Time to Shop for Valentine's Day: Shopping Occasions and Sequential Recommendation in E-commerce

Jianling Wang
Texas A&M University
jlwang@tamu.edu

Raphael Louca, Diane Hu, Caitlin Cellier
Etsy Inc.
{rlouca,dhu,ccellier}@etsy.com

James Caverlee
Texas A&M University
caverlee@tamu.edu

Liangjie Hong
Etsy Inc.
lhong@etsy.com

ABSTRACT

Currently, most sequence-based recommendation models aim to predict a user's next actions (e.g. next purchase) based on their past actions. These models either capture users' intrinsic preference (e.g. a comedy lover, or a fan of fantasy) from their long-term behavior patterns or infer their current needs by emphasizing recent actions. However, in e-commerce, intrinsic user behavior may be shifted by occasions such as birthdays, anniversaries, or gifting celebrations (Valentine's Day or Mother's Day), leading to purchases that deviate from long-term preferences and are not related to recent actions. In this work, we propose a novel next-item recommendation system which models a user's default, intrinsic preference, as well as two different kinds of occasion-based signals that may cause users to deviate from their normal behavior. More specifically, this model is novel in that it: (1) captures a *personal occasion* signal using an attention layer that models reoccurring occasions specific to that user (e.g. a birthday); (2) captures a *global occasion* signal using an attention layer that models seasonal or reoccurring occasions for many users (e.g. Christmas); (3) balances the user's intrinsic preferences with the personal and global occasion signals for different users at different timestamps with a gating layer. We explore two real-world e-commerce datasets (Amazon and Etsy) and show that the proposed model outperforms state-of-the-art models by 7.62% and 6.06% in predicting users' next purchase.

CCS CONCEPTS

• **Information systems** → **Recommender systems**;

KEYWORDS

Recommendation; E-commerce; Temporal Effects; Occasion

ACM Reference Format:
Jianling Wang, Raphael Louca, Diane Hu, Caitlin Cellier, James Caverlee, and Liangjie Hong. 2020. Time to Shop for Valentine's Day: Shopping Occasions and Sequential Recommendation in E-commerce. In *The Thirteenth ACM International Conference on Web Search and Data Mining (WSDM '20), February 3–7, 2020, Houston, TX, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3336191.3371836

WSDM '20, February 3–7, 2020, Houston, TX, USA

ACM ISBN 978-1-4503-6822-3/20/02...$15.00
https://doi.org/10.1145/3336191.3371836

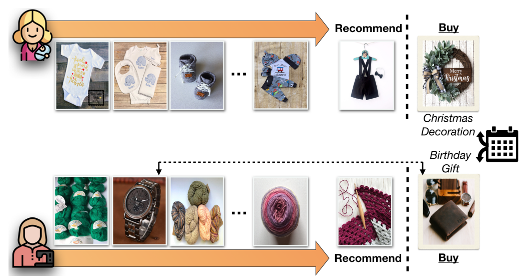

Figure 1: Example of Occasion-driven Purchases. User behavior in E-commerce is not always related to their recent actions or long-term intrinsic preferences, as assumed by many previous sequential recommendation systems. For example, a mom who frequently buys clothing for her infant will look for Christmas decorations near Christmas. A buyer who routinely purchases crochet supplies may purchase a birthday gift for her son every year.

1 INTRODUCTION

Recommendations act as an important component in e-commerce helping users discover interesting items that fit their needs. A well-performing recommendation system needs to infer and recommend items that are closely related to user preferences at a certain time. To handle the complex situation where user preferences can develop and change along time, recent efforts have focused on modeling users in a dynamic manner, which can adjust the recommendation based on the sequential behaviors of users [9, 18]. They either rely on the sequential transition between recent purchases [4, 21] or model the intrinsic preferences of users with different neural structures based on their historic sequential behaviors [10, 24, 33].

However, in e-commerce, users' shopping decisions can also be influenced by different occasions that lead to behavior which is not related to their recent actions or long-term intrinsic preferences. For example, a user who buys a pair of sandals in June would not want to be recommended an item for "Summer vacation" during the user's next shopping session in December. A "boho" style lover may purchase clothes or accessories that match her style, however, she may occasionally purchase a birthday gift for a friend whose

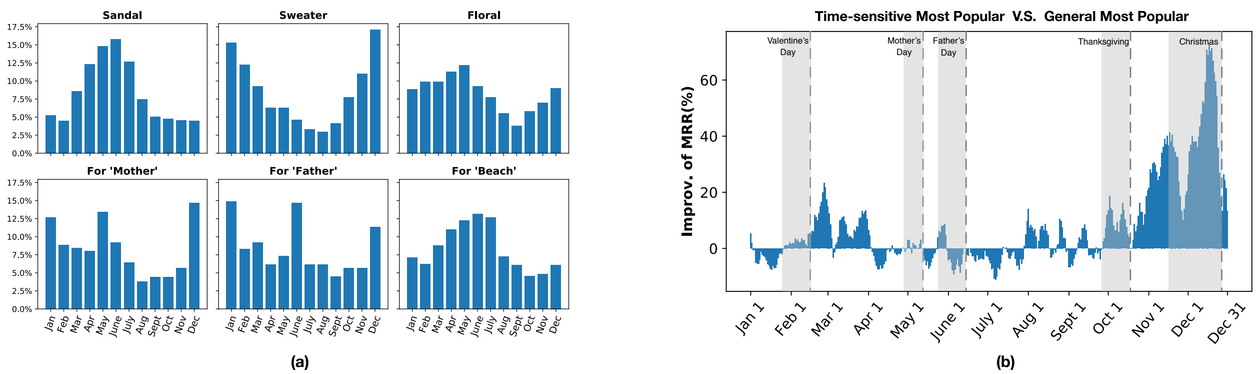

Figure 2: (a) In Amazon, users' shopping preferences are dynamic and can reflect reoccurring occasions (festivals, holidays, seasonal activities). We can detect occasion-based shopping trends from crowd behavior. (b) Recommending temporally popular items works better than recommending general popular items when there is an intense shopping trend for a specific occasion.

style is not "boho". Previous works assuming that users' actions are coherent or change smoothly along time can not handle such scenarios where users' behaviors can also be driven by different occasions (as illustrated in Figure 1).

