

# DGTN: Dual-channel Graph Transition Network for Session-based Recommendation

Yujia Zheng\*, Siyi Liu\*, Zekun Li<sup>†‡</sup>, Shu Wu<sup>§¶</sup>

\*University of Electronic Science and Technology of China

<sup>†</sup>School of Cyber Security, University of Chinese Academy of Sciences

<sup>‡</sup>Institute of Information Engineering, Chinese Academy of Sciences

<sup>§</sup>School of Artificial Intelligence, University of Chinese Academy of Sciences

<sup>¶</sup>Institute of Automation and Artificial Intelligence Research, Chinese Academy of Sciences

{yjzheng19, ssui.liu1022, lizekunlee}@gmail.com, shu.wu@nlpr.ia.ac.cn

**Abstract**—The task of session-based recommendation is to predict user actions based on anonymous sessions. Recent research mainly models the target session as a sequence or a graph to capture item transitions within it, ignoring complex transitions between items in different sessions that have been generated by other users. These item transitions include potential collaborative information and reflect similar behavior patterns, which we assume may help with the recommendation for the target session. In this paper, we propose a novel method, namely Dual-channel Graph Transition Network (DGTN), to model item transitions within not only the target session but also the neighbor sessions. Specifically, we integrate the target session and its neighbor (similar) sessions into a single graph. Then the transition signals are explicitly injected into the embedding by channel-aware propagation. Experiments on real-world datasets demonstrate that DGTN outperforms other state-of-the-art methods. Further analysis verifies the rationality of dual-channel item transition modeling, suggesting a potential future direction for session-based recommendation.

**Index Terms**—Session-based Recommendation, Graph Neural Network

## I. INTRODUCTION

Session-based Recommendation System (SRS) has attracted much attention for its highly practical value, especially in some real-world scenarios that concentrated with multitudes of anonymous interactive data. Different from most of the other recommendation tasks that need explicit user demographic profiles, SRS only relies on anonymous user action logs (e.g., clicks) in an ongoing session to predict the user’s next action.

Under these circumstances, several methods are proposed to tackle the SRS task. Markov Chains (MC) [1] is a representation of traditional methods. It predicts the users next action based on the previous one thus introduces sequentiality into SRS. Recently, neural network-based methods have become popular due to their strong abilities to model sequential data, such as the methods based on Recurrent Neural Networks (RNN) [2]–[4]. Unfortunately, they can only model the uni-directional transitions between consecutive items but neglect those among other contextual items in the same session. To solve that, SR-GNN [5] models the target session in the graph structure and utilizes Graph Neural Networks (GNN) to model complex transitions among item nodes on the graph.

Despite the success of these methods based on RNN or GNN, they only focus on the item transitions within the target session but ignore those in the neighbor sessions. As a result, they are deficient in modeling the complex item transitions among different sessions, which contains potential fine-grained collaborative information for prediction.

In this paper, we propose a novel method, namely Dual-channel Graph Transition Network (DGTN), to model item transitions within not only the target session but also its neighbor ones. We first construct the target session and its neighbor sessions into a single graph. To consider the difference between the item transitions within the target sessions and the neighbor sessions, we leverage intra- and inter-session channels for embedding propagation, respectively. Then we use fusion function to aggregate features from the two channels.

The main contributions of our work are summarized as follows:

- We propose a novel model DGTN, which explicitly exploits item transitions among different sessions by constructing the target and neighbor sessions as a single graph and propagating embeddings on it in a channel-aware manner.
- We evaluate our model on two real-world datasets. Extensive experiments demonstrate the state-of-the-art performance of DGTN and its rationality of explicitly modeling dual-channel item transitions.

The rest of this paper is structured as follows. In § II we introduce related works about session-based recommendation and graph embedding. In § III we describe the necessary preliminaries. In § IV we elaborate the proposed model. The experiments and analysis are presented in § V. And the concluding takeaways are in § VI.

## II. RELATED WORK

In this section, we introduce some related works in Session-based Recommendation, Collaborative Filtering, and Graph Embedding.

### A. Session-based Recommendation

Traditional methods for session-based recommendation are mainly based on Markov Chains (MC) [1], [6]–[8], which introduces the sequentiality in SRS by predicting the user’s next action based on the last action. Zimdars *et al.* [6] apply probabilistic decision-tree models to study the way to extract the sequentiality. Mobasher *et al.* [7] choose the contiguous sequential patterns for SRS after studying the effect of different patterns. Shani *et al.* [8] employ the Markov Decision Processes that consider the long-term effect and the expected value of each recommendation. However, MC-based methods lose a balance between user’s general preference and sequential behavior, for they seldom consider sequentiality between items that are not consecutively adjacent in the same session. To achieve that balance, Rendle *et al.* [1] propose a hybrid method taking account of the combination of Matrix Factorization and MC, namely FPMC.

