Using Gaussian Spatial Processes to Model and Predict Interests in Museum Exhibits

Fabian Bohnert, Ingrid Zukerman, and Daniel F. Schmidt

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

{fabianb,ingrid,dschmidt}@infotech.monash.edu.au

Abstract

This paper adapts models from the area of spatial statistics to the task of predicting a user's interests (i. e., implicit item ratings) within a recommender system in the museum domain. We develop a model based on Gaussian spatial processes, and discuss two ways of computing item-to-item distances in the museum setting. Our model was evaluated with a real-world dataset collected by tracking visitors in a museum. Overall, our model attains a higher predictive accuracy than nearest-neighbour collaborative filters. In addition, the model variant using physical distances outperforms that using distances computed from item-to-item similarities.

1 Introduction

Spatial processes (*random fields*) are a subclass of *stochastic processes* which are applied to domains that have a geospatial interpretation, e. g., [Diggle *et al.*, 1998; Banerjee *et al.*, 2004]. They are typically used in the field of spatial statistics to model spatial associations between a set of observations made at certain locations, and to predict values at locations where no observations have been made. This paper applies such models to the prediction of a user's interests or item ratings in recommender systems (RS). We develop our *Spatial Process Model (SPM)* by adapting a Gaussian spatial process model to the RS scenario, and demonstrate our model's applicability to the task of predicting implicit ratings in the museum domain. The use of spatial processes requires a measure of distance between items in addition to users' ratings. This measure, which is non-specific (e. g., it may be a physical or a conceptual distance), can be readily obtained in most cases. For example, distances could be computed from feature vectors representing the items (similarly to content-based RS), from item-to-item similarities (similarly to item-to-item collaborative filtering [Sarwar *et al.*, 2001]), or from physical distance. In this paper, we explore the latter two measures.

Our application scenario is motivated by the need to automatically recommend exhibits to museum visitors, based on non-intrusive observations of their actions in the physical space. Employing RS in this scenario is challenging due to (1) the physical nature of the domain, (2) having exhibit viewing times rather than explicit ratings, and (3) predictions differing from recommendations (we do not want to recommend exhibits that visitors are going to see anyway). We turn the first challenge into an advantage by exploiting the fact that physical distances between exhibits are meaningful, enabling the use of walking distance between exhibits to calculate (content) distance. This supports the direct, interpretable application of spatial processes by using a simple parametric Gaussian spatial process model (with the ensuing low variance in parameter estimates), compared to more complex non-parametric approaches, e. g., [Schwaighofer *et al.*, 2005]. The second challenge, which stems from the variable semantics of viewing times (time t for different exhibits could mean interest or boredom), is naturally addressed by *SPM*'s structure. The third challenge can be addressed by (a) using *SPM* to build a model of a visitor's interests in unseen exhibits, (b) inferring a predictive model of a visitor's pathway through the remainder of the museum [Bohnert *et al.*, 2008], and (c) combining these models to recommend exhibits of interest that may be overlooked if the predicted pathway is followed.

SPM was evaluated with a real-world dataset of time spans spent by museum visitors at exhibits (viewed as implicit ratings). We compared our model's performance to that of (1) a baseline model which delivers a non-personalised prediction, and (2) a nearest-neighbour collaborative filter incorporating performance-enhancing modifications, e. g., [James and Stein, 1961; Herlocker *et al.*, 1999]. Our results show that *SPM* significantly outperforms both models.

The paper is organised as follows. In Section 2, we discuss related research. Section 3 describes our domain and dataset. Our spatial processes approach for modelling and predicting exhibit interests is developed in Section 4. In Section 5, we present the results of our evaluation, followed by a discussion in Section 6 and our conclusions in Section 7.

