SMINet: State-Aware Multi-Aspect Interests Representation Network for Cold-Start Users Recommendation

Wanjie Tao,[1] **Yu Li,**[2] **Liangyue Li,**[1] **Zulong Chen,**[1] **Hong Wen,**[1]
Peilin Chen,[1] **Tingting Liang,**[2] **Quan Lu**[1]
[1] Alibaba Group,Hangzhou,China
[2] Hangzhou Dianzi University,Hangzhou,China
{wanjie.twj,liliangyue.lly,zulong.czl,qinggan.wh,peilin.cpl,luquan.lq}@alibaba-inc.com, {liyucomp,liangtt}@hdu.edu.cn

Abstract

Online travel platforms (OTPs), e.g., bookings.com and Ctrip.com, deliver travel experiences to online users by providing travel-related products. Although much progress has been made, the state-of-the-arts for cold-start problems are largely sub-optimal for user representation, since they do not take into account the unique characteristics exhibited from user travel behaviors. In this work, we propose a State-aware Multi-aspect Interests representation Network (SMINet) for cold-start users recommendation at OTPs, which consists of a multi-aspect interests extractor, a co-attention layer, and a state-aware gating layer. The key component of the model is the multi-aspect interests extractor, which is able to extract representations for the user's multi-aspect interests. Furthermore, to learn the interactions between the user behaviors in the current session and the above multi-aspect interests, we carefully design a co-attention layer which allows the cross attentions between the two modules. Additionally, we propose a travel state-aware gating layer to attentively select the multi-aspect interests. The final user representation is obtained by fusing the three components. Comprehensive experiments conducted both offline and online demonstrate the superior performance of the proposed model at user representation, especially for cold-start users, compared with state-of-the-art methods.

Introduction

Online travel platforms (OTPs), e.g., bookings.com and Ctrip.com, deliver travel experiences to online users by providing travel-related products (e.g., flight tickets, hotels, and package tours, etc). With the large-scale product portfolio available on the platform, high-quality personalized recommendations are essential for delivering superb user experiences. The recommender system usually follows the classic two-stage paradigm, which consists of a matching phase that generates candidate items for the user and a ranking phase that ranks the items according to conversion rate (CVR) or click-through rate (CTR) (Xu et al. 2021; Su et al. 2020). During both phases, learning a good user representation is the key.

One key characteristic of user behaviors on OTPs that differentiate themselves from other e-commerce platforms is that user's behaviors are quite sparse, since travel is a low-frequency demand compared with shopping. The lack of rich user behaviors renders the recommender system ineffective at delivering high-quality recommendation results. Such cold start problem is one of the major challenges we face at online travel platforms.

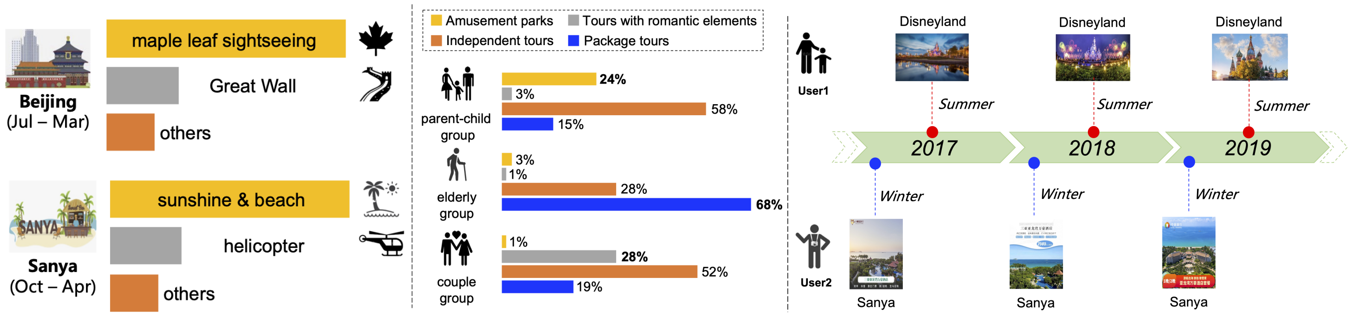
Recent advances in recommender systems have been focused on recommending for cold-start users (Silva et al. 2019; Wang et al. 2020; Lu, Fang, and Shi 2020; Chae et al. 2020). To alleviate this issue, the general idea is to leverage the *side information* to enrich the user representation, thereby improving the recommendation accuracy (Liang et al. 2020; Sedhain et al. 2017). To address the issue that users with similar profiles are recommended with same items, meta-learning has been proposed to address the cold-start problem by leveraging user's few behaviors, with optimization-based meta-learning being the most dominant (Dong et al. 2020; Lee et al. 2019).

Although much progress has been made, the state-of-the-arts for cold-start problems are largely sub-optimal for user representation at OTPs, as the user travel behaviors exhibit some unique characteristics that are not taken into account in these approaches.

C1 *spatial temporal interests:* users' travel behaviors demonstrate general preferences with respect to a specific destination during a given time period. For instance, as illustrated in Figure 1(a), most travellers in Beijing prefer to view maple leaves during July while most visitors in Sanya prefer to enjoy the beach in Oct. Such spatial temporal preferences can potentially complement the user representations, especially for users with few behaviors.

C2 *user group interests:* users' travel behaviors exhibit the herding phenomenon, i.e., people in the similar social status tend to make similar traveling choices. For instance, as shown in Figure 1(b), the parent-child group prefers to go to the amusement parks, while the elderly group tends to go with a package tour. Benefiting from the preference similarity among users in the same group, we could utilize such user group interest to enrich the user representations.