Like most other fields [9]–[13], deep learning methods [14]–[24] frequently appear in recent SRS models and obtain new state-of-the-art performance in terms of accuracy, especially RNN-based methods [2]–[4], [25]. Hidasi *et al.* [2] employ RNN with the Gated Recurrent Unit (GRU) into SRS and outperform traditional methods. Tan *et al.* [3] further improve it by introducing data augmentation, distillation integrating privileged information, and a pre-training approach to account for temporal shifts in the data distribution. Later, attention mechanism is applied by an encoder-decoder recommendation method (NARM) to combine sequentiality and user’s general preference [4]. However, STAMP [25] also adopts the concept combining general and current interest, but the difference is that STAMP explicitly models the current interest reflected by the last click to emphasize the importance of the last click, while NARM considers them as equally important. Most recently, geometric deep learning has become popular in a variety of tasks. SR-GNN [5] transforms sessions into the graph-structured data and applies GNN based on that. The significant improvement in recommendation performance proves the potential of geometric deep learning in SRS, and the motivation behind this work is enlightening to our work. However, all aforementioned deep learning methods only consider item transitions within a single session, which limits the upper bound of performance because of the lack of collaborative information.

Moreover, Collaborative Filtering (CF) idea-based methods are also popular in SRS. Unlike traditional user-based [26] or item-based [27]–[30] CF models in other recommendation tasks, modifications need to be made for them to perform well in SRS. Simply using item neighborhood information [31] cannot extract the integrity and sequentiality of items in the current session, which are extremely important for SRS application scenarios because of the lack of auxiliary data. Thus, SKNN [32] is proposed to consider each session as a whole and its improved version KNN-RNN [33] integrates GRU4REC to extract the sequentiality. Later, an end-to-end neural model (CSRMM [34]) outperforms KNN-RNN with

learnable latent session representations. The major difference between our method and theirs (KNN-RNN and CSRMM) is that they only stay at the minimum granularity of session as the collaborative information, while we dig deeper and integrate item transitions among different sessions into SRS.

### B. Graph Embedding

The most important part of aforementioned deep learning-based methods is embedding because generating more accurate and meaningful session embedding directly decides the performance. Thus, graph embedding becomes a critical component of the method when it comes to graph-structured data. However, traditional kernel-based methods (e.g., Weisfeiler-Lehman kernel [35], Deep Graph Kernels [36]) focus more on the unsupervised tasks and have trouble scaling to large graphs, so we mainly introduce neural network-based graph embedding methods here.

The concept of Graph Neural Networks (GNN) is first purposed by Gori *et al.* [37], then developed and deepened by Scarselli *et al.* [38] and Micheli *et al.* [39]. These early methods mainly generate representations of target nodes by using the recurrent neural unit to aggregate information of neighbor nodes. Inspired by the success of Convolutional Neural Network (CNN) in the image classification task, Bruna *et al.* [40] propose the spectral Graph Convolutional Neural Network (GCN). Then Defferrard *et al.* propose a variant model by introducing fast localized spectral filtering [41], and Kipf *et al.* improve it with a first-order approximation of spectral graph convolutions to motivate the choice of convolutional architecture [42]. Moreover, Message Passing Neural Network (MPNN) generalizes these GCN-based methods and introduces a two-step framework: message passing and readout [43]. Gated Graph Neural Networks (GGNN) [44] extends GNN to the sequential output, which is of great significance for sequential recommendations, such as SRS. However, as a large number of studies have shown that the attention mechanism improves the performance of deep learning-based methods in various tasks, it is therefore natural for researchers to import it on graphs [45], [46]. Most recently, Wu *et al.* [47] discovers that there is lots of unnecessary computation in GCN, and reduces its complexity by developing the simplified GCN (SGCN). After removing nonlinearities and collapsing weight matrices, SGCN can have a better efficiency without hurting the accuracy performance. Thus, we integrate this simplifying approach into our model.

## III. PRELIMINARY

The goal of SRS can be defined as using users current sequential session data to predict users next click items. Let  $\mathcal{V} = \{v_1, v_2, \dots, v_m\}$  represents a set of unique items in all sessions.  $s = \{v_1, v_2, \dots, v_n\}$  represents an anonymous session which contains items ordered by timestamps.  $S = \{s_1, s_2, \dots, s_{|S|}\}$  denotes the whole session set. For each item in  $\mathcal{V}$ , we embed it into a unified embedding space. Let  $\mathbf{v}_i \in \mathbb{R}^d$  denotes the latent vector of corresponding item  $v_i \in \mathcal{V}$ .