2 Related Research

Recommender systems (RS) are designed to direct users to personally interesting items in situations where the amount of available information exceeds the users' processing capability [Resnick and Varian, 1997; Burke, 2002]. Typically, such systems (1) use information about a user (i. e., a user model) to predict ratings of items that the user has not yet considered, and (2) recommend suitable items based on these predictions. *Collaborative* modelling techniques constitute one of the main model classes applied in RS [Albrecht and Zukerman, 2007]. They base their predictions upon the assumption that users who have agreed in their behaviour in the past will agree in the future.

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, for which features are not readily apparent. The two main collaborative approaches are *memory-based* and *model-based*. Previous research has mainly focused on memory-based approaches, such as nearest-neighbour models (classic collaborative filtering), e. g., [Herlocker *et al.*, 1999]. The main drawback of memory-based algorithms is that they operate over the entire user database to make predictions. In contrast, model-based approaches use techniques such as Bayesian networks, latent-factor models and artificial neural networks, e. g., [Breese *et al.*, 1998; Bell *et al.*, 2007], to first learn a statistical model in an offline fashion, and then use it to make predictions and generate recommendations. This decomposition can significantly speed up the recommendation generation process.

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 [Cheverst *et al.*, 2002] 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 [Aroyo *et al.*, 2007] 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 models that require an explicit initialisation.

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 are usually updated from a user's interactions with the system, the focus being on adapting content presentation as opposed to predicting and recommending exhibits to be viewed. For example, *HyperAudio* [Petrelli and Not, 2005] 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* [Hatala and Wakkary, 2005] 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 [Stock *et al.*, 2007] 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.

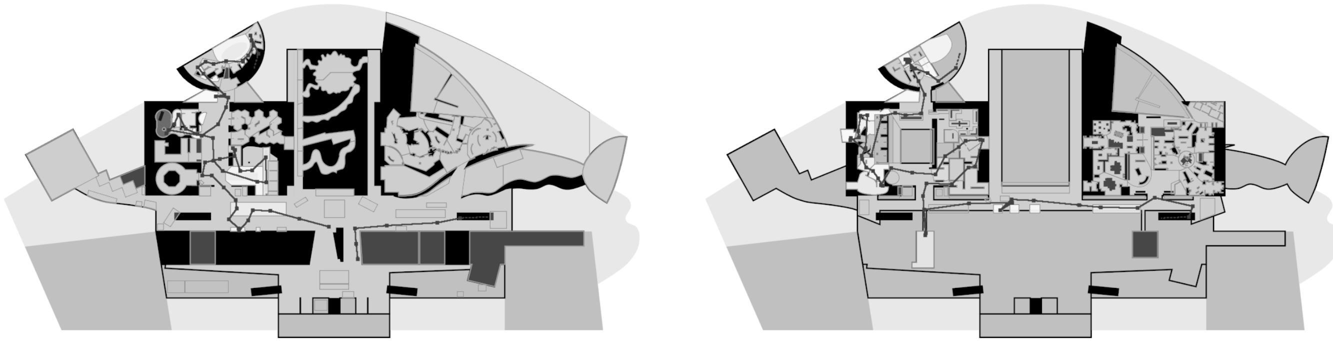
These systems, like most systems in the museum domain, rely on knowledge-based user models in some way, and hence, require an explicit, a-priori engineered representation of the domain knowledge. In contrast, our research investigates non-intrusive statistical user modelling and recommendation techniques that do not require such an explicit domain knowledge representation [Albrecht and Zukerman, 2007].

3 Domain and Dataset

The GECKO project endeavours to develop user modelling and personalisation techniques for information-rich physical spaces, relying on non-intrusive observations of users' behaviour [Bohnert *et al.*, 2008]. 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 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.

Museums such as Melbourne Museum (Melbourne, Australia) display thousands of exhibits distributed over many separate galleries and exhibitions. Normally, visitors do not require recommendations to travel between individual, logically related exhibits in close physical proximity. Rather, they may prefer recommendations regarding physically separated 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. Figure 1 depicts the site map of Melbourne Museum showing these exhibit areas, together with one of the visitor pathways we collected.

