Using Interest and Transition Models to Predict Visitor Locations in Museums

Fabian Bohnert [a,*], Ingrid Zukerman [a], Shlomo Berkovsky [a,b], Timothy Baldwin [b], and Liz Sonenberg [b]
[a] *Monash University, Clayton, Victoria 3800, Australia*
E-mail: {fabianb,ingrid}@csse.monash.edu.au
[b] *The University of Melbourne, Parkville, Victoria 3010, Australia*
E-mail: {shlomo,tim}@csse.unimelb.edu.au
E-mail: l.sonenberg@unimelb.edu.au

Museums offer vast amounts of information, but a visitor's receptivity and time are typically limited — providing the visitor with the challenge of selecting the (subjectively) interesting exhibits to view within the time available. Mobile, context-aware computer systems offer the opportunity to improve a visitor's experience by recommending exhibits of interest, and personalising the delivered content. A first step in this process is the prediction of a visitor's activities and interests. In this paper we study non-intrusive, adaptive user modelling techniques that include consideration of the physical constraints of the exhibition layout. We present two collaborative models for predicting a visitor's locations in a museum, and an ensemble model that combines their predictions. These models were trained and tested on a small dataset of museum visits. Our results are encouraging, with the ensemble model yielding the best performance overall.

Keywords: collaborative models, location prediction, museums, physical spaces.

1. Introduction

Museums offer vast amounts of information, but a visitor's receptivity and time are typically limited. The possibility of information overload is evident, as the visitor is confronted with the challenge of selecting the personally interesting exhibits to view within the time available. As a result, s/he might miss out on items of interest. These problems can be addressed by judicious, advance selection of the exhibits to be viewed. However, precise information about the exhibits in different spaces is often not readily available, and also, visitors might change their mind after viewing some exhibits in which they thought they were interested. A personal human guide who is aware of a visitor's interests and time limitations could easily solve all of these problems. However, the provision of such human guides is outside the scope of most museums.

Advances in mobile, context-aware technologies have made electronic handheld guides possible. Similarly to human guides, electronic guides should not have to be explicitly informed about a visitor's interests, but have the potential to infer these interests by observing (tracking) the visitor's behaviour within the museum. The Kubadji project (`http://www.kubadji.org`) is investigating user modelling and language technologies to support the creation of such guides. Key processes to facilitate are to (1) infer a visitor's interests from observing his/her behaviour, (2) make recommendations about items of interest, and (3) personalise the content delivered for these items. The physicality of the domain poses practical challenges for these processes [14]. For example, the spatial layout of the environment influences the curator's decisions about the positioning of the exhibits, and both influence a visitor's decisions about which exhibits to view and in which order. Hence, the spatial arrangement of items is an input that should improve the accuracy of predictions of a visitor's behaviour. To our knowledge, this factor has not been considered to date.

In this paper, we describe a first step of this project, i. e., the prediction of a visitor's interests and locations in a museum on the basis of observed behaviour. Specifically, we consider two collaborative predictive models of visitor behaviour, *Interest* and *Transition*, and an ensemble model that combines their predictions.[1] The *Interest Model* predicts a visitor's next lo-

[*]Corresponding author: Fabian Bohnert, Faculty of Information Technology, Monash University, Clayton, Victoria 3800, Australia. E-mail: fabianb@csse.monash.edu.au

[1]In the future, we also intend to consider content-based models, and their ensemble combination with the collaborative models.

cations (exhibits), on the basis of his/her observed viewing times in the context of the viewing times of other museum visitors. The *Transition Model* predicts locations based on the transitions between exhibits made by other visitors. These models are employed to predict the next K exhibits ($K = 1$ and $K = 3$), using two types of prediction approaches: *set*, which predicts a set of exhibits, and *sequence*, which predicts an ordered sequence.

Our models were trained and tested on a small dataset collected from visitors to the Marine Life Exhibition in Melbourne Museum. Our results show that the *Transition Model* outperforms the *Interest Model*, indicating that the layout of a physical space with homogeneous exhibits (e. g., marine life) is a key factor influencing visitor behaviour.[2] However, the ensemble model yields the best performance overall, demonstrating the importance of considering also a visitor's interests. Its average accuracy for predicting the next exhibit is 68%, and its average accuracy for predicting the next three exhibits is 59%. Our results also indicate that, when predicting the next three exhibits ($K = 3$) to be viewed, a model that predicts a sequence of items has a higher accuracy than a model that predicts a set (59% vs. 49%). Surprisingly, this is not the case when predicting a single exhibit ($K = 1$), where the accuracy of a set-based model is comparable to the accuracy of a sequence-based model (66% vs. 68%).

The rest of this paper is organised as follows. In Section 2, we outline related work. Our collaborative user models are described in Section 3. The probability of visiting exhibits is calculated in Section 4, followed by different models which use these probabilities to predict actual locations (Section 5). In Section 6, we present the results of our evaluation, followed by our conclusions in Section 7.

2. Related Research

Recommender systems are designed to decrease the burden of information overload in situations where the amount of available information vastly exceeds a user's processing capacity. Generalising the initial definition of recommender systems given by Resnick and Varian [23], Burke [5] defines a recommender system as a "system that produces individualised recommendations as output or has the effect of guiding the user in a personalised way to interesting or useful objects in a large space of possible options". Two steps can be identified within the recommendation process: user model construction, and recommendation generation. The user models provide the knowledge source containing explicit information and assumptions about those aspects of the user that are relevant to the system, e. g., preferences and interests, goals or plans, and beliefs or expertise. Acquisition of the user models can be done either explicitly, by requesting input from the user, or implicitly, by observing a user's behaviour or interaction with the system. Later on, the acquired user models are exploited for predicting the user's future interests, preferences, or activities. These predictions form the basis for recommendations.

Although intensive research on recommender systems was initiated by Resnick and Varian [23], the algorithms exploited in state-of-the-art recommender systems were devised earlier [17]. Malone *et al.* [17] discuss the process of preventing irrelevant information from reaching the user, and propose two basic recommendation techniques: (1) cognitive filtering, which is nowadays referred to as content-based filtering [18], and (2) social filtering, which is nowadays known as collaborative filtering [22]. Other popular recommendation techniques, which were added later, include knowledge-based recommendation [5] and demographic filtering [15].

Content-based recommendation techniques make the assumption that a user's previous preferences or interests are reliable indicators for his/her future behaviour. Their main shortcoming is that the features selected when building a content-based model have a substantial effect on the usefulness of this model [29]. In contrast, *collaborative* recommendation techniques base their predictions upon the assumption that users who have agreed in their behaviour in the past, will agree in the future, making predictions based on user-to-user similarities. The greatest strength of collaborative approaches is that they are independent of any representation of the items being recommended, and work well for complex objects such as audio and visual items, where variation in taste is responsible for much of the variation in preferences. However, being dependent on an overlap of ratings across users, collaborative techniques suffer from the *sparsity problem*, i. e., they have difficulty generating recommendations when the space of ratings is sparse [12].

Zukerman and Albrecht [29] and Burke [5] analyse in detail the advantages and shortcomings of the

[2] In the future, we intend to investigate the predictive power of a visitor's interests in a physical space with heterogeneous exhibits, e. g., the entire museum space showing diverse exhibitions.

above modelling techniques. In addition, Burke discusses possible methods for their hybridisation, and he experimentally validates the hypothesis that combining recommendation techniques can improve the accuracy of the generated recommendations. This hypothesis is further validated by Lekakos and Giaglis [16], who explore a number of hybridisation approaches to address the sparsity problem in collaborative filtering algorithms.

Our work lies at the intersection of statistical user modelling and personalised guide systems for physical museum spaces. Personalised guide systems in physical domains have often employed adaptable user models, which require visitors to explicitly state their interests in some form. For example, the *GUIDE* project [7] developed a handheld tourist guide for visitors to Lancaster, UK. It employed a user model obtained from explicit user input to generate a dynamic and user-adapted city tour, where the order of the visited items could be varied. In the museum domain, the *CHIP* project [2] investigates how Semantic Web techniques can be used to provide personalised access to digital museum collections both online and in the physical museum, based on explicitly initialised user models.

Less attention has been paid to predicting preferences from non-intrusive observations, and to utilising adaptive user models that do not require explicit user input. In the museum domain, adaptive user models have usually been updated from the user's interactions with the system, with a focus on adapting content presentation rather than recommending suitable exhibits. For example, the *Hippie/HIPS* project [19] developed an electronic handheld guide for navigating both the physical space and the informational space when visiting a museum. The system provided the user with information personalised and contextualised from diverse sources, including interaction history. Rules were used to reason about a user's interests and knowledge, and a domain hierarchy was exploited for content matching. *HyperAudio* [21] dynamically adapted the presentation content and hyperlinks to stereotypical assumptions about a user, and to what the user has already accessed and seems interested in. The augmented audio reality system for museums *ec(h)o* [11] treated user interests in a dynamic manner, and adapted its user models on the basis of the users' interactions with the system. The collected user modelling data were used to deliver personalised information associated with exhibits via audio display. The *PEACH* project [27] developed a multimedia handheld guide, which adapted its user models both from explicit visitor feedback and implicit observations of visitor interactions with the device, and used the information stored in these user models to generate personalised multimedia presentations.

These systems, like most systems in the museum domain, primarily rely on knowledge-based user models, which require an explicit, a-priori engineered representation of the domain knowledge. In contrast, this work investigates non-intrusive statistical user modelling techniques that do not require an explicit representation of the domain knowledge, and takes into account spatial constraints — a factor that has not been considered to date. As far as we are aware, the only instance of the application of a statistical technique [1,29] for predicting a visitor's behaviour in a museum is described in [10].

3. Building Collaborative User Models using Spatio-Temporal Information

In this work, we consider two collaborative approaches for predicting a visitor's locations: temporal and transitional. The temporal approach predicts a visitor's locations on the basis of the interest in unseen exhibits ascribed to him/her, which in turn is estimated from the time the visitor spent at the exhibits s/he saw. The transitional approach predicts a visitor's locations on the basis of the pathways followed by other visitors to the museum. The application of these models to predict visitor locations is discussed in Section 4.

