

A Survey on Side Information-driven Session-based Recommendation: From a Data-centric Perspective

Xiaokun Zhang, Bo Xu, Chenliang Li, Bowei He, Hongfei Lin, Chen Ma, and Fenglong Ma

Abstract—Session-based recommendation is gaining increasing attention due to its practical value in predicting the intents of anonymous users based on limited behaviors. Emerging efforts incorporate various side information to alleviate inherent data scarcity issues in this task, leading to impressive performance improvements. The core of side information-driven session-based recommendation is the discovery and utilization of diverse data. In this survey, we provide a comprehensive review of this task from a data-centric perspective. Specifically, this survey commences with a clear formulation of the task. This is followed by a detailed exploration of various benchmarks rich in side information that are pivotal for advancing research in this field. Afterwards, we delve into how different types of side information enhance the task, underscoring data characteristics and utility. Moreover, we discuss the usage of various side information, including data encoding, data injection, and involved techniques. A systematic review of research progress is then presented, with the taxonomy by the types of side information. Finally, we summarize the current limitations and present the future prospects of this vibrant topic.

Index Terms—Session-based recommendation, Side information, Benchmarks, Data-centric perspective.

I. INTRODUCTION

RECOMMENDER systems (RS), as an effective tool to handle information overload, play a vital role in the current information era. Generally, we can summarize the utility of RS from two views, *i.e.*, information-consumer view and information-provider view. On one hand, for information consumers (*i.e.*, users), RS can significantly reduce their time and effort required to find information of interest among vast amounts of data, thus lowering the cost of obtaining valuable information. On the other hand, as for information providers, these systems can effectively identify potential consumers for their products, thereby facilitating information distribution and transaction completion.

General recommender systems typically rely on *users' profiles* and *long-term behaviors* to mine their behavior patterns,

Manuscript received 10 Oct. 2024; revised 6 Mar. 2025; accepted 28 Apr. 2025. This work was supported by the Natural Science Foundation of China (No.62376051, No.62076046), the Early Career Scheme (No. CityU 21219323) and the General Research Fund (No. CityU 11220324) of the University Grants Committee (UGC), and the NSFC Young Scientists Fund (No. 9240127). (Corresponding authors: Bo Xu and Chen Ma.)

Xiaokun Zhang, Bo Xu and Hongfei Lin are with the School of Computer Science and Technology, Dalian University of Technology, Dalian, China. E-mail: dawnkun1993@gmail.com, xubo@dlut.edu.cn, hflin@dlut.edu.cn

Bowei He, Chen Ma are with the Department of Computer Science, City University of Hong Kong. E-mail: boweihe2-c@my.cityu.edu.hk, chenma@cityu.edu.hk

Chenliang Li is with the School of Cyber Science and Engineering, Wuhan University, Wuhan, China. E-mail: cllee@whu.edu.cn

Fenglong Ma is with the College of Information Sciences and Technology, Pennsylvania State University, Pennsylvania, USA. E-mail: fenglong@psu.edu

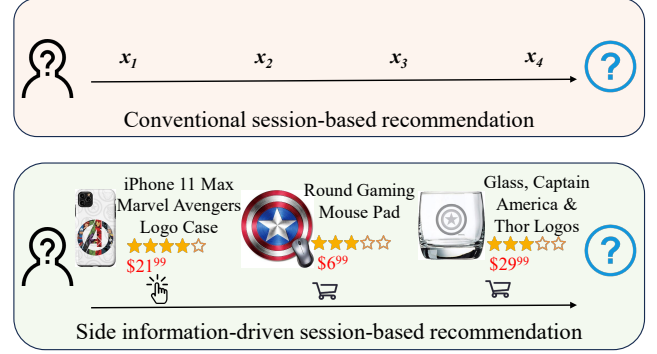


Fig. 1. Conventional SBR v.s. side information-driven SBR.

like collaborative relations and co-occurrence signals, for personalized recommendations [1]. However, due to stringent privacy policies, it has become increasingly difficult to access such sensitive identity information, leading to the unsatisfactory performance of general recommendation methods.

Session-based recommendation (SBR) is proposed to address this prevalent scenario, where SBR can predict the next items of interest for an *anonymous* user based on her limited behaviors within a *short* session [3], [2]. Owing to possessing significantly practical value, SBR has garnered widespread attention since its inception [3], [2], [4]. In the meantime, unfortunately, SBR inevitably suffers from serious data scarcity issues [5], [6]. As shown in the first row of Fig. 1, conventional SBR mainly relies on a limited number of user-item interactions merely represented by item IDs to infer a user's intents. In fact, these methods can just mine co-occurrence associations among items while failing to reveal insightful user behavior intents [7], [8]. As a result, this paradigm greatly limits the effectiveness of SBR.

To this end, as shown in the second row of Fig. 1, emerging efforts tend to incorporate various types of side information, such as item images [9], [7], title text [10], [11], price [12], [13] and behavior types [14], [15], to facilitate user intent understanding under the circumstances of SBR. The incorporation of side information into SBR can undoubtedly enrich session data, and its advantages can be summarized into the following aspects: (1) Depicting item detailed features. For example, images can convey a clothing's style, such as whether it is tight or loose. Such detailed features offer a comprehensive portrayal of an item, encompassing crucial factors in determining users' choices. (2) Revealing user fine-grained preferences. As shown in the second row of Fig. 1, we could infer that the user may be a Marvel fan since

there are commonly Marvel factors in the images or text of items this user has bought. Such insights about user fine-grained preferences open up the possibility of providing highly personalized services, like recommending a T-shirt featuring an Iron Man logo to this user.

Despite the increasing trend of incorporating side information into SBR, unfortunately, there is no corresponding survey thoroughly reviewing this vibrant field of side information-driven session-based recommendation abbreviated as SIDSBR. Especially, the data is of great significance in the domain, as it affects the design and evaluation of recommendation algorithms. Under SBR scenarios, different datasets apply distinct methods to generate session data (please refer to Section III for details on session generation), which greatly impacts recommendation performance. More importantly, data determines from what view corresponding approaches aim to model user behaviors. Specifically, different types of side information portray user behaviors from different aspects. For instance, the item price signifies a user's price sensitivity, while the image converts her preferences on an item's appearance. As a result, there is an urgent need to examine the SIDSBR with a focus on data perspective.

In this survey, therefore, we aim to provide a comprehensive review of the SIDSBR from a data-centric perspective. Specifically, we first give a clear formulation of SIDSBR and clarify the differences between two frequently confused tasks, *i.e.*, session-based recommendation and sequential recommendation. Available benchmarks supporting the research of SIDSBR are then outlined, which elaborates on the types of side information they contain. Next, we delve into the data characteristics and utility of each type of information, presenting their potential to improve the SBR task. The usage of side information is then discussed, with an emphasis on data encoding, data injection, and involved techniques. We thoroughly examine current methods for SIDSBR using a new taxonomy based on the types of information they rely on, offering convenience for future work in replicating and comparing these methods. Finally, we discuss the limitations of current SIDSBR approaches and look forward to the future directions of this vibrant topic. The main contributions of this survey can be summarized as follows:

- We comprehensively review side-information-driven session-based recommendations from a data-centric perspective. To the best of our knowledge, this survey marks the first effort to examine this vibrant topic.
- This survey thoroughly presents formulation, available benchmarks, data utility and usage, as well as research progress for side information-driven session-based recommendation.
- We highlight and discuss several promising research directions for this topic, looking forward to inspiring the community.

Paper collection. This survey provides a comprehensive review of over 60 papers that focus on the topic of side information-driven session-based recommendation. For liter-

ature collection, DBLP¹ and Google Scholar² are primary means of searching for relevant works. Specifically, we rely on keywords, such as “session-based recommendation”, “session recommendation” and “sequential recommendation”, to search related works from these literature libraries. Besides, we focus on works published after 2016, following the first proposal and wide acceptance of SBR in GRU4Rec [3] presented at ICLR 2016. To ensure the quality of collected works, we prioritize papers from top-tier conferences and journals, including SIGIR, KDD, WWW, ICDE, AAAI, IJCAI, WSDM, CIKM, ICDM, RecSys, IEEE TKDE, and ACM TOIS. Additionally, we also pay attention to the unpublished preprints on arXiv, selecting those with interesting ideas or novel techniques, especially the large language models-related methods, to present a more inclusive panorama for SIDSBR.

Differences from existing surveys. Our survey distinguishes itself from existing works of several key areas, including reviews on sequential recommendation (SR) [16], [17], multi-modal recommendation (MMRec) [18], [19] and session-based recommendation (SBR) [20], [4], [21]. Firstly, as clarified in Section II-B, SBR is different from SR in several perspectives, including task definition, data distribution, and experiment settings. Different from existing reviews on SR, this survey is specifically tailored to explore various approaches that are uniquely suited to the task of SBR. Secondly, MMRec methods typically require user profiles to guide model implementation, leading to their inapplicability to SBR. Moreover, different from current MMRec mainly focusing on text and images, this survey extends the scope to include extra information particularly prevalent in SBR, such as time and category. The time information sheds light on user dynamic intents within a session [22], [25], while category information is popular in SBR to reveal hierarchical transition patterns among items [26], [27]. Thirdly, existing reviews on SBR [20], [4] mainly focus on a single kind of information, *i.e.*, item ID, while ignoring the significance of various side information in revealing user intents within short sessions. In contrast, our survey embraces a wider range of data types, clarifying the potential of rich data in revealing user intents. This comprehensive review on SBR from a data-centric perspective adapts to current requirements of the community, facilitating understand, exploration and implementation of researchers and practitioners on this vibrant topic.

Content arrangement. The remaining content of this survey is organized as follows. Section II presents the paradigm of side information-driven session-based recommendation and details the distinction between session-based recommendation and sequential recommendation. Section III introduces the available datasets to the task of SIDSBR in detail. Section IV outlines the characteristics and utility of various types of side information on SBR. Section V presents the encoding manners of side information, illustrates the methods for incorporating side information from item-level, session-level and prompt-level, and summarizes techniques involved in the domain. Section VI elaborates on the relevant works in SIDSBR from

¹<https://dblp.org>

²<https://scholar.google.com>

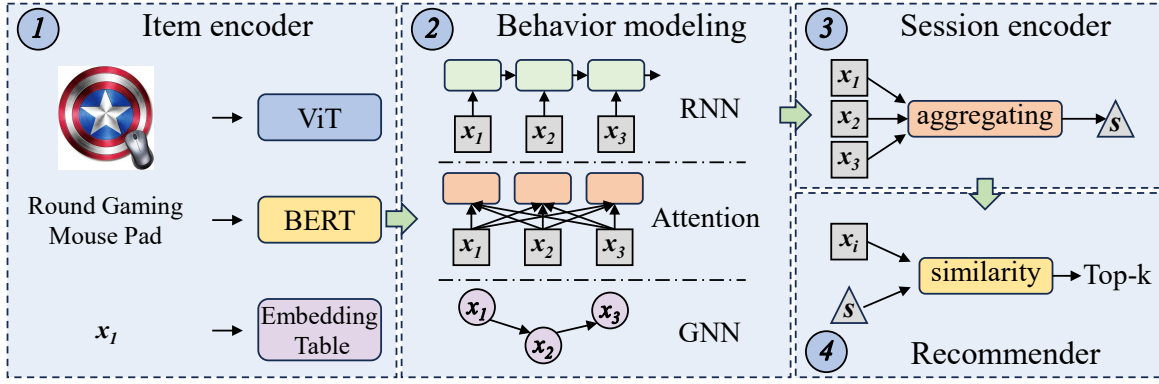


Fig. 2. The workflow of side information-driven session-based recommendation.

a new taxonomy, *i.e.*, types of side information they used. Section VII discusses the limitations of existing SIDSBR and provides several promising research directions about this popular field. Section VIII concludes this survey.

II. TASK FORMULATION

A. Paradigm of SIDSBR

The emergence of session-based recommendation has coincided with the rise of deep learning techniques, leading to the dominance of neural models in this field. Generally, neural models focus on obtaining item and session embeddings by modeling user behaviors within sessions, basing recommendations on their similarity. As shown in Fig. 2, the common workflow of SIDSBR can be summed up as four steps: (1) Item encoder, which converts item information into dense embeddings. For instance, look-up embedding table is usually used to map discrete item IDs into ID embeddings. (2) Behavior modeling, which captures transition patterns among items in sessions for item embedding learning. Various neural structures are modified in this stage to handle SIDSBR, such as Recurrent Neural Network (RNN) [3], [9], [28], attention mechanism [167], [29], [30], [31] or Graph Neural Network (GNN) [32], [34], [35]. (3) Session encoder, which encodes intents of the anonymous user as a dense vector by aggregating item embeddings within the session. (4) Recommender, which usually calculates the similarity between session embedding and item embeddings, *e.g.*, cosine similarity, to formulate the recommendation list.

Formally, let \mathcal{I} denote an item set, where $n = |\mathcal{I}|$ is the total number of unique items. An item $x_i \in \mathcal{I}$ may contain various kinds of side information, such as category, text (title and description), image, and so on. Besides, a session $S = [x_1, x_2, \dots, x_t]$ ($x_j \in \mathcal{I}$) consists of t items an anonymous user interacted with within a short period. Current methods of SIDSBR typically encode item information as dense embeddings $\mathbf{v}_i \in \mathbb{R}^d$ (d is the embedding dimension), model item transition relations throughout sessions, derive session embedding \mathbf{s} via various approaches, and evaluate similarity scores between the current session embedding and candidate item embeddings. Finally, the items with top- k scores are shown to users for personalized recommendation.

B. Distinction between SBR and SR

Some researchers and practitioners in the community often indiscriminately conflate two popular recommendation tasks, namely session-based recommendation (SBR) and sequential recommendation (SR). As a result, it is crucial to clarify the distinction between SBR and SR before delving into the SIDSBR in detail. Specifically, we will comprehensively distinguish between SBR and SR from a macro-view to a micro-view across three aspects: task definition, data characteristics and experimental implementations.

For the task definition, SBR is dedicated to predicting actions of an *anonymous* user based solely on her *short-term* behaviors happened in a short period, such as 30 minutes or one day [36], [37], [38]. In contrast, SR assumes the availability of user profiles, and can employ rich *user information* like demographics and their entire *long-term behaviors* spanning from several months to several years to provide personalized suggestion [39], [40], [41].

From the perspective of data characteristics, the sequence length in SBR is much shorter than that of SR [42], [43], [44]. Generally, a session in SBR commonly contains no more than five items, whereas the sequence length in SR is usually longer than ten. Additionally, there are significantly more sequences in SBR compared to SR. Under the scenarios of SR, the number of sequences corresponds to the number of users. However, in SBR situations, a single user sequence may be divided into multiple sessions based on pre-defined time span, resulting in a large amount of unique sessions. These characteristics, *i.e.*, short and numerous sessions, also make SBR inherently more challenging than SR.

