

Modelling and Predicting Movements of Museum Visitors: A Simulation Framework for Assessing the Impact of Sensor Noise on Model Performance

Fabian Bohnert, Ingrid Zukerman,
and David W. Albrecht

Faculty of Information Technology
Monash University, Australia

firstname.lastname@monash.edu

Timothy Baldwin

Dept. of Comp. Sci. and Soft. Eng.
The University of Melbourne, Australia

tb@ldwin.net

Abstract

We present a simulation framework to examine the impact of sensor noise on the performance of user models in the museum domain. Our contributions are: (1) models to simulate noisy visit trajectories as time-stamped sequences of (x, y) positional coordinates which reflect *walking* and *hovering* behaviour; (2) a discriminative inference model that distinguishes between hovering and walking on the basis of (simulated) noisy sensor observations; (3) a model that infers viewed exhibits from hovering coordinates; and (4) a model that predicts the next exhibit on the basis of inferred (rather than known) viewed exhibits. Our staged evaluation assesses the effect of these models (in combination with sensor noise) on inferential and predictive performance, thus shedding light on the reliability attributed to inferences drawn from sensor observations.

1 Introduction

The construction of models of visitors to public spaces, in particular museums, has been of interest to the user modelling and cultural tourism communities for some time [Cheverst *et al.*, 2002; Hatala and Wakkary, 2005; Stock *et al.*, 2007]. These models are used to predict visitors' interests in order to personalise the content of presentations, or make recommendations of locations (e. g., exhibits) to be visited. In most systems developed to date, these user models are acquired through the active participation of the visitors, e. g., by providing feedback through a device. This requirement imposes a burden on the visitors, which in turn may reduce the reliability of the obtained information, e. g., if visitors provide feedback only occasionally.

Recent advances in mobile computing and sensing technologies have enabled the instrumentation of physical public spaces, which in turn has enabled the automatic tracking of visitors' movements [Hazas *et al.*, 2004; Lassabe *et al.*, 2009; Philipose *et al.*, 2004]. Information regarding visitors' whereabouts and the time spent at different locations supports the automatic inference of visitors' interests and the prediction of their trajectories [Bohnert and Zukerman, 2009]. Clearly, inferences from positional and timing information are more indirect and uncertain than visitors' direct feedback. However,

the information stream is reliable, as opposed to information obtained from visitors' direct participation.

In order to personalise content and generate recommendations on the basis of information provided by unobtrusive sensors (rather than from user participation), questions of interest include: (1) how to infer a visitor's viewed exhibits solely from sensor readings; and (2) how to predict the next exhibit(s) a visitor is likely to view. In this paper, we present a realistic simulation model which offers some insights to answer these questions, and may be employed to make decisions regarding the instrumentation of a space.

In previous research, we offered a simulation framework for investigating the impact of different sensing technologies on the predictive performance of user models [Schmidt *et al.*, 2009]. The aim was to provide a practical solution to the problem of assessing the accuracy of the user models that can be derived from a sensor-based system prior to actually deploying a particular technology. However, that work made strong simplifying assumptions that affected the realism of the framework, and hence the significance and usefulness of its results, viz: (1) sensors can detect, with some error, a single square (in a grid representation of the museum floor) where a visitor is *statically positioned* while viewing an exhibit i_k ; and (2) the previously viewed exhibits i_1, \dots, i_{k-1} are known (not just the previous coordinates of a visitor) when predicting the next exhibit i_{k+1} . In reality, people tend not to remain stationary at an exhibit, and they certainly do not 'teleport' between squares on the floor. Rather, they *walk* between exhibits, and often *hover* around an exhibit to view it from different angles or distances. Thus, when sensing a visitor's movements in a museum, the best we can hope for is a time-stamped trajectory of (x, y) coordinates (sampled at a particular rate), where the observed coordinates diverge from the true positions of the visitor by some sensor error. As a result, the sequence of previously viewed exhibits cannot be known with certainty — at best a likely sequence of exhibits can be inferred from the sensor observations.

