Knowledge-Based Systems 211 (2021) 106511

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Neural Attentive Travel package Recommendation via exploiting long-term and short-term behaviors

Guixiang Zhu [a], Youquan Wang [b,*], Jie Cao [a,b], Zhan Bu [b], Shuxin Yang [c], Weichao Liang [a], Jingting Liu [b]

[a] College of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
[b] Jiangsu Provincial Key Laboratory of E-Business, Nanjing University of Finance and Economics, Nanjing, China
[c] School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, China

ARTICLE INFO

Article history:
Received 17 March 2020
Received in revised form 2 October 2020
Accepted 6 October 2020
Available online 23 October 2020

Keywords:
Recommender systems
Travel recommendations
Sequential behaviors
Neural networks
Personalized attention

ABSTRACT

Travel package recommendation is a critical task in the tourism e-commerce recommender systems. Recently, an increasing number of studies proposed various travel package recommendation algorithms to improve Online Travel Agencies (OTAs) service, such as collaborative filtering-based, matrix factorization-based and neural network-based methods. Despite their value, however, the main challenges that incorporating complex descriptive information of the travel packages and capturing complicated users' long-term preferences for fine-grained travel package recommendation are still not fully resolved. In terms of these issues, this paper propose a novel model named Neural Attentive Travel package Recommendation (NATR) for tourism e-commerce by combining users' long-term preferences with short-term preferences. Specifically, NATR mainly contains two core modules, namely, *travel package encoder* and *user encoder*. The *travel package encoder* module is developed to learn a unified travel package representation by an attentive multi-view learning approach including word-level and view-level attention mechanisms. The *user encoder* module is designed to study long-term and short-term preference of the user by Bidirectional Long Short-Term Memory (Bi-LSTM) neural networks with package-level attention mechanism. In addition, we further adopt a gated fusion approach to coalesce these two kinds of preferences for learning high-quality the user's representation. Extensive experiments are conducted on a real-life tourism e-commerce dataset, the results demonstrate the proposed model yields significant performance advantages over several competitive methods. Further analyses from different attention weights provide insights of attentive multi-view learning and gated fusion network, respectively.

1. Introduction

Tourism is one of the largest and fastest growing service industries in the world. Report in UNWTO[1] shows that tourism is an important industry to the world economy and contributes more than 1.34 trillion U.S. dollars in receipts. Moreover, the overall international tourist arrivals grew by 7 percent to 1.33 billion in 2017, the largest rise since the 2009 global economic crisis. It is also noted that Chinese tourists spent 258 billion U.S. dollars on international tourism in 2017, almost one fifth of the world's total tourism spending. As a result, a large amount of travel data has been readily available to the OTAs. However, it is hard for tourists to find the most suitable travel information quickly from millions of travel information candidates [1]. Therefore, travel package recommendations are proposed to increase scales and serve the tourists with a highly personalized manner [2,3].

The term "travel package" in the clickstream usually refers to an integrated package containing a set of necessary travel-related ingredients [4]. For example, Table 1 lists three typical travel packages used in this research and each travel package is mainly made up of four kinds of attributes. Specifically, the title attribute is much longer and more detailed than the others, while the destination and the category attributes (i.e., travel region and travel type) are usually labeled with a few words.

In the literatures, a large number of personalized recommendation methods have sprang up for travel package recommendation over the past several years [4–9]. These methods can effectively capture the non-linear relationship between users and travel packages. Since the interaction matrix between users and travel packages data is much sparser than that of traditional e-commerce data [4], several previous works [6,9] attempt to incorporate more auxiliary information to alleviate the cold-start

* Corresponding author.
E-mail address: youq.wang@gmail.com (Y. Wang).

1 http://www2.unwto.org/

https://doi.org/10.1016/j.knosys.2020.106511
0950-7051/

Table 1
Three typical travel packages.

Title	Destination	Travel region	Travel type
Maldives Kurumba 4 nights *6 days Hong Kong Airlines deluxe room* sale	*Maldives*	*Short-term travel abroad*	*DIY tour*
Lijiang Shangri-La 6 days tour Kunming to Lijiang *train* local tour	*Yunnan*	*Domestic local apply*	*Local apply tour*
Suzhou Nanjing Hangzhou 3 days tour explore *ancient capital*	*Eastern China*	*Travel nearby*	*Group tour*

Note: Important words are showed in italic type.

problem. Recently, neural-based methods such as neural collaborative filtering and recurrent neural networks (RNNs) have become increasingly popular on recommendation domain [10–13]. These neural-based methods can effectively capture the user preferences and item attributes compared with traditional algorithms. Unfortunately, due to the difficulty of modeling multifaceted tourism contexts and dynamic user preferences, few works focus on investigating issues of deep learning-based travel package recommendation.

There are three key challenges for modeling travel package recommendation, which motivate us to build a deep learning-based recommendation model for tourism e-commence via exploiting users' long-term preferences with short-term preferences.

Firstly, unlike traditional e-commerce recommendation systems, the description information of travel package is complicated, making difficulties for researchers to directly incorporate it in existing recommendation systems. Most of the previously proposed methods [8,14] build topic models or sentiment models to learn package classification based on various package information. Intuitively, there are significant differences in the degree of travel entity words attention for different travel information. Consider the former example in Table 1, "Maldives" contains more information than "Kurumba" in first travel package. Meanwhile, the tourists may pay different attention to different package information in different browsing stages. For instance, if a user browses an extremely popular travel package of "Eastern China" and a less popular travel package of "Maldives", then the latter may be more informative for inferring his/her preference than the former.

Secondly, the preferences of users may dynamic change over time. However, most of previous works [15–17] only model static user–item interactions [18]. Intuitively, due to the streaming nature of users' behaviors, their long-term and short-term preferences from the historical and current clickstream can be dynamically captured. For example, Fig. 1 illustrates a typical scenario of tourism e-commerce, where the user's current clickstream implies this user might be interested in a short tour to Nanjing. If we adopt a session-based model which only considers current behaviors, another similar travel package about Nanjing may be recommended as shown in the red chart. By contrast, in the long-term view, the user's historical clickstream implies that this user might be interested in a short tour to Shanghai. Similarly, for one session from a user's historical or current clickstream, the preferences of this user can be also dynamically captured as the length of his/her session increases. For example, a user browsed two travel packages about "Sanya" in the first two clicks, which can reveal that he/she may interested in tour to "Sanya". While, the user browsed several travel packages about "Phuket Island" in the rest clicks of the session, which indicated that the preferences of users may dynamically transform from "Sanya" to "Phuket Island".

Thirdly, only simply coalescing users' short-term with long-term preferences is insufficient to achieve satisfactory recommendation results. For example, as shown in the green chart of Fig. 1, if we simply concatenate long-term and short-term clickstream as the input of the general CF [15] or MF [16] methods, another similar or popular travel packages about "Nanjing" and "Shanghai" would be recommended. In fact, the user may plan a travel route which contains a series of nearby cities implied by the historical and current clickstream as shown in the blue chart of Fig. 1. Thus, how to perfectly integrate long-term preferences with short-term preferences remains to be explored.

To address the above three challenges, we propose a novel model named **N**eural **A**ttentive **T**ravel package **R**ecommendation (NATR) for OTAs. NATR mainly contains two core modules, namely, *travel package encoder* and *user encoder*. Specifically, in the *travel package encoder* module, we learn a unified travel package representation from its attributes. And, in the *user encoder* module, we learn the dynamic evolution of the user's preference with Recurrent Neural Networks (RNN) models. In addition, we develop a gated fusion approach to coalesce the user's long-term and short-term preferences for learning high-quality user representations. It is also worth noting that the real-life travel dataset used in this paper has a lot of potential to help OTAs improve services, attract and retain customers, and eventually increase conversions from browsers to buyers. In summary, the threefold contributions of this paper are summarized as follows:

- To accurately learn travel packages representations from its four kinds of attributes, in the *travel package encoder* module, we propose an attentive multi-view learning approach to learn the unified travel package representations from the different attributes by incorporating them as different views. Different from previous methods which incorporate special auxiliary information to improve recommendation accuracy, the word-level and view-level attention mechanisms in such module are used to effectively select important and informative words and views, respectively.
- To accurately capture the dynamic evolution in whole sequential behaviors of users, in the *user encoder* module, we develop two Bi-directional Long Short-Term Memory (Bi-LSTM) neural networks with a package-level attention mechanism to dynamically learn users' long-term and short-term preferences from their historical and current clickstream, respectively. Different from the traditional session-based recommendation methods, the Bi-LSTM neural network with an attention mechanism can select important packages to precisely represent users' preferences from their sequential clickstream.
- To effectively capture correlation of users' long-term and short-term preferences as well as users' attention to them, we further apply a gated fusion network which incorporates their correlation information rather than simple combinations. Unlike the scalar weight of attention-like networks, the gated vector has more powerful representation ability to control the importance of these two kinds of preferences.