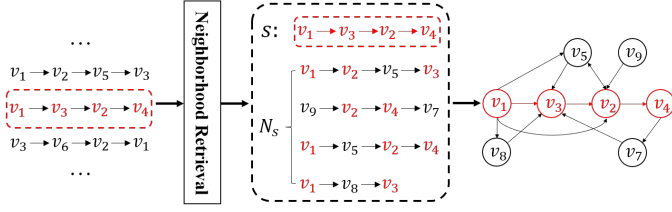


Fig. 1. The construction of the graph. We select the neighbor set  $N_s$  from previous sessions based on their similarities to the target session (red) and construct them as a session graph.

Given a session  $s$ , our model aims to predict the user's possible next click item  $v_{n+1}$ . We generate probabilities  $\hat{y}$  for all possible items based on the input target session  $s$ . Each element's value of vector  $\hat{y}$  is the recommendation score of the corresponding item. The items with a top- $T$  recommendation score will be recommended as our model's output.

#### IV. OUR PROPOSED METHOD: DGTN

##### A. Graph Construction

To model item transitions among different sessions in an explicit manner, we model the sessions in the graph structure (Figure 1). Given a session  $s$ , we first determine its neighbor set  $N_s$  consisting of its  $r$  most similar previous sessions. To simplify the computation, we use the number of duplicate items between sessions to calculate the similarity. Based on the ranking of the number of duplicate items between each session and the target session, we sample the top- $r$  sessions to constitute the neighborhood set  $N_s$ .

Then we model the target session  $s$  and its neighbor sessions in  $N_s$  as a session graph  $\mathcal{G}_s$ , where each node represents either an item that appears in the session  $s$  or any neighbor session in  $N_s$ . Each edge  $(v_i, v_j)$  denotes a user clicking on item  $v_j$  after  $v_i$  in the session  $s$  or any neighbor session in  $N_s$ . We denote  $\mathcal{V}_s$  as the set containing items in session  $s$ . And  $\mathcal{V}_{N_s}$  denotes the set of items belonging to  $N_s$ .

##### B. Item Embedding Learning

To encode the item transition signals into item embeddings, we follow the message-passing structure of GNN [48]. Each item node  $v_i$  in the graph aggregates messages passed from its neighbor item nodes in the graph. However, the constructed session graph contains item nodes from target session  $s$  and neighbor sessions  $N_s$ , which reflect the users' current behaviors and the global collaborative information, respectively. They ought to have different effects on the recommendation. Therefore, we propagate the item embeddings in two channels for the two types of neighbor item nodes. In the intra-session channel, we aggregate messages from only the neighbor nodes in  $\mathcal{V}_s$ . In the inter-session channel, we aggregate messages from the neighbor nodes in  $\mathcal{V}_{N_s}$ . Inspired by SGCN [47], we remove the nonlinearities and collapse the weight matrix

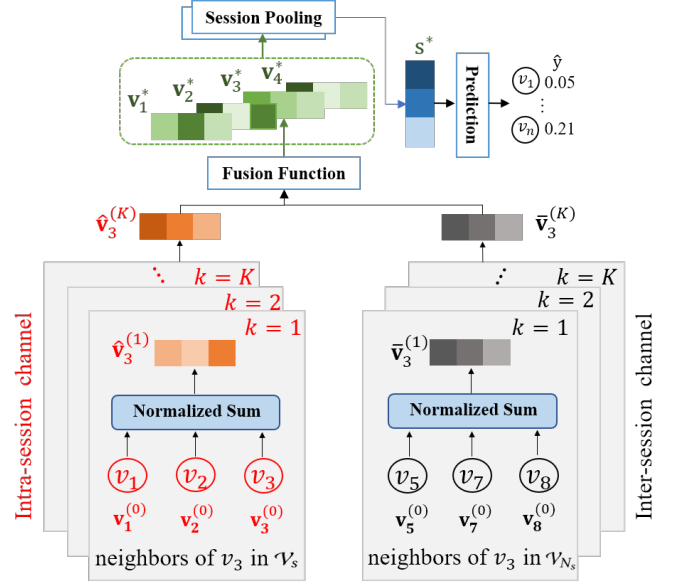


Fig. 2. An illustration of the model architecture. Taking item  $v_3$  for example, we propagate the embeddings of neighbor items in the target session  $s$  and the neighbor session set  $N_s$  towards the next layer in the intra- and inter-session channel, respectively. Then the embeddings of the final layer in the two channels are fed into the fusion function to obtain the final item embedding  $\hat{v}_3^*$ . Based on the learned item embeddings, we generate the session embedding via session pooling. Finally, we apply a prediction layer to generate the recommendation probability  $\hat{y}$ .