Regarding experimental implementations, SR relies on user profiles, such as user ID or demographics, to derive user embeddings (*i.e.*, sequence embeddings). Conversely, SBR aggregates item embeddings within a session to obtain the session embedding without accessing any extra user information. Moreover, SR typically applies a leave-one-out operation to split all sequences into training, validation and test sets [45], [46], [47]. This means that for a single sequence, the last item is used for testing, the penultimate item for validation, and the remaining for training. In contrast, SBR formulates experiment sets via grouping sessions instead of splitting any single session [48], [49], [50].

TABLE I
AVAILABLE DATASETS FOR SIDE INFORMATION-DRIVEN SBR.

Dataset	Time*	Category	Brand	Price	Title	Description	Image	Address	Rating*	Review*	Behavior*
Amazon	♡	♠	♠	♡	♠	♠	♠		♡	♠	
Beer	♡	♠			♠				♡	♠	
Cosmetics	♡	♡	♠	♡							♠
Diginetica	♡	♡		♡	♡						
Foursquare	♡	♠						♡			
Goodreads	♡				♠	♠	♠		♡	♠	
Instacart	♡	♠			♠						♠
JingDong	♡	♡									♠
LastFM	♡	♠			♠						
MovieLens	♡	♠			♠				♡		
NineRec	♡				♠	♠	♠				
RetailRocket	♡	♡									♠
Steam	♡	♠		♡	♠					♠	
Taobao	♡	♡									♠
Tmall	♡	♡	♡								♠
Trivago	♡			♡							♠
XING	♡	♠			♠			♡			♠
Yelp	♡	♠						♡	♡	♠	
Yoochoose	♡	♡		♡							

The symbol * denotes **user-item interaction** information, while other types of information without this symbol belong to items' **intrinsic properties**. The symbol ♡ indicates the information is presented by **numerical** format, and the symbol ♠ is for **descriptive** format, while the blank means the relevant information is absent.

In order to cover as many methods as possible suited to the task of SBR, this survey includes SBR methods as well as certain SR models that can be modified with reasonable efforts to handle the situation of SBR. Specifically, the selected SR models should be effective when trained using the aforementioned SBR experimental implementations. This incorporation of certain SR methods facilitates the expansion of our review's scope, providing a broad reference for the community.

III. BENCHMARKS

Datasets, as collections of user-item interaction data, are the foundation for recommender system research. They serve as the essential groundwork for designing, training, and evaluating recommendation algorithms. This is particularly true for SIDSBR, where the types of information in a dataset dictate research questions it supports. For instance, studying user price sensitivity is infeasible with datasets lacking price information. Similarly, exploring multi-modal recommendation algorithms is only possible with datasets containing multi-modal information. Unfortunately, there is no work comprehensively reviewing important datasets in the field. To this end, this section provides a detailed introduction to the available datasets for promoting future research in SIDSBR.

A list of available datasets for SIDSBR, along with their associated information types, is outlined in Table I. These datasets are selected from literature in SBR since its inception in 2016 [3]. Additionally, we scrutinize works of general recommender systems in the last five years (2020–2025) and select datasets that are appropriate to our research topic.

We can categorize various kinds of side information into the following two types: (1) interaction information, which refers to data generated during user-item interactions, such as interaction time, user rating or reviewing for an item, as

marked with * in Table I; and (2) item intrinsic properties, which describe the inherent features of the item, including category, brand, price and so on. Moreover, relying on different vehicles for conveying item features, various information can also be grouped into two categories: *numerical information* of ♡ and *descriptive information* of ♠, as illustrated in Table I. On one hand, the numerical information delivers the abstract meaning of an item using numbers, such as integer IDs for the category or real numbers for the item price. On the other hand, descriptive information characterizes an item via modality information like text or images, which can encompass rich semantics such as color, style and so on. The details of these datasets, including contents, session formulation paradigms and related methods, are outlined as follows:

- **Amazon**³ is scratched from the E-commerce website Amazon, capturing product reviews along with its rich metadata including category, price, and title/description text. This dataset covers various domains such as Beauty, Cellphones and Accessories, Toys and Games and more. Typically, data belonging to each domain is treated as a distinct dataset for recommendation tasks. Current models usually view user behaviors occurring within one day as a session in the Amazon dataset due to the time granularity in this dataset being precise to the day. Containing various kinds of information, the Amazon dataset is widely used in the community to evaluate recommendation algorithms [51], [52], [12], [13], [7].
- **Beer**⁴ contains user reviews about beer from Ratebeer and Beeradvocate websites. Besides the information listed in Table I, it also records sensory aspects of beer, such as taste, look, feel, and smell. The user-item interactions

³<http://jmcauley.ucsd.edu/data/amazon/>

⁴[https://cseweb.ucsd.edu/\\$\sim\\$jmcauley/datasets.html\#multi_aspect](https://cseweb.ucsd.edu/\simjmcauley/datasets.html\#multi_aspect)

are presented in the granularity of second in this dataset, enabling the creation of sessions with fine-grained time spans, such as 20-second intervals. As far as we know, there is a few methods using this dataset to examine recommendation performance, *e.g.*, CoCoRec [26].

- **Cosmetics**⁵ records behavior data over 5-month period spanning from October 2019 to February 2020 in a medium-sized cosmetics online store. Note that, this dataset explicitly highlights session IDs, eliminating the need for extra data partitioning for session formulation. Generally, data from each month is used as a separate dataset in existing studies [12], [53], [54].
- **Diginetica**⁶, used in the CIKM Cup 2016 competition, consists of sessions extracted from e-commerce search engine logs. In this dataset, item title is presented via discrete token IDs instead of raw text. This dataset is popular in the task of SBR [24], [55], [56], [57].
- **Foursquare**⁷ records digital footprints of users in location based social networks, crawled from Foursquare. It includes global-scale check-in data from restaurant venues across over 400 cities, covering a period of around 18 months, from April 2012 to September 2013. This dataset contains detailed geographic information about the restaurants, making it a valuable resource for studying point-of-interest (POI) recommendation, where the goal is to suggest places based on a user's interests and location. In practice, data from different cities in this dataset are often used as separate benchmarks for evaluating recommendation models [58], [59], [60], [61].
- **Goodreads**⁸ contains users' reviews towards books from the Goodreads website [62]. This dataset records user-item interactions with minute-level granularity, allowing for the generation of sessions with arbitrary time intervals, such as 30 minutes. Additionally, since the dataset is originally presented in the form of a book graph, corresponding methods often utilize GNN to handle the recommendation tasks [63], [64], [65].
- **Instacart**⁹ describes grocery orders of users on the online grocery marketplace Instacart. Specifically, this anonymized dataset includes over 3 million grocery orders from more than 200,000 users. Owing to the vast volume of data within this dataset, it is a common practice to sample a subset from it for analysis and experimentation in existing methods [?], [8].
- **JingDong**¹⁰ captures various micro-behaviors of users, such as ordering, adding to cart, and commenting, from the Chinese e-commerce website JD.com. Besides recording the timestamp of user behaviors, this dataset also highlight the dwell time at the second level that users spend on each item. The dwell time can serve as an indicator of user preference, where the longer a user stays on an item, the stronger their interest in it [22].

Moreover, this dataset includes two item categories, *i.e.*, 'Appliances' and 'Computers', which are commonly used as separate datasets in related research [67], [14], [15].

- **LastFM**¹¹, as a music-artist dataset, recording user interactions (listening) with music artists, collected from online music service Last.fm. In this dataset, user-provided tags for the artists are used as the category information. Due to its unique focus on music-related events, this dataset is particularly suitable for music interest recommendation [68], [69], [70], [71].
- **MovieLens**¹² a classical rating dataset in the community, compiles user ratings about movies from the MovieLens website. The dataset is available in various sizes, each forming a distinct dataset used in different studies. For instance, MovieLens-100k and MovieLens-1m are commonly used variants, each named after the number of user-movie ratings they contain [72], [73], [74].
- **NineRec**¹³ consists of the user-watching behaviors on short videos from prominent short-video platforms in China, such as Bilibili, Kuaishou, and Douyin [168]. This dataset is unique because it includes data from different channels within a single platform (*e.g.*, vertical channels on Bilibili) as well as across different platforms (*e.g.*, Bilibili versus Douyin). This feature makes the dataset particularly valuable for exploring both cross-domain and cross-platform recommendation scenarios [75], [76]. As a newly released dataset, it has yet to be introduced and utilized in the field of SBR.
- **RetailRocket**¹⁴, which is generated from the Retailrocket online shopping site, spans a period over 4.5 months and captures various kinds of user-item interactions, such as 'view', 'add to cart' and 'transaction'. Note that, all words in text values were normalized and hashed in this dataset. Additionally, this dataset highlights the hierarchical relationships among categories using a category tree structure, which can be a valuable resource in recommendation research. It has been a popular benchmark used in various studies, including [77], [78], [79].
- **Steam**¹⁵, crawled from a famous online video game distribution platform Steam, contains user reviews about computer games from October 2010 to January 2018. This dataset includes unique information about video games, such as users' play hours, media scores and developer details. Such information can provide game manufacturers with valuable insights into user behaviors within the gaming domain. Some methods evaluate their recommendation performance under this game dataset [80], [81], [82], [83].
- **Taobao**¹⁶ captures users' online shopping behaviors from Taobao which is one of China's largest e-commerce platforms. This dataset is widely used in current research to address implicit feedback recommendation scenarios,

⁵<https://www.kaggle.com/mkechinov/e-commerce-events-history-in-cosmetics-shop>

⁶<https://competitions.codalab.org/competitions/11161>

⁷<https://sites.google.com/site/yangdingqi/home/foursquare-dataset>

⁸<https://mengtingwan.github.io/data/goodreads>

⁹<https://www.kaggle.com/c/instacart-market-basket-analysis>

¹⁰<https://tinyurl.com/ybo8z4yz>

¹¹<http://ocelma.net/MusicRecommendationDataset/index.html>

¹²<https://grouplens.org/datasets/movielens/>

¹³<https://github.com/westlake-repl/NineRec>

¹⁴<https://www.kaggle.com/retailrocket/e-commerce-dataset>

¹⁵<https://github.com/kang205/SASRec/tree/master/data>

¹⁶<https://tianchi.aliyun.com/dataset/dataDetail?dataId=649>

providing valuable insights into user interactions and preferences [79], [84], [85], [86], [87].

- **Tmall**¹⁷ collects anonymized users' shopping logs over six months, including data before and on the 'Double 11' day which is a major promotional event in Chinese e-commerce platforms. Originating from Tmall, one of the largest B2C platforms in China, this dataset includes the label information indicating whether a user is a repeat buyer for the same item. This feature is particularly valuable for studying repeat consumption behaviors. In addition, this dataset was initially published for the IJCAI 2015 competition and has since been extensively used in related research [88], [89], [90], [91].
- **Trivago**¹⁸ is a session-based hotel recommendation dataset that records user actions during online hotel searches. Released by Trivago which is a global hotel search platform, this dataset was used in the RecSys Challenge 2019. It specifies the type of actions performed by users like view item rating, item information, item image and so on, providing valuable insights into user behaviors when booking hotels. Despite the richness of its user interaction data related to hotel bookings, it has been underutilized in academic research, with relatively few studies leveraging its potential [15].
- **XING**¹⁹ collects user interactions with job postings, tailored specifically for job recommendation tasks [169], [170]. Created by XING, a social network focused on job searching and recruitment, this dataset was used in the ACM RecSys Challenge 2016. It includes rich information about job postings, such as career levels (e.g., beginner, experienced, manager), industry types (e.g., Internet, Automotive, Finance), and employment types (e.g., full-time, part-time). Given the inherent sequential dependencies among user-job interactions, many efforts attempt to handle the job recommendation task with techniques in SBR [92], [93], [94], [95].
- **Yelp**²⁰ captures users' reviews for businesses across eight metropolitan areas from the online review platform Yelp.com. This dataset includes a wide range of business attributes, such as opening hours, parking availability, and ambience. Notably, its location information makes it well-suited for point-of-interest (POI) recommendation tasks. Moreover, due to its large size, many existing studies only utilize a subset of this dataset for their analyses and implementation [96], [97], [98], [40].
- **Yoochoose**²¹ contains six months of click-stream data from a European e-commerce website that sells a variety of items like garden tools, electronics and more. This dataset explicitly segments a user's behavior sequence into multiple sessions, making it a popular benchmark for SBR. Released in the RecSys Challenge 2015, Yoochoose is one of the earliest datasets designed for SBR and remains widely used in the domain [99], [100], [101].

IV. DATA CHARACTERISTICS AND UTILITY

The goal of side information-driven session-based recommendation is to incorporate various types of data to enhance user intent capturing. Specifically, each type of information possesses its unique characteristics and can shed light on different behavior patterns of users, thereby improving SBR from various aspects. For instance, 'Time' information is instrumental in modeling sequential dependencies among items, whereas the 'Price' factor can provide insights into user price sensitivity. Furthermore, these various types of data allow for the creation of distinct recommendation scenarios. Location data, such as the address information, is crucial for the formulation of point-of-interest recommendations. In the following contents of this section, thus, we will delve into the details of currently available side information. Specifically, we will explain the characteristics of each type information, highlight user behavior patterns that they help reveal, point out the specific recommendation scenarios that certain types of information could support, and finally discuss the collective relations of various information in affecting SIDSBR.

A. Time Information

Time information records the exact timestamps of user-item interactions, often represented in UNIX time format. This data can be easily converted into a detailed 'year-month-hour-minute-second' structure, offering a fine-grained view for user activities. In the context of SBR, time information plays a pivotal role in identifying the sequential dependencies among items, which is essential for capturing the evolving nature of user preferences [102], [103], [104], [105]. Moreover, the varying granularity of time data facilitates the detection of diverse sequential dependencies, which is of importance in interpreting user behaviors. For instance, by analyzing time data at the day-level granularity, it becomes intuitive to recommend quick meal options during weekdays when users may be pressed for time, and more leisurely dining options on weekends when they might prefer a more relaxed experience. Because it enables modeling of dynamic user intent, time information is of great importance in handling both SBR and sequential recommendation tasks.

