respect to various filter thresholds.



(a) (b)

Fig. 5. Effect of Contrastive Learning.

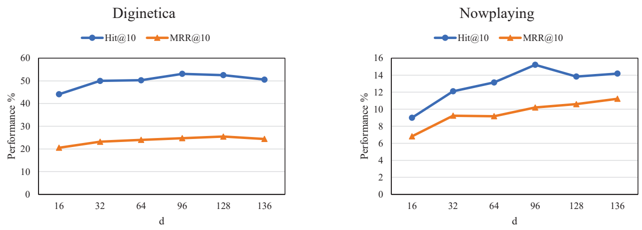
D. Hyper-parameter Study

In this subsection, we further study how the threshold of filter a, the magnitude of contrastive learning task λ_1 and the embedding size d influence the performance of ASA-NPP.

Effect of Filter Threshold. To investigate the impact of semantics-inconsistent collaborative signal filtering, we set the threshold of filter a in the range of $\{0, 0.2, 0.4, 0.5, 0.6, 0.8, 1.0\}$. The experimental results are demonstrated in Fig. 4, from which we can observe that setting an appropriate threshold of filter can improve the performance of ASA-NPP. Increasing the filter threshold does not always result in a better performance, since the larger the threshold a, the greater the possibility of erroneously filtering out the relevant collaborative signals. For Diginetica, the best performance of ASA-NPP on both Hit@10 and MRR@10 is achieved while setting a to 0.4. On Nowplaying, ASA-NPP yields best Hit@10 result and MRR@10 result with a as 0.8 and 0.5, respectively.

Effect of Contrastive Learning. To explore how the contrastive learning affects the performance of ASA-NPP, we tune the magnitude of the contrastive learning task λ_1 in $\{0, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.5\}$ and plot the results in Fig. 5. From Fig. 5, we can see that the performance of ASA-NPP improved while increasing the value of λ_1, and dropped after reaching to a peak in a relatively small value of λ_1. This indicates that the recommendation task is sensitive to the magnitude of the contrastive learning task, even setting it to a small value can boost the performance. However, the performance declines with larger λ_1.

Effect of Embedding Size. To study the influence of embedding size d, we range it in $\{16, 32, 64, 96, 128, 136, 144\}$. As showed in Fig. 6, with the increase of embedding size, the performance of the model first improves and then decreases. The reason is that a small embedding size may limit the learning capacity of model, and a large embedding size may cause model overfitting. Increasing the embedding size does not always result in a performance improvement.



(a) (b)

Fig. 6. Performance of ASA-NPP with respect to different embedding sizes.

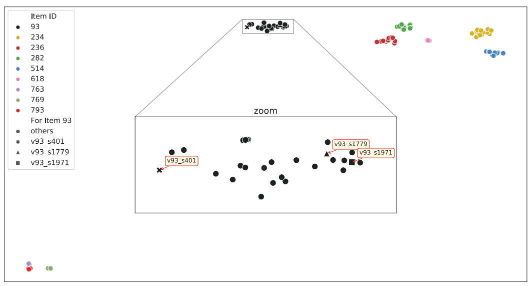


Fig. 7. t-SNE visualization of the learned session-specific embeddings for items' appearances in three sessions randomly sampled from Diginitica.

E. Case Study

To show the benefits of performing session-specific embedding learning with the help of SGAL, i.e., preventing preference negative transfer between irrelevant sessions, we visualize the learned session-specific embedding for items' appearances from three sampled sessions using t-SNE in Fig. 7. Specifically, items appearing in sessions with the same next item are considered to have similar semantics. We randomly select three sessions that all contain Item 93. Among them, two sessions (Session $1971 = \{282, 514, 618, 93\}$ and Session $1779 = \{236, 93, 282\}$) with same next item (Item 234) have similar semantics, and the remaining session (Session $401 = \{93, 793, 769\}$) has different next item (Item 763). The session-specific embeddings of Item 93 in three sessions are marked with different shape. From Fig. 7, we can observe that the distribution of anchor item Item 93's session-specific embeddings is relatively decentralized and the embeddings of Item 93 in Session 1971 and Session 1779 are situated closer, while the embedding of Item 93 in Session 401 is far from them. Hence, the collaborative signals from Session 1779 can be passed to Session 1971 and those from Session 401 are blocked. By exploiting session-specific embeddings, we are able to depict and distinguish various semantics of an item appeared in different sessions and block the cross-session collaborative signals with inconsistent semantics, which can help to prevent a preference negative transfer between irrelevant sessions and improves the accuracy of the learned representation.

