A Graph ATtention Networks Model for Session-Based Recommender Systems

Boudjemaa Boudaa
Department of Computer Science
University of Tiaret
Tiaret, Algeria
boudjemaa.boudaa@univ-tiaret.dz

Hadil Touhami
Department of Computer Science
University of Tiaret
Tiaret, Algeria
hadil.touhami@univ-tiaret.dz

2023 International Conference on Networking and Advanced Systems (ICNAS) | 979-8-3503-1917-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICNAS59892.2023.10330439

***Abstract*—In recent years, session-based recommender systems (SBRSs) have emerged as a new paradigm of recommender systems to help users in their different decision-making processes. The goal of SBRSs is to capture dynamic and short-term user preferences within sessions to provide more timely and accurate next-item recommendations that are sensitive to be adapted in different contexts. In the literature, the proposed approaches for SBRS development are limited to some models that lack more precision and efficacy, which are the primary purposes for this kind of recommender system. This paper aims to formally present a graph neural networks model for session-based recommender systems based on Graph ATtention networks (GAT). GAT can capture complex transitions between items through a session modelled as graph-structured data. The effectiveness of the proposed GAT-based model is extensively evaluated on three public real-world datasets. Experimental results show the superiority of our model against some used baselines in the SBRS field.**

***Index Terms*—Session-Based Recommender System, Graph ATtention Network, Attention Mechanism.**

I. Introduction

Recommender systems (RS) suggest useful items (products or services) to users in order to help them in their different decision-making processes [1]. Nowadays, the recommendation systems' effectiveness is clearly confirmed in diverse application domains (e.g., YouTube, Amazon, ResearchGate, and social networks).

In the last few years, session-based recommender systems (SBRSs) have emerged as a new recommender system type [2]. As a kind of sequence-aware recommender system [3], SBRSs capture dynamic and short-term user preferences within sessions to provide more timely and accurate next-item recommendations according to different session contexts. In the literature, the proposed development approaches of SBRSs [4] are limited to traditional deep learning models such as CNN and RNN that lack capturing complex transitions between user-item iterations and consequently decrease the accuracy and effectiveness of recommendations.

This paper aims to present a new design model for session-based recommender systems using the graph neural networks (GNN) across its architecture of graph attention network (GAT). In addition to representing the data in non-Euclidean space and handling the complex transitions between interacted items, GAT can handle graphs of varying sizes and structures without the need to prior knowledge about graph topology. This makes it a valuable tool for a wide range of applications, including social network analysis, protein structure prediction, and recommendation systems.

979-8-3503-1917-0/23/$31.00 ©2023 IEEE

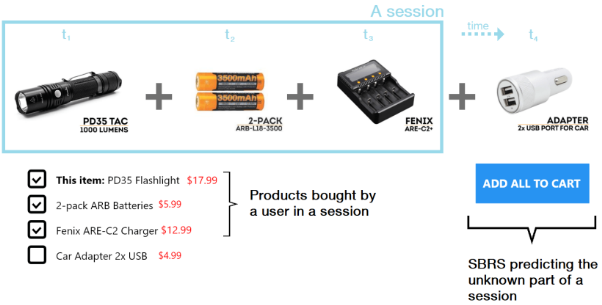
A session
t₁
t₂
t₃
time
t₄
PD35 TAC
1000 LUMENS
2-PACK
ARB-L18-3500
FENIX
ARE-C2+
ADAPTER
2x USB PORT FOR CAR
☑ This item: PD35 Flashlight $17.99
☑ 2-pack ARB Batteries $5.99
☑ Fenix ARE-C2 Charger $12.99
☐ Car Adapter 2x USB $4.99
Products bought by a user in a session
ADD ALL TO CART
SBRS predicting the unknown part of a session

Figure 1. An example of an e-commerce website using SBRS.

The remainder of this paper is organized as follows. Theoretical background about SBRS and GAT is given in Section 2. In Section 3, some important related works are discussed. Section 4 details our proposed GAT-based model for SBRS. Sections 5 and 6 present the conducted experiments and discuss the obtained results. Finally, Section 7 concludes this article by showing our future work.

II. Fundamentals

A. Session-Based Recommender Systems

Session-based recommender systems (SBRS) [2] represent a promising trend in the field of recommender systems. SBRS take into account short-term user preferences (or intents). They work on observing the interactions of the user with previous items in the current session to predict recommendations about what would happen next on the items (e.g., next video to watch, POI to visit, or product to purchase). Their recommendation process relies on tracking the intra-session dependencies between time-stamped user-item interactions.

