Graph Neural Networks for Session-based Recommender Systems: A Brief Review*

Boudjemaa Boudaa[1,*,†], Abdelhafid Abouaissa[2,†] and Mouloud Amine Djenane[1,†]

[1]Computer Science Department, University of Tiaret, Tiaret, 14000, Algeria
[2]IRIMAS, University of Haute-Alsace, Mulhouse, 68093, France

Abstract
Over the past few years, session-based recommender systems (SBRSs) have garnered growing attention from both researchers and industry professionals. Diverging from conventional recommender systems, SBRSs rely on user-item interactions occurring within the current session to forecast recommendations for the next item. Academic literature has introduced numerous methods for developing SBRSs, including those that leverage Graph Neural Networks (GNNs). Due to their ability to handle intricate relationships such as in social network analysis, GNN has demonstrated their effectiveness in comparison to alternative approaches. This paper provides a first review of the prominent SBRSs based on GNNs and sheds some light on further research. This initiative seeks to support researchers by providing valuable insights into the latest advancements and prospects of GNNs, with the overarching goal of enhancing session-based recommender systems.

Keywords
Session-Based Recommender System, Graph Neural Network, Next-Item Recommendation.

1. Introduction

Sequence-aware recommender systems (SARS) [1] are designed to consider the chronological sequence of user interactions or activities when generating personalized recommendations. In contrast to conventional recommender systems that treat user-item interactions as separate events, SARS analyse the temporal patterns and sequences of user actions. These systems are particularly relevant in scenarios where the order of user interactions is crucial in understanding user preferences and intents.

As a sort of SARS, a Session-Based Recommender System (SBRS) [2, 3] focuses on providing personalized recommendations based on short-term interactions within a user's current session. Rather than relying on the entirety of a user's historical behaviour, SBRSs concentrate on the user's immediate context and preferences within a single session.

Session-based recommender systems are commonly used in various applications [4], including e-commerce, news websites, and online advertising, where user interactions occur in real time, and personalized recommendations can greatly enhance user engagement and satisfaction.

In the existing literature, various approaches are proposed in the context of SBRS [2, 5]. Notably, there has been a growing trend of exploring the integration of Graph Neural Networks (GNN) in the advancement of SBRS [2, 6, 7, 8].

GNN is a type of artificial neural network designed to work with graph-structured data [9]. Graphs are composed of nodes linked by edges, and GNNs are custom-designed to process and analyse data represented in this structure [6, 7]. GNNs have gained significant popularity in various domains, including social network analysis, recommendation systems, molecular chemistry, and natural language processing [10]. They demonstrate exceptional proficiency in tasks requiring the comprehension of relationships and interdependencies among interconnected entities within a graph.

In the absence of similar work, this paper aims to review the main proposed approaches of SBRS based on GNN. However, its contributions can be summarized in the following:

1. Exhibit the use of different GNN architectures for building SBRS,
2. Review of GNN-Based approaches for SBRS,
3. and, highlighting some open research issues for future directions on SBRS.

Regarding the remainder of the paper, its structure will be as follows: In Section 2, our attention will be towards providing comprehensive introductions to SBRS and GNN. Section 3 exhibits the SBRS design using GNN and the main implementation guidelines. The current GNN's architectures in SBRS are detailed in Section 4. In Section 5, a categorized review of proposed GNN-based SBRS

6th International Hybrid Conference On Informatics And Applied Mathematics, December 6-7, 2023 Guelma, Algeria
*Corresponding author.
†These authors contributed equally.
boudjemaa.boudaa@univ-tiaret.dz (B. Boudaa); abdelhafid.abouaissa@uha.fr (A. Abouaissa); mouloudamine.djenane@univ-tiaret.dz (M. A. Djenane)
0000-0002-1356-4775 (B. Boudaa); 0000-0003-0459-8081 (A. Abouaissa); 0009-0002-0176-3977 (M. A. Djenane)

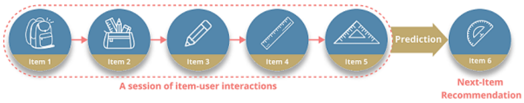

Figure 1: A Shopping Session example on school supplies.

approaches is given. Section 6 emphasizes certain emerging future directions on SBRS mainly with GNN. Finally, Section 7 concludes this review paper and unveils our nearest research.

2. Preliminaries

To introduce the current work, this section provides a theoretical background with some details on SBRS and GNN.

2.1. Session-Based Recommender Systems

Within the expanding landscape of recommender systems, session-based recommender systems (SBRS) are emerging as an intriguing trend [11]. SBRSs fall under the category of sequence-aware recommendation systems (SARS) [1] and specifically consider short-term user preferences or intentions. They make use of user interactions with previously viewed items during the current session to anticipate what item or items might be of interest next. This could involve predicting the next video to watch, the point of interest (POI) to visit, or the product to purchase. A crucial element of their recommendation process involves tracking the relationships and dependencies between time-stamped interactions of users with items during a session.

Currently, SBRSs are gaining considerable attention from both researchers and industry professionals [2]. They are becoming increasingly viable recommendation tools across a wide spectrum of real-world domains, including e-commerce, tourism, and leisure. For instance, as illustrated in Fig. 1, SBRSs can provide suggestions for future purchases based on a series of earlier user activities within the same online shopping session.

