

Enhancing Repeat-Aware Recommendation from a Temporal-Sequential Perspective

Shigang Quan
Shanghai Jiao Tong University
Shanghai, China
quan123@sjtu.edu.cn

Zhenzhe Zheng
Shanghai Jiao Tong University
Shanghai, China
zhengzhenzhe@sjtu.edu.cn

Shui Liu
Alibaba Group
Hangzhou, Zhejiang, China
shui.lius@alibaba-inc.com

Fan Wu
Shanghai Jiao Tong University
Shanghai, China
fwu@cs.sjtu.edu.cn

ABSTRACT

Repeat consumption, such as repurchasing items and relistening songs, is a common scenario in daily life. To model repeat consumption, the repeat-aware recommendation has been proposed to predict which item will be re-interacted based on the user-item interactions. In this paper, we investigate various inherent characteristics to enhance the repeat-aware recommendation. Specifically, we explore these characteristics from two aspects: one is from the temporal aspect where we consider the time interval relationship in the user behavior sequence; the other is from the sequential aspect where we consider the sequential-level relationship in the user behavior sequence. And our intuition is that both the *temporal pattern* and *sequential pattern* will reflect users' intentions of repeat consumption.

By utilizing these two patterns, a novel model called Temporal and Sequential repeat-aware Recommendation (TSRec for short) is proposed to enhance repeat-aware recommendation. TSRec has three main components: 1) User-specific Temporal Representation Module (UTRM), which encodes and extracts user historical repeat temporal information. 2) Item-specific Temporal Representation Module (ITRM), which incorporates item time interval information as side information to alleviate the data sparsity problem of user repeat behavior sequence. 3) Sequential Repeat-Aware Module (SRAM), which represents the similarity between the user's current and the last repeat sequences. Extensive experimental results on three public benchmarks demonstrate the superiority of TSRec over state-of-the-art methods. The implementation code is available ¹.

CCS CONCEPTS

• Information systems → Recommender systems.

¹<https://anonymous.4open.science/r/TSRec-2306/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CIKM '23, October 21–25, 2023, Birmingham, United Kingdom.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0124-5/23/10...\$15.00
<https://doi.org/10.1145/XXXXXX.XXXXXX>

KEYWORDS

Repeat-Aware Recommendation; Sequence Matching; Temporal Pattern Modeling

1 INTRODUCTION

The recommendation system has achieved great success due to its ability to capture dynamic user interests across various online platforms, such as E-commerce [26], social media [33, 36], music recommendation [16, 24], etc. The fundamental objective of a recommendation system is to predict the user's next item by estimating the similarity between the target item and the user's previously interacted items.

Recently, rather than just modeling user behavior, the repeat-aware recommendation task has been proposed to explore repeat consumption, a phenomenon where users repeatedly purchase or click on certain items. Modeling users' repeat behavior can improve the accuracy of next-item prediction. To this end, some researchers attempt to model users' temporal behavior using distribution models, such as the Negative Binomial Distribution model [5] and Point Process Model [31]. However, these methods only focus on modeling user repeat behavior and neglect user dynamic preferences for new items, thus limiting the applicability and generalizability to scenarios with relatively low repeat ratios. Conversely, other methods, as mentioned in [1, 2, 25], aim to improve upon these limitations by incorporating an additive neural component to predict the intention of user repeat consumption, or by introducing a supervised loss, with a ground truth repeat label, to regulate model training. These methods are characterized by the absence of modeling intentions behind repeat consumption, which impedes their effectiveness in striking a balance between recommending new and repeat items.

In this paper, we investigate various inherent characteristics to enhance the repeat-aware recommendation. Specifically, we explore these characteristics from two aspects. One is from the temporal aspect: we consider the time interval relationship in the user behavior sequence; the other is from the sequential aspect: we consider the sequential-level relationship in the user behavior sequence. And our intuition from these two aspects is as follows:

- **Intuition 1:** for repeatedly consumed products, users typically exhibit a fixed consumption cycle for these products. For example,

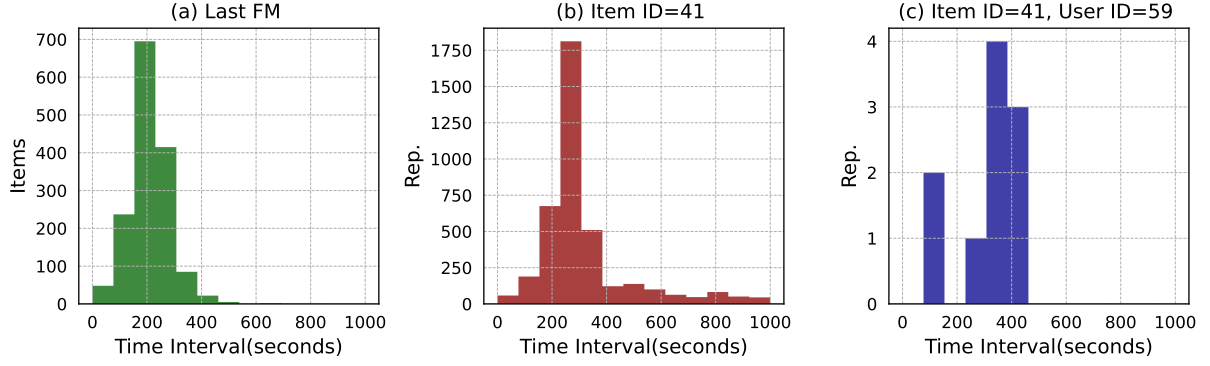


Figure 1: Data analyses to illustrate the inherent characteristics related to repeat consumption from the temporal aspect on the LastFM dataset. The x-axis is the repeat time interval of an item and the y-axis is the repeat frequency for different time intervals.

Table 1: The calculated Jaccard similarity on three different datasets. It is derived by calculating the union and intersection between the user’s current interacted item sequence sets and the former repeat interacted item sequence sets.

Dataset	RetailRocket	LastFM	Diginetica
Jacc.(%)	49.69	58.10	45.23

after users consume milk, they are likely to repurchase it during a certain period, based on their practical demand.

- **Intuition 2:** we may infer the likelihood of users making repeat consumption based on the inherent correlation within a subset of user behavior sequences. For example, when people listen to music, they tend to repeat a series of songs rather than a single track.

The above intuitions motivate us to conduct preliminary data analysis on public datasets. After preliminary data analysis, we identified patterns in the data that align with the intuitions mentioned above from two aspects: for intuition 1, we calculated the distribution of item repeat frequency with fixed repeat time intervals on the LastFM dataset in Figure 1(a) and the proportion of these items is 13.15%. The *temporal pattern* is that repeat consumption typically exhibits repeat behavior within a defined time interval; for intuition 2, we calculated the Jaccard similarity of item sets between the user behavior sequence preceding the item and the last repeat behavior sequence of the target item across three datasets and found a relatively high degree of item overlap (45%-60%), as shown in Table 1. The *sequential pattern* is that repeat consumption typically manifests as a high correlation across the entire sequence.

In this paper, we propose a novel model called **Temporal and Sequential repeat-aware Recommendation (TSRec for short)** to automatically model repeat consumption from temporal and sequential aspects. The variants of TSRec include User-specific Temporal Representation Module (UTRM), Item-specific Temporal Representation Module (ITRM), and Sequential Repeat-Aware Module (SRAM). Specifically, to capture the *temporal pattern*, we design a UTRM based on the target attention layer to encode and extract

user historical repeat temporal information. But the above module suffers from data sparsity issues, because changes in the user’s preferences may lead to a low frequency of repeat consumption of specific items. As depicted in Figure 1(b), even users exhibiting high-frequency repeat consumption behavior (e.g., User ID=59) have a relatively low overall frequency of repeat consumption for a target item (Item ID=41), with a frequency of nearly 10. Therefore we propose UTRM, which incorporates item time interval information from other users as side information. Our preliminary experiment in Figure 1 has shown that the item repeat time intervals for different users were relatively uniform. Based on this finding, we define the Repeat Time Interval Matrix based on the time intervals of the target item and design an external ITRM to encode the item-specific temporal information. To represent the user multi-scale temporal information, we refer to the idea of "Network In Network" (NIN) [21] and design ITRM, comprising CNN and pooling layers, to extract local and global information. To capture the *sequential pattern*, we design the SRAM, which consists of the Sequence Encoder and Sequence Matching Module, to represent the similarity between the user’s current sequence and the former sequence. The Sequence Encoder will encode these two sequences through a hierarchical Bi-GRU [9] and one-dimension convolution[17] layer. Then following the "matching-aggregation" [6, 7] framework widely used in natural language processing (NLP) tasks, we feed the encoded sentence representations into the Sequence Matching Module, which consists of several complex alignment operations, to enhance the model’s ability to extract and fuse mutual relation between two sentences. Finally, we evaluate the effectiveness of the proposed TSRec on three public datasets with different characteristics. To provide benchmarks for comparison, we consider several well-known sequential and repeat-aware recommendation methods. The empirical results show that TSRec can significantly outperform all the benchmarks in terms of all the evaluation metrics.