Concretely, an *occasion* is a particular time or instance of an event that causes or triggers a purchase. There are *global occasions* which happen at the same time for a large number of users; examples include festivals or celebrations (like Christmas, Valentines' Day, Mother's Day) or seasonal events (like buying a snowboard in the Winter and a surfboard in the Summer). These global occasions are able to encourage or lead to similar shopping decisions for crowds of users. On the other hand, there are also *personal occasions*, which may happen at different timestamps for different users; examples include birthdays (for themselves or friends) and anniversaries. Those occasions usually occur in a periodic and repeated pattern for a specific user.

It is important to exploit the linkage between different occasions and shopping behaviors in e-commerce, so that we can: (i) recommend more time or season-aware candidates (like recommending a surfboard in the Summer while recommending snowboard in the Winter), which may alleviate the cold-start problem; (ii) reduce the noise in modeling users' intrinsic preferences since occasion-driven purchases (like gifts for others) may show different patterns compared to normal purchases from the same users; (iii) recommend relevant items to the user for upcoming reoccurring occasions. (Though the user may not purchase the exact same item for a reoccurring event, like consecutive Mother's Days events, the items purchased for Mother's Day previously will likely be related.)

There are several key challenges with using occasion signals in recommendation systems: (i) Are there traceable patterns distinguishing different occasions that we can use to holistically model a user's preference? (ii) Can we capture reoccurring shopping trends based on large crowd behavior? (iii) Can we model a flexible time-window for when occasions may reoccur? (iv) Can we properly balance a user's intrinsic preference versus the impact of a particular occasion in order to accurately predict their next purchase? Solutions to these challenges lead to a novel recommendation framework. Our major contributions in this work are:

- We uncover the patterns of shopping occasions and explore how they can change users' behaviors from both a global and a personal perspective.
- We propose to model the repeated personal occasion signals with attention layers, while modeling the global occasion signals by memorizing the temporal trends of shopping behaviors.
- With a gating component, we balance global and local effects of different occasions and propose *OAR* – an ***O**ccasion-**A**ware **R**ecommender* system for e-commerce while centering around each user's intrinsic preferences.

We conduct extensive experiments on real-world datasets from Etsy and Amazon and find that the proposed *OAR* outperforms the state-of-the-art approach in sequential recommendation.

2 MOTIVATION

In this section, we motivate the problem by showing evidence of different personal and global occasions that may influence the intrinsic purchase behavior of users. We collect data from Etsy and Amazon, two large, online marketplaces that sell products that are relevant to different occasions.

2.1 Temporal Shopping Trends

In Amazon [15], we can roughly infer users' shopping occasions or intentions with keywords that were mentioned in the reviews. Thus we summarize the occurrences of different keywords over different calendar months and show several examples in Figure 2(a) of possible occasion influences and how they change with time. For example, in the summer, users are likely to look for sandals instead of sweaters, while floral items are more popular in the spring and summer. As for gifting, we find that people tend to purchase for their mothers for Mother's Day (happening in May) or for Christmas. While approaching Father's Day, purchases peak in June, at which time people tend to purchase gifts for their fathers. We can conclude that users have changing preferences within a year for different occasions (festivals, holidays, seasonal activities) and crowds of users tend to purchase related items during similar

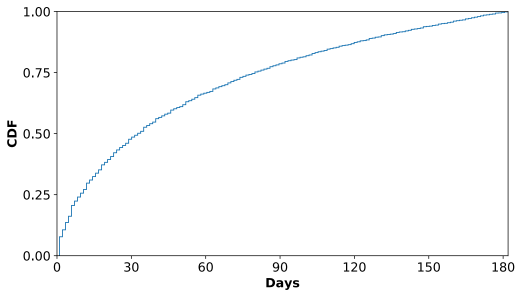

Figure 3: Time Gap between Purchases for Wedding and Anniversary within a year. More than 50% of purchases for anniversary are near the date of wedding purchase within a time window less than 30 days.

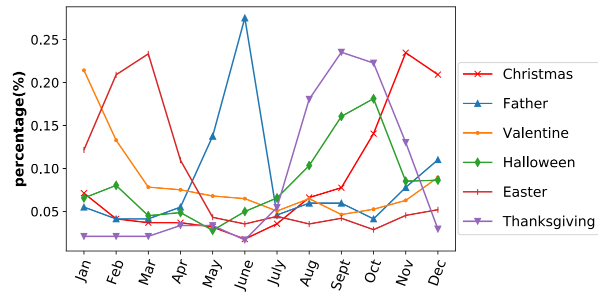

Figure 4: The reasons an infant's items shopper changes his/her shopping behaviors.

occasions. This analysis shows that capturing shopping trends as a function of time and season is useful for understanding purchase preference.

2.2 Occasion Signals for Recommendation

Next, we turn to Etsy to understand how global, annual occasions may influence normal shopping trends. Consider a simple Most Popular (MP) model that ranks products based on their overall popularity and then recommends the most popular items to users. A naive improvement to capture some occasion signals is a Temporal Most Popular (TMP) model, in which items are ranked based on their popularity within a short, recent time window, aggregated over all previous years. Here, we set the size of the time window to be 5 days, meaning that a prediction for January 10, 2018 would be obtained by ranking items that were most purchased between the dates of January 5 and January 15 over the course of the last 11 years.

We plot the improvement of TMP over MP for each day in the test set in Figure 2(b). When this ratio is positive (e.g. TMP is a better predictor than MP), it indicates that there is a strong (annually occurring) occasion-based shopping trend. As expected, we tend to see this pattern around big American holidays such as Christmas and Thanksgiving. We also note that the duration of this improvement is variable (e.g., the impact of Christmas lasts longer than Valentine's Day), thus indicating that hard-coded time window (5-days in this analysis) may not be flexible enough to model occasions with varying time effects. Conversely, in Figure 2(b), there are also time periods when TMP performs worse than MP, indicating that the crowd may not have strong time-dependent preferences during this time. Most of these areas fall during times when there are no globally celebrated occasions.

Note that recommending globally popular items to new, unseen users is a general solution for the cold-start problem. This analysis shows that using an occasion-aware global model can improve the accuracy of recommendations for new users who come to purchase for a special global occasion, such as Christmas or Thanksgiving.

2.3 Occasion Signals from Personal Perspective

In addition to global occasions, there may be occasions which may or may not be related to trending behavior, but can reoccur for individual users. For example, a user may look for birthday gifts for a parent every year as the birthday is approaching. These reoccurring occasions may lead to similar shopping behaviors across years, which we define as *personal occasions*. We want to explore the patterns of these personal occasions and whether they are traceable.