In order to perform non-intrusive, adaptive user modelling, we process observations acquired without the user's intervention. In this paper, an observation comprises the museum exhibit (location) visited by a user, associated with a time duration — the time that the visitor spent at that exhibit. This information, which was obtained by tracking people manually, is of the same type as information obtainable from instruments in a real-world setting.[3] Thus, for each visitor u, we have an ordered sequence of viewing durations $t_{ui_1}, t_{ui_2}, \ldots$ for items $i_1, i_2, \ldots$ respectively. These data were obtained from 44 visitors to the Marine Life Exhibition in Melbourne Museum, which contains 22 exhibits. In total, our dataset comprises 317 data points (Section 6.1). This is a rather small dataset, in particular for statistical modelling.[4]

[3]The consideration of the impact of instrument accuracy on user models is outside the scope of this work.

[4]In Section 7, we discuss some of the difficulties associated with collecting and processing data in physical settings.

3.1. Interest Model

In an information-seeking context, users are expected to spend more time on relevant than on irrelevant information, as viewing time correlates positively with preference and interest [20]. That is, the time spent at a given exhibit can be used as a measure of interest. However, viewing time is also positively correlated with item complexity. Additionally, viewing times vary over different visitors depending on their available time.[5] Hence, in order to infer the interests of visitors in different items, the observed viewing times cannot be used directly, but must be transformed into a measure that takes these factors into account. To this effect we have devised the relative interest measure, which reflects the interest of a visitor in an exhibit in the context of the time s/he has spent on previously seen exhibits, and the time spent by other visitors on this exhibit. This measure implicitly takes into account item complexity, as complex items are likely to be viewed for a longer time than simpler items.

Definition 1 (Relative Interest (RI))
The *relative interest* of visitor u in exhibit i is calculated as follows.

$$RI_{ui} = \frac{t_{ui}}{\bar{t}_{u\cdot}} - \frac{1}{n_{\cdot i}} \sum_{v \in U} n_{vi} \frac{t_{vi}}{\bar{t}_{v\cdot}} \tag{1}$$

where

t_{ui} is the time visitor u spent at exhibit i,
$\bar{t}_{u\cdot}$ is the average viewing time of visitor u,
$n_{\cdot i}$ is the number of visitors that viewed exhibit i,
U is the set of visitors (users), and

$$n_{vi} = \begin{cases} 1 & \text{if visitor } v \text{ has viewed exhibit } i \\ 0 & \text{otherwise} \end{cases}$$

The first term in Equation 1 reflects a visitor u's viewing time of item i relative to his/her average viewing time, and the second term indicates the average relative viewing time spent at item i (over all the visitors that viewed this item). Hence, RI_{ui} measures whether visitor u is (relative to his/her average viewing time) more or less interested in item i than the average interest in item i.[6]

The collaborative *Interest Model (IM)* is built by calculating RI_{ui}, the relative interest of visitor u in exhibit i, for all visitors $u = 1, \ldots, |U|$ and all items $i = 1, \ldots, |I|$, where $|U|$ is the number of visitors and $|I|$ is the number of exhibits. This yields a relative interest matrix $\mathcal{RI}$ of size $|U| \times |I|$, which contains defined values for all combinations of visitors u and items i that occurred, i. e., combinations referring to an item i viewed by a visitor u. These values, which may be viewed as implicit ratings given by visitors to exhibits, do not take into account the order in which the exhibits were visited. In Section 4.1, we discuss how missing interest values of the active visitor a can be predicted collaboratively from values in $\mathcal{RI}$, and in Section 7 we consider the incorporation of spatial information into our *Interest Model*.

3.2. Transition Model

As mentioned above, the *Interest Model* considers only a visitor's relative interests and does not take into account the order in which the exhibits were visited. Here we describe an alternative model, denoted *Transition Model (TM)*, which considers the visit order.

The *Transition Model* is represented by a stationary 1-stage Markov model, where the transition matrix $\mathcal{TM}$ approximates the probabilities of moving between exhibits. Specifically, the element $\mathcal{TM}(i, j)$ approximates the probability of a visitor going from exhibit i to exhibit j, where $i, j = 1, \ldots, |I|$ and $|I|$ is the number of exhibits. This probability is estimated on the basis of the frequency count of transitions between i and j. In order to overcome the data sparseness problem (which is exacerbated by our small dataset) and to smooth out outliers, we added a flattening constant ε ($= 1/|I|$) to each frequency count. We have chosen this flattening constant, instead of the traditional values of 0.5 or 1 [9], as these larger values would distort the counts obtained from our small dataset.

The *Transition Model* implicitly captures the physical layout of the museum space, i. e., the physical proximity of items, on the basis of the assumption that transitions to spatially close items occur more frequently than movements to items that are further away. However, in the future, we will also experiment with spatial models that represent more directly the distance between exhibits (Section 7).

4. Using Collaborative Models to Predict Location Probabilities

In this section, we describe how our *Interest Model* and *Transition Model* are used to estimate the probability of a user visiting a particular exhibit from his/her

[5] Viewing time was also found to be negatively correlated with familiarity, positively correlated with novelty, and decreases from beginning to end within a sequence of stops [20]. However, these factors are not yet considered in our models.

[6] Other measures of interest are possible. For instance, Bohnert and Zukerman [4] explored a different variant of relative interest, which was slightly outperformed by the measure presented here.

observed visit so far, and propose an ensemble approach [16] to combine the predictions generated by these models (called *weighted hybridisation* in [5]). The use of these estimations to predict actual locations is described in Section 5.

4.1. Interest Model

In Section 3.1, we used viewing time to build a model that estimates a visitor's interest in viewed exhibits relative to his/her own viewing patterns and the interests of other visitors. Here we employ this model to estimate the probability of visiting unseen exhibits.[7] To this effect, we first collaboratively predict missing relative interest values for the active visitor a from the relative interest matrix $\mathcal{RI}$, and then use these values to estimate the probability of visiting unseen exhibits (Algorithm 1).

Step 1: Estimating $\widehat{RI}_{ui}^{\text{visit}}$, a visitor's relative interest in visited items.
The estimation of $\widehat{RI}_{ui}^{\text{visit}}$ uses Equation 1 in Definition 1 (Section 3.1). Note that for the active user, the viewed items are those that s/he has viewed so far during his/her ongoing visit, while for the rest of the visitors, the viewed items are those viewed during their entire visit.

Step 3: Finding a set of item mentors for the active visitor.
The *item mentors* of a visitor are the visitors that have viewed item i and whose relative interests are most similar to those of the active visitor with respect to the commonly viewed items.

Definition 2 (Mentor)
Let $M_a = \{u \in U, u \neq a : \sum_{i \in I} n_{ai} n_{ui} \geq K\}$ be the set of *mentors* of the active visitor a, i. e., the set of visitors who have visited at least K (= 3) items that the active visitor has viewed (n_{ui} is defined in Definition 1).

Definition 3 (Item Mentor)
Let $\tilde{M}_{ai} = \{u \in M_a : s_{au} n_{ui} > 0\}$ be the set of *item mentors* of the active visitor a for item i. Following [8], we have limited $\tilde{M}_{ai}$ to at most $\tilde{K}$ item mentors which are most similar to the active visitor. We use $\tilde{K} = 15$.

[7] Although visitors sometimes return to previously viewed exhibits, our observations indicate that this barely happens. Hence, we focus on unseen exhibits.

Algorithm 1 Estimating the probability of visiting an unseen exhibit

1: Estimate from the observed viewing times the relative interests of all visitors — including the active visitor a — in the exhibits viewed during their visit.
2: **for all** i such that i is an unvisited exhibit **do**
3: Find a set of *item mentors*, who have viewed item i, and whose relative interests are most similar to those of the active visitor.
4: Estimate the active visitor's relative interest in item i as the weighted mean of the relative interests of his/her item mentors in i, where the weights are visitor-to-mentor similarities.
5: **end for**
6: Calculate the probabilities of the active visitor visiting unseen exhibits.

The similarity s_{au} between the active visitor a and visitor u is calculated using *Pearson's correlation coefficient*, which expresses the linear correlation of the ratings of two visitors [22,25].[8]

$$s_{au} = \begin{cases} \dfrac{c_{au}}{s_a\, s_u} & \text{if denominator} > 0 \\ 0 & \text{otherwise} \end{cases}$$

where

c_{au} = $\frac{1}{n-1} \sum_{i \in I} n_{ai} n_{ui} (RI_{ai} - \overline{RI}_{a\cdot}^{u})(RI_{ui} - \overline{RI}_{u\cdot}^{a})$ is the covariance of the relative interests of the commonly rated items of visitors a and u,

s_a = $\sqrt{\frac{1}{n-1} \sum_{i \in I} n_{ai} n_{ui} (RI_{ai} - \overline{RI}_{a\cdot}^{u})^2}$ is the standard deviation of all ratings RI_{ai} of the active visitor a for which corresponding ratings were also given by visitor u (s_u is defined similarly),

n = $\sum_{i \in I} n_{ai} n_{ui}$ is the number of commonly rated items,

I is the set of exhibits, and

$\overline{RI}_{a\cdot}^{u}$ is the arithmetic mean of all ratings RI_{ai} of the active visitor a for which corresponding ratings were also given by visitor u ($\overline{RI}_{u\cdot}^{a}$ is defined similarly).