To obtain our dataset, we manually tracked visitors to Melbourne Museum from April to June 2008, using a custom-made tracking tool running on laptop computers [Bohnert and Zukerman, 2009]. In total, we recorded over 170 visitor pathways. We only tracked 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 our tracking software, 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 a post-processing tool we developed. For instance, we removed tracking 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 computer 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



(a) Melbourne Museum – Ground level (b) Melbourne Museum – Upper level

Figure 1: Visitor pathway visualised on a site map of Melbourne Museum

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

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 an automated RS in a museum requires suitable positioning technologies to non-intrusively track visitors, and models to infer which exhibits are being viewed. Although our dataset was obtained manually, it provides information of the type that may be inferred from sensing data (the work described in [Schmidt *et al.*, 2009] links sensory and manually obtained information). 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 model.

4 Using Gaussian Spatial Processes to Model and Predict Visitors' Exhibit Interests

In this section, we first describe how we use viewing time to quantify interest in exhibits (Section 4.1), and discuss the applicability of spatial process models [Banerjee *et al.*, 2004] to the prediction of a visitor's interest in exhibits in our RS scenario (Section 4.2). We then propose a model-based collaborative approach based on the theory of Gaussian spatial processes for predicting a visitor's ($\log$) viewing times (viewed as exhibit interests) from non-intrusive observations of his/her ($\log$) viewing times at visited exhibits (Section 4.3).

4.1 From Viewing Time to Exhibit Interest

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 [Parsons *et al.*, 2004]. 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), 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)* [Schwarz, 1978]. 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). Hence, we transformed all viewing times to their $\log$-equivalent to obtain approximately 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.

4.2 Spatial Statistics in the Context of Our Application Scenario

Spatial statistics is concerned with the analysis and prediction of geographic data [Banerjee *et al.*, 2004]. 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 RS 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 (i. e., exhibit interests). We use $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ to denote the locations of exhibits $i, j \in I = \{1, \ldots, n\}$ in a space providing such a distance measure, i. e., $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$. For example, $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ can be computed from feature vectors representing the items (similarly to content-based RS), from item-to-item similarities (similarly to item-to-item collaborative filtering [Sarwar *et al.*, 2001]), or from physical distance.

In this paper, we explore the two latter options: *Item-to-Item Distance* and *Physical Distance*.

- **Item-to-Item Distance (I2I).** Item-to-item collaborative filtering [Sarwar *et al.*, 2001] utilises a database of ratings to compute item-to-item similarities, and predicts a current user's rating of an unseen item from his/her ratings of those items that are most similar to the item in question. Inspired by how item-to-item similarities are computed in this process, we use the observed $\log$ viewing times to derive the *I2I* distance measure as follows. We first transform the $\log$ viewing times into z-scores by normalising the values for each visitor separately. This ensures that varying viewing behaviour does not affect the similarity computation.[1] Secondly, we calculate item-to-item similarities using Pearson's correlation coefficient on the normalised $\log$ viewing times of exhibits i and j (using only the normalised $\log$ viewing times of those visitors that have viewed both exhibits i and j). The resulting similarity value from within the interval $[-1, 1]$ is finally transformed into a distance measure by mapping it onto a value in $[0, 1]$ (a similarity value of -1 yields a distance of 1, and a similarity of 1 yields a distance of 0).
- **Physical Distance (PD).** 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 *PD* as a measure of distance between exhibits. Specifically, a SVG file-based representation of Melbourne Museum was used to calculate the walking distances by mapping the site map (Figure 1) 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 $[0, 1]$.

4.3 Our Gaussian Spatial Process Model

In this section, we utilise theory from the area of spatial statistics (Section 4.2) to formulate a Gaussian spatial process model, called *Spatial Process Model (SPM)*, for predicting a museum visitor's interests in unseen exhibits (i. e., $\log$ viewing times) from his/her viewing behaviour at visited exhibits.

