Where To Go Next?
Exploiting Behavioral User Models in Smart Environments

Seyyed Hadi Hashemi
University of Amsterdam
Amsterdam, The Netherlands
hashemi@uva.nl

Jaap Kamps
University of Amsterdam
Amsterdam, The Netherlands
kamps@uva.nl

Abstract
There is a growing interest in using the Internet of Things (IoT) to create smart environments, which hold the promise to provide personalized experience based on the trail of user interactions with smart devices. We experiment with behavioral user models based on interactions with smart devices in a museum, and investigate the personalized recommendation of what to see after visiting an initial set of Point of Interests (POIs), a key problem in personalizing museum visits or tour guides. We have logged users’ onsite physical information interactions of visits in a museum. Moreover, to have a better understanding of users’ information interaction behaviors and their preferences, we have collected and studied query logs of a search engine of the same collection, and we have found similarities between users’ online digital and onsite physical information interaction behaviors. We exploit user modeling based on users’ different information interaction behaviors and experiment with a novel approach to a critical one-shot POI recommendation using deep neural multilayer perceptron based on explicitly given users’ contextual information, and set-based extracted features using users’ physical information interaction behaviors and similar users’ digital information interaction behaviors. Experimental results indicate that our proposed behavioral user modeling, using both physical and digital user information interaction behaviors, improves the onsite POI recommendation baselines’ performances in all common Information Retrieval evaluation metrics. Our proposed approach provides an effective way to achieve a high precision at rank 1 in onsite critical one-shot POI recommendation problem.

Keywords
Human information interaction; Onsite logs; Behavioral user models; POI recommendation; Internet of things

ACM Reference format:
DOI: http://dx.doi.org/10.1145/3079628.3079687

1 INTRODUCTION
The last decade witnessed a tremendous interests in implementing of Internet of Things (IoT) in different applications [2, 3, 10, 16-18, 25], such as smart shopping malls and smart museums, which provides the infrastructure for understanding users’ physical interaction behavior and consequently their preferences in interacting with smart environments. This prompts the question: How tracking people in their real-life and understanding their interaction behaviors would be helpful? Is it possible to give effective suggestions to users by user tracking using IoT but without getting any explicit information about their preferences like ratings?

Imagine you are at a huge museum like the Louvre Museum in Paris and you want to explore the museum. Usually, it is impossible to visit the whole objects of some big museums like the Louvre Museum in one day. Moreover, museum free roaming is more desirable in comparison to the traditional fixed walking route designed in a non-personalized way. Providing personalized experiences for users is so valuable in this context and will help them to visit all interesting objects of the museum according to their preferences. In this case, how amazing would it be if the contextual recommender system can tell you accurately what to visit without using any history or explicit feedback from you?

Emergence of the above applications leads to rise interests in logging users’ onsite physical information interactions, which creates a new potentially exponentially growing data like search engine query logs. Although understanding users’ search behavior and their information needs based on query logs is well-studied [7, 29], to the best of our knowledge, there is not any study on how to understand users’ behaviors and their information needs based on similarities between users’ onsite physical and online digital information interaction behaviors. Addressing this research problem by
learning a behavioral user model using both onsite physical and online digital user behaviors is our main contribution in this paper.

To this aim, users’ onsite physical interactions of visits in a museum and users’ online query logs of a search engine of the same collection are logged. Onsite physical information interactions are based on unlocking contents of an installed iPod at each POI using RFID tags. Figure 1 shows an example of the museum space with the mentioned installation. In this way, we log users’ interactions with POIs and track users’ visits in the museum.

As it is shown in Figure 2, users behave differently after visiting a set of POIs. Figure 2 plots walk-through graph of 3 real users after checking in at POI1 and POI2. The blue and red paths show walk-through behaviors of two users tend to check-in at POIs one after the other but with different preferences. The green path shows a user who behaves completely different from the other two and does not check-in at POIs one after the other. This figure shows an example of how different are users onsite physical behaviors, which indicates understanding and prediction of users’ onsite physical behaviors are challenging and difficult.