Our major contributions can be highlighted as follows:

- A novel model TSRec is proposed to enhance repeat-aware recommendation. TSRec is the first to incorporate inherent patterns of repeat consumption from temporal and sequential aspects into the design of the repeat-aware recommendation.

Table 2: Some notations and descriptions used in the paper.

Notation	Description
$\mathcal{U} = \{u\}$	the whole user set
$\mathcal{I} = \{i\}$	the whole item set
$\mathcal{S}^u = \{i^u\}$	the interacted item sequence of user u
$\mathcal{T}^u = \{t^u\}$	the timestamp sequence of user u corresponding \mathcal{S}^u
$\mathcal{P}^{u,i} = \{p^{u,i}\}$	the time interval sequence of user u for item i
$d \in \mathbb{N}$	the dimensionality of the latent representations
$L \in \mathbb{N}$	the length of truncated sequence
$\mathbf{e}^u \in \mathbb{R}^d$	the embedded user vector
$\mathbf{e}^v \in \mathbb{R}^d$	the embedded target item vector
$\mathbf{e}^t \in \mathbb{R}^d$	the embedded target time interval vector
$\mathbf{M}^i \in \mathbb{N}^{m \times n}$	the Repeat Time Interval Matrix of item i
$\mathbf{E}^h \in \mathbb{R}^{L \times d}$	the embedded historical time interval matrix
$\mathbf{E}^b \in \mathbb{R}^{L \times d}$	the embedded user behavior matrix
$\mathbf{E}^l \in \mathbb{R}^{L \times d}$	the embedded former repeat behavior matrix

- To extract the *temporal pattern* and *sequential pattern*, we have devised two innovative sub-modules: UTRM and SRAM, respectively. Additionally, we have designed an ITRM sub-module to address the data sparsity issue in one user's repeat consumption sequence.
- We conducted extensive experiments on three publicly available datasets to validate the superiority of our proposed method over various state-of-the-art recommendation methods. Furthermore, ablation and case studies provide additional evidence of the benefits of our model.

2 PRELIMINARIES

If we denote the user and item sets as \mathcal{U} and \mathcal{I} , respectively, their corresponding sizes can be represented by $|\mathcal{U}|$ and $|\mathcal{I}|$. In sequential recommendation, the interaction sequence of user $u \in \mathcal{U}$ is organized chronologically as $\mathcal{S}^u = \{i_1^u, i_2^u, \dots, i_{|\mathcal{S}^u|}^u\}$ and time sequence is $\mathcal{T}^u = \{t_1^u, t_2^u, \dots, t_{|\mathcal{S}^u|}^u\}$, which means the user u interacts with the item i^u at time t^u .

Task of repeat-aware recommendation. Given $\mathcal{U}, \mathcal{I}, \mathcal{S}$ and \mathcal{T} , the task of repeat-aware recommendation is to predict the score that user u clicks on the next (either new or repeat) item i at time $t_{|\mathcal{S}^u|+1}^u$:

$$\hat{y}_{t_{|\mathcal{S}^u|+1}^u}^u = p(i_{|\mathcal{S}^u|+1}^u | \mathcal{S}^u, \mathcal{T}^u) \quad (1)$$

Definition 2.1. Repeat Consumption. Given $\mathcal{U}, \mathcal{I}, \mathcal{S}$ and \mathcal{T} , repeat consumption exists in user u if and only if user u 's interacted item $i_{|\mathcal{S}^u|+1}^u = i^u \in \mathcal{S}^u$.

3 METHODOLOGY

In this section, we propose the TSRec in Figure 2 to model repeat consumption in the sequential recommendation. TSRec consists of three sub-modules: User-specific Temporal Representation Module (UTRM), Item-specific Temporal Representation Module (ITRM), and Sequential Repeat-Aware Module (SRAM). And representation vectors from these three sub-modules will be concatenated with embedded user vectors as the user's complete preference for the

target item. We list some key notations and their descriptions in Table 2.

3.1 Feature Representation

3.1.1 Sequence Embedding. We first transform the u 's interacted item sequence $\mathcal{S}^u = \{i_1^u, i_2^u, \dots, i_{|\mathcal{S}^u|}^u\}$ into a fixed-length sequence $\mathbf{s} = \{i_{|\mathcal{S}^u|-L+1}^u, i_{|\mathcal{S}^u|-L+2}^u, \dots, i_{|\mathcal{S}^u|}^u\}$, where L is the predefined sequence length. If the length of \mathcal{S}^u exceeds L , we keep only the most recent L behaviors. Conversely, if the sequence length is less than L , we insert padding items to the left until the desired length L is reached. For each item $i \in \mathcal{I}$, we first transform id categories into sparse one-hot vectors. Then we transform these high-dimensional representations into dense vectors as the embedded vectors.

For $\mathcal{S}^u = \{i_1^u, i_2^u, \dots, i_{j-1}^u, i_j^u, \dots, i_{k-1}^u, i_k^u, \dots, i_{|\mathcal{S}^u|}^u\}$, if $i_j^u = i_k^u$, we define $\mathcal{S}^{u,b} = \{i_{j+1}^u, i_{j+2}^u, \dots, i_{k-1}^u\}$ as the user behavior sequence preceding the target item i_k^u , and $\mathcal{S}^{u,l} = \{i_1^u, i_2^u, \dots, i_{j-1}^u\}$ as the last repeat behavior sequence of the target item i_k^u . As previous operation, $\mathcal{S}^{u,b}, \mathcal{S}^{u,l}$ will be cut off and padded into fixed length L as $\mathbf{v}^b, \mathbf{v}^l$. And these two sequence vectors will be transformed into two embedded matrices by utilizing the embedding layer, named $\mathbf{E}^b, \mathbf{E}^l \in \mathbb{R}^{L \times d}$.

3.1.2 Time Interval Embedding. To represent the temporal information of repeat consumption, we need to pre-process the timestamp sequence \mathcal{T}^u and transform it into the learnable matrix. We first define the Repeat Time Interval Matrix as follows:

Definition 3.1. Repeat Time Interval Matrix. For user u , the timestamp sequence of the interacted item i can be denoted as $\mathcal{T}^{u,i} = \{t_1^{u,i}, t_2^{u,i}, \dots, t_n^{u,i}, t_{n+1}^{u,i}\}$. The k -th time interval of item i is $p_k^{u,i} = \left\lfloor \frac{t_{k+1}^{u,i} - t_k^{u,i}}{p_{\min}} \right\rfloor$ ($k = 1, 2, \dots, n$), where p_{\min} is the minimal time interval and $p_k^{u,i}$ is a discrete value. The time interval sequence of user u for item i is $\mathcal{P}^{u,i} = \{p_1^{u,i}, p_2^{u,i}, \dots, p_n^{u,i}\}$. The Repeat Time Interval Matrix of item i can be defined as $\mathbf{M}^i \in \mathbb{N}^{m \times n}$:

$$\mathbf{M}^i = \begin{bmatrix} \mathcal{P}^{u_1,i} \\ \mathcal{P}^{u_2,i} \\ \dots \\ \mathcal{P}^{u_m,i} \end{bmatrix} = \begin{bmatrix} p_1^{u_1,i} & p_2^{u_1,i} & \dots & p_n^{u_1,i} \\ p_1^{u_2,i} & p_2^{u_2,i} & \dots & p_n^{u_2,i} \\ \vdots & \vdots & \ddots & \vdots \\ p_1^{u_m,i} & p_2^{u_m,i} & \dots & p_n^{u_m,i} \end{bmatrix}. \quad (2)$$

Like sequence embedding, we can convert discrete time interval values into one-hot vectors and then transform them into dense embedded vectors by the embedding layer. Specifically, the target time interval (time interval between current timestamp and last behavior's timestamp) $p^{u,i}$, user historical time interval matrix $\mathcal{P}^{u,i}$, and Repeat Time Interval Matrix \mathbf{M}^i will be embedded by using the embedding layer named $\mathbf{e}^t \in \mathbb{R}^d$, $\mathbf{E}^h \in \mathbb{R}^{n \times d}$ and $\mathbf{E}^P \in \mathbb{R}^{m \times n \times d}$.