Let $U = \{1, \ldots, m\}$ be the set of all visitors, and $I = \{1, \ldots, n\}$ be the set of all items. Typically, for a visitor $u \in U$, we have viewing times for only a subset of I, say for n_u exhibits. Denoting a visitor's $\log$ viewing time vector with $\boldsymbol{r}_u$, we collect all observed $\log$ viewing times into a vector $\boldsymbol{r} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m)$ of dimension $\sum_{u=1}^{m} n_u$. Associated with each exhibit $i \in I$ is a $\log$ viewing time mean μ_i and a standard deviation σ_i. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ be the vector of mean $\log$ viewing times, and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ the vector of standard deviations. Furthermore, $\boldsymbol{\mu}_u$ and $\boldsymbol{\sigma}_u$ are the vectors of means and standard deviations respectively for only those exhibits viewed by a visitor u. For example, if visitor 1 viewed exhibits 2, 3, 7 and 9, then $\boldsymbol{\mu}_1 = (\mu_2, \mu_3, \mu_7, \mu_9)$ and $\boldsymbol{\sigma}_1 = (\sigma_2, \sigma_3, \sigma_7, \sigma_9)$.

Similarly to spatial processes, *SPM* assumes a special correlation structure between the viewing times of different exhibits. In our experiments, we use a powered exponential [Banerjee *et al.*, 2004]:

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

where $\phi > 0$ and $0 < \nu < 2$. That is, $\rho(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|; \phi, \nu)$ models the correlation between the $\log$ viewing times of exhibits i and j ($\rho(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|; \phi, \nu)$ depends on the sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ of exhibits i and j only through the distance $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$). Let $H(\phi, \nu)$ be a correlation matrix with components $(H(\phi, \nu))_{ij} = \rho(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|; \phi, \nu)$ collecting all these correlations, and let $H_u(\phi, \nu)$ denote a visitor u's correlation matrix (dimension $n_u \times n_u$). That is, $H_u(\phi, \nu)$ corresponds to $H(\phi, \nu)$ without the rows and columns for unvisited exhibits. Also, let $\boldsymbol{\theta} = \left(\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau^2, \phi, \nu\right)$ be a vector representing the $2n + 3$ model parameters, where τ^2 denotes the variance of non-spatial error terms necessary to fully specify the model (these terms model non-spatial variation in the data). Then, modelling the data using Gaussian spatial processes (a detailed derivation appears in [Bohnert *et al.*, 2009]), $\boldsymbol{r}$ given $\boldsymbol{\theta}$ is multivariate normal of dimension $\sum_{u=1}^{m} n_u$. As the viewing times of different visitors $u = 1, \ldots, m$ are independent, the model simplifies to

$$\boldsymbol{r}_u \,|\, \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}_u, \Sigma_u\right) \text{ for all } u = 1, \ldots, m, \tag{1}$$

where $\Sigma_u = \boldsymbol{\sigma}_u \mathit{1}_{n_u} H_u(\phi, \nu) \boldsymbol{\sigma}_u \mathit{1}_{n_u} + \tau^2 \mathit{1}_{n_u}$ is a visitor u's covariance matrix, and $\mathit{1}_{n_u}$ is the identity matrix of dimension $n_u \times n_u$.

We employ Bayesian inference using *SPM*'s likelihood function derived from Equation 1 to estimate $\boldsymbol{\theta}$ from $\boldsymbol{r}$ (in particular, we use *slice Gibbs sampling* [Neal, 2003]). 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.