into one weight matrix. Only the normalized sum of neighbor embeddings are propagated towards next layers:

$$\begin{aligned}\hat{v}_i^{(k)} &= \frac{1}{d_i + 1} \mathbf{v}_i^{(k-1)} + \sum_{v_j \in \mathcal{V}_s} \frac{a_{ij}}{\sqrt{(d_i + 1)(d_j + 1)}} \mathbf{v}_j^{(k-1)}, \\ \bar{v}_i^{(k)} &= \frac{1}{d_i + 1} \mathbf{v}_i^{(k-1)} + \sum_{v_j \in \mathcal{V}_{N_s}} \frac{a_{ij}}{\sqrt{(d_i + 1)(d_j + 1)}} \mathbf{v}_j^{(k-1)},\end{aligned}\quad (1)$$

where  $d_i$  is the degree of node  $v_i$  in the adjacency matrix.  $a_{ij} = 1$  denotes that there is an edge between node  $v_i$  and  $v_j$ , and a missing edge is represented through  $a_{ij} = 0$ . The updated set of item embeddings from the intra-session channel is denoted as  $\hat{\mathbf{V}}_s = \{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n\}$ , where  $\hat{v}_i = \hat{v}_i^{(K)}$  and  $K$  is the number of layers. By contrast, the set from the inter-session channel is denoted as  $\bar{\mathbf{V}}_s = \{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n\}$ . Likewise,  $\bar{v}_i = \bar{v}_i^{(K)}$ .

Then we need to fuse the information of these two channels together and extract effective features (Figure 2). Here we explore five different fusion functions: (a) Mean Pooling (Mean) fuses the intra- and inter-session channel information by taking the mean value of every dimension of two embeddings. (b) Max Pooling (Max) takes the maximum value of every dimension of two embeddings. (c) Concatenation (Concat) is the concatenation of two item embeddings. (d) Fusion Gating (FG) is a linear interpolation between the two item embeddings  $\hat{v}_i$  and  $\bar{v}_i$ , inspired from the session fusion

gating in CSRM:

$$\begin{aligned} \mathbf{f} &= \sigma(\mathbf{W}_f^1 \hat{\mathbf{v}}_i + \mathbf{W}_f^2 \bar{\mathbf{v}}_i + \mathbf{b}_f), \\ \mathbf{v}_i^* &= \mathbf{f} \hat{\mathbf{v}}_i + (1 - \mathbf{f}) \bar{\mathbf{v}}_i, \end{aligned} \quad (2)$$

where  $\mathbf{W}_f^1, \mathbf{W}_f^2$  denote weight matrices in fusion gate and  $\mathbf{b}_f$  denotes the bias vector.  $\mathbf{v}_i^*$  is the final embedding for item  $v_i$ . (e) Convolution Neural Networks (CNN) Pooling is the strategy employed in our method, which will be elaborated later. Among multiple fusion functions, we employ CNN pooling according to experimental results. Suppose we have  $f$  vertical convolution filters,  $\mathbf{F}^k \in \mathbb{R}^{2 \times 1}, 1 \leq k \leq f$ . Each of them interacts with the columns of  $\mathcal{M}_i$ , the concatenation of two item embeddings, by sliding from left to right. Then the results of  $f$  convolution filters are concatenated and transformed to the final item embedding  $\mathbf{v}_i^*$ :

$$\begin{aligned} \mathcal{M}_i &= \hat{\mathbf{v}}_i^{(K)} \parallel \bar{\mathbf{v}}_i^{(K)}, \\ \mathbf{v}_i^* &= \mathbf{W}_i (c^1 \parallel c^2 \parallel \dots \parallel c^f), \end{aligned} \quad (3)$$

where  $\mathcal{M}_i \in \mathbb{R}^{2 \times d}$  is the concatenation of two item embeddings,  $c^k \in \mathbb{R}^{1 \times d}, 1 \leq k \leq f$  is the convolution result of filter  $\mathbf{F}^k$ ,  $\mathbf{W}_i \in \mathbb{R}^{1 \times f}$  projects  $\mathbf{o}_i \in \mathbb{R}^{f \times d}$  into embedding space  $\mathbb{R}^d$ .