V. Related Work

A. Session-based Recommendation

The early studies of session-based recommendation mostly utilize shallow machine learning techniques including neighborhood based methods [27], matrix factorization based methods [28], [29], and Markov chains based methods [4], [5]. With the advancement of deep neural networks, many deep learning methods such as RNN-based methods [6]–[8], CNN-based methods [9], [10], and attention mechanism based methods [11], [12] have been proposed for session-based recommendation and achieved promising results.

Recently, GNN-based approaches become popular in session-based recommendation since GNN have a great ability to model the complex transitions embedded in session graph data [3]. SR-GNN [1] is the pioneering work which models the session sequences as directed session graphs and applies the gated GNN based on that. Following the success of SR-GNN, GC-SAN [13] utilizes both GNN and self-attention mechanism to capture local and long-range dependencies respectively. In order to exploit cross-session collaborative signals, some works take account of the other sessions when constructing session graph. GCE-GNN [14] learns session-level and global-level item embeddings by employing attention mechanism and GNN on session graph and global graph. DHCN [2] adopts hypergraph and line graph to capture intra- and inter-session information separately. However, the existing methods suffer from a drawback that they all neglect the semantic inconsistency problem, which may result in a preference negative transfer and reduce the accuracy of recommendation. Different from existing works, our proposed model is able to judge the semantics similarity of cross-session collaborative signals and integrate the relevant one, while blocking those with inconsistent semantics.

B. Contrastive Learning for Recommendation

Contrastive learning aims to learn high-quality representations by making comparison with positive and negative samples in self-supervised manner [30]. Recent works introduce contrastive learning to recommender system [31]–[33]. S3-Rec [31] utilizes the mutual information maximization (MIM) principle to learn the correlations among attribute, item, subsequence, and sequence for sequential recommendation. In order to construct samples for comparison, CL4Srec [32] uses three data augmentation approaches (crop, mask, and reorder) to obtain different views of user interaction sequences. Similarly, SGL [33] applies three operators (node dropout, edge dropout, and random walk) on user-item graph to generate multiple views of a node and then exploits contrastive learning to improve node representation learning. SGL is the most relevant work to our model, which utilizes the augmentation operators to mitigate the influence of high-degree nodes and improve self-discrimination. However, the proposed augmentation operators of SGL is not applicable to session-based recommendation since the dropout operator may further aggravate the data sparsity problem of session data. Unlike it, we use an elaborate sampling strategy rather than change the original session data.

VI. Conclusion

In this paper, we propose a novel Anchor-Semantics-Aware Neural Preference Propagation (ASA-NPP) model for session-based recommendation, which is able to capture semantics-consistent collaborative signals across sessions and learn more accurate and personalized representations. ASA-NPP models the session data with an elaborately designed Session Graph with Anchor Links (SGAL) and exploits an Anchor-Semantics Attention Network (ASAN) on it to incorporate semantics-consistent collaborative signals for session-specific embedding learning. To enhance the semantic discriminability, a Session Semantics Enhancement (SSE) module is designed to preserve the discriminative information. Comprehensive experiments on real-world datasets demonstrate the effectiveness and superiority of ASA-NPP in session-based recommendation.

References

[1] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 346–353.

[2] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, and X. Zhang, "Self-supervised hypergraph convolutional networks for session-based recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 4503–4511.

[3] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, and D. Lian, "A survey on session-based recommender systems," *ACM Computing Surveys (CSUR)*, vol. 54, no. 7, pp. 1–38, 2021.

[4] G. Shani, D. Heckerman, R. I. Brafman, and C. Boutilier, "An mdp-based recommender system." *Journal of Machine Learning Research*, vol. 6, no. 9, 2005.

[5] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing personalized markov chains for next-basket recommendation," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 811–820.

[6] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," *arXiv preprint arXiv:1511.06939*, 2015.