In this paper, we propose a simulation framework that eschews the above assumptions, significantly extending our previous work and the insights obtained from it. Specifically, our contributions are: (1) models to simulate noisy visit trajectories as time-stamped sequences of (x, y) positional coordinates which reflect *walking* and *hovering* behaviour; (2) a discriminative inference model that distinguishes be-

tween hovering and walking on the basis of noisy sensor observations; (3) a model that infers likely viewed exhibits from time-stamped sequences of hovering coordinates (instead of a single static grid square per exhibit as done in our previous work); and (4) a model that predicts the next exhibit on the basis of these inferred (rather than known) viewed exhibits. At present, we assume that the sensors can only track a visitor’s position. However, our models may be extended to incorporate orientation information and occasional user feedback to improve the accuracy of inferences obtained from sensor readings, and hence the predictions of subsequent exhibits.

The research in this paper builds on the framework described by Schmidt *et al.* [2009], which comprises a predictive user model of exhibits to be viewed, and a spatial viewing model of positions from which each exhibit can be seen. Like Schmidt *et al.*, we evaluate our framework in the context of the Marine Life Exhibition at Melbourne Museum. In this paper, we augment the evaluations done by Schmidt *et al.*, presenting the results of a *staged* evaluation which examines the effect of different information-based models, in combination with sensor noise, on inferential and predictive performance.

This paper is organised as follows. Section 2 discusses related research. Section 3 briefly summarises the key components of our previous simulation framework. Our approach for simulating detailed coordinate-based visit trajectories is presented in Section 4, and our inference and prediction models are described in Section 5. The results of our evaluation are presented in Section 6, followed by concluding remarks in Section 7.

2 Related Research

The research community has initiated a wealth of projects that investigate user modelling and personalisation technology in the context of physical spaces. For example, in the museum domain, *HyperAudio* dynamically adapted hyperlinks and presented content to stereotypical assumptions about a visitor, and to what the visitor has already accessed through a mobile device and seems interested in [Petrelli and Not, 2005]. The *CHIP* project harnessed Semantic Web techniques to provide personalised access to digital museum collections both online and in the physical museum [Wang *et al.*, 2009]. This was done by using explicitly initialised user models. The *Kubadji* project investigated user and language modelling techniques that rely on mobile technology deployed in museums [Bohnert and Zukerman, 2009]. While the focus was on modelling visitors based on non-intrusive observations that can be derived from sensor readings, the project did not evaluate its models with real-world sensing technology.

In contrast to these projects, which did not employ real-world sensing technology, other research projects incorporated wireless technology or sensor networks. The *GUIDE* project developed a handheld tourist guide for visitors to the city of Lancaster, UK [Cheverst *et al.*, 2002]. It employed user models obtained from explicit user input to generate dynamic and user-adapted city tours, where the order of the visited locations could be varied. The project used wireless access points to stream content data to a user’s device, but did not employ the wireless network to localise the user. The

PEACH project developed technology which adapts its user model on the basis of both explicit user feedback and implicit observations of a user’s interactions with a mobile device [Stock *et al.*, 2007]. This user model was used to generate personalised multimedia presentations for museum visitors. The *PEACH* project also explored simple localisation technology, but did not derive user modelling information from sensor readings. The augmented audio reality system for museums *ec(h)o* adapted its user model on the basis of a visitor’s movements through the exhibition space and his/her interactions with the system [Hatala and Wakkary, 2005]. The collected user modelling data were used to deliver personalised information associated with exhibits via audio display. However, the project did not investigate the effect of localisation accuracy on the quality of the resultant user modelling information.

In contrast to the above research, this paper investigates the impact of using sensing technology as a means for gathering information about a user, i. e., to learn a user model. To this effect, we offer a simulation framework which generates noisy visit trajectories that reflect walking and hovering behaviour, and investigate the relationship between sensor noise and inferential and predictive user model performance.