Given the model parameters $\boldsymbol{\theta} = \left(\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau^2, \phi, \nu\right)$, our model is fully specified, and we can use standard multivariate normal theory to predict a current visitor a's $\log$ viewing times of unseen exhibits, say $\boldsymbol{r}_{a,1}$, from a vector of observed $\log$ viewing times $\boldsymbol{r}_{a,2}$. This is because $(\boldsymbol{r}_{a,1}, \boldsymbol{r}_{a,2}) \,|\, \boldsymbol{\theta}$ is normally distributed (similarly to Equation 1). If we use the following notation

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

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

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

where $\mathbb{E}\left(\boldsymbol{r}_{a,1} | \boldsymbol{r}_{a,2}, \boldsymbol{\theta}\right)$ represents a personalised prediction of the $\log$ viewing times $\boldsymbol{r}_{a,1}$. Additionally, a measure of confidence in this prediction can be easily derived from

[1] We also tested a variant of the *I2I* measure without visitor-wise normalisation. However, this variant yielded inferior results.

$\mathrm{Cov}\left(\boldsymbol{r}_{a,1}|\boldsymbol{r}_{a,2},\boldsymbol{\theta}\right)$, i. e., by using the variances on the diagonal of this matrix.

Being a model-based approach, *SPM* offers advantages over memory-based collaborative filters. For instance, the model parameters $\boldsymbol{\theta} = \left(\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau^2, \phi, \nu\right)$ 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 decoupling the model-fitting phase from the prediction phase.

5 Evaluation

This section reports on the results of an evaluation performed with our dataset (Section 3), including comparison with a nearest-neighbour collaborative filter.[2]

5.1 Experimental Setup

To evaluate the predictive performance of our *Spatial Process Model (SPM)*, we implemented two additional models: *Mean Model (MM)* and *Collaborative Filter (CF)*. *MM*, which we use as a baseline, predicts the $\log$ viewing time of an exhibit area i to be its (non-personalised) mean $\log$ viewing time μ_i. For *CF*, we implemented a nearest-neighbour collaborative filtering algorithm, and added modifications from the literature that improve its performance, such as shrinkage to the mean [James and Stein, 1961] and significance weighting [Herlocker *et al.*, 1999]. Additionally, to ensure that varying exhibit area complexity does not affect the similarity computation for selecting the nearest neighbours (viewing time increases with exhibit complexity), we transformed the $\log$ viewing times into z-scores by normalising the values for each of the exhibit areas separately. Visitor-to-visitor differences with respect to their mean viewing durations were removed by transforming predictions to the current visitor's viewing-time scale [Herlocker *et al.*, 1999]. Refer to [Bohnert and Zukerman, 2009] for a detailed description of *CF*. We tested several thousand different parameterisations, but in this paper, we report only on the performance of the best one.

Due to the relatively small dataset, we used leave-one-out cross validation to evaluate the performance of the different models. That is, for each visitor, we trained the models with a reduced dataset containing the data of 157 of the 158 visit trajectories, and used the withheld visitor pathway for testing. To train and instantiate the *SPM* variants (i. e., *SPM-I2I* and *SPM-PD*), we obtained a sample of $\boldsymbol{\theta} = \left(\boldsymbol{\mu}, \boldsymbol{\sigma}, \tau^2, \phi, \nu\right)$ from $p(\boldsymbol{\theta}|\boldsymbol{r})$ by performing slice Gibbs sampling [Neal, 2003] on the training data. For each of the 129 free model parameters,[3] we used (uninformative) independent uniform prior distributions. We used every 20-th sample after a burn-in phase of 1000 iterations as a sample of $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{r})$, and stopped the sampling procedure after 8000 iterations. Thus, in total, we obtained 350 samples of $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{r})$. This procedure was followed to obtain samples of $\boldsymbol{\theta}$ for both *SPM* variants, i. e., for both distance measures *I2I* and *PD* (Section 4.2). We then used the posterior means estimated from these samples to compute predictions by conditioning a multivariate normal distribution (Section 4.3). We improved *SPM-I2I*'s predictive performance by using the (non-personalised) mean $\log$ viewing time μ_i as a prediction whenever the conditioning would have been based on fewer than K $\log$ viewing times (in our case, $K = 19$). This modification was not applied to *SPM-PD*. For *CF*, predictions were computed from the ratings of the nearest neighbours; and for *MM*, we used μ_i, estimated from the appropriate reduced dataset, as a prediction.