The primary attributes of SBRSs [2, 12, 3] are, (i) Identification of users is not feasible (because the majority of website traffic is from first-time or non-logged-in users), (ii) due to (i), Long-term user preferences information are unavailable, and (iii) User preferences must be deduced based on a limited set of consecutive interactions within the session.

Formally, a session s is a non-empty bounded observations list of user-item interactions o_i generated over a continuous timespan that may be linked with a particular user via some form of identification (e.g., user ID, session-ID, cookie). It can be represented as : $s = \{o_1, o_2, \ldots, o_{|s|}\}$.

In the literature, a multitude of methods and models have been proposed for the development of SBRSs [2]. Among them, Graph Neural Networks (GNNs) stand out as a remarkable method that offers enhanced performance for SBRSs. GNNs have showcased their capacity to model intricate transitions occurring within and between sessions, treating them as graph-structured data and employing deep neural networks. The following section will provide an introduction to GNNs.

2.2. Graph Neural Networks

Graph neural networks (GNNs) are a novel advanced model for employing deep learning techniques that operate on graph-structured data, as for social networks (relationships domain), citation networks (scientific domain), protein-protein interaction networks (biology domain), user-item interaction networks (recommendation domain), and others [10]. In contrast to traditional neural networks, which operate on structured input data like images or sequences, GNNs are designed to operate on non-Euclidean data, where data points are represented as nodes in a graph and the edges connecting nodes symbolize the connections and associations between them.

Over the past few years, GNNs have become popular because of their handling of complex and highly interconnected data structures. They use message-passing algorithms to propagate information through the graph, updating node representations based on the aggregated representations of their neighbouring nodes. GNNs can be used for a wide range of tasks, including node prediction, link prediction, and graph classification [13].

Recently, the recommendation field has benefited from GNNs for enhancing its systems [14, 6]. In the context of SBRSs, several approaches based on GNN have been proposed by incorporating various GNN architectures for effective proposals [2]. Using GNN have improved recommendation accuracy by modelling sequential interactions between users and items. Generally, graphs for SBRSs are directed in such a way that they maintain the order in which user-item interactions occur (see Fig. 2).

In the absence of a comprehensive survey paper on Session-Based Recommender Systems (SBRS) utilising Graph Neural Networks (GNNs), this work serves this purpose by conducting a review of the most significant contributions in this field.

3. How to Use GNN for SBRS

From a technical perspective, this section offers a reference framework which serves as a core model for any

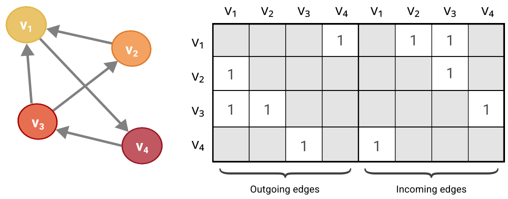

Figure 2: A Directed Graph with Adjacency Matrix of Sequential Data in SBRS.

GNN-Based SBRS proposal. In addition, the main implementation instructions in this direction were mentioned.

3.1. Overall Framework for GNN-Based SBRS

The literature in the last years has witnessed an increasing adoption of GNN for developing SBRSs via different proposed frameworks and architectures. Fig. 3 depicts a reference architecture for SBRS based on GNN. This showcases three main steps:

1. Graph Construction. To apply GNN in the sequential recommendation, it is essential to convert sequential data into a sequential graph format in which nodes and edges can be featured with information vectors as their hidden states.
2. Information Propagation. To capture the transition patterns, a propagation mechanism should be adopted. Generally, each GNN layer is designed to execute a specific set of operations on each node within the graph, namely:
    - Message Passing: this is defined as the procedure of collecting the features of neighbouring nodes, transforming them, and then transmitting them to the source node. In parallel, this process is repeated for every node in the graph, thereby ensuring a comprehensive examination of all neighbourhoods by the conclusion of this step.
    - Aggregation: neighbourhood aggregation involves the exchange of data between nodes within their respective neighbourhoods. A source node, equipped with its initial embeddings, receives input from its neighbours, and this information is transmitted through edge neural networks.
    - Update: upon receiving these messages, each node updates its features by considering both its current features and the aggregated information received from its neighbouring nodes.
3. Recommendation task. In GNN scope, the prediction can be allowed on node, edge or graph, and using a scoring and ranking function to recommend appropriate next-item for SBRSs.

3.2. Implementation Guidelines for GNN-Based SBRS

From a practical standpoint, the session-based recommender systems development process with GNN can be generally carried out using the main following activities:

1. Data Representation: Arrange your data into sessions, where each session denotes a series of user engagements (e.g., clicks, views, purchases) with items. Create a graph structure in which items are depicted as nodes, and the connections between items, like co-occurrences, similarities, or pertinent associations, are represented as edges. Encode both session-specific details and item attributes as features for the nodes in the graph.
2. Graph Neural Network Architecture: Select an appropriate Graph Neural Network (GNN) architecture for session-based recommendation (see Section 4). Customize the GNN to effectively manage sequential data and dynamic graph structures.
3. Session Representation: Leverage the GNN to acquire session representations through the aggregation of information from items within a session. Achieve this by executing message-passing across the graph while considering the relationships between items. Additionally, contemplate the integration of time decay or attention mechanisms to assign greater significance to recent interactions.
4. Item Embeddings: Derive item embeddings by employing the GNN individually to process each item within the graph, effectively capturing their interconnections with other items within the context of sessions.
5. Candidate Generation: Using the acquired session and item representations, produce a roster of potential items for recommendation. This can be accomplished by scoring items with a function that takes into account their pertinence to the context of the session.
6. Loss Function and Training: Establish a suitable loss function for your recommendation objective, such as the Bayesian Personalized Ranking (BPR) loss or Triplet loss, which incentivizes the model to prioritize positive items over negative ones in ranking. Proceed to train the GNN-based recommender system using session-level data and historical user-item interactions.