First, we sample 7,000 users in Etsy who have one purchase for "Wedding" and at least one purchase for "Anniversary" in the following years, which we can assume to be a relevant personal occasion. (In this exploration, we roughly infer the purchase intents based on the tags and description for the products.) We calculate the absolute difference between the date of the "wedding" purchase and "anniversary" purchase (aggregated over multiple years) of each user and plot the cumulative density (CDF) in Figure 3. We find that more than 50% of these users will purchase for "anniversary" occasions near the date of the original "wedding" purchase with a time gap of less than 30 days in the following years. So we can conclude that *there are occasions that may reoccur within a certain period (e.g. annually or monthly) over the course of multiple years and trigger relevant purchase*, revealing the traceable patterns in these personal occasion signals.

In Figure 4, we focus on users who purchase items for "infants/newborns/toddlers" in more than 50% of their transactions. We can assume that "buying products for infants" is their intrinsic preference for shopping, which are not related to occasions. Then, we summarize the tags/occasions of their "abnormal" purchases, e.g. the transactions without any infants items. While deviating from their intrinsic preference, these users tend to shop for Father's Day around June and Valentine's in January. Those occasions may reoccur each year and influence their purchase preference at a similar timestamp each year. Additionally, we find that preparation time for different occasions can vary. Users tend to start shopping for Christmas earlier than Valentine's or Father's Day. From a personal perspective, each user can deviate from their intrinsic preference and desire for different occasions. It is important to capture these personal occasion signals and adjust the recommendation when the reoccurring occasions is approaching.

3 OCCASION-AWARE RECOMMENDATION

In this section, we start with the problem setting and introduce the attention mechanism as our preliminary. Then we step though the development of the proposed *OAR* model by answering several research questions.

3.1 Problem Setting

Let $\mathbf{U} = \{u_1, u_2, ..., u_N\}$ represent the set of N users and $\mathbf{P} = \{p_1, p_2, ..., p_C\}$ represent the set of C products in a platform. In addition, let $\mathbf{T} = \{t_1, t_2, ..., t_M\}$ be the set of timestamps, which can be days, weeks or months in a calendar year. We sort the set of products user u has purchased in chronological order as $\mathbf{H}^u = ((p_1^u, t_1^u), (p_2^u, t_2^u), ..., (p_{|\mathbf{H}^u|}^u, t_{|\mathbf{H}^u|}^u))$. Each pair (p_n^u, t_n^u), $n \in [1, |\mathbf{H}^u|]$ denotes that user u purchases product p_n^u at time t_n^u. In e-commerce, we want to predict what a user want to purchase when he/she starts a (shopping) session at a future timestamp. Following the problem setting as in [4, 10] for sequential next-item recommendation, the goal of our work is to generate a list of top-k interesting items for user u at a future timestamp $t_{|\mathbf{H}^u|+1}^u$.

3.2 Preliminary: Attention Mechanism

To provide accurate recommendation, our goal is to understand how to aggregate the purchase record of a user in the past to infer the user's preferences in a future timestamp. The neural attention mechanism [2, 14, 25, 31] can be applied to capture the correlation between the target query (recent purchased items or the future timestamp for prediction) and the context contents (purchase history). For different types of attention modules, the input usually consists of a *Query*, and *Key-Value* pairs. The goal is to map the query with a set of key-value pairs to generate the output (as shown in Figure 5). An attention module can be divided into two steps. The first step entails computing the relationship/similarity scores between the *query* and a set of *keys*, which are used as the attention weights to aggregate the corresponding set of *values* [25]. Mathematically, given the input query $\mathbf{q}$ and a set of key-value pairs $\mathbf{P} = \{(\mathbf{k}_l, \mathbf{v}_l) \mid l \in [1, L]\}$, the resulted output $\mathbf{o}$ is calculated as:

$$\mathbf{o} = \sum_{l=1}^{L} \alpha_{ql} \mathbf{v}_l, \quad where \quad \alpha_{ql} = \frac{\exp(s(\mathbf{q}, \mathbf{k}_l))}{\sum_{l=1}^{L} \exp(s(\mathbf{q}, \mathbf{k}_l))} \tag{1}$$

where $s(\cdot, \cdot)$ is the similarity scoring function used to calculate the correlation between a query and a key. Based on our analysis in Section 2, we propose to make use of the attention mechanism for computing user profiling by taking different types of occasion signals into consideration, in addition to their intrinsic preferences. In the following sections, we will explain the details of each component in our OAR model (in Figure 6) through a discussion on the *design for query, key-value pair and the appropriate weight scoring function* to answer the following questions:

- **RQ1:** How to utilize the correlation between recent and historic purchased items to identify a user's intrinsic preferences which are mainly driven by a user's personal taste and self-desire?
- **RQ2:** How to model and predict user preferences for reoccurring personal occasions by tracing their personal shopping history?
- **RQ3:** How to memorize the crowd behavior at different time periods and perform dynamic mapping to aggregate the relevant global occasion signals?
- **RQ4:** How can we fuse intrinsic user preferences and different types of occasion signals to obtain a complete user profile that will inform what to recommend next?

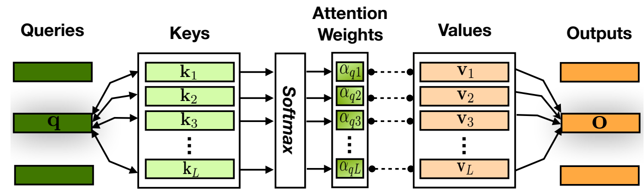

Figure 5: Attention Module.

3.3 RQ1: Intrinsic Preference Modeling

Users' intrinsic preferences on items are comparatively stable or change smoothly [12]. Thus previous works in recommendation usually model users in a static way with collaborative filtering-based methods [6, 22, 23, 30], or in a dynamics way by capturing the behavior patterns with the chronological order of user-item interactions via Markov Chains [5, 21], RNNs [8, 17, 28], and CNNs [24, 33]. Recently, self-attention [25], has demonstrated its effectiveness in sequential recommendation by capturing both the long-term semantics and relevant items with the recent interactions [10]. In a similar way, we try to model users' dynamic intrinsic preferences based on the correlation between the most recent purchase and the personal historic purchases.