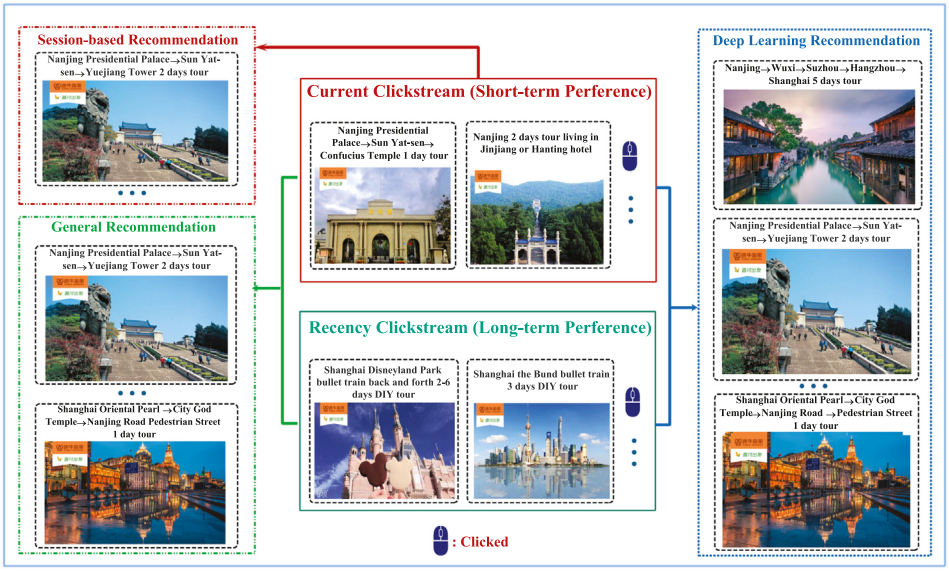


Fig. 1. Recommending by mergeing long-term behaviors and short-term behaviors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Related work

In this section, we detail relevant previous work on tourism-oriented recommendations and neural-based recommender systems.

2.1. Tourism-oriented recommendations

Recent studies in tourism-oriented Recommendations can be categorized as Point of Interest (POI)/travel route and travel product/package recommendation methods. POI/travel route recommendation methods focus on understanding users' travel behaviors by mining the human mobility data in daily life and predicting the next location he/she may visit or to generate itinerary under trip constraints. Examples of tourism-oriented Recommendation models include the stochastic approach [19], support vector machine-based models [20–22], topic models [8, 23]. However, users' preferences may vary dramatically with respect to the geographical regions due to different urban compositions and cultures. Therefore, some recent studies tried to leverage deep learning methods to address these challenges [24, 25]. For instance, Yin et al. [24] developed a novel POI recommendation model to jointly perform deep representation learning for POIs from heterogeneous features and hierarchically additive representation learning for spatial-aware personal preferences.

Although a wide array of studies fall within the aforementioned field, this work is highly-related to the second sub-stream: travel product or package recommendation. Since some existing studies [4,5,7–9,26] have shown that travel data has some unique characteristics, such as extremely sparse, involving a large number of cold-start users or unavailable ratings, traditional recommendation methods, such as collaborative filtering (CF) [15] or matrix factorization (MF) [16] are not applicable to travel data directly. To this end, latent factor models like matrix factorization models [5,6,9] with auxiliary content information have been widely proposed to alleviate the above problems. Essentially, the above methods are all content-based recommendations which rely on CF [15] or MF [16]. Thus, some problems of travel recommendation still need to be solved, such as sequential behaviors. Along this line, Zhu et al. [8] designed a novel recommendation engine for travel products based on topic sequential patterns. However, previous studies on this regard focused on exploiting some specific type of factor to improve the recommendation quality and little attention has been paid to design a systematic and flexible framework to incorporate all-round knowledge for travel recommendation. Furthermore, they primarily concerned about modeling static user–item interactions, and it cannot well capture the dynamic evolution in users' whole sequential behaviors [27]. To the best of our knowledge, this work is the first one which investigates personalized travel package recommendation via exploiting current and historical clickstream provided from an Online Travel Agency (OTA) platform.

2.2. Neural-based recommender systems

Considering the sparsity of interaction matrix between users and items, neural-based recommender systems aims at combining classical CF and MF with deep neural networks to learn the relationships between users and items [28–33]. For example, Wang et al. [29] used a hierarchical representation model to capture both sequential behaviors and users' general tastes by involving transaction and user representations in next-basket prediction. Guo et al. [31] employed a DeepFM approach to combine a component of factorization machines and a component of deep neural networks that shared the input. However, most of these methods provide recommendations by mining the static relevancy between users and items. The dynamic and evolution of users' preferences as well as their present consumption motivations are not be taken seriously enough. In addition, those methods cannot effectively capture the important contexts of users and item representations.

To address aforementioned issues, some studies tried to use various sequential recommendation scenarios via exploiting the clickstream of users' online behaviors, such as session-based [34, 35], next-basket [36] and next-item recommendations [27,37]. For instance, Quadrana et al. [35] proposed a seamless way to personalize Recurrent Neural Networks (RNN) models with cross-session information transfer and devise a Hierarchical RNN (HRNN) model to learn latent hidden states of the RNNs across user sessions. Although these models have taken users' sequential information into consideration, issues related to the coherence of customers' sequential behaviors and the dynamics of historical preferences are still far from being fully explored. Along this line, Li et al. [37] proposed a BINN model by concatenating

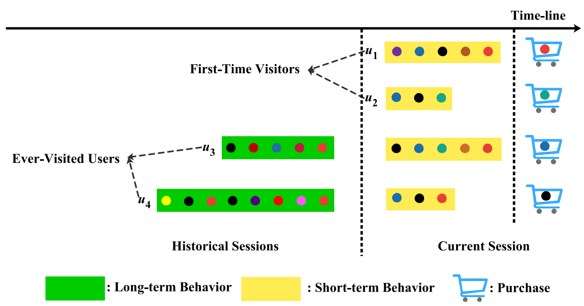


Fig. 2. Long-term and short-term behaviors with a cut time line.

users' session behavior representations and stable preferences of historical purchasing behaviors. However, these models only exploit a single kind of traditional products information provided by e-retailers (e.g., Taobao[2] and Jingdong[3]), which is not suitable to learn the travel packages representations containing complicated and various information. In addition, these methods neglect different products browsed or purchased by the same user may have different informativeness for learning user representations, which limits recommendation performances of neural-based recommendation methods.

3. The proposed approach

In this section, we introduce the proposed NATR model for personalized travel package recommendation. First, we give the problem definition and present the overall framework of NATR in Section 3.1. Then, we describe two core components of NATR in details, i.e., *travel package encoder* module in Section 3.2 and *user encoder* module in Section 3.3, respectively. Finally, we provide the model training and loss function in Section 3.4.

3.1. Problem formulation

The proposed NATR model is a task of predicting what a tourist would likely to purchase next based on his/her long-term and short-term behaviors. Here, some fundamental definitions and a formulation for the NATR model is defined as following.

Definition 1 (*Travel Package x_j*)**.** The travel package x_j usually refer to an integrated package containing a set of necessary travel-related ingredients, which can be described as a triple: $x_j = \langle \texttt{Title}, \texttt{Destination}, \texttt{Categories} \rangle$.

`Title` is a brief text description of the travel package including the countries/cities, scenic spots, hotels, vehicles, and number of days, etc. `Destination` is labeled with one or few words, which is usually represented as a city or a country name. `Categories` is a description of type from two angles (i.e., travel region and travel type). In detail, the first category of travel package is according to the travel region and thus each package is described as local/around tour, domestic short/long haul and oversea short/long haul. Another category of travel package addresses the travel type including Tuniu special tour, package tour, self-driving tour, self-guided tour, company's package tour, and local-attended tour.

Definition 2 (*Session S^u*)**.** Session S^u is a clickstream of travel packages views for a single user u, which refers to the sequential travel package browsed/purchased by user u during a certain period.