Currently, SBRS have gotten more and more interest from researchers and industrialists [4], as are becoming suitable recommendation tools in many real-life domains such as e-commerce, tourism, and leisure. For instance, the smart supermarket TRIAL[1] recommends the next purchases depending on a sequence of consumers' previous actions in the same e-commerce session. Fig. 1 illustrates the SBRS process through an example of a shopping session.

The main characteristics of SBRSs [2], [4] are, (i) users are not identifiable (because the majority of website traffic is from first-time or non-logged-in users), (ii) due to (i), long-term preferences information is unavailable, and (iii) user preferences should be inferred from a small set of sequential interactions during a session.

In the literature [4], various approaches and models have been proposed to develop SBRSs. This paper presents a design model for session-based recommender systems using Graph Neural Networks (GNN) via its promising Graph ATtention network (GAT) architecture. This has demonstrated the ability to model complex transitions within and between sessions as graph-structured data using deep neural networks to provide better performance for SBRS. The next section will introduce the GAT architecture.

B. *Graph Attention Networks*

Non-Euclidean structural data (such as interactions in social networks or between chemical molecules) cannot be captured by conventional deep learning models (e.g., RNN and CNN). For this end and based on graphs theory (nodes and edges), GNN [5], [6] have been proposed to represent the data in non-Euclidean space and to handle their complex transitions.

The GNN process involves propagating information through the graph by passing messages between neighbouring nodes. At any iteration of the message-passing algorithm, each node receives messages from its neighbours, aggregates those messages, and then updates its own state based on the aggregated information. This process is repeated for a predetermined number of iterations until convergence is attained.

GNN can be designed through various architectures of which three main ones are considered in the SBRS literature [4], namely: Gated Graph Neural Network (GGNN), Graph Convolutional Network (GCN), and Graph Attention Network (GAT). The latter will be used in this paper to present our proposed model (see Section 4).

GAT [7] is a neural network architecture that utilizes attention mechanisms to handle graph data. Nodes and edges in a graph have features that represent different characteristics of the graph. These features are used as input by GAT, which learns to assign weights to them using attention mechanisms. Similar to GCN, the attention mechanism in GAT is based on a graph convolution operation. For each neighbouring node, it computes a weight based on its features and those of the current node. This enables the model to process the graph while concentrating on the most significant features.

Moreover, GAT computes numerous sets of weights for every node using multiple attention heads [7], which aids in gathering various facets of the graph structure. The multi-head attention enables the model to capture diverse patterns of relationships between nodes by using multiple parallel attention mechanisms, each with its own weight matrix. Each attention head's outputs are combined and sent through a non-linear activation function.

GAT is now a strong model for processing graph-structured data [8] and has been effectively applied in several applications, including recommender systems.

III. Related Work

To the best of our knowledge, there are only two researches which have used GAT with its entire process for developing session-based recommender systems. In [9], the authors model user-item sessions using GAT in what is called PSR-GAT for Personalized Session-based Recommendation using Graph Attention Networks. The PSR-GAT model combines a user's past preferences at several scales in addition to the available information in item transitions. Furthermore, the new model of Knowledge-enhanced Graph Attention Network for Session-based Recommendation (KGAT-SR) is proposed in [10]. It exploits the knowledge about items via a knowledge graph attention network to generate a knowledge-enhanced session graph (KESG). The latter is aggregated via weighted graph attention, and the node features and graph topology in the graph are utilized to generate accurate session embedding for recommending the next item. These two works presented complex architectures based on GAT which could take a long time in computing operations.

Other research works in the literature have been partially based on GAT by only using the attention mechanism layer. Here, we will mention the most important ones. SR-GNN [11] is the pioneer model that has utilized GNN for capturing the complex structure and transitions between items within a session. Formally, it is based on another architecture of GNN namely, gated graph neural networks (GGNN) and just using the attention mechanism layer of GAT architecture to combine current interests with long-term preferences for predicting the next behaviours of users. Whereas, our contribution in this article will be based on the whole functional framework of GAT including its attention mechanism and without any hybridization with other GNN architecture to present a new design model for building SBRSs. TAGNN [12] captures rich item transitions in sessions and learns the node vectors using GGNN. It extends SR-GNN by proposing a novel target-aware attention mechanism that learns different user interests with respect to varied target items.