Given the sequence of items user u has purchased $\mathbf{P}^u = (p_1^u, p_2^u, ..., p_{|P^u|}^u)$ in chronological order, we use the combination $\mathbf{m}_{p_d^u} = \mathbf{e}_{p_d^u} + \mathbf{x}_d$ to represent the item at position d (the dth item in the sequence). Here, $\mathbf{e}_{p_d^u}$ is the embedding for item p_d^u and $\mathbf{x}_d$ is the positional embedding of position d, which is used to retain the order information. Self-attention [25] is designed to match a sequence against itself and thus uses the same objects as the queries, keys and values. In our case, we will map the query item p_d^u to the sequence of items $(p_1^u, p_2^u, ..., p_d^u)$, which have been purchased by u no later than p_d^u. Before calculating the attention weights and aggregation, we conduct linear projections for each $\mathbf{m}_{p_d^u}$, $p_d^u \in \mathbf{P}^u$ with matrices $\mathbf{W}^Q$, $\mathbf{W}^K$, $\mathbf{W}^V$ to generate embedding $\hat{\mathbf{m}}_{p_d^u}^Q = \mathbf{m}_{p_d^u}\mathbf{W}^Q$, $\hat{\mathbf{m}}_{p_d^u}^K = \mathbf{m}_{p_d^u}\mathbf{W}^K$, $\hat{\mathbf{m}}_{p_d^u}^V = \mathbf{m}_{p_d^u}\mathbf{W}^V$ for queries, keys and values correspondingly. Thus, we have:

$$Query: \hat{\mathbf{m}}_{p_d^u}^Q \qquad Scoring: s(\mathbf{q}, \mathbf{k}_j) = \frac{\mathbf{q}\mathbf{k}_j^T}{\sqrt{D}}$$

$$(Key,\ Value): (\hat{\mathbf{m}}_{p_1^u}^K, \hat{\mathbf{m}}_{p_1^u}^V), (\hat{\mathbf{m}}_{p_2^u}^K, \hat{\mathbf{m}}_{p_2^u}^V), ..., (\hat{\mathbf{m}}_{p_d^u}^K, \hat{\mathbf{m}}_{p_d^u}^V)$$

Here we adopt the scaled dot-product to calculate the score between keys and queries. D denotes the dimension of the embedding. The output $\mathbf{o}_{u,t_{d+1}^u}^I$ based on the most recent item p_d^u can represent the dynamic intrinsic preference of user u after purchasing p_d^u, and will be used to infer the user's next purchase at t_{d+1}^u.

3.4 RQ2: Personal Occasion Elicitation

Based on our exploration in Section 2, we know that users can deviate from their intrinsic preferences because of some personally reoccurring occasions. For each user, the shopping behaviors driven by the same personal occasions are likely to fall into a small time window. For example, a user will often purchase a birthday gift two to three weeks in advance of the birthday. Thus while

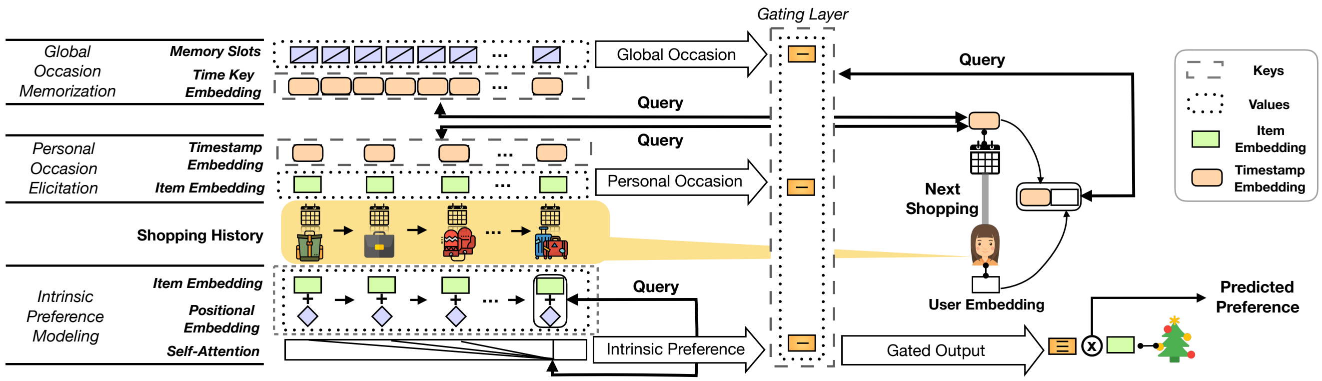

Figure 6: The proposed *Occasion-Aware Recommendation* (OAR) model.

predicting a user's preference, we also need to elicit the personal occasion signal by tracing the user's previous shopping behavior in the neighboring days. In this component, we want to map the upcoming timestamp (query) with the timestamps of the user's previous purchases (keys) and the corresponding items (values). Given the $\mathbf{H}^u = ((p_1^u, t_1^u), (p_2^u, t_2^u), ..., (p_{|\mathbf{H}^u|}^u, t_{|\mathbf{H}^u|}^u))$, we use $\mathbf{t}_{t_d^u}$ to denote the embedding of timestamp t_d^u. As in Section 3.3, the time embedding for queries or keys will be multiplied with the matrices $\mathbf{W}^{Q'}$ and $\mathbf{W}^{K'}$ respectively, with $\hat{\mathbf{t}}_{t_{d+1}^u}^{Q'} = \mathbf{t}_{t_{d+1}^u}\mathbf{W}^{Q'}$ and $\hat{\mathbf{t}}_{t_d^u}^{K'} = \mathbf{t}_{t_d^u}\mathbf{W}^{K'}$. We also apply linear projection for the item embedding with $\mathbf{W}^{V'}$ to generate the embedding for values. While predicting for u at a future time t_{d+1}^u, the personal occasion preference can be obtained with the attention operation below:

$$\textit{Query}: \hat{\mathbf{t}}_{t_{d+1}^u}^{Q'} \qquad \textit{(Key, Value)}: (\hat{\mathbf{t}}_{t_1^u}^{K'}, \hat{\mathbf{e}}_{p_1^u}^{V'}), (\hat{\mathbf{t}}_{t_2^u}^{K'}, \hat{\mathbf{e}}_{p_2^u}^{V'}), ..., (\hat{\mathbf{t}}_{t_d^u}^{K'}, \hat{\mathbf{e}}_{p_d^u}^{V'})$$

We use the same similarity function $s(\cdot,\cdot)$ as in Section 3.3. While generating the output $\mathbf{o}_{u,t_{d+1}^u}^P$, items which were purchased a long time ago but within a small time window with the query's upcoming timestamp can also get high attention from the model. In this way, *OAR* can capture personally reoccurring occasions.