Here, session is automatically identified by a `SessionID` of clickstream provided by Tuniu. Let $\mathcal{U}$ and $\mathcal{X}$ denote the set of users and travel packages, respectively. For user $u \in \mathcal{U}$, his/her interaction sequence $\mathcal{S}^u = \{S_1^u, S_2^u, \dots, S_N^u\}$ with the ascending order of time can be obtained, where S_n^u denotes the n-th session of user u, and $N = |\mathcal{S}^u|$ is the number of sessions of user u. Each session S_n^u of user u can be represented as $S_n^u = \{(x_1^u, a_1^u), (x_2^u, a_2^u), \dots, (x_M^u, a_M^u)\}$, where $M = |S_n^u|$ is the number of travel packages in session S_n^u, x_j^u denotes the j-th travel package that user u operates and a_j^u denotes the operation type (such as a click or a purchase). As shown in Fig. 2, the latest session S_N^u that before latest purchase of user u is regarded as the short-term behaviors $\mathcal{S}^u$, while the rest of session of user u is regarded as the long-term behaviors $\mathcal{L}^u = \{S_1^u, S_2^u, \dots, S_{N-1}^u\}$. Based on these preliminaries, the NATR Recommendation task can be defined as follow:

Definition 3 (*Travel Package Recommendation*)**.** Given the target of user u with his/her short-term behaviors $\mathcal{S}^u$ and long-term behaviors $\mathcal{L}^u$, the travel package recommendation task is to predict the item x_j that target user u most likely to purchase in his/her next visit.

In this paper, we address this task with a novel personalized travel package recommendation model, i.e., Neural Attentive Travel package Recommendation (NATR). As shown in Fig. 3, the NATR model contains two core components: the *travel package encoder* module with attentive multi-view learning networks to learn representations of travel packages, the *user encoder* module with attentive and gated fusion networks to learn representations of users' preferences. In detail, our model firstly takes short-term behaviors $\mathcal{S}^u$ and long-term behaviors $\mathcal{L}^u$ of user u as input. $\mathcal{S}^u$ and $\mathcal{L}^u$ are encoded as short-term preference representation $\boldsymbol{s}_u$ and long-term preference representation $\boldsymbol{l}_u$ of user u, respectively. Then, we develop a gated-fusion network to coalesce two kinds of the user's preference representations from $\boldsymbol{s}_u$ and $\boldsymbol{l}_u$, and obtain the user's encoder vector $\boldsymbol{o}_u \in \mathbb{R}^d$. Let $\mathcal{V} \in \mathbb{R}^{d \times |\mathcal{X}|}$ denote the travel package's encoder vector of $\mathcal{X}$, where $|\mathcal{X}|$ is the number of all travel packages and d is the embedding size of each vector. Our goal is to predict top-K item (travel package) candidates based on the recommendation scores of inner travel package between $\boldsymbol{o}_u$ and each column vector $\boldsymbol{r}_j$ in $\mathcal{V}$ as:

$$\boldsymbol{z}_k = \boldsymbol{o}_u{}^{\mathrm{T}} \boldsymbol{r}_j, \tag{1}$$

where $\boldsymbol{r}_j \in \mathbb{R}^d$ is the j-th travel package's encoder vector.

3.2. Travel package encoder

The *travel package encoder* module aims to learn a unified travel package representation from the title, destination and categories (i.e., travel region and travel type) by incorporating them as different views of travel package, as shown in Fig. 3. Since different kinds of travel package information should be processed differently, as well as different words in the same title may have different important weights, we develop word-level and view-level attention networks to select important words and views for learning informative travel package representations. As shown in Fig. 3, the *travel package encoder* contains four core components as following.

3.2.1. Title encoder

The *title encoder* module is used to learn travel packages representations from their titles with three layers. The first layer is a title embedding layer, which is used to transform a package title into a low-dimensional dense vector. Specifically, we

[2] https://www.taobao.com/
[3] https://www.jd.com/

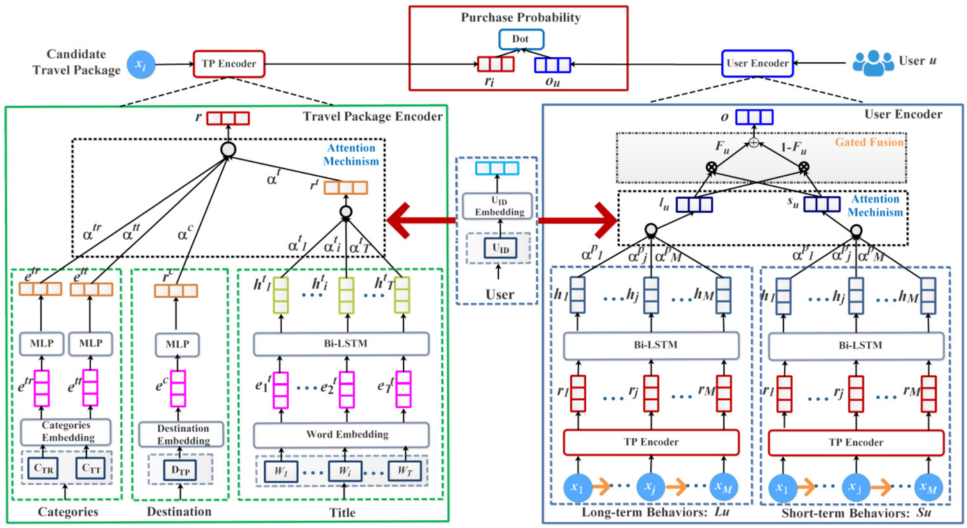


Fig. 3. The framework of the proposed model.

first use word2vec[4] to train an embedding model from a corpus of travel packages including 63 hundred thousand cases. Then, we can get the pre-trained word embedding dictionary $\mathcal{W} \in \mathbb{R}^{V \times D_w}$, where V and D_w are the vocabulary size and the word embedding dimension, respectively. Finally, let $\{w_1, w_2, \ldots, w_I\}$ denote the word sequence of a travel package's title, the word sequence can be converted into a sequence of word embedding vector $\{\boldsymbol{e}_1^t, \boldsymbol{e}_2^t, \ldots, \boldsymbol{e}_I^t\}$ by using the pre-trained word embedding dictionary $\mathcal{W}$, where I is number of words in this title.

The second layer is a Bidirectional Long Short-Term Memory (Bi-LSTM) [38] network. Since titles are the most comprehensive text description for all travel packages information, it is crucial important to learn representations of travel packages by exploiting their titles. For instance, the title of a travel package after processing by Chinese word segmentation is "the Forbidden City, the Temple of Heaven, the Badaling Great Wall, 3 days, tour, classic, the Imperial City and Beijing". On the whole, local context of "Beijing" and "3 days" are more useful for representing that it is a 3 days tour to Beijing. Meanwhile, at the micro-level, "the Forbidden City", "Temple of Heaven", and "the Badaling Great Wall" are more useful to represent the visited scenic spots. Therefore, Bi-LSTM is used to learn contextual word representation $\boldsymbol{e_i^t} \in \mathbb{R}^{\boldsymbol{D_w}}$ with their sequential and memory contexts. Let $\overrightarrow{h_i^t}$ and $\overleftarrow{h_i^t}$ denote the t-th forward and backward hidden layer representation of word w_i, which are calculated by:

$$\overrightarrow{h_i^t} = \overrightarrow{LSTM}(e_i^t), \quad \overleftarrow{h_i^t} = \overleftarrow{LSTM}(e_i^t) \tag{2}$$

We can get the final hidden layer vector of word w_i, denoted as $h_i^t = [\overrightarrow{h_i^t}, \overleftarrow{h_i^t}]$, by concatenating forward and backward LSTM representation, respectively. The output of the Bi-LSTM layer is the sequence of contextual word embedding vectors, denoted as $[h_1^t, h_2^t, \ldots, h_I^t]$.

The third layer is a word-level attention network [39]. Actually, different words in the title of same travel package usually have different informativeness to learn travel package representation. For example, in the travel package title "3 days, tour, classic, the Imperial City, Beijing", the word "Beijing" is more informative than "tour" in representing this travel package. Thus, recognizing the important words in different travel packages titles has the potential to learn more informative travel package representation. In addition, in traditional non-personalized attention networks, the attention weights are only calculated based on the input representation sequence via a fixed attention vector, and the users' preferences are not considered [40]. To model the informativeness of each word for the recommendation of different users, a personalized attention network is proposed to select important words in the packages titles with the user's preferences, as shown in Fig. 3. Specifically, we first embed identifiers (ID) of user u into a representation vector as the user's preference vector $\boldsymbol{q_u} \in \mathbb{R}^{D_q}$, where D_q is the embedding size of user representation vector. Let α_i^t denote the attention weight of the i-th word in the j-th travel package title, which is formulated as:

$$a_i^t = \boldsymbol{q}_{\boldsymbol{u}}^{\mathrm{T}} \tanh(W_t h_i^t + b_t), \tag{3}$$

$$\alpha_i^t = \frac{\exp(a_i^t)}{\sum_{m=1}^{I} \exp(a_m^t)}, \tag{4}$$

where W_t and b_t are the projection parameters. The final title representation of travel package x_j is the summation of the contextual representations of its words weighted by their attention weights, which is calculated by:

$$\boldsymbol{r}_j^t = \sum_{i=1}^{I} \alpha_i^t h_i^t. \tag{5}$$

3.2.2. Destination encoder

The *destination encoder* module is used to learn travel packages representations from their destinations. In contrast with title of travel package, destination is a just label with very few words, such as "USA", "Eastern China" and "Beijing". Actually, tourists usually begin to visit travel packages on OTA after they travel destinations and schedules have been arranged. For example, if a user browsed a number of travel packages with the destination of "Beijing", then we can infer that this user may be interested in tours of "Beijing". Here, we only design a module containing two layers for the *destination encoder* module. The inputs of the *destination encoder* are the identifiers (IDs) of the destinations. The first layer is a destination embedding layer, which can transform the discrete IDs of destination into low-dimensional dense representation vector $\boldsymbol{e}^c \in \mathbb{R}^{D_d}$, where D_d is the embedding dimension of the destination ID. The second layer is a Multi-Layer Perceptron (MLP) [41] which is used to learn the hidden destination representation of travel package x_j:

$$\boldsymbol{r}_j^c = \mathrm{ReLU}(W_c \boldsymbol{e}^c + b_c), \tag{6}$$

[4] https://radimrehurek.com/gensim/models/word2vec.html

where W_c and b_c represent weight matrix and bias vector in the *destination encoder*.

3.2.3. Categories encoder

The *categories encoder* module is used to learn travel packages representations from their categories. As introduced in Section 3.1, travel package is usually labeled with categories from two angles by the OTA, i.e., Travel Region (TR) and Travel Type (TT). For example, if a user browsed amounts of travel packages with the travel region of "around tour" and travel type of "self-driving tour". Then, we can infer that this user may be interested in a self-driving tour to the region where is close to his/her living city. Similar with the *destination encoder* module, there are also two layers in the *categories encoder* module. In detail, we propose to incorporate both two categories information to learn travel package representation. The first layer is a category embedding layer, which can convert the discrete IDs of Travel Region (TR) and Travel Type (TT) into low-dimensional dense representation vectors denoted as $\boldsymbol{e}^{tr} \in \mathbb{R}^{D_{tr}}$ and $\boldsymbol{e}^{tt} \in \mathbb{R}^{D_{tt}}$, where D_{tr} and D_{tt} are the embedding dimension of travel region ID and travel type ID, respectively. The second layer is also a MLP layer, and it learns the category representations $\boldsymbol{r}_j^{tr}$ and $\boldsymbol{r}_j^{tt}$, respectively.

3.2.4. View-level attention

The *view-level attention* approach is used to model the informativeness of different kinds of travel package information to learn a unified travel package representation. For example, as shown in Table 1, title of the first travel package including destination, scenic spot, number of days, vehicle, hotel and sale is precise and contains rich information. Thus, it should have a high weight in representing this travel package. However, the title of the third travel package just includes destination, number of days, and characteristic is short and vague. Thus, the weight of title in the third travel package should be smaller than in the first travel package for representing this travel package. Motivated by these observations, a view-level attention mechanism is proposed to model the informativeness of different kinds of package information for learning the package representation, as shown in Fig. 3. Let $\alpha_t, \alpha_c, \alpha_{tr}$ and α_{tt} denote the attention weights of title, destination, travel region and travel type of travel package x_j, respectively. Taking the attention weight of the title view α_t as an example, α_t can be calculated by:

$$\begin{aligned} a_t &= \boldsymbol{q}_{\boldsymbol{u}}^{\mathrm{T}} \tanh(W_v \boldsymbol{r}_j^t + b_v), \\ \alpha_t &= \frac{\exp(a_t)}{\exp(a_t) + \exp(a_c) + \exp(a_{tr}) + \exp(a_{tt})}, \end{aligned} \tag{7}$$

where W_v and b_v are parameters in the *view-level attention* approach, $\boldsymbol{q}_u \in \mathbb{R}^{D_q}$ is the user preference query as mentioned in the *title encoder* module. The attention weights α_c, α_{tr} and α_{tt} of the destination, travel region and travel type of travel package can be computed in a similar way.

The final unified representation of package x_j is the summation of the travel package representations from different views weighted by their attention weights as follows:

$$\boldsymbol{r}_j = \alpha_t \boldsymbol{r}_j^t + \alpha_c \boldsymbol{r}_j^c + \alpha_{tr} \boldsymbol{r}_j^{tr} + \alpha_{tt} \boldsymbol{r}_j^{tt}. \tag{8}$$

At last, if user u has a session denoted as $\{(x_1^u, a_1^u), (x_2^u, a_2^u), \dots, (x_M^u, a_M^u)\}$, we apply the *travel package encoder* module to learn the representations of all the browsed/purchased travel packages, the representations vector can be denoted as $\boldsymbol{r}^u = \{\boldsymbol{r}_1^u, \boldsymbol{r}_2^u, \dots, \boldsymbol{r}_M^u\}$.

3.3. User encoder

The *user encoder* module aims to learn the representations of users, as shown in Fig. 3. In this module, package-level attention mechanism and gated fusion approach are used to build informative user representation, respectively. As introduced in Section 1, different travel packages browsed by the same user have different informativeness for representing this user preference. Thus, we adopt a travel package attention mechanism to learn more informative user representations by selecting important travel packages. In addition, since users' long-term preferences always influence the decisions at present [27], it is crucial to consider both long-term and short-term preferences in the *user encoder* module. To this end, we design a gated fusion network to coalesce both long-term and short-term preference of users, which are learned from representation of browsed/purchased travel packages in their current and historical clickstream, respectively. As shown in Fig. 3, there are three core components in the *user encoder* module.

3.3.1. Short-term preference encoder

The *short-term preference encoder* module is used to learn the short-term preference of the target user u from his/her short-term behaviors $\mathcal{S}^u$. Aiming at the alignment of session of user u, we develop a Bi-LSTM model to learn short-term preference motivations of user u. With the *travel package encoder* module, we can obtain the representation vectors of short-term behaviors $r^{\mathcal{S}^u} = \{\boldsymbol{r}_1^u, \boldsymbol{r}_2^u, \dots, \boldsymbol{r}_M^u\}$. At the t-th interaction step, the forward hidden representation $\overrightarrow{h_j^t}$ and backward hidden representation $\overleftarrow{h_j^t}$ of current travel package x_j is updated by the Bi-LSTM model. Here, the detailed computation is similar to Section 3.2.1. As a result, we can obtain the final out state $c_j^p = [\overrightarrow{h_j^t}, \overleftarrow{h_j^t}]$ as the representation of the j-th travel package. To model the different informativeness of the same travel package for different users, we also apply personalized attention mechanism to represent the travel package clicked or purchased by the same user. We denote the attention weight of the j-th package browsed/purchased by the user u as α_j^p, which is calculated by evaluating the importance of the interactions between user preference and package representation as follows:

$$\begin{aligned} a_j^p &= \boldsymbol{q}_{\boldsymbol{u}}^{\mathrm{T}} \tanh(W_p c_j^p + b_p), \\ \alpha_j^p &= \frac{\exp(a_j^p)}{\sum_{m=1}^{M} \exp(a_m^p)}, \end{aligned} \tag{9}$$

where W_p and b_p are the projection parameters. The final short-term preference representation $\boldsymbol{s}_u$ of user u is the summation of the packages contextual representations weighted by their attention weights, which is calculated by:

$$\boldsymbol{s}_u = \sum_{j=1}^{|\mathcal{S}^u|} \exp(\alpha_j^p c_j^p). \tag{10}$$

3.3.2. Long-term preference encoder

The *long-term preference encoder* module is used to learn the long-term preference of the target user u from his/her long-term behaviors $\mathcal{L}^u$. The first layer is *travel package encoder* module, which is used to convert long-term behaviors $\mathcal{L}^u$ from the ID sequence of travel package into a sequence of representation vector $\boldsymbol{r}^u$. Let $\boldsymbol{r}_j^u$ denote the j-th travel package representation vector in long-term behaviors $\mathcal{L}^u$. Similar with the *short-term preference encoder* module, we also use the Bi-LSTM and the personalized attention mechanism to obtain the final long-term preference representation $\boldsymbol{l}_u$ of user u.