3 Prerequisites

This section briefly summarises four key components of the simulation framework introduced by Schmidt *et al.* [2009], which is extended in this paper: (1) frequency-based *Transition Model*; (2) *Spatial Exhibit Viewing Model*; (3) generation of exhibit tours; and (4) generation of exhibit squares.

Frequency-based Transition Model. We use a frequency-based *Transition Model* to represent visitors’ movements between museum exhibits [Bohnert *et al.*, 2008; Schmidt *et al.*, 2009]. This model, which is implemented as a 1-stage Markov model, estimates the transition probabilities $P_{i,j}$ between exhibits i and j from frequency counts of exhibit transitions that are derived from observed visit trajectories. When estimating the transition probabilities, *additive smoothing* is applied in light of our small dataset of 44 observed trajectories (Section 6.1):

$$\hat{P}_{i,j} = \frac{n_{i,j} + \alpha_i}{N_i + M\alpha_i} \quad \text{for } i, j = 1, \dots, M$$

where $n_{i,j}$ counts the transitions from exhibit i to exhibit j , α_i is a smoothing constant, $N_i = \sum_{k=1, \dots, M} n_{i,k}$ is the total number of times exhibit i was viewed, and M is the number of exhibits.

Spatial Exhibit Viewing Model. Our modelling framework employs a probabilistic model of the viewing areas for each exhibit in the museum space, which divides the space into a grid of squares (for the Marine Life Exhibition, the grid size is $47 \times 61 = 2,867$ squares, where a square is approximately $30 \text{ cm} \times 30 \text{ cm}$; Figure 1). The model specifies a discrete probability distribution which represents $P(i|x, y)$, the probability of a visitor viewing each exhibit i from a square at position (x, y) .

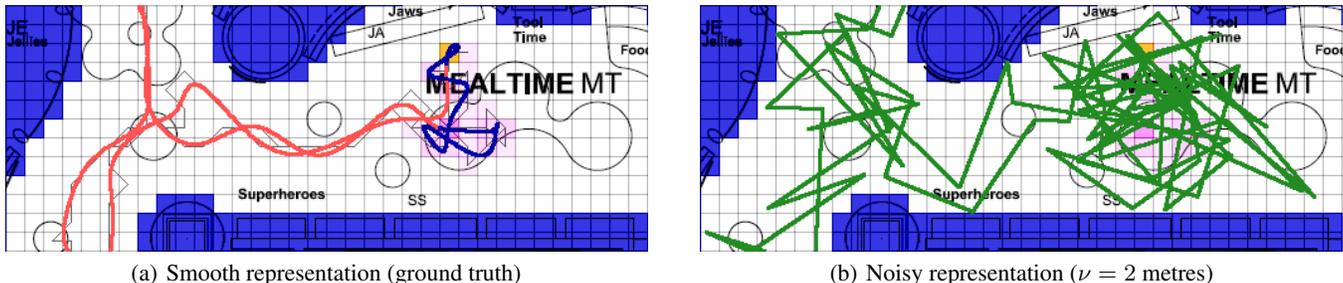


Figure 1: Two representations of part of a simulated visitor pathway

Generation of exhibit tours. We generate tours of viewed exhibits as follows. Each tour begins at a fictitious *start exhibit* i_0 and ends at a fictitious *end exhibit* i_{end} . For each exhibit i_{k-1} already in the tour ($k = 1, 2, \dots$), the next exhibit i_k is generated by sampling from a categorical distribution specified by the transition probabilities P_{i_{k-1}, i_k} . This step is repeated for each added exhibit i_k until the end exhibit i_{end} is reached.

In addition to this sequence of exhibits, our walking/hovering model (Section 4) requires the time that a visitor spends at each viewed exhibit. We generate a viewing time T_i at exhibit i by randomly drawing from an exponential distribution, i. e., $T_i \sim \text{Exp}(\lambda_i)$, where the average viewing time λ_i at each exhibit i is estimated by maximum likelihood from the 44 observed tours in the Marine Life Exhibition dataset.