Recently and innovatively, a session-enhanced graph neural network (SE-GNNRM) model is proposed in [13]. In this model, the complex transition relationship between items and item features is captured using a self-attention mechanism. Then, the attention mechanism is employed to combine short-term and long-term preferences to construct a global session graph. The authors in [14] propose AutoGSR, a NAS (Neural Architecture Search) framework for automatically searching suitable graph architectures to be adopted in different session-based recommendation scenarios. To determine the optimal graph neural network architecture, two novel GNN operations

[1] https://www.trial-net.co.jp/shops/view/319

Table I
NOTATIONS AND DESCRIPTIONS

Notation	Description
W	Weight matrix of the linear transformation
a	Weight vector of the attention mechanism
h_i, h_j	Transformed feature vectors of nodes i and j
e_{ij}	Attention coefficient between nodes i and j
α_{ij}	Attention score for the edge $(i$,$j)$
$softmax$	Normalization function
$\|$	Concatenation operation
$\overrightarrow{a}^T$	Transposition of a
$LeakyRelu$	Activation function with negative input slope $\alpha = 0.2$
$\overrightarrow{h'_i}$	Final embedding vector of node i
σ	Activation function

(namely, Relational GGNN and Mixup which combines the relational GGNN with the relational GAT) are added to build a complete and expressive search space, and a differentiable search algorithm is used.

Otherwise, to add other information sources (such as social information) in session-based recommendation processes, the authors in [15] propose a novel session-based social recommendation model named GNNRec, in which a GGNN is first used to represent the current session information of the user. Next, an attention mechanism is utilized to aggregate social information on users and their friends on social networks for modelling users' interests.

Very recently, GPAN or Graph Positional Attention Network has been presented in [16]. It is based on position attention in response to the use of the user's higher-order features and to address the impact of item position information on the current session, enhancing predictions in SBRSs.

IV. A GAT-BASED MODEL FOR SBRS

Our proposed model based on GAT for building SBRS is depicted in Fig. 2. After representing a session with a graphical structure, this model processes the resulting graph with its adjacency matrix through several layers iteratively. The formulas that govern operations of these layers have been drawn from [7]. In Table I, the formal notations used in this process are described.

1) Input layer: It receives the graph as a set of node features vectors as inputs to the model,

$$h = \left\{\overrightarrow{h_1}, \overrightarrow{h_2}, ..., \overrightarrow{h_N}\right\}, \overrightarrow{h_i} \in R^F$$

where N is the number of nodes, and F is the number of features in each node.

2) Linear Transformation layer: To represent each node in a lower dimension, the feature matrix x(set of $\overrightarrow{h_i}$) is transformed using a shared weight matrix W and a bias b to produce the output Y.

$$Y = xW + b$$

This linear transformation allows making the required features simpler to identify and classify to compute the attention coefficients on a reduced feature space.

3) Attention mechanism Layer: an attention mechanism is used to compute the importance of each neighbouring node for a given node i. Two main operations are carried out in this layer, namely:

- Attention Coefficients: The attention coefficients are computed by a shared neural network with parameters of a. The output of this network is a scalar value e_{ij} that represents the compatibility between nodes i and j.

$$e_{ij} = a\left(W\overrightarrow{h_i}, W\overrightarrow{h_j}\right)$$

- Attention Scores: The softmax function is then used to normalize the attention coefficients in order to produce a probability distribution for node i's neighbours. The model can then focus on the nodes that are the most important for each node.

$$\alpha_{ij} = softmax\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{k_{i \in N}} \exp\left(e_{ik}\right)}$$

The attention mechanism is a single-layer neural network, the input for this network are the two transformed node features vectors for an edge, and applying the *LeakyReLU* nonlinearity (with negative input slope α=0.2). The output indicates the importance of these nodes.

$$\alpha_{ij} = softmax\left(e_{ij}\right) = \frac{\exp\left(LeakyReLu(\overrightarrow{a}^T\left[\left(W\overrightarrow{h_i}\|W\overrightarrow{h_j}\right)\right])\right)}{\sum_{k_{i \in N}} \exp\left(LeakyReLu(\overrightarrow{a}^T\left[\left(W\overrightarrow{h_i}\|W\overrightarrow{h_j}\right)\right])\right)}$$

4) Aggregation and Update Layer: By aggregating the embeddings of node i's neighbours, weighted by their attention scores, the final embedding vector can be updated. This is done by using a weighted and nonlinearity sum operation σ.

$$\overrightarrow{h'} = \sigma \sum_{k_{j \in N_i}} \left(\alpha_{ij} W\overrightarrow{h_j}\right)$$

This layer outputs a new set of node features in the form of (potentially, with a different cardinality F'):

$$h' = \left\{\overrightarrow{h'_1}, \overrightarrow{h'_2}, ..., \overrightarrow{h'_N}\right\}$$

5) Next-item recommendation layer: The final embedding vector with final scores are used to presents a ranked list of next-item recommendations.