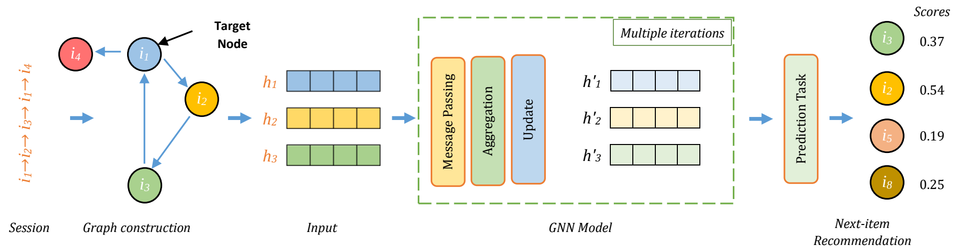

Figure 3: An Overall Framework for GNN-based SBRS.

7. Evaluation: Assess the recommendation system's performance by employing metrics such as Recall, Precision, NDCG (Normalized Discounted Cumulative Gain), or Hit Rate. To gauge the model's ability to generalize effectively, partition your data into training, validation, and test sets for evaluation purposes.
8. Regularization and Hyper-parameter Tuning: Implement regularization methods like dropout or L2 regularization to counteract over-fitting. Conduct experiments to fine-tune hyper-parameters, including learning rate, the number of layers, hidden dimensions, and the number of epochs, to optimize the model's performance.

It is worth emphasizing that the decision regarding the GNN architecture and hyper-parameters should be thoughtfully tailored to suit the unique characteristics of your dataset and the specific requirements of your problem domain. The optimal choices may vary based on factors such as data distribution, the nature of user interactions, and the desired level of recommendation accuracy. Therefore, a thorough understanding of your specific context is crucial in making these decisions.

4. Current GNN Architectures in SBRS

GNN refers to a variety of different architecture models, including mainly gated graph neural networks (GGNNs) [15], graph convolutional networks (GCNs) [16], and graph attention networks (GATs) [17]. GGNNs use recurrent neural networks with GRU (Gated Recurrent Unit) cells to propagate information through the graph. While GCNs are one of the most commonly used GNN architectures and are based on a convolutional-like operation that aggregates information from neighbouring nodes. However, GATs use attention mechanisms to weigh the importance of each neighbouring node when updating a node's representation. These three architectures are intensively used in the SBRS field [2] to handle complex user-item interactions and provide accurate next-item recommendations. Formally, this section will separately explain the utilization of each architectural design in the SBRS context.

4.1. Gated Graph Neural Networks

Graph Gated Neural Networks (GGCNs) [18] are an extension of Scarselli et al.'s previous work on GNN [19]. They use gated recurrent units (GRUs) and compute gradients using the back-propagation through time (BPTT) algorithm.

Let $G = (V, E)$ be a directed graph, with V and E representing the set of vertices and edges respectively. An edge is a tuple that contains the source and target nodes (cf. Fig. 2). The connections between the nodes are better stated in an adjacency matrix $A \in \mathbb{R}^{|V|\times 2|V|}$, which is a matrix structured with its rows and columns corresponding to the labelled vertices in the graph and a 0 or 1 in position (v_i, v_j) indicating the nature of adjacency between v_i and v_j (Fig. 2).

Typically, an SBRS approach based on GGNN [2] first builds a directed graph from all the previously sorted items within a session, with the direction of each edge denoting the sequence of subsequent interactions. Then, GGNN processes the session graph successively to produce the embedding $\boldsymbol{n}_i$ of node n_i, specifically the embedding of the appropriate interaction o_i. At the end of the process, the embeddings of all interactions are acquired, which are subsequently utilized to construct an embedding of the session context in order to develop recommendations for that session. A GRU is specifically utilised in GGNN to learn each node's embedding by updating the embedding recurrently. In particular, the embedding (or hidden state) $\boldsymbol{h}_i^t$ of node n_i at step t is updated (eq. 1) by the preceding hidden state of itself and its neighbourhood nodes, i.e., $\boldsymbol{h}_i^{(t-1)}$ and $\boldsymbol{h}_j^{(t-1)}$,

$$\boldsymbol{h}_i^t = GRU\left(\boldsymbol{h}_i^{(t-1)}, \sum_{n_j \in N(n_i)} \boldsymbol{h}_j^{(t-1)}, \boldsymbol{A}\right) \quad (1)$$

where $N(n_i)$ is the set of neighbourhood nodes of n_i in the session graph, and A is the adjacency matrix constructed based on the session graph. Upon reaching a stable equilibrium over several iterations, the hidden state at the final step of node n_i is considered as its embedding $\boldsymbol{n}_i$.