B. Category Information

Category information refers to the hierarchical taxonomy used to classify and group items based on their types, characteristics, or other relevant attributes. In current datasets, category information could be represented either as integers or raw text, as illustrated in Table I. Typically, the item categories are organized as a tree structure, such as 'Electronics' branching into subcategories like 'Computer' and 'Smartphones', with further subdivisions like 'Laptops' and 'Desktops' under the 'Computer' branch. This hierarchical structure mirrors the users' decision-making process, as they progressively narrow down their intentions, such as navigating from 'Clothes' to 'Men's Clothing' and finally to 'Pants'. Leveraging these category hierarchies can enhance recommender systems by significantly reducing the number of candidate items, leading

¹⁷<https://tianchi.aliyun.com/dataset/dataDetail?dataId=42>

¹⁸<http://www.recsyschallenge.com/2019/>

¹⁹<http://2016.recsyschallenge.com>

²⁰<https://www.yelp.com/dataset>

²¹<http://2015.recsyschallenge.com/challenge.html>

to more accurate recommendations. For instance, suggesting phone cases from the ‘phone-accessories’ subcategory would be highly relevant for users who recently bought a smartphone from the ‘Smartphones’ category.

C. Brand Information

Brand information encompasses details about the brand or manufacturer of an item, reflecting intangible item attributes, such as reputation or quality. It serves as a unique indicator of manufacturer state among the currently available information. Similar to item categories, brand information is typically represented as integers or raw text. The brand of an item plays a unique role in guiding customers’ purchasing decisions. Specifically, the ‘brand effect’, widely studied in economics, underscores how brand loyalty influences a consumer’s behaviors. Recommender systems can harness this effect to deliver more personalized and relevant suggestions. For example, users who exhibit a strong preference for a particular brand are likely to respond favorably to recommendations featuring products from that brand. This alignment with user preferences can significantly enhance the effectiveness of recommendation strategies, making ‘Brand’ information a valuable asset to meet individual tastes.

D. Price Information

Price denotes to the monetary cost assigned to an item, representing the amount a user is required to pay for its acquisition. In current datasets, item prices are commonly presented with the format of real numbers, such as \$19.99. An item’s price is undeniably a key factor influencing a user’s decision when selecting products [106], [107], [108]. Typically, users approach shopping with a predetermined budget or price range in mind. No matter how interested they may be in a particular item, if its price exceeds their budget, they are likely to abandon the purchase, emphasizing the importance of aligning recommendations with users’ financial expectations. Owing to its critical influence on user decisions, the integration of item price into SBR task holds significant potential to boost recommendation effectiveness and its applicability in real-world scenarios. By taking price factors into considerations, SBR models can offer recommendations that are not only aligned with user interest but also their financial expectations, thereby greatly improving user satisfaction [12], [13].

E. Text Information

Text information provides detailed descriptions of items using natural language, typically presented in the form of raw words. As outlined in Table I, we primarily focus on two common types of text: item *title* and *description*. The item title serves as a concise descriptive snippet, summarizing the item with a focus on its key features or selling points, such as size, capacity, and material. In contrast, the description offers a more comprehensive overview, giving potential buyers detailed insights into aspects like usage instructions and support information. With rich semantics, these textual elements serve as a valuable resource for enhancing recommender systems.

On the one hand, they help articulate item features, such as highlighting Marvel-related factors in Fig. 1. On the other hand, they can reveal users’ fine-grained preferences, like identifying Marvel fans in Fig. 1. Moreover, with achieving tremendous success across various tasks, Large Language Models (LLMs) have been widely applied in the field of recommender systems [109], [110], [111]. Obviously, text information is the foundation in the development of LLM-based recommender systems. Furthermore, text also plays a vital role in advancing research within multi-modal recommender systems [112], [113], [114].

F. Image Information

Image information refers to the visual representations of items, commonly in the form of digital photographs or graphics. These images provide users with a clear and intuitive understanding of what the items look like, effectively showcasing key features such as color and style. Visual content plays a unique role in reducing the uncertainty that often accompanies online shopping, where users cannot physically inspect the products. By harnessing the detailed insights gained from images, recommender systems are able to offer more precisely tailored suggestions, matching closely to individual user interests. As depicted in Fig. 1, item images can be leveraged to identify a user’s specific interests on Marvel-themed items, thereby significantly enhancing the effectiveness of recommendations. Similar to text information, item images are a vital resource supporting the research on multi-modal recommendation scenarios [7], [115], [116].

G. Address Information

Address information signifies the geographic location of certain items such as restaurants, tourist attractions, or other points of interest. This information is typically provided in varying levels of granularity, ranging from precise coordinates (latitude and longitude) to broader regions like ‘Northeast’ or ‘Southwest’. Address information is crucial in influencing user decisions, especially on location-based services including location-based advertising and online food delivery [61]. For instance, under restaurant recommendations, the proximity of a restaurant plays a significant role in a user’s decision to visit since that users generally prefer dining at places that are conveniently located nearby. Moreover, address information forms the basis for the specialized task of point-of-interest (POI) recommendation, where the goal is to suggest new places, such as restaurants or tourist spots, tailored to a user’s preferences and location [117], [60], [61].

H. Rating Information

Rating information serves as a quantifiable indicator of a user’s overall attitude toward an item, typically expressed on a scale from 0 to 5. Generally, higher ratings indicate a stronger preference or satisfaction with the item. Unlike other types of information provided by manufacturers, ratings are uniquely generated by users themselves. This user-generated nature makes ratings particularly valuable for crafting personalized

recommendations. For instance, by analyzing users' explicit ratings across various items, recommender systems can identify and suggest items similar to those that have received high ratings, while avoiding items with lower ratings. The incorporation of ratings aligns recommendations more closely with the user's demonstrated preferences, resulting in a more tailored and satisfying user experience. Additionally, rating information is central to the classical task of rating prediction, which was a primary focus in the early development of recommender systems [118], [119].

I. Review Information

Review information consists of written feedback provided by users who have purchased and experienced the items. Unlike ratings, which offer a numerical summary of user satisfaction, reviews delve into the specifics of user preferences, item features, and overall experiences through raw text. This detailed content offers valuable insights that can significantly enhance the precision and relevance of recommendations. For instance, if a user mentions a preference for loose-fitting clothing in a review, the system can prioritize recommending items that match this preference, avoiding the suggestion of unsuitable tight-fitting options. Besides, reviews allow for an evaluative perspective on the item itself. Specifically, if the majority of reviews criticize an item for poor quality, it suggests that the item might not be suitable for users seeking high-quality products. By integrating these insights from reviews, recommender systems can deliver more reasonable and user-aligned recommendations. In addition, the rich, descriptive content found in reviews serves as a valuable resource for LLM-based approaches to analyze and understand user preferences at a deeper level.

J. Behavior Information

Behavior information captures the specific types of user-item interactions, such as clicks, add-to-cart, or purchases, typically presented in raw text format. Diverse types of behavior offer distinct insights into users' intents and preferences regarding items. For instance, a user clicking on an item may not necessarily indicate a strong liking. Instead, it could just be a part of casual browsing. In contrast, a behavior such as making a purchase is a strong indicator of user satisfaction with the item. The consideration of these diverse behaviors allows for a more fine-grained and deeper interpretation of user-specific preferences, thus holding the potential for more effective recommendations. This focus on different user behaviors has also given rise to a new research topic known as 'multi-behavior recommendation' whose aim is to enhance user intent modeling by integrating various types of user interactions [120], [121], [122].

K. Discussion

Beyond their individual effects, various types of side information also exhibit collective relations in user behavior modeling within SIDSBR, which can be either complementary or conflicting. On one hand, different types of information

can be complementary in capturing user behavior patterns. For example, users may be more willing to accept higher prices for well-known brands, making the joint consideration of item price and brand valuable for understanding their price preferences [12], [13]. Similarly, ratings quantify user satisfaction, whereas reviews provide explanations and sentiment. Their fusion may improve understanding of user preferences beyond numerical scores. On the other hand, conflicts may arise among different types of side information. For instance, image and text descriptions of an item do not always align, *e.g.*, misleading promotional images or exaggerated textual descriptions can introduce inconsistencies, making it challenging to extract coherent item representations [7]. Overall, effectively modeling these relations requires careful design, like modality alignment techniques or adaptive fusion mechanisms, to resolve conflicts and leverage complementary signals.

V. USAGE OF SIDE INFORMATION

Side information-driven session-based recommendation is centered around the strategic use of diverse data to deliver personalized suggestions catering to user preferences. In previous sections, we elaborated on the characteristics of various side information and their utility in revealing user intents—essentially addressing what side information is and why it is valuable. This section shifts the focus to the practical aspect of side information, *i.e.*, how to use these side information. Drawing from a comprehensive review of SIDSBR, we plan to detail the usage of side information across three key areas: (1) Data encoding, *i.e.*, the pre-processing stage, where we explain how to encode various side information to serve as inputs for subsequent neural models; (2) Data injection, an overview of manners for integrating side information into the recommendation task; and (3) Involved techniques, a summary of common techniques used in SIDSBR, such as RNN, attention mechanism, GNN and so on.

A. Data Encoding

As discussed in Section III, we classify various kinds of side information into two groups, *i.e.*, numerical information and descriptive information, based on their vehicles for portraying item features. In the meantime, these two types of information are typically encoded in different ways by current neural models in SIDSBR.

For numerical information, the data is first converted into consecutive integers and then mapped into dense embeddings using a look-up embedding table [12], [55], [123]. This transformation allows raw numerical values to be seamlessly integrated into neural networks. However, recent methods have recognized that original numerical information may contain unique relationships, such as the hierarchical structure of item categories. To preserve these relationships during the integer transformation, some approaches have been developed. For instance, CoHHN [12] and BiPNet [13] have observed that item price distribution often follows a logistic distribution, so they convert item price into discrete price levels based on this distribution instead of commonly used uniform one. This ensures that the encoded numerical data preserves its intrinsic

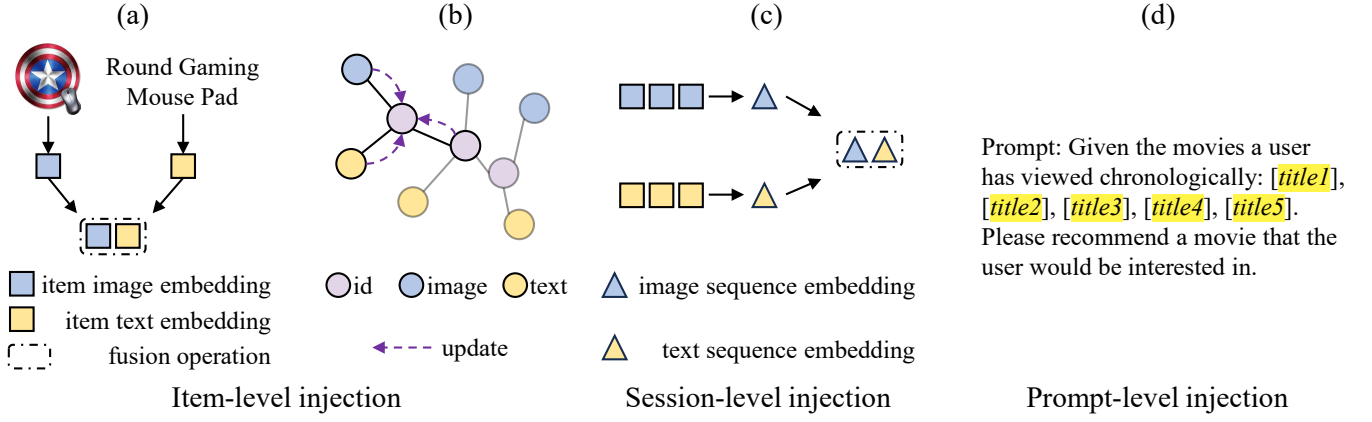


Fig. 3. Three manners for side information injection: item-level injection, session-level injection and prompt-level injection

characteristics, enabling the generation of more meaningful and informative representations.

In contrast, descriptive information, such as text and images, requires more sophisticated processing techniques. Specifically, existing approaches tend to employ advanced pre-trained models from Natural Language Processing (NLP) and Computer Vision (CV) to represent this type of information [172], [124], [125]. For example, text data might be processed using models like GloVe [126] or BERT [127], which are capable of generating rich text embeddings based on their extensive training on massive language corpora. Similarly, image data can be represented using models like GoogLeNet [128] or ViT [129], which are well-trained on vast image datasets and are highly effective at capturing intricate visual features. These pre-trained models are equipped with rich knowledge about their respective modalities, allowing them to generate robust embeddings for text and image data. By leveraging these embeddings, recommender systems can enhance the quality and relevance of their recommendations.

B. Data Injection

The key points of neural SBR lie in two aspects: item representation learning and session representation learning. Once these embeddings are established, recommendation lists can be generated based on the similarity between them. Relying on a single type of item ID information, conventional methods in SBR usually resort to a straightforward manner to obtain item ID embeddings and aggregate them into session embeddings to achieve personalized suggestions. However, the complexity increases significantly in side information-driven SBR due to the involvement of the utilization for various kinds of side information. Obviously, how to effectively inject these types of side information into neural models possesses huge impact on recommendation performance. Unfortunately, the injection of side information in SBR has received little attention in current research.

To this end, we summarize the main manners of injecting side information within SIDSBR into three types, including item-level, session-level and prompt-level injection, as illustrated in Fig. 3. The first two methods are commonly applied

in neural models and are organized based on when the injection occurs within the aforementioned workflow of SIDSBR, *i.e.*, Item encoder, Behavior modeling, Session encoder and Recommender. The last method, prompt-level injection, is particularly relevant for approaches using Large Language Models (LLMs). In addition, we will discuss the limitations associated with each type of side information injection.

1) Item-level Injection: Item-level injection incorporates various kinds of side information into SBR on item representation learning, specifically occurring during the Item encoder and Behavior modeling stages, *i.e.*, the first two steps of standard workflow in SIDSBR.

In practice, item-level injection first combines various item side information into unified item embeddings. These unified embeddings then serve as the foundation for representing an anonymous user's intent within a session [56], [112], [72], [123]. For instance, item encoder converts an item's image and text into dense embeddings, respectively. These embeddings can then be merged into a unified item embedding using simple concatenation [130], [131], [56], [10], as illustrated in Fig. 3 (a). Another common approach for item-level injection involves the use of Graph Neural Network (GNN). As presented in Fig. 3 (b), some works [12], [15], [13], [52] create a heterogeneous graph with nodes representing different types of side information. Through performing node embedding updating on this graph, unified item embeddings integrating various side information can be learned.

Although intuitive and straightforward, it is non-trivial to conduct effective item-level side information injection. At first, the inherent heterogeneity and noise across different types of information make it challenging to obtain robust embeddings for these information [7], [132], [133]. In addition, this item-level injection incorporates various types of information before modeling user behavior sequences. Consequently, it falls short of capturing user preferences specific to each type of information. Taking the important price factor as an example, when price is blended with other information at the item level, the model may struggle to discern a user's price sensitivity. This lack of clarity impedes the model's ability to accurately understand user behaviors, which in turn negatively impacts the overall recommendation performance.