### C. Session Pooling

Inspired by STAMP, we combine both users' long-term preference and short-term interests of the session to generate the final session embedding. For the session  $s = \{v_1, v_2, \dots, v_n\}$ , we use the last click item's embedding to represent user's short-term interests, i.e.,  $\mathbf{p}_s = \mathbf{v}_n^*$ . Then we obtain the long-term preference  $\mathbf{p}_l$  by adopting a soft-attention mechanism to draw dependencies between the short-term interest  $\mathbf{p}_s$  and each item in the session. Specifically, we derive  $\mathbf{p}_l$  by the following calculation:

$$\begin{aligned} \alpha_i &= \text{softmax}(\mathbf{q}^T (\mathbf{W}_a^1 \mathbf{p}_s + \mathbf{W}_a^2 \mathbf{v}_i^* + \mathbf{W}_a^3 \mathbf{v}_{avg}^* + \mathbf{b}_a)), \\ \mathbf{p}_l &= \sum_{i=1}^n \alpha_i \mathbf{v}_i^*, \end{aligned} \quad (4)$$

where  $\mathbf{q}^T \in \mathbb{R}^d$  is a projection vector,  $\mathbf{W}_a^1, \mathbf{W}_a^2 \in \mathbb{R}^{d \times d}$  are learnable weighted parameters,  $\mathbf{v}_{avg}^* = \sum_i (\mathbf{v}_i^*) / n$  is the average item embedding, and  $\mathbf{b}_a$  is a bias vector.

Finally, we combine  $\mathbf{p}_l$  and  $\mathbf{p}_s$  to generate the final session embedding  $\mathbf{s}^*$ :

$$\mathbf{s}^* = \mathbf{W}_c [\mathbf{p}_l \parallel \mathbf{p}_s], \quad (5)$$

where  $\mathbf{W}_c \in \mathbb{R}^{d \times 2d}$  transfers the concatenation vector from latent space  $\mathbb{R}^{2d}$  to  $\mathbb{R}^d$ .

### D. Prediction Module and Objective Function

After obtaining the final session representation  $\mathbf{s}^*$ , we use it to multiply each candidate item vector  $\mathbf{v}_i$  to generate recommendation score  $\hat{\mathbf{z}}_i$  for corresponding item:

$$\hat{\mathbf{z}}_i = \mathbf{v}_i^\top \mathbf{s}^*. \quad (6)$$

Then we apply a softmax function to generate the output vector  $\hat{\mathbf{y}}$  of the model:

$$\hat{\mathbf{y}} = \text{softmax}(\hat{\mathbf{z}}), \quad (7)$$

where  $\hat{\mathbf{z}} \in \mathbb{R}^m$  represents the recommendation scores over all candidate items, and  $\hat{\mathbf{y}} \in \mathbb{R}^m$  denotes the probabilities of items becoming the next-click item in session  $s$ .

In the training process, we apply cross-entropy as the loss function:

$$\mathcal{L}(\hat{\mathbf{y}}) = - \sum_{i=1}^m \mathbf{y}_i \log(\hat{\mathbf{y}}_i) + (1 - \mathbf{y}_i) \log(1 - \hat{\mathbf{y}}_i), \quad (8)$$

where  $\mathbf{y}$  denotes the one-hot encoding vector of the ground truth item. Finally, our model is trained by Back-Propagation Through Time (BPTT) algorithm.

## V. EXPERIMENTS

In this section, we aim to answer the following research questions:

- **RQ1.** How does DGTN perform as compared with state-of-the-art SRS methods?
- **RQ2.** How does the number of neighbor sessions affect the model performance?
- **RQ3.** How do different fusion functions affect the model performance?
- **RQ4.** What is the training efficiency of DGTN?

We conduct experiments on two real-world datasets: Yoochoose<sup>1</sup> and Diginetica<sup>2</sup>. We apply the same preprocessing as [5], [25]. We filter out sessions of length one and items that appear less than five times for both datasets as same as previous studies [5], [25]. Furthermore, we use the last one day in YOOCHOOSE and last seven days in Diginetica to generate the test data. The data static is shown in Table I. Because collaborative filtering-based methods cannot recommend an item which has not appeared before [2], we filter out items from test set which do not appear in the training set. Following previous studies [5], [25], We only use the most recent 1/64 of the training sequence of Yoochoose. We use Recall@20 and MRR@20 as evaluation metrics.