4.2. Graph Convolutional Networks

Graph convolutional networks (GCNs) have established themselves as one of the most widely adopted and highly effective GNN models. The convolution principle is borrowed from CNN [16] enabling GCNs to aggregate data based on local graph neighbourhoods [9].

GCN-based SBRSs primarily employ the pooling operation to incorporate information from node n_i's neighbourhood node n_j in the graph, and then to proceed to update the hidden state of n_i as formulated by equations 2, 3:

english

$$\hat{\boldsymbol{h}}_i^t = pooling\left(\left\{\boldsymbol{h}_j^{(t-1)}, n_j \in N(n_i)\right\}\right) \quad (2)$$

britishwhere $N(n_i)$ is the set of neighbourhood nodes of node n_i. Various particular pooling techniques (such as mean pooling and max pooling) can be employed, contingent on particular situations. Subsequently, the combined neighbourhood information can be integrated into the iterative update of the node's hidden state n_i:

english

$$\boldsymbol{h}_i^t = \boldsymbol{h}_i^{(t-1)} + \hat{\boldsymbol{h}}_i^t \quad (3)$$

Ultimately, once a stable equilibrium is attained, the final hidden state of node britishn_i is adopted (eq. 3) as its embedding british$\boldsymbol{n}_i$. british

4.3. Graph ATtention Networks

Graph ATtention Networks (GATs) utilize an attention mechanism to measure the significance of nearby nodes' features while updating a node's representation. The attention mechanism allows GATs to effectively model intricate relationships between nodes in a graph. Every node in a GAT has a hidden representation (features) that is updated depending on the neighbour representations using an attention mechanism [17].

The fundamental principle of GATs is to learn an attention coefficient for each nearby node that specifies how much weight to give to that node's representation while updating the target node's representation. The attention coefficients are learned using a single-layer neural network that takes as input the features of the target node as well as the features of its neighbouring nodes. A softmax function is used to calculate the attention coefficients, ensuring that the weights given to each neighbouring node add up to one.

Moreover, the GAT computes numerous sets of weights for every node using multiple attention heads, which aids in capturing 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 weight matrix. Each attention head's output is combined and sent through a non-linear activation function.

Formally, equation 4 summarizes the key operation of GAT in which the general module ***attention*** can be designated to different operations including self-attention, multi-head attention, etc.

english

$$\boldsymbol{h}_i^t = attention\left(\left\{\boldsymbol{h}_j^{(t-1)}, n_j \in N(n_i)\right\}\right) \quad (4)$$

britishIn essence, the operations within $attention$ can be partitioned into two stages: (1) computing the significance weights for each neighbouring node, and (2) aggregating the hidden states of neighbouring nodes based on their significance weights.

Finally, after the completion of forward propagation across multiple attention layers, the hidden state of each node n_i within a session graph at the final layer is adopted as its embedding $\boldsymbol{n}_i$.

5. Literature Review

Using the classification reported in [2], this section presents the state-of-the-art methods for developing session-based recommendations that have used GNN with one or more mentioned architectures above. Three research categories can be distinguished, namely:

5.1. Session graph modelling

Initially, SBRS only used directed graphs to model each session sequence. SR-GNN [20] is the pioneer model that has utilized GNN in SBRS development for obtaining the latent vectors of nodes in each session graph. It is based on GGNN architecture including an attention mechanism layer. Aside from capturing the complex structure and transitions between items within a session, SR-GNN employs a strategy that combines current interests with long-term preferences to predict next behaviours of users. Another work in [21] proposed a graph contextualized self-attention network (GC-SAN) which begins by constructing dynamic directed graphs from session sequences and then applies GGNN to pick up local dependencies between items including contextual information of sessions. Next, a self-attention network (SAN) has been used to capture global dependencies between input and output sequences without regard to their distances. Finally, the local short-term and global self-attended relationships are combined for the session representation.

In [22], the authors propose GACOforREC model for session-based recommendation in order to handle the long-term and short-term preferences of users and preserve the hierarchy of potential preferences. This model has used convolution operations of GCN to understand the sequence within the session and the spatial characteristics within the network for capturing the user's short-term preferences. For learning long-term preferences, it applied ConvLSTM, a variant of the LSTM network. In addition, GACOforREC proposes a new pair adaptive attention mechanism (Long-Attention and Short-Attention) based on GCNs to give consideration to the impact of different propagation distances of GCNs. To improve the model's hierarchical learning of a variety of preferences, ON-LSTM has been introduced, it is a network structure that focuses more on hierarchy and neuron ordering. This ordering is crucial to the comprehensive perception of the model's user preferences for accurate recommendations. Another GCN-based SBRS Model called AUTOMATE was presented in [23]. It integrates a graph convolutional layer based on Auto-Regressive Moving Average (ARMA) filters. It can capture complex transformations between items through sessions modelled as graph-structured data. The core principle behind AUTOMATE revolves around leveraging the ARMAConv layer, which allows us to merge enduring user preferences with real-time session interests to generate the graph transfer signal.

Thereafter, TAGNN [24] 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. In [25], 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 [26]. 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. While the node features and graph topology in the graph are used to generate appropriate session embedding for recommending the next item.

Just recently and innovatively, the authors in [27] 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 (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. As part of AutoGSR, learning items are associated with meta-knowledge that contains comprehensive session information.