3.3.3. Long-term and short-term preference fusion

To combine the long-term and short-term preferences of user u, we use a gated-fusion network to measure the usefulness of the long-term and short-term preference vectors and aggregates information accordingly. Given the user preference q_u, short-term preference vector $\boldsymbol{s}_u$ and long-term preference vector $\boldsymbol{l}_u$ as inputs, the gate vector $\boldsymbol{F}_u \in \mathbb{R}^{D_q}$ is developed to control the contribution between long-term and short-term preferences:

$$\boldsymbol{F}_u = sigmoid(W_q \boldsymbol{q}_u + W_s \boldsymbol{s}_u + W_l \boldsymbol{l}_u + b_u), \tag{11}$$

where W_q, W_s, W_l, b_u are projection parameters. The final output of preference vector $\boldsymbol{o}_u$ of user u is computed by:

$$\boldsymbol{o}_u = (1 - \boldsymbol{F}_u) \odot \boldsymbol{s}_u + \boldsymbol{F}_u \odot \boldsymbol{l}_u, \tag{12}$$

where $\odot$ is element-wise multiplication.

3.4. Model training and test stage

During training process, for short-term behaviors $\mathcal{S}^u$ and long-term behaviors $\mathcal{L}^u$ of user u, the positive label is the next purchased travel package x_j^u. While, the negative labels are travel packages sampled from $\mathcal{X}$ excluding x_j^u by the log-uniform sampler in a real-world application. After obtaining the user preference representation vector $\boldsymbol{o}_u$ and the travel package representation vector $\boldsymbol{r}_j$, the recommendation score $\boldsymbol{z}_k$ in the candidate travel packages $\boldsymbol{z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{|\mathcal{K}|}\}$ can be calculated by Eq. (1), where $\mathcal{K}$ is the sampled subset of $\mathcal{X}$ including positive and negative labels, $\hat{\boldsymbol{y}} = \{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_{|\mathcal{K}|}\}$ is the prediction probabilities over each travel package in $\mathcal{K}$. Then, we apply a sampled-softmax function [27] to get the output vector of the NATR model, i.e., $\hat{\boldsymbol{y}} = \text{softmax}(\boldsymbol{z})$, where $\hat{\boldsymbol{y}}$ is the probabilities of travel packages appearing to be purchased in session $\mathcal{S}^u$.

For each user, the loss function is defined as the cross-entropy of the prediction probability and the ground truth. It can be calculated by:

$$\mathcal{L}(\hat{\boldsymbol{y}}) = -\sum_{x_j \in \mathcal{K}} y_j \log(\hat{y}_j), \tag{13}$$

where y_j is the truly probability distribution of package x_j. Specifically, if x_j is the positive label, then $y_j = 1$, otherwise $y_j = 0$.

By minimizing the loss function $\mathcal{L}(\hat{\boldsymbol{y}})$ via Adam optimizer, all parameters can be tuned in the NATR model. Compared with existing recommendation methods, the proposed model can effectively exploit the useful information in negative samples, and further reduce the computational cost for model training. Thus, NATR can be trained more easily on a large collection of clickstream in tourism e-commerce.

4. Experiments

In this section, we first present our experimental setup. Then, we demonstrate the effectiveness of proposed model from the following aspects: (1) the performance comparisons between the proposed NATR model and other benchmark methods; (2) the effectiveness of multi-view learning and attention module inside the NATR model; (3) the visualization of attention weights inside the NATR model; (4) the analysis on cold-start scenarios.

4.1. Experimental setup

4.1.1. Data

Our dataset is provided by Tuniu,[5] a large tourism e-commerce in China. This dataset is mainly made up of the page-to-page clickstream data recorded in server logs, which has been widely adopted in the research on recommender systems for e-commerce [27,37]. The clickstream data involved in this study spans two months (i.e., Jul. and Aug., 2013), which are denoted as D_1 and D_2 respectively. Then, we use the first 29 days data for training and the rest 2 days data for testing. For the reliability of the experimental results, several preprocessing tasks are conducted: (i) to focus on the interactions on product pages (i.e., travel packages) which can reflect users' preferences, we filter out a large number of non-product pages, such as main pages and anonymous pages visited by the users; (ii) we filter out the users whose short-term behaviors lengths are less than 2, and items that appear less than 5 times. The statistics of two datasets after preprocessing are shown in Table 2.

4.1.2. Baseline methods

We compare the proposed NATR model with four traditional recommendation methods (i.e., POP, User-KNN, Item-KNN and SVD), four state-of-the-art neural-based recommendation methods (i.e., CDL, DeepFM, BINN and SDM), and two variants of NATR model (i.e., NATR-NoL and NATR-NoA) for ablation study.

- **POP**. Popularity (POP) predictor always recommends the most popular items in the training set. Despite its simplicity, it is usually used as a baseline in various domains.
- **User-KNN** [42]. User-KNN is a user-based collaborative filtering method, which is one of the major candidate generation approaches in industry.
- **Item-KNN** [43]. Item-KNN is an item-based collaborative filtering method, which is widely used in recommendation for e-commerce websites.
- **SVD** [16]. Singular Value Decomposition (SVD) is another famous recommendation model with the matrix factorization method.
- **CDL** [29]. Collaborative Deep Learning (CDL) jointly performs deep representation learning for the content information and collaborative filtering for the rating (feedback) matrix.
- **DeepFM** [31]. DeepFM is also a widely used neural-based recommendation method that combines factorization machines with deep neural networks.
- **BINN** [37]. BINN applies RNN-based method to encode present consumption motivations and historical purchase behaviors, respectively. It generates a unified representation by concatenation operation.
- **SDM** [27]. Sequential Deep Matching (SDM) model uses multi-head self-attention module to capture multiple types of interests, and long-short term gated fusion approach to incorporate long-term preferences. Successive items can be recommended after matching between sequential user behavior vector and item embedding vectors.
- **NATR-NoL**. NATR-NoL removes the long-term preference encoder from NATR model.
- **NATR-NoA**. NATR-NoA removes the word-level, view-level and package-level attention networks from NATR.

To be specific, for the traditional methods, cosine similarity is selected as the similarity measure, and set the number of nearest neighbors as 80 in User-KNN and Item-KNN; the dimension of latent variables is set as 10 in SVD, and all of these algorithms are implemented in Mahout.[6] For the neural-based methods, all the hidden units sizes in the RNN-based models are set to 256, and their dropout probabilities and learning rates are set to 0.2. In the proposed NATR model, the hidden unit size of LSTM is set to 128,

5 http://www.tuniu.com

6 http://mahout.apache.org/

Table 2
Statistics of two datasets.

Dataset	Data type	Time interval	#Users	#Items	#Records	#Sessions	S.Len	S.Len	#Purchased items
D_1	Training	1 to 29 Jul.	22,699	24,834	420,315	25,988	4.86	11.32	6,101
	Test	30 to 31 Jul.	1,699	7,971	38,559	1,862	4.46	16.25	1,097
D_2	Training	1 to 29 Aug.	25,704	18,419	422,512	31,754	4.81	9.64	6,348
	Test	30 to 31 Aug.	1,028	4,402	16,045	1,112	4.63	9.80	780

Note: (1) "#" indicates the number of someone object;
(2) "S.Len" and "L.Len" indicate the average length of short-term and long-term clickstream, respectively.

the batch size is set to 128, the attention query size D_q is set to 256, and the embedding dimension D_d, D_{tr} and D_{tt} are set to 256. In the pre-trained word embedding model, the word embedding size D_w is set to 300. The NATR and all the compared neural-based models are defined and trained on a Windows server with 3.60 GHz Intel I9-9900k CPU and 11 GB Nvidia GeForce RTX 2080 Ti GPU, and implemented in Pytorch.[7]

4.1.3. Evaluation metrics

To evaluate the performance of NATR and baseline methods, we adopt HitRate [27], Item-coverage [44] and Mean Reciprocal Rank (MRR) [37] as the evaluation metrics.

HitRate@20 measures the accuracy of the recommendation. It is defined as the proportion of the cases which have the correctly recommended items among the top-20 items in all test cases:

$$\text{HitRate@20} = \frac{1}{|\mathcal{T}|}\sum_{u\in\mathcal{T}} 1 \text{ (if } R_{u,g_u} \le 20), \tag{14}$$

where $\mathcal{T}$ is the test set, and g_u is the item that was purchased in the current session of user u. R_{u,g_u} is the rank of this desired item. If the item g_u occurs in the top 20 list, the indicator function will be set to 1, otherwise 0.

Item-coverage@20 is the percentage of distinct items that are correctly recommended among the top-20 items in all test cases. The definition of Item-coverage@20 is as follows:

$$\text{Item-coverage@20} = \frac{Dis(\sum_{u\in\mathcal{T}} R_{u,g_u})}{Dis(\sum_{u\in\mathcal{T}} g_u)} \text{ (if } R_{u,g_u} \le 20), \tag{15}$$

where *Dis* is the function counting the number of distinct items in a dataset.