V. EXPERIMENTS

A. Datasets

Our experiments have been conducted on three publicly available datasets from movies and e-commerce domains. Table II gives statistics about these datasets.

[2] https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset
[3] https://www.kaggle.com/datasets/retailrocket/ecommerce-dataset
[4] https://www.kaggle.com/datasets/phhasian0710/yoochoose

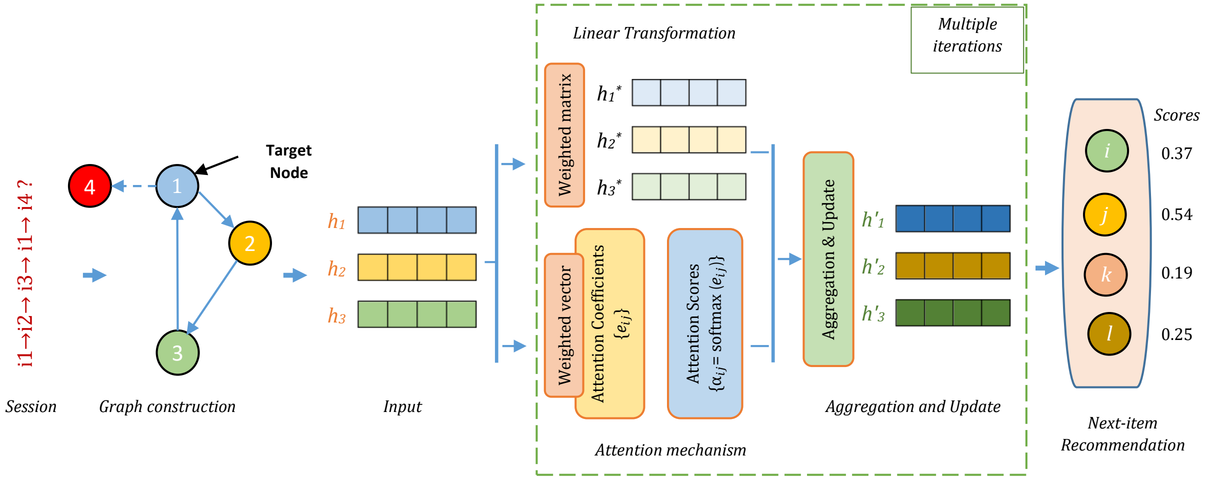

Figure 2. Overview on the proposed GAT-based model for SBRS.

Table II
CHARACTERISTICS OF THE DATASETS

	MovieLens[2]	RetailRocket[3]	YooChoose[4]
Rows	100226	35309	74708
Sessions	22202	3636	14514
items	9723	15144	9255

B. Evaluation metrics

The performances of the proposed model and compared baselines are evaluated according to the following metrics:

- Accuracy: it reprsents a highly intuitive metric. The percentage of cases that were successfully predicted out of all instances is used to calculate accuracy.

$$Accuracy = \frac{Number\,of\,correct\,predictions}{Total\,number\,of\,predictions}$$

- Precision: it measures the relevant retrieved recommendations by calculating the proportion of correctly recommended items to the total of recommended items.

$$Precision\,(P) = \frac{Correctly\,recommended\,items}{Total\,recommended\,items}$$

- Recall (a.k.a Sensitivity or True Positive Rate): it evaluates the retrieved relevant recommendations by the fraction of correctly recommended items that are also part of the collection of useful recommended items.

$$Recall\,(R) = \frac{Correctly\,recommended\,items}{Total\,useful\,recommended\,items}$$

- F1-Score (a.k.a F-measure): it is used to evaluate the overall performance by giving a balanced measurement between the precision and recall.

$$F1 - Score = \frac{2PR}{(P+R)}$$

C. Baselines

We have compared our proposed model with the following standard baselines:

1) POP: is a simple recommendation approach that suggests items to users based on their overall popularity or frequency of occurrence in a dataset.
2) S-POP: in SBRS scope, S-Pop focuses on recommending items that are popular within the current session.
3) Random recommendations: it is a straightforward approach where recommendations are generated randomly without considering historical data, user preferences or item features.