3.5 RQ3: Global Occasion Memorization

By only tracing the personal purchase history, the model is still unable to predict upcoming global occasions. However, these occasion signals can be captured from the behaviors of the crowd from a neighboring time period in the past. Under a global occasion, the crowd of users tends to have similar purchases, like shopping for costumes before Halloween or green shirts near St Patrick's day. We aim to memorize the shopping behaviors of the crowd under different global occasions, which can be used to enrich the preference representation of individual users when a certain occasion is coming. Following a similar idea as in the key-value memory network [16], we use the timestamps as keys and pair each of the keys with a memory slot to represent preferences of the crowd at the timestamp.

Let $\mathbf{T} = \{\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_M\}$ denote the set of embedding for each timestamp. We use $\hat{\mathbf{t}}_i, i \in [1, M]$, which is the projected version of the embedding for timestamp $\hat{\mathbf{t}}_i = \mathbf{t}_i\mathbf{W}^{K''}$, to be the key. Furthermore, we set a separate memory slot $\mathbf{r}_i, i \in [1, M]$ to store global behaviors. Given a query timestamp, we will multiply its embedding $\mathbf{t}_{t_{d+1}^u}$ with matrix $\mathbf{W}^{Q''}$ to get $\hat{\mathbf{t}}_{t_{d+1}^u}^{Q''}$. Then we want to map it with all the key-value memory slots to get the corresponding global occasion representation. As in Section 3.3, we use scaled dot-product as the similarity scoring function $s(\cdot,\cdot)$ and:

$$\textit{Query}: \hat{\mathbf{t}}_{t_{d+1}^u}^{Q''} \qquad \textit{(Key, Value)}: (\hat{\mathbf{t}}_1, \mathbf{r}_1), (\hat{\mathbf{t}}_2, \mathbf{r}_2), ..., (\hat{\mathbf{t}}_M, \mathbf{r}_M)$$

The output $\mathbf{o}_{t_{d+1}^u}^G$ of the attention operation can be the representation of global occasions at t_{d+1}^u.

3.6 RQ4: Gating Layer

Lastly, we discuss how to balance a user's intrinsic preferences with occasion signals for personalization? Here we turn to an attention (gating) layer which can control how we assign different weights to each of the components we have developed in the previous sections. The query will be a user-timestamp pair because the status for a user at different timestamps will be different. For example, there are users who have strong personal desire for handcrafted supplies and seldom purchase other items on a site like Etsy. Or users may tend to be influenced by their surroundings in December but may stick to their own intrinsic preference in June (as shown in Figure 2(b)). While predicting for user u at timestamp t_{d+1}^u, with embedding $\mathbf{u}$ and $\mathbf{t}_{d+1}^u$ for u and t_{d+1}^u:

$$\textit{Query}: \mathbf{u}||\mathbf{t}_{d+1}^u \qquad \textit{Scoring}: s(\mathbf{q}, \mathbf{k}_j) = \mathbf{a}^T \mathit{tanh}(\mathbf{W}[\mathbf{q}||\mathbf{k}_j])$$

$$\textit{(Key, Value)}: (\mathbf{o}_{u,t_{d+1}^u}^I, \mathbf{o}_{u,t_{d+1}^u}^I), (\mathbf{o}_{u,t_{d+1}^u}^P, \mathbf{o}_{u,t_{d+1}^u}^P), (\mathbf{o}_{t_{d+1}^u}^G, \mathbf{o}_{t_{d+1}^u}^G)$$

in which $||$ denotes concatenation, and $\mathbf{a}$ and $\mathbf{W}$ represent the transform vector and matrix, respectively, for this additive attention operation. Thus we get the output $\mathbf{o}_{u,t_{d+1}^u}$, which can be used to accurately represent u's preference at future timestamp t_{d+1}^u.

3.7 Prediction and Loss

After generating $\mathbf{o}_{u,t_{d+1}^u}$ as the complete representation of a user's current status, we can predict the preference score on item i with $\hat{y}_{ui}^{t_{d+1}^u} = \mathbf{o}_{u,t_{d+1}^u}\mathbf{e}_i$. We adopt the Bayesian Pairwise Loss [20] to

Dataset	#Users	#Items	#Purchases	Density	Cutting Time
Amazon	84,191	100,946	1.0M	0.0124%	2013/8/1
Etsy	118,668	80,214	5.3M	0.0561%	2018/1/1

Table 1: Dataset Statistics.

maximize the gap between the ground truth positive user-item pair and negative sampled pairs. The loss function is:

$$L = \sum_{(u,t,i,j)\in\mathbf{D}} -\ln\sigma(\bar{y}_{ui}^{t} - \bar{y}_{uj}^{t}) + \lambda||\theta||^2$$

where $||\theta||^2$ is a regularization term and $\sigma(\cdot)$ is the Sigmoid function. Each element (u, t, i, j) in the training data set $\mathbf{D}$ is generated by combining the ground truth interaction pair (u, t, i), which means u purchased i at t, with a negative sampled item j that u did not purchase at time t.

4 EXPERIMENTS

In this section, we conduct experiments on two real-world datasets from e-commerce platforms to answer several research questions: (i) How does the proposed *OAR* model perform compared with other sequential models in real-world e-commerce scenarios? (ii) How does each component contribute to the user modeling and impact next-item recommendation? (iii) Does the proposed structure of *OAR* successfully capture different occasion signals? and (iv) Can we visualize the patterns of shopping occasions learned by the attention mechanisms?

4.1 Data

To avoid data leakage while modeling the crowd behaviors in the global occasion component, we split the datasets for training and testing with a cutting time. We only use data before the cutting date to train the model. In both datasets, we keep users who purchased at least twice after the cutting time, so that we can use the first purchase of each user after the cutting date as a validation case and the second purchase as a test case. The detailed information is summarized in Table 1. In the experiments, we consider each day in the calender year as a timestamp, that is $t_1 \in \mathbf{T}$ means the first day in a year (January 1).

Etsy. We collect purchase data from November 2006 to December 2018 in Etsy, which is one of the largest e-commerce platform selling handmade items. We filter out users with fewer than 5 purchases before the cutting time. To examine the long-term effects, we keep only users who are active for at least two years, requiring that the time gap between their last purchase and their first purchase be larger than 365 days.

Amazon. We test over a public Amazon review dataset [15], containing product reviews from May, 1996 to July, 2014. We treat each review as a purchase record and use the time they input the review to approximate the purchase time. We filter out users who purchased fewer than 5 items before the cutting time.

