

Non-Intrusive Personalisation of the Museum Experience

Fabian Bohnert and Ingrid Zukerman

Faculty of Information Technology, Monash University
Clayton, VIC 3800, Australia

{fabian.bohnert, ingrid.zukerman}@infotech.monash.edu.au

Abstract. The vast amount of information presented in museums is often overwhelming to a visitor, making it difficult to select personally interesting exhibits. Advances in mobile computing and user modelling have made possible technology that can assist a visitor in this selection process. Such a technology can (1) utilise non-intrusive observations of a visitor’s behaviour in the physical space to learn a model of his/her interests, and (2) generate personalised exhibit recommendations based on interest predictions. Due to the physicality of the domain, datasets of visitors’ behaviour (i. e., visitor pathways) are difficult to obtain prior to deploying mobile technology in a museum. However, they are necessary to assess different modelling techniques. This paper reports on a methodology that we used to conduct a manual data collection, and describes the dataset we obtained. We also present two collaborative models for predicting a visitor’s viewing times of unseen exhibits from his/her viewing times at visited exhibits (viewing time is indicative of interest), and evaluate our models with the dataset we collected. Both models achieve a higher predictive accuracy than a non-personalised baseline.

1 Introduction

Cultural heritage spaces such as museums offer a vast amount of information. However, a visitor’s receptivity and time are typically limited, posing the challenge of selecting personally interesting exhibits to view within the available time. Advances in mobile computing and user modelling provide the opportunity to assist a visitor in this selection process — by means of personalised mobile technology. Such a technology can (1) utilise non-intrusive observations of a visitor’s behaviour in the physical space to learn a model of his/her interests, and (2) generate personalised exhibit recommendations based on interest predictions. The physicality of the domain poses practical challenges for developing predictive user models. For example, datasets of visitors’ behaviour in the museum (i. e., visitor pathways) are difficult to obtain prior to deploying mobile technology (e. g., positioning technology).

In this paper, we describe a computer-supported methodology that we used to manually collect a dataset of visitor pathways in Melbourne Museum (Melbourne, Australia), and the dataset we obtained. We then present two collaborative models for predicting a visitor’s viewing times of unseen exhibits from his/her viewing times at visited exhibits: (1) a memory-based nearest-neighbour collaborative filter, and (2) a model-based approach utilising the theory of Gaussian spatial processes. Our models were evaluated with the dataset we collected, by comparing their predictive accuracy with that of a non-personalised baseline. Both models attain a higher predictive accuracy than the baseline, with our spatial process model outperforming the nearest-neighbour collaborative filter.

The paper is organised as follows. In Section 2, we outline related research. Section 3 describes our methodology for collecting visitor pathways in a physical museum and the dataset we obtained, followed by Section 4 where we discuss our models for predicting a visitor’s viewing times. Section 5 summarises the results of our evaluation, and in Section 6, we discuss ways to utilise our predictive models in a personalised museum handheld guide. We conclude in Section 7.

2 Related Research

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 [1] developed a handheld tourist guide for visitors to the city of 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 a user’s interactions with the system, with a focus on adapting content presentation, rather than predicting and recommending exhibits to be viewed. For example, *HyperAudio* [3] dynamically adapted the presented content and hyperlinks to stereotypical assumptions about a user, and to what a user has already accessed and seems interested in. The augmented audio reality system for museums *ec(h)o* [4] treated user interests in a dynamic manner, and adapted its user model on the basis of a user’s interactions with the system. The collected user modelling data were used to deliver personalised information associated with exhibits via audio display. The *PEACH* project [5] developed a multimedia handheld guide which adapts its user model on the basis of both explicit visitor feedback and implicit observations of a visitor’s interactions with the device. This user model was then used to generate personalised multimedia presentations.

These systems, like most systems in the museum domain, rely on knowledge-based user models in some way, and hence, require an explicit, a-priori engineered representation of the domain knowledge. In contrast, our research investigates non-intrusive statistical user modelling and recommendation techniques that do not require such an explicit domain knowledge representation [6].

3 Data Collection and Dataset

This section describes our methodology for collecting a dataset of visitor pathways (Section 3.1), and the dataset we obtained (Section 3.2).