3.2 User-specific Temporal Representation Module

The proposed module aims to extract the *temporal pattern* from user historical repeat consumption behavior primarily through the target attention layer. To achieve this, we first obtain the target time interval embedding vector \mathbf{e}^t and historical time interval embedding matrix $\mathbf{E}^h = \{\mathbf{e}_1^h, \dots, \mathbf{e}_k^h, \dots, \mathbf{e}_n^h\}$ through the embedding

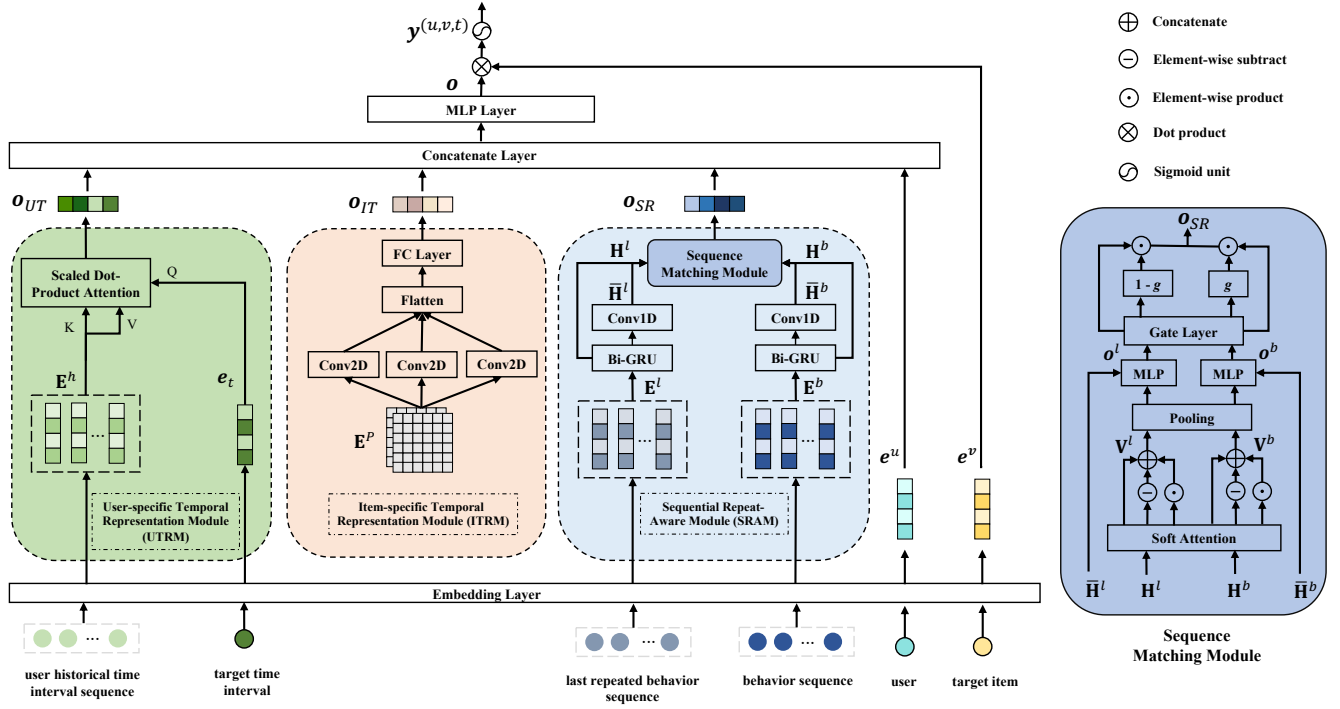


Figure 2: The overview architecture of TSRec, which consists of three parts: User-specific Temporal Representation Module (UTRM), Item-specific Temporal Representation Module (ITRM), and Sequential Repeat-Aware Module (SRAM). The UTRM captures *temporal pattern* by modeling the relation of repeat time intervals in user sequence. The ITRM captures the *temporal pattern* by extracting the temporal information in Repeat Time Interval Matrix. The SRAM first encodes the user’s current behavior sequence and former behavior sequence and then aligns these two encoded vectors. All these representations will be integrated into the prediction layer for the next-item prediction.

layer. We can figure out if the target item is the clicked item of the user, the target time interval e^t will be the next element of E^h . We explicitly incorporate temporal information into the attention procedure, utilizing the widely used attention methodology - *Scaled Dot-Product Attention* [30]. The *Scaled Dot-Product Attention* computes weights of the sequence representations of user historical time intervals, after converting them by three learnable weight matrices, $W^Q, W^K, W^V \in \mathbb{R}^{d \times d}$, as:

$$\text{sim}(e^t, e_k^h) = \frac{(W^Q e^t) \cdot (W^K e_k^h)^T}{\sqrt{d}}, \quad (3)$$

$$\text{Attention}(e^t, e_k^h) = \frac{\exp(\text{sim}(e^t, e_k^h))}{\sum_{k'=1}^n \exp(\text{sim}(e^t, e_{k'}^h))}, \quad (4)$$

$$o_{UT} = \sum_{k=1}^n \text{Attention}(e^t, e_k^h) \cdot (W^V e_k^h), \quad (5)$$

where $o_{UT} \in \mathbb{R}^d$ denotes the weighted sum pooling of the historical time interval embedding matrix. By performing the aforementioned operations, we can capture the *temporal pattern* in the user’s historical sequence and their personalized repeat consumption intention for different items.

3.3 Item-specific Temporal Representation Module

The user historical repeat consumption sequence for item i is inherently sparse, necessitating the incorporation of other user time intervals as side information. Therefore, we define the Repeat Time Interval Matrix for target item i in definition 3.1. However, capturing the complex temporal repeat consumption information in user behavior is challenging. To address this, we draw inspiration from the "Network In Network" (NIN) [21, 35] concept and propose a multi-scale feature extraction approach. We design $C = c_1, c_2, \dots, c_k, \dots, c_n$ two-dimensional convolutional filters with different kernel sizes to extract *temporal patterns* of different scales and aggregate these representations using both mean-pooling and max-pooling operations:

$$m_{\text{mean}}^{c_k} = \text{Mean_Pooling}(\text{Conv2D}^{c_k}(E^P)), \quad (6)$$

$$m_{\text{max}}^{c_k} = \text{Max_Pooling}(\text{Conv2D}^{c_k}(E^P)). \quad (7)$$

The pooling representation vectors of the c -th two-dimensional convolutional filter $m^{c_k} = [m_{\text{mean}}^{c_k}; m_{\text{max}}^{c_k}]$. We aggregate these representation vectors of different scales as follows:

$$o_{IT} = \text{MLP}([m^{c_1}; \dots; m^{c_k}; \dots; m^{c_n}]), \quad (8)$$

where $\mathbf{o}_{IT} \in \mathbb{R}^d$.