In this paper, our main aim is to study the question: How to model users’ information interaction behavior with IoT having an aim of providing a personalized onsite POI recommendation? Specifically, we answer the following research questions:

1. How to understand users’ onsite physical behavior and create a behavioral user model that is able to effectively predict relevant unseen POIs?
2. How strong are different users’ interaction behaviors with IoT in understanding users’ preferences?
   (a) Are online digital behaviors similar to onsite physical behaviors? Does understanding online digital users’ information interaction behaviors have a positive effect in learning a model to predict unseen relevant POIs and complete users’ personalized onsite visits?
   (b) What are the relative importance of each feature extracted based on different users’ interaction behaviors in effectiveness of POI recommendation systems?
3. How effective is behavioral POI recommendation system in one-shot POI recommendation problem?

The rest of the paper is organized as follows. In Section 2, we review some related work on context aware recommendation and POI recommendation systems. Section 3 is devoted to stating the problem and discussing baselines. Our proposed onsite POI recommendation approach is detailed in Section 4. The experimental setup and results are discussed in Section 5 and 6. Finally, we present the conclusions and future work in Section 7.
2 RELATED WORK

In this section, we discuss related work on context aware recommendation systems, POI recommendation systems, and recommendation systems in museums.

Traditionally, recommendation systems deal with applications having just two types of entities, users and items. However, creation of more complex and realistic applications leads to interests in a new line of research about how to incorporate contextual information as an extra dimension to the recommendation systems. There are 3 ways of incorporating context in the recommender systems: contextual pre-filtering, contextual post-filtering, and contextual modeling [1]. As the later approach is closer to our study in this paper, we will discuss some of the related research in the contextual modeling.

In order to contextually model the context aware recommendation system, Karatzoglou et al. proposed a multiverse recommendation method based on tensor factorization [21], which integrate contextual information by modeling data as User-Item-Context N-dimensional tensor instead of traditional 2-dimensional User-Item matrix. One problem of this method is the data sparseness, which is proportional to the number of defined context in their method. Liu et al. [22] proposed to partition the User-Item matrix by grouping ratings of similar context, which could be helpful to decrease the data sparseness. The other problem of the multiverse recommendation method is that it only works for categorical features. To overcome this problem, Rendle et al. [26] proposed to use factorization machines to model contextual information. The above studies are done to model contextual information, however none of them are really scalable and effective for the recent exponentially growing data.

There have also been lots of studies to solve the POI recommendation problem in both academia and industry [14, 38]. They generally try to adapt traditional recommendation algorithms to the POI recommendation problem. One line of research includes collaborative filtering and matrix factorization approaches in location-based social networks (LBSNs). Berjani et al. in [5] proposed regularized matrix factorization, in which they apply personalized collaborative filtering on dimensionally reduced user-POI matrices to minimize the squared regularized error. In addition to the geographical aspects, they are some researches in POI recommendation that in addition to the geographical dimension, tried to include temporal dimension in the matrix factorization framework [11, 13].

Within the POI recommendation literature, there are some studies that are related to ours in the sense that they studied users’ check-in behavior [24, 27, 30–34, 36, 37]. As three interesting examples of these related works, Zheng et al. proposed collaborative location activity filtering [35]. Particularly, they used collective factorization to recommend locations or activities to users. To this aim, they used comments having GPS data in a web-based GPS management system as a data source. Moreover, Ye et al. in [31] proposed a collaborative POI recommendation algorithm based on geographical influence. To this aim, they used users check in activities in LBSNs. At last, Scholz et al. [27] studied talk attendance prediction in an academic conference using a link prediction approach. To this aim, they logged talk attendance behavior using RFID tags. However, none of the above studies used both the actual users’ onsite physical information interaction behaviors and users’ online digital click-through behaviors.

As another line of related research, there are several researches that study recommender systems for museum visitors. Grieser et al. [12] studied next exhibition recommendation problem in the museum space using visitors history. They applied Naive Bayes learning model using textual description, geospatial proximity and popularity of exhibitions. In their study, popularity baseline, which is one of our defined baseline in this paper, was reported as the most successful next exhibition recommendation model.

Bohnert et al. [6] studied unseen exhibition recommendation using nearest-neighbor content-based filtering approach by taking visitors explicit ratings of exhibitions as inputs. They did the study using 41 museum visitors as participants. Moreover, in a recent work of Bartolini et al. [4], they study recommendation of diverse multimedia data across several web repositories, and arrangement of them in visiting paths. They consider location, number of persons and weather condition as context in their contextual pre-filtering system, and they did the study based on 90 users as participants.