We performed two types of experiments: *Individual Exhibit* and *Progressive Visit*.

- **Individual Exhibit (IE).** *IE* evaluates predictive performance for a single exhibit. For each observed visitor-exhibit area pair (u, i), we removed the observation r_{ui} from the vector of visitor u's $\log$ viewing durations, and computed a prediction $\hat{r}_{ui}$ from the other observations. This experiment is lenient in the sense that all available observations except the observation for exhibit area 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 exhibit areas increases. For each visitor, we started with an empty visit, and iteratively added each viewed exhibit area to the visit history, together with its $\log$ viewing time. We then predicted the $\log$ viewing times of all yet unvisited exhibit areas.

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

$$\mathrm{MAE} = \frac{1}{\sum_{u \in U} |I_u|} \sum_{u \in U} \sum_{i \in I_u} |r_{ui} - \hat{r}_{ui}|,$$

where I_u denotes a visitor u's set of exhibit areas for which predictions were computed. For *IE*, we calculated the total MAE for all valid visitor-exhibit area pairs; and for *PV*, we computed the MAE for the yet unvisited exhibit areas for all visitors at each time fraction of a visit (to account for different visit lengths, we normalised all visits to a length of 1).

5.2 Results

Table 2 shows the results for the *IE* experiment, where both spatial models (*SPM-I2I* and *SPM-PD*) outperform both *MM* and *CF*. Specifically, *SPM-I2I* achieves an MAE of 0.7756 (stderr 0.0068), and *SPM-PD* attains an MAE of 0.7548 (stderr 0.0066), outperforming *SPM-I2I* as well. The pairwise performance differences are statistically significant with $p \ll 0.01$ for all model pairings.

The performance of *SPM-PD*, *SPM-I2I*, *CF* and the baseline *MM* for the *PV* experiment is depicted in Figure 2. *CF* outperforms *MM* slightly (statistically significantly for visit fractions 0.191 to 0.374 and for several shorter intervals later on, $p < 0.05$). More importantly, both *SPM-I2I* and *SPM-PD* perform significantly better than *MM* and *CF*. For *SPM-I2I*, this performance increase is statistically significant for visit fractions 0.189 to 0.960 when comparing to *MM*, and except

[2]For our experiments, we ignore travel between exhibit areas, and collapse multiple viewing events of one area into one event.

[3]We set $\sigma_i = \sqrt{\sigma^2_{\boldsymbol{r},i} - \tau^2}$ to speed up the sampling process, where $\sigma^2_{\boldsymbol{r},i}$ denotes the sample variance of the $\log$ viewing times at exhibit i, calculated from the observed $\log$ viewing times r_{ui}. This reduces the number of free parameters from 255 (126×2+3) to 129.

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

	MAE	Stderr
Mean Model (MM)	0.8618	0.0071
Collaborative Filter (CF)	0.7868	0.0068
Spatial Process Model using *I2I* (SPM-*I2I*)	0.7756	0.0068
Spatial Process Model using *PD* (SPM-*PD*)	0.7548	0.0066

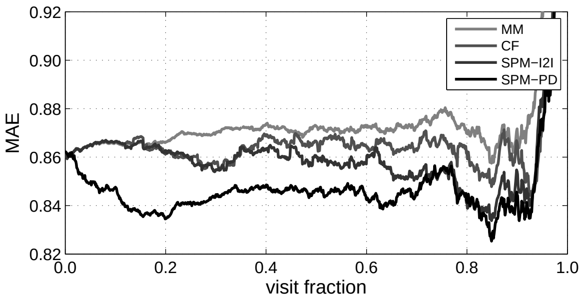


Figure 2: Model performance for the *PV* experiment (MAE)