4.2 Experimental Setup

Metrics. Following the evaluation strategy as in [6, 10], for each user u, we randomly sample 100 negative items, with which we rank the ground-truth items in the test set of u while generating the top-K recommendation. We adopt the metrics commonly used for next-item recommendation task for evaluation, including Normalized Discounted Cumulative Gain (NDCG@K), Hit Rate (HR@K), and Mean Reciprocal Rank (MRR).

Since there is only one item in test or validation set for each user (leave-one-out task), Hit Rate (HR@K) is equivalent to recall, indicating whether the ground-truth item is among the top-K ranked list of items. Also for each user, the ideal discounted cumulative Gain (IDCG) is equal to 1. Let $rank_u$ represent the predicted ranking of the ground-truth item in the test for user u. In top-K evaluation, if $rank_u \leq K$, then $NDCG_u@K = DCG_u@K = \frac{1}{\log_2(rank_u+1)}$. Otherwise, $NDCG_u@K = 0$. We also use the mean reciprocal rank $MRR = \sum_{u\in\mathbf{U}} \frac{1}{rank_u}$ to evaluate the positions of recommendation.

Baselines.

- **MP:** *Most Popular.* It ranks all the products based on their overall popularity and recommends the most popular products.
- **MF-BPR:** *Matrix Factorization with Bayesian Personalized Ranking* [20]. This model predicts user's preference on a product based on the multiplication between their latent factors (MF) and is optimized with Bayesian personalized ranking (BPR) loss.
- **Fossil:** *Fusing Similarity Models with Markov Chains* [5]. It improves the method of factorizing personalized Markov Chain (FPMC) with item similarity-based algorithm (FISM) to capture the long-term and short-term dynamics of users simultaneously.
- **GRU4Rec+**: *Recurrent Neural Networks with Top-k Gains* [7]. It is similar to GRU4Rec [8] in utilizing GRU model to capture the sequential patterns, but with a modified loss function and sampling strategy to achieve better performance in the Top-K recommendation task.
- **TCN**: *A Simple Convolutional Generative Network for Next Item Recommendation* [33]. This is an improved dilated convolution neural network (CNN) modeling both short and long-range item dependencies in a sequence to recommend the next item.
- **HPMN**: *Lifelong Sequential Modeling with Hierarchical Periodic Memory Network* [18]. It capture the multi-scale sequential patterns of users in e-commerce with a hierarchical and periodical updating mechanism. It is able to model users' periodic behavior patterns appearing in both long-term or short-term.
- **SARec**: *Self-attentive sequential recommendation* [10] With the self-attention layers, this model is able to balance the long-term effect of a sequence and from recent products.

Parameters. All experiments are conducted with a single Nvidia TITAN Xp GPU. For HPMN, SARec, TCN and GRU4Rec+, we use the implementations provided in their original papers. We implement other baselines and the proposed OAR model with TensorFlow.

For fair comparison, we adopt BPR loss [20] and set the negative sampling rate to be 1 for all the models. The maximum length of shopping record is fixed on 50 in Amazon data and 100 in Etsy data. Batch size is set be 128 for all the models. We grid search for the best size of hidden layer or latent factor over {10, 20, 50, 100, 150, 200}. The learning rate is searched over {0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1}, the coefficient of $L2$ regularization (λ in loss function) is over $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ and the optimization

Model	*Etsy*					*Amazon*				
	NDCG		HR		MRR	NDCG		HR		MRR
	K=5	K=10	K=5	K=10		K=5	K=10	K=5	K=10	
MP	0.1531	0.1919	0.2304	0.3511	0.1673	0.2129	0.2509	0.3020	0.4199	0.2195
MF-BPR	0.4519	0.5001	0.5947	0.7434	0.4376	0.2663	0.3012	0.3619	0.4698	0.2668
Fossil	0.4946	0.5354	0.5511	0.7630	0.4746	0.2160	0.2483	0.2967	0.3969	0.2221
TCN	0.5199	0.5726	0.6698	0.8059	0.5090	0.2632	0.3029	0.3664	0.4893	0.2650
GRU4Rec+	0.5346	0.5771	0.6830	0.8136	0.5126	0.2763	0.3169	0.3828	0.5087	0.2770
HPMN	0.5480	0.5883	0.6962	0.8201	0.5245	0.2820	0.3216	0.3881	0.5109	0.2819
SARec	0.5665	0.6047	0.7102	0.8278	0.5433	0.3009	0.3385	0.4085	0.5251	0.2984
OAR	**0.6078***	**0.6415***	**0.7425***	**0.8462***	**0.5847***	**0.3200***	**0.3580***	**0.4301***	**0.5476***	**0.3165***

Table 2: Comparison of Different Models. $*$ indicates that the improvement of the best result is statistically significant compared with second best result with $p < 0.01$.

Model	Etsy		Amazon	
	NDCG@5	MRR	NDCG@5	MRR
Global (G)	0.1816	0.1953	0.2238	0.2294
Intrinsic (I)	0.5665	0.5433	0.3009	0.2984
Personal (P)	0.5791	0.5582	0.3069	0.3047
I + G	0.5885	0.5642	0.3099	0.3063
I + P	0.5916	0.5677	0.3136	0.3108
Remove Gate	0.5859	0.5618	0.3074	0.3039
OAR	**0.6078**	**0.5847**	**0.3200**	**0.3165**

Table 3: Ablation Test Results.

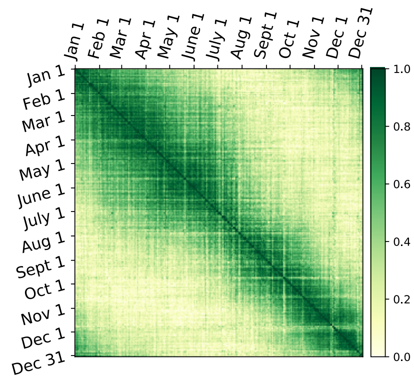

Figure 7: Similarity between different calendar days.

methods is over {Adam, Adagrad, SGD}. We also fine-tune all the model-specific hyperparameters and report the best performance in the following sections.

4.3 Model Comparison

We summarize the best performance of all the baseline models and the proposed model in Table 2. We can see that OAR achieves the best performance under different metrics in both datasets. It gains 7.62% and 6.06% MRR improvement in Etsy and Amazon compared with the state-of-the-art.