[7] Y. K. Tan, X. Xu, and Y. Liu, "Improved recurrent neural networks for session-based recommendations," in *Proceedings of the 1st workshop on deep learning for recommender systems*, 2016, pp. 17–22.

[8] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural attentive session-based recommendation," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1419–1428.

[9] J. Tang and K. Wang, "Personalized top-n sequential recommendation via convolutional sequence embedding," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ser. WSDM '18, 2018, pp. 565–573.

[10] C. Xu, P. Zhao, Y. Liu, J. Xu, V. S. S.Sheng, Z. Cui, X. Zhou, and H. Xiong, "Recurrent convolutional neural network for sequential recommendation," in *The World Wide Web Conference*, ser. WWW '19, 2019, pp. 3398–3404. [Online]. Available: https://doi.org/10.1145/3308558.3313408

[11] C. Ma, L. Ma, Y. Zhang, J. Sun, X. Liu, and M. Coates, "Memory augmented graph neural networks for sequential recommendation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 5045–5052.

[12] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, "Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, ser. CIKM '19, 2019, pp. 1441–1450.

[13] C. Xu, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, F. Zhuang, J. Fang, and X. Zhou, "Graph contextualized self-attention network for session-based recommendation." in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, ser. IJCAI-19, 2019, pp. 3940–3946.

[14] Z. Wang, W. Wei, G. Cong, X.-L. Li, X.-L. Mao, and M. Qiu, "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*, 2020, pp. 169–178.

[15] Y. Pang, L. Wu, Q. Shen, Y. Zhang, Z. Wei, F. Xu, E. Chang, B. Long, and J. Pei, "Heterogeneous global graph neural networks for personalized session-based recommendation," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, ser. WSDM '22, 2022, pp. 775–783.

[16] Y. Zheng, S. Liu, Z. Li, and S. Wu, "Dgtn: Dual-channel graph transition network for session-based recommendation," in *2020 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2020, pp. 236–242.

[17] R. Qiu, Z. Huang, J. Li, and H. Yin, "Exploiting cross-session information for session-based recommendation with graph neural networks," *ACM Transactions on Information Systems (TOIS)*, vol. 38, no. 3, pp. 1–23, 2020.

[18] Q. Li, Z. Han, and X.-M. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," in *Thirty-Second AAAI conference on artificial intelligence*, ser. AAAI 2018, 2018, pp. 3538–3545.

[19] C. Yang, R. Wang, S. Yao, and T. Abdelzaher, "Hypergraph learning with line expansion," 2020.

[20] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "Lightgcn: Simplifying and powering graph convolution network for recommendation," in *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 2020, pp. 639–648.

[21] J. Wang, K. Ding, L. Hong, H. Liu, and J. Caverlee, "Next-item recommendation with sequential hypergraphs," in *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, ser. SIGIR '20, 2020, pp. 1101–1110.

[22] J. Wang, K. Ding, Z. Zhu, and J. Caverlee, "Session-based recommendation with hypergraph attention networks," in *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, 2021, pp. 82–90.

[23] L. Guo, H. Yin, Q. Wang, T. Chen, A. Zhou, and N. Quoc Viet Hung, "Streaming session-based recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1569–1577.

[24] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, 2001, pp. 285–295.

[25] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "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*, 2018, pp. 1831–1839.

[26] K. Zhou, H. Yu, W. X. Zhao, and J.-R. Wen, "Filter-enhanced mlp is all you need for sequential recommendation," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 2388–2399.

[27] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston *et al.*, "The youtube video recommendation system," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 293–296.

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

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

[30] Z. Liu, Y. Ma, Y. Ouyang, and Z. Xiong, "Contrastive learning for recommender system," *arXiv preprint arXiv:2101.01317*, 2021.

[31] K. Zhou, H. Wang, W. X. Zhao, Y. Zhu, S. Wang, F. Zhang, Z. Wang, and J.-R. Wen, "S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 1893–1902.

[32] X. Xie, F. Sun, Z. Liu, J. Gao, B. Ding, and B. Cui, "Contrastive pre-training for sequential recommendation," *arXiv e-prints*, pp. arXiv–2010, 2020.

[33] J. Wu, X. Wang, F. Feng, X. He, L. Chen, J. Lian, and X. Xie, "Self-supervised graph learning for recommendation," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 726–735.