Step 4: Estimating $\widehat{RI}_{ai}$, the active user's relative interest in unvisited items.
We estimate $\widehat{RI}_{ai}$ as the weighted mean of the relative interests of visitors $u \in \tilde{M}_{ai}$ [25], i. e., visitors who

[8] Any consistent measure that is large for close visitors and small for distant visitors can be used to measure the similarities.

are item mentors of the active visitor a for item i. The weights are set to the similarities between these visitors and the active visitor. If no item mentors are found for visitor a and item i (cold-start problem [12,29]), we use the (non-personalised) arithmetic mean $\overline{RI}_{\cdot i}$ of the relative interests in item i to estimate $\widehat{RI}_{ai}$.[9]

$$\widehat{RI}_{ai} = \begin{cases} \dfrac{\sum_{u \in \tilde{M}_{ai}} s_{au} RI_{ui}}{\sum_{u \in \tilde{M}_{ai}} s_{au}} & \text{for } \tilde{M}_{ai} \neq \emptyset \\ \overline{RI}_{\cdot i} \ (= 0) & \text{for } \tilde{M}_{ai} = \emptyset \end{cases} \tag{2}$$

Step 6: Calculating the probabilities of visiting unseen exhibits.

Given a visit where a visitor a has viewed k items so far, the probability of the $(k+1)$-th item being item i is represented by the expression $\Pr(X_{k+1} = i \mid \boldsymbol{v}_a^k)$, where $\boldsymbol{v}_a^k$ is visitor a's visit history. Using our *Interest Model* to approximate this expression yields the following formula.

$$\Pr\left(X_{k+1} = i \mid \boldsymbol{v}_a^k\right) \approx \Pr_{IM}\left(X_{k+1} = i \mid \boldsymbol{t}_a^k\right)$$

where $\boldsymbol{t}_a^k$ is the time component of the visit history $\boldsymbol{v}_a^k$ (the *Interest Model* depends on viewing times, rather than transitions between locations). The calculation of this probability from the estimated relative interests in unseen items is done by normalising the relative interests to the interval $[0, 1]$ as follows.

$$\Pr_{IM}\left(X_{k+1} = i \mid \boldsymbol{t}_a^k\right) = \frac{f\left(\widehat{RI}_{ai}\right)}{\sum_{j \in I \setminus I_a^k} f\left(\widehat{RI}_{aj}\right)}$$

where $I \setminus I_a^k$ is the set of exhibits not yet visited by the active visitor, and f is a linear transformation that ensures that $f(\widehat{RI}_{ai}) \geq 0$ for all unvisited exhibits.

4.2. Transition Model

In contrast to the *Interest Model*, the *Transition Model* described in Section 3.2 depends on transitions between locations, rather than viewing times. Thus, employing the *Transition Model* to approximate the probability that the $(k+1)$-th exhibit viewed by the active visitor a is item i, we obtain the formula

$$\Pr\left(X_{k+1} = i \mid \boldsymbol{v}_a^k\right) \approx \Pr_{TM}\left(X_{k+1} = i \mid I_a^k\right)$$

where I_a^k are the exhibits visited by the active visitor.

Since our *Transition Model* is a 1-stage Markov model, the probability of the next exhibit being item i is further approximated by

$$\begin{aligned} \Pr_{TM}\left(X_{k+1} = i \mid I_a^k\right) &\approx \Pr_{TM}\left(X_{k+1} = i \mid X_k = i_k\right) \\ &= \mathcal{TM}(i_k, i) \end{aligned}$$

where i_k is the current item.

As mentioned above, our observations indicate that visitors rarely return to previously viewed exhibits. Hence, prior to calculating these probabilities, we set to 0 the entries of $\mathcal{TM}$ that correspond to the visited items (items in I_a^k), and appropriately renormalise the rows.

4.3. Combining Interest Model and Transition Model

As indicated above, the predictions made by the *Interest Model* are based on temporal information, while the predictions made by the *Transition Model* implicitly capture spatial information. Additionally, while the *Interest Model* adapts to the behaviour of a visitor, the *Transition Model* is not personalised. In this section, we propose an ensemble model [5,16] called *Hybrid Model (HM)* that combines the predictions made by these models, thereby jointly taking into account transition and temporal information.

We use the probabilities generated by our ensemble model to approximate $\Pr(X_{k+1} = i \mid \boldsymbol{v}_a^k)$ as follows.

$$\Pr\left(X_{k+1} = i \mid \boldsymbol{v}_a^k\right) \approx \Pr_{HM}\left(X_{k+1} = i \mid \boldsymbol{v}_a^k\right)$$

This probability in turn is calculated by means of a weighted average of the predictions generated by our *Interest Model* and our *Transition Model*.

$$\begin{aligned} &\Pr_{HM}\left(X_{k+1} = i \mid \boldsymbol{v}_a^k\right) \\ &= \quad \omega \ \Pr_{IM}\left(X_{k+1} = i \mid \boldsymbol{t}_a^k\right) \\ &\quad + (1 - \omega) \Pr_{TM}\left(X_{k+1} = i \mid I_a^k\right) \end{aligned}$$

where $0 \leq \omega \leq 1$. We experimented with different values for ω, with the assignment $\omega = \beta/(\alpha + \beta)$ yielding the best performance,[10] where

$$\alpha = \min_{i \in I \setminus I_a^k} \Pr_{IM}\left(X_{k+1} = i \mid \boldsymbol{t}_a^k\right), \text{ and}$$

$$\beta = \min_{i \in I \setminus I_a^k} \Pr_{TM}\left(X_{k+1} = i \mid I_a^k\right)$$

This choice of ω is motivated as follows. Our observations showed that the minima α and β are often outliers of the probability distributions generated by the *Interest Model* and *Transition Model* respectively. To reduce the influence of these outliers on the predictions generated by our models, we divide the probabilities produced by the *Interest Model* and *Transition Model*

[9] By definition, $\overline{RI}_{\cdot i} = 0$ for our measure of relative interest.

[10] In the future, we propose to apply machine learning techniques to learn the optimal value of ω.

by α and β respectively, compute their sum for each unvisited exhibit, and renormalise the resultant values to probabilities, so that their sum over all unvisited exhibits adds up to 1. This amplifies the predicted probabilities relative to the outliers, and gives more weight to the model with the smaller minimum.

5. Building Predictive Models

In this section, we describe two approaches for using the probabilities estimated in Section 4 to predict the next K exhibits to be viewed by a visitor: *TopK* predicts the next K items as a ranked set, and *SequenceK/N* predicts the next K items as the initial portion of a sequence of N items.

5.1. TopK Prediction

This approach assumes that the current history of the active visitor a is sufficient to predict his/her future behaviour, and that it is unnecessary to consider the impact of future transitions on a visitor's subsequent behaviour. Hence, in order to predict the next K items to be visited (having visited k items), we find the set of K unvisited items $i_{k+1}, \ldots, i_{k+K}$ which maximises the product of their visit probabilities.

$$\underset{i_{k+1},\ldots,i_{k+K} \in I \backslash I_a^k}{\arg\max} \prod_{m=1}^{K} \Pr\left(X_{k+1} = i_{k+m} \mid \boldsymbol{v}_a^k\right)$$

This approach is equivalent to computing the probabilities $\Pr(X_{k+1} = i \mid \boldsymbol{v}_a^k)$ for all unvisited exhibits $i \in I \backslash I_a^k$ (pretending that each of these exhibits is the next one — hence the subscript $k+1$), then sorting these items in descending order of their visit probability, and selecting the top K items.

5.2. SequenceK/N Prediction

In contrast to the *TopK* approach, the *SequenceK/N* approach assumes that future transitions influence a visitor's subsequent behaviour. Hence, in order to predict the next K items to be visited (having visited k items), we find the maximum-probability sequence of N unvisited items $i_{k+1}, \ldots, i_{k+N}$, and then select the first K items $i_{k+1}, \ldots, i_{k+K}$ within this sequence.

$$\underset{i_{k+1},\ldots,i_{k+N} \in I \backslash I_a^k}{\arg\max} \Pr\left(X_{k+1} = i_{k+1}, \ldots, X_{k+N} = i_{k+N} \mid \boldsymbol{v}_a^k\right)$$

This probability is decomposed as follows, assuming that X_{k+m} only depends on the past.

$$\begin{aligned} &\Pr\left(X_{k+1} = i_{k+1}, \ldots, X_{k+N} = i_{k+N} \mid \boldsymbol{v}_a^k\right) \\ &= \prod\nolimits_{m=1}^{N} \Pr\left(X_{k+m} = i_{k+m} \mid \boldsymbol{v}_a^{k+m-1}\right) \end{aligned} \tag{3}$$

Due to this decomposition, the joint probability in Equation 3 can be maximised by recursively spanning a search tree of depth $N-1$, and performing an exhaustive search for a maximising path from its root to one of the leaves.

5.3. Predicting Viewing Times

The probability $\Pr(X_{k+m} = i_{k+m} \mid \boldsymbol{v}_a^{k+m-1})$ in Equation 3 depends on the active user's visit history up to exhibit i_{k+m-1}, but in practice this history is available only up to item i_k. Future exhibits can be incorporated into a "potential history" for the *Transition Model* by iteratively adding possible unseen exhibits to construct different possible future sequences. However, in order to incorporate such a history into the *Interest Model* (and hence the *Hybrid Model*), we also need to predict viewing times.

We denote the active visitor a's estimated viewing time of exhibit i given his/her visit history $\boldsymbol{v}_a^k$ by $\hat{t}_{ai \mid \boldsymbol{v}_a^k}$, and calculate it as follows. This calculation is similar to that performed in Equation 2.

$$\hat{t}_{ai \mid \boldsymbol{v}_a^k} = \begin{cases} \bar{t}_{a\cdot} \times \dfrac{\sum_{u \in \tilde{M}_{ai}^{(t)}} s_{au} \dfrac{t_{ui}}{\bar{t}_{u\cdot}}}{\sum_{u \in \tilde{M}_{ai}^{(t)}} s_{au}} & \text{for } \tilde{M}_{ai}^{(t)} \neq \emptyset \\ \bar{t}_{a\cdot} \times \dfrac{1}{n_{\cdot i}} \sum_{u \in U} n_{ui} \dfrac{t_{ui}}{\bar{t}_{u\cdot}} & \text{for } \tilde{M}_{ai}^{(t)} = \emptyset \end{cases}$$

where t_{ui}, $\bar{t}_{u\cdot}$, n_{ui} and $n_{\cdot i}$ are defined as in Definition 1, and s_{au} is defined as in Definition 3. Further, $\tilde{M}_{ai}^{(t)} = \{u \in \tilde{M}_{ai} : t_{ui} > 0\}$ is the set of visitor a's item mentors with positive viewing times t_{ui} for item i.