Generation of exhibit squares. Once a tour of exhibits has been simulated, Schmidt *et al.* [2009] generate a *single* viewing square at position (x, y) for each viewed exhibit i in the tour. This is done by sampling from the categorical distribution $P(x, y | i)$ over all exhibit squares, where $P(x, y | i)$ is derived by applying Bayes’ theorem to the viewing probabilities $P(i | x, y)$ obtained from the Spatial Exhibit Viewing Model.

In this work, we use Schmidt *et al.*’s model to generate the *first* hovering square for each viewed exhibit (Section 4.2).

4 Simulation of Coordinate-based Visitor Pathways

The previous section outlined our method for generating exhibit tours with a single static grid square per exhibit. In this section, we simulate (smooth and noisy) coordinate-based visit trajectories which reflect two types of behaviour: *walking* between exhibits, and *hovering* at exhibits. Our approach comprises the following four steps, which are described below: (1) generation of connected paths of *walking* squares between exhibits; (2) generation of connected paths of *hovering* squares to simulate viewing behaviour at exhibits; (3) smoothing of the obtained square trajectory; and (4) simulation of noisy sensor observations from this smooth pathway representation.

Figure 1 depicts two representations of part of a simulated visit trajectory (we show the part for the Tool Time exhibit in the Mealtime section of the Marine Life Exhibition). Figure 1(a) shows the trajectory obtained after simulation (walk-

ing is represented by a red/grey line, hovering is represented by a blue/dark-grey line on pink/shaded squares, and wall squares are coloured in blue/grey), and Figure 1(b) is the representation obtained by applying Gaussian sensor noise at a level of $\nu = 2$ metres.

4.1 Generating Walking Squares

In Section 3, we generated one viewing square for each exhibit in a visitor’s tour. However, visitors do not simply teleport between squares. To produce a more realistic continuous visit trajectory, we must build a path that links these squares. At first glance, it seems that a shortest-path algorithm may be used for this task. However, trajectories generated in this way exhibit an unnatural level of repetition and purposefulness, tending to run directly along exhibition walls. In practice, visitors tend to move more erratically. To simulate these behaviours, we incorporate stochastic effects into the shortest-path procedure. Specifically, we model the probability of moving into a square as being proportional to the probability of viewing the destination exhibit from this square, moderated by the visitor’s propensity to avoid walls and to meander. Our approach uses parameters that control two behavioural aspects of visitors: (1) how erratic or purposeful their movement is; and (2) their propensity to avoid walls.¹ These considerations are implemented as follows.

Assume we want to generate a sequence of walking squares to connect two exhibits i and j in a tour. Let (x_s, y_s) denote the end square of exhibit i (i. e., the *source square*), and (x_d, y_d) the starting square of exhibit j (i. e., the *destination square*). Also, treating diagonal squares as adjacent, let the *candidate squares* of a square (x, y) be the eight squares surrounding this square. We start by employing Dijkstra’s algorithm [Dijkstra, 1959] to generate a distance matrix D whose elements $D_{x,y}$ correspond to the shortest-path distances from each square (x, y) to the destination square (x_d, y_d) . Then, we generate a sequence of walking squares as follows. For each square (x_n, y_n) (starting from the source square (x_s, y_s)), the next square (x_{n+1}, y_{n+1}) that a visitor moves into while walking is sampled from among

¹In our evaluation, we use fixed parameter values. Alternatively, one could sample the values for each trajectory simulation. Also, certain parameter values in combination with different transition models may yield different types of museum visitors, e. g., the *ant*, *fish*, *butterfly* and *grasshopper* types [Véron and Levasseur, 1983; Zancanaro *et al.*, 2007].