**Baseline.** We compare DGTN with frequency based method (POP), two RNN-based methods (GRU4REC [2] and NARM [4]), attention-based method (STAMP [25]), two traditional Matrix Factorization or Markov Chain approaches (FPMC [1] and BPR-MF [49]), and GNN-based method (SR-GNN [5]). We also compare the model with some neighborhood-based methods. Although they consider the collaborative information among multiple sessions, they either ignore the sequentiality of items (Item-KNN [31] and SKNN [32]) or only consider coarse-grained session-level collaborative information (KNN-RNN [33] and CSRM<sup>3</sup> [34]), failing to model the complex item transitions between different sessions.

<sup>1</sup><http://2015.recsyschallenge.com/challenge.html>

<sup>2</sup><http://cikm2016.cs.iupui.edu/cikm-cup>

<sup>3</sup>Following previous studies [5], [25], we set the hidden size of all baseline to 100, which is different from the original setting of CSRM.

TABLE I  
STATISTICS OF DATASETS USED IN OUR EXPERIMENTS.

Datasets	# of clicks	# of training sessions	# of testing sessions	# of items	average length
YOOCHOOSE 1/64	565,552	375,043	55,405	17,319	6.07
Diginetica	982,961	719,470	68,977	43,097	5.12

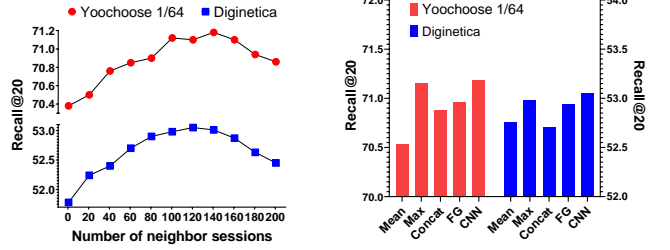
TABLE II  
PERFORMANCE COMPARISON OF DIFFERENT METHODS.

Methods	Diginetica		Yoochoose 1/64	
	MRR@20	Recall@20	MRR@20	Recall@20
POP	0.20	0.89	1.65	6.71
Item-KNN	11.57	35.75	21.82	51.60
SKNN	16.77	45.45	24.03	57.27
FPMC	6.95	26.53	15.01	45.62
BPR-MF	1.98	5.24	12.08	31.31
GRU4REC	8.33	29.45	22.89	60.64
NARM	16.17	49.70	28.63	68.32
STAMP	14.32	45.64	29.67	68.74
KNN-RNN	9.65	31.89	24.05	62.36
CSRM	17.16	52.00	30.48	70.79
SR-GNN	17.59	50.73	30.94	70.57
DGTN	<b>18.07</b>	<b>53.05</b>	<b>31.35</b>	<b>71.18</b>
Improv.(%)	2.72	2.02	1.33	0.56

#### A. Comparison with baseline methods (RQ1)

The experimental results of all methods are illustrated in Table II, and the following observations stand out:

- Two KNN-based methods, Item-KNN and SKNN, consistently achieve better performance than other conventional methods. Both RNN-based methods (GRU4REC and NARM) underperform their neighborhood-enhanced versions (KNN-RNN and CSRM). This illustrates the effectiveness of adopting collaborative information from other sessions.
- All of the neural network-based methods distinctly outperform other conventional recommendation methods, demonstrating the superiority of adopting deep learning technology to make recommendations. The key reason for this may be RNNs ability to process sequentiality and thus model the item transitions within the target session.
- On the whole, graph-based methods (SR-GNN, and DGTN) outperform RNN-based methods (GRU4REC, NARM, KNN-RNN, and CSRM). This indicates the significance of explicitly modeling complex contextual item transitions owing to the strong power of GNN in modeling graph-structured data. On the contrary, RNN can only deal with unidirectional transitions between consecutive items.
- DGTN consistently yields the best performance on all datasets (RQ1), which verifies the superiority of the proposed method. By propagating embedding on the constructed graph, DGTN is capable of exploring complex item transitions among different sessions. And the improvement over SR-GNN indicates the significance of that.



(a) Number of neighbor sessions (b) Different fusion functions  
Fig. 3. Performance w.r.t. number of neighbor sessions (a), and fusion functions (b).

#### B. Effect of the number of neighbors (RQ2)

Focusing on the current session  $s$  allows the model to dig deeper into the current user preference, but loses valuable collaborative information with other sessions; Introducing a large set of neighbor sessions  $N_s$  widens the range of information but leads to more noises. Thus, we vary the number of neighbor sessions  $r$  from zero<sup>4</sup> to 200, to study the trade-off between them (RQ2). In Figure 3(a), the performance of DGTN improves with the increase of  $r$  at first since more collaborative information is introduced. However, it starts to drop after  $r = 140$  on Yoochoose 1/64, and  $r = 120$  on Diginetica, because of the increasing noise introduced by less similar sessions (the similarity decreases as  $r$  increases). Differences in the optimal numbers might be caused by the statistical differences between two datasets. And it is clear that neighbor sessions significantly increase the model performance, which verifies the rationality and effectiveness of integrating dual-channel item transitions.