Compared with the basic general MP, we can see that MF-BPR which represents users and items with static latent factors can achieve a 177.43% and 39.67% improvement on average in Etsy and Amazon. Then by introducing the Markov Chains to capture the transition of users among different items, we find that Fossil works better than MF in Etsy but performs worse in Amazon. Presumably it is because the Amazon data is extremely sparse and results in an unstable factorized Markov Chains component in Fossil.

Comparing the recent neural-based sequential models, we find that GRU4Rec+ works slightly better than TCN, which is based on dilated CNN. And HPMN utilizing hierarchical multi-layer memory networks outperforms GRU4Rec+ in both data, which proves that there are periodic pattern in users' shopping behaviors. However, HPMN model assumes that the period of shopping behavior is constant for all the users along the time and thus lack of flexibility to handle the real-world scenarios. We find that SARec, which is utilizing self-attention to model users' intrinsic preference, works even better than HPMN. This shows that attention mechanisms are a good fit for modeling sequential behaviors. And by carefully eliciting the occasion signals and combining them with the intrinsic preferences, OAR achieves the best performance in the next-item prediction via an accurate user model.

4.4 Evaluation of OAR

To examine whether each component in OAR achieves its goal and to understand how it contributes to the recommendation, we analyze their impacts with an ablation test (in Table 3).

The Global occasion component (G), in which we set up a certain number of memory slots to record the crowd behavior in different occasions, does not provide personalized recommendation individually. It can outperform the general Most Popular (MP) model by 17.17% and 4.89% in Etsy and Amazon, which demonstrates that it can capture the temporal global occasion signals hidden in the crowd behavior. Additionally, we can infer that the users in Etsy are more likely to follow the temporal global trends in shopping. Both Intrinsic and Personal components can provide personalized next-item recommendation. In Intrinsic (I), it maps the most recent purchase to the items purchased before to infer the "relevant" items in the future. While in Personal (P), the main idea is to trace back to the previous behaviors in the related time periods. We can see that P performs slightly better than I, which means that in e-commerce, it is important to predict the shopping occasion and pay more attention to the items purchased around similar occasions while inferring the next purchase. While combining the I and G or I and P, we can see the joint models can improve each of the individual components. Thus we find that in e-commerce platforms,

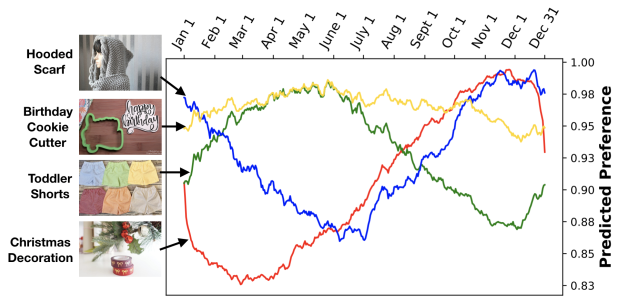

Figure 8: The average preferences predicted by OAR.

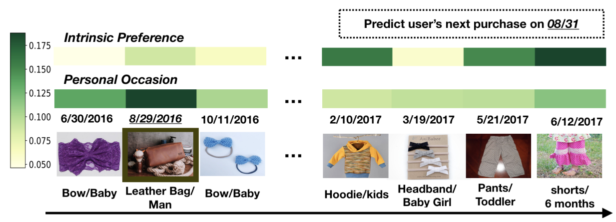

Figure 9: The attention weights by different components.

it is necessary to take the occasion signals into consideration while making recommendations.

To examine the impact of the gating component, which is designed for a personalized and temporal-aware fusing of intrinsic preference and the occasion signals, we replace it with a simple addition layer. That is we use $\mathbf{o}^{I}_{u,t^u_{d+1}} + \mathbf{o}^{P}_{u,t^u_{d+1}} + \mathbf{o}^{G}_{t^u_{d+1}}$ as a representation of user u at time t^u_{d+1} while removing the Gating Layer. We find that there is a large drop in recommendation quality, which supports the assumption that the influence of different occasions does vary for different users at different timestamps. Thus, it is important to take the personalization and temporal information into consideration simultaneously while utilizing the occasion signals.

4.5 Case Study

4.5.1 Temporal Information and Occasions. To examine whether the proposed model is able to capture the occasions by linking the neighboring time periods to the relevant occasions, we plot out the attention weights (or similarity scores) between each timestamp (in Figure 7). Near the diagonal (similarity between the exact same timestamp), we can see there are many dark regions, indicating the strong correlation between nearby time periods. For some of the regions, the dark color diffuses to a large area (like around March and April), meaning that the occasions at that time have a continuous lasting influence. Since the occasion calendar is a loop, there is high correlation between dates in December and dates in January, which results in the dark region at the left bottom and right top corner. Thus we find that *OAR* is capable in modeling the occasion signals along time.

4.5.2 Visualization of Occasion-driven Purchases in Etsy. To explore whether *OAR* captures the occasion signals to adjust the recommendation at different timestamps, we predict users' preferences on several items every day in the test year. We calculate and plot out the average predicted preference scores for all the users in Figure 8. We find that the preferences for the hooded scarf drop down when the weather gets warmer and increases in Fall and Winter time, while the preferences for shorts are in a totally opposite pattern. And for the Christmas decoration tab (red line), the preferences on it reach the peak in early December but drop down rapidly after Christmas, meaning the product is sensitive to the occasion. However, for items which are fit for occasions that can happen all year round (like birthdays), the average preference on it is flat during the year.

4.5.3 How Occasion Signals compensate the intrinsic preference. We show the results for an Etsy user as an example (in Figure 9) to examine how the occasion signals supplement the intrinsic preference for improved recommendation. In intrinsic preference modeling, a high score will be assigned to the most recent purchase (shorts for the 6-month) and items relevant to that (baby's clothing). Thus, while predicting for August 31 with the intrinsic preference individually, we will keep recommending similar items. However, in personal occasion elicitation, it traces the history and assigns high score to items which are purchased in the related time windows. So that in this case, though the user purchased lots of baby clothing, by capturing the occasion signals, the "leather bag for man" purchased on August 29 two years ago still receives high attention. We can see that *OAR* is able to recall the purchase for the leather bag and thus recommend some related items for the upcoming occasion.