Apart from different recommendation methods being used in the above studies in the museum domain, they are very limited in term of number of participants. In addition, none of them log and study users’ onsite physical information interactions behaviors. In this paper, we log more than 21,000 users’ visits of a museum in a 5 months period, and our proposed model is based on users’ both online digital and onsite physical information interaction behaviors.

Closest in spirit to our work is [17], in which users’ onsite physical behaviors in the existence of a crowd of users have been studied. They studied skip or stay behavior prediction in checking in different POIs as a classification problem. Their study is different from ours as they do not investigate on similarities between users’ physical and digital behaviors. Moreover, we study a POI ranking problem in this paper but they did research on onsite physical interaction behavior classification problem.

3 BACKGROUND AND PRELIMINARIES

In this section, we state the behavioral unseen POI recommendation problem and the best baselines comparable with our proposed model.

3.1 Problem Statement

Let \( U = \{u_1, u_2, ..., u_t\} \subset U^l \) be a subset of users visited a smart environment, \( C_{\text{seen}} = \{c_1, c_2, ..., c_j\} \subset C^l \) a subset of seen or occurred contexts, and \( P_{\text{seen}} = \{p_{1,1}, p_{2,1}, ..., p_{k,1}\} \subset P^k \) a subset of seen POIs. Then, let \( R_{\text{seen}} \subset \mathbb{R}^{P_{\text{seen}} \times C_{\text{seen}}} \) be a user-context-POI matrix containing \( i \) users, \( j \) seen contexts and \( k \) seen POIs. Value \( r_{i,j,k} \in R_{\text{seen}} \) refers to the visit frequency of user \( i \) in context \( j \) to the POI \( k \). In this paper, due to the fact that museum visitors rarely check in to a POI more than once, we have used binary seen or unseen values rather than considering the frequency.

Having above information about users, given a subset of unseen contexts (i.e., \( C_{\text{unseen}} = \{c_1, c_2, ..., c_m\} \subset C^m_{\text{unseen}} \)), and a subset of unseen POIs (i.e., \( P_{\text{unseen}} = \{p_{1,1}, p_{2,1}, ..., p_n\} \subset P^n_{\text{unseen}} \)), the behavioral unseen POI recommendation problem is estimation of \( r_{i,m,n} \in R_{\text{unseen}} \) based on users interaction behaviors with the
seen POIs, in which \( R_{unseen} \in \mathbb{R}^{m \times n} \) is a user-context-POI matrix containing \( i \) users, \( m \) unseen contexts and \( n \) unseen POIs.

### 3.2 Baselines

In this section, baselines created for the evaluation purposes are detailed.

#### 3.2.1 Popularity

The popularity based recommendation ranks POIs candidates according to their popularity scores. The popularity is computed as the number of users who checked in at each POI. The popularity baseline is usually used in evaluation of personalized recommendation systems and it is informed as a very challenging and hard-to-beat baseline [23].

#### 3.2.2 Bias-Based Filtering

As Hashemi et al. discussed in [16], there are some biases in onsite user information interaction logs. They introduce the walk-through position-bias that shows users tend to visit POIs one after the other from check-in to check-out stations. They also observed time-rank bias that indicates users tend to spend less time at the end of exhibitions. Considering these two biases, the probability of checking in at a POI is proportional to the distance from the Check-out station. Therefore, Bias-based baseline ranks POIs based on their distance from the check-out station.

#### 3.2.3 Content-Based Filtering

As descriptions of POIs in museums are well curated, they are very informative source of information that makes the content-based filtering as a very effective baseline in this domain. In this study, each POI contains 3 museum objects with reach descriptions. In order to build a content-based filtering model, we build a profile of each user after visiting a set of POIs using Language Modeling framework. Each profile’s language model is based on all seen objects of \( P_{seen} \).

Since we have profiles of users at each context, KL-Divergence of each unseen POI’s language model and the profile is considered as content-based filtering scores for ranking unseen POIs.

### 4 POI RECOMMENDATION USING USERS’ BEHAVIORS

This section studies how to predict relevant POIs to the given user and context based on users’ interaction behaviors, aiming to answer our first research question: How to understand users’ onsite physical behavior and create a behavioral user model that is able to effectively predict relevant unseen POIs?