3.1 Data Collection Methodology

The *GECKO* project endeavours to develop user modelling techniques which rely on non-intrusive observations of users’ behaviour in physical spaces [7]. Developing such non-intrusive user modelling and personalisation techniques for museums requires datasets about visitor behaviour in the physical museum space (i. e., visitor pathways). Datasets that are suitable for the development phase can be obtained by manually tracking

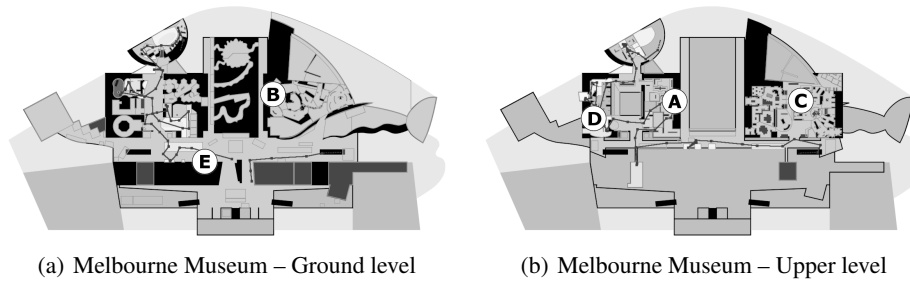


Fig. 1. Visitor pathway visualised on a site map of Melbourne Museum

museum visitors. Such a data collection methodology is clearly inappropriate for model deployment, but it facilitates model development by eschewing issues related to technology selection and instrumentation accuracy.

In the museum domain, traditional manual tracking methodologies include using printed site maps and a stopwatch to record visitors' pathways and the time spent at various exhibits [8]. However, depending on the required level of detail and frequency of events, such logging techniques can overwhelm a tracker, potentially yielding tracking errors. Additionally, they require a substantial transcription effort to digitise the data. This motivated us to develop a computer-supported methodology for recording museum visitors' time-annotated pathways. Hence, in the framework of the GECKO project, we developed two Java-based tools for manual tracking and visualisation of datasets, *GECKOtracker* and *GECKOvisualiser* respectively.

- *GECKOtracker* is a clickable interface showing a digitised site map of the physical space encoded in the *Scalable Vector Graphics (SVG)* file format. *GECKOtracker* resides on portable computers carried by (human) trackers — one tracker follows one museum visitor at a time. When following a visitor, a tracker logs the visitor's position by clicking on the map, while the computer clock delivers the time. A 'viewing event' is registered when the tracker clicks on an exhibit. Figure 1 depicts the site map for Melbourne Museum, together with one of the visitor pathways we collected.
- *GECKOvisualiser* is used for post-collection visualisation and analysis of the gathered data. It supports different views of the data (e. g., showing a pathway or the distribution of viewing times) in two linked formats: visualisation on the site map (Figure 1) and textual log. *GECKOvisualiser* was used to gain a better understanding of our dataset, and to correct obvious mistakes made by our trackers.

Museums such as Melbourne Museum display thousands of exhibits distributed over many separate galleries and exhibitions. However, normally visitors do not require recommendations to travel between individual, logically related exhibits in close physical proximity. Rather, they may prefer recommendations regarding physically separate areas. In order to gather data for assessing predictive models that support appropriate recommendations, we grouped Melbourne Museum's individual exhibits into semantically coherent and spatially confined *exhibit areas*. This task, which was performed with the assistance of museum staff, yielded 126 exhibit areas.

Table 1. Dataset statistics

	Mean	Stddev	Min	Max
Visit length (hrs)	1:50:39	0:47:54	0:28:23	4:42:12
Viewing time (hrs)	1:31:09	0:42:05	0:14:09	4:08:27
Exhibit areas / visitor	52.70	20.69	16	103
Visitors / exhibit area	66.09	25.36	6	117

GECKO*tracker* was used by 16 trackers in total, comprising university students and museum staff. Feedback from trackers and other museum staff indicates that they value our software. The trackers particularly liked the software’s ease of operation. Feedback regarding the digital maps of the museum indicates that our maps encode sufficient information for the trackers to correctly identify exhibit areas — a key requirement for accurate tracking. Feedback from participants shows that most visitors did not feel disturbed by a tracker following them through the museum. In fact, some participants stated that quite early into their visit, they forgot that they were being tracked (despite being approached at the start of the visit to obtain their approval).

3.2 Dataset

Using GECKO*tracker*, we recorded the pathways of over 170 visitors to Melbourne Museum from April to June 2008. We restricted ourselves to tracking first-time adult visitors travelling on their own, to ensure that neither prior knowledge about the museum nor other visitors’ interests influenced a visitor’s decisions about which exhibits to view. Prior to the data collection, we briefed our trackers on the usage of the tracking tool, the layout of the museum, and its digital representation on the site map. Additionally, we clarified what should be considered a viewing event. After the data collection, the visitor pathways were post-processed using GECKO*visualiser*. For instance, we removed mis-clicks reflecting viewing events that could not have possibly occurred, e. g., visitor transitions from one end of the museum to the other and back within a few seconds, or transitions outside the museum walls and back. We also removed incomplete visitor pathways, e. g., due to a laptop running out of battery, or a visitor leaving unexpectedly. The resulting dataset comprises 158 complete visitor pathways in the form of time-annotated sequences of visited exhibit areas, with a total visit length of 291:22:37 hours, and a total viewing time of 240:00:28 hours. The dataset also contains demographic information about the visitors, which was obtained by means of post-visit interviews conducted by our trackers. In total, we obtained 8327 viewing durations at the 126 exhibit areas, yielding an average of 52.7 exhibit areas per visitor (41.8% of the exhibit areas). Hence, on average 58.2% of the exhibit areas were not viewed by a visitor. This indicates that there is potential for pointing a visitor to relevant but unvisited exhibit areas. Table 1 summarises further statistics of the dataset.