MRR@20 is the average reciprocal ranks of the correctly recommended item, which indicates how well the model ranks the item [34]. Intuitively, ranking the correctly recommended item higher is more preferable in practice. The definition of MRR@20 is as follows:

$$\text{MRR@20} = \frac{1}{|\mathcal{T}|}\sum_{u\in\mathcal{T}} \frac{1}{R_{u,g_u}}, \tag{16}$$

where if the rank is large than 20, the reciprocal rank will be set to 0.

4.2. Performance evaluation

In this subsection, we compare the overall performance between the NATR model and other baseline methods to demonstrate the effectiveness of the proposed NATR model. The results are shown in Table 3. As a whole, NATR achieves the best performance on both two travel datasets indicated by all of the evaluation metrics. Specifically, we have several observations.

Firstly, it is clear to observe that neural-based collaborative filtering and matrix factorization methods (i.e., CDL and DeepFM) are noticeably superior to traditional methods (e.g., Item-NN, User-KNN and SVD). In particular, the improvement rates of CDL and DeepFM on HitRate@20, Item-coverage@20, and MRR@20 are 7.1%–19.8%, 16.1%–32.1%, and 49.4%–160.7% respectively, compared with the best traditional method (User-KNN) on both two datasets. This is probably because user–package interaction matrices in travel datasets are so sparse that traditional methods cannot be applicable to travel package recommendation directly, and neural models can learn better users and travel packages representations than traditional methods. In addition, POP performs the worst, which indicates the necessity of modeling users personalized preferences rather than just recommending popular items to users.

Next, the neural-based methods using RNN (e.g., BINN, SDM and NATR) outperform most of the methods using collaborative filtering and matrix factorization (i.e., CDL and DeepFM), indicating that RNN-based models do have better abilities to dynamically learn users' preferences than traditional methods. For example, SDM has an improvements of 10.7%–13.7% on HitRate@20, 24.6%–38.9% on Item-coverage@20, and 20.6%–35.1% on MRR@20 respectively, compared with CDL on both two datasets. This clearly shows that RNN-based models are more suitable for travel package recommendations by exploit amounts of users' clickstream.

Then, RNN-based models (i.e., BINN, SDM and NATR) using long-term preference learning outperform those without (i.e., CDL, DeepFM and NATR-NoL). In particular, the NATR model has an improvements of 26.0%–27.3% on HitRate@20, 35.0%–41.4% on Item-coverage@20, and 38.2%–45.2% on MRR@20, compared with the NATR-NoL model on both two datasets. This is because long-term behaviors always represent personal preferences that can affect users' current purchasing decisions. Moreover, from the view of Item-coverage@20, the recommendation list could be more diverse if long-term preference learning is fused with short-term preference learning.

Finally, RNN-based models with attention networks (i.e., SDM and NATR) outperform most of those without (e.g., BINN and NATR-NoA). In particular, the NATR model has an improvements of 12.8%–14.3% on HitRate@20, 21.2%–22.6% on Item-coverage@20, and 18.5%–20.2% on MRR@20, compared with NATR-NoA model on both two datasets. This phenomenon is consistent with the result of neural news recommendation with personalized attention approach [40], which is probably because different packages and their contexts usually have different informativeness for recommendation and selecting important features of packages and users is useful for achieving better recommendation performance.

4.3. Effectiveness of multi-view learning and attention networks

In this subsection, we try to reveal the run-time mechanism and validate the effectiveness of multi-view learning approach and attention networks inside the NATR model. To this end, we conduct several experiments to observe the performance of NATR and its variants with different combinations of view learning and attention networks. As a result, the experimental results on two datasets are shown in Figs. 4 and 5, respectively.

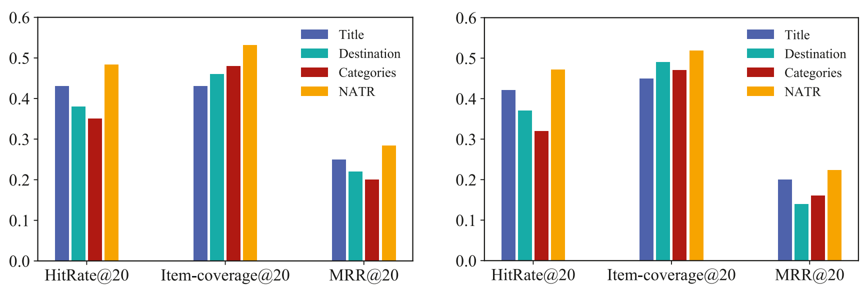
According to Fig. 4, we have several observations. First, the model with title view achieves better performance than those

[7] https://pytorch.org/

Table 3
Performance comparison of different methods of on two datasets.

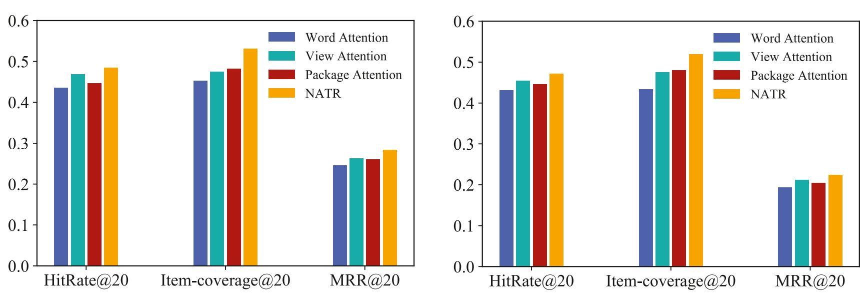
Methods	D_1			D_2		
	HR@20 (%)	Item-c@20 (%)	MRR@20 (%)	HR@20 (%)	Item-c@20 (%)	MRR@20 (%)
POP	8.13	0.35	1.92	9.05	0.35	1.96
Item-KNN	4.88	13.72	1.11	5.30	14.52	1.20
User-KNN	33.06	31.30	7.49	36.54	31.55	8.67
SVD	18.69	16.43	5.31	21.20	16.78	5.92
CDL	38.91	36.33	19.53	39.12	38.57	12.95
DeepFM	39.62	38.51	17.78	39.95	41.67	18.73
BINN	40.61	41.18	25.33	41.94	48.05	17.26
SDM	44.24	45.25	23.56	43.30	**53.59**	21.19
NATR-NoL	38.41	37.59	19.53	37.09	38.47	16.21
NATR-NoA	42.32	43.84	23.59	41.87	42.36	18.91
NATR	**48.39**	**53.14**	**28.35**	**47.21**	51.92	**22.41**

Note: HR@20 and Item-c@20 denote HitRate@20 and Item-coverage@20, respectively.



(a) D_1 (b) D_2

Fig. 4. Effectiveness of multi-view learning approach in NATR.



(a) D_1 (b) D_2

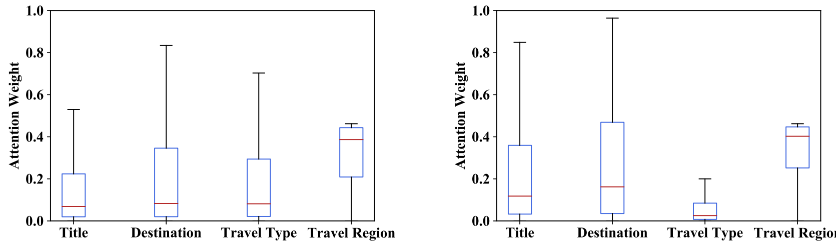
Fig. 5. Effectiveness of attention networks in NATR.

with destination or categories. This is because titles of travel packages usually contain rich and various descriptions, such as hotels, scenic spots, vehicles and time duration, and so on. Therefore, these titles can provide rich information to model different packages topics. Second, the destination and categories views are also informative for travel package recommendations. This is probably because destination and categories (i.e., travel region and travel type) are important clues of package topics, and it is also useful to incorporate destination and categories views for learning travel packages and users representations. Third, the NATR model combining all three views can further improve the performance of this approach. In summary, these results validate the effectiveness of the multi-view learning approach in NATR.

According to Fig. 5, we have several observations. First, the word-level attention network can effectively improve the performance of this approach. One possible explanation could be that words are basic units in titles to convey their meanings, and different words usually have different informativeness for learning travel packages representations. Therefore, word-level attention network can recognize and highlight important words, which is useful to learn more informative package representations. Second, the view-level attention network can also improve the performance of our approach. Since different views can also have different informativeness for learning travel package and users representations, it may be useful to evaluate the importance of views for improving the performance of our approach. Third, the package-level attention network can also improve the performance of our approach. This is probably because packages browsed by the same user usually have different informativeness for learning the representations of this user, and it is very critical to recognize the important packages in long-term and short-term clickstream to learn high-quality users representations. Fourth, integrating all attention networks can further improve the performance of our approach. In summary, these results validate the effectiveness of attention networks in NATR.