C3 *user periodic interests:* individual users traveling behaviors usually exhibit some periodic pattern. For example, as shown in Figure 1(c), some users continue to book



(a) spatial temporal interest

(b) user group interest

(c) user periodic interest

Figure 1: User behavior characteristics at one OTP. (Bar length in (a) (b) indicates popularity of vacation items)

hotels in Sanya during the Spring Festival to enjoy the beach there for years in a row. Such periodic patterns could span different time scales and could be leveraged to enhance the user representation from the user's historic behaviors.

With these in mind, we propose a state-aware multi-aspect interests representation network (SMINet in short), for cold-start users recommendation at OTPs, which consists of a multi-aspect interest extractor, a co-attention layer, and a state-aware gating layer. First, the key component in the proposed model is the multi-aspect interests extractor, which is able to extract representations for the above spatial temporal interests, user group interests, and user periodic interests. In addition, we propose to also extract representations for the *user long-term interest* that is able to activate user's travel experiences a few years back and *user short-term interest* to emphasize user's recent behaviors. Such multi-aspect interests extractor is built on top of a multi-aspect interest search unit (MAISU) to extract the item set relevant to the multi-aspect interests. Second, we design a co-attention layer between user's behaviors in the current session and the multi-aspect interests representations. This is because the current session behaviors and the multi-aspect interests can influence each other. For example, a frequent business traveler's current click behavior on a hotel might activate more of the user group interest; conversely, the user periodic interest might also influence the user's click behaviors. Such co-attention mechanism allows better interactions between the two modules. Third, we design a travel state-aware gating layer to attentively select the multi-aspect interests conditioned on the travel state (exact definition of travel state is defined in Table 1). The intuition is that, users could make different traveling decisions depending on the travel state. For instance, the user might not want to interact with an item from the city she just finishes visiting.

The main contributions are summarized as follows:

- To the best of our knowledge, this is the first work to address the cold-start users problem for OTPs, where traditional approaches are not generalizable due to the unique characteristics of user travel-related behaviors.
- We address the cold-start user problem by taking into account the characteristics of user travel-related behaviors and propose a novel state-aware multi-aspect interests representation network.

  The proposed approach can be generalized to other application domains, e.g., intelligent transportation, ride-sharing services, etc.
- Comprehensive experiments conducted both offline and online demonstrate the effectiveness of our proposed model. The model has been deployed online to serve real traffic of an OTP.

Related Work

Cold Start Recommendation

The common solution to address cold start problem is to utilize the side information, such as auxiliary and contextual information (Barjasteh et al. 2016; Li et al. 2019) and knowledge from other domains (Fu et al. 2019; Song et al. 2017; Hu et al. 2019; Wang et al. 2020). DecRec (Barjasteh et al. 2016) decouples the rating sub-matrix completion and knowledge transduction from ratings to exploit the side information. Some recent works take into account the user behaviors for the user representation. Based on the intuition that the users with similar preferences might have similar consumption habits, a zero-shot learning (ZSL) model is proposed to address the cold start issue (Li et al. 2019). The key component of this model is the low-rank linear auto-encoder that connects user behaviors via user attributes. Silva et al. (Silva et al. 2019) argue that integrating multiple non-personalized recommender systems could address the cold-start user problem effectively. To address the issue that users with similar profile are recommended with same items, meta-learning has been proposed to address the cold-start problem by leveraging user's few behaviors, with optimization-based meta-learning being the most dominant (Dong et al. 2020; Lee et al. 2019). Besides, based on the assumption that similar users in different domains might have similar preferences, transfer-learning based methods (Hu et al. 2019; Wang et al. 2020; Zhao et al. 2020) use cross-domain knowledge to mitigate cold-start problems in the target domain.

Session-Based Recommendation

Session-based recommendation focuses on modeling user's implicit feedback within the current user session. Liu et al. (Liu et al. 2016) consider the time intervals between adjacent interactions and external situation (i.e., time, location, weather) where the interactions happen as context information. This information is used to help RNN tune state transitions. However, RNN based approaches might suffer from exploding or vanishing gradients. GRU-based RNN is proposed to process multiple sessions in parallel in a mini-batch manner (Hidasi et al. 2015). It learns the common characterization of session behaviors from multiple sessions. Neural Attentive Recommendation Machine (NARM) (Li et al. 2017) conducts session representation through learning the interest evolution process in the session. It uses the attention module to learn the importance of different behavior entities in the session, and detects the global expression and local expression to better represent the session. STAMP (Liu et al. 2018) improves (Li et al. 2017) by separately introducing the user's last click as input. It focuses on strengthening the influence of the user's last behavior, thereby harnessing the user's immediate interest. SR-GNN (Wu et al. 2019) models all user session sequences as session graphs, and obtain the embeddings of the nodes in the session via gated graph neural networks.

Problem Definition

In this section, we formally define the user representation problem for the matching stage of the recommendation system at OTPs.

Given a user $u \in \mathcal{U}$, where $\mathcal{U}$ is the set of users, we want to retrieve a set of items from the item pool $\mathcal{I}$ that might be of interest to the user. Let p_u represent the basic user profile information, e.g., user id, gender, age, purchase level, etc, and x_i represent the information of an item $i \in \mathcal{I}$, e.g., item id, category id, city id, etc. For each of the user u, we can observe the user's behavior sequence ordered by time as $\mathcal{B}_u = \{x_1^u, x_2^u, \ldots, x_n^u\}$, where x_i^u indicates the i-th item interacted by the user u and the interaction could be click, purchase, etc. For cold-start users, the cardinality of $\mathcal{B}_u$, i.e., $|\mathcal{B}_u|$ could be very small. In addition, we also have context features denoted as c_u, e.g., location, time, etc.

The core task of this paper is to learn a mapping function to obtain user representations from the raw input features, which can be formulated as:

$$\mathbf{h}_u = f_{\mathcal{U}}(p_u, \mathcal{B}_u, c_u, \mathcal{U}). \tag{1}$$

Note that for a particular user u, $\mathbf{h}_u$ is also a function of $\mathcal{U}$, since to learn a good representation for cold-start users, we might also leverage other users' engagements.