6. Evaluation

In this section, we discuss our experimental evaluation. The experimental setting is described in Section 6.1, and our evaluation measures in Section 6.2. Our results are presented in Sections 6.3 and 6.4, and summarised in Section 6.5.

6.1. Experimental Setting

Thanks to a cooperation with Museum Victoria, we could access data from several observational studies conducted at Melbourne Museum. For our evaluation, we used the data collected at the Marine Life Exhibition in 2006. This exhibition comprises 56 exhibits in four sections displaying considerably homogeneous marine-related topics. Figure 1 depicts the layout of the exhibition space, including the exhibition high-

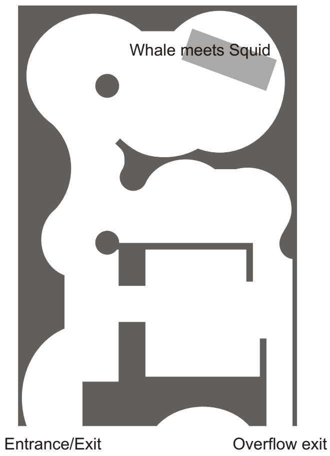


Fig. 1. Layout of the Marine Life Exhibition

light "Whale meets Squid". To collect the dataset used in this study, 44 visitors were observationally tracked within the exhibition. Their pathways and viewing times for exhibits were manually recorded, which allowed us to reconstruct time-annotated visitor pathways. The overall number of viewed exhibits is 401, such that the average pathway contains 9.11 exhibits.

According to the museum's classification, the exhibits are partitioned such that every item on display (e. g., an exhibit and a panel describing it) are considered separate exhibits. This partitioning may skew the results, as the events of viewing such exhibits are highly correlated. Hence, museum curators helped us unify logically related exhibits into grouped exhibits. This transformed the original set of 56 exhibits into a smaller set of 22 grouped exhibits. For the experimental evaluation, we totalled the viewing times for the individual exhibits that belong to the same grouped exhibit. In summary, we obtained 317 observations, such that the length of the average visitor pathway is 7.20 exhibits, and the shortest and longest visits comprise 3 and 16 exhibits respectively. Table 1 shows the distribution of the pathway lengths among the visitors.

As described in Section 5, predictions were generated in two modes: *TopK*, which predicts a ranked set of K exhibits, and *SequenceK/N*, which predicts an ordered sequence of K exhibits (in fact, the first K exhibits in a sequence of N predicted exhibits). In our experiments, we evaluated the performance of the proposed approaches for two values of K ($K = 1$ and $K = 3$) and a fixed value for N ($N = 3$), yielding the four variants *Top*1, *Sequence*1/3, *Top*3 and *Sequence*3/3. For every combination of prediction mode (set or sequence) and value of K ($K = 1$ or $K = 3$), we considered the three prediction models defined in Section 4: *Interest Model (IM)*, *Transition Model (TM)* and *Hybrid Model (HM)*. Hence, we evaluated a total of 12 variants.

Table 1
Distribution of pathway lengths

stops	3-4	5-6	7-8	9-10	11-12	13-14	15-16
users	11	8	12	6	3	3	1

Due to the small size of our dataset, we used leave-one-out evaluation, i. e., we trained our prediction models on 43 of the 44 visitors in our dataset, and tested them on the remaining visitor (the active visitor). Additionally, since our research does not focus on the cold-start problem of recommender systems [12,29], we considered only the portion of a museum visit for which sufficient data could be gathered about a visitor to support the construction of a collaborative *Interest Model* (i. e., to compute the visitor's similarity with the other visitors). Hence, we disregarded the results obtained for the initial stages of the visit, and report on the results obtained only after at least three observations have been made for the active visitor. Additionally, to be able to compute values for our evaluation measures (Section 6.2), we also disregarded the final three observations of the visit. Finally, to obtain statistically valid results, we considered only visit percentages where at least 10 users were observed. Owing to these considerations, the results presented in this paper pertain to the middle part of a museum visit, spanning between 25% and 70% of the visit.[11]

6.2. Evaluation Measures

We applied two types of measures to evaluate the performance of the proposed approaches: *Classification Accuracy* (*CA*) and *Ranking Accuracy* (*RA*) [13]. For each percentage of a visit, we averaged the value of these measures for all the active visitors in the test set (we considered visit percentages, rather than the actual number of viewed exhibits, because different visitors have visit histories of different lengths). Clearly, the percentage of a visit reflects the amount of information available about a visitor, which increases as the visit proceeds, thereby adding evidence to the *Interest Model* of the visitor.

[11] In the future, we intend to consider different models that address starting and ending conditions, and to apply machine learning techniques to determine the points in a visit where these models can be deployed.

6.2.1. Classification Accuracy

We use Classification Accuracy to verify whether the predicted exhibits were actually viewed by the visitor. Given a set (or sequence) $\mathcal{K}$ of the next K predicted exhibits, and a set (or sequence) $\mathcal{M}$ of the next M exhibits that were actually viewed, $CA(K,M)$ is computed by

$$CA(K,M) = \frac{|\mathcal{K} \cap \mathcal{M}|}{\min\{K, M\}} \tag{4}$$

We used four main variants of $CA(K,M)$ to evaluate the predictions generated by our models, two for $K = 1$ and two for $K = 3$, as follows.[12]

- $CA(1,all)$ – Was the first predicted exhibit viewed during the remainder of the visit?
- $CA(1,1)$ – Was the first predicted exhibit actually viewed immediately?
- $CA(3,all)$ – What proportion of the three predicted exhibits was viewed during the rest of the visit?
- $CA(3,1)$ – Were any of the three predicted exhibits viewed immediately?

The first three measures, where $K \leq M$, are analogous to the precision measure used in Information Retrieval [24]. In agreement with Herlocker *et al.*'s observations regarding the impracticality of using traditional recall in recommender systems [13], we eschew the calculation of recall. The reason for this is that due to the large number of exhibits left to be viewed at most stages of a visit (i. e., $M \gg K$), the settings $(1, all)$ and $(3, all)$ would yield low recall values, which are not comparable to the precision values.

Note that for any value of K, the values of $CA(K,M)$ decrease monotonically as M decreases. For instance, $CA(1,all)$, which indicates whether the first predicted exhibit was viewed during the remainder of the visit, is higher than $CA(1,1)$, which indicates whether the predicted exhibit was viewed next.

6.2.2. Ranking Accuracy

Unlike the above Classification Accuracy measures, which measure inclusion or exclusion of the predicted exhibits in the list of the viewed exhibits, the Ranking Accuracy measures reflect the accuracy of the ranking in a ranked set or a sequence of exhibits. These measures are calculated by comparing the ranking of exhibits in the predicted sequence with the actual order of exhibits viewed by the visitor in the remainder of the visit. We implemented three Ranking Accuracy measures: RA_{pred}, RA_{real}, and mSP.

- RA_{pred} represents the ranking of the first predicted exhibit in the sequence of exhibits actually viewed by a visitor, normalised to $[0, 1]$ such that a ranking of 1 results in a value of 1.

$$RA_{pred} = 1 - \frac{rank_{pred} - 1}{|\mathcal{M}|} \tag{5}$$

  where $\mathcal{M}$ denotes the sequence of exhibits actually viewed by the visitor in the remainder of the visit, and $rank_{pred}$ denotes the ranking of the first predicted exhibit within this sequence. If the first predicted exhibit is not viewed by the visitor in the remainder of the visit at all, $rank_{pred}$ is set to a value of $|\mathcal{M}| + 1$.
- RA_{real} represents the ranking of the next exhibit actually viewed by a visitor in the sequence of the predicted exhibits, normalised to $[0, 1]$ such that a ranking of 1 leads to a value of 1. RA_{real} is computed as follows.

$$RA_{real} = 1 - \frac{rank_{real} - 1}{|\mathcal{K}|} \tag{6}$$

  where $\mathcal{K}$ denotes the list of predicted exhibits, and $rank_{real}$ denotes the ranking of the next viewed exhibit within this sequence. If the next viewed exhibit does not appear in $\mathcal{K}$, $rank_{real}$ is assigned the penalty value $|\mathcal{K}| + 1$.
- mSP, a modified version of Spearman's rank correlation, evaluates the quality of a predicted sequence as a whole, rather than just the position of the next viewed exhibit or the first predicted exhibit. The original form of Spearman's rank correlation measures the correlation between two sequences of equal length K that contain the same items as follows [26].

$$\rho = 1 - \frac{6 \sum_{i=1}^{K} {d_i}^2}{K\,(K^2 - 1)} \tag{7}$$

  where d_i represents the difference between the ranks (positions) of item i in the two sequences. However, this form cannot be directly used in our setting, because the length K of the predicted sequence is usually different from the length M of the remaining visit, and one sequence may contain exhibits that do not occur in the other. In addition, in our domain, the correlations between predicted and actual items that appear earlier in

[12] We also considered $CA(1,3)$ and $CA(3,3)$, but the results obtained for these variants were consistent with those obtained for $CA(1,all)$ and $CA(3,all)$ respectively, and hence do not merit a separate discussion.

the sequences are more important than the correlations between items that appear later in the sequences. To take into account these factors, we modify the computation of the rank differences d_i in Spearman's rank correlation as follows.[13]

$$d_i = \begin{cases} \log(p_i) - \log(r_i) & \text{if } 1 \leq r_i \leq K \\ \log(K + \gamma) & \text{if } K < r_i \leq M \\ \log(K + \delta) & \text{if } r_i \text{ undefined} \end{cases}$$

where $1 \leq p_i \leq K$ is the rank of the i-th item in the predicted sequence, and $1 \leq r_i \leq M$ is the rank of the i-th item in the remaining visit. This calculation penalises the correlation value with $\log(K + \gamma)$, where $\gamma = 1$, if an item i that was included in the predicted sequence is actually viewed by the visitor beyond the K-th position. The penalty increases to $\log(K + \delta)$, where $\delta = 2$, if an item i is not viewed at all in the remainder of the visit.[14] Finally, we perform a linear transformation of the modified Spearman's rank correlation to obtain values in the interval $[0, 1]$.