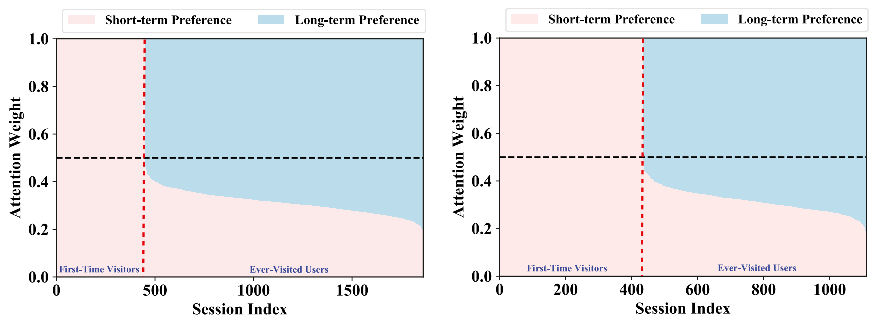
4.4. Visualization of attention weights

In this subsection, we take further analysis on the visualization of attention weights inside the NATR model. To this end, we conduct several experiments to observe attention weights of travel



(a) D_1 (b) D_2

Fig. 6. Visualization of the view-level attention weights in NATR.



(a) D_1 (b) D_2

Fig. 7. Visualization of the long-term and short-term preferences attention weights in NATR. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

packages views and long-term and short-term preferences on two test datasets. The results are shown in Figs. 6 and 7, respectively.

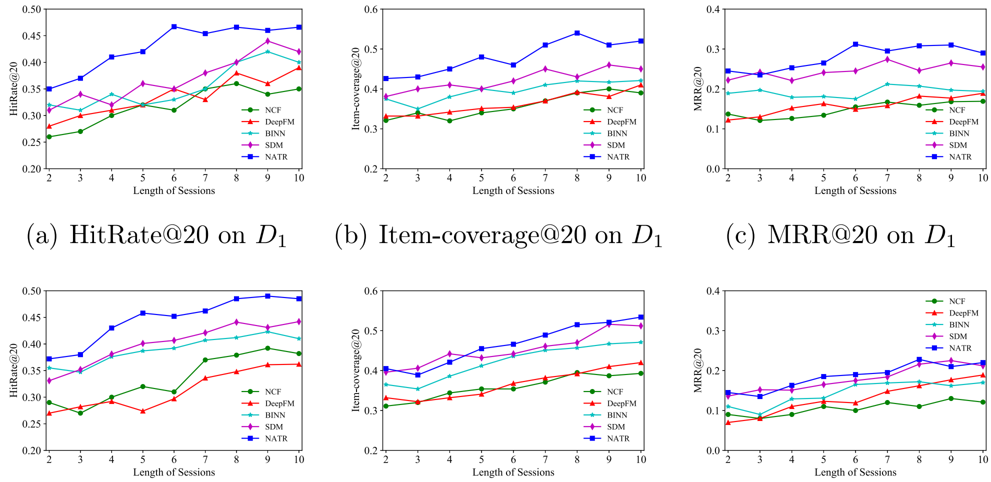
First, from Figs. 6(a) and 6(b), it is clear to observe that the travel region view gains the highest attention weights. Specifically, the medians of its attention weights are 0.40 and 0.39 in D_1 and D_2, respectively. The reason for this is pretty obvious: the majority of users who have strong purchase intents are likely to browse travel packages taking their interested travel region as target, such as nearby tour, short-term domestic tour and long-term abroad tour, etc. Similarly, we observe that average attention weights of the destination view are slightly higher than those of title and travel type. This phenomenon indicates that destination is a very important clue for inferring travel topics, which is very informative for learning high-quality travel package representations. Furthermore, it is noteworthy that attention weights of the title view are small for many samples. This is mainly because part of titles are vague and uninformative for learning travel packages representations.

Second, Figs. 7(a) and 7(b) show the visualization of long-term and short-term preferences attention weights on two test datasets. Note that users can be divided into two disjoint categories (i.e., `first-time visitors` and `ever-visited users`), solely according to the browsing records. Since first-time visitors have no long-term behaviors for learning users representations, so long-term preferences weights of those users are set to 0 by default, while their short-term preferences weights are set to 1 by default (see the left region of red dotted line in Figs. 7(a) and 7(b)). It is noteworthy that the percentage of first-time visitors in D_1 and D_2 are 18.54% and 33.00%, respectively, which indicates that the majority of users have long-term preferences. We further look inside short-term preferences attention weights of first-time visitors. As shown by the right region of red dotted line in Figs. 7(a) and 7(b), we can observe that attention weights of long-term preferences are significantly higher than those of short-term preferences, falling into [0.50, 0.82] and [0.51, 0.81], respectively. While, those of short-term preferences are falling into [0.18, 0.50] and [0.19, 0.49], respectively. This implies that, compared with short-term preferences, long-term preferences are more important for learning ever-visited users representations. Therefore, it is crucial to coalesce long-term preferences with short-term preferences for learning high-quality users representations.

4.5. Cold start of new users

Cold start is a common problem of recommender systems that new users or items have not yet gathered sufficient information to recommend or be recommended [37]. Actually, previous studies [2,4,8,9] have verified that the interaction matrix between users and travel packages is much sparser than that of traditional products. In the travel datasets, an overwhelming majority of users clicked less than 10 travel packages in the current session. To this end, we extract these cold start users from test datasets and focus on examining the performance of the NATR model on the cold start problem caused by these users.

Indeed, since new users have no interaction to be pre-trained and recommender systems cannot generate user profiles, many user profile-based recommendation methods cannot work well, especially collaborative filtering and factorization models [37]. However, for neural-based recommendations, we can use a trained neural network to learn preferences of new users and recommend items that users may be interested in. It is noteworthy that we do not change any training process and just use the selected cold-start users for testing, thus all the testing do not need retraining. For better illustration, we report results of all neural-based models on two datasets in Fig. 8, respectively, and the results yield the following conclusions. First, in most cases, the NATR significantly outperforms benchmark methods. For example, for all of the users whose current session length is 5, the improvement of NATR achieves 14.2%–16.7% on HitRate@20, 5.3%–20.0% on Item-coverage@20 and 8.2%–12.1% on MRR@20, compared with the second-best method (SDM). This indicates the effectiveness of multi-view learning and attention network



(a) HitRate@20 on D_1 (b) Item-coverage@20 on D_1 (c) MRR@20 on D_1

(d) HitRate@20 on D_2 (e) Item-coverage@20 on D_2 (f) MRR@20 on D_2

Fig. 8. Performance of cold start users on two datasets.

in NATR. Second, on the whole, with the length of users' sessions growing, the NATR model achieves far greater improvement on recommendation performance. That can illustrate the effectiveness of modeling long-term preferences of users with the NATR model. Third, all the RNN-based models have shown strong capacity to face the cold-start challenges of new users, compared with neural-based models using collaborative filtering and matrix factorization. Therefore, it is noteworthy that all the neural-based models can work well for new users.

5. Conclusions

With the rapid development of tourism e-commerce industry, massive amounts of travel data has been accumulated, thus providing opportunities for decision makers to understand tourists' behaviors and provide information about travel recommendation. To tackle the travel package recommendation issues, we put forward a neural model named NATR including two core components, i.e., *travel package encoder* and *user encoder* in this paper. Different from existing methods, in the *travel package encoder* module, a unified travel package representation can be learned by using word-level and view-level attention networks to select important words and views from their attributes. Meanwhile, in the *user encoder* module, the dynamic evolution of users' preferences can be learned by Bi-directional Long Short-Term Memory (Bi-LSTM) neural networks with package-level attention mechanism, instead of traditional collaborative filtering and matrix factorization in matching stage. Moreover, a gated fusion network is developed to coalesce long-term and short-term preferences for learning high-quality users' representations, which shows that the network has more expressive power to control the importance of both preferences.

In this paper, we conduct extensive experiments based on two real-life travel datasets. Experimental results show the effectiveness of the proposed model in personalized travel package recommendation. In particular, multi-view learning and attention mechanism can remarkably improve the performance of the proposed approach. Meanwhile, attentive multi-view learning and gated fusion network can provide an informative visualization of attention weights inside the NATR model, which is helpful to identify the key views that accurately represent travel package and give importance to the two kinds of preferences for learning users representations accurately. Despite our current work provides a solid foundation for neural attentive travel package recommendation, a neural multi-task learning framework that supports both purchase prediction and recommendation for tourism e-commerce still warrants further investigation.