In order to model the set-based contextual POI recommendation, we cast the context-aware recommendation problem to a binary classification problem, in which relevant POIs are labeled 1 and irrelevant ones labeled 0. In this way, we try to learn a behavioral model to predict relevant unseen POIs to the given user and context based on the user’s interaction behaviors in the context. Then, relevance probability of POIs to the user and context pairs will be used to rank the unseen POIs. To this aim, a set of features that represent users’ interaction behaviors in given contexts is defined.

#### 4.1 Feature Set

In order to learn an effective model to rank POIs, we have extracted 18 different features. As it is shown in Table 1, we have classified features to 3 sets, namely, explicit context, onsite and online.

Explicit context refers to information explicitly given by users about the context. In our study, we collected users’ gender, their preferred language, their age range and their chosen perspective of the narratives at the exhibition. Previous study on these explicit contexts [16] shows that users behave differently in these different contexts. For example, as it is discussed in [16], children tends to spend less time in front of the POI about the death. Therefore, it seems a reasonable set of features to consider as explicit contexts.

The second group of features is the one gathered onsite without asking users to give any information about their preferences. These features extracted based on users walk-through data. \( f_3 \) is the number of seen POIs, which can be considered as a confidence indicator of some other features’ scores like \( f_6 \). \( f_4 \) is the context-based filtering score of POI candidate based on the profile built using the seen POIs. This content-based filtering score is calculated based on the onsite POI descriptions and users’ onsite interactions. That is why it is considered as one of the onsite features in our feature classification.

In addition to \( f_3 \) and \( f_6 \), we build users’ walk-through graph using their onsite interactions with POIs based on the train set onsite information interaction logs, and calculate \( f_1, f_5, f_6, f_{10}, f_{11}, f_{12} \) and \( f_{13} \) features. Details of these features are available in Table 1.

The third group of features is defined based on an onsite selected POIs using the onsite users’ interactions logs. However, the feature calculation is based on online click-through graph of the museum search engine. Therefore, we classified them as online features. The online click-through graph is filtered to the objects available at onsite POIs. In this study, each onsite POI contains 3 different museum objects. We merge all the objects related to each POI as one node, and the click-through graph’s edges are aggregated from all the edges of POIs’ objects. As a result, same as onsite walk-through graph, the online click-through graph has onsite POIs as nodes. Details of these features are available in Table 1.

#### 4.2 Learning Model

In order to learn a set-based behavioral POI recommendation model, we have implemented a logistic regression classifier and a deep neural multilayer perceptron with dropouts to estimate relevance of each POI to the given user after visiting a set of POIs. The logistic regression classifier and the deep multilayer perceptron have been trained separately based on each group of features extracted using different users’ information interaction behaviors to study which user information interaction behavior is more effective in understanding users’ preferences in their interactions with the IoT in smart environment. In the rest of this section, we will detail the logistic regression and the deep multilayer perceptron implemented for the set-based behavioral POI recommendation.

##### 4.2.1 Logistic Regression

Logistic regression classifier is a linear classifier that transparently helps to understand contribution of each feature in estimation of POIs relevancy. In fact, we would like to know which trained logistic classifier performs better and why. To this aim, we train different logistic regression classifiers based on different feature sets using different users’ interaction behaviors.