3.4 Sequential Repeat-Aware Module

3.4.1 Sequence Encoder. After the embedding layer, we obtain the user behavior matrix $\mathbf{E}^b = \{\mathbf{e}_1^b, \mathbf{e}_2^b, \dots, \mathbf{e}_L^b\}$ and the last repeat behavior matrix $\mathbf{E}^l = \{\mathbf{e}_1^l, \mathbf{e}_2^l, \dots, \mathbf{e}_L^l\}$, which exhibit strong temporal relations in Table 1. To encode these two embedded matrices, a simple pooling operation is intuitive but this encoding operation may lose the local temporal information and can not reflect user dynamic preferences. In this paper, we adopt the stacked encoded layers: Bidirectional Gated Recurrent Unit (Bi-GRU) [9] and one-dimension convolution [17], which extract different features from multi-dimensional perspectives. Bi-GRU focuses on sequence and context features, while also being able to memorize long-term dependencies through its use of gating mechanisms. Also, Bi-GRU can processes input data from both forward and backward directions simultaneously, enabling better understanding and representation of the input data. In contrast, one-dimensional convolution is better at extracting local patterns across the entire sequence, regardless of position, through its use of sliding windows. The formulations of the forward GRU encoder for $\mathbf{E}^l = \{\mathbf{e}_1^l, \mathbf{e}_2^l, \dots, \mathbf{e}_j^l, \dots, \mathbf{e}_L^l\}$ are listed as follows:

$$\mathbf{r}_j = \sigma(\mathbf{W}_r \mathbf{e}_j^l + \mathbf{U}_r \vec{\mathbf{h}}_{j-1}^l), \quad (9)$$

$$\mathbf{z}_j = \sigma(\mathbf{W}_z \mathbf{e}_j^l + \mathbf{U}_z \vec{\mathbf{h}}_{j-1}^l), \quad (10)$$

$$\tilde{\mathbf{h}}_j = \tanh(\mathbf{W}_h \mathbf{e}_j^l + \mathbf{U}_h (\mathbf{r}_j \odot \vec{\mathbf{h}}_{j-1}^l)), \quad (11)$$

$$\vec{\mathbf{h}}_j^l = \mathbf{z}_j \odot \vec{\mathbf{h}}_{j-1}^l + (1 - \mathbf{z}_j) \odot \tilde{\mathbf{h}}_j, \quad (12)$$

where σ denotes the sigmoid function, $\mathbf{W}_r, \mathbf{W}_z, \mathbf{W}_h, \mathbf{U}_r, \mathbf{U}_z, \mathbf{U}_h \in \mathbb{R}^{d \times d}$ are learnable weight matrices, \mathbf{r}_j is the reset gate, \mathbf{z}_j is the update gate and $\tilde{\mathbf{h}}_j$ is the candidate activation. The j -th hidden state of the Bi-GRU layer is obtained by concatenating the hidden state in the forward GRU and the backward GRU as $\mathbf{h}_j^l = [\vec{\mathbf{h}}_j^l; \overleftarrow{\mathbf{h}}_j^l]$ and the encoded representation matrix is $\mathbf{H}^l = \{\mathbf{h}_1^l, \mathbf{h}_2^l, \dots, \mathbf{h}_j^l, \dots, \mathbf{h}_L^l\}$. Like the encoding operation of the Bi-GRU layer for \mathbf{E}^l , we can obtain the Bi-GRU encoded representation matrix for \mathbf{E}^b , named \mathbf{H}^b . To extract local information of $\mathbf{H}^l, \mathbf{H}^b \in \mathbb{R}^{d \times L}$, we use one-dimensional convolutional encoders, and these convolutional encoders will use the sliding windows of different lengths on $\mathbf{H}^b, \mathbf{H}^l$:

$$\tilde{\mathbf{H}}^b = \text{Conv1D}(\mathbf{H}^b) \quad (13)$$

$$\tilde{\mathbf{H}}^l = \text{Conv1D}(\mathbf{H}^l) \quad (14)$$

3.4.2 Sequence Matching Module. The Bi-GRU encoded representation matrices $\mathbf{H}^l, \mathbf{H}^b \in \mathbb{R}^{d \times L}$ need to be aligned and extracted the relation between the above two parts. In this paper, we adopt the soft attention alignment [3, 22], with the alignment matrix computed as follows:

$$\mathbf{A} = (\mathbf{H}^l)^\top \cdot \mathbf{H}^b, \quad (15)$$

where each element in $\mathbf{A} = \{a_{jk}\} \in \mathbb{R}^{L \times L}$ represents the alignment score of the element between \mathbf{H}^l and \mathbf{H}^b . To get the alignment representation, the encoded matrix will be transformed by the

weighted sum of the alignment matrix:

$$\tilde{\mathbf{h}}_k^b = \sum_{j=1}^L \frac{\exp(a_{jk})}{\sum_{j'=1}^L \exp(a_{j'k})} \mathbf{h}_j^b, \quad (16)$$

$$\tilde{\mathbf{h}}_j^l = \sum_{k=1}^L \frac{\exp(a_{jk})}{\sum_{k'=1}^L \exp(a_{jk'})} \mathbf{h}_k^l, \quad (17)$$

where the aligned matrices are $\tilde{\mathbf{H}}^l, \tilde{\mathbf{H}}^b \in \mathbb{R}^{d \times L}$. Intuitively, $\tilde{\mathbf{H}}^b$ is a weighted sum of \mathbf{H}^l with the weights being determined by the similarity score between \mathbf{H}^b and \mathbf{H}^l . To further enhance the alignment information, we employ the element-wise subtract and product operation between $\tilde{\mathbf{H}}^b$ and \mathbf{H}^b , as well as $\tilde{\mathbf{H}}^l$ and \mathbf{H}^l . We expect that such operations could help sharpen local information between elements in the sentence representations and capture relationships such as contradiction. The representations of element-wise subtract and product operations will be concatenated with encoded and aligned representations as follows:

$$\mathbf{V}^b = \text{MLP}([\mathbf{H}^b; \tilde{\mathbf{H}}^b; \mathbf{H}^b - \tilde{\mathbf{H}}^b; \mathbf{H}^b \odot \tilde{\mathbf{H}}^b]), \quad (18)$$

$$\mathbf{V}^l = \text{MLP}([\mathbf{H}^l; \tilde{\mathbf{H}}^l; \mathbf{H}^l - \tilde{\mathbf{H}}^l; \mathbf{H}^l \odot \tilde{\mathbf{H}}^l]), \quad (19)$$

where $\mathbf{V}^l \in \mathbb{R}^{d \times L}, \mathbf{V}^b \in \mathbb{R}^{d \times L}$, and MLP is a Multi-Layer Perception [27]. The above operations aim to transform the representations from different perspectives. These matrix representations will be projected into vectors by mean-pooling and max-pooling operations:

$$\mathbf{v}_{\text{mean}}^b = \sum_{j=1}^L \frac{\mathbf{V}^b}{L}, \mathbf{v}_{\text{max}}^b = \max_{j=1}^L \mathbf{V}^b, \quad (20)$$

$$\mathbf{v}_{\text{mean}}^l = \sum_{j=1}^L \frac{\mathbf{V}^l}{L}, \mathbf{v}_{\text{max}}^l = \max_{j=1}^L \mathbf{V}^l, \quad (21)$$

where $\mathbf{v}_{\text{mean}}^l, \mathbf{v}_{\text{max}}^l, \mathbf{v}_{\text{mean}}^b, \mathbf{v}_{\text{max}}^b \in \mathbb{R}^d$. We flatten the encoded matrices $\tilde{\mathbf{H}}^l, \tilde{\mathbf{H}}^b$ as $\tilde{\mathbf{h}}^l, \tilde{\mathbf{h}}^b \in \mathbb{R}^{d \times L}$ and concatenate them with $\mathbf{v}_{\text{mean}}^l, \mathbf{v}_{\text{max}}^l, \mathbf{v}_{\text{mean}}^b, \mathbf{v}_{\text{max}}^b$ as two vectors:

$$\mathbf{o}^b = \text{MLP}([\mathbf{v}_{\text{mean}}^b; \tilde{\mathbf{h}}^b; \mathbf{v}_{\text{max}}^b]), \quad (22)$$

$$\mathbf{o}^l = \text{MLP}([\mathbf{v}_{\text{mean}}^l; \tilde{\mathbf{h}}^l; \mathbf{v}_{\text{max}}^l]), \quad (23)$$

To fuse \mathbf{o}^l and \mathbf{o}^b , we refer to the gated-based fusion operation [32] and obtain the fused vectors:

$$\mathbf{g} = \sigma(\mathbf{W}_1 \mathbf{o}^l + \mathbf{W}_2 \mathbf{o}^b), \quad (24)$$

$$\mathbf{o}_{SR} = \mathbf{o}^l \odot \mathbf{g} + \mathbf{o}^b \odot (1 - \mathbf{g}), \quad (25)$$

where $\mathbf{W}_1 \in \mathbb{R}^{d \times d}, \mathbf{W}_2 \in \mathbb{R}^{d \times d}$ are learnable weight matrices of the gate layer and σ is the sigmoid activation function. $\mathbf{g} \in \mathbb{R}^d$ is a vector rather than a scalar, which enables the neural gate to regulate the contributions of \mathbf{o}^l and \mathbf{o}^b at the dimension granularity.