D. Loss function

An essential aspect of determining the effectiveness of SBRS with GAT is the selection of the loss function.

Cross-Entropy Loss is frequently employed in architectural work to evaluate the difference between the probabilities of the predicted class and the target one. In this case, the predicted class probabilities are derived from the GAT model's output, whereas the target class is the subsequent item in each session. This loss function works by computing the negative logarithmic-likelihood of the anticipated class probabilities concerning the target class. It is defined with the following equation:

$$L_{CE} = -\sum_{i=1}^{n} t_i log(p_i)$$

for n classes, where t_i is the truth label and p_i is the softmax probability for the i^{th} class. While, the hyper-parameters can include:

In_features: The input feature dimension of each node in the graph. In this case, it is set to 1 because node characteristics are represented by the position of each element in the session.

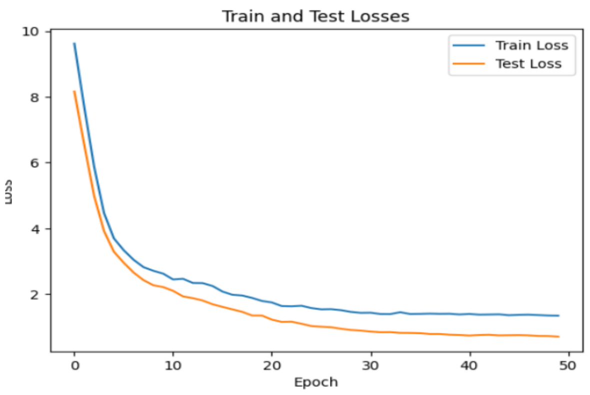

Figure 3. Train and test losses for RetailRocket dataset.

Table III
ACCURACY OF THE PROPOSED GAT MODEL

Epochs	MovieLens	RetailRocket	YooChoose
100	0.5751	0.7923	0.6158
200	0.5743	0.7933	0.6165

Table IV
COMPARISON OF THE GAT-BASED SBRS MODEL WITH BASELINES

	RetailRocket			YooChoose		
Baseline	P	R	F1	P	R	F1
POP	0.4540	0.6462	0.5333	0.5343	0.9241	0.6771
S-POP	0.0792	0.2229	0.1169	0.1107	0.1944	0.1410
Random Recom.	0.0721	0.2920	0.1156	0.1014	0.1842	0.1308
GAT Model	0.7084	0.7807	0.7428	0.9505	0.6036	0.7384

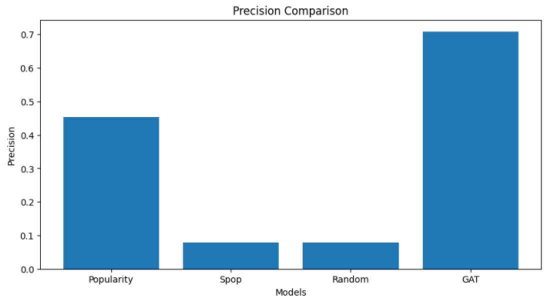

Figure 4. Performance comparison with precision metric on RetailRocket.

Out_features: The output feature dimension of each node after applying GAT convolution. In this code, it is set to 100.

Num_heads: The number of attention heads used in GAT convolution. It controls how often the attention mechanism is applied. The code uses 8 attention heads.

Dropout: The dropout rate applied after GAT convolution to prevent overfitting. Setting it to 0.1 means that 10% of node features are randomly set to zero during training.

Fig. 3 depicts the losses in training and test of our GAT model on the RetailRocket dataset.

VI. RESULTS AND DISCUSSION

Table III shows the results of the proposed GAT-based SBRS model on the aforementioned evaluation datasets using the accuracy metric. Several architectures were examined and a single layer of GAT was found to be the best performer.

A. *Comparing Results with Baselines*

Due to the low precision of the Movielens dataset with our proposed model, as shown in Table III, only the other two datasets will be considered to allow comparisons between the three basic lines with the results of the SBRS model based on metrics (precision, recall and F1-score). Table IV displays that our model dominates other methods.

Fig. 4 and 5 vividly illustrate some performance differences summarized in Table IV.