Note that the values of RA_{real} and RA_{pred} are not directly comparable: RA_{real} is computed with respect to the predefined length of the predicted pathway (i. e., $|\mathcal{K}| = K$), whereas RA_{pred} is computed with respect to the length of the pathway in the remainder of the visit (i. e., $|\mathcal{M}| = M$). Since $M \geq K$, the fraction's denominator in the two cases will almost always differ. However, RA_{pred} is a multi-valued variant of the binary measure *CA(*1*,all)*, and RA_{real} is a multi-valued variant of *CA(K,*1*)*.

6.3. Evaluation for $K = 1$

This section presents the results of the evaluation for $K = 1$, i. e., the prediction of the next single exhibit to be viewed by the active user. As described earlier, we consider two ways of performing this task: predicting an individual exhibit with the highest probability of being visited (*Top*1), and predicting the first exhibit in a maximum probability sequence of three exhibits (*Sequence*1/3). For each percentage of the visit, we generated predictions using our three prediction models: *IM*, *TM* and *HM*. Our predictions were evaluated using the Classification Accuracy measures *CA(*1*,all)* and *CA(*1*,*1*)* and the Ranking Accuracy measures RA_{pred} described in Section 6.2 (we do not use RA_{real} and *mSP* as only one exhibit is predicted).

6.3.1. Classification Accuracy

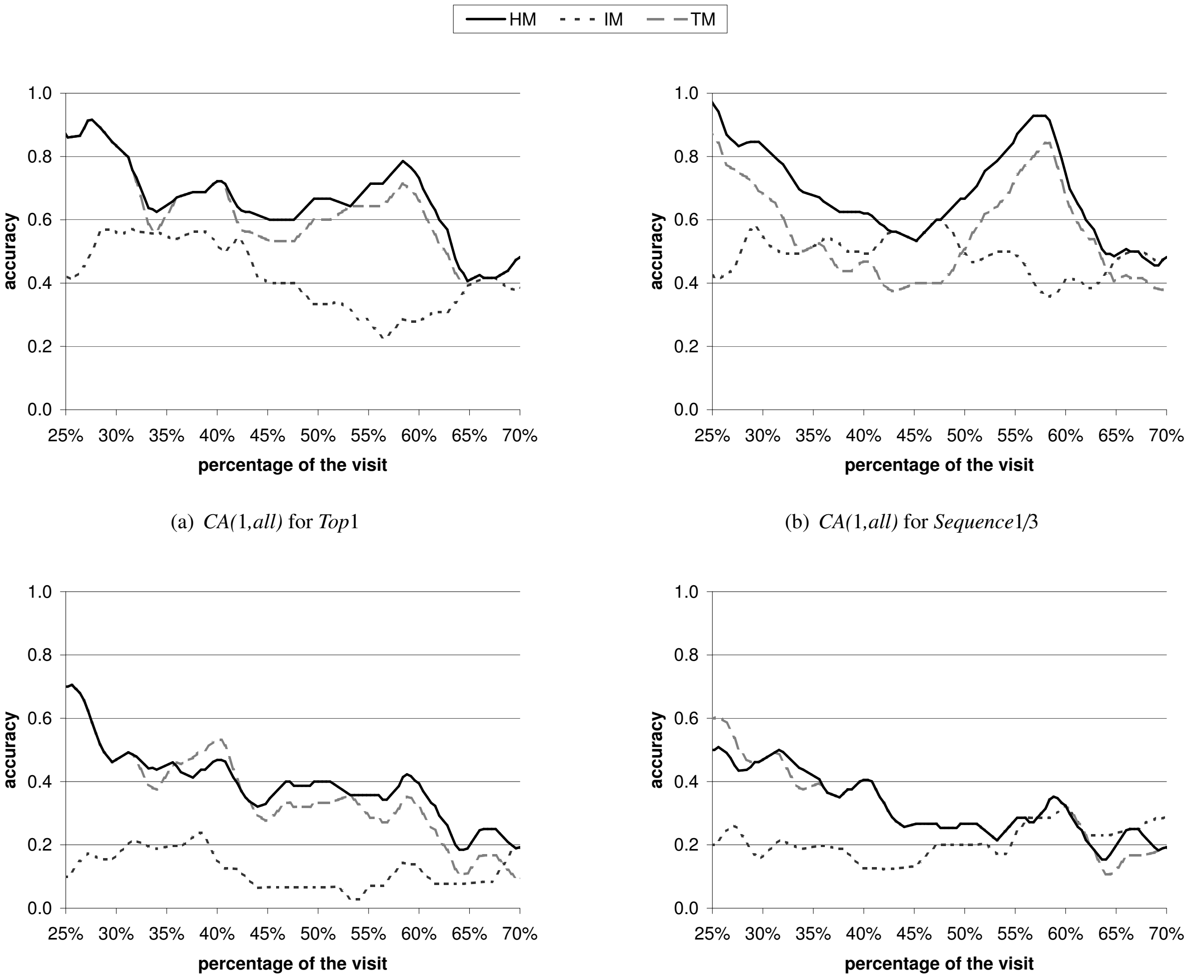
In this section, we present the results for *CA(*1*,all)* and *CA(*1*,*1*)*. Figures 2(a) and 2(b) depict *CA(*1*,all)* for the *Top*1 and *Sequence*1/3 modes of operation respectively, and Figures 2(c) and 2(d) depict *CA(*1*,*1*)* for *Top*1 and *Sequence*1/3 respectively. Figures 2(a) and 2(b) show that the accuracy of *HM* equals or exceeds that of *TM*, which in turn is mostly superior to the accuracy of *IM*. The difference between *HM* and *IM* is statistically significant for both *Top*1 and *Sequence*1/3 ($p < 0.05$).[15] The difference between *HM* and *TM* is statistically significant with $p < 0.1$ for *Sequence*1/3, but it is not significant for *Top*1. The results shown in Figures 2(c) and 2(d) are not as clear cut, since the accuracy of *HM* is occasionally lower than that of *TM* and even *IM*. The performance difference between *HM* and *TM* is not statistically significant, but the difference in performance between *HM* and *IM* is statistically significant ($p < 0.05$) for visit percentages up to about 50%. In summary, despite the relatively low predictive accuracy of *IM*, the overall superior performance of *HM* shows that a visitor's predicted interests should also be taken into account.

The dominance of *TM* over *IM* can be explained by the joint influence of two factors: the structure of the physical space and the homogeneity of the exhibition. The structure of the space constraints the accessible exhibits by means of walls, signs and written guidance. In addition, the Marine Life Exhibition is highly homogeneous, thereby reducing the impact of a visitor's interests. The influence of the structure of the space (Figure 1) also explains the divergent predictive performance of *TM* and *IM* at different portions of the visit. When the space is highly constrained, the performance of *TM* improves, as there are only a few exhibits from which to choose, while the performance of *IM* deteriorates, as interest can only have a limited effect on a visitor's behaviour. In contrast, when the space is less constrained, the opposite tends to happen. Particular instances of this general observation are the initial stage of the visit, and a period of time around 55%-60% of the visit. In the initial stage, visitors pass through a long entrance area where *TM* makes accurate predic-

[13] See [28] for an alternative formulation of Spearman's rank correlation for top-N evaluation that is compatible with the setting of $K = M$.

[14] In the future, we intend to consider penalty values that reflect more accurately how late in the visit an exhibit is viewed, if at all.

[15] Throughout this paper, the statistical tests performed are paired two-tailed t-tests. Also, we consider $p \geq 0.1$ to indicate a lack of statistical significance. We use this lenient p-value due to the small size of our dataset.



(a) *CA(*1*,all)* for *Top*1

(b) *CA(*1*,all)* for *Sequence*1/3

(c) *CA(*1*,*1*)* for *Top*1

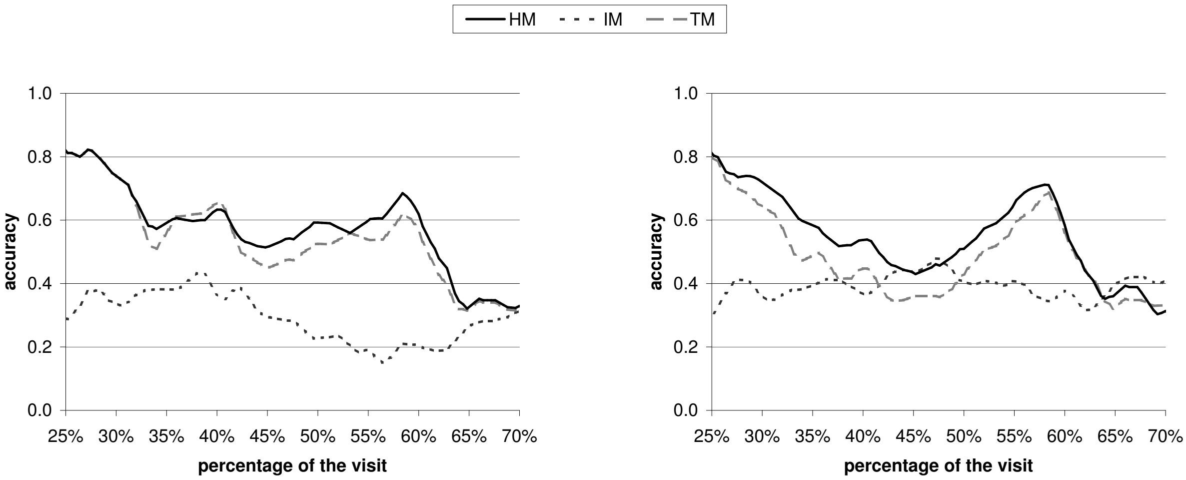
(d) *CA(*1*,*1*)* for *Sequence*1/3

Fig. 2. Comparison of *Top*1 and *Sequence*1/3 with respect to Classification Accuracy

tions. This area leads to a space with several smaller rooms where visitor behaviour is harder to predict. Figures 2(a) and 2(b) show a large peak in the predictive accuracy of *TM* (and *HM*) for the period of time around 55%-60% of the visit (and a corresponding drop for *IM*). Analysis of visitor trajectories revealed that during this part of their visit, people enter the area from which the "Whale meets Squid" exhibit is visible (this is arguably the highlight of the Marine Life Exhibition). At that point, the probability of the visitors viewing this exhibit is high, and hence the *Transition Model* is capable of generating very accurate predictions.