3.5 Network Training

After different modules, we get different representation vectors $\mathbf{o}_{UT}, \mathbf{o}_{IT}, \mathbf{o}_{SR}$. These vectors, together with user embedding vector \mathbf{e}^u , will be combined as:

$$\mathbf{o} = \text{MLP}([\mathbf{o}_{UT}; \mathbf{o}_{IT}; \mathbf{o}_{SR}; \mathbf{e}^u]), \quad (26)$$

where $\mathbf{o} \in \mathbb{R}^d$ denotes the fused representation. Finally, we adopt a dot-product of \mathbf{o} with the target item \mathbf{e}^v as a normalized prediction score of user u rating for target item v at the timestamp t :

$$\mathbf{y}^{(u,v,t)} = \sigma(\mathbf{o}^T \cdot \mathbf{e}^v). \quad (27)$$

In this paper, since the length of user sequences may vary, we adopt a sliding window strategy [34]. This approach involves dividing each sequence into windows of size L if the sequence length exceeds L , or padding the sequence if the length is less than L . If $\tilde{\mathcal{T}}^u = \{L+1, L+2, \dots, |\mathcal{S}_t^u|\}$ is the timestamp set we need to predict, our optimized objective loss for this prediction task can be defined with Cross-Entropy loss as follows:

$$\mathcal{L} = \sum_u \sum_{t \in \tilde{\mathcal{T}}^u} -\log(\mathbf{y}^{(u,v^+,t)}) + \sum_u \sum_{t \in \tilde{\mathcal{T}}^u} \sum_{v^-} -(1 - \log(\mathbf{y}^{(u,v^-,t)})). \quad (28)$$

Here, v^+ represents the positive sample where user u interacted at timestamp t , and v^- represents the randomly sampled negative sample that does not exist in the positive sample set.

4 EXPERIMENTS

In this section, our objective is to address the following research questions (RQs):

- **RQ1:** How does TSRec perform compared to various state-of-the-art recommendation methods with different settings?
- **RQ2:** How effective are the key modules (e.g., User-specific Temporal Representation Module, Item-specific Temporal Representation Module) in TSRec?
- **RQ3:** How do different hyper-parameters, such as embedding layer dimension d and user sequence length L , affect the TSRec's performance?
- **RQ4:** How is the capture capability of TSRec considering the temporal information modeling?

4.1 Experimental Settings

4.1.1 Datasets. We conduct our experiments on the following datasets:

- **RetailRocket**² This dataset is generated from an online shopping site-Retailrocket. It is a collection of user browsing activities on an online shopping site spanning six months.
- **LastFM**³ The LastFM dataset is a rich collection of music listening histories generated by users of the LastFM platform. We employ the 1K version in our experiments.
- **Diginetcia**⁴ The Diginetcia dataset, which was released by CIKM Cup 2016, is a collection of users' clicked stream data captured in an e-commerce website. We only used the transaction data that was released.

Following previous work [13, 15, 28], we only consider implicit feedback (click/order) and we treat explicit feedback (ratings) as implicit feedback in the datasets. We group the interaction records for each user and build the interaction sequence for each user by sorting these interaction records according to the timestamps. For each dataset, we divide the data by time into a training set (70%),

validation set (10%), and testing set (20%). To address the cold-start issue, we filter out items that have been interacted with fewer than 10 times and only retain users whose behavior sequences consist of more than one interaction. The detailed statistics of experimental datasets are shown in Table 4.

4.1.2 Evaluation Metrics. Following [15, 28], we use the *Top-K Hit Rate* (HR@K) and *Top-K Normalized Discounted Cumulative Gain* (NDCG@K), which are commonly used in related research [15, 28]. Our reported results are based on HR@{5, 10, 20} and NDCG@{5, 10, 20}. We calculate all metrics based on item ranking and report the average scores. We conduct a paired t-test, with statistical significance being indicated when the p-value is below 0.01.

4.1.3 Baselines. To evaluate the superiority and effectiveness of TSRec, we benchmark it with two types of baselines:

General Sequential Recommendation Methods.

- **PopRec.** It recommends items to users based on their ranked frequency interacted by users.
- **GRU4Rec** [13]. This method utilizes the gated recurrent unit as a sequence encoder to learn user dynamic preference with a ranking-based loss.
- **NARM** [19]. It integrates the RNN layers with the attention mechanism to capture the local and global preferences of users.
- **Caser** [34]. This sequential recommendation method integrates the convolutional neural layers from both vertical and horizontal views to encode user time-varied behavior sequences.
- **SASRec** [15]. This method utilizes self-attention mechanisms to automatically learn long-term dependencies in the sequence. It also uses techniques such as negative sampling and masked loss to improve the training efficiency and recommendation performance of the model.
- **BERT4Rec** [28]. It is a recommendation model based on the Bidirectional Encoder Representations from Transformers (BERT) architecture. The model is optimized with the Cloze objective and has produced state-of-the-art performance among sequential learning-based baselines.

Repeat-aware Recommendation Methods.

- **RepeatNet** [25]. RepeatNet is an attention-based model, which aims to address the problem of repeat consumption. In addition to the standard attention mechanism used in sequence-to-sequence models, RepeatNet also employs a repeat-aware attention mechanism to model the probability of a user revisiting an item in the sequence.
- **SLRC_NCF** [31]. SLRC_NCF is proposed with a combination of two techniques: Collaborative Filtering [18] and Hawkes Process [11]. The model adaptively incorporates two item-specific temporal dynamics of repeat consumption into the kernel function of the Hawkes Process.
- **ReCANet** [2]. ReCANet is a novel neural architecture designed for Next Basket Recommendation (NBR) in grocery shopping, with a specific focus on repeat items.
- **NovelNet** [20]. NovelNet is a recommendation model designed specifically for online novels, which encodes user interactions by taking into account fine-grained interaction attributes. It utilizes

²<https://www.kaggle.com/retailrocket/ecommerce-dataset>

³<https://www.lastfm.com>

⁴<http://cikum2016.cs.iupui.edu/cikum-cup>

Table 3: Performance comparison of all baseline methods. The best results of all methods are indicated in bold, while the second best results are indicated in underlined. Improvements over baselines are statistically significant with $p < 0.01$.

Datasets	Metric	PopRec	GRU4Rec	NARM	Caser	SASRec	BERT4Rec	RepeatNet	SLRC_NCF	ReCANet	NovelNet	TSRec	Improv.
RetailRocket	HR@5	0.3265	0.6923	0.7276	0.6545	<u>0.7798</u>	0.7729	0.5418	0.6412	0.5149	0.6934	0.8112	4.03%
	HR@10	0.4667	0.7662	0.7980	0.7368	<u>0.8307</u>	0.8244	0.6408	0.7002	0.6097	0.7679	0.8620	3.77%
	HR@20	0.6369	0.8415	0.8669	0.8198	<u>0.8849</u>	0.8808	0.7575	0.7679	0.7329	0.8435	0.9087	2.69%
	NDCG@5	0.2249	0.6057	0.6388	0.5525	<u>0.7125</u>	0.7103	0.4656	0.5867	0.4466	0.5977	0.7240	0.91%
	NDCG@10	0.2701	0.6296	0.6616	0.5791	<u>0.7310</u>	0.7270	0.4975	0.6058	0.4772	0.6218	0.7405	0.89%
	NDCG@20	0.3130	0.6486	0.6790	0.6001	<u>0.7427</u>	0.7412	0.5269	0.6229	0.5082	0.6410	0.7523	0.62%
LastFM	HR@5	0.4452	0.8162	0.8298	0.8460	<u>0.8543</u>	0.8484	0.7407	0.8541	0.8460	0.8110	0.8660	1.37%
	HR@10	0.5766	0.8816	0.8891	0.8982	<u>0.9045</u>	0.8940	0.8140	0.8993	0.8934	0.8750	0.9128	0.92%
	HR@20	0.7051	0.9292	0.9334	0.9237	<u>0.9418</u>	0.9328	0.8791	0.9392	0.9269	0.9246	0.9508	1.22%
	NDCG@5	0.3216	0.7223	0.7451	0.7397	<u>0.7754</u>	0.7775	0.6675	0.7776	<u>0.7808</u>	0.7246	0.7903	1.54%
	NDCG@10	0.3641	0.7435	0.7644	0.7568	0.7918	0.7923	0.6913	0.7923	<u>0.7935</u>	0.7554	0.8056	1.52%
	NDCG@20	0.3966	0.7557	0.7757	0.7655	0.8013	0.8021	0.7078	0.8025	<u>0.8026</u>	0.7580	0.8152	1.57%
Diginetica	HR@5	0.2142	0.7460	<u>0.8899</u>	0.8315	0.8810	0.8410	0.6392	0.6890	0.2703	0.8840	0.9386	5.47%
	HR@10	0.3268	0.8348	<u>0.9300</u>	0.8837	0.9165	0.8907	0.7519	0.7462	0.3743	0.9269	0.9575	2.96%
	HR@20	0.4764	0.9004	<u>0.9573</u>	0.9243	0.9435	0.9284	0.8123	0.8055	0.5151	0.9536	0.9699	1.32%
	NDCG@5	0.1418	0.6070	0.7609	0.7391	<u>0.7848</u>	0.7373	0.4068	0.6172	0.2044	0.7471	0.8054	2.62%
	NDCG@10	0.1780	0.6359	0.7740	0.7561	<u>0.7964</u>	0.7535	0.4441	0.6358	0.2378	0.7612	0.8116	1.91%
	NDCG@20	0.2156	0.6525	0.7810	0.7664	<u>0.8032</u>	0.7630	0.4594	0.6508	0.2732	0.7679	0.8147	1.43%