5.2. Graph structure enrichment

Other methods in the field of SBRSs aim to enhance the relationships and information within the session graph by incorporating data from other sessions or supplementary sources. A-PGNN model [28] is proposed to capture firstly the complex dependencies between the session's items, and at the same time, the user's long-term performance by modelling the effect of historical sessions on the ongoing session. In addition to the attention mechanism and Transformer network, A-PGNN has used GGNN with closed recurrent units (GRUs) to update information about each node's hidden state. SGNN-HN [29] applies a star graph neural network (SGNN) to model the complex transition relationship between items (adjacent and non-adjacent) in a current session to generate accurate item embeddings. To represent the propagation information, GGNN is used to feed satellite nodes in the star graph. In this work, the over-fitting problem of GNN in SBRS is tackled using highway networks (HN) which can dynamically merge information from item embeddings before and after multi-layer SGNNs. Subsequently, the generated item embeddings are aggregated through an attention mechanism to represent a user's final preference which is then combined with her recent interest expressed (i.e., last clicked items in the session) for next-items prediction. In [30], the authors propose a position-aware gated graph attention network (PA-GGAN) model to assign the position information of items in the session sequences into session graphs. The enhanced GGNN that underpins this model has been supplied with a self-attention mechanism for aggregating features from nodes. Similarly, [31] proposes MKM-SR which incorporates user Micro-behaviours and item Knowledge simultaneously into Multi-task learning for Session-based Recommendation. Instead of a sequence of items, a session in MKM-SR is modelled on the micro-behaviour level accompanied by a sequence of operations on each item to sufficiently capture the transition pattern in the session. GGNN and GRU are used to learn item and operation embeddings, and the multi-task learning processes involve learning knowledge embeddings to promote the major task of SBRS. Otherwise, to add other information sources (such as social information) in session-based recommendation processes, the authors in [32] propose a novel session-based social recommendation model named GNNRec, in which a gated graph neural network (GGNN) is first used to represent the current session information of the user. Next, a GAT is utilized to aggregate social information on users and their friends on social networks for modelling user's interests. However, to make session-based social recommendations more efficiently, the research work in [33] has proposed a model called Social-aware Efficient Recommender (SERec) that implements the SEFrame framework in which a heterogeneous knowledge graph is constructed from the social network and historical user behaviours. In this study, a GGNN is utilized to derive a contextualized feature vector by integrating information from neighbouring nodes and the initial feature vectors.

Recently, and based on GNN and attention networks, a session-enhanced graph neural network (SE-GNNRM) model has been proposed in [34]. During the encoding phase of this model, the intricate transitional relationships between items and item features are captured in GNN and SAN. Then, the attention mechanism is employed to combine short-term and long-term preferences to construct a global session graph. Furthermore, a GAT is devoted to recognising features between similar sessions and integrating similarity information between sessions in which GGNNs are used to extract node features.

5.3. High-order relation

In order to capture the complex high-order information between items in real-world scenarios, the authors in [35] have proposed DHCN (Dual Channel Hyper-graph Convolutional Networks). This model is based on a hypergraph using convolution operations and integrating self-supervised learning to generate high-quality session-based recommendations.

In order to acquire more refined item representations, GNN-based models can extract additional information from high-order neighbours over the graph structure. New methods aim to enhance the recommendation by modelling the high-order relations in session data. Very recently, Disen-GNN [36] brought up the problem of anonymous user purpose in a session and demonstrated the promise of using disengaged learning to solve this problem. This model captures the session's intent by considering factor-level attention for each item in the session, and it employs a disentangled learning technique to transform item embeddings into multiple-factor embeddings. The embedding of each factor is learned separately via GGNN based on the item adjacent similarity matrix computed for each factor. Additionally, the distance correlation is employed as a means of enhancing the independence between each pair of factors. As each item is represented with independent factors, an attention mechanism is designed to determine the user's intent regarding the different factors of each item. Subsequently, the session embedding is computed by aggregating all the items that have been assigned factor-level attention. The user intents at the factor level are taken into account when determining the purpose of a session. As a solution for the over-smoothing issue associated with session-based recommendation using GNN (i.e., all nodes reach the same value), SR-HGNN [37] is proposed. It is based a hybrid-order GGNN, where insignificant patterns are avoided and complex interactions between items are captured. Furthermore, an attention mechanism is adopted to learn different weights of orders in the propagation.

Very recently, GPAN or Graph Positional Attention Network has been presented in [38]. 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. Still with the use of hyper-graphs, the study in [39] has introduced HyperS2Rec, where it takes into account both item consistency and sequential item dependence simultaneously. This model leverages hypergraph-structured data through HGCN and captures sequential information using GRU to collectively model user preferences. In this proposal, the reversed position embedding mechanism and soft attention mechanism are combined to derive session representations.

6. Discussion and Future Directions

Based on the literature review conducted in the preceding section, we can initiate a discussion on specific aspects and identify potential avenues for future research.

6.1. Discussion

Table 1 displays the primary GNN architecture employed in each reviewed study, notwithstanding the incorporation of additional layer types, such as attention mechanisms. In addition, Fig 4 shows an overview of the distribution of the usage of GNN architectures among the summarized approaches.

It is clear that GGNN architecture has been more applied than GCN and GAT to address various issues raised in the cited research works. Evenly, attention mechanisms represent another common point between most of the mentioned works.