Comparing immediate predictive accuracy (*CA(*1*,*1*)*) with eventual predictive accuracy (*CA(*1*,all)*), immediate predictions (Figures 2(c) and 2(d)) unsurprisingly have a lower accuracy than the more lenient eventual predictions (Figures 2(a) and 2(b)). Comparing *Top*1 (Figures 2(a) and 2(c)) and *Sequence*1/3 (Figures 2(b) and 2(d)), the performance is comparable for both eventual predictions (*CA(*1*,all)*) and immediate predictions (*CA(*1*,*1*)*). In fact, the performance differences are almost always statistically insignificant, except for *Sequence*1/3 *HM* outperforming *Top*1 *HM* at around 57% with respect to *CA(*1*,all)* (statistically significant with $0.05 < p < 0.1$), and *Top*1 *TM* which performs slightly better than *Sequence*1/3 *TM* at around 50% of the visits (statistically significant with $p < 0.1$). Additional observations are required to draw more compelling conclusions.

Except for the peak at 55%-60% of the visits, Figures 2(a)-(d) show a downwards trend for the accuracy of *IM*, *TM* and *HM* as the visit progresses. At



(a) RA_{pred} for *Top*1

(b) RA_{pred} for *Sequence*1/3

Fig. 3. Comparison of *Top*1 and *Sequence*1/3 with respect to Ranking Accuracy

first glance, this behaviour seems counter-intuitive, as the accuracy of predictions should typically increase as more evidence becomes available. However, this trend can be explained by the structure of the exhibition (Figure 1). In the initial phase of the visit, after the visitor has entered through the only entrance, accurate predictions are relatively easy to achieve. In the middle phase until the peak, the visitor can choose from different pathways, which makes accurate predictions more difficult as the visitor proceeds. The same applies to the final portion of the visit, after the peak at which the visitor behaviour is dominated by the exhibition highlight.

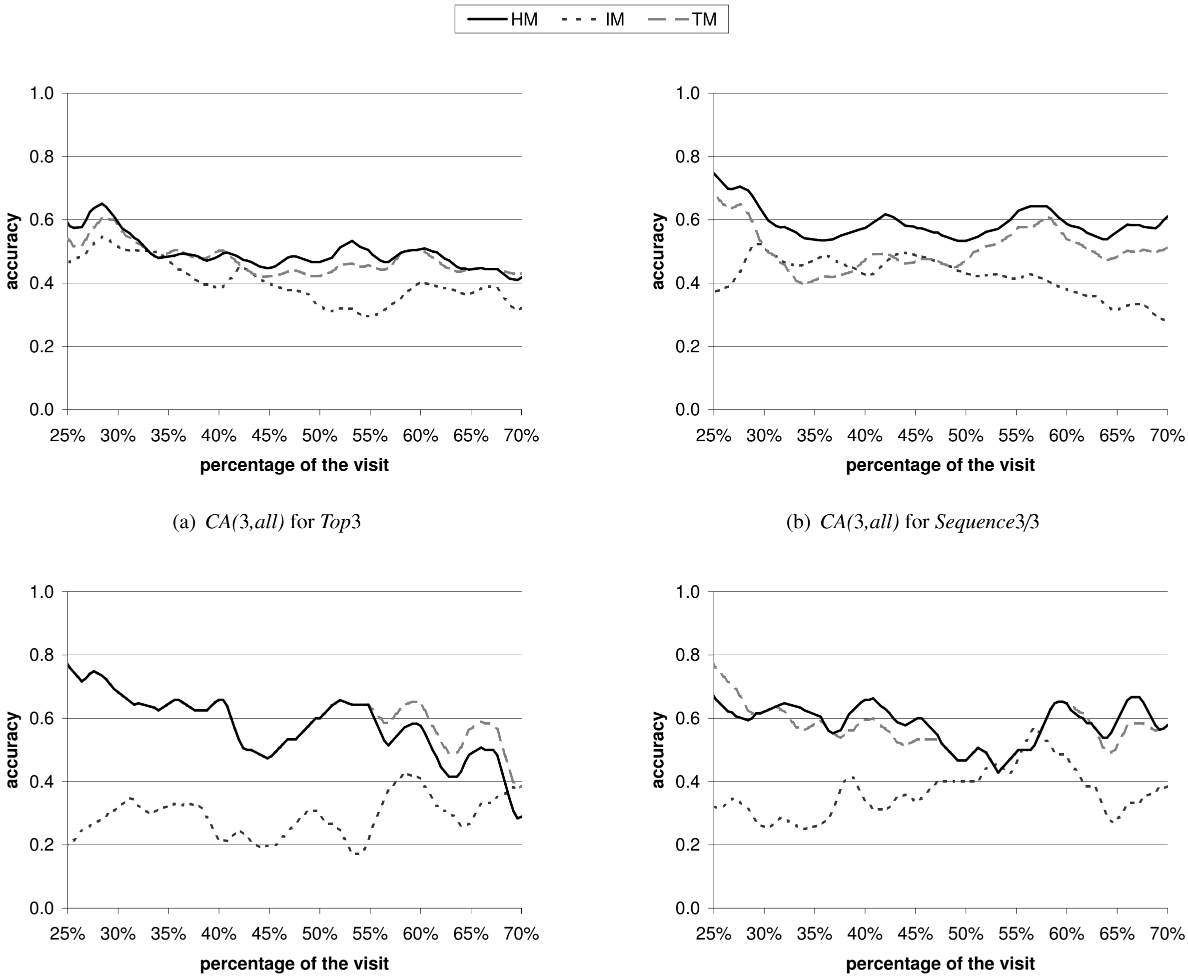
6.3.2. *Ranking Accuracy*

Here we present the results for RA_{pred}, the ranking of the first predicted exhibit in the sequence of exhibits that were actually viewed by the visitor in the remainder of the visit. The results for *Top*1 and *Sequence*1/3 are shown in Figure 3. In both modes of operation, the accuracy of *HM* is comparable to the accuracy of *TM* (the difference is almost always not statistically significant), both of which mostly outperform *IM* (statistically significant with $p < 0.05$). These results provide additional empirical evidence for our earlier observation that in a physical space with homogeneous information items, the spatial structure (rather than the interests of the visitors) dominates predictive accuracy. Also, although *HM* mostly outperforms *TM* and *IM* individually, the closeness between *HM* and *TM* indicates that hybridisation has a minor benefit with respect to Ranking Accuracy. Comparing the operation modes *Top*1 and *Sequence*1/3 across Figures 3(a) and 3(b), the variants *HM* and *IM* perform comparably, i. e., the differences are not statistically significant. However, *Top*1 performs slightly better than *Sequence*1/3 for *TM* (and *HM*) at around the 40% point (statistically significant with $p < 0.1$). Additional observations are necessary to draw more compelling conclusions regarding this behaviour.

Similarly to *CA(*1*,all)*, RA_{pred} shows an accuracy peak at 55%-60% of the visit. The explanation is similar to above. Once a visitor enters the areas from where the highlight exhibit is visible, predictions of its ranking in the sequence of actually visited exhibits are accurate. Also, as for *CA(*1*,M)* except for the peak, RA_{pred} decreases as the visit progresses. Similarly to the above argument for *CA(*1*,M)*, this trend can be explained by the structure of the exhibition.

6.4. *Evaluation for* $K = 3$

This section summarises the analysis of the evaluation for $K = 3$, i. e., the prediction of the next three exhibits to be viewed by the visitor. As for $K = 1$, we consider two modes of operation, *Top*3 and *Sequence*3/3. For each percentage of the visit, we generated predictions using our three prediction models: *IM*, *TM* and *HM*. Our predictions were evaluated using the Classification Accuracy measures *CA(*3*,all)* and *CA(*3*,*1*)* and the Ranking Accuracy measures RA_{real} and *mSP* described in Section 6.2.