In order to learn a logistic classifier, we use variable \( c \in \{0, 1\} \) to show relevance of a POI to a user in a context. Specifically, \( P_\theta(c = 1|u, c, p) \) is the relevance score of the POI \( p \) to the user \( u \).
Table 1: Defined features to predict relevant unseen POIs to users after visiting a set of POIs.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_1</td>
<td>Explicit Context</td>
<td>Gender (e.g., Female)</td>
</tr>
<tr>
<td>f_2</td>
<td>Explicit Context</td>
<td>Language (e.g., English)</td>
</tr>
<tr>
<td>f_3</td>
<td>Explicit Context</td>
<td>Visitor age range (e.g., Adults)</td>
</tr>
<tr>
<td>f_4</td>
<td>Explicit Context</td>
<td>Chosen perspective (e.g., Roman)</td>
</tr>
<tr>
<td>f_5</td>
<td>Onsite</td>
<td>Seen POIs set size</td>
</tr>
<tr>
<td>f_6</td>
<td>Onsite</td>
<td>Content-based relevance score of a POI candidate to a profile created using seen POIs’ content</td>
</tr>
<tr>
<td></td>
<td></td>
<td>that was shown onsite</td>
</tr>
<tr>
<td>f_7</td>
<td>Onsite</td>
<td>Unseen POI’s PageRank in onsite visits walk-through weighted graph built based on a train set</td>
</tr>
<tr>
<td>f_8</td>
<td>Onsite</td>
<td>Unseen POI’s PageRank in onsite visits walk-through unweighted graph built based on a train set</td>
</tr>
<tr>
<td>f_9</td>
<td>Onsite</td>
<td>Unseen POI’s centrality in onsite visits walk-through graph built based on a train set</td>
</tr>
<tr>
<td>f_{10}</td>
<td>Onsite</td>
<td>Minimum distance of the seen set of POIs to the POI candidate in the onsite visits walk-through graph built based on a train set</td>
</tr>
<tr>
<td>f_{11}</td>
<td>Onsite</td>
<td>Median distance of the seen set of POIs to the POI candidate in the online visits walk-through weighted graph built based on a train set</td>
</tr>
<tr>
<td>f_{12}</td>
<td>Onsite</td>
<td>Mean distance of the seen set of POIs to the POI candidate in the onsite visits walk-through graph built based on a train set</td>
</tr>
<tr>
<td>f_{13}</td>
<td>Online</td>
<td>Unseen POI’s PageRank in online click-through weighted graph built based on a train set</td>
</tr>
<tr>
<td>f_{14}</td>
<td>Online</td>
<td>Unseen POI’s PageRank in online click-through unweighted graph built based on a train set</td>
</tr>
<tr>
<td>f_{15}</td>
<td>Online</td>
<td>Unseen POI’s Centrality in Online click-through graph built based on a train set</td>
</tr>
<tr>
<td>f_{16}</td>
<td>Online</td>
<td>Minimum distance of the seen set of POIs to the POI candidate in the online click-through weighted graph built based on a train set</td>
</tr>
<tr>
<td>f_{17}</td>
<td>Online</td>
<td>Median distance of the seen set of POIs to the POI candidate in the online click-through graph built based on a train set</td>
</tr>
<tr>
<td>f_{18}</td>
<td>Online</td>
<td>Mean distance of the seen set of POIs to the POI candidate in the online click-through graph built based on a train set</td>
</tr>
</tbody>
</table>

and the context \( c \), in which \( \theta \) is unknown parameters learned using maximum likelihood estimation (MLE) based on the train set. Given the relevance judgments \( r \) of each POI \( p_k \) to a user \( u_i \) and context \( c_j \) in the train set, the likelihood \( L \) of the train set is as follows:

\[
L = \prod_{i=1}^{U} \prod_{j=1}^{|C|} \prod_{k=1}^{|P_{seen}|} P_{\theta}(r = 1|u_i, c_j, p_k)^{r} P_{\theta}(r = 0|u_i, c_j, p_k)^{1-r},
\]

in which we assume relevance judgments \( r \) are generated independently. We model \( P_{\theta}(r = 1|u_i, c_j, p_k) \) by logistic function on a linear combination of features created based on each specific group of users’ information interaction behaviors. Then, we optimize the unknown parameters \( \theta \) by maximizing the following log likelihood function:

\[
\theta^* = \arg\max_{\theta} \sum_{i=1}^{U} \sum_{j=1}^{|C|} \sum_{k=1}^{|P_{seen}|} r \log P_{\theta}(r = 1|u_i, c_j, p_k)
+ (1-r) \log P_{\theta}(r = 0|u_i, c_j, p_k).
\]

In order to turn the logistic classifier scores to probabilities, we have used the softmax function:

\[
S(y_i) = \frac{e^{y_{i,j}}}{\sum_{j} e^{y_{j}}},
\]

in which \( y_{i,j} \) is the logistic classifier score, and \( S(y_{i}) \) is the output relevance probability of our behavioral POI recommendation model.

At last, we rank unseen POIs based on the logistic classifier output probability of POIs’ relevancy being estimated based on features created using interaction behaviors of a given user in a context.