Clearly, the deployment of non-intrusive personalised visitor support in a museum requires suitable positioning technology to track visitors, and models to infer visitors’ interests. Although our dataset was obtained manually, it provides information that is of the same type as information inferable from sensing data. Additionally, the results obtained from experiments with this dataset are essential for model development, as they provide an upper bound for the predictive performance of our models.

4 Viewing Time Prediction from Non-Intrusive Observations

In an information-seeking context, people usually spend more time on relevant information than on irrelevant information, as viewing time correlates positively with preference and interest [9]. Hence, viewing time can be used as an indirect measure of interest. We propose to use log viewing time (instead of raw viewing time), due to the following reasons. When examining our dataset (Section 3.2), we found the distributions of viewing times at exhibits to be positively skewed (we use the terms ‘exhibit’ and ‘exhibit area’ synonymously in the remainder of this paper). Thus, the usual assumption of a Gaussian model did not seem appropriate. To select a more appropriate family of probability distributions, we used the *Bayesian Information Criterion (BIC)*. We tested exponential, gamma, normal, log-normal and Weibull distributions. The log-normal family fitted best, with respect to both number of best fits and average BIC score (averaged over all exhibits). By transforming all viewing times to their log-equivalent, we obtained normally distributed data. This transformation fits well with the idea that for high viewing times, an increase in viewing time indicates a smaller increase in the modelled interest than a similar increase in the context of low viewing times.

In this section, we propose two models for predicting a visitor’s (log) viewing times from non-intrusive observations of his/her (log) viewing times at visited exhibits: a memory-based nearest-neighbour collaborative filter [10] (Section 4.1), and a model-based approach based on the theory of Gaussian spatial processes [11] (Section 4.2).

4.1 Nearest-Neighbour Collaborative Filter

Our *Collaborative Filter Model (CFM)* for predicting a visitor’s viewing times of unseen exhibits is a nearest-neighbour collaborative filter [10]. The predictive model is built by first collecting all observed log viewing times into a matrix of size $m \times n$, where m is the cardinality of the set V of all visitors, and n is the cardinality of the set I of all exhibits (we use $v \in V$ to denote a visitor, and $i \in I$ to denote an exhibit). To ensure that varying exhibit complexity does not affect the similarity computation for selecting the nearest neighbours (viewing time increases with exhibit complexity), we then normalise all these values by calculating exhibit-wise z-scores. That is, we normalise the log viewing time of a visitor for an exhibit by subtracting its log viewing time mean $\bar{r}_{\cdot i}$ and dividing by its standard deviation σ_i . The resultant normalised log viewing times r_{vi} , which are stored in a matrix R of size $m \times n$, may be regarded as implicit ratings given by visitors to exhibits.

We calculate \tilde{r}_{ai} , a personalised prediction of a current visitor a ’s unobserved (normalised log) viewing time r_{ai} , from the values in R as follows (we unnormalise afterwards to obtain a log viewing time):

$$\tilde{r}_{ai} = \bar{r}_{a\cdot} + \frac{\sum_{v \in N(a,i)} s_{av} (r_{vi} - \bar{r}_{v\cdot})}{\sum_{v \in N(a,i)} |s_{av}|}, \quad (1)$$

where $\bar{r}_{a\cdot}$ denotes the current visitor a ’s average normalised log viewing time, $N(a, i)$ is the set of nearest neighbours, and s_{av} is the similarity between visitors a and v (calculated using Pearson’s correlation coefficient on the normalised log viewing times of visitors a and v). The set of nearest neighbours $N(a, i)$ for the current visitor a and

exhibit i is constructed by (1) calculating s_{av} for all visitors v who viewed exhibit i , and (2) selecting the visitors most similar to current visitor a — those for whom $|s_{av}|$, the absolute similarity to visitor a , is above a certain threshold. When calculating \tilde{r}_{ai} , we use a weighted mean of deviations from each neighbour’s average normalised log viewing duration \bar{r}_v , in order to neutralise viewing behaviour differences between visitors. This weighted mean is then added to the current visitor’s average normalised log viewing time \bar{r}_a . Our experiments suggest that these calculations should be performed only after enough evidence has been gathered for obtaining a good estimate of \bar{r}_a . In our case, this happens after 20 observations. Prior to that, we estimate r_{ai} using only a (personalised) similarity-weighted mean of the r_{vis} .