for a few short intervals, for visit fractions 0.375 to 0.902 when comparing to *CF*. In comparison, *SPM-PD* performs significantly better than both *MM* and *CF* for visit fractions 0.019 to 0.922 (statistically significantly, $p < 0.05$). Additionally, *SPM-PD* outperforms *SPM-I2I* until visit fraction 0.660 (statistically significantly, $p < 0.05$). Drawing attention to the initial portion of the visits, *SPM-PD*'s MAE decreases rapidly, whereas the MAE for *MM* and *CF* 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 RS, and to guide a visitor in a phase of the visit where such guidance is most likely needed. A similar improvement in performance cannot be observed for *SPM-I2I*, which suggests that a visitor's exhibit interests observed in close physical proximity are better predictors of interests in unseen exhibits than interests in exhibits with positively correlated viewing times. As expected, *MM* performs at a relatively constant MAE level. For *CF*, *SPM-I2I* and *SPM-PD* we expected to see an improvement in performance (relative to *MM*) as the number of visited exhibit areas increases. However, this trend is rather subtle (it can be observed when plotting the models' performance relative to *MM*). Additionally, for all four 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 exhibit areas 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 2 for clarity of presentation). However, this behaviour requires further, more rigorous investigation.

6 Discussion

SPM offers advantages over other model-based approaches in that, unlike neural networks (and memory-based techniques), it returns the confidence in a prediction, and its parameters have a clear interpretation; unlike Bayesian networks, our model does not require a domain-specific adaptation, such as designing the network topology. In addition, the distance measure endows our model with capabilities of hybrid RS [Burke, 2002; Albrecht and Zukerman, 2007] by seamlessly supporting the incorporation of other types of models (e. g., content-based). The distance measure also alleviates the *cold-start problem*. The *new-item problem* is addressed by utilising the (distance-based) correlation between this item and the other items. The *new-user problem* is similarly handled through the correlation between items rated by a user and the other items (when utilising *Physical Distance* as the distance measure, our model can make useful personalised predictions after only one item has been rated).

Our dataset is relatively small compared to other real-world RS applications. Although a high number of ratings per user slows down the slice Gibbs sampler due to repeated inversion of matrices of high dimension, employing our model with larger datasets should not represent a problem in practice. This is because the number of ratings per user is usually small compared to the number of users and items, and the computational complexity of evaluating the likelihood function depends only linearly on the number of users in the database.

7 Conclusions and Future Work

In this paper, we utilised the theory of spatial processes to develop a model-based approach for predicting users' interests in exhibits (i. e., implicit item ratings) within a RS in the museum domain. We applied our model to a real-world dataset collected by tracking visitors in a museum, using two measures of item-to-item (content) distance: (1) distances computed from item-to-item similarities (as in item-to-item collaborative filtering), and (2) physical walking distance. For both distance measures, our model attains a higher predictive accuracy than nearest-neighbour collaborative filters. Additionally, the model variant using physical distances outperforms that using distances computed from item-to-item similarities. Under the realistic *Progressive Visit* setting, our model using physical distance to measure item-to-item distance rapidly adapts to a user's ratings (starting from as little as one rating), thus alleviating the *new-user problem* common to collaborative filtering. This is not the case for the model variant based on distance computed from item-to-item similarities, which suggests that a visitor's interests observed for exhibits in close physical proximity are better predictors of interests in unseen exhibits than those interests for exhibits with positively correlated viewing times.

In the future, we intend to hybridise our model by incorporating content-based item features into our distance measure, and to explore hybrids of models utilising a variety of item-to-item distances. We also plan to extend our model to fit non-Gaussian item ratings, e. g., [Diggle *et al.*, 1998; Yu *et al.*, 2006].

Acknowledgments

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 their assistance; and David Abramson, Jeff Tan and Blair Bethwaite for their help with the computer cluster.