Besides, the representation of an item i, $\mathbf{x}_i$ can be obtained via an embedding layer as $\mathbf{x}_i = \mathcal{E}(x_i)$. For each of the categorical features of an item, e.g., category id, we can get its embedding vector via the embedding layer, and the representation of the item is the concatenation of these embedding vectors. With the user representation $\mathbf{h}_u$ and the item representation $\mathbf{x}_i$, we can retrieve the top items to recommend to the user via the following scoring function:

$$f_{score} = \mathbf{h}_u^T \mathbf{x}_i, \tag{2}$$

which is an inner product between the two vectors. With the above notations, we formally define the USER REPRESENTATION LEARNING FOR MATCHING problem as follows:

PROBLEM 1. USER REPRESENTATION LEARNING FOR MATCHING

Given: *the basic user profile information p_u, the user's behavior sequence $\mathcal{B}_u$ ordered by time, for $u \in \mathcal{U}$, the context features c_u;*

Output: *the representations $\mathbf{h}_u$ for each $u \in \mathcal{U}$.*

Proposed Method

In this section, we present our state-aware multi-aspect interests extraction model, named SMINet, for cold-start users recommendation at OTPs.

The overall architecture of our model is presented in Fig. 3. The model consists of a multi-aspect interest extractor, a co-attention layer, and a state-aware gating layer. The key component of the model is the multi-aspect interests extractor, which is designed to alleviate the cold-start user issue. The interests extractor produces five different vectors that describe different aspects of the user's traveling preferences, including spatial temporal interest, user group interest, user periodic interest, long-term interest and short-term interest. Such interest extractor is built on top of a multi-aspect interest search unit (MAISU) to extract the relevant item set for the ensuing interests representation. Since a user's behavior in the current session and her multi-aspect interests are not independent of each other, we design a co-attention layer to learn the cross-attentions between the two modules. In addition, depending on the travel states the user is in (e.g., pre-travel/post-travel, etc), the user can make different decisions. To learn a better representation for the user conditioned on the travel states, we design a travel state-aware gating layer to attentively select the multi-aspect interests. The final user representation is further interacted with the item embedding to optimize the click-through rate. We detail each of the components in the following subsections.

Multi-Aspect Interest Search Unit

The ensuing multi-aspect interests extractor is based on a Multi-Aspect Interest Search Unit (MAISU) that extracts relevant item set for the purpose of interests representation. As illustrated in Figure 2, given all users' behaviors as $\mathcal{B} = \{\mathcal{B}_u | u \in \mathcal{U}\}$, and a query, MAISU would return the top-k relevant items that matches the query from all the behaviors. For instance, the query could be a tuple of context location and time, then MAISU would return the top-k items

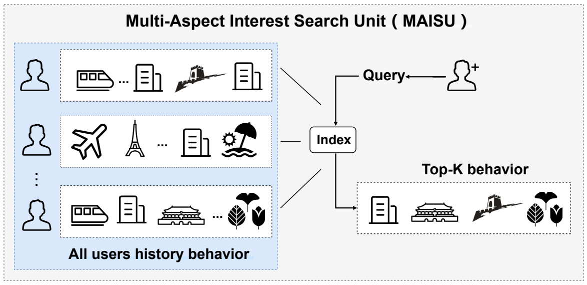

Multi-Aspect Interest Search Unit (MAISU)
Query
Index
Top-K behavior
All users history behavior

Figure 2: Multi-Aspect Interest Search Unit.