2) *Session-level Injection*: Session-level injection relies on session representation learning to introduce side information into SIDSBR, where the side information injection is conducted in Session encoder and Recommender stages.

In this injection manner, neural models prioritize user behavior sequence over a single action (*i.e.*, item). Specifically, they first build sequences for each type of information within the session, then represent each sequence independently, and finally generate recommendation list by considering all sequence embeddings [68], [55], [134]. As illustrated in Fig. 3 (c), for a session, its item image sequence and item text sequence, which have been processed by Behavior modeling separately, are converted into image and text sequence embeddings. Afterwards, these sequence embeddings are combined, like using concatenation, to represent user overall preferences for recommendation generation [135], [86].

This paradigm holds the merit of capturing users' specific preferences related to certain types of information. However, lacking in fusing side information at the item level, they struggle to obtain informative item embeddings. Another challenge lies in synthesizing a comprehensive understanding of users' overall preferences based on their unique interests in each type of information. In reality, users' final decisions are typically influenced by a combination of various factors. Unfortunately, there are not explicit signals to indicate how these factors interrelate to drive user actions, making it difficult to perform effective predictions in this manner.

3) *Prompt-level Injection*: Prompt-level injection is uniquely designed for LLM-based methods in SIDSBR (refer to Section V-C6 for more details on LLM-based SIDSBR), distinguishing itself from item- and session-level injection tailored for neural models. A prompt acts as a structured template containing raw text, designed to trigger an LLM's ability to complete a specific task [136], [110], [111]. Generally, a prompt consists of a task description, an instruction, and the side information to be injected. For instance, as illustrated in Fig. 3 (d), prompt-level injection may involve placing item titles into the template (*i.e.*, injecting side information into the task context) and then leveraging the extensive knowledge encoded within the LLM to generate personalized recommendations.

Despite facilitating the application of LLM on recommendation tasks, the prompt-level injection still faces several challenges that limit the full potential of LLMs in this domain. Firstly, a prompt based solely on a single user's behavior lacks the collaborative knowledge derived from the collective actions of a large user base. This collaborative knowledge is crucial in recommendation systems for identifying shared preferences and trends among users [8], [137]. Furthermore, designing an effective and universally applicable prompt template is difficult, as subtle differences in template structure can lead to significant variations in LLM's performance. Lastly, prompt-level injection is uniquely suitable for text information, making it challenging to integrate other types of data, such as images or numerical attributes. This limitation restricts the versatility and effectiveness of LLM in recommendation domain.

C. Involved Techniques

Various neural structures have been leveraged in SIDSBR, including RNN, attention mechanisms, GNN, LLM, and others. These techniques are utilized in the behavior modeling stage to handle user actions within sessions. This section will examine these techniques from several perspectives: formulations, unique contributions to uncovering user preferences, representative models that instantiate these techniques, and the limitations associated with each technique. Note that, we exclude the machine learning based methods in SBR, such as matrix factorization [138], markov chain [139] and nearest-neighbor-based approaches [140], as these methods focus on modeling item ID sequences rather than incorporating side information for performing SBR.

1) *Recurrent Neural Networks (RNN)*: RNN is one of the earliest neural structures applied in SBR [3]. Given its inherent ability to model ordered data, RNN-based methods are able to capture sequential dependencies among user actions. Taking the representative Gated Recurrent Unit (GRU) as an example, we can formulate its processing as follows,

$$\begin{aligned} \mathbf{h}_j &= (1 - \mathbf{z}_j)\mathbf{h}_{j-1} + \mathbf{z}_j\hat{\mathbf{h}}_j, \\ \mathbf{z}_j &= \sigma(W_z\mathbf{v}_j + U_z\mathbf{h}_{j-1}), \\ \hat{\mathbf{h}}_j &= \tanh[W\mathbf{v}_j + U(\mathbf{r}_j \odot \mathbf{h}_{j-1})], \\ \mathbf{r}_j &= \sigma(W_r\mathbf{v}_j + U_r\mathbf{h}_{j-1}), \end{aligned} \quad (1)$$

where W, W_r, W_z, U_r, U_z are weight matrices; \mathbf{v}_j is the embedding of j -th item in the session; $\mathbf{h}_j \in \mathbb{R}^k$ denotes the output of j -th hidden state in GRU layer ($\mathbf{h}_0 = \mathbf{0}$); and σ denotes the sigmoid function.

In the early stage of SBR's development, RNN and its variants are popular choices for handling side information in the task. For example, MV-RNN [141] and TMRN [67] rely on item-level injection to incorporate side information. P-RNN [9] explores both item-level and session-level injection methods, attempting to capture the contextual impact of side information at different granularities. Nowadays, RNN mostly severs as the basic component for deriving session embeddings, and is combined with other techniques, like GNN [130], [15] and attention mechanism [10], to generate final suggestions.

The limitations of RNN in SIDSBR are notable: (a) The assumption of strict order between any successive items in RNN models is not always valid in real-world scenarios [16], [142]; and (b) Inherent recursive nature of RNN makes it hard to parallelize, leading to high time consumption and poor efficiency [167], [23], [31].

2) *Convolutional Neural Networks (CNN)*: CNN is another pioneering neural structure used in the early stage of SBR [143], [58]. Generally, We can formulate the processing of GNN as follows,

$$\mathbf{o} = \text{Conv}([\mathbf{v}_1; \mathbf{v}_2; \dots; \mathbf{v}_t]), \quad (2)$$

where $[\cdot]$ denotes concatenation operation, $\text{Conv}(\cdot)$ is the convolutional layer including stacked filter and pooling operations, and \mathbf{o} is the output, which is expected to encapsulate the local patterns of user actions.

Compared with the widely used RNN, CNN offers three main advantages. Firstly, CNN does not impose a strict order assumption on successive items in sessions, enabling more robust recommendation performance. Secondly, CNN is inherently designed for parallel processing, improving model efficiency. Lastly, CNN is able to effectively learn local patterns within small groups of items in a session, facilitating understanding of item relationships.

Despite these benefits, CNN has seen limited applications in SBR. Notable examples include Caser [58] and NextIt-Net [144]. Besides, only a few efforts, like 3D-CNN [143], have attempted to handle side information via CNN. In this approach, an item's embeddings of various side information are stacked into a matrix and fed into convolutional layers for side information incorporation.

One of the main challenges with CNN in this context is the need to determine an optimal filter size, which indicates the range of local patterns CNN can capture. Obviously, it is hard to determine optimal range in practice, thereby significantly limiting CNN's broader applications in the domain of SIDSBR [145].

3) *Attention Mechanisms*: Attention mechanism is good at discerning the importance of distinct tokens within a sequence, making it particularly useful for SBR. In this context, attention mechanism is typically employed to highlight interactions that are relevant to user preferences while downplaying less relevant ones, thereby enhancing the learning of user interests [30], [167], [31]. We can formulate the attention mechanism via representative SASRec [30] as follows,

$$\begin{aligned} \mathbf{o} &= \text{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V, \\ Q &= W^Q[\mathbf{v}_1; \mathbf{v}_2; \dots; \mathbf{v}_t], \\ K &= W^K[\mathbf{v}_1; \mathbf{v}_2; \dots; \mathbf{v}_t], \\ V &= W^V[\mathbf{v}_1; \mathbf{v}_2; \dots; \mathbf{v}_t], \end{aligned} \quad (3)$$

where W^Q , W^K and W^V are used to project inputs into Query, Key and Value spaces, respectively.

Currently, attention mechanism is widely adopted for analyzing user behaviors on various side information. For instance, NOVA [72] integrates side information by using item IDs as the Value and side attributes as the Query and Key. M2TRec [131] and MMSBR [7] first merge side information into unified item embeddings, and capture user intents via transformer architecture. In contrast, CBML [55] handles different types of side information separately, and performs information fusion at session-level.

However, attention-based models depend heavily on additional position information to capture sequential dependencies among items [20], [4]. The effectiveness of these models is highly sensitive to the schema of position encoding, leading to their unstable performance [146].

4) *Graph Neural Networks (GNN)*: GNN has become a dominant neural structure in SBR recently. Its strength lies in effectively capturing complex transition relationships among items within sessions, leading to the generation of informative item embeddings. Generally, GNN-based methods operate on

a specially constructed graph \mathcal{G} derived from session data, updating an item embedding (\mathbf{v}_j) via,

$$\mathbf{v}_j^* = g(\mathbf{v}_j, f_{x_k \in \mathcal{G}_j}(\mathbf{v}_k)), \quad (4)$$

where \mathcal{G}_j is a node set containing nodes adjacent to x_j within the \mathcal{G} , $f(*)$ is used to extract relevant semantic information from these neighboring nodes, and $g(*)$ updates the embedding of target node via aggregating the extracted semantics.

Under SIDSBR, existing methods usually resort to heterogeneous graphs to capture diverse types of side information such as MKM-SR [14], FAPAT [57], REKS [52] and GHTID [120], using item-level injection as shown in Fig. 3(b). Besides, some efforts extend general graphs to hypergraphs with degree-free hyperedges for capturing high-order relations among different kinds of nodes, as seen in models like CoHHN [12], BiPNet [13] and MBHT [79].

Despite their significant contributions to advancing SIDSBR, GNN-based methods suffer from some limitations, such as the need for manually constructing graphs and relying on heuristic aggregating algorithms. These limitations pose major obstacles to their practical implementations, restricting their broader applicability in real-world cases.

5) *Contrastive Learning (CL)*: CL, as a classical paradigm of self-supervised learning, has been a potent remedy for alleviating the inherent data sparsity issue in SBR [147], [45]. Typically, contrastive loss of CL-based methods can be formulated as follows,

$$\mathcal{L}_{con} = - \sum \frac{\langle \mathbf{r}, \mathbf{r}^+ \rangle}{\langle \mathbf{r}, \mathbf{r}^+ \rangle + \sum_k \langle \mathbf{r}, \mathbf{r}_k^- \rangle}, \quad (5)$$

where $\langle * \rangle$ is similarity function, \mathbf{r}^+ and \mathbf{r}_k^- are positive and negative samples about \mathbf{r} , respectively. CL-based methods typically obtain positive sessions by cropping, masking, or reordering certain items within a session, while viewing other sessions in the same batch as negatives. Subsequently, they are able to refine session embeddings via pulling positives closer and pushing negatives apart.

In the context of SIDSBR, current CL-based methods often leverage sequences from different types of side information to construct positives, such as in S^3 -Rec [68], HPM [148], MSSR [149] and Ti-model [150].

The primary limitation of this technique lies in its inability to guarantee the generation of robust positives and negatives. Concretely, it is challenging to generate positives with truly similar semantics or negatives with clearly dissimilar meanings, leading to less effective self-supervised learning.

6) *Large Language Models (LLM)*: LLM, known for its extensive world knowledge, is an emerging technique in the domain of Natural Language Processing (NLP), achieving impressing performance in various tasks. Generally, the current usage of LLM in recommendation tasks can be summarized into two manners: LLM as an assistant and LLM as a recommender. In the first approach, relevant methods utilize well-trained LLM to extract specific information from raw data to improve their understanding of user intent. For instance, FineRec [40] employs a LLM to process user-item reviews and extract item attributes that users care about. In the other approach, LLMs are directly trained using fine-tuning,

prompt tuning or instruction tuning based on recommendation data [151], [152], [153]. Specifically, these methods focus on creating informative prompts, as discussed in prompt-level injection of Section V-B3, to evoke LLM's capability in generating personalized suggestions.

Despite opening up a promising avenue to handle SBR, LLM still suffers from several obstacles that hinder their effective applications in the task. Firstly, while LLMs are trained on vast datasets from general domains, they lack the specific knowledge required for recommendation tasks, *e.g.*, collaborative knowledge, leading to their limitations in accurately understanding user behaviors. Secondly, the hallucination issue poses a significant challenge, where LLM-generated recommendations can sometimes deviate from reality. Although this may occur infrequently, it can negatively impact users' experience and diminish their trust in the system. Lastly, with a massive number of parameters, often in the billions, LLM's implementation in recommendation scenarios poses challenges related to both efficiency and effectiveness, making practical deployment difficult.

VI. RESEARCH PROGRESS

With substantial potential in revealing user intents from various perspectives, side information has received increasing attention in SBR recently. As discussed previously, the types of side information are of great importance in guiding the design and implementation of recommendation algorithms. To this end, we introduce a new taxonomy that places particular emphasis on the type of side information to examine research progress within SIDSBR. Our objective with this novel taxonomy is to offer a unique and insightful reference for future research, enabling researchers and practitioners to conveniently and efficiently identify, replicate, and evaluate algorithms that focus on specific types of side information. This unique approach not only facilitates a comprehensive understanding of existing methods but also highlights areas for potential advancements related to various information in the field. In the subsequent sections, we will outline the representative methods categorized by this taxonomy and point out their limitations. For clarity, Table II offers a comprehensive summary of representative models within SIDSBR, detailing the publication venues, types of side information integrated, and datasets utilized in these approaches.

A. Time-driven SBR

Time has been recognized as a crucial indicator of users' changing preferences and has been widely studied [3], [22], [2]. There are two primary manners to integrate time information into SBR: *temporal order* and *time interval*.

Temporal order models only take the order of user-item interactions into account, emphasizing the relative position of interactions within a session. Specifically, these methods can be further divided into two categories: implicit and explicit modeling. On one hand, implicit methods rely on RNN and its variants, which inherently process data in sequential order, to implicitly capture sequential patterns among items without needing extra time-related inputs, such as GRU4Rec [3],

NARM [2] and RepeatNet [99]. On the other hand, explicit efforts directly inject position embeddings into original item embeddings to achieve time-aware item embeddings including [77] with KNN, [163], [130], [164] with GNN and [30], [167] with self-attention. In contrast, time interval models focus on the specific duration between interactions, assuming that longer browsing times could indicate stronger user preferences. These methods usually discretize continuous time intervals into one-hot encoding and leverage self-attention [80], [92], [165] or contrastive learning [96], [150] to capture temporal patterns.

Despite the extensive investigation of time information, most methods primarily concentrate on straightforward sequential dependencies, typically from the single perspective of item order. In reality, time information offers rich insights into user behavior from various perspectives. For example, considering time from a daily perspective can reveal users' habits, such as a tendency for fast-food at noon and full-meals at night. Incorporating diverse time views, such as daily routines, seasonal changes, or even yearly trends, could unlock deeper insights into user preferences and lead to more contextually appropriate recommendations.

B. Category-driven SBR

The category feature suggests a user's hierarchical decision-making process, offering the potential to narrow down the range of candidate items. As the pioneering efforts incorporate item categories into this task, CoCoRec [26] relies on self-attention layers to explore user in-category preferences for improving recommendation performance. Other approaches model item ID sequences and category sequences simultaneously using GNN [56] or self-attention [68] for comprehensive session representation learning.