4.2.2 Deep Neural Multilayer Perceptron. In this subsection, we investigate on a deep neural multilayer perceptron by an aim of improving effectiveness of the POI recommendation to be used in critical one-shot POI recommendation applications. The motivation behind the critical one-shot POI recommendation is that an irrelevant recommendation sometimes has a very negative effect in users’ experience in a way that they might be incorrectly guided to an uninteresting department of a museum that leads to a dissatisfied experience. In this model, for each user in a context, our main goal is to recommend a POI which is highly relevant to them. In the one-shot POI recommendation, we do not care about relevant POIs retrieved after rank 1. In the rest of this section, we detail our deep multilayer perceptron with an aim of improving effectiveness of POI recommendation to be used for the critical one-shot POI recommendation problem.

In order to learn a set based behavioral POI recommendation and learn users’ onsite complicated physical behaviors, we have used a deep neural network with 3 hidden layers having 326 units. To learn an effective model and overcome overfitting problem, we have used a dropout feedforward neural network. Let \( l \in \{1, 2, 3\} \) be the index of the hidden layers of the network. Let \( z^{(l)} \) be the vector of input to layer \( l \) and \( y^{(l)} \) be the vector of outputs from layers.
layer \( l \). The dropout neural network is modelled as follows for any hidden unit \( i \) and \( l \in \{0, 1, 2\} \)\cite{28}:

\[
\begin{align*}
\tilde{r}_i^{(l)} &\sim \text{Bernoulli}(p), \\
\tilde{g}_i^{(l)} &= r_i^{(l)} \ast g_i^{(l)}, \\
\tilde{z}_i^{(l+1)} &= w_l^{(l+1)} \tilde{g}_i^{(l)} + b_l^{(l+1)}, \\
y_i^{(l+1)} &= f(\tilde{z}_i^{(l+1)}),
\end{align*}
\]

where \( r_i^{(l)} \) denotes a vector of independent Bernoulli random variables having probability \( p \) of being 1, \( \tilde{g}_i^{(l)} \) is thinned outputs created by multiplying a sample of \( r_i^{(l)} \) vector by outputs of layer \( l \) (i.e., \( g_i^{(l)} \)) and used as input for the next layer \( l + 1 \), \( w_l^{(l+1)} \) and \( b_l^{(l+1)} \) are weights and biases at layer \( l \), and \( f \) is an activation function, which is rectified linear units (ReLUs) in our setup. This process is done at each layer.

Following many researches in neural network domain, we have used \( p = 0.5 \) in our dropout network. This value is reported as a close to optimal value for a wide range of networks in different applications \cite{28}.

In the learning phase, the derivatives of the loss function are back-propagated through the dropout network. The dropout network is trained using the stochastic gradient descent (SGD) algorithm with mini-batches, which is widely used algorithm for training neural networks. The learning rates are adjusted based on adaptive gradient algorithm (AdaGrad) \cite{8}. In the test phase, the sub-network is used without dropout, but the weights are scaled as \( W_{test}^{(l)} = pW^{(l)} \).

For the classification purpose and having probabilities as outputs, we have used Logistic classifier in the last layer. The logistic classifier in the last layer is trained same as the logistic regression classifier being discussed in previous subsection. The only difference is that, in the logistic classifier being used in the last layer, we model \( P(c = 1|u, i, j, p_k) \) by logistic function on a linear combination of inputs from the last hidden layer units’ outputs. At last, the final relevance probability of \( P(c = 1|u, i, j, p_k) \) is used to rank unseen POIs based on features created using interaction behaviors of a given user in a context.

## 5 EXPERIMENTAL SETUP

In this section, we describe our experimental setup. We first describe the data set used in this paper, and second detail the evaluation methodology used in this study.

### 5.1 Dataset

The dataset of this study is based on onsite and online interaction logs collected at an archeological museum. In this archeological museum, RFID tags are provided as a key to access some additional information about objects being shown in the museum. Users can enter their preferences at the beginning of the museum exhibition in order to personalize the content being shown in all of the POIs. These preferences are perspectives of the narratives, language, gender and the user’s age range.

After checking in, users are free to put their tags on RFID readers of POIs to unlock contents being shown about objects at the POIs. Each POI contains 3 different archeological objects. Users are free to interact with POIs in any order. They can watch short movies, interact with 3D photos of POIs’ objects, or read contents about objects being shown at POIs. At last, users might check out in a summary station, in which they might leave their name, birth date and email. In this paper. 5 months onsite logs of the museum with more than 21,000 sessions is used, which leads to 3,925 high-quality onsite sessions to be used for evaluation purposes.