(a) *CA(*3*,all)* for *Top*3

(b) *CA(*3*,all)* for *Sequence*3/3

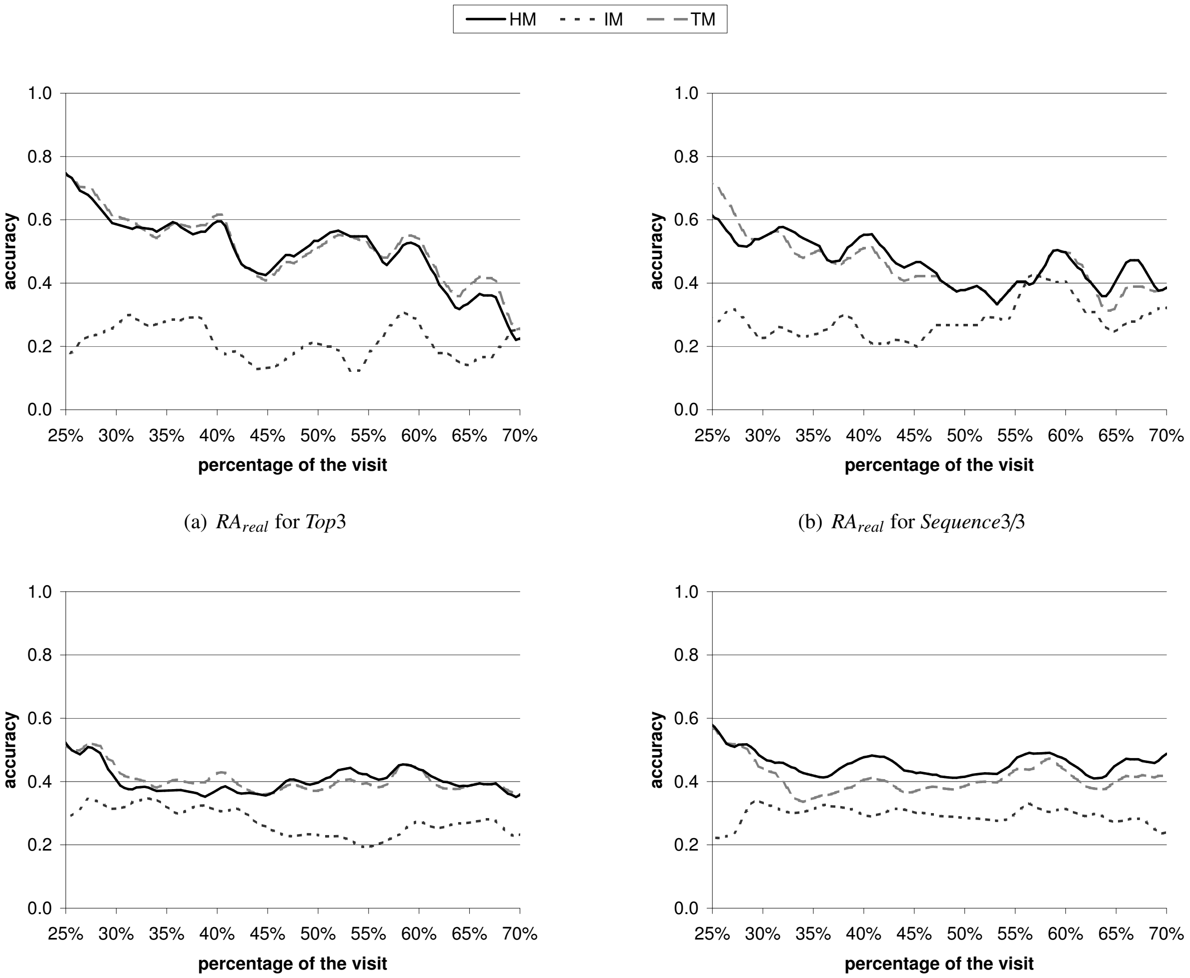
(c) *CA(*3*,*1*)* for *Top*3

(d) *CA(*3*,*1*)* for *Sequence*3/3

Fig. 4. Comparison of *Top*3 and *Sequence*3/3 with respect to Classification Accuracy

6.4.1. *Classification Accuracy*

In this section, we present the results for *CA(*3*,all)* and *CA(*3*,*1*)*. Figures 4(a) and 4(b) depict *CA(*3*,all)* for the *Top*3 and *Sequence*3/3 modes of operation respectively, and Figures 4(c) and 4(d) depict *CA(*3*,*1*)* for *Top*3 and *Sequence*3/3. The relative performance of *HM* and *TM* is similar to that shown in Figures 2(a) and 2(b), but the differences between *IM* and *TM* (and hence *HM*) are less pronounced than for *CA(*1*,all)*. Often, however, *HM* still outperforms *IM* significantly ($p < 0.05$). The performance difference between *HM* and *TM* for the *Top*3 mode of operation is not statistically significant, whereas *HM* slightly outperforms *TM* for the *Sequence*3/3 mode of operation ($p < 0.1$) up to 50% of the way through the visit. The divergence between *TM* and *IM* observed in Figures 2(a)-(b) also appears in the initial stages of the visit for *Sequence*3/3 (Figure 4(b)) and around the 55%-60% point for both *Top*3 and *Sequence*3/3. The results depicted in Figures 4(c) and 4(d) are more clear cut than those in Figures 4(a) and 4(b) (but recall that, in contrast to the other Classification Accuracy measures, *CA(*3*,*1*)* does not measure precision). The performance of *TM* and *HM* is comparable (the differences are not statistically significant), with both *TM* and *HM* significantly outperforming *IM* ($p < 0.05$). Overall, as for $K = 1$, the performance of *HM* is mostly equal to or better than the performance of *TM*, which in turn outperforms *IM*. The Classification Accuracy for *Sequence*3/3 is higher than for *Top*3, with *HM* yielding the best performance overall (Figures 4(a) and 4(b)). In particular, when comparing Figures 4(a)



(a) RA_{real} for *Top*3

(b) RA_{real} for *Sequence*3/3

(c) *mSP* for *Top*3

(d) *mSP* for *Sequence*3/3

Fig. 5. Comparison of *Top*3 and *Sequence*3/3 with respect to Ranking Accuracy

and 4(b), *Top*3 *IM* and *Sequence*3/3 *IM* almost always perform comparably. The same holds for *TM*. However, *Sequence*3/3 *HM* often significantly outperforms *Top*3 *HM* ($p < 0.1$). Comparing *Top*3 and *Sequence*3/3 across Figures 4(c) and 4(d) with respect to *CA(*1,3*)*, there is no performance difference for any of *HM*, *IM* and *TM*, except for *IM* around 50%.

Note that the peak observed at 55%-60% of the visit for the *CA(*1,*all)* experiments is less prominent for *CA(*3,*all)*. We posit that this is because when three exhibits are predicted (rather than a single exhibit), the influence of the highlight exhibit on the Classification Accuracy is reduced, as it is moderated by the accuracy of the predictions for two other exhibits. Also, as expected, for the *TopK* operating mode, the Classification Accuracy of *CA(*3,*all)* is lower than for *CA(*1,*all)*, since the former predicts three independent exhibits, while the latter predicts a single exhibit (Section 5.1). Interestingly, the discrepancy between *CA(*3,*all)* and *CA(*1,*all)* is lower for the *SequenceK/N* operating mode, since in this mode a sequence of three items is predicted anyway, with only the first item being selected for *CA(*1,*all)* (Section 5.2).

6.4.2. Ranking Accuracy

Finally, we present the results for RA_{real}, i. e., the ranking of the next exhibit actually viewed by the visitor in the sequence of predicted exhibits, and *mSP*, i. e., the fit between the predicted exhibit sequence as a whole relative to the actual sequence of visited exhibits. The results are shown in Figure 5.

Figures 5(a) and 5(b) show the performance with respect to RA_{real}. Comparing the different model variants of *Top*3 and *Sequence*3/3 with respect to RA_{real} it is evident that *HM* and *TM* perform identically (the difference is almost always not statistically significant), and both mostly outperform *IM* ($p < 0.05$). The closeness of *HM* and *TM* shows that, as for *Top*1 and *Sequence*1/3, hybridisation has a minor benefit with respect to Ranking Accuracy. Comparing the operation modes *Top*3 and *Sequence*3/3 across Figures 5(a) and 5(b), *HM* and *TM* perform comparably, i. e., the differences are not statistically significant. However, *Sequence*3/3 slightly outperforms *Top*3 for *IM* ($p < 0.1$).

Figures 5(c) and 5(d) depict the accuracy of *Top*3 and *Sequence*3/3 with respect to *mSP*, our modified version of Spearman's rank correlation. Comparing the different model variants of *Top*3, *HM* and *TM* perform identically (the difference is not statistically significant). However, *HM* outperforms *IM* for small visit percentages and visit percentages larger than 50% ($p < 0.05$). For *Sequence*3/3, *HM* generally outperforms *IM* ($p < 0.05$). *HM* also often outperforms *TM* ($p < 0.05$) for visit percentages of less than 50%. Comparing the operation modes *Top*3 and *Sequence*3/3 across Figures 5(c) and 5(d), *TM* and *IM* perform comparably (the differences are almost always not statistically significant). However, *Sequence*3/3 *HM* outperforms *Top*3 *HM* ($p < 0.1$) for visit percentages less than 50%.

The above results show that the ranking generated by *HM* is the most accurate for any amount of information available about the visitor. The performance of *HM* is slightly higher than that of *TM* overall, both of which are considerably higher than that of *IM*. This result validates our previous observation regarding the benefits of hybridisation.

6.5. Summary

The main findings of this work are as follows.

- When generating predictions for the next ($K = 1$) exhibit to be viewed by a visitor in a constrained physical museum space: (1) the accuracy of simple *Top*1 predictions is comparable to the accuracy of more complex *Sequence*1/3 predictions, and there is no need to apply a complex sequence prediction mechanism; and (2) although the hybridisation of *TM* and *IM* can be beneficial, the spatial structure of the space dominates, and any improvement in predictive accuracy as a result of hybridisation is slight.
- When predicting a sequence of $K = 3$ exhibits: (1) *Sequence*3/3 is superior to simple *Top*3, meaning that sequence information aids prediction; and (2) individual *TM* and *IM* predictions should be hybridised, as their combined predictive accuracy surpasses that of the individual methods.