Nevertheless, existing methods typically convert item categories into one-hot encoding to serve as inputs for neural networks. This paradigm, while straightforward, falls short of capturing the hierarchical associations among categories, where this hierarchical feature is important in refining user intents. As a result, it leads to a loss of valuable contextual information and impedes the effective utilization of category information in the task.

C. Brand-driven SBR

The brand serves as a distinctive indicator of a manufacturer's reputation among available side information. It holds significant influence over user behavior, particularly in cases driven by the 'fandom economy' as highlighted in economic research. Despite its potential impact, unfortunately, none of the existing methods have specifically focused on incorporating brand information into the task of SIDSBR. Instead, brand data is usually bundled with other types of side information within current methods. A common practice involves combining brand attributes with categories, often employing separate self-attention layers to model item ID and side information independently, thereby enhancing the representation of sessions and improving recommendation performance [135], [55], [123].

TABLE II
AN OVERVIEW OF THE SIDE INFORMATION-DRIVEN SESSION-BASED RECOMMENDATION APPROACHES.

Models	Venues	Types of Incorporated Information									Datasets
		Category	Brand	Price	Text	Image	Address	Rating	Review	Behavior	
P-RNN [9]	RecSys16'				✓	✓					Private
3D-CNN [143]	RecSys17'	✓			✓						Private
DER [154]	AAAI19'								✓		Amazon, Yelp
FDSA [135]	IJCAI19'	✓	✓								Amazon, Tmall
RNS [33]	IJCAI19'								✓		Amazon
TMRN [67]	WSDM19'	✓									Amazon, JingDong, LastFm
S ³ -Rec [68]	CIKM20'	✓	✓								Amazon, LastFm, Yelp
GeoSAN [59]	KDD20'						✓				Foursquare
MKM-SR [14]	SIGIR20'	✓	✓							✓	JingDong
MV-RNN [141]	TKDE20'				✓	✓					Private
ESRM-KG [10]	WWW20'				✓						Private
NOVA [72]	AAAI21'	✓						✓			MovieLens
CBML [55]	CIKM21'	✓	✓								Yoochoose, Diginetica
CoCoRec [26]	SIGIR21'	✓									Beer, Taobao
MML [155]	CIKM22'				✓	✓					Private
NextIP [156]	CIKM22'									✓	Tmall
EMBSR [15]	ICDE22'									✓	JingDong, Trivago
MBHT [79]	KDD22'									✓	Retailrocket
UniSRec [11]	KDD22'	✓	✓		✓						Amazon
CARCA [104]	RecSys22'	✓	✓			✓					Amazon
GPG4HSR [157]	RecSys22'									✓	MovieLens, Yoochoose, Tmall
M2TRec [131]	RecSys22'	✓			✓						Diginetica
CoHHN [12]	SIGIR22'	✓		✓							Amazon, Cosmetics, Diginetica
DIF-SR [158]	SIGIR22'	✓									Amazon, Yelp
MB-STR [87]	SIGIR22'									✓	Taobao, Yelp
MGS [56]	SIGIR22'	✓	✓								Diginetica, Tmall
DETAIn [27]	WWW22'	✓	✓	✓				✓			Amazon, LastFM
MMSR [112]	CIKM23'				✓	✓					Amazon
M3SRec [51]	CIKM23'				✓	✓					Amazon
TASTE [97]	CIKM23'				✓						Amazon, Yelp
REKS [52]	ICDE23'	✓	✓								Amazon, MovieLens
CCA [159]	ICDM23'								✓		Amazon
Recformer [160]	KDD23'	✓	✓		✓						Amazon
FAPAT [57]	NeurIPS23'	✓	✓								Diginetica, Tmall, Private
DLFS-Rec [123]	RecSys23'	✓	✓								Amazon
HPM [148]	SIGIR23'	✓									Amazon
KMVG [53]	SIGIR23'	✓	✓								Amazon, Cosmetics, Yelp
LP-MRGNN [86]	TKDE23'									✓	Yoochoose, Taobao, Private
MMSBR [7]	TKDE23'			✓	✓	✓					Amazon
BiPNet [13]	TOIS23'	✓	✓	✓							Amazon, Cosmetics
DUVRec [74]	TOIS23'	✓									Amazon, MovieLens, Steam
DisenPOI [61]	WSDM23'						✓				Foursquare
MMMLP [134]	WWW23'				✓	✓					MovieLens
DWSRec [42]	AAAI24'				✓						Amazon
DIMO [8]	SIGIR24'				✓	✓					Amazon, Instacart
LLM4ISR [153]	SIGIR24'				✓						Amazon, MovieLens
GHTID [120]	WSDM24'									✓	MovieLens, Yoochoose, Tmall
MSSR [149]	WSDM24'	✓									Amazon, Yelp
ASIF [171]	WWW24'	✓	✓								Amazon, Yelp, Private
END4Rec [85]	WWW24'									✓	Diginetica, Yoochoose, Taobao
LLM-TRSR [152]	WWW24'				✓						Amazon
M-KGHT [161]	WWW24'	✓	✓	✓	✓						Diginetica, Yoochoose, Private
SLIM [162]	WWW24'				✓						Amazon
SAID [151]	WWW24'				✓						Amazon

D. Price-driven SBR

Price, as a pivotal factor users concern when purchasing items, plays a crucial role in influencing their decisions. Surprisingly, limited research has incorporated price information

into the domain of SBR. A notable attempt is made by CoHHN [12], which constructs a heterogeneous hypergraph encoding item ID, price, and categories simultaneously to simultaneously capture a user's price and interest preferences.

Building on this, BiPNet [13] introduces additional item brand information to further enhance the model's ability to capture user sensitivity to price variations. Given its unique and critical role in the user decision-making process, we believe that there is considerable potential for further research to more effectively integrate price factors into solutions of SIDSBR.

E. Text-driven SBR

As a primary way to display items on the website, text elements, including item titles and descriptions, enable users to quickly grasp the essence of items, directly influencing their choices. Various approaches have been developed to incorporate this textual data into SBR, aiming to leverage its rich semantics for pinpointing user intents. For instance, ESRM-KG [10] enhances recommendation performance through a keyword generation task which is designed to uncover shared semantics among items with distinct IDs. Some recent models, like UniSRec [11] and Reformer [160], introduce pre-trained NLP techniques into user behavior modeling, aiming at transferring linguistic knowledge to recommendation domain. These approaches typically represent an item via modeling its text, train models following the training-fine-tuning paradigm, and achieve item and sequence representation learning. Due to its remarkable success in the field of NLP, Transformer architecture is prevalent in these text-driven methods, as seen in models such as M2TRec [131] and DWSRec [42]. Additionally, increasing efforts attempt to fine-tune Large Language Models based on item sequences for LLM's adaption on recommendation task. In these solutions, an item is also represented by its text. In this emerging paradigm for recommendation task, recent methods usually focus on constructing effective prompts based on item text to unleash LLM's power on capturing user intents. Representative instances include summarization prompts in LLM-TRSR [152], chain-of-thought prompting in SLIM [162], as well as generative prompts in LLM4ISR [153].

Despite the advancements in text-driven SBR methods, existing methods, no matter based on neural networks or LLMs, typically process item text data as a whole, *i.e.*, they often convert lengthy descriptions into a single dense embedding. This paradigm overlooks to distinguish between relevant and irrelevant information and fails to prioritize key features that are critical to user preferences. As a result, valuable semantic details that could offer insights into user intents are lost, thereby limiting the model's ability to generate effective recommendations.

F. Image-involved Multi-modal SBR

Images offer a vivid portrayal of an item's appearance, highlighting visual features like color and style. Due to their intuitive and appealing nature, images typically occupy a prominent position on websites. In practice, images are frequently combined with text, formulating multi-modal recommendation tasks. A representative approach specially designed for multi-modal SBR is MMSBR [7]. This model accounts for distinct user behavior patterns across different modalities, aiming to capture user preferences holistically. In

addition, various techniques have been utilized to fuse multi-modal embeddings for a comprehensive understanding of user intents, including simple weighted sum [9], Multi-Layer Perceptron (MLP) [134], self-attention mechanism [155], [51] and GNN [112].

Despite the promising potential, multi-modal SBR is still in its infancy. Several critical challenges remain in this rapidly evolving domain, such as handling modality heterogeneity, addressing noisy information inherent in modalities, uncovering user preferences unique to each modality, and generating explainable recommendations.

G. Point-of-interest SBR

Address information, which provides the geographical context of items, is essential for a specialized recommendation scenario known as Point-of-Interest (POI) SBR [166]. In the task of POI SBR, user actions are influenced by both sequential behaviors and geographical proximity. Recent research has increasingly focused on integrating spatial and temporal information to enhance Point-of-interest SBR. For instance, GeoSAN [59] emphasizes the exact GPS coordinates of locations, while STiSAN [60] prioritizes time intervals between consecutive visits. DisenPOI [61] takes a step further by disentangling sequential and geographical relationships using contrastive learning techniques, aiming to enhance the quality of embeddings for more accurate recommendations.

H. Rating-driven SBR

Rating data provides a direct and explicit measure of user preferences towards items, offering valuable insights into the user attitudes. While rating prediction was a primary focus in the early stages of recommender system's research, its application has been somewhat overlooked in the realm of session-based recommendation. There are only a few methods taking rating data into account in the topic. For instance, NOVA [72] presents a non-invasive fusion mechanism based on self-attention to fuse ID, rating, and category information for user sequential behavior modeling. In this setup, the fused information serves as the Key and Query vectors, while item IDs act as the Value vectors. Given that ratings can explicitly deliver user preferences, we argue that this information warrants deeper exploration in SBR. For example, replacing implicit user actions with explicit ratings as supervised signals could provide a more direct and meaningful guide for model training. This approach would better leverage user-stated preferences, providing clearer indications of user intent and enhancing the effectiveness of SBR models.

I. Review-driven SBR

Reviews, posted by users after consuming products or services, offer detailed insights into both user preferences and item characteristics [40]. To this end, incorporating user-item reviews into recommendation models has attracted increasing attention recently. As the pioneering work in this topic, RNS [33] represents items with the reviews they received, employs an attention mechanism to learn a user's short- and

long-term interests, and merges these interests to strengthen model's expressive ability. DER [154] focuses on profiling items by extracting informative reviews while also considering time intervals between actions to better capture dynamic user preferences, improving both accuracy and explainability of its suggestions. CCA [159] further exploits the complementary nature of individual reviews (*e.g.*, a single review of an item) and aggregated reviews (*e.g.*, all reviews an item received) to maximize the rich information of review data.

A major challenge in review-driven recommendation lies in handling review heterogeneity. Specifically, distinct users may provide significantly different, even contradictory, opinions on the same item, making it difficult to learn a consistent and robust representation for that item. Besides, existing approaches often treat reviews as a singular unit, aiming to capture user preferences from the review as a whole. However, within a single review, a user may express conflicting opinions over different attributes of an item. For instance, a review for a T-shirt might say: 'The color is perfect, while I just do not like the patterns'. This highlights the need for future research to develop methods that can analyze user-item reviews in a more fine-grained, attribute-specific manner.

J. Multi-behavior SBR

Reflecting diverse user intents, different behavior types have become a focal point in SBR recently, leading to the emergence of a new recommendation task known as 'Micro-Behavior' or 'Multi-behavior' session-based recommendation. Current solutions in this emerging area generally involve constructing separate sequences for items and behavior types, such as item sequences and operation sequences. These sequences are then jointly analyzed to forecast the user's next actions, accounting for the unique context and purpose associated with each type of behavior. For instance, MKM-SR [14] applies GNN and GRU to handle item sequences and operation sequences respectively, while EMBSR [15] processes item sequences with GNN and uses self-attention for operation sequences. Recent advances have also sought to mitigate the noise that arises from the incorporation of various behaviors, as seen in END4Rec [85] and GHTID [120].

At its initial stage, multi-behavior SBR suffers from several limitations. Firstly, by modeling item and behavior sequences separately, current methods are limited to capturing only the intra-dependencies within a single type of sequence. This approach overlooks the complex inter-dependencies that exist between different types of sequences, which are crucial for a comprehensive understanding of user behaviors. Secondly, the choice of neural architectures such as RNNs, GNNs, or Transformers significantly influences recommendation performance. However, current models typically select these structures based on empirical guidelines, making it challenging to identify the optimal architecture for a given scenario. Thirdly, current methods often focus on a predefined set of behavior types, which restricts their adaptability to new or evolving behaviors. This rigidity can impair the model's performance in real-world applications, where user interactions are continually evolving, requiring a more flexible and robust approach to handle the changing behavioral landscape.

K. Others

In addition to the aforementioned side information, recent studies in SBR have begun to delve into more complex types of data. This includes not just item side information, but also the relations between different types of information. For instance, a 'belong-to' relation might indicate that an item belongs to a particular category or brand. Knowledge Graphs (KGs), consisting of entity nodes and relation edges, are especially well-suited for representing such intricate relationships. In this context, an entity in a KG could be an item, brand, or category, while relations represent connections like 'belong-to' between these entities. Current methods, such as KSR [69], REKS [52], KMVG [53] and M-KGHT [161], integrate KGs into SBR to leverage the wealth of knowledge embedded in these graphs, aiming at generating more informative item embeddings for better recommendations.

A key challenge in KG-based SBR lies in the requirement for a well-structured KG, which often demands extensive domain expertise and considerable manual effort to construct and maintain. This issue seriously limits the development of KG in SBR.

VII. OPEN CHALLENGES AND FUTURE DIRECTIONS

In recent years, side information-driven session-based recommendation has witnessed significant growth, driven by advancements in neural network technologies ranging from RNN, CNN, attention mechanism, GNN and contrastive learning to LLMs. After offering a detailed review of the research practices in SIDSBR, we also identify and discuss several open research directions, highlighting opportunities for future breakthroughs and advancements in the field.

A. Strengthen the Utilization of Available Information in SBR

As previously discussed, each type of information, ranging from time and price to images and reviews, plays a distinct role in capturing user preferences. These diverse data sources are instrumental in uncovering specific user interests and delivering personalized services. Unfortunately, as noted in the section of Research Progress (Section VI), certain types of information, such as brand, price, and reviews, remain relatively under-explored in current SBR research. To address this gap, more efforts should be dedicated to incorporating these less-exploited data types into SBR, leveraging their unique value to enhance the understanding of user behaviors in this challenging task.