5 RELATED WORK

Sequential Recommendation. Recently there is an increasing attention on predicting the next interesting items based on users' sequential actions in the past [8, 13, 29]. Previous efforts have explored various methods to model the sequences of users' behaviors with or without user identification information. By placing users and items on the same embedding space, TransRec [4] treats users as "translation vectors" that transit between items (points on the space). And in [24, 33], they propose to utilize a CNN to aggregate the sequential behaviors of users. SR-GNN [29] constructs a graph for each behavior sequence based on the transition of items and predict for the next item with the embedding resulted from graph neural networks (GNN). GRU4Rec [8] consists of GRU layers to learn the pattern from users' feedback sequences to generate recommendation. NARM [13] enriches RNN with the local information generated from an attention network which aggregates the hidden output of RNN at each timestamp. Based on the success in replacing RNN with attention networks and transformer [25], SARec [10] is proposed to use a self-attention based model to infer relevant items based on users' action history, which outperforms various state-of-the-art sequential models in recommendation. However, none of these models take the occasion signals into consideration and are not a good fit for the occasion-driven scenarios in e-commerce.

Temporal Effects & Dynamic User Modeling in E-commerce. There are works which have been done on dynamic user modeling considering the temporal effects. Koren proposes to divide the long time series into slices and training for different latent representations at each slice in TimeSVD++ [11]. Utilizing the explicit time stamp, in [27, 28], they use parallel RRN structure to model the dynamics of users and items simultaneously. The work in [9] explores how users' shopping decisions can be influenced by the life-stage along time, and proposes to select corresponding recommendation model after labeling consumer's life-stage.

While focusing on the sequential behavior patterns of users in e-commerce, there are previous works assuming that a user would behave centering around the intense shopping intent and tend to interact with the exact same items repeatedly [1, 3]. RepeatNet [19] predicts the probability of being repeated for a user at each timestamp, and then decide whether to recommend from the purchased items or new items. In [26], they model the repeat consumption of different products with Hawkes Process and integrate the resulting signals into Collaborative Filtering to generate recommendations. However, these models can not be generalized to many shopping platforms where a user seldom purchases the exact same item repeatedly (like clothes, accessories and books).

There are also works trying to capture both the long-term dynamics and short-term effects simultaneously building on top of hierarchical structures. HRNN [17] consists of a two-layer hierarchical RNN, which learns the representation for each short-term session with a lower layer RNN and then aggregates the resulting outputs from the same user with a higher layer RNN. The work of [32] achieves a similar goal with hierarchical attention layers. HPMN [18] is proposed to model the periodic patterns of users with a hierarchical recurrent memory network. Although these methods can model the dynamic users preferences, they do not take the influence of different occasions into consideration.

6 CONCLUSION

Shopping decisions can be influenced by different occasions, leading to purchases that deviate from a user's intrinsic preferences. Over Amazon and Etsy, we gain insights into the traceable patterns of personal and global occasion signals. We propose to utilize different attention mechanisms to elicit different occasion signals for recommendation. Through experiments, we find the proposed ***O***ccasion-***A***ware ***R***ecommender model can outperform the state-of-the-art model in two real-world e-commerce datasets. Next, we are interested in introducing more context information to characterize the occasions explicitly and provide explainable recommendations.

REFERENCES

[1] Ashton Anderson, Ravi Kumar, Andrew Tomkins, and Sergei Vassilvitskii. 2014. The dynamics of repeat consumption. In *Proceedings of the 23rd international conference on World wide web*. ACM, 419–430.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *ICLR*.

[3] 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*. ACM, 62–70.

[4] Ruining He, Wang-Cheng Kang, and Julian McAuley. 2017. Translation-based recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 161–169.

[5] Ruining He and Julian McAuley. 2016. Fusing similarity models with markov chains for sparse sequential recommendation. In *ICDM*. IEEE, 191–200.

[6] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 173–182.

[7] Balázs Hidasi and Alexandros Karatzoglou. 2018. Recurrent neural networks with top-k gains for session-based recommendations. In *CIKM*.

[8] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based recommendations with recurrent neural networks. In *ICLR*.

[9] Peng Jiang, Yadong Zhu, Yi Zhang, and Quan Yuan. 2015. Life-stage prediction for product recommendation in e-commerce. In *Proceedings of the 21th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1879–1888.

[10] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *ICDM*.

[11] Yehuda Koren. 2009. Collaborative filtering with temporal dynamics. In *KDD*.

[12] Huayu Li, Yong Ge, Defu Lian, and Hao Liu. 2017. Learning User's Intrinsic and Extrinsic Interests for Point-of-Interest Recommendation: A Unified Approach.. In *IJCAI*. 2117–2123.

[13] 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 on Conference on Information and Knowledge Management*. ACM, 1419–1428.

[14] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *EMNLP*.

[15] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 43–52.

[16] Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. 2016. Key-value memory networks for directly reading documents. *arXiv preprint arXiv:1606.03126* (2016).

[17] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing session-based recommendations with hierarchical recurrent neural networks. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 130–137.

[18] Kan Ren, Jiarui Qin, Yuchen Fang, Weinan Zhang, Lei Zheng, Weijie Bian, Guorui Zhou, Jian Xu, Yong Yu, Xiaoqiang Zhu, et al. 2019. Lifelong Sequential Modeling with Personalized Memorization for User Response Prediction. In *SIGIR*.

[19] 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 *AAAI*.

[20] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI Press, 452–461.

[21] 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*. ACM, 811–820.

[22] Badrul Munir Sarwar, George Karypis, Joseph A Konstan, John Riedl, et al. 2001. Item-based collaborative filtering recommendation algorithms.. In *WWW*.

[23] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. 2015. Autorec: Autoencoders meet collaborative filtering. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 111–112.

[24] Jiaxi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 565–573.

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[26] 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*. ACM, 1977–1987.

[27] Jianling Wang and James Caverlee. 2019. Recurrent Recommendation with Local Coherence. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 564–572.

[28] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J Smola, and How Jing. 2017. Recurrent recommender networks. In *Proceedings of the tenth ACM international conference on web search and data mining*. ACM, 495–503.

[29] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based Recommendation with Graph Neural Networks. In *AAAI*.

[30] Yao Wu, Christopher DuBois, Alice X Zheng, and Martin Ester. 2016. Collaborative denoising auto-encoders for top-n recommender systems. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 153–162.

[31] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*. 2048–2057.

[32] Haochao Ying, Fuzhen Zhuang, Fuzheng Zhang, Yanchi Liu, Guandong Xu, Xing Xie, Hui Xiong, and Jian Wu. 2018. Sequential recommender system based on hierarchical attention networks. In *the 27th International Joint Conference on Artificial Intelligence*.

[33] 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*. ACM, 582–590.