In addition to the users’ onsite interaction logs, we also collected query logs of the museum search engine. The online features, detailed in Table 1, have been extracted based on 18,001 high-quality sessions created by filtering bot sessions.

### 5.2 Evaluation Methodology

In our collected onsite interaction logs, about 16,000 out of 21,000 sessions either did not have any interactions with POIs or they did not check out at the summary station, and about 1,000 of them had interactions with all the POIs. In order to avoid bias over users who are interested in visiting all or none of the POIs at the museum, we exclude all sessions have checked in at all or none of the POIs at the exhibition. As a result of this preprocessing step, 3,925 out of 21,000 high-quality onsite interaction sessions remains for creating the test collection.

Considering the walk-through graph, for each user in a session and at each checked in POI during their visit, we created a test collection using the seen set of POIs, the user and the explicit contexts as the query and the unseen POIs as the candidates, for which we have judgments based on the user’s session. Basically, we know which POI candidates are visited by the user and consider them as relevant POIs. The rest of the POIs are considered as irrelevant POIs.

Doing the above procedure in building the test collection leads to create a contextual set-based POI recommendation test collection having 1,083,623 judgments. Table 2 shows an example of records created using a user session. To test our proposed model, in order to avoid overfitting, we have done 5-fold cross-validation, in which for each fold as a test set, 3 out of the 4 remained folds randomly sampled and used as a train set, and the remained fold used as a

<table>
<thead>
<tr>
<th>Context</th>
<th>Query Seen POI set</th>
<th>Candidate</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI3</td>
<td>0</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI4</td>
<td>1</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI5</td>
<td>0</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI6</td>
<td>0</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI7</td>
<td>1</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI8</td>
<td>0</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI3</td>
<td>0</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI5</td>
<td>0</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI6</td>
<td>0</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI7</td>
<td>1</td>
</tr>
<tr>
<td>c1</td>
<td>&lt;POI1,POI2&gt;</td>
<td>POI8</td>
<td>0</td>
</tr>
</tbody>
</table>
5.3 Evaluation Metrics

For the evaluation of the defined set-based behavioral POI recommendation task, we cast the problem to a ranking task and use mean reciprocal-rank (MRR), mean average precision (MAP) and R-precision (R-Prec) as metrics that are effective to evaluate proposed models. Moreover, in order to evaluate the one-shot POI recommendation systems, we use precision at rank 1 (P@1) as an evaluation metric.

The MRR is the average of the reciprocal ranks of the first relevant result for a set of queries $Q$ as $MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$, where $\text{rank}$ is the rank of a POI in the ranking. Equation (4) is defined as the average of the $p@n$ values (i.e., $p@n = \frac{\text{# relevant POIs in top } n \text{ results}}{n}$) for all relevant POIs as $AP = \frac{\sum_i^N p@n \times \text{rel}(n)}{R}$, in which $n$ is the rank, $N$ is the number of retrieved POIs candidates, and rel(n) is a binary function indicating the relevance of a given rank. MAP is the mean value of the $AP$s computed for all queries. R-Prec is precision at rank $R$ where $R$ is the number of relevant candidates for the given query. At last, $p@1$ is the precision at rank 1.

6 EXPERIMENTAL RESULTS

In this section, we provide answer to the research questions stated in the introduction section.

6.1 POI Recommendation Using Users’ Information Interaction Behaviors

This section answer our second research question: How strong are different users’ interaction behaviors with IoT in understanding users’ preferences?

To this aim, we have used each of the three groups of features extracted based on each information interaction behaviors to train a POI recommendation system. Specifically, we have trained three different logistic regression classifiers, which are trained based on: 1) the explicit context features (i.e., Logistic Regression-Explicit Context) 2) the onsite features (i.e., Logistic Regression-Onsite) and 3) the online features (i.e., Logistic Regression-Online).

In the rest of this subsection, we first investigate whether users’ online digital interaction behaviors are similar to the users’ onsite physical behavior. Then, we detail relative importance of each feature extracted based on features’ weights being learned by logistic regression classifiers using each type of users’ interaction behaviors with an aim of understanding users’ behaviors.