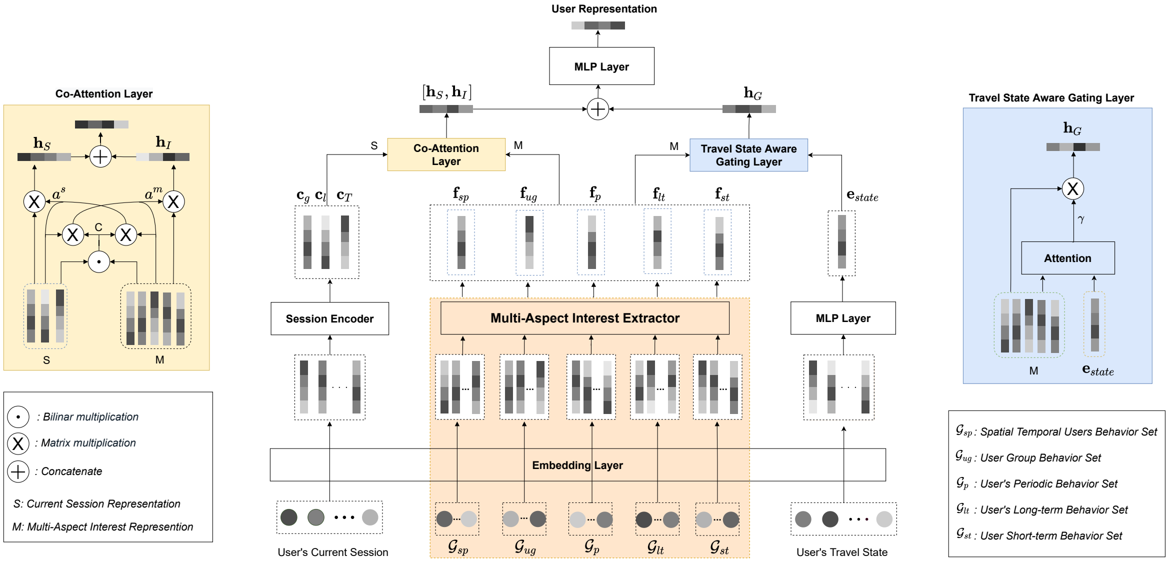

User Representation
MLP Layer
$[\mathbf{h}_S, \mathbf{h}_I]$
$\mathbf{h}_G$
Co-Attention Layer
$\mathbf{h}_S$
$\mathbf{h}_I$
a^s
a^m
C
S
M
Co-Attention Layer
Travel State Aware Gating Layer
$\mathbf{c}_g$ $\mathbf{c}_l$ $\mathbf{c}_T$
$\mathbf{f}_{sp}$
$\mathbf{f}_{ug}$
$\mathbf{f}_p$
$\mathbf{f}_{lt}$
$\mathbf{f}_{st}$
$\mathbf{e}_{state}$
Session Encoder
Multi-Aspect Interest Extractor
MLP Layer
Embedding Layer
User's Current Session
$\mathcal{G}_{sp}$ $\mathcal{G}_{ug}$ $\mathcal{G}_p$ $\mathcal{G}_{lt}$ $\mathcal{G}_{st}$
User's Travel State
Travel State Aware Gating Layer
$\mathbf{h}_G$
γ
Attention
M
$\mathbf{e}_{state}$
⊙ : Bilinear multiplication
⊗ : Matrix multiplication
⊕ : Concatenate
S: Current Session Representation
M: Multi-Aspect Interest Represention
$\mathcal{G}_{sp}$: Spatial Temporal Users Behavior Set
$\mathcal{G}_{ug}$: User Group Behavior Set
$\mathcal{G}_p$: User's Periodic Behavior Set
$\mathcal{G}_{lt}$: User's Long-term Behavior Set
$\mathcal{G}_{st}$: User Short-term Behavior Set

Figure 3: Overall architecture of the proposed SMINet model.

in the location and time, as

$$\{x_1, x_2, \ldots, x_k\} = \mathsf{MAISU}((c_u.city, c_u.time))$$
$$= \{x_i | x_i.city = c_u.city \& x_i.behaviorTime = c_u.time, x_i \in \mathcal{B}\}.$$

In this way, we can extract top-k items in Hangzhou during March as $\mathsf{MAISU}((Hangzhou, March))$. Next, we detail how to extract multi-aspect interest with MAISU.

Multi-Aspect Interests Extraction

To address the cold-start users on OTPs, we propose to extract users' multi-aspect interests that describe the users' traveling preferences from multiple aspects. We introduce how to extract each of the multi-aspect interests as follows.

Spatial Temporal Interest. Users' travel decisions are heavily influenced by the traveling trend in their cities during a specific season. For example, visitors in Hangzhou tend to view cherry blossoms in March, while travelers in Salt Lake City tend to go snowboarding during winter. We propose to supplement the user's representation by extracting such spatial temporal interest. Specifically, given the context location (e.g., Hangzhou) and the context time (e.g., March), we use MAISU to obtain the top-k items in the location and time as $\mathcal{G}_{sp} = \mathsf{MAISU}((c_u.city, c_u.time))$. The embeddings of these k items are denoted as $\mathbf{X}^u_{sp} = \mathcal{E}(\mathcal{G}_{sp}) = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k]$. We feed them to a multi-head self-attention layer (Vaswani et al. 2017) to capture the self-interactions among these items and obtain the output as

$$\begin{aligned} \mathbf{F}_{sp} &= \mathbf{MultiHead}(\mathbf{X}^u_{sp}) \\ &= concat(head_1, head_2, \ldots, head_h)\mathbf{W}^H, \end{aligned}$$

where in the multi-head self-attention layer, the output from each head is concatenated followed by a linear projection with projection matrix $\mathbf{W}^H$, and the output from each head is through a self-attention layer as:

$$head_i = Attention(\mathbf{X}^u\mathbf{W}^Q, \mathbf{X}^u\mathbf{W}^K, \mathbf{X}^u\mathbf{W}^V),$$
$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})\mathbf{V}$$

where $\mathbf{W}^Q$, $\mathbf{W}^K$ and $\mathbf{W}^V$ are the projection matrices for the query, keys and values, respectively. Finally, we perform an aggregation operation on the self-attended embeddings of the items in the set and extract the spatial temporal interest as $\mathbf{f}_{sp} = Agg(\mathbf{F}_{sp})$, is an aggregation operation, which could be average pooling or max pooling.

User Group Interest. We categorize users into different user groups according to their basic profile and historical behaviors. For example, business group travel frequently for business and thus have a lot of orders for flight/train tickets and hotels on the platform; while the parent-child crowd prefer to go to the amusement parks or the museums and thus purchase the tickets of those places on the platform. User grouping can be done by first labeling some seed users and then train a supervised model to categorize users. To extract the user group interest, we obtain the set of the top k popular items sold among a given user group as $\mathcal{G}_{ug} = \mathsf{MAISU}(u.userGroup)$. Similar to spatial temporal interest extraction, we perform a multi-head self-attention mechanism followed by an aggregation on the set of items and obtain the final user group interest representation as $\mathbf{f}_{ug} = Agg(\mathbf{MultiHead}(\mathcal{E}(\mathcal{G}_{ug})))$.

User Periodic Interest. Users' traveling behaviors usually present a cyclic pattern, e.g., some users travel to Sanya, the southernmost city on Hainan Island, in the winter every year to enjoy the warm weather there, or some people go hiking every weekend. Such periodic pattern presents itself at different time granularities, at weekly level, monthly level, or yearly level. To represent the user periodic interest, we extract the set of top k items preferred by the user at different time granularities. For example, to represent the user's monthly periodic interest, we can obtain the user's top preferred items during a given month (e.g., Februaries) in recent years, denoted as $\mathcal{G}_p = \mathsf{MAISU}((u.userID, c_u.month))$. Afterwards, the multi-head self-attention and aggregation layers are applied on top to obtain the final user periodic interest representation as $\mathbf{f}_p = Agg(\mathbf{MultiHead}(\mathcal{E}(\mathcal{G}_p)))$.