B. Joint Incorporation of Various Information into SBR

Current methods in SBR, as outlined in Table II, predominantly utilize only a limited subset of the available side information. The potential of combining multiple diverse information sources remains largely unexplored. This integration of various information could significantly enhance session data and provide a deeper insight into user behaviors. For example, merging time information with color attributes derived from item images could uncover seasonal trends in user interests, such as a tendency towards lighter colors in summer

apparel and darker tones in winter wear. Obviously, such joint incorporation endows recommender systems with the potential to capture user preferences in a fine-grained manner, enabling the delivery of highly personalized and, at times, unexpectedly relevant suggestions. Furthermore, leveraging a broader range of information sources can effectively mitigate the data sparsity challenge that is inherent in SBR.

C. Side Information Enhanced Cold-start SBR

The cold-start problem remains a persistent challenge in recommender systems, where the systems fall short of offering suggestions about new items. Despite its prominence, this long-standing issue has received limited attention within SBR. The root cause of the cold-start problem lies in the lack of collaborative relationships between new and existing items, *i.e.*, item-level co-occurrence patterns. Consequently, the mainstream ID-based methods struggle to recommend newly added items effectively. Fortunately, the incorporation of side information presents a promising avenue to tackle this issue by enabling recommendations based on collaborative relations at fine-grained feature-level instead of coarse-grained item-level. For instance, it becomes feasible to recommend a new action movie starring Jackie Chan to a user who has previously enjoyed films featuring him, by leveraging detailed side information like the cast. This approach allows the system to draw meaningful connections and deliver relevant suggestions, even in the absence of direct historical data.

D. Side Information Driven Explainable SBR

Providing explanations alongside recommendations can significantly enhance user satisfaction and trust in recommender systems. Nevertheless, explainable SBR has not yet received extensive attention. Side information, with its rich semantics, serves as a valuable source for uncovering user specific interests, contributing to generating personalized and insightful explanations. For instance, as illustrated in the second row of Fig. 1, a user's preference for Marvel-related items can be inferred from the images and text of his previously purchased items. Based on these fine-grained interests, the system can generate personalized explanations such as, 'You have purchased items featuring Marvel themes, so we recommend a T-shirt with a Marvel logo'. This approach offers clear justifications for its suggestions, enhancing system transparency and user engagement. As a result, leveraging side information to create explainable SBR would emerge as an important direction for future research.

E. LLM-based SBR with Side Information

Large Language Models (LLMs) have shown remarkable capabilities across various tasks and are increasingly being explored in the field of recommender systems. However, their initial applications in this domain have not been as successful as anticipated. We contend that while LLMs encapsulate rich general knowledge, they struggle to learn the collaborative signals reflected by item IDs, which are of great significance under the scenario of recommendation. This challenge is

particularly evident in SBR, where sessions often involve limited user actions. As a result, current LLM-based methods have yet to match the performance of mainstream ID-based neural models. Fortunately, side information discussed in this survey, especially item text, presents a significant opportunity to unleash the full potential of LLMs in SBR. Since LLMs are adept at handling modality-rich data, like item titles and images, they can better grasp user preferences hidden in such information. Therefore, further investigation into LLM-based SBR with side information is not only warranted but essential for advancing this field.

F. Side Information-based Benchmarks for SBR

As outlined in Table I, none of the existing datasets cover all types of information required for a holistic evaluation of methods in this field. This limitation poses a significant challenge to conducting fair performance comparisons. For instance, a method may excel in leveraging behavior types using 'Behavior' information on the Cosmetics, while another one might concentrate on capturing user preferences from item images using 'Images' information on the Amazon. Lacking a dataset including both 'Behavior' and 'Images' information simultaneously, it becomes impossible to directly compare the performance of these methods. This gap underscores the urgent need for high-quality datasets that integrate diverse side information. However, a major challenge in collecting such data arises from privacy policies enforced by various online platforms. We encourage researchers and platform providers to consider publicly releasing such datasets, thereby supporting the development of this promising research topic.

VIII. CONCLUSIONS

The session-based recommendation is a hot-spot research topic due to its substantial application value in offering personalized services for anonymous users based on their limited actions. In recent times, there has been a surge in efforts to integrate side information into SBR as a means to address its inherent issues of data sparsity, giving rise to a promising topic of side information-driven session-based recommendation. In this domain, data plays a crucial role as it directly influences the design, implementation, and evaluation of recommendation algorithms. As a result, we elaborated on side information-driven SBR from a data-centric perspective in this work. This survey commences with formulating the paradigm of SIDSBR, details the available benchmarks, discusses the data characteristics and utility, illustrates the usage of various side information, comprehensively reviews the research progress, and points out future directions. We sincerely hope that our insights and analyses will serve as a valuable resource for researchers and practitioners, encouraging further innovations and developments in this vibrant topic.

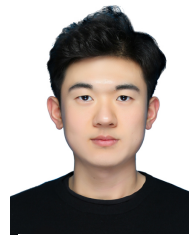
REFERENCES

- [1] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 5:1–5:38, 2019.
- [2] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural attentive session-based recommendation," in *CIKM*, 2017, pp. 1419–1428.
- [3] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," in *ICLR*, 2016.
- [4] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, and D. Lian, "A survey on session-based recommender systems," *ACM Comput. Surv.*, pp. 154:1–154:38, 2022.
- [5] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, and X. Zhang, "Self-supervised hypergraph convolutional networks for session-based recommendation," in *AAAI*, 2021, pp. 4503–4511.
- [6] H. Yang, Y. Choi, G. Kim, and J. Lee, "LOAM: improving long-tail session-based recommendation via niche walk augmentation and tail session mixup," in *SIGIR*, 2023, pp. 527–536.
- [7] X. Zhang, B. Xu, F. Ma, C. Li, L. Yang, and H. Lin, "Beyond co-occurrence: Multi-modal session-based recommendation," *IEEE Trans. Knowl. Data Eng.*, 2023.
- [8] X. Zhang, B. Xu, L. Ren, X. Wang, H. Lin, and F. Ma, "Disentangling ID and modality effects for session-based recommendation," in *SIGIR*. ACM, 2024, pp. 1883–1892.
- [9] B. Hidasi, M. Quadrana, A. Karatzoglou, and D. Tikk, "Parallel recurrent neural network architectures for feature-rich session-based recommendations," in *RecSys*. ACM, 2016, pp. 241–248.
- [10] Y. Liu, Z. Ren, W. Zhang, W. Che, T. Liu, and D. Yin, "Keywords generation improves e-commerce session-based recommendation," in *WWW*, 2020, pp. 1604–1614.
- [11] Y. Hou, S. Mu, W. X. Zhao, Y. Li, B. Ding, and J. Wen, "Towards universal sequence representation learning for recommender systems," in *KDD*, 2022, pp. 585–593.
- [12] X. Zhang, B. Xu, L. Yang, C. Li, F. Ma, H. Liu, and H. Lin, "Price DOES matter!: Modeling price and interest preferences in session-based recommendation," in *SIGIR*, 2022, pp. 1684–1693.
- [13] X. Zhang, B. Xu, F. Ma, C. Li, Y. Lin, and H. Lin, "Bi-preference learning heterogeneous hypergraph networks for session-based recommendation," *ACM Trans. Inf. Syst.*, vol. 42, no. 3, 2023.
- [14] W. Meng, D. Yang, and Y. Xiao, "Incorporating user micro-behaviors and item knowledge into multi-task learning for session-based recommendation," in *SIGIR*, 2020, pp. 1091–1100.
- [15] J. Yuan, W. Ji, D. Zhang, J. Pan, and X. Wang, "Micro-behavior encoding for session-based recommendation," in *ICDE*, 2022, pp. 2886–2899.
- [16] S. Wang, L. Hu, Y. Wang, L. Cao, Q. Z. Sheng, and M. A. Orgun, "Sequential recommender systems: Challenges, progress and prospects," in *IJCAI*. ijcai.org, 2019, pp. 6332–6338.
- [17] H. Fang, D. Zhang, Y. Shu, and G. Guo, "Deep learning for sequential recommendation: Algorithms, influential factors, and evaluations," *ACM Trans. Inf. Syst.*, vol. 39, no. 1, pp. 10:1–10:42, 2020.
- [18] Q. Liu, J. Hu, Y. Xiao, X. Zhao, J. Gao, W. Wang, Q. Li, and J. Tang, "Multimodal recommender systems: A survey," *ACM Comput. Surv.*, vol. 57, no. 2, pp. 26:1–26:17, 2025.
- [19] Z. Sun, Q. Guo, J. Yang, H. Fang, G. Guo, J. Zhang, and R. Burke, "Research commentary on recommendations with side information: A survey and research directions," *Electron. Commer. Res. Appl.*, vol. 37, 2019.
- [20] M. Ludewig, N. Mauro, S. Latifi, and D. Jannach, "Empirical analysis of session-based recommendation algorithms," *UMUAI*, pp. 1–33, 2020.
- [21] Z. Li, C. Yang, Y. Chen, X. Wang, H. Chen, G. Xu, L. Yao, and M. Sheng, "Graph and sequential neural networks in session-based recommendation: A survey," *ACM Comput. Surv.*, vol. 57, no. 2, pp. 40:1–40:37, 2025.
- [22] A. Dallmann, A. Grimm, C. Pölitz, D. Zoller, and A. Hotho, "Improving session recommendation with recurrent neural networks by exploiting dwell time," *CoRR*, vol. abs/1706.10231, 2017.
- [23] X. Zhang, H. Lin, B. Xu, C. Li, Y. Lin, H. Liu, and F. Ma, "Dynamic intent-aware iterative denoising network for session-based recommendation," *Inf. Process. Manag.*, vol. 59, p. 102936, 2022.
- [24] X. Zhang, H. Lin, L. Yang, B. Xu, Y. Diaio, and L. Ren, "Dual part-pooling attentive networks for session-based recommendation," *Neurocomputing*, vol. 440, pp. 89–100, 2021.
- [25] Y. Lv, L. Zhuang, P. Luo, H. Li, and Z. Zha, "Time-sensitive collaborative interest aware model for session-based recommendation," in *ICME*, 2020, pp. 1–6.
- [26] R. Cai, J. Wu, A. San, C. Wang, and H. Wang, "Category-aware collaborative sequential recommendation," in *SIGIR*. ACM, 2021, pp. 388–397.
- [27] K. Lin, Z. Wang, S. Shen, Z. Wang, B. Chen, and X. Chen, "Sequential recommendation with decomposed item feature routing," in *WWW*, 2022, pp. 2288–2297.
- [28] B. Hidasi and A. Karatzoglou, "Recurrent neural networks with top-k gains for session-based recommendations," in *CIKM*, 2018, pp. 843–852.
- [29] S. Wang, L. Hu, L. Cao, X. Huang, D. Lian, and W. Liu, "Attention-based transactional context embedding for next-item recommendation," in *AAAI*, 2018, pp. 2532–2539.
- [30] W. Kang and J. J. McAuley, "Self-attentive sequential recommendation," in *ICDM*, 2018, pp. 197–206.
- [31] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, "Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer," in *CIKM*, 2019, pp. 1441–1450.
- [32] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *AAAI*, 2019, pp. 346–353.
- [33] C. Li, X. Niu, X. Luo, Z. Chen, and C. Quan, "A review-driven neural model for sequential recommendation," in *IJCAI*, 2019, pp. 2866–2872.
- [34] Z. Wang, W. Wei, G. Cong, X. Li, X. Mao, and M. Qiu, "Global context enhanced graph neural networks for session-based recommendation," in *SIGIR*, 2020, pp. 169–178.
- [35] Z. Pan, F. Cai, W. Chen, H. Chen, and M. de Rijke, "Star graph neural networks for session-based recommendation," in *CIKM*, 2020, pp. 1195–1204.
- [36] F. Yuan, X. He, H. Jiang, G. Guo, J. Xiong, Z. Xu, and Y. Xiong, "Future data helps training: Modeling future contexts for session-based recommendation," in *WWW*, 2020, pp. 303–313.
- [37] C. Wei, B. Bai, K. Bai, and F. Wang, "Gsl4rec: Session-based recommendations with collective graph structure learning and next interaction prediction," in *WWW*. ACM, 2022, pp. 2120–2130.
- [38] X. Xin, L. Yang, Z. Zhao, P. Ren, Z. Chen, J. Ma, and Z. Ren, "On the effectiveness of unlearning in session-based recommendation," in *WSDM*. ACM, 2024, pp. 855–863.
- [39] N. Li, X. Ban, C. Ling, C. Gao, L. Hu, P. Jiang, K. Gai, Y. Li, and Q. Liao, "Modeling user fatigue for sequential recommendation," in *SIGIR*. ACM, 2024, pp. 996–1005.
- [40] X. Zhang, B. Xu, Y. Wu, Y. Zhong, H. Lin, and F. Ma, "Finerrec: Exploring fine-grained sequential recommendation," in *SIGIR*. ACM, 2024, pp. 1599–1608.
- [41] P. Li, M. Que, and A. Tuzhilin, "Dual contrastive learning for efficient static feature representation in sequential recommendations," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 2, pp. 544–555, 2024.
- [42] L. Zhang, X. Zhou, Z. Zeng, and Z. Shen, "Dual-view whitening on pre-trained text embeddings for sequential recommendation," in *AAAI*. AAAI Press, 2024, pp. 9332–9340.
- [43] M. Choi, H. Kim, H. Cho, and J. Lee, "Multi-intent-aware session-based recommendation," in *SIGIR*. ACM, 2024, pp. 2532–2536.
- [44] J. Wang, K. Ding, Z. Zhu, and J. Caverlee, "Session-based recommendation with hypergraph attention networks," in *SDM*, 2021, pp. 82–90.
- [45] X. Xie, F. Sun, Z. Liu, S. Wu, J. Gao, J. Zhang, B. Ding, and B. Cui, "Contrastive learning for sequential recommendation," in *38th IEEE International Conference on Data Engineering*, 2022, pp. 1259–1273.
- [46] Z. Li, A. Sun, and C. Li, "Diffurec: A diffusion model for sequential recommendation," *ACM Trans. Inf. Syst.*, vol. 42, no. 3, pp. 66:1–66:28, 2024.
- [47] Y. Lin, C. Wang, Z. Chen, Z. Ren, X. Xin, Q. Yan, M. de Rijke, X. Cheng, and P. Ren, "A self-correcting sequential recommender," in *WWW*. ACM, 2023, pp. 1283–1293.
- [48] C. Huang, J. Chen, L. Xia, Y. Xu, P. Dai, Y. Chen, L. Bo, J. Zhao, and J. X. Huang, "Graph-enhanced multi-task learning of multi-level transition dynamics for session-based recommendation," in *AAAI*, 2021, pp. 4123–4130.
- [49] W. Chen, F. Cai, H. Chen, and M. de Rijke, "A dynamic co-attention network for session-based recommendation," in *CIKM*, 2019, pp. 1461–1470.
- [50] C. Xu, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, F. Zhuang, J. Fang, and X. Zhou, "Graph contextualized self-attention network for session-based recommendation," in *IJCAI*, 2019, pp. 3940–3946.
- [51] S. Bian, X. Pan, W. X. Zhao, J. Wang, C. Wang, and J. Wen, "Multi-modal mixture of experts representation learning for sequential recommendation," in *CIKM*. ACM, 2023, pp. 110–119.