#### C. Effect of fusion functions (RQ3)

To analyze how different fusion functions affect the results (RQ3), we test the model with five different fusion functions. From Figure 3, observations of the results can be listed as follows:

- CNN pooling outperforms others on both datasets in terms of Recall@20. And the same phenomenon occurs in the MRR@20 evaluation, which is omitted for the space limit. This indicates that CNN has a stronger ability to fuse intra- and inter-session channel features, which might be attributed to its representational capacity.
- Mean Pooling and Concatenation show bad performance in both datasets. The reason may be that a simple overall

<sup>4</sup> $r = 0$  means that the model only consider transitions within the target session.

smoothness does not well deal with the differences between two different channels for the diversity of different user behaviors.

- Max Pooling is also a good choice because of its relatively high performance and computational efficiency, which illustrates the effectiveness of capturing the most important feature.

#### D. Efficiency (RQ4)

TABLE III  
EFFICIENCY (TRAINING TIME PER EPOCH AND GPU MEMORY USAGE)  
COMPARISON BETWEEN SR-GNN AND DGTN.

Methods	Diginetica		Yoochoose 1/64	
	Time (s)	Memory (MB)	Time (s)	Memory (MB)
SR-GNN	601	973	522	961
DGTN	895	1031	679	995

We evaluate the training efficiency of DGTN in this section. As DGTN trains on a larger session graph than previous graph-based methods, we are curious about its efficiency compared with SR-GNN. To make a fair comparison, we set the batch size as 128 and the hidden size as 100 for both methods, following the common implementation [5], [25]. All experiments are conducted on a single GeForce RTX 2080ti GPU and the same computation environment. Both methods are training with 15 epochs and we report the average training time per epoch. And the GPU memory usages are recorded when the training stabilizes. The results of the training time per epoch and the GPU memory usage are shown in Table III.

From Table III, we can observe that DGTN performs worse than SR-GNN, which is reasonable as DGTN uses a bigger session graph. However, the difference between GPU memory usage is minor and the training time of DGTN is also acceptable considering the performance improvement.

#### VI. CONCLUSION

We proposed a novel method, DGTN, to explicitly model item transitions among different sessions. We constructed the complex transitions between items among different sessions into a single graph. Then the transition signals are injected into the embedding in the channel-aware learning process. Extensive experiments on two real-world datasets demonstrate the effectiveness of DGTN.

#### ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China (61772528, U19B2038) and National Key Research and Development Program (2016YFB1001000, 2018YFB1402600).