User Long-Term Interest. User long-term interest extraction intends to extract users interests from their life-long behavioral sequences. This is useful for recommending items that may resonate with the user's travel experiences a few years back. We obtain the set of top k items preferred by the user considering the user's life-long behavioral sequences as $\mathcal{G}_{lt} = \mathsf{MAISU}(u.userID)$. The final user long-term interest is obtained using the the multi-head self-attention and aggregation layers as $\mathbf{f}_{lt} = Agg(\mathbf{MultiHead}(\mathcal{E}(\mathcal{G}_{lt})))$.

User Short-Term Interest. Different from user long-term interest, user short-term interest extraction aims to emphasize the user's recent behaviors. We extract the set of top k items preferred by the user from the most recent sessions (excluding the current session), denoted as $\mathcal{G}_{st} = \{x_i | x_i.behaviorTime > T, x_i \in \mathcal{B}_u\}$, where T is a time threshold picked manually. The multi-head self-attention and aggregation layers are applied to get the final short-term interest representation $\mathbf{f}_{st} = Agg(\mathbf{MultiHead}(\mathcal{E}(\mathcal{G}_{st})))$.

Co-Attention with Current Session Behaviors

User's behavior in the current session and her multi-aspect interests are not independent of each other, but can affect each other. For example, a frequent business traveler's current click behavior on a hotel might activate more of her user group interest; conversely, her user periodic interest might also influence her click behaviors. With this intuition in mind, we design a co-attention layer between the current session behaviors and the multi-aspect interests.

Current Session Representation To represent the user's behavioral sequence in the current session, we adopt the hybrid encoder with two different attention mechanisms similar to the Neural Attentive Recommendation Machine (NARM) (Li et al. 2017). This encoder outputs three vectors, namely, the global session vector $\mathbf{c}_g$, the local session vector $\mathbf{c}_l$ and the time-aware session vector $\mathbf{c}_T$.

We follow the steps in NARM to produce $\mathbf{c}_g$ and $\mathbf{c}_l$ (the details are omitted for brevity and included in appendix) and additionally propose a time-aware attention mechanism to encode the current session. The intuition is that items engaged recently have more influence on the next decision than the items engaged a long time ago. We denote the behaviors in the current session as $\{x_{i1}, x_{i2}, \ldots, x_{it}\}$ and feed the behavior sequence into a recurrent neural network (RNN) with Gated Recurrent Units (GRU) (Chung et al. 2014) and get the hidden states at different time steps as $\mathbf{h}_i^T$. In order to capture the time-aware attention, we represent the time-aware session vector $\mathbf{c}_T$ as $\mathbf{c}_T = \sum_{i=1}^{t} \beta_{it} \mathbf{h}_i^T$, where β_{it} is the time-aware attention assigned to the i-th item as

$$\beta_{it} = f(T_i, T_t) = \frac{1/\sigma(log(T_t - T_i + 1))}{\sum_{k=1}^{t} 1/(\sigma(log(T_t - T_k + 1)))} \tag{3}$$

where T_i is the timestamp that the user interacts with the i-th item. Intuitively, the items that the user interacts more recently get assigned more weights.

Co-Attention Layer User's behavior in the current session and her multi-aspect interests are not independent of each other, but can affect each other. For example, the user's spatial temporal interests, namely, the traveling trend in the user's city can affect what the user is likely to click in the current session. Conversely, the engagement behaviors in the user's session also impact her short term interests. Inspired by the co-attention mechanism for visual question answering tasks (Lu et al. 2016), we propose to jointly attend to the session representations as well as the multi-aspect interests representations. Given the session representations $\mathbf{S}$ as $\mathbf{S} = [\mathbf{c}_g, \mathbf{c}_l, \mathbf{c}_T]$ and the multi-aspect interests representations $\mathbf{M}$ as $\mathbf{M} = [\mathbf{f}_{sp}, \mathbf{f}_{ug}, \mathbf{f}_p, \mathbf{f}_{lt}, \mathbf{f}_{st}]$, the affinity matrix between $\mathbf{S}$ and $\mathbf{M}$ can be calculated as $\mathbf{C} = tanh(\mathbf{S}\mathbf{W}_b\mathbf{M})$, where $\mathbf{W}_b$ is the learnable weight matrix. With the affinity matrix as a feature, we can compute the attention maps for both the session as well as the multi-aspect interests as:

$$\mathbf{H}^s = tanh(\mathbf{W}_s\mathbf{S} + (\mathbf{W}_m\mathbf{M})\mathbf{C})$$
$$\mathbf{H}^m = tanh(\mathbf{W}_m\mathbf{M} + (\mathbf{W}_s\mathbf{S})\mathbf{C}^T)$$
$$\mathbf{a}^s = softmax(\mathbf{w}_{hs}^T\mathbf{H}^s), \quad \mathbf{a}^m = softmax(\mathbf{w}_{hm}^T\mathbf{H}^m)$$

where $\mathbf{a}^s$ and $\mathbf{a}^m$ are the attention probabilities of each session vector and each multi-aspect interest vector, and $\mathbf{W}_s$, $\mathbf{W}_m$, $\mathbf{w}_{hs}$ and $\mathbf{w}_{hm}$ are the learnable weight parameters. With the attention weights, we can compute the session vector as well as the multi-aspect interest vector after co-attention as follows:

$$\mathbf{h}_S = (\mathbf{a}^s)^T\mathbf{S}, \quad \mathbf{h}_I = (\mathbf{a}^m)^T\mathbf{M}$$

where $\mathbf{h}_S$ is the co-attended session representation and $\mathbf{h}_I$ is the co-attended multi-aspect interests representation.