- [52] H. Wu, H. Fang, Z. Sun, C. Geng, X. Kong, and Y. Ong, "A generic reinforced explainable framework with knowledge graph for session-based recommendation," in *ICDE*. IEEE, 2023, pp. 1260–1272.
- [53] Q. Chen, Z. Guo, J. Li, and G. Li, "Knowledge-enhanced multi-view graph neural networks for session-based recommendation," in *SIGIR*. ACM, 2023, pp. 352–361.
- [54] Q. Chen, J. Li, Z. Guo, G. Li, and Z. Deng, "Attribute-enhanced dual channel representation learning for session-based recommendation," in *CIKM*. ACM, 2023, pp. 3793–3797.
- [55] J. Song, J. Xu, R. Zhou, L. Chen, J. Li, and C. Liu, "CBML: A cluster-based meta-learning model for session-based recommendation," in *CIKM*, 2021, pp. 1713–1722.
- [56] S. Lai, E. Meng, F. Zhang, C. Li, B. Wang, and A. Sun, "An attribute-driven mirror graph network for session-based recommendation," in *SIGIR*, 2022, pp. 1674–1683.
- [57] X. Liu, Z. Li, Y. Gao, J. Yang, T. Cao, Z. Wang, B. Yin, and Y. Song, "Enhancing user intent capture in session-based recommendation with attribute patterns," in *NeurIPS*, 2023.
- [58] J. Tang and K. Wang, "Personalized top-n sequential recommendation via convolutional sequence embedding," in *WSDM*, 2018, pp. 565–573.
- [59] D. Lian, Y. Wu, Y. Ge, X. Xie, and E. Chen, "Geography-aware sequential location recommendation," in *KDD*. ACM, 2020, pp. 2009–2019.
- [60] E. Wang, Y. Jiang, Y. Xu, L. Wang, and Y. Yang, "Spatial-temporal interval aware sequential POI recommendation," in *ICDE*. IEEE, 2022, pp. 2086–2098.
- [61] Y. Qin, Y. Wang, F. Sun, W. Ju, X. Hou, Z. Wang, J. Cheng, J. Lei, and M. Zhang, "Disenpoi: Disentangling sequential and geographical influence for point-of-interest recommendation," in *WSDM*. ACM, 2023, pp. 508–516.
- [62] M. Wan and J. J. McAuley, "Item recommendation on monotonic behavior chains," in *RecSys*. ACM, 2018, pp. 86–94.
- [63] Y. Zhang, Y. Liu, Y. Xu, H. Xiong, C. Lei, W. He, L. Cui, and C. Miao, "Enhancing sequential recommendation with graph contrastive learning," in *IJCAI*. ijcai.org, 2022, pp. 2398–2405.
- [64] C. Ma, L. Ma, Y. Zhang, J. Sun, X. Liu, and M. Coates, "Memory augmented graph neural networks for sequential recommendation," in *AAAI*. AAAI Press, 2020, pp. 5045–5052.
- [65] Y. Li, H. Chen, X. Sun, Z. Sun, L. Li, L. Cui, P. S. Yu, and G. Xu, "Hyperbolic hypergraphs for sequential recommendation," in *CIKM*. ACM, 2021, pp. 988–997.
- [66] Y. Qin, P. Wang, and C. Li, "The world is binary: Contrastive learning for denoising next basket recommendation," in *SIGIR*, 2021, pp. 859–868.
- [67] J. Huang, Z. Ren, W. X. Zhao, G. He, J. Wen, and D. Dong, "Taxonomy-aware multi-hop reasoning networks for sequential recommendation," in *WSDM*. ACM, 2019, pp. 573–581.
- [68] K. Zhou, H. Wang, W. X. Zhao, Y. Zhu, S. Wang, F. Zhang, Z. Wang, and J. Wen, "S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization," in *CIKM*. ACM, 2020, pp. 1893–1902.
- [69] J. Huang, W. X. Zhao, H. Dou, J. Wen, and E. Y. Chang, "Improving sequential recommendation with knowledge-enhanced memory networks," in *SIGIR*, 2018, pp. 505–514.
- [70] P. Wang, Y. Fan, L. Xia, W. X. Zhao, S. Niu, and J. X. Huang, "KERL: A knowledge-guided reinforcement learning model for sequential recommendation," in *SIGIR*. ACM, 2020, pp. 209–218.
- [71] W. Ye, S. Wang, X. Chen, X. Wang, Z. Qin, and D. Yin, "Time matters: Sequential recommendation with complex temporal information," in *SIGIR*, 2020, pp. 3940–3946.
- [72] C. Liu, X. Li, G. Cai, Z. Dong, H. Zhu, and L. Shang, "Noninvasive self-attention for side information fusion in sequential recommendation," in *AAAI*, 2021, pp. 4249–4256.
- [73] H. Liu, Y. Zhu, and Z. Wu, "Knowledge graph-based behavior denoising and preference learning for sequential recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 6, pp. 2490–2503, 2024.
- [74] L. Xue, D. Yang, S. Zhai, Y. Li, and Y. Xiao, "Learning dual-view user representations for enhanced sequential recommendation," *ACM Trans. Inf. Syst.*, vol. 41, no. 4, pp. 86:1–86:26, 2023.
- [75] J. Fu, X. Ge, X. Xin, A. Karatzoglou, I. Arapakis, J. Wang, and J. M. Jose, "IISAN: efficiently adapting multimodal representation for sequential recommendation with decoupled PEFT," in *SIGIR*. ACM, 2024, pp. 687–697.
- [76] Y. Li, H. Du, Y. Ni, P. Zhao, Q. Guo, F. Yuan, and X. Zhou, "Multi-modality is all you need for transferable recommender systems," in *ICDE*. IEEE, 2024, pp. 5008–5021.
- [77] D. Garg, P. Gupta, P. Malhotra, L. Vig, and G. M. Shroff, "Sequence and time aware neighborhood for session-based recommendations: STAN," in *SIGIR*. ACM, 2019, pp. 1069–1072.
- [78] P. Gupta, D. Garg, P. Malhotra, L. Vig, and G. Shroff, "Niser: Normalized item and session representations to handle popularity bias," *arXiv preprint arXiv:1909.04276*, 2019.
- [79] Y. Yang, C. Huang, L. Xia, Y. Liang, Y. Yu, and C. Li, "Multi-behavior hypergraph-enhanced transformer for sequential recommendation," in *KDD*. ACM, 2022, pp. 2263–2274.
- [80] J. Li, Y. Wang, and J. J. McAuley, "Time interval aware self-attention for sequential recommendation," in *WSDM*. ACM, 2020, pp. 322–330.
- [81] Y. Zheng, S. Liu, Z. Li, and S. Wu, "Cold-start sequential recommendation via meta learner," in *AAAI*. AAAI Press, 2021, pp. 4706–4713.
- [82] J. J. Seol, Y. Ko, and S. Lee, "Exploiting session information in bert-based session-aware sequential recommendation," in *SIGIR*. ACM, 2022, pp. 2639–2644.
- [83] Y. Zhang, X. Wang, H. Chen, and W. Zhu, "Adaptive disentangled transformer for sequential recommendation," in *KDD*. ACM, 2023, pp. 3434–3445.
- [84] H. Zhu, D. Chang, Z. Xu, P. Zhang, X. Li, J. He, H. Li, J. Xu, and K. Gai, "Joint optimization of tree-based index and deep model for recommender systems," in *NeurIPS*, 2019, pp. 3973–3982.
- [85] Y. Han, H. Wang, K. Wang, L. Wu, Z. Li, W. Guo, Y. Liu, D. Lian, and E. Chen, "Efficient noise-decoupling for multi-behavior sequential recommendation," in *WWW*. ACM, 2024, pp. 3297–3306.
- [86] W. Wang, W. Zhang, S. Liu, Q. Liu, B. Zhang, L. Lin, and H. Zha, "Incorporating link prediction into multi-relational item graph modeling for session-based recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 2683–2696, 2023.
- [87] E. Yuan, W. Guo, Z. He, H. Guo, C. Liu, and R. Tang, "Multi-behavior sequential transformer recommender," in *SIGIR*, 2022, pp. 1642–1652.
- [88] Y. Hou, B. Hu, Z. Zhang, and W. X. Zhao, "CORE: simple and effective session-based recommendation within consistent representation space," in *SIGIR*. ACM, 2022, pp. 1796–1801.
- [89] Q. Han, C. Zhang, R. Chen, R. Lai, H. Song, and L. Li, "Multi-faceted global item relation learning for session-based recommendation," in *SIGIR*, 2022, pp. 1705–1715.
- [90] K. Zhou, H. Yu, W. X. Zhao, and J. Wen, "Filter-enhanced MLP is all you need for sequential recommendation," in *WWW*, 2022, pp. 2388–2399.
- [91] S. Qiao, W. Zhou, J. Wen, H. Zhang, and M. Gao, "Bi-channel multiple sparse graph attention networks for session-based recommendation," in *CIKM*. ACM, 2023, pp. 2075–2084.
- [92] J. Wu, R. Cai, and H. Wang, "Dějà vu: A contextualized temporal attention mechanism for sequential recommendation," in *WWW*, 2020, pp. 2199–2209.
- [93] N. R. Kermany, J. Yang, J. Wu, and L. Pizzato, "Fair-srs: A fair session-based recommendation system," in *WSDM*. ACM, 2022, pp. 1601–1604.
- [94] Y. Pang, L. Wu, Q. Shen, Y. Zhang, Z. Wei, F. Xu, E. Chang, B. Long, and J. Pei, "Heterogeneous global graph neural networks for personalized session-based recommendation," in *WSDM*, 2022, pp. 775–783.
- [95] M. Zhang, S. Wu, M. Gao, X. Jiang, K. Xu, and L. Wang, "Personalized graph neural networks with attention mechanism for session-aware recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3946–3957, 2022.
- [96] C. Tian, Z. Lin, S. Bian, J. Wang, and W. X. Zhao, "Temporal contrastive pre-training for sequential recommendation," in *CIKM*. ACM, 2022, pp. 1925–1934.
- [97] Z. Liu, S. Mei, C. Xiong, X. Li, S. Yu, Z. Liu, Y. Gu, and G. Yu, "Text matching improves sequential recommendation by reducing popularity biases," in *CIKM*, 2023, pp. 1534–1544.
- [98] H. Chang, N. Agarwal, and A. McCallum, "To copy, or not to copy; that is a critical issue of the output softmax layer in neural sequential recommenders," in *WSDM*. ACM, 2024, pp. 67–76.
- [99] P. Ren, Z. Chen, J. Li, Z. Ren, J. Ma, and M. de Rijke, "Repeatnet: A repeat aware neural recommendation machine for session-based recommendation," in *AAAI*, 2019, pp. 4806–4813.
- [100] M. Choi, J. Kim, J. Lee, H. Shim, and J. Lee, "S-walk: Accurate and scalable session-based recommendation with random walks," in *WSDM*, 2022, pp. 150–160.
- [101] J. Chen, G. Zhu, H. Hou, C. Yuan, and Y. Huang, "Autogrs: Neural architecture search for graph-based session recommendation," in *SIGIR*, 2022, pp. 1694–1704.