Whenever a similarity-weighted personalised prediction is not possible (e. g., when the set of nearest neighbours is empty), we estimate r_{ai} using an unweighted average of the deviations from the neighbours’ (normalised log viewing time) means [10]:

$$\tilde{r}_{.i}^a = \bar{r}_a + \frac{\sum_{v \in N(\cdot, i)} (r_{vi} - \bar{r}_v)}{|N(\cdot, i)|},$$

where $N(\cdot, i)$ denotes the set of visitors who viewed exhibit i . As above, we use a simple (non-personalised) mean of the r_{vis} for less than 20 observations in visitor a ’s profile.

We added further modifications from the literature to improve *CFM*’s performance. For instance, we use *significance weighting* [10] to decrease the influence of nearest neighbours whose similarity value is computed from a small number of co-viewed exhibits. We also employ *shrinkage to the mean* [12], which has been shown to often improve statistical estimation, whenever we compute a personalised prediction of r_{ai} (replacing Equation 1):

$$\hat{r}_{ai} = \tilde{r}_{.i}^a + \omega (\tilde{r}_{ai} - \tilde{r}_{.i}^a), \quad (2)$$

where $\omega \in [0, 1]$ is chosen such that an error measure of choice is minimised. We use the *mean absolute error (MAE)* (Section 5.1).

4.2 Gaussian Spatial Process Model

Spatial statistics is concerned with the analysis and prediction of geographic data [11]. Utilising spatial processes, the field deals with tasks such as modelling the associations between observations made at certain locations, and predicting values at locations where no observations have been made. The assumption made for spatial processes, that correlation between observations increases with decreasing site distance, fits well with our scenario, where viewing times are usually more correlated the more related exhibits are. Hence, by introducing a notion of spatial distance between exhibits to functionally specify this correlation structure, we can use spatial process models for predicting viewing times. We use s_1, \dots, s_n to denote the locations of exhibits $i, j \in I = \{1, \dots, n\}$ in a space providing such a distance measure, i. e., $\|s_i - s_j\|$. This distance measure can be easily obtained for the museum domain. That is, museums are carefully themed by curatorial staff, such that closely-related exhibits are in physical proximity. Based on this observation, we hypothesise that physical walking distance between exhibits is inversely proportional to their (content) similarity. Thus, we use physical walking distance as our distance measure between exhibits. Specifically, our SVG file-based representation of the museum (Section 3.1) was used to calculate the walking distances by

mapping the site map onto a graph structure which preserves the physical layout of the museum (i. e., preventing paths from passing through walls or ceilings). We normalised the resulting distances to the interval $[0, 1]$.

Typically, for a visitor $v \in V$, we have viewing times for only a subset of I , say for n_v exhibits. Denoting a visitor's log viewing time vector with \mathbf{r}_v , we collect all observed log viewing times into a vector $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_m)$ of dimension $\sum_{v=1}^m n_v$.¹ Associated with each exhibit i is a log viewing time mean $\bar{r}_{\cdot i}$ and a standard deviation σ_i . Let $\boldsymbol{\mu} = (\bar{r}_{\cdot 1}, \dots, \bar{r}_{\cdot n})$ be the vector of mean log viewing times, and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)$ the vector of standard deviations. Furthermore, $\boldsymbol{\mu}_v$ and $\boldsymbol{\sigma}_v$ are the vectors of means and standard deviations respectively for only those exhibits viewed by a visitor v . For example, if visitor 1 viewed exhibits 2, 3, 7 and 9, then $\boldsymbol{\mu}_1 = (\bar{r}_{\cdot 2}, \bar{r}_{\cdot 3}, \bar{r}_{\cdot 7}, \bar{r}_{\cdot 9})$ and $\boldsymbol{\sigma}_1 = (\sigma_2, \sigma_3, \sigma_7, \sigma_9)$.

Similarly to spatial processes, our *Spatial Process Model (SPM)* assumes a special correlation structure between the viewing times of different exhibits. In our experiments, we use a powered exponential [11]:

$$\rho(\|\mathbf{s}_i - \mathbf{s}_j\|; \phi, \nu) = \exp(-(\phi\|\mathbf{s}_i - \mathbf{s}_j\|)^\nu),$$

where $\phi > 0$ and $0 < \nu < 2$. That is, $\rho(\|\mathbf{s}_i - \mathbf{s}_j\|; \phi, \nu)$ models the correlation between the log viewing times of exhibits i and j . Let $H(\phi, \nu)$ be a correlation matrix with components $(H(\phi, \nu))_{ij} = \rho(\|\mathbf{s}_i - \mathbf{s}_j\|; \phi, \nu)$ collecting all these correlations, and let $H_v(\phi, \nu)$ denote a visitor v 's correlation matrix (dimension $n_v \times n_v$). That is, $H_v(\phi, \nu)$ corresponds to $H(\phi, \nu)$ having removed those rows and columns that correspond to unvisited exhibits. Also, let $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau^2, \phi, \nu)$ be a vector collecting the $2n + 3$ model parameters, where τ^2 denotes the variance of non-spatial error terms necessary to fully specify the model (they model non-spatial variation in the data). Then, modelling the data using Gaussian spatial processes (a detailed derivation appears in [13]), \mathbf{r} given $\boldsymbol{\theta}$ is multivariate normal of dimension $\sum_{v=1}^m n_v$. As the viewing times of different visitors $v = 1, \dots, m$ are independent, the model simplifies to