Table 4: Statistics of our experimented datasets. Inter. means the number of interactions; Avg. means the average number of interactions per user; Rep.(%) calculates the ratio of repeated user-item interactions.

Dataset	Users	Items	Inter.	Avg.	Rep.(%)
RetailRocket	217,508	51,586	910,947	4.00	35.03
LastFM	728	11,543	1,108,999	15.41	89.29
Diginetica	187,938	24,700	861,859	4.59	16.53

a pointer network with pointwise and listwise loss functions to model user repeat consumption behavior.

4.1.4 Hyper-parameter Settings. Hyper-parameters are usually established by default baselines. For sequential recommendation methods, we adopt the preceding research of [15] and apply Adam with a learning rate of 0.001 to balance gradients. The batch size is 256, and the dropout rate is set at 0.5. We utilize the activation functions of sigmoid and ReLU. The number of two-dimensional convolutional layers $C = 3$, with kernel sizes of 1×1 , 1×3 , and 3×1 . The grid search method is applied to find the optimal settings of hyper-parameters using the validation set. These include the latent representation dimension d from $\{10, 20, 50, 100, 200\}$, and the sequence length L is from $\{5, 10, 15, 20\}$. To construct the Repeat Time Interval Matrix, we sort each user by item repeat frequencies and choose the Top-M ($M=10$) users. We tune hyper-parameters using the validation set and terminate training if validation performance doesn't improve for 20 epochs. Following the common strategy [15], we pair the ground-truth item with 3 randomly sampled negative items that the user has not interacted with in the validation set and 100 in the test set.

4.2 Overall Performance Comparison (RQ1)

4.2.1 Comparison with Baselines. We report the detailed performance comparison with 10 competitive baselines on different datasets in Table 3 and summarize the observations as follows:

- **TSRec has been empirically validated to achieve state-of-the-art performance.** The results on three datasets indicate that our TSRec method consistently outperforms all types of baselines by a significant margin. Notably, our methodology leads to an improvement of up to 4.03% in HR@5 for the RetailRocket dataset and up to 5.47% in HR@5 for the Diginetica dataset. Our superior performance can be attributed to the following factors: 1) through SRAM, TSRec is able to capture the *sequential pattern* from coarse-grained to fine-grained granularities; 2) with the target-attention-based User-specific Temporal Representation Module, we endow TSRec with the capability of capturing long-range item repeat correlations over time. 3) The Repeat Time Interval Matrix can have multi-perspective time-interval information and the corresponding UTRM is effective to extract the above information.
- **Our method demonstrates robustness across various recommendation scenarios.** Our methodology exhibits significant improvement despite the diverse dataset characteristics, including differences in average sequence length and repeat ratio. Also, the items and users in these three datasets are totally different and this difference will lead to preference characteristics (different distributions of repeat time intervals and different lengths of repeat sequences). This improvement compared with SOTA benchmarks indicates that our method possesses strong generalization capabilities and is robust in practical scenarios.
- **Existing repeat-aware recommendation methods exhibit inferior performance compared to general sequential recommendation techniques.** In the RetailRocket and Diginetica datasets, the evaluated metrics of repeat-aware recommendation methods (e.g., RepeatNet, ReCANet) are significantly lower than those of sequential recommendation methods (e.g., SASRec,

BERT4Rec). This outcome may be attributed to the relatively lower repeat ratio (35.03% and 16.53%) in these datasets, which causes repeat-aware recommendation methods to overly prioritize repeat items and neglect the sequential relation for new items. In contrast, in the LastFM dataset where the repeat ratio is 89.29%, the repeat-aware recommendation method can achieve competitive results. In addition, the incorporation of repeat-explore hybrid prediction parts will reduce the representation ability of user historical preference.

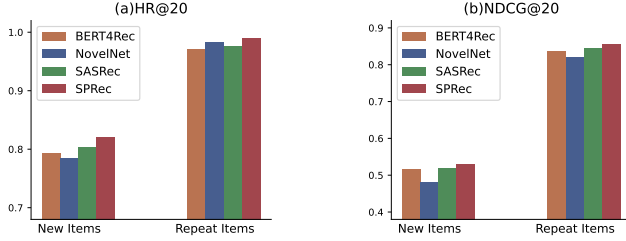


Figure 3: The performance of TSRec is evaluated on both new and repeat items, and compared against various benchmarks on the RetailRocket dataset using HR@20 and NDCG@20 metrics.

4.2.2 Analysis of Recommending New and Repeat Items. To further compare our method with SOTA benchmarks, we split the prediction accuracy according to the repeat frequency of items - new items and repeat items. In Figure 3, we present the comparative results with various analogous benchmarks (BERT4Rec, NovelNet, SASRec). Due to spatial constraints, we have only enlisted the commensurate results on the RetailRocket dataset and summarized the following findings:

- **Our proposed method outperforms other benchmarks both in new items and repeat items.** Compared with other recommendation methods, our method performs well in new and repeat items. This finding reveals that our sub-modules can capture fine-grained repeat characteristics of users more effectively, thereby balancing predictions between exploitation and exploration.
- **The approach based on repeat consumption has high accuracy for recommending repeat items, but it leads to a decline in the accuracy rate of new items.** All models perform better for repeat items than for new items because these methods underrate new items as they do not exist in the user’s historical sequence. The repeat-aware recommendation methods cannot be compared with sequential recommendation methods for predicting new items. For example, for users who have a preference for purchasing new items, the repeat-aware recommendation methods can not fully predict these users’ demands. To summarize, these methods do not strike a balance between exploitation and exploration.

4.3 Ablation Study(RQ2)

To validate the effectiveness of different components of TSRec, we remove three components (UTRM, ITRM, and SMM) of TSRec and make a comparison with our TSRec method as shown in Table 5:

Table 5: Effectiveness validation of TSRec and its components on three datasets. SMM means Sequence Matching Module; UTRM means User-specific Temporal Representation Module; ITRM means Item-specific Temporal Representation Module. The Best performance is in bold font.

Datasets	Metric	w/o SMM	w/o UTRM	w/o ITRM	TSRec
RetailRocket	HR@20	0.9066	0.9061	0.9050	0.9087
	NDCG@20	0.7371	0.7365	0.7333	0.7523
LastFM	HR@20	0.9466	0.9455	0.9443	0.9508
	NDCG@20	0.8138	0.7913	0.8041	0.8152
Diginetica	HR@20	0.9641	0.9617	0.9598	0.9699
	NDCG@20	0.8127	0.8036	0.7909	0.8147

- **All the components in the TSRec method are useful to improve the recommendation performance.** Three of these components yield significant drops in performance, indicating their essentiality. More specifically, removing SMM results in an incomplete exploration of the *sequential pattern*, causing the model to only capture basic sequential data, leading to a decline in performance. Additionally, the exclusion of ITRM and UTRM results in a failure to leverage the *temporal pattern*, which hinders the model’s ability to predict temporal intentions of repeat consumption.
- **ITRM achieves better performance than UTRM.** Removing ITRM and UTRM will decrease the performance of TSRec, but ITRM is more significant actually. The reason for the performance gap is data sparsity issues in one user repeat sequence. The Repeat Time Interval Matrix has rich *temporal pattern* and the *temporal pattern* from other users is coherent with that of the current user.