Furthermore, to harness the capabilities of GNN architectures, some research work have undertaken the approach of amalgamating two or more GNN architectures.

Table 1
Summary of GNN-Based Approaches for SBRS.

Class	Approaches	GGNN	GCN	GAT
Session graph modelling	SR-GNN [20]	✓		
	GC-SAN [21]	✓		
	GACOforREC [22]		✓	
	AUTOMATE [23]		✓	
	TAGNN [24]	✓		
	PSR-GAT [25]			✓
	KGAT-SR [26]			✓
	AutoGSR [27]	✓		✓
Graph structure enrichment	A-PGNN model [28]	✓		
	SGNN-HN [29]	✓		
	PA-GGAN [30]	✓		
	MKM-SR [31]	✓		
	GNNRec [32]	✓		✓
	SERec [33]	✓		
	SE-GNNRM [34]	✓		✓
High-order relation	DHCN [35]		✓	
	Disen-GNN [36]	✓		
	SR-HGNN [37]	✓		
	GPAN [38]			✓
	HyperS2Rec [39]	✓	✓	✓

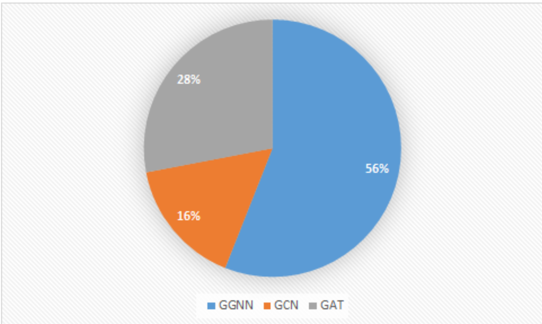

Figure 4: An overview of the distribution of GNN architectures' utilization.

However, researchers in this field have not fully embraced the utilization of graph convolutional networks (GCN), despite their adeptness in extracting patterns of user-item interactions.

Also, these techniques may not adequately capture intricate dependencies and still have poor ability against the cold start problem.

6.2. Future Directions

Graph Neural Networks (GNNs) serve as a potent instrument for examining and characterizing intricate data structures. Nonetheless, they continue to pose certain difficulties, such as issues related to scalability, and adaptability to novel graph instances. Additionally to what is mentioned in [2], we outline the main potential directions for future research and development in session-based recommender systems, including when developing them with advanced GNN-based models:

- Contextual Embeddings: SBRSs may benefit from incorporating additional contextual information such as user demographics, device information, or contextual data related to the session. This requires investigation of advanced methods for grasping and leveraging session context more effectively, including the application of contextual embeddings like transformer-based models (such as BERT and GPT) to gain a deeper comprehension of user intentions and preferences.
- Sequential Attention Mechanisms: Develop enhanced attention mechanisms that can handle long-term dependencies within sessions more effectively, potentially by incorporating memory-augmented neural networks or adaptive attention mechanisms.
- Hybrid Models: Explore hybrid recommender systems that merge session-based techniques with conventional collaborative filtering or content-based strategies, effectively harnessing both short-term and long-term user preferences.
- Transfer Learning: Examine transfer learning methodologies to utilize insights gained from one

domain or dataset, thereby enhancing recommendations in another domain, particularly in scenarios where session data is scarce (as with short sessions).

- Online Learning and Personalization: Develop online learning algorithms that can adapt and personalize recommendations in real-time as user preferences change during a session.
- Scalability: Optimize models and algorithms for scalability to handle large-scale session data efficiently, especially in platforms with a massive user base.
- Leveraging advanced GNN models: Ongoing research in GNNs is continually enhancing the field with emerging architectures, including attention-based graphs [40], Graph Auto-encoders (GAEs) [41], and Graph Transformer Networks [42]. SBRS stands to gain significant advantages from these advancements in GNN architecture, resulting in increased effectiveness and accuracy in next-item recommendations.

These future directions can help session-based recommender systems become more accurate, personalized, and capable of meeting the evolving needs and challenges of recommendation tasks.

7. Conclusion

The field of session-based recommender systems (SBRS) research is thriving, marked by a continuous stream of innovative techniques and emerging approaches. Among them, graph neural networks (GNN) have emerged as a powerful deep learning technique to perform inference on non-Euclidean data described by graphs. In this paper, we present contemporary GNN architectures and conduct a comprehensive review of prominent approaches that employ GNNs for advancing Session-Based Recommender Systems (SBRSs). Furthermore, we have elucidated potential research directions in this evolving field for future exploration.

Our forthcoming research will concentrate on comparing and evaluating the performance of these reviewed approaches.

References

[1] M. Quadrana, P. Cremonesi, D. Jannach, Sequence-aware recommender systems, ACM Comput. Surv. 51 (2018). URL: https://doi.org/10.1145/3190616. doi:10.1145/3190616.

[2] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, D. Lian, A survey on session-based recommender systems, ACM Comput. Surv. 54 (2021). URL: https://doi.org/10.1145/3465401. doi:10.1145/3465401.

[3] S. Wang, Q. Zhang, L. Hu, X. Zhang, Y. Wang, C. Aggarwal, Sequential/session-based recommendations: Challenges, approaches, applications and opportunities, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 3425–3428. URL: https://doi.org/10.1145/3477495.3532685. doi:10.1145/3477495.3532685.