$$\mathbf{r}_v | \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_v, \Sigma_v) \text{ for all } v = 1, \dots, m, \quad (3)$$

where $\Sigma_v = \boldsymbol{\sigma}_v \mathbf{1}_{n_v} H_v(\phi, \nu) \boldsymbol{\sigma}_v \mathbf{1}_{n_v} + \tau^2 \mathbf{1}_{n_v}$ is a visitor v 's covariance matrix, and $\mathbf{1}_{n_v}$ is the identity matrix of dimension $n_v \times n_v$.

Given the model parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau^2, \phi, \nu)$, our model is fully specified. We employ Bayesian inference using *SPM*'s likelihood function derived from Equation 3 to estimate $\boldsymbol{\theta}$ from \mathbf{r} (in particular, we use *slice Gibbs sampling* [14]). This solution offers attractive advantages over the classic frequentist approach, such as the opportunity of incorporating prior knowledge into parameter estimation via the prior distribution, and capturing the uncertainty about the parameters via the posterior distribution.

We can now use multivariate normal theory to predict a current visitor a 's log viewing times of unseen exhibits, say $\mathbf{r}_{a,1}$, from a vector of observed viewing times $\mathbf{r}_{a,2}$. This is because $(\mathbf{r}_{a,1}, \mathbf{r}_{a,2}) | \boldsymbol{\theta}$ is normally distributed (similarly to Equation 3). If we use the following notation

$$\begin{bmatrix} \mathbf{r}_{a,1} \\ \mathbf{r}_{a,2} \end{bmatrix} | \boldsymbol{\theta} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_{a,1} \\ \boldsymbol{\mu}_{a,2} \end{bmatrix}, \begin{bmatrix} \Sigma_{a,11} & \Sigma_{a,12} \\ \Sigma_{a,12}^T & \Sigma_{a,22} \end{bmatrix} \right),$$

¹ The information in \mathbf{r} is the unnormalised equivalent of the information in R (Section 4.1).

then the conditional distribution $p(\mathbf{r}_{a,1}|\mathbf{r}_{a,2},\boldsymbol{\theta})$ is normal with mean vector and covariance matrix

$$\begin{aligned}\hat{\mathbf{r}}_{a,1} &= \mathbb{E}(\mathbf{r}_{a,1}|\mathbf{r}_{a,2},\boldsymbol{\theta}) = \boldsymbol{\mu}_{a,1} + \Sigma_{a,12}\Sigma_{a,22}^{-1}(\mathbf{r}_{a,2} - \boldsymbol{\mu}_{a,2}), \\ \text{Cov}(\mathbf{r}_{a,1}|\mathbf{r}_{a,2},\boldsymbol{\theta}) &= \Sigma_{a,11} - \Sigma_{a,12}\Sigma_{a,22}^{-1}\Sigma_{a,12}^T,\end{aligned}\quad (4)$$

where $\hat{\mathbf{r}}_{a,1} = \mathbb{E}(\mathbf{r}_{a,1}|\mathbf{r}_{a,2},\boldsymbol{\theta})$ represents a personalised prediction of the log viewing times $\mathbf{r}_{a,1}$. Additionally, a measure of confidence in this prediction can be derived from $\text{Cov}(\mathbf{r}_{a,1}|\mathbf{r}_{a,2},\boldsymbol{\theta})$, e. g., by using the variances on the diagonal of this matrix.

Being a model-based approach, *SPM* offers advantages over memory-based techniques such as *CFM*. For instance, the model parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau^2, \phi, \nu)$ have a clear interpretation, and the confidence measure provided by the model supports an informed interpretation of the model’s predictions. Additionally, recommendation generation is sped up by uncoupling the model-fitting phase from the prediction phase.

5 Evaluation

We describe the experimental setup in Section 5.1, and discuss our results in Section 5.2.