- [102] Z. Yang, X. He, J. Zhang, J. Wu, X. Xin, J. Chen, and X. Wang, "A generic learning framework for sequential recommendation with distribution shifts," in *SIGIR*. ACM, 2023, pp. 331–340.
- [103] J. Li, T. Zhao, J. Li, J. Chan, C. Faloutsos, G. Karypis, S. Pantel, and J. J. McAuley, "Coarse-to-fine sparse sequential recommendation," in *SIGIR*. ACM, 2022, pp. 2082–2086.
- [104] A. Rashed, S. Elsayed, and L. Schmidt-Thieme, "Context and attribute-aware sequential recommendation via cross-attention," in *RecSys*, 2022, pp. 71–80.
- [105] C. Long, H. Yuan, J. Fang, X. Xian, G. Liu, V. S. Sheng, and P. Zhao, "Learning global and multi-granularity local representation with MLP for sequential recommendation," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 4, pp. 87:1–87:15, 2024.
- [106] H. Wu, H. Zhang, L. Li, Z. Chen, F. Zhu, and X. Fang, "Cheaper is better: Exploring price competitiveness for online purchase prediction," in *38th IEEE International Conference on Data Engineering, ICDE*, 2022, pp. 3399–3412.
- [107] Y. Zheng, C. Gao, X. He, Y. Li, and D. Jin, "Price-aware recommendation with graph convolutional networks," in *ICDE*, 2020, pp. 133–144.
- [108] M. Wan, D. Wang, M. Goldman, M. Taddy, J. Rao, J. Liu, D. Lymberopoulos, and J. J. McAuley, "Modeling consumer preferences and price sensitivities from large-scale grocery shopping transaction logs," in *WWW*, 2017, pp. 1103–1112.
- [109] Y. Wang, Z. Chu, X. Ouyang, S. Wang, H. Hao, Y. Shen, J. Gu, S. Xue, J. Zhang, Q. Cui, L. Li, J. Zhou, and S. Li, "LLMRG: improving recommendations through large language model reasoning graphs," in *AAAI*. AAAI Press, 2024, pp. 19 189–19 196.
- [110] X. Lin, W. Wang, Y. Li, S. Yang, F. Feng, Y. Wei, and T. Chua, "Data-efficient fine-tuning for llm-based recommendation," in *SIGIR*. ACM, 2024, pp. 365–374.
- [111] Y. Zhao, J. Wu, X. Wang, W. Tang, D. Wang, and M. de Rijke, "Let me do it for you: Towards LLM empowered recommendation via tool learning," in *SIGIR*. ACM, 2024, pp. 1796–1806.
- [112] H. Hu, W. Guo, Y. Liu, and M. Kan, "Adaptive multi-modalities fusion in sequential recommendation systems," in *CIKM*. ACM, 2023, pp. 843–853.
- [113] X. Zhou, H. Zhou, Y. Liu, Z. Zeng, C. Miao, P. Wang, Y. You, and F. Jiang, "Bootstrap latent representations for multi-modal recommendation," in *WWW*. ACM, 2023, pp. 845–854.
- [114] Y. Wei, X. Wang, L. Nie, X. He, and T. Chua, "Graph-refined convolutional network for multimedia recommendation with implicit feedback," in *MM*, 2020, pp. 3541–3549.
- [115] S. Geng, J. Tan, S. Liu, Z. Fu, and Y. Zhang, "VIP5: towards multimodal foundation models for recommendation," in *Findings of EMNLP*. Association for Computational Linguistics, 2023, pp. 9606–9620.
- [116] K. Liu, F. Xue, D. Guo, L. Wu, S. Li, and R. Hong, "MEGCF: multimodal entity graph collaborative filtering for personalized recommendation," *ACM Trans. Inf. Syst.*, vol. 41, no. 2, pp. 30:1–30:27, 2023.
- [117] C. Ma, Y. Zhang, Q. Wang, and X. Liu, "Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence," in *CIKM*. ACM, 2018, pp. 697–706.
- [118] J. Tang, H. Gao, X. Hu, and H. Liu, "Context-aware review helpfulness rating prediction," in *RecSys*. ACM, 2013, pp. 1–8.
- [119] K. Wang, W. X. Zhao, H. Peng, and X. Wang, "Bayesian probabilistic multi-topic matrix factorization for rating prediction," in *IJCAI*. IJCAI/AAAI Press, 2016, pp. 3910–3916.
- [120] X. Li, H. Chen, J. Yu, M. Zhao, T. Xu, W. Zhang, and M. Yu, "Global heterogeneous graph and target interest denoising for multi-behavior sequential recommendation," in *WSDM*. ACM, 2024, pp. 387–395.
- [121] M. Yan, F. Liu, J. Sun, F. Sun, Z. Cheng, and Y. Han, "Behavior-contextualized item preference modeling for multi-behavior recommendation," in *SIGIR*. ACM, 2024, pp. 946–955.
- [122] L. Guo, S. Zhou, H. Tang, X. Zheng, and Y. Luo, "Multi-behavior hypergraph contrastive learning for session-based recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 3, pp. 1325–1338, 2025.
- [123] H. Liu, Z. Deng, L. Wang, J. Peng, and S. Feng, "Distribution-based learnable filters with side information for sequential recommendation," in *RecSys*. ACM, 2023, pp. 78–88.
- [124] R. Sun, X. Cao, Y. Zhao, J. Wan, K. Zhou, F. Zhang, Z. Wang, and K. Zheng, "Multi-modal knowledge graphs for recommender systems," in *CIKM*. ACM, 2020, pp. 1405–1414.
- [125] D. Wang, X. Zhang, Y. Yin, D. Yu, G. Xu, and S. Deng, "Multi-view enhanced graph attention network for session-based music recommendation," *ACM Trans. Inf. Syst.*, vol. 42, no. 1, pp. 16:1–16:30, 2024.
- [126] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *EMNLP*, 2014, pp. 1532–1543.
- [127] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT*, 2019, pp. 4171–4186.
- [128] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015, pp. 1–9.
- [129] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *ICLR*, 2021.
- [130] R. Qiu, J. Li, Z. Huang, and H. Yin, "Rethinking the item order in session-based recommendation with graph neural networks," in *CIKM*. ACM, 2019, pp. 579–588.
- [131] W. Shalaby, S. Oh, A. Afsharnejad, S. Kumar, and X. Cui, "M2trec: Metadata-aware multi-task transformer for large-scale and cold-start free session-based recommendations," in *RecSys*, 2022, pp. 573–578.
- [132] T. Baltrusaitis, C. Ahuja, and L. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 423–443, 2019.
- [133] A. Nagrani, S. Yang, A. Arnab, A. Jansen, C. Schmid, and C. Sun, "Attention bottlenecks for multimodal fusion," in *NeurIPS*, 2021, pp. 14 200–14 213.
- [134] J. Liang, X. Zhao, M. Li, Z. Zhang, W. Wang, H. Liu, and Z. Liu, "MMMLP: multi-modal multilayer perceptron for sequential recommendations," in *WWW*, 2023, pp. 1109–1117.
- [135] T. Zhang, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, D. Wang, G. Liu, and X. Zhou, "Feature-level deeper self-attention network for sequential recommendation," in *IJCAI*, 2019, pp. 4320–4326.
- [136] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023.
- [137] Z. Yuan, F. Yuan, Y. Song, Y. Li, J. Fu, F. Yang, Y. Pan, and Y. Ni, "Where to go next for recommender systems? ID- vs. modality-based recommender models revisited," in *SIGIR*, 2023, pp. 2639–2649.
- [138] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: bayesian personalized ranking from implicit feedback," in *UAI*. AUAI Press, 2009, pp. 452–461.
- [139] G. Shani, D. Heckerman, and R. I. Brafman, "An mdp-based recommender system," *J. Mach. Learn. Res.*, pp. 1265–1295, 2005.
- [140] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *WWW*. ACM, 2001, pp. 285–295.
- [141] Q. Cui, S. Wu, Q. Liu, W. Zhong, and L. Wang, "MV-RNN: A multi-view recurrent neural network for sequential recommendation," *IEEE Trans. Knowl. Data Eng.*, pp. 317–331, 2020.
- [142] 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, pp. 3771–3777.
- [143] T. X. Tuan and T. M. Phuong, "3d convolutional networks for session-based recommendation with content features," in *RecSys*, 2017, pp. 138–146.
- [144] F. Yuan, A. Karatzoglou, I. Arapakis, J. M. Jose, and X. He, "A simple convolutional generative network for next item recommendation," in *WSDM*, 2019, pp. 582–590.
- [145] T. Chen and R. C. Wong, "Session-based recommendation with local invariance," in *ICDM*, 2019, pp. 994–999.
- [146] R. Qiu, Z. Huang, T. Chen, and H. Yin, "Exploiting positional information for session-based recommendation," *ACM Trans. Inf. Syst.*, vol. 40, no. 2, pp. 35:1–35:24, 2022.
- [147] X. Xia, H. Yin, J. Yu, Y. Shao, and L. Cui, "Self-supervised graph co-training for session-based recommendation," in *CIKM*, 2021, pp. 2180–2190.
- [148] C. Huang, S. Wang, X. Wang, and L. Yao, "Dual contrastive transformer for hierarchical preference modeling in sequential recommendation," in *SIGIR*. ACM, 2023, pp. 99–109.
- [149] X. Lin, J. Luo, J. Pan, W. Pan, Z. Ming, X. Liu, S. Huang, and J. Jiang, "Multi-sequence attentive user representation learning for side-information integrated sequential recommendation," in *WSDM*. ACM, 2024, pp. 414–423.
- [150] Y. Dang, E. Yang, G. Guo, L. Jiang, X. Wang, X. Xu, Q. Sun, and H. Liu, "Uniform sequence better: Time interval aware data augmentation for sequential recommendation," in *AAAI*. AAAI Press, 2023, pp. 4225–4232.

- [151] J. Hu, W. Xia, X. Zhang, C. Fu, W. Wu, Z. Huan, A. Li, Z. Tang, and J. Zhou, "Enhancing sequential recommendation via llm-based semantic embedding learning," in *WWW*. ACM, 2024, pp. 103–111.
- [152] Z. Zheng, W. Chao, Z. Qiu, H. Zhu, and H. Xiong, "Harnessing large language models for text-rich sequential recommendation," in *WWW*. ACM, 2024, pp. 3207–3216.
- [153] Z. Sun, H. Liu, X. Qu, K. Feng, Y. Wang, and Y. S. Ong, "Large language models for intent-driven session recommendations," in *SIGIR*. ACM, 2024, pp. 324–334.
- [154] X. Chen, Y. Zhang, and Z. Qin, "Dynamic explainable recommendation based on neural attentive models," in *AAAI*. AAAI Press, 2019, pp. 53–60.
- [155] X. Pan, Y. Chen, C. Tian, Z. Lin, J. Wang, H. Hu, and W. X. Zhao, "Multimodal meta-learning for cold-start sequential recommendation," in *CIKM*. ACM, 2022, pp. 3421–3430.
- [156] J. Luo, M. He, X. Lin, W. Pan, and Z. Ming, "Dual-task learning for multi-behavior sequential recommendation," in *CIKM*. ACM, 2022, pp. 1379–1388.
- [157] W. Chen, M. He, Y. Ni, W. Pan, L. Chen, and Z. Ming, "Global and personalized graphs for heterogeneous sequential recommendation by learning behavior transitions and user intentions," in *RecSys*. ACM, 2022, pp. 268–277.
- [158] Y. Xie, P. Zhou, and S. Kim, "Decoupled side information fusion for sequential recommendation," in *SIGIR*, 2022, pp. 1611–1621.
- [159] B. Huang, J. Luo, W. Du, W. Pan, and Z. Ming, "Cascaded cross attention for review-based sequential recommendation," in *ICDM*. IEEE, 2023, pp. 170–179.
- [160] J. Li, M. Wang, J. Li, J. Fu, X. Shen, J. Shang, and J. J. McAuley, "Text is all you need: Learning language representations for sequential recommendation," in *SIGKDD*. ACM, 2023, pp. 1258–1267.
- [161] Y. Wang, A. Javari, J. Balaji, W. Shalaby, T. Derr, and X. Cui, "Knowledge graph-based session recommendation with session-adaptive propagation," in *WWW*. ACM, 2024, pp. 264–273.
- [162] Y. Wang, C. Tian, B. Hu, Y. Yu, Z. Liu, Z. Zhang, J. Zhou, L. Pang, and X. Wang, "Can small language models be good reasoners for sequential recommendation?" in *WWW*. ACM, 2024, pp. 3876–3887.
- [163] T. Chen and R. C. Wong, "Handling information loss of graph neural networks for session-based recommendation," in *KDD*, 2020, pp. 1172–1180.
- [164] Y. Li, C. Gao, X. Du, H. Wei, H. Luo, D. Jin, and Y. Li, "Spatiotemporal-aware session-based recommendation with graph neural networks," in *CIKM*. ACM, 2022, pp. 1209–1218.
- [165] L. Chen, N. Yang, and P. S. Yu, "Time lag aware sequential recommendation," in *CIKM*. ACM, 2022, pp. 212–221.
- [166] Y. Zhang, G. Liu, A. Liu, Y. Zhang, Z. Li, X. Zhang, and Q. Li, "Personalized geographical influence modeling for POI recommendation," *IEEE Intell. Syst.*, vol. 35, no. 5, pp. 18–27, 2020.
- [167] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "STAMP: short-term attention/memory priority model for session-based recommendation," in *KDD*, 2018, pp. 1831–1839.
- [168] J. Zhang, Y. Cheng, Y. Ni, Y. Pan, Z. Yuan, J. Fu, Y. Li, J. Wang, and F. Yuan, "Ninerec: A benchmark dataset suite for evaluating transferable recommendation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–12, 2024.
- [169] Y. Du, D. Luo, R. Yan, X. Wang, H. Liu, H. Zhu, Y. Song, and J. Zhang, "Enhancing job recommendation through llm-based generative adversarial networks," in *AAAI*. AAAI Press, 2024, pp. 8363–8371.
- [170] F. Gutiérrez, S. Charleer, R. D. Croon, N. N. Htun, G. Goetschalckx, and K. Verbert, "Explaining and exploring job recommendations: a user-driven approach for interacting with knowledge-based job recommender systems," in *RecSys*. ACM, 2019, pp. 60–68.
- [171] S. Wang, B. Shen, X. Min, Y. He, X. Zhang, L. Zhang, J. Zhou, and L. Mo, "Aligned side information fusion method for sequential recommendation," in *WWW*. ACM, 2024, pp. 112–120.
- [172] Y. Wei, X. Wang, L. Nie, X. He, R. Hong, and T. Chua, "MMGCN: multi-modal graph convolution network for personalized recommendation of micro-video," in *MM*, 2019, pp. 1437–1445.



Xiaokun Zhang is currently a postdoc research fellow in the Department of Computer Science, City University of Hong Kong. He received his B.S., M.S. and Phd degrees in the School of Computer Science and Technology, Dalian University of Technology. His research interests include data mining and natural language processing, especially focusing on intelligent recommender systems.



Bo Xu received the BSc and PhD degrees from the Dalian University of Technology, China, in 2011 and 2018. He is currently an associate professor in School of Computer Science and Technology of Dalian University of Technology. His current research interests include information retrieval and natural language processing.



Chenliang Li is a full Professor with School of Cyber Science and Engineering, Wuhan University China. His research areas include Information Retrieval, Recommendation System, Natural Language Processing and Social Computing. He has published over 80 papers in leading conferences and journals, and was the recipient of SIGIR 2016 Best Student Paper Award Honorable Mention. Currently, He serves as an Associate Editor for ACM TOIS, ACM TALLIP, and an editorial board member of JASIST and IPM. He has been a PC member of many leading conference, such as SIGIR, WWW, ACL, WSDM, AAAI, CIKM.



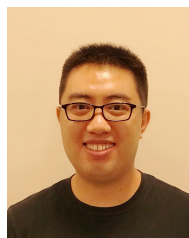
Bowei He is a Ph.D. candidate in Computer Science at City University of Hong Kong. His research focuses on advanced ML (e.g., RL, continual learning, and Auto-ML) and causal inference for data mining and information retrieval. His recent interests include efficient and personalized large language model. He has published papers on KDD, WWW, NeurIPS, ICDE, TKDE, CVPR, ICCV, CIKM, ICDM, UAI, SDM, etc. He interned and collaborated with Didi AI Labs, Huawei Noah's Ark Lab, Alibaba, Kuaishou, and Tencent Financial Technology.



Hongfei Lin received the BSc degree from the Northeastern Normal University in 1983, the MSc degree from the Dalian University of Technology in 1992, and the PhD degree from the Northeastern University in 2000. He is currently a professor in School of Computer Science and Technology at the Dalian University of Technology. He has published more than 500 research papers in various journals, conferences, and books. His research interests include information retrieval, text mining for biomedical literatures, biomedical hypothesis generation, information extraction from huge biomedical resources, learning-to-rank. He is the director of Information Retrieval Lab. at Dalian University of Technology.



Chen Ma is currently an Assistant Professor in the Department of Computer Science, City University of Hong Kong. His research interests lie in the theory of data mining and machine learning and their applications in recommender systems, knowledge graphs, social computing, and social good. He received his B.S. and M.S. degrees in Software Engineering from Beijing Institute of Technology in 2013 and 2015, respectively. He received his PhD degree in Computer Science from McGill University.



Fenglong Ma is an assistant professor in the College of Information Sciences and Technology at the Pennsylvania State University. He received his Ph.D. from the Department of Computer Science and Engineering, University at Buffalo (UB) in 2019. His research interests lie in data mining and machine learning, with an emphasis on mining health-related data. His research interests include natural language processing, federated learning, and AI security.