References

[Albrecht and Zukerman, 2007] D.W. Albrecht and I. Zukerman. Introduction to the special issue on statistical and probabilistic methods for user modeling. *User Modeling and User-Adapted Interaction*, 17(1-2):1–4, 2007.

[Aroyo *et al.*, 2007] L. Aroyo, N. Stash, Y. Wang, P. Gorgels, and L. Rutledge. CHIP demonstrator: Semantics-driven recommendations and museum tour generation. In *Proceedings of the Sixth International Conference on the Semantic Web (ISWC-07)*, pages 879–886, 2007.

[Banerjee *et al.*, 2004] S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, 2004.

[Bell *et al.*, 2007] R. Bell, Y. Koren, and C. Volinsky. Chasing \$1,000,000: How we won the Netflix progress prize. *ASA Statistical and Computing Graphics Newsletter*, 18(2):4–12, 2007.

[Bohnert and Zukerman, 2009] F. Bohnert and I. Zukerman. Non-intrusive personalisation of the museum experience. In *Proceedings of the 1st and 17th International Conference on User Modeling, Adaptation, and Personalization (UMAP-09)*, pages 197–209, 2009.

[Bohnert *et al.*, 2008] F. Bohnert, I. Zukerman, S. Berkovsky, T. Baldwin, and L. Sonenberg. Using interest and transition models to predict visitor locations in museums. *AI Communications*, 21(2-3):195–202, 2008.

[Bohnert *et al.*, 2009] F. Bohnert, D.F. Schmidt, and I. Zukerman. Spatial processes for recommender systems. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-09)*, 2009.

[Breese *et al.*, 1998] J.S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th International Conference on Uncertainty in Artificial Intelligence (UAI-98)*, pages 42–52, 1998.

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

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

[Diggle *et al.*, 1998] P.J. Diggle, J.A. Tawn, and R.A. Moyeed. Model-based geostatistics. *Applied Statistics*, 47(3):299–350, 1998.

[Hatala and Wakkary, 2005] M. Hatala and R. 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.

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

[James and Stein, 1961] W. James and C.M. Stein. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1*, pages 361–379, 1961.

[Neal, 2003] R.M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003.

[Parsons *et al.*, 2004] J. Parsons, P. Ralph, and K. Gallager. Using viewing time to infer user preference in recommender systems. In *Proceedings of the AAAI Workshop on Semantic Web Personalization (SWP-04)*, pages 52–64, 2004.

[Petrelli and Not, 2005] D. Petrelli and E. 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.

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

[Sarwar *et al.*, 2001] B. Sarwar, G. Karypis, J.A. Konstan, and J.T. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on the World Wide Web (WWW-01)*, pages 285–295, 2001.

[Schmidt *et al.*, 2009] D.F. Schmidt, I. Zukerman, and D.W. Albrecht. Assessing the impact of measurement uncertainty on user models in spatial domains. In *Proceedings of the 1st and 17th International Conference on User Modeling, Adaptation, and Personalization (UMAP-09)*, pages 210–222, 2009.

[Schwaighofer *et al.*, 2005] A. Schwaighofer, V. Tresp, and K. Yu. Learning Gaussian process kernels via hierarchical Bayes. In *Advances in Neural Information Processing Systems 17 (NIPS-04)*, pages 1209–1216, 2005.

[Schwarz, 1978] G.E. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[Stock *et al.*, 2007] O. Stock, M. Zancanaro, P. Busetta, C. Callaway, A. Krüger, M. Kruppa, T. Kuflik, E. Not, and C. Rocchi. Adaptive, intelligent presentation of information for the museum visitor in PEACH. *User Modeling and User-Adapted Interaction*, 18(3):257–304, 2007.

[Yu *et al.*, 2006] S. Yu, K. Yu, V. Tresp, and H.-P. Kriegel. Collaborative ordinal regression. In *Proceedings of the 23rd International Conference on Machine Learning (ICML-06)*, pages 1089–1096, 2006.