7. Conclusions and Future Work

We have offered two models for predicting visitor locations in a museum — a *Transition Model* implicitly capturing spatial information, and an *Interest Model* based on viewing times —, and we have experimented with different combinations of these models into a hybrid ensemble model. The performance of our models was tested on a small dataset of museum visits collected at the Marine Life Exhibition in Melbourne Museum. Our results show that the *Transition Model* outperforms the *Interest Model*, indicating that the layout of a physical space with homogeneous exhibits is a key factor influencing visitor behaviour. However, the hybrid model yields the best performance overall, with an average accuracy of 68% with respect to *CA(*1*,all)*, and 59% with respect to *CA(*3*,all)*, which demonstrates the importance of also considering a visitor's interests. Moreover, our results indicate that, when predicting the next three exhibits to be viewed, a model that predicts a sequence of items has a higher accuracy than a model that predicts a set (59% vs. 49%). Surprisingly, this is not the case when predicting a single item, where the performance of the simpler set-based model is comparable to the performance of the sequence-based model (66% vs. 68%). Nonetheless, if only one of the models is to be used to predict pathways of different lengths, then the sequence-based model seems preferable.

Our results raise several issues pertaining to modelling users in physical spaces in general, and museums in particular. These issues pertain to exhibit diversity, amount and quality of data, user modelling strategies, and recommendations.

Exhibit diversity. Our experiments were conducted on the Marine Life Exhibition, which contains homogeneous exhibits. This means that the visitors who enter the exhibition were already interested in marine life. Thus, as corroborated by our results, a key factor influencing visitor behaviour is the spatial factor, reflecting the layout of the exhibition. However, this conclusion may not be valid in a space of heterogeneous exhibits, such as the entire Melbourne Museum, which has exhibits relating to flora, fauna, Australian history and modern life, or a smaller museum with diverse exhibit topics. These observations motivate future experiments at different levels of granularity, e. g., inter-exhibition versus intra-exhibition, while considering the link between granularity and topic diversity. These experiments are expected to shed light on the influence of exhibition size and exhibit diversity on the applicability of our models.

Amount and quality of data. Our dataset comprised traces of trajectories (plus viewing durations) of 44 museum visitors. This is a rather small amount of data, in particular in the context of probabilistic models. However, in contrast to web-based data collections, the collection of visit traces in a physical space is an expensive and time-consuming process, both when done by human trackers and by electronic devices. Electronic devices are still relatively inaccurate, which may affect the accuracy of the derived user models and the quality of the personalisation provided to the users [6]. This problem should be considered and addressed prior to deploying such devices. At the same time, this problem obfuscates basic user modelling issues, and should be avoided during initial model development. In contrast, human tracking is precise, and hence ideal for initial model development, but clearly cannot be used during model deployment.

In the near future, we intend to collect additional traces of museum visits using human trackers. However, the amount of data we can feasibly collect will still be relatively small, which brings us to the consideration of the robustness of the resultant models. Although our statistical models have yielded good experimental results, their estimates have a high variance, due to the limited number of observed visitor traces. An important avenue of research for settings where data collection is problematic involves determining the type of model that can be built with the data that can be feasibly collected. To address this issue we propose to investigate simple spatial models that require few parameters, and hence are statistically robust, and compare their performance with that of the *Transition Model*. One such model is a nearest-neighbour model, where we group all the exhibits that are next to a location, all the exhibits that are T steps away, and so on, and calculate the probability of transitioning to a group (where all the exhibits would be equiprobable). Another option is to make the probability of visiting an exhibit inversely proportional to the distance from the current exhibit.

User modelling strategies. Our current approach for combining user models belongs to the ensemble category, where the predictions made by two models are combined [16] (also called a weighted hybrid [5]). However, the models themselves are built separately — the *Transition Model* from trajectory information, and the *Interest Model* from temporal information. In the future, we propose to combine these information sources and conduct model hybridisation at the model acquisition stage. For example, this can be done by considering the distance from a current exhibit when computing a visitor's *Interest Model*. That is, the farther a newly visited exhibit is from the last visited exhibit, the higher the interest in the new exhibit. Similarly to [3], we plan to investigate a combination of collaborative user models with content-based models. We also intend to address the cold-start problem [12,29] by applying machine learning techniques to determine the point in a visit at which personalised models can be deployed. These techniques will also be applied to find the optimal weight of the individual models in ensemble models.

Recommendations. In this paper, we have focused on the prediction of a visitor's locations in a museum. Accurately predicting these locations will enable us to make recommendations about exhibits to visit, and to personalise the content delivered for these exhibits. However, in this domain, the transition from prediction to recommendation is not trivial. On one hand, our system should recommend exhibits the visitor intends to visit in order to build trust in the system. On the other hand, trivial recommendations, such as exhibits along a path prescribed by the spatial layout, or too many of these recommendations are likely to annoy the visitor. The physicality of the museum space must also be taken into account, as an item of interest that is far from a visitor's current location should only be recommended under special circumstances, e. g., if the museum is about to close. In the future, we plan to investigate recommendation generation strategies which strike a balance between these factors.

Acknowledgements

This research was supported in part by Discovery grant DP0770931 from the Australian Research Council. The authors thank Enes Makalic for his assistance with ensemble models. Thanks also go to Carolyn Meehan and her team from Museum Victoria for fruitful discussions, their support of this research, and the dataset.

References

[1] David W. Albrecht and Ingrid Zukerman, editors. Special issue on statistical and probabilistic methods for user modeling. *User Modeling and User-Adapted Interaction*, 17(1-2), 2007.

[2] Lora Aroyo, Natalia Stash, Yiwen Wang, Peter Gorgels, and Lloyd Rutledge. CHIP demonstrator: Semantics-driven recommendations and museum tour generation. In *Proc. of the Sixth Intl. Semantic Web Conf. (ISWC-07)*, pages 879–886, 2007.

[3] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. Cross-technique mediation of user models. In *Proceedings of the 4th Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 21–30, 2006.

[4] Fabian Bohnert and Ingrid Zukerman. Using viewing time for theme prediction in cultural heritage spaces. In *Proceedings of the 20th Australian Joint Conference on Artificial Intelligence (AI-07)*, pages 367–376, 2007.

[5] Robin Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, 2002.

[6] David J. Carmichael, Judy Kay, and Bob Kummerfeld. Consistent modelling of users, devices and sensors in a ubiquitous computing environment. *User Modeling and User-Adapted Interaction*, 15(3-4):197–234, 2005.

[7] Keith Cheverst, Keith Mitchell, and Nigel Davies. The role of adaptive hypermedia in a context-aware tourist guide. *Communications of the ACM*, 45(5):47–51, 2002.

[8] Josef Fink and Alfred Kobsa. User modeling for personalized city tours. *Artificial Intelligence Review*, 18(1):33–74, 2002.

[9] Irving John Good. *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. Research Monograph No. 30. MIT Press, Cambridge, Massachusetts, 1965.

[10] Karl Grieser, Timothy Baldwin, and Steven Bird. Dynamic path prediction and recommendation in a museum environment. In *Proceedings of the ACL 2007 Workshop on Language Technology for Cultural Heritage Data (LaTeCH-07)*, pages 49–56, 2007.

[11] Marek Hatala and Ron Wakkary. Ontology-based user modeling in an augmented audio reality system for museums. *User Modeling and User-Adapted Interaction*, 15(3-4):339–380, 2005.

[12] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John T. Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-99)*, pages 230–237, 1999.

[13] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):5–53, 2004.

[14] Anthony Jameson and Antonio Krüger, editors. Special issue on user modeling in ubiquitous computing. *User Modeling and User-Adapted Interaction*, 15(3-4), 2005.

[15] Bruce Krulwich. Lifestyle Finder: Intelligent user profiling using large-scale demographic data. *Artificial Intelligence Magazine*, 18(2):37–45, 1997.

[16] George Lekakos and George M. Giaglis. A hybrid approach for improving predictive accuracy of collaborative filtering algorithms. *User Modeling and User-Adapted Interaction*, 17(1-2):5–40, 2007.

[17] Thomas W. Malone, Kenneth R. Grant, Franklyn A. Turbak, Stephen A. Brobst, and Michael D. Cohen. Intelligent information-sharing systems. *Communications of the ACM*, 30(5):390–402, 1987.

[18] Masahiro Morita and Yoichi Shinoda. Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-94)*, pages 272–281, 1994.

[19] Reinhard Oppermann and Marcus Specht. A context-sensitive nomadic information system as an exhibition guide. In *Proceedings of the Second International Symposium on Handheld and Ubiquitous Computing (HUC-00)*, pages 127–142, 2000.

[20] Jeffrey Parsons, Paul Ralph, and Katherine Gallager. Using viewing time to infer user preference in recommender systems. In *Proc. of the AAAI Workshop on Semantic Web Personalization (SWP-04), in conjunction with AAAI-04*, pages 52–64, 2004.

[21] Daniela Petrelli and Elena Not. User-centred design of flexible hypermedia for a mobile guide: Reflections on the HyperAudio experience. *User Modeling and User-Adapted Interaction*, 15(3-4):303–338, 2005.

[22] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW-94)*, pages 175–186, 1994.

[23] Paul Resnick and Hal R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997.

[24] Gerard Salton and Michael J. McGill. *An Introduction to Modern Information Retrieval*. McGraw Hill, 1983.

[25] Upendra Shardanand and Pattie Maes. Social information filtering: Algorithms for automating “word of mouth”. In *Proceedings of the 1995 SIGCHI Conference on Human Factors in Computing Systems (CHI-95)*, pages 210–217, 1995.

[26] Sidney Siegel and N. John Castellan. *Non-Parametric Statistics for the Behavioral Sciences*. McGraw–Hill, Inc., second edition, 1988.

[27] Oliviero Stock, Massimo Zancanaro, Paolo Busetta, Charles Callaway, Antonio Krüger, Michael Kruppa, Tsvi Kuflik, Elena Not, and Cesare Rocchi. Adaptive, intelligent presentation of information for the museum visitor in PEACH. *User Modeling and User-Adapted Interaction*, 18(3):257–304, 2007.

[28] Ichiro Yamada and Timothy Baldwin. Automatic discovery of telic and agentive roles from corpus data. In *Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation (PACLIC-04)*, pages 115–26, 2004.

[29] Ingrid Zukerman and David W. Albrecht. Predictive statistical models for user modeling. *User Modeling and User-Adapted Interaction*, 11(1-2):5–18, 2001.