(x_n, y_n) 's eight candidate squares, provided that the move does not take the visitor farther away from (x_d, y_d) (the distance information is obtained from \mathbf{D}). In this procedure, the sampling is performed from a categorical distribution over the eight candidate squares, whose probabilities are proportional to the probabilities of viewing the destination exhibit from each square, moderated by the visitor's propensity to avoid walls and to meander (the probabilities are zero for the squares that take the visitor farther away from (x_d, y_d)). The visitor moves in this fashion until (x_d, y_d) is reached. At that point, the trajectory between (x_s, y_s) and (x_d, y_d) is complete, and timestamps are iteratively added to the trajectory assuming a constant walking speed v_w for the visitor.

4.2 Generating Hovering Squares

Once at an exhibit, visitors usually observe the exhibit for some time before moving on to the next one. Additionally, visitors typically do not remain static, but move around to examine the exhibit from different angles and distances. This so-called *hovering* behaviour is included in our simulation framework by varying the movement model described in Section 4.1, so that a visitor is more likely to move towards a square from which the exhibit is more likely to be viewed, but may not move at all.

Timestamps are added to the generated hovering squares assuming a hovering speed of $v_h < v_w$ (as for the walking case, we assume a constant hovering speed). The hovering behaviour continues until the sampled viewing time T_i for the current exhibit i is exceeded (viewing time sampling is described in Section 3).

4.3 Smoothing the Square Trajectory

To obtain a smooth positional tour representation from a time-stamped trajectory of squares, i.e., $(\langle t_n, x_n, y_n \rangle; n = 1, 2, \dots)$, we fit piecewise cubic splines to the coordinate-individual trajectories $\langle t_n, x_n \rangle$ and $\langle t_n, y_n \rangle$ (one piecewise cubic spline each). We do this by applying the `splinefit` package from the Matlab Central File Exchange [Lundgren, 2007]. This approach uses the method of least squares to fit splines with reduced degrees of freedom (we reduce the number of spline pieces by 70% compared to direct interpolation), and generates a smooth representation of the trajectory in the sense that (x, y) , (\dot{x}, \dot{y}) and (\ddot{x}, \ddot{y}) are all continuous in time.

The resultant representation may be interpreted as a continuous positional representation of the visit trajectory, enabling us to obtain a visitor's position at any point in time. Figure 1(a) depicts part of one such smooth visit trajectory.

4.4 Simulating Sensor Noise

The visit trajectories obtained so far are smooth and continuous. However, in practice, any trajectory-based input to a user modelling system would be acquired through sensors that deliver only a visitor's *approximate* position (due to measurement error) at a certain sampling rate.

In this paper, we explore sensor noise that may be attributed to *range-based positioning technology*, e.g., WiFi and ultra-wide band (UWB) [Zhao and Guibas, 2004]. We follow a widely accepted model for sensor noise in this setting, and assume that the *measured* coordinates (x', y') are

obtained by distorting the true coordinates (x, y) through additive Gaussian noise and sampling at regular time intervals (for our experiments, we use a constant sampling rate of one second). Specifically, the measured coordinates are found by sampling from a bivariate normal distribution $N((x, y), \sigma^2 \mathbf{I})$ with mean (x, y) and covariance $\sigma^2 \mathbf{I}$, where σ is a constant which reflects the expected accuracy of the sensing infrastructure, and \mathbf{I} is the identity matrix. For example, if the infrastructure is able to deliver positions within an accuracy level of ν metres 95% of the time, then $\sigma = \nu/2$ would be a suitable value, as this places approximately 95% of the probability mass within the circle defined by $(x' - x)^2 + (y' - y)^2 = \nu^2$. Figure 1(b) depicts part of a noisy visit trajectory which was sampled by following this procedure for the pathway shown in Figure 1(a) at a sampling rate of one second with $\nu = 2$ metres.

5 Inference and Prediction of Viewed Exhibits from Positional Coordinates

When information on a visitor's movements is automatically gathered through sensors, all that is available is a sequence of (typically noisy) time-stamped (x, y) coordinates (Section 4.4).² Assuming that we have a method for detecting whether a visitor is hovering (and hence viewing an exhibit), we can decompose the complete (x, y) sequence into sub-sequences of (x, y) coordinates that pertain to hovering behaviour (Section 5.1). From