Travel State-Aware Gating Layer

Users could be in multiple states, e.g., depending on the behaviors, the user could be in the state of browsing, decision-making, focus, etc; depending on the user's LBS information, the user could be in the state of pre-travel, on-travel or post-travel. We define the user states in Table 1. Depending on the travel states the user is in, the user can make different decisions. For instance, if the user has completed the travel, she might not want to interact with the items from the city she has just visited. As a result, the multi-aspect interests might contribute differently to the user's decision conditioned on the travel state. We propose to apply a travel state-aware gating layer to implement such intuition. This gating layer can learn to select the important multi-aspect interests conditioned on the user's travel state as $\mathbf{h}_G = \sum_{i=1}^{5} \gamma_i \mathbf{M}_i$,

where γ_i is the weight assigned to the i-th multi-aspect interest vector and is computed via a softmax layer as:

$$\gamma = softmax(\mathbf{W}_g\mathbf{e}_{state}) \tag{4}$$

where $\mathbf{e}_{state} = MLP(\mathcal{E}(s))$ is the representation for the user's multiple states, s is the binary encoding of the user's states, and $\mathbf{W}_g$ is a trainable weight matrix.

State Category	State Value	Definitions
Behaviors	Silence	The user has no behavior in the past one year.
	Browsing	The user has few behaviors in the last 30 days.
	Decision -making	The user has rich behaviors over items from diverse cities.
	Focus	The user has rich behaviors, and focused on items from one city.
Location	Pre-travel	The user has placed an order but has not started the trip.
	On-travel	The user has placed an order, and is traveling in the destination.
	Post-travel	The user has completed the trip and returned to the resident city.

Table 1: Definition of User Travel States.

Final MLP Layer

To obtain the final representation for the user, we concatenate the following three intermediate vectors, namely, $\mathbf{h}_S$, the co-attended current session representation; $\mathbf{h}_I$, the co-attended multi-aspect interests representation; and $\mathbf{h}_G$, the travel-state aware multi-aspect interests representation. To learn the interactions of these vectors, we apply a multi-layer perceptron (MLP) layer to generate the final user representation $\mathbf{h}_u$ as $\mathbf{h}_u = MLP(Concat(\mathbf{h}_S, \mathbf{h}_I, \mathbf{h}_G))$. This final user representation is used to compute the user's probability of clicking on an item along with the item embeddings.

Model Training

The collection of training data $\mathcal{D}$ usually consists of users' engagement logs as $\{(u, i, y_{ui}, t)\}$, which indicates the interaction of user u with item i at time t and $y_{ui} \in \{0, 1\}$ indicates whether the user has clicked the item. To train the model, we minimize the following cross-entropy loss as:

$$\mathcal{L} = \sum_{(u,i)\in\mathcal{D}} -(y_{ui}log(p_{ui}) + (1 - y_{ui})log(1 - p_{ui})),$$

where p_{ui} is the predicted click probability of user u on item i and is computed as $p_{ui} = \sigma(\mathbf{h}_u^T \mathbf{x}_i)$,
where $\sigma(\cdot)$ is the sigmoid function.

Experiments

Experimental Settings

Datasets We use two datasets[1]. (1) **Fliggy:** our proprietary dataset extracted from user's behavior logs at Fliggy, one of the largest OTP in China. The clicked samples are labeled as positive and those impressed but unclicked samples are treated as negative. The dataset is further split into training set, test set and validation set. (2) **Foursquare:** a public dataset that contains check-in data of a user at a particular location at a specific timestamp, along with attribute information of users and locations. The statistics of the datasets are in Table 2.

Datasets	#samples	#users	#items	#cities
Fliggy	224M	5.74M	0.26M	341
Foursquare	33M	0.27M	3.68M	415

Table 2: Statistics of the dataset.

Evaluation Metrics In the experiments, two metrics, i.e., MRR@k and Recall@k are adopted to measure the recommendation performance of different methods, which are also widely used in other related works.

Comparison Methods We compare SMINet with the following models: (1) *POP*, a statistical model that recommends the most popular items; (2) *GRU4Rec* (Hidasi et al. 2015), *NARM* (Li et al. 2017), *STAMP* (Liu et al. 2018) and *JNN* (Guo et al. 2020), *SASRec* (Kang and McAuley 2018),which utilize RNN based approaches to model users' sequential behaviors; (3) *NETA* (Lv, Zhuang, and Luo 2019), *HERS* (Hu et al. 2019) and *LHRM* (Wang et al. 2020) that leverage additional side information, e.g., session neighborhood, user LBS information and cross-domain information; (4) *SR-GNN* (Wu et al. 2019), GNN based approach that learns user representation from user's behavior graph; (5) *DVN-V2* (Wang et al. 2021), a deep and cross network for learning feature interactions.

Performance Comparison

We compare the performances between our proposed model and the competitors in this section.

Evaluation of Comparison Methods Table 3 shows the results of all the comparison methods. In general, the proposed SMINet outperforms all the state-of-the-art methods. We have the following observations: (1) among the session-based methods, NETA achieves the best performance as it complements the current session using similar neighborhood sessions, which is effective on our dataset since user behaviors are sparse. (2) HERS and LHRM address the cold start issue by leveraging user social network relations or user behavior clustering to complement the user behaviors, leading to better performance than those not (e.g., STAMP). (3)

Methods	Foursquare		Fliggy	
	Recall@10	MRR@10	Recall@10	MRR@10
POP	0.0475	0.0128	0.0519	0.0158
GRU4Rec	0.4641	0.1192	0.5357	0.2043
NARM	0.5122	0.1654	0.6032	0.2607
STAMP	0.5379	0.1838	0.6184	0.2647
NETA	0.5402	0.1875	0.6303	0.2735
JNN	0.5480	0.1951	0.6235	0.2778
SR-GNN	0.5546	0.2089	0.6284	0.2758
HERS	0.5617	0.2170	0.6279	0.2746
LHRM	0.5748	0.2397	0.6301	0.2793
SASRec	0.5716	0.2365	0.6298	0.2753
DCN-V2	0.5815	0.2402	0.6303	0.2794
SMINet	**0.6083**	**0.2674**	**0.6582**	**0.2983**

Table 3: Performance comparison of different methods.