#### REFERENCES

- [1] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing personalized markov chains for next-basket recommendation," in *WWW*, 2010.
- [2] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," in *ICLR*, 2015.
- [3] Y. K. Tan, X. Xu, and Y. Liu, "Improved recurrent neural networks for session-based recommendations," in *DLRS*, 2016.
- [4] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural attentive session-based recommendation," in *CIKM*, 2017.
- [5] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *AAAI*, 2019.
- [6] A. Zimdars, D. M. Chickering, and C. Meek, "Using temporal data for making recommendations," in *UAI*, 2001.
- [7] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Using sequential and non-sequential patterns in predictive web usage mining tasks," in *ICDM*, IEEE, 2002.
- [8] G. Shani, D. Heckerman, and R. I. Brafman, "An mdp-based recommender system," *Journal of Machine Learning Research*, vol. 6, no. Sep, pp. 1265–1295, 2005.
- [9] Z. Li, Z. Cui, S. Wu, X. Zhang, and L. Wang, "Fi-gnn: Modeling feature interactions via graph neural networks for ctr prediction," in *CIKM*, 2019.
- [10] Y. Zhang, X. Yu, Z. Cui, S. Wu, Z. Wen, and L. Wang, "Every document owns its structure: Inductive text classification via graph neural networks," *ACL*, 2020.
- [11] F. Hu, Y. Zhu, S. Wu, W. Huang, L. Wang, and T. Tan, "Graphair: Graph representation learning with neighborhood aggregation and interaction," *arXiv preprint arXiv:1911.01731*, 2019.
- [12] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Deep graph contrastive representation learning," *arXiv preprint arXiv:2006.04131*, 2020.
- [13] X. Chen, C. Du, X. He, and J. Wang, "Jit2r: A joint framework for item tagging and tag-based recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1681–1684, 2020.
- [14] F. Yu, Y. Zhu, Q. Liu, S. Wu, L. Wang, and T. Tan, "Tagnn: Target attentive graph neural networks for session-based recommendation," in *SIGIR*, 2020.
- [15] F. Mi, X. Lin, and B. Faltings, "Ader: Adaptively distilled exemplar replay towards continual learning for session-based recommendation," in *RecSys*, 2020.
- [16] S. Liu and Y. Zheng, "Long-tail session-based recommendation," in *RecSys*, 2020.
- [17] S. Wang, L. Hu, Y. Wang, Q. Z. Sheng, M. A. Orgun, and L. Cao, "Modeling multi-purpose sessions for next-item recommendations via mixture-channel purpose routing networks," in *IJCAI*, 2019.
- [18] S. Wang, L. Cao, and Y. Wang, "A survey on session-based recommender systems," *arXiv preprint arXiv:1902.04864*, 2019.
- [19] S. Wang, L. Hu, Y. Wang, X. He, Q. Z. Sheng, M. Orgun, L. Cao, N. Wang, F. Ricci, and P. S. Yu, "Graph learning approaches to recommender systems: A review," *arXiv preprint arXiv:2004.11718*, 2020.
- [20] S. Wang, L. Hu, Y. Wang, Q. Z. Sheng, M. Orgun, and L. Cao, "Intention nets: psychology-inspired user choice behavior modeling for next-basket prediction," in *AAAI*, 2020.
- [21] S. Wang, L. Hu, Y. Wang, Q. Z. Sheng, M. Orgun, and L. Cao, "Intention2basket: A neural intention-driven approach for dynamic next-basket planning," in *IJCAI*, 2020.
- [22] S. Wu, M. Zhang, X. Jiang, X. Ke, and L. Wang, "Personalizing graph neural networks with attention mechanism for session-based recommendation," *arXiv preprint arXiv:1910.08887*, 2019.
- [23] X. Chen, H. Xu, Y. Zhang, J. Tang, Y. Cao, Z. Qin, and H. Zha, "Sequential recommendation with user memory networks," in *WSDM*, 2018.
- [24] W. Ye, S. Wang, X. Chen, X. Wang, Z. Qin, and D. Yin, "Time matters: Sequential recommendation with complex temporal information," in *SIGIR*, 2020.
- [25] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "Stamp: short-term attention/memory priority model for session-based recommendation," in *KDD*, 2018.
- [26] R. Jin, J. Y. Chai, and L. Si, "An automatic weighting scheme for collaborative filtering," in *SIGIR*, 2004.
- [27] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, 2003.
- [28] D. Billsus and M. J. Pazzani, "Learning collaborative information filters," in *ICML*, 1998.
- [29] J. Pareek, M. Jhaveri, A. Kapasi, and M. Trivedi, "Snetrs: social networking in recommendation system," in *Advances in Computing and Information Technology*, Springer, 2013.

- [30] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Application of dimensionality reduction in recommender system-a case study," tech. rep., Minnesota Univ Minneapolis Dept of Computer Science, 2000.
- [31] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *WWW*, 2001.
- [32] G. Bonnin and D. Jannach, "Automated generation of music playlists: Survey and experiments," *ACM Computing Surveys (CSUR)*, 2015.
- [33] D. Jannach and M. Ludewig, "When recurrent neural networks meet the neighborhood for session-based recommendation," in *RecSys*, 2017.
- [34] M. Wang, P. Ren, L. Mei, Z. Chen, J. Ma, and M. de Rijke, "A collaborative session-based recommendation approach with parallel memory modules," in *SIGIR*, 2019.
- [35] N. Shervashidze, P. Schweitzer, E. J. v. Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *Journal of Machine Learning Research*, 2011.
- [36] P. Yanardag and S. Vishwanathan, "Deep graph kernels," in *KDD*, ACM, 2015.
- [37] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *IJCNN*, 2005.
- [38] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, 2008.
- [39] A. Micheli, "Neural network for graphs: A contextual constructive approach," *IEEE Transactions on Neural Networks*, 2009.
- [40] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013.
- [41] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *NIPS 2016*, 2016.
- [42] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
- [43] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *ICML*, 2017.
- [44] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," *arXiv preprint arXiv:1511.05493*, 2015.
- [45] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *ICLR*, 2018.
- [46] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, "Gram: graph-based attention model for healthcare representation learning," in *KDD*, 2017.
- [47] F. Wu, T. Zhang, A. H. d. Souza Jr, C. Fifty, T. Yu, and K. Q. Weinberger, "Simplifying graph convolutional networks," in *ICML*, 2019.
- [48] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017.
- [49] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," in *UAI*, 2009.