CRediT authorship contribution statement

Guixiang Zhu: Conceptualization, Methodology, Software, Writing. **Youquan Wang:** Methodology, Writing - review & editing. **Jie Cao:** Supervision, Formal analysis, Funding acquisition. **Zhan Bu:** Supervision, Investigation. **Shuxin Yang:** Investigation, Validation. **Weichao Liang:** Writing, Validation. **Jingting Liu:** Visualization, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 71701089, Grant 71871109, Grant 91646204, the National Center for International Joint Research on E-Business Information Processing under Grant 2013B01035, and the International Innovation Cooperation Project of Jiangsu Province under Grant BZ2020008.

References

[1] G. Zhu, Z. Wu, Y. Wang, S. Cao, J. Cao, Online purchase decisions for tourism e-commerce, Electron. Commer. Res. Appl. 38 (2019) 100887.
[2] Q. Liu, Y. Ge, Z. Li, E. Chen, H. Xiong, Personalized travel package recommendation, in: Proceedings of the 11th International Conference on Data Mining, IEEE, 2011, pp. 407–416.
[3] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, Recommender system application developments: A survey, Decis. Support Syst. 74 (2015) 12–32.
[4] Q. Liu, E. Chen, H. Xiong, Y. Ge, Z. Li, X. Wu, A cocktail approach for travel package recommendation, IEEE Trans. Knowl. Data Eng. 26 (2) (2014) 278–293.
[5] Y. Ge, Q. Liu, H. Xiong, A. Tuzhilin, J. Chen, Cost-aware travel tour recommendation, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2011, pp. 983–991.
[6] Y. Ge, H. Xiong, A. Tuzhilin, Q. Liu, Cost-aware collaborative filtering for travel tour recommendations, ACM Trans. Inf. Syst. 32 (1) (2014) 4.

[7] J. He, H. Liu, H. Xiong, SocoTraveler: Travel-package recommendations leveraging social influence of different relationship types, Inf. Manag. 53 (8) (2016) 934–950.
[8] G. Zhu, J. Cao, C. Li, Z. Wu, A recommendation engine for travel products based on topic sequential patterns, Multimedia Tools Appl. 76 (16) (2017) 17595–17612.
[9] L. Chen, Z. Wu, J. Cao, G. Zhu, Y. Ge, Travel recommendation via fusing multi-auxiliary information into matrix factorization, ACM Trans. Intell. Syst. Technol. (TIST) 11 (2) (2020) 22.
[10] Y. Li, S. Wang, Q. Pan, H. Peng, T. Yang, E. Cambria, Learning binary codes with neural collaborative filtering for efficient recommendation systems, Knowl.-Based Syst. 172 (2019) 64–75.
[11] C. Angulo, Z. Falomir, D. Anguita, N. Agell, E. Cambria, Bridging cognitive models and recommender systems, Cognit. Comput. 12 (2) (2020) 426–427.
[12] C. Xu, M. Liao, P. Li, Z. Liu, S. Yuan, New results on pseudo almost periodic solutions of quaternion-valued fuzzy cellular neural networks with delays, Fuzzy Sets and Systems (2020) http://dx.doi.org/10.1016/j.fss.2020.03.016.
[13] C. Xu, P. Li, Y. Pang, Exponential stability of almost periodic solutions for memristor-based neural networks with distributed leakage delays, Neural Comput. 28 (12) (2016) 2726–2756.
[14] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107.
[15] J.S. Breese, D. Heckerman, C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, in: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, ACM, 1998, pp. 43–52.
[16] Y. Koren, Factorization meets the neighborhood: A multifaceted collaborative filtering model, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2008, pp. 426–434.
[17] Q. Zhang, J. Lu, D. Wu, G. Zhang, A cross-domain recommender system with kernel-induced knowledge transfer for overlapping entities, IEEE Trans. Neural Netw. Learn. Syst. 30 (7) (2018) 1998–2012.
[18] B. Wu, T. Mei, W.-H. Cheng, Y. Zhang, Unfolding temporal dynamics: Predicting social media popularity using multi-scale temporal decomposition, in: Proceedings of the 30th AAAI Conference on Artificial Intelligence, 2016, pp. 272–278.
[19] S. Ahmad, I. Ullah, F. Mehmood, M. Fayaz, D. Kim, A stochastic approach towards travel route optimization and recommendation based on users constraints using Markov chain, IEEE Access 7 (2019) 90760–90776.
[20] L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang, H. Li, B. Yang, D. Liu, Evolving support vector machines using fruit fly optimization for medical data classification, Knowl.-Based Syst. 96 (2016) 61–75.
[21] M. Wang, H. Chen, B. Yang, X. Zhao, L. Hu, Z. Cai, H. Huang, C. Tong, Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses, Neurocomputing 267 (2017) 69–84.
[22] M. Wang, H. Chen, Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis, Appl. Soft Comput. 88 (2020) 105946.
[23] J. Liu, B. Liu, Y. Liu, H. Chen, L. Feng, H. Xiong, Y. Huang, Personalized air travel prediction: A multi-factor perspective, ACM Trans. Intell. Syst. Technol. (TIST) 9 (3) (2018) 30.
[24] H. Yin, W. Wang, H. Wang, L. Chen, X. Zhou, Spatial-aware hierarchical collaborative deep learning for POI recommendation, IEEE Trans. Knowl. Data Eng. 29 (11) (2017) 2537–2551.
[25] H. Li, Y. Ge, D. Lian, H. Liu, Learning user's intrinsic and extrinsic interests for point-of-interest recommendation: A unified approach, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017, pp. 2117–2123.
[26] C. Tan, Q. Liu, E. Chen, H. Xiong, X. Wu, Object-oriented travel package recommendation, ACM Trans. Intell. Syst. Technol. (TIST) 5 (3) (2014) 43.
[27] F. Lv, T. Jin, C. Yu, F. Sun, Q. Lin, K. Yang, W. Ng, SDM: Sequential deep matching model for online large-scale recommender system, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, ACM, 2019, pp. 2635–2643.
[28] G. Pang, X. Wang, F. Hao, J. Xie, X. Wang, Y. Lin, X. Qin, ACNN-FM: A novel recommender with attention-based convolutional neural network and factorization machines, Knowl.-Based Syst. 181 (2019) 104786.
[29] H. Wang, N. Wang, D.-Y. Yeung, Collaborative deep learning for recommender systems, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 1235–1244.
[30] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in: Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 173–182.
[31] H. Guo, R. Tang, Y. Ye, Z. Li, X. He, DeepFM: A factorization-machine based neural network for CTR prediction, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017, pp. 1725–1731.
[32] C. Xu, Q. Zhang, Y. Wu, Bifurcation analysis of two-neuron networks with discrete and distributed delays, Cognit. Comput. 8 (6) (2016) 1103–1118.
[33] J. Lian, X. Zhou, F. Zhang, Z. Chen, X. Xie, G. Sun, xDeepFM: Combining explicit and implicit feature interactions for recommender systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2018, pp. 1754–1763.
[34] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, J. Ma, Neural attentive session-based recommendation, in: Proceedings of the 26th ACM on Conference on Information and Knowledge Management, ACM, 2017, pp. 1419–1428.
[35] M. Quadrana, A. Karatzoglou, B. Hidasi, P. Cremonesi, Personalizing session-based recommendations with hierarchical recurrent neural networks, in: Proceedings of the 11th ACM Conference on Recommender Systems, ACM, 2017, pp. 130–137.
[36] P. Wang, J. Guo, Y. Lan, J. Xu, S. Wan, X. Cheng, Learning hierarchical representation model for nextbasket recommendation, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pp. 403–412.
[37] Z. Li, H. Zhao, Q. Liu, Z. Huang, T. Mei, E. Chen, Learning from history and present: Next-item recommendation via discriminatively exploiting user behaviors, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2018, pp. 1734–1743.
[38] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, B. Xu, Attention-based bidirectional long short-term memory networks for relation classification, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, pp. 207–212.
[39] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, X. Xie, Neural news recommendation with attentive multi-view learning, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019, pp. 3863–3869.
[40] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, X. Xie, NPA: Neural news recommendation with personalized attention, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019, pp. 2576–2584.
[41] J. Tang, C. Deng, G.-B. Huang, Extreme learning machine for multilayer perceptron, IEEE Trans. Neural Netw. Learn. Syst. 27 (4) (2015) 809–821.
[42] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, GroupLens: An open architecture for collaborative filtering of netnews, in: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, ACM, 1994, pp. 175–186.
[43] G. Linden, B. Smith, J. York, Amazon.com recommendations: Item-to-item collaborative filtering, IEEE Internet Comput. 7 (1) (2003) 76–80.
[44] M. Ge, C. Delgado-Battenfeld, D. Jannach, Beyond accuracy: Evaluating recommender systems by coverage and serendipity, in: Proceedings of the 4th ACM Conference on Recommender Systems, ACM, 2010, pp. 257–260.