[1] The details for preprocessing the datasets along with the data and code are released at https://github.com/wanjietao/Fliggy-SMINet-AAAI2022

multi-aspect interests	Recall @10	MRR @10
all multi-aspect interests	0.6582	0.2983
remove spatial temporal interest	0.6493	0.2869
remove user group interest	0.6517	0.2924
remove user periodic interest	0.653	0.2931
remove user long-term interest	0.6509	0.2938
remove user short-term interest	0.6238	0.2723

Table 4: Ablation study results of multi-aspect interests.

The proposed model comprehensively considers all the additional information to complement the behaviors in the current session, e.g., spatial temporal interest and user group interest, which are especially helpful on our dataset.

Ablation Study of the Multi-Aspect Interests We conduct an ablation study on the multi-aspect interests, where we remove the multi-aspect interests one at a time to see how it affects the final performance. Table 4 shows the results of the ablation study on Fliggy. From the results, we can see that removing user short-term interest brings the most drop on the performance. This is because the user's next click behavior is more closely related to the user's most recent behaviors. Removing spatial temporal interest, user group interest, and user periodic interest all bring similar level of performance drop, which show the importance of modeling spatial temporal information, user group information and user's periodic information on the travel platform. In addition, removing the user long-term interest also has a negative effect.

Ablation Study of Different Modules We perform an ablation study on the two layers and show the results on Fliggy in Table 5. In the table, 'without co-attention layer' means that we concatenate the session representations $\mathbf{S}$ and multi-aspect interest representations $\mathbf{M}$ directly in the last layer without going through the co-attention layer, and 'without the travel state-aware gating layer' means we concatenate the co-attended session vector $\mathbf{h}_S$ and interest vector $\mathbf{h}_I$ with the travel state vector $\mathbf{e}_{state}$ without the state-aware interest vector $\mathbf{h}_G$. As can be seen from the table, removing either could lead to significant drop in performance. This shows that both the co-attention layer and the state-aware gating layer can bring additional gain, as they allow better interactions between the multi-aspect interests, the current session behaviors, and the travel state.

Attentions in the Travel State-Aware Gating Layer The attention weights in the gating layer reflects the importance of each of the multi-aspect interests depending on the user state. In Figure 4, we show the average attention weights of users with different behaviors. For those users with few behaviors, the spatial temporal interest and user group interest play an significantly more important roles than other interests to complement the user preferences which is difficult to extract from the few behaviors. For users who have placed orders and are traveling, the spatial temporal interests can guide the user to pick travel items that are popular in the local city during the current season.

Model	Recall@10	MRR@10
full method	0.6582	0.2983
without co-attention layer	0.6469	0.2894
without state-aware gating fusion	0.6481	0.2902

Table 5: Ablation study of different modules.

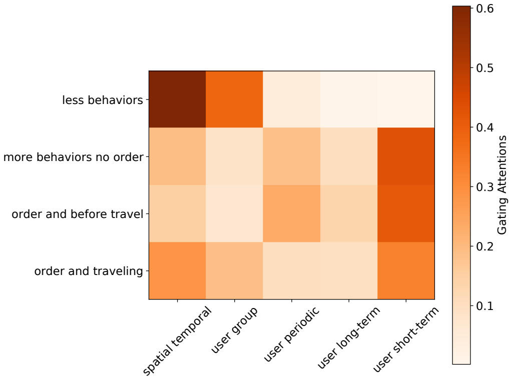


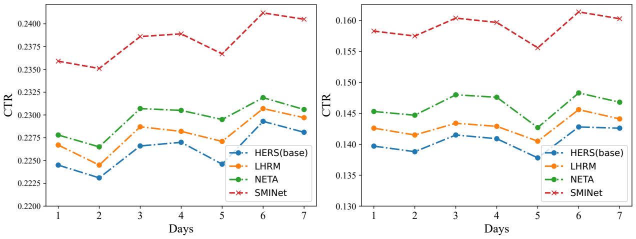
Figure 4: Attentions in the travel state-aware gating layer.

Online A/B Test

We conduct online experiments by deploying SMINet to handle real traffic in the personalized recommendation interface of an OTP for one week in Jan 2021. We deploy all comparison methods concurrently in the matching phase of the platform under an A/B test framework, and each method gets equal share of traffic. The click-through rate (i.e., CTR) is employed to evaluate the performance of online experiments and the results are shown in Figure 5. We can see that the proposed SMINet consistently outperforms the other methods, and for cold-start users, the improvement margin is even larger. This demonstrates the effectiveness of the multi-aspect interests in alleviating the cold-start user issue.

Conclusion

We propose a travel state-aware multi-aspect interests representation network (SMINet) for cold-start user recommendation at OTPs. To extract user's interests from multiple aspects, we design a multi-aspect interests extractor. We additionally design a co-attention layer and a travel state-aware gating layer to fuse the multi-aspect interests with the current session behaviors and user's travel state.



(a) CTRs of all users.

(b) CTRs of cold-start users.

Figure 5: Online CTRs of different methods in one week.

Acknowledgments

This research work is supported by The National Natural Science Foundation of China under Grant No.U20A20173.

References

Barjasteh, I.; Forsati, R.; Ross, D.; Esfahanian, A.; and Radha, H. 2016. Cold-Start Recommendation with Provable Guarantees: A Decoupled Approach. *IEEE Trans. Knowl. Data Eng.*, 28(6): 1462–1474.