5.1 Experimental Setup

We used the dataset discussed in Section 3.2 to evaluate the predictive performance of our models *CFM* and *SPM* (Section 4). For our experiments, we ignored travel time between exhibits, and collapsed multiple viewing events of one exhibit into one event. Due to the relatively small size of our dataset, we used leave-one-out cross validation. That is, for each visitor, we trained the models with the data from 157 of the 158 visit trajectories, and used the withheld visit pathway for testing. For *CFM*, we evaluated several thousand parameterisations (e. g., varying the maximum number of nearest neighbours and the shrinkage weight ω), and used the best-performing one for our final experiments. As mentioned above, *SPM*’s model parameters $\boldsymbol{\theta}$ were estimated from the training data using slice Gibbs sampling [14]. For *CFM*, we computed predictions of a visitor’s log viewing times of unseen exhibits from the (normalised) log viewing times of the nearest neighbours (Equation 2). For *SPM*, log viewing times were predicted by conditioning a multivariate normal distribution (Equation 4), using the parameter estimates for $\boldsymbol{\theta}$ to instantiate the model. In addition to *CFM* and *SPM*, we implemented a baseline *Mean Model (MM)* which predicts the log viewing time of an exhibit i to be its (non-personalised) mean log viewing time $\bar{r}_{\cdot i}$.

We performed three types of experiments: *Individual Exhibit*, *Progressive Visit* and *Recommendation Potential*.

- **Individual Exhibit (IE)**. *IE* evaluates predictive performance for a single exhibit. For each observed visitor-exhibit pair (v, i) , we removed the log viewing time r_{vi} from the vector of the visitor’s log viewing durations, and computed a prediction \hat{r}_{vi} from the other observations. This experiment is lenient in the sense that all available observations except the observation for exhibit i are kept in a visitor’s viewing duration vector.
- **Progressive Visit (PV)**. *PV* evaluates performance as a museum visit progresses, i. e., as the number of viewed exhibits increases. For each visitor, we started with

an empty visit, and iteratively added each viewed exhibit to the visit history, together with its log viewing time. We then predicted the log viewing times of all yet unvisited exhibits.

- **Recommendation Potential (RP)**. *RP* assesses the recommendation potential of our models, i. e., it gives an indication as to whether our models can discover unvisited but personally interesting exhibits. We predicted the log viewing times of all unvisited exhibits for each visitor, given his/her complete visit history. We then counted the predicted log viewing durations that were significantly above the corresponding exhibit’s average log viewing time $\bar{r}_{.i}$. For this purpose, we used the 95% credible interval around $\bar{r}_{.i}$.

For the first two experiments, we used the *mean absolute error* (MAE) to measure predictive accuracy as follows:

$$\text{MAE} = \frac{1}{\sum_{v \in V} |I_v|} \sum_{v \in V} \sum_{i \in I_v} |r_{vi} - \hat{r}_{vi}|,$$

where I_v denotes a visitor v ’s set of exhibits for which predictions were computed. For *IE*, we calculated the total MAE for all visitors and all exhibits; and for *PV*, we computed the MAE across the yet unvisited exhibits for all visitors for each time fraction of a visit (to account for different visit lengths, we normalised all visits to a length of 1).

Figure 2 shows a plot of the relationship between the shrinkage weight ω and the MAE for *CFM*. To obtain the plot, we used the best-performing *CFM* parameterisation and varied ω over $[0, 1]$. For each ω , we averaged the MAEs obtained for the *IE* and *PV* experiments. The minimum MAE is achieved for $\omega \approx 0.75$. This value of ω was used in our comparative evaluations.

5.2 Results

For the *IE* experiment, *CFM* outperforms *MM*, achieving an MAE of 0.7868 (stderr 0.0068). *SPM* outperforms both *MM* and *CFM*, achieving an MAE of 0.7548 (stderr 0.0066). The performance differences are statistically significant with $p \ll 0.01$ (upper portion of Table 2, column ‘log t MAE’).

Computing MAEs with respect to log viewing times penalises errors for higher viewing times less than errors for lower viewing times, which is reasonable in our context. To illustrate the meaning of our results in terms of raw viewing times, we give a few exhibit-specific MAEs (calculated as for the *IE* experiment, but on raw viewing times). The lower portion of Table 2 shows these values for five exhibit areas *A* to *E*, which were selected on the basis of the variability of their viewing times and locations in Melbourne Museum (marked in Figure 1). The first column in Table 2 designates the exhibit area, the second and third column the mean and standard deviation of the distribution of viewing times for this area (in seconds) respectively (we used the parameters of the fitted log-normal models to compute estimates of the means and standard deviations), and the last column shows the MAE with respect to log viewing times. In addition, for each exhibit area, we split the data at the median of the fitted log-normal model, separating low and high viewing times. We then computed the MAEs for each half separately (fourth and fifth columns). As expected, the MAEs for the lower half are smaller than the MAEs for the upper half. For instance, *CFM* achieves an MAE of 52.6 seconds for the

Table 2. Model performance for the *IE* experiment (MAE)