[4] D. Jannach, M. Ludewig, L. Lerche, Session-based item recommendation in e-commerce: On short-term intents, reminders, trends and discounts, User Modeling and User-Adapted Interaction 27 (2017) 351–392. URL: https://doi.org/10.1007/s11257-017-9194-1. doi:10.1007/s11257-017-9194-1.

[5] M. Ludewig, N. Mauro, S. Latifi, D. Jannach, Empirical analysis of session-based recommendation algorithms: A comparison of neural and non-neural approaches, User Modeling and User-Adapted Interaction 31 (2021) 149–181. URL: https://doi.org/10.1007/s11257-020-09277-1. doi:10.1007/s11257-020-09277-1.

[6] S. Wu, F. Sun, W. Zhang, X. Xie, B. Cui, Graph neural networks in recommender systems: A survey 55 (2022). URL: https://doi.org/10.1145/3535101. doi:10.1145/3535101.

[7] Z. Sheng, T. Zhang, Y. Zhang, S. Gao, Enhanced graph neural network for session-based recommendation, Expert Systems with Applications 213 (2023) 118887. URL: https://www.sciencedirect.com/science/article/pii/S0957417422019054. doi:https://doi.org/10.1016/j.eswa.2022.118887.

[8] S. Wang, Y. Wang, F. Sivrikaya, S. Albayrak, V. W. Anelli, Data science for next-generation recommender systems, International Journal of Data Science and Analytics 16 (2023) 135–145. URL: https://doi.org/10.1007%2Fs41060-023-00404-w. doi:10.1007/s41060-023-00404-w.

[9] I. Ullah, M. Manzo, M. Shah, M. G. Madden, Graph convolutional networks: analysis, improvements and results, Applied Intelligence 52 (2022) 9033–9044. URL: https://doi.org/10.1007/s10489-021-02973-4. doi:10.1007/s10489-021-02973-4.

[10] B. Boudaa, I. Bestani, N. Benadjrouda, Graph convolutional networks for designing collaborative filtering-based health recommender systems, in: 2022 2nd International Conference on New Technologies of Information and Communication (NTIC), IEEE, 2022. URL: https://doi.org/10.1109%2Fntic55069.2022.10100590. doi:10.1109/

ntic55069.2022.10100590.

[11] D. Jannach, B. Mobasher, S. Berkovsky, Research directions in session-based and sequential recommendation, User Modeling and User-Adapted Interaction 30 (2020) 609–616. URL: https://doi.org/10.1007/s11257-020-09274-4. doi:10.1007/s11257-020-09274-4.

[12] W. Song, S. Wang, Y. Wang, S. Wang, Next-item recommendations in short sessions, in: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 282–291. URL: https://doi.org/10.1145/3460231.3474238. doi:10.1145/3460231.3474238.

[13] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: A review of methods and applications, AI Open 1 (2020) 57–81. URL: https://www.sciencedirect.com/science/article/pii/S2666651021000012. doi:https://doi.org/10.1016/j.aiopen.2021.01.001.

[14] C. Gao, Y. Zheng, N. Li, Y. Li, Y. Qin, J. Piao, Y. Quan, J. Chang, D. Jin, X. He, Y. Li, A survey of graph neural networks for recommender systems: Challenges, methods, and directions, ACM Trans. Recomm. Syst. 1 (2023). URL: https://doi.org/10.1145/3568022. doi:10.1145/3568022.

[15] Y. Li, D. Tarlow, M. Brockschmidt, R. S. Zemel, Gated graph sequence neural networks, in: Y. Bengio, Y. LeCun (Eds.), 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, 2016. URL: http://arxiv.org/abs/1511.05493.

[16] S. Zhang, H. Tong, J. Xu, R. Maciejewski, Graph convolutional networks: a comprehensive review, Computational Social Networks 6 (2019) 11. URL: https://doi.org/10.1186/s40649-019-0069-y. doi:10.1186/s40649-019-0069-y.

[17] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, in: International Conference on Learning Representations, 2018. URL: https://openreview.net/forum?id=rJXMpikCZ.

[18] Y. Li, D. Tarlow, M. Brockschmidt, R. S. Zemel, Gated graph sequence neural networks, in: Y. Bengio, Y. LeCun (Eds.), 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, 2016. URL: http://arxiv.org/abs/1511.05493.

[19] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Transactions on Neural Networks 20 (2009) 61–80. doi:10.1109/TNN.2008.2005605.

[20] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, T. Tan, Session-based recommendation with graph neural networks, in: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'19/IAAI'19/EAAI'19, AAAI Press, 2019. URL: https://doi.org/10.1609/aaai.v33i01.3301346. doi:10.1609/aaai.v33i01.3301346.

[21] C. Xu, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, F. Zhuang, J. Fang, X. Zhou, Graph contextualized self-attention network for session-based recommendation, in: International Joint Conference on Artificial Intelligence, 2019.

[22] M. Zhang, Z. Yang, GACOforRec: Session-based graph convolutional neural networks recommendation model, IEEE Access 7 (2019) 114077–114085. URL: https://doi.org/10.1109%2Faccess.2019.2936461. doi:10.1109/access.2019.2936461.