Chae, D.-K.; Kim, J.; Chau, D. H.; and Kim, S.-W. 2020. AR-CF: Augmenting Virtual Users and Items in Collaborative Filtering for Addressing Cold-Start Problems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, 1251–1260. New York, NY, USA: Association for Computing Machinery. ISBN 9781450380164.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Dong, M.; Yuan, F.; Yao, L.; Xu, X.; and Zhu, L. 2020. MAMO: Memory-Augmented Meta-Optimization for Cold-Start Recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, KDD '20, 688–697. New York, NY, USA: Association for Computing Machinery. ISBN 9781450379984.

Fu, W.; Peng, Z.; Wang, S.; Xu, Y.; and Li, J. 2019. Deeply Fusing Reviews and Contents for Cold Start Users in Cross-Domain Recommendation Systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33: 94–101.

Guo, Y.; Zhang, D.; Ling, Y.; and Chen, H. 2020. A joint neural network for session-aware recommendation. *IEEE Access*, 8: 74205–74215.

Hidasi, B.; Karatzoglou, A.; Baltrunas, L.; and Tikk, D. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*.

Hu, L.; Jian, S.; Cao, L.; Gu, Z.; and Amirbekyan, A. 2019. HERS: Modeling Influential Contexts with Heterogeneous Relations for Sparse and Cold-Start Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33: 3830–3837.

Kang, W.-C.; and McAuley, J. 2018. Self-attentive sequential recommendation. In *2018 IEEE International Conference on Data Mining (ICDM)*, 197–206. IEEE.

Lee, H.; Im, J.; Jang, S.; Cho, H.; and Chung, S. 2019. MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, KDD '19, 1073–1082. New York, NY, USA: Association for Computing Machinery. ISBN 9781450362016.

Li, J.; Jing, M.; Lu, K.; Zhu, L.; and Huang, Z. 2019. From Zero-Shot Learning to Cold-Start Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33: 4189–4196.

Li, J.; Ren, P.; Chen, Z.; Ren, Z.; Lian, T.; and Ma, J. 2017. Neural attentive session-based recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1419–1428.

Liang, T.; Xia, C.; Yin, Y.; and Yu, P. S. 2020. Joint Training Capsule Network for Cold Start Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, 1769–1772. New York, NY, USA: Association for Computing Machinery. ISBN 9781450380164.

Liu, Q.; Wu, S.; Wang, D.; Li, Z.; and Wang, L. 2016. Context-aware sequential recommendation. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 1053–1058. IEEE.

Liu, Q.; Zeng, Y.; Mokhosi, R.; and Zhang, H. 2018. STAMP: short-term attention/memory priority model for session-based recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1831–1839.

Lu, J.; Yang, J.; Batra, D.; and Parikh, D. 2016. Hierarchical Question-Image Co-Attention for Visual Question Answering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, 289–297. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510838819.

Lu, Y.; Fang, Y.; and Shi, C. 2020. Meta-Learning on Heterogeneous Information Networks for Cold-Start Recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, KDD '20, 1563–1573. New York, NY, USA: Association for Computing Machinery. ISBN 9781450379984.

Lv, Y.; Zhuang, L.; and Luo, P. 2019. Neighborhood-Enhanced and Time-Aware Model for Session-based Recommendation. *arXiv preprint arXiv:1909.11252*.

Sedhain, S.; Menon, A.; Sanner, S.; Xie, L.; and Braziunas, D. 2017. Low-rank linear cold-start recommendation from social data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Silva, N.; Carvalho, D.; Pereira, A. C. M.; Mouro, F.; and Rocha, L. 2019. The Pure Cold-Start Problem: A deep study about how to conquer first-time users in recommendations domains. *Information Systems*, 80(FEB.): 1–12.

Song, T.; Peng, Z.; Wang, S.; Fu, W.; and Yu, P. S. 2017. Review-Based Cross-Domain Recommendation Through Joint Tensor Factorization. *Springer, Cham.*

Su, Y.; Zhang, L.; Dai, Q.; Zhang, B.; Yan, J.; Wang, D.; Bao, Y.; Xu, S.; He, Y.; and Yan, W. 2020. An Attention-based Model for Conversion Rate Prediction with Delayed Feedback via Post-click Calibration. In Bessiere, C., ed., *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, 3522–3528. International Joint Conferences on Artificial Intelligence Organization. Main track.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In *Advances in Neural Information*

Processing Systems 30: Annual Conference on Neural Information Processing Systems, 5998–6008.

Wang, R.; Shivanna, R.; Cheng, D.; Jain, S.; Lin, D.; Hong, L.; and Chi, E. 2021. DCN V2: Improved Deep amp; Cross Network and Practical Lessons for Web-Scale Learning to Rank Systems. WWW '21, 1785–1797. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383127.

Wang, Z.; Xiao, W.; Li, Y.; Chen, Z.; and Jiang, Z. 2020. LHRM: A LBS Based Heterogeneous Relations Model for User Cold Start Recommendation in Online Travel Platform. In *Neural Information Processing - 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 23-27, 2020, Proceedings, Part III*, 479–490.

Wu, S.; Tang, Y.; Zhu, Y.; Wang, L.; Xie, X.; and Tan, T. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 346–353.

Xu, J.; Wang, Z.; Chen, Z.; Lv, D.; Yu, Y.; and Xu, C. 2021. Itinerary-aware Personalized Deep Matching at Fliggy. In *WWW*.

Zhao, C.; Li, C.; Xiao, R.; Deng, H.; and Sun, A. 2020. CATN: Cross-Domain Recommendation for Cold-Start Users via Aspect Transfer Network. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, 229–238. New York, NY, USA: Association for Computing Machinery. ISBN 9781450380164.