Exhibit area	Viewing time			lowerMAE	upperMAE	log <i>t</i> MAE
	Mean	Stddev		Mean (Stderr)	Mean (Stderr)	Mean (Stderr)
Total <i>MM</i>						0.8618 (0.0071)
Total <i>CFM</i>						0.7868 (0.0068)
Total <i>SPM</i>						0.7548 (0.0066)
Area A	220	246	<i>CFM</i>	60.8 (6.3)	119.1 (12.4)	0.6024 (0.0492)
			<i>SPM</i>	53.2 (5.8)	111.4 (11.5)	0.5513 (0.0463)
Area B	115	156	<i>CFM</i>	29.8 (4.0)	84.4 (11.1)	0.6844 (0.0590)
			<i>SPM</i>	26.9 (3.7)	76.4 (10.7)	0.6245 (0.0588)
Area C	177	240	<i>CFM</i>	52.6 (6.5)	118.7 (19.2)	0.7038 (0.0587)
			<i>SPM</i>	42.8 (5.4)	110.7 (18.0)	0.6308 (0.0531)
Area D	87	149	<i>CFM</i>	20.7 (2.7)	64.9 (11.3)	0.8066 (0.0733)
			<i>SPM</i>	20.6 (3.3)	60.6 (10.7)	0.7588 (0.0719)
Area E	75	125	<i>CFM</i>	18.1 (3.5)	47.0 (10.8)	0.7896 (0.1118)
			<i>SPM</i>	18.1 (3.5)	49.3 (10.1)	0.7921 (0.1047)

low viewing times at area *C* (29.7% of the mean), and 118.7 seconds for the high viewing times (67.1% of the mean). For *CFM*, the average MAE (when averaged over the five exhibits) as a percentage of mean exhibit viewing time is 26.2% for the lower half, and 66.4% for the upper half. In contrast, *SPM* achieves 23.9% and 63.0% respectively.

The performance of *SPM*, *CFM* and the baseline *MM* for the *PV* experiment is depicted in Figure 3. *CFM* outperforms *MM* slightly (statistically significantly for visit fractions 0.191 to 0.374 and for several shorter intervals later on, $p < 0.05$). There is a significant improvement in performance for *SPM*, compared to both *MM* and *CFM* (statistically significant for visit fractions 0.019 to 0.922, $p < 0.05$). Drawing attention to the initial portion of the visits, *SPM*'s MAE decreases rapidly, whereas the MAE for *MM* and *CFM* remains at a higher level. Generally, the faster a model adapts to a visitor's interests, the more likely it is to quickly deliver personally useful recommendations. Such behaviour in the early stages of a museum visit is essential in order to build trust in the system's recommendations, and to guide a visitor in a phase of his/her visit where such guidance is most likely needed. As expected, *MM* performs at a relatively constant MAE level. For *CFM* and *SPM*, we expected to see a relative improvement in performance as the number of visited exhibits increases. However, this trend is rather subtle. Additionally, for all three models, there is a performance drop towards the end of a visit. We postulate that these phenomena may be explained, at least partially, by the increased influence of outliers on the MAE as the number of exhibits remaining to be viewed is reduced with the progression of a visit. This influence in turn offsets potential gains in performance obtained from additional observations. Our hypothesis is supported by a widening in the standard error bands for all models as a visit progresses, in particular towards the end (not shown in Figure 3 for clarity of presentation).

For the *RP* experiment, we obtained the following results. Per visitor (on average), *CFM* discovers 29.3 exhibits with predicted viewing times that are significantly higher than the average. This corresponds to 37.0% of the predictions per visitor (on average).

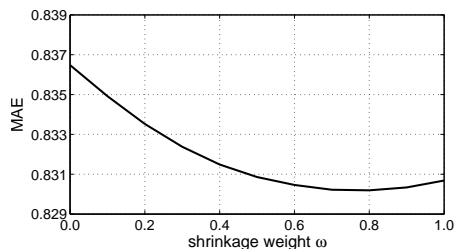


Fig. 2. *CFM* performance (shrinkage weight vs. MAE)

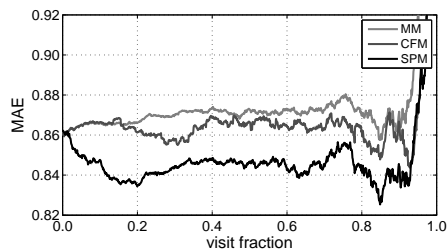


Fig. 3. Model performance for the *PV* experiment (MAE)

In comparison, *SPM* predicts 23.6 such exhibits (30.1% of the predictions). These numbers indicate that our models discover exhibits which visitors appear to be interested in but did not view. As *SPM* significantly outperforms *CFM* with respect to predictive accuracy, *SPM*'s percentage is most likely a more realistic estimate of the true potential of our models with respect to visitor support. A more conclusive interpretation requires a further, more rigorous investigation.