[23] H. Wang, G. Xiao, N. Han, H. Chen, Session-based graph convolutional arma filter recommendation model, IEEE Access 8 (2020) 62053–62064. doi:10.1109/ACCESS.2020.2984039.

[24] F. Yu, Y. Zhu, Q. Liu, S. Wu, L. Wang, T. Tan, TAGNN: target attentive graph neural networks for session-based recommendation, CoRR abs/2005.02844 (2020). URL: https://arxiv.org/abs/2005.02844. arXiv:2005.02844.

[25] Y. Xie, Z. Li, T. Qin, F. Tseng, K. Johannes, S. Qiu, Y. L. Murphey, Personalized session-based recommendation using graph attention networks, in: 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021. URL: https://doi.org/10.1109%2Fijcnn52387.2021.9533533. doi:10.1109/ijcnn52387.2021.9533533.

[26] Q. Zhang, Z. Xu, H. Liu, 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, 2021. URL: https://doi.org/10.1109%2Fictai52525.2021.00164. doi:10.1109/ictai52525.2021.00164.

[27] J. Chen, G. Zhu, H. Hou, C. Yuan, Y. Huang, Autogsr: Neural architecture search for graph-based session recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 1694–1704. URL: https://doi.org/10.1145/3477495.3531940. doi:10.1145/3477495.3531940.

[28] M. Zhang, S. Wu, M. Gao, X. Jiang, K. Xu, L. Wang, Personalized graph neural networks with attention mechanism for session-aware recommenda-

tion, IEEE Transactions on Knowledge and Data Engineering 34 (2022) 3946–3957. doi:10.1109/TKDE.2020.3031329.

[29] Z. Pan, F. Cai, W. Chen, H. Chen, M. de Rijke, Star graph neural networks for session-based recommendation, in: Proceedings of the 29th ACM International Conference on Information and Knowledge Management, CIKM '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1195–1204. URL: https://doi.org/10.1145/3340531.3412014. doi:10.1145/3340531.3412014.

[30] J. Wang, Q. Xu, J. Lei, C. Lin, B. Xiao, Pa-ggan: Session-based recommendation with position-aware gated graph attention network, in: 2020 IEEE International Conference on Multimedia and Expo (ICME), 2020, pp. 1–6. doi:10.1109/ICME46284.2020.9102758.

[31] W. Meng, D. Yang, Y. Xiao, Incorporating user micro-behaviors and item knowledge into multi-task learning for session-based recommendation, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1091–1100. URL: https://doi.org/10.1145/3397271.3401098. doi:10.1145/3397271.3401098.

[32] C. Liu, Y. Li, H. Lin, C. Zhang, Gnnrec: gated graph neural network for session-based social recommendation model, Journal of Intelligent Information Systems (2022). URL: https://doi.org/10.1007/s10844-022-00733-5. doi:10.1007/s10844-022-00733-5.

[33] T. Chen, R. C.-W. Wong, An efficient and effective framework for session-based social recommendation, in: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 400–408. URL: https://doi.org/10.1145/3437963.3441792. doi:10.1145/3437963.3441792.

[34] L. Yin, P. Chen, G. Zheng, Session-enhanced graph neural network recommendation model (se-gnnrm), Applied Sciences 12 (2022). URL: https://www.mdpi.com/2076-3417/12/9/4314. doi:10.3390/app12094314.

[35] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, X. Zhang, Self-supervised hypergraph convolutional networks for session-based recommendation, Proceedings of the AAAI Conference on Artificial Intelligence 35 (2021) 4503–4511. URL: https://doi.org/10.1609%2Faaai.v35i5.16578. doi:10.1609/aaai.v35i5.16578.

[36] A. Li, Z. Cheng, F. Liu, Z. Gao, W. Guan, Y. Peng, Disentangled graph neural networks for session-based recommendation, ArXiv abs/2201.03482 (2022).

[37] Y.-H. Chen, L. Huang, C.-D. Wang, J.-H. Lai, Hybrid-order gated graph neural network for session-based recommendation, IEEE Transactions on Industrial Informatics 18 (2022) 1458–1467. doi:10.1109/TII.2021.3091435.

[38] L. Dong, G. Zhu, Y. Wang, Y. Li, J. Duan, M. Sun, A graph positional attention network for session-based recommendation, IEEE Access 11 (2023) 7564–7573. URL: https://doi.org/10.1109%2Faccess.2023.3235353. doi:10.1109/access.2023.3235353.

[39] C. Ding, Z. Zhao, et al., Session-based recommendation with hypergraph convolutional networks and sequential information embeddings, Expert Systems with Applications 223 (2023) 119875. doi:10.1016/j.eswa.2023.119875.

[40] C. Sun, C. Li, X. Lin, T. Zheng, F. Meng, X. Rui, Z. Wang, Attention-based graph neural networks: a survey, Artificial Intelligence Review (2023). URL: https://doi.org/10.1007%2Fs10462-023-10577-2. doi:10.1007/s10462-023-10577-2.

[41] M. Lin, K. Wen, X. Zhu, H. Zhao, X. Sun, Graph autoencoder with preserving node attribute similarity, Entropy 25 (2023) 567. URL: https://doi.org/10.3390%2Fe25040567. doi:10.3390/e25040567.

[42] L. Müller, M. Galkin, C. Morris, L. Rampasek, Attending to graph transformers, 2023. arXiv:2302.04181.