4.4 The Impact of Hyper-parameters(RQ3)

In the subsequent study, we will examine the effects of the hyper-parameters, namely L and d , individually, while preserving the optimal configurations of the remaining hyper-parameters.

4.4.1 Impact of sequence length L . We validate the robustness of our method and choose different sequence lengths L as shown in Figure 4a. Across the three datasets, it is observed that the truncation of distinct sequences has minimal impact on the HR of the TSRec’s prediction, whereas it has a more notable effect on the final ranking position. This finding indicates that TSRec is robust in accurately identifying relevant items with different sequence lengths. Notably, employing larger sequence lengths may reduce model efficiency. Therefore, to optimize both performance and efficiency, we designate $L = 10$ as the hyperparameter for TSRec.”

4.4.2 Impact of embedding dimension d . The performance of TSRec over a range of embedding dimensions (from 10 to 200) is presented in Figure 4b. Examining the results, we observe that the HR and NDCG of our model initially increase as the embedding dimension expands. Specifically, HR reaches its maximum at the embedding size of 100, owing to the ability of a larger dimension to retain and transmit more information. However, further increasing the embedding dimension will likely lead to overfitting, diminishing the performance of TSRec. As our model achieves satisfactory performance with embedding dimensions of $d \leq 100$, we choose $d = 100$ as the optimal hyper-parameter of TSRec.

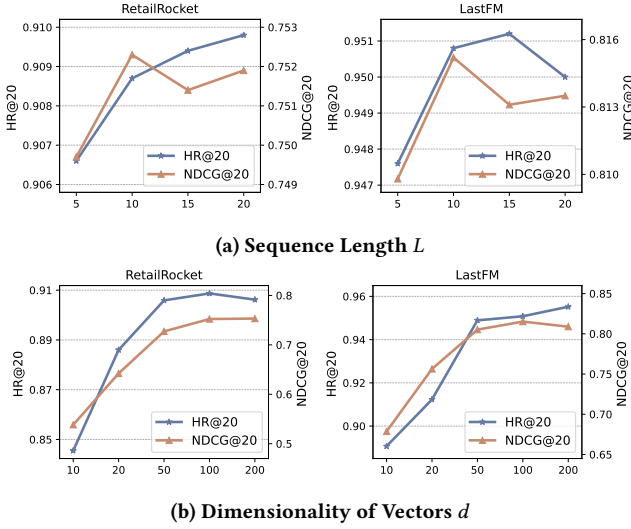


Figure 4: The impact of hyper-parameters, namely L and d on two datasets in terms of HR@20 and NDCG@20.

4.5 Case Study of Temporal Modeling(RQ4)

In order to verify that TSRec has learned the temporal information of user repeat consumption, we randomly selected a few test cases. In Figure 5, the solid dots are the time interval of the user’s last repeat consumption while the hollow dots are virtually generated target time intervals around the user’s last repeat time interval. Among these dots, we can observe that the solid dots reach the maximum $y^{(u,v,t)}$, and hollow dots have a relatively lower $y^{(u,v,t)}$. This observation indicates that TSRec has indeed learned the relationship between the time intervals of the user’s historical repeat consumption and the target repeat time interval.

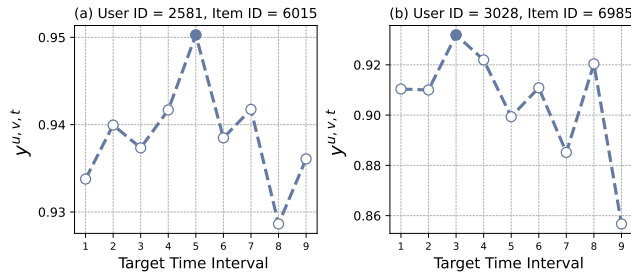


Figure 5: Test cases to verify that TSRec has indeed learned the temporal information of user repeat consumption. Hollow dots represent the target repeat time interval that we virtually generated, while solid dots represent the user’s last repeat consumption time interval.

5 RELATED WORK

Sequential Recommendation. Sequential Recommendation is a task used to predict users’ future preferences based on their personalized sequential behaviors. There are various works designed to

exploit and model users’ history to capture sequential relations and forecast their future behaviors. In these works, Factorized Personalized Markov Chain (FPMC) [26] integrated Matrix Factorization with Markov Chains for next-basket recommendations, embedding the information of adjacent behaviors into item latent factors. [14] followed this line and extend it to high-order Markov chains.

We summarized several works most related to our method. To model sequential behaviors, He et al. [12] proposed a sparse sequential recommendation approach with Markov chains. Meanwhile, Hidasi et al. proposed GRU4REC [13, 29] utilizing RNNs and an improved pairwise loss to model short-term preferences for session-based recommendations. Additionally, Li et al. proposed NARM [19] leveraging a global and local encoder with an attention module to model both user short- and long-term interest. Furthermore, other works employed memory networks [8] and CNNs [34] for sequential recommendations. Meanwhile, SASRec [15] and BERT4Rec[28] followed the structures of Transformer [30] and BERT [10] to bridge the item-item relationship and gain the benefits in the next-item prediction task. As these works are similar to our repeat-aware recommendation approach, we have selected several crucial or state-of-the-art recommendation models as comparative baselines.

Repeat-aware Recommendation. The repeat consumption has been extensively investigated in areas such as music listening [16, 24], E-commerce [31], grocery shopping [2], online novel reading [20], and live streaming [23]. In these fields, repeat consumption has been a vital indicator for modeling repeated user-item interaction. In recent years, there has been a series of studies that focus on modeling intrinsic patterns related to repeat consumption through simplified tasks. For instance, Anderson et al. [1] analyzed repeat consumption patterns and behaviors and identified recency as a crucial factor in predicting user purchasing patterns through a cache-based method. Their findings suggest that users tend to purchase recently-purchased items. Meanwhile, Benson et al. [4] demonstrated the impact of users getting bored with an item by analyzing the temporal gaps between users’ consecutive purchases of identical products. They found that greater temporal gaps between purchases indicated user boredom with the item, which eventually leads to abandonment. Wang et al. proposed SLRC [31] exploring the short-term effect and the lifetime effect to model repeat consumption by integrating Collaborative Filtering and Hawkes Process methods. While these approaches mainly model repeat consumption by exploring its characteristic, some studies have considered the design of model structures. For example, Ren et al. proposed RepeatNet [25], which used the copy attention mechanism to predict the probability of repetition and exploration. Ariannazhad et al. proposed ReCANet [2], a framework with LSTM layers to explicitly model the repeat behavior of users. Li et al. proposed NovelNet [20] which used a pointer network combined with a pointwise loss to model user repeat consumption behavior. Though the above works have considered modeling repeat consumption, they overlooked the fine-grained characteristics behind repeat consumption, which might lead to the degradation of model prediction performance.

6 CONCLUSION

In this paper, we investigate various inherent characteristics to enhance the repeat-aware recommendation. A novel approach called

Temporal and Sequential repeat-aware Recommendation (TSRec for short) is presented to capture the *temporal pattern* and *sequential pattern*. Offline experimental results on three public datasets demonstrate the superiority of TSRec over several state-of-the-art methods for both new and repeat items. Ablation studies illustrate the effectiveness of the three components of TSRec. Additionally, case studies demonstrate TSRec's ability to capture meaningful temporal information via time interval modeling. The phenomenon of repeat consumption is prevalent in many recommendation scenarios (e.g., e-commerce, music, and social media). The temporal and sequential patterns underlying repeat consumption represent the inherent characteristics of user repeat purchase behavior. We believe that modeling these patterns holds significant potential to enhance the performance of the recommendation system.

ACKNOWLEDGMENTS

To Robert, for the bagels and explaining CMYK and color spaces.