6.1.1 Onsite Physical Behavior vs. Online Digital Behavior.

We first look at the question: Are online digital behaviors similar to onsite physical behaviors? Does understanding online digital users’ information interaction behaviors have a positive effect in learning a model to predict unseen relevant POIs and complete users’ personalized onsite visits?

In order to answer this research question, we compare POI recommendation systems trained based on each type of interaction behavior. As it is shown in Figure 3, the POI recommendation system trained based on users’ online digital interaction behavior is not only as good as the other POI recommendation systems being trained based on either explicit context or onsite interaction behaviors, but also is performing better than them in terms of all common tested information retrieval metrics. This experiment indicates that availability of the considerable amount of online interaction logs in comparison to onsite interaction logs leads to training an effective onsite POI recommendation system based on users’ online digital interaction behaviors. As we achieve an effective onsite POI recommendation system based on users’ online digital interaction behaviors, we conclude that there is a similarity between onsite physical and online digital information interaction behaviors.

6.1.2 Features Relative Importance in Understanding Users’ Interaction Behaviors. We now look at the question: What are the relative importance of each feature extracted based on different users’ interaction behaviors in effectiveness of POI recommendation systems?

To this aim, we normalize features’ weights being learned in each logistic regression classifier trained for each group of features separately. Then, average of the normalized features’ weights over the 5-fold cross-validation are reported and compared in Figure 4.

As it is shown in Figure 4, among the explicit context interaction, the chosen language (i.e., $f_2$) at the start of museum visits is relatively more important in comparison to other explicit context based features. Moreover, mean distance of the seen POIs to a POI candidate in the onsite visits’ walk-through graph (i.e., $f_2$) has relatively more importance in comparison to other onsite interaction behavior based features. Regarding the online interaction behaviors, median distance of the seen set of POIs to the given candidate in the online click-through graph (i.e., $f_7$) is relatively more important than other online features in the effectiveness of the POI recommendation systems.

6.2 One-Shot POI Recommendation Using Users’ Interaction Behaviors.

This section answer our third research question: How effective is behavioral POI recommendation system in one-shot POI recommendation problem? To this aim, we study effectiveness of the
The main focus of this paper is the study of how to build a behavioral user model for the set-based POI recommendation problem using users’ both onsite and online information interaction behaviors. A study on the strength of using each type of users’ interaction behaviors with IoT in understanding users’ onsite information interaction preferences shows that POI recommendation systems trained using features extracted from a combination of both onsite physical and online digital information interaction behaviors (i.e., online features) performs better than the ones trained by explicitly given context or onsite information interaction behavior. Therefore, we conclude that there is a similarity between onsite physical and online digital interaction preferences that causes an improvement on the onsite POI recommendation effectiveness. Moreover, we have studied the critical one-shot POI recommendation problem. According to our analysis, the learned models based on just basic explicit given contexts or onsite users’ behaviors do not improve the hard-to-beat defined baselines (i.e., popularity and bias-based filtering). However, using a deep multilayer perceptron based on features extracted by online interaction behaviors leads to a significant improvement over the best baseline in all the defined evaluation metrics. Specifically, it has a statistically significant improvement over all baselines with 23% improvement in term of $p@1$ and 11% improvement in term of $MRR$. Therefore, our proposed approach is very effective in critical one-shot POI recommendation. The proposed behavioral user model is very general and can be widely used in any environment with an integrated Internet of Things (IoT) infrastructure. Specifically, in the Cultural Heritage domain, the implemented technology being used in this study and implemented within the European meSch project, is listed in [9, 20] as one of the implemented technologies in museums that provides a more interactive and multisensory experiences for visitors. This is one of the technologies mentioned in museum edition of the NMC Horizon 2015 and 2016 reports as a technology being integrated in museums in four to five years time-to-adoption horizon. As a future work, we are going to increase number of POIs in the museum and see how effective is the proposed behavioral set-based POI recommendation model for the bigger datasets. Moreover, in addition to the evaluation detailed in this paper based on a high volume of real users, we are eager to do a user study to test our proposed behavioral user model. As another line of future work, we will investigate on using recursive neural network to improve our proposed behavioral user models to be used in contextual suggestion problem [15].

**ACKNOWLEDGMENTS**

This research is funded in part by the European Community’s FP7 (project meSch, grant # 600851).
REFERENCES