B. *Results discussion*

The results demonstrate the superiority of our GAT-based SBRS model over the baselines. Its performance improvements can be attributed to several factors. mainly: (i) the GAT architecture employed in this model has the ability to capture complex relationships and dependencies among items in sessions, (ii) By leveraging attention mechanisms [8], the model assigns higher weights to important items in the session, thereby enabling more accurate recommendations. This is particularly beneficial in session-based recommendation scenarios where user preferences can change dynamically.

VII. CONCLUSION

This paper highlights the effectiveness of using graph attention networks for the developpement of session-based recommendation systems. The proposed GAT-based SBRS model demonstrates a superior performance compared to some baselines which clearly validates its potential contribution in this recommendation type. The practical implications of accurate session-based recommendations emphasize the value of our model in improving user experience and driving business outcomes. However, further research advancements and comparisons in this area will continue to refine and extend the capabilities of GAT for SBRSs. Our nearest future research will be on evaluating the model on additional datasets to

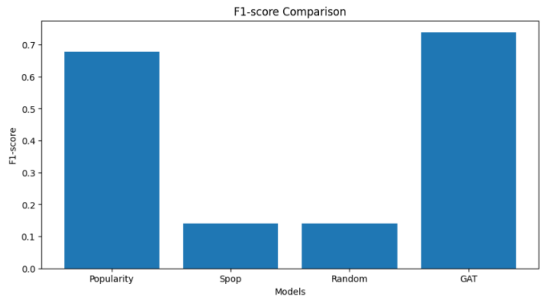

Figure 5. Performance comparison with F1-score metric on YooChoose.

validate its performance across different contexts with other baselines and related work.

REFERENCES

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, jul 2013.

[2] D. Jannach, B. Mobasher, and S. Berkovsky, "Research directions in session-based and sequential recommendation," *User Modeling and User-Adapted Interaction*, vol. 30, pp. 609–616, aug 2020.

[3] M. Quadrana, P. Cremonesi, and D. Jannach, "Sequence-aware recommender systems," *ACM Computing Surveys*, vol. 51, pp. 1–36, jul 2018.

[4] 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*, vol. 54, pp. 1–38, jul 2021.

[5] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, "Graph neural networks in recommender systems: A survey," *ACM Computing Surveys*, vol. 55, pp. 1–37, dec 2022.

[6] C. Gao, Y. Zheng, N. Li, Y. Li, Y. Qin, J. Piao, Y. Quan, J. Chang, D. Jin, X. He, and Y. Li, "A survey of graph neural networks for recommender systems: Challenges, methods, and directions," *ACM Transactions on Recommender Systems*, vol. 1, pp. 1–51, mar 2023.

[7] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *International Conference on Learning Representations*, 2018.

[8] C. Sun, C. Li, X. Lin, T. Zheng, F. Meng, X. Rui, and Z. Wang, "Attention-based graph neural networks: a survey," *Artificial Intelligence Review*, aug 2023.

[9] Y. Xie, Z. Li, T. Qin, F. Tseng, K. Johannes, S. Qiu, and Y. L. Murphey, "Personalized session-based recommendation using graph attention networks," in *2021 International Joint Conference on Neural Networks (IJCNN)*, IEEE, jul 2021.

[10] Q. Zhang, Z. Xu, H. Liu, and Y. Tang, "KGAT-SR: Knowledge-enhanced graph attention network for session-based recommendation," in *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, nov 2021.

[11] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 346–353, jul 2019.

[12] F. Yu, Y. Zhu, Q. Liu, S. Wu, L. Wang, and T. Tan, "TAGNN: Target attentive graph neural networks for session-based recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, jul 2020.

[13] L. Yin, P. Chen, and G. Zheng, "Session-enhanced graph neural network recommendation model (SE-GNNRM)," *Applied Sciences*, vol. 12, p. 4314, apr 2022.

[14] J. Chen, G. Zhu, H. Hou, C. Yuan, and Y. Huang, "AutoGSR," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, jul 2022.

[15] C. Liu, Y. Li, H. Lin, and C. Zhang, "GNNRec: gated graph neural network for session-based social recommendation model," *Journal of Intelligent Information Systems*, vol. 60, pp. 137–156, aug 2022.

[16] L. Dong, G. Zhu, Y. Wang, Y. Li, J. Duan, and M. Sun, "A graph positional attention network for session-based recommendation," *IEEE Access*, vol. 11, pp. 7564–7573, 2023.