REFERENCES

- [1] Ashton Anderson, Ravi Kumar, Andrew Tomkins, and Sergei Vassilvitskii. 2014. The dynamics of repeat consumption. In *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014*, Chin-Wan Chung, Andrei Z. Broder, Kyuseok Shim, and Torsten Suel (Eds.). ACM, 419–430.
- [2] Mozhddeh Ariannezhad, Sami Jullien, Ming Li, Min Fang, Sebastian Schelter, and Maarten de Rijke. 2022. ReCANet: A Repeat Consumption-Aware Neural Network for Next Basket Recommendation in Grocery Shopping. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11–15, 2022*. 1240–1250.
- [3] Dmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.).
- [4] Austin R. Benson, Ravi Kumar, and Andrew Tomkins. 2016. Modeling User Consumption Sequences. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11–15, 2016*, Jacqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao (Eds.). ACM, 519–529.
- [5] Rahul Bhagat, Srevatsan Muralidharan, Alex Lobzhanidze, and Shankar Vishwanath. 2018. Buy It Again: Modeling Repeat Purchase Recommendations. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, Yike Guo and Faisal Farooq (Eds.). ACM, 62–70.
- [6] Qian Chen, Zhen-Hua Ling, and Xiaodan Zhu. 2018. Enhancing Sentence Embedding with Generalized Pooling. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, Emily M. Bender, Leon Derczynski, and Pierre Isabelle (Eds.). Association for Computational Linguistics, 1815–1826.
- [7] Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017. Enhanced LSTM for Natural Language Inference. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, Regina Barzilay and Min-Yen Kan (Eds.). Association for Computational Linguistics, 1657–1668.
- [8] Xu Chen, Hongteng Xu, Yongfeng Zhang, Jiaxi Tang, Yixin Cao, Zheng Qin, and Hongyuan Zha. 2018. Sequential Recommendation with User Memory Networks. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018*, Yi Chang, Chengxiang Zhai, Yan Liu, and Yoelle Maarek (Eds.). ACM, 108–116.
- [9] Kyunghyun Cho, Bart van Merriënboer, Çağlar Gülçehre, Dmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.)*. ACL, 1724–1734.
- [10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186.
- [11] ALAN G. HAWKES. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.
- [12] Ruining He and Julian J. McAuley. 2016. Fusing Similarity Models with Markov Chains for Sparse Sequential Recommendation. In *IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*, Francesco Bonchi, Josep Domingo-Ferrer, Ricardo Baeza-Yates, Zhi-Hua Zhou, and Xindong Wu (Eds.). IEEE Computer Society, 191–200.
- [13] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.
- [14] Balázs Hidasi, Massimo Quadrana, Alexandros Karatzoglou, and Domonkos Tikk. 2016. Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016*, Shilad Sen, Werner Geyer, Jill Freyne, and Pablo Castells (Eds.). ACM, 241–248.
- [15] Wang-Cheng Kang and Julian J. McAuley. 2016. Self-Attentive Sequential Recommendation. In *IEEE International Conference on Data Mining, ICDM 2016, Singapore, November 17-20, 2016*. 197–206.
- [16] Komal Kapoor, Karthik Subbian, Jaideep Srivastava, and Paul R. Schrater. 2015. Just in Time Recommendations: Modeling the Dynamics of Boredom in Activity Streams. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM 2015, Shanghai, China, February 2-6, 2015*, Xueqi Cheng, Hang Li, Evgeniy Gabrilovich, and Jie Tang (Eds.). ACM, 233–242.
- [17] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.)*. ACL, 1746–1751.
- [18] Yehuda Koren and Robert M. Bell. 2015. Advances in Collaborative Filtering. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 77–118.
- [19] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural Attentive Session-based Recommendation. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06–10, 2017*. 1419–1428.
- [20] Yuncong Li, Cunxiang Yin, Yancheng He, Guoqiang Xu, Jing Cai, Leeven Luo, and Sheng-hua Zhong. 2022. Modeling User Repeat Consumption Behavior for Online Novel Recommendation. In *RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18–23, 2022*, Jennifer Golbeck, F. Maxwell Harper, Vanessa Murdock, Michael D. Ekstrand, Bracha Shapira, Justin Basilico, Keld T. Lundgaard, and Even Oldridge (Eds.). 14–24.
- [21] Min Lin, Qiang Chen, and Shuicheng Yan. 2014. Network In Network. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.).
- [22] Yang Liu, Matt Gardner, and Mirella Lapata. 2018. Structured Alignment Networks for Matching Sentences. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsuiji (Eds.). Association for Computational Linguistics, 1554–1564.
- [23] Jérémie Rappaz, Julian J. McAuley, and Karl Aberer. 2021. Recommendation on Live-Streaming Platforms: Dynamic Availability and Repeat Consumption. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021*, Humberto Jesús Corona Pampin, Martha A. Larson, Martijn C. Willemsen, Joseph A. Konstan, Julian J. McAuley, Jean Garcia-Gathright, Bouke Huurnink, and Even Oldridge (Eds.). ACM, 390–399.
- [24] Markus Reiter-Haas, Emilia Parada-Cabaleiro, Markus Schedl, Elham Motamedi, Marko Tkalcic, and Elisabeth Lex. 2021. Predicting Music Relisting Behavior Using the ACT-R Framework. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021*, Humberto Jesús Corona Pampin, Martha A. Larson, Martijn C. Willemsen, Joseph A. Konstan, Julian J. McAuley, Jean Garcia-Gathright, Bouke Huurnink, and Even Oldridge (Eds.). ACM, 702–707.
- [25] Pengjie Ren, Zhumin Chen, Jing Li, Zhaochun Ren, Jun Ma, and Maarten de Rijke. 2019. RepeatNet: A Repeat Aware Neural Recommendation Machine for Session-Based Recommendation. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. 4806–4813.
- [26] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized Markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, Michael Rappa, Paul Jones, Juliana Freire, and Soumen Chakrabarti (Eds.). ACM, 811–820.

- [27] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. 1988. *Learning Representations by Back-Propagating Errors*. MIT Press, Cambridge, MA, USA, 696–699.
- [28] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3–7, 2019*. 1441–1450.
- [29] Yong Kiam Tan, Xinxing Xu, and Yong Liu. 2016. Improved Recurrent Neural Networks for Session-based Recommendations. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2016, Boston, MA, USA, September 15, 2016*, Alexandros Karatzoglou, Balázs Hidasi, Domonkos Tikk, Oren Sar Shalom, Haggai Roitman, Bracha Shapira, and Lior Rokach (Eds.). ACM, 17–22.
- [30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4–9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 5998–6008.
- [31] Chenyang Wang, Min Zhang, Weizhi Ma, Yiqun Liu, and Shaoping Ma. 2019. Modeling Item-Specific Temporal Dynamics of Repeat Consumption for Recommender Systems. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13–17, 2019*. 1977–1987.
- [32] Wei Wang, Chen Wu, and Ming Yan. 2018. Multi-Granularity Hierarchical Attention Fusion Networks for Reading Comprehension and Question Answering. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15–20, 2018, Volume 1: Long Papers*, Iryna Gurevych and Yusuke Miyao (Eds.). Association for Computational Linguistics, 1705–1714.
- [33] Hongzhi Yin, Bin Cui, Ling Chen, Zhiting Hu, and Xiaofang Zhou. 2015. Dynamic User Modeling in Social Media Systems. *ACM Trans. Inf. Syst.* 33, 3 (2015), 10:1–10:44.
- [34] Fajie Yuan, Alexandros Karatzoglou, Ioannis Arapakis, Joemon M. Jose, and Xiangnan He. 2019. A Simple Convolutional Generative Network for Next Item Recommendation. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11–15, 2019*. 582–590.
- [35] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Zhi Zhang, Haibin Lin, Yue Sun, Tong He, Jonas Mueller, R. Manmatha, Mu Li, and Alexander J. Smola. 2020. ResNeSt: Split-Attention Networks. *CoRR* abs/2004.08955 (2020).
- [36] Xiangmin Zhou, Dong Qin, Xiaolu Lu, Lei Chen, and Yanchun Zhang. 2019. Online Social Media Recommendation Over Streams. In *35th IEEE International Conference on Data Engineering, ICDE 2019, Macao, China, April 8–11, 2019*. IEEE, 938–949.