6 Discussion

Recommender systems have often been employed in virtual (i. e., non-physical) domains, where personalised recommendations are directly derived from predicted ratings, e. g., by recommending the items with the highest ratings. In contrast, in a physical domain, the transition from predicting a visitor's interests to recommendation generation is not trivial, as we do not want to recommend exhibits that visitors are going to see anyway. We suggest the following approach to address this problem. Firstly, use the predictions generated by interest-based predictive models (Section 4) to build a list of areas in the museum that a visitor is likely to be interested in, e. g., by determining whether a predicted interest is significantly higher than the (non-personalised) average interest. Secondly, form a list of exhibits from a location-based prediction of a visitor's pathway through the physical museum [7]. Then, after merging the lists appropriately, one can recommend exhibits that a visitor may be interested in but is likely to overlook. This approach requires a strategy for merging the lists, e. g., whether locations that a visitor is likely to visit anyway should be included (to help build trust in the system) or excluded (to avoid over-communication). The modality of the presentation, e. g., visualised on a site map or provided in textual or audio form, should be taken into account when selecting the exhibits to be recommended.

7 Conclusions and Future Work

In this paper, we proposed a computer-supported methodology that we used to collect pathways of visitors to Melbourne Museum. We presented two collaborative models for predicting a visitor's viewing times of unseen exhibits from his/her viewing times at visited exhibits — a memory-based nearest-neighbour collaborative filter (called *CFM*), and a model-based approach utilising the theory of Gaussian spatial processes (called *SPM*). Our models were evaluated with the dataset we collected. Our results show that both models attain a higher predictive accuracy than a non-personalised baseline, with

SPM outperforming the other models. Additionally, in the realistic *Progressive Visit* setting, *SPM* rapidly adapts to observed visitor behaviour, addressing the *new-user problem* of collaborative approaches.

In the future, we intend to investigate ways of hybridising *SPM* by incorporating content-based exhibit features into our distance measure. We also plan to combine our models with a model that predicts a visitor's pathway (i. e., a sequence of exhibits), and develop strategies for delivering useful personalised recommendations about exhibits.

Acknowledgements. This research was supported in part by grant DP0770931 from the Australian Research Council. The authors thank Carolyn Meehan and her team from Museum Victoria for fruitful discussions and their support; and David Abramson, Jeff Tan and Blair Bethwaite for their assistance with using their computer cluster.

References

1. Cheverst, K., Mitchell, K., Davies, N.: The role of adaptive hypermedia in a context-aware tourist guide. *Communications of the ACM* **45**(5) (2002) 47–51
2. Aroyo, L., Stash, N., Wang, Y., Gorgels, P., Rutledge, L.: CHIP demonstrator: Semantics-driven recommendations and museum tour generation. In: *Proc. of the Sixth Intl. Semantic Web Conf. (ISWC-07)*. (2007) 879–886
3. Petrelli, D., Not, E.: User-centred design of flexible hypermedia for a mobile guide: Reflections on the HyperAudio experience. *User Modeling and User-Adapted Interaction* **15**(3-4) (2005) 303–338
4. Hatala, M., Wakkary, R.: Ontology-based user modeling in an augmented audio reality system for museums. *User Modeling and User-Adapted Interaction* **15**(3-4) (2005) 339–380
5. Stock, O., Zancanaro, M., Busetta, P., Callaway, C., Krüger, A., Kruppa, M., Kuflik, T., Not, E., Rocchi, C.: Adaptive, intelligent presentation of information for the museum visitor in PEACH. *User Modeling and User-Adapted Interaction* **18**(3) (2007) 257–304
6. Albrecht, D.W., Zukerman, I.: Special issue on statistical and probabilistic methods for user modeling. *User Modeling and User-Adapted Interaction* **17**(1-2) (2007)
7. Bohnert, F., Zukerman, I., Berkovsky, S., Baldwin, T., Sonenberg, L.: Using interest and transition models to predict visitor locations in museums. *AI Communications* **21**(2-3) (2008) 195–202
8. Diamond, J.: *Practical Evaluation Guide – Tools for Museums and Other Informal Educational Settings*. AltaMira Press (1999)
9. Parsons, J., Ralph, P., Gallager, K.: Using viewing time to infer user preference in recommender systems. In: *Proc. of the AAAI Workshop on Semantic Web Personalization (SWP-04)*. (2004) 52–64
10. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.T.: An algorithmic framework for performing collaborative filtering. In: *Proc. of the 22th Annual Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR-99)*. (1999) 230–237
11. Banerjee, S., Carlin, B.P., Gelfand, A.E.: *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC (2004)
12. James, W., Stein, C.M.: Estimation with quadratic loss. In: *Proc. of the Fourth Berkeley Symp. on Mathematical Statistics and Probability, Vol. 1*. (1961) 361–379
13. Bohnert, F., Schmidt, D.F., Zukerman, I.: Spatial processes for recommender systems. In: *Proc. of the 21st Intl. Joint Conf. on Artificial Intelligence (IJCAI 2009)*. (2009) 2022–2027
14. Neal, R.M.: Slice sampling. *The Annals of Statistics* **31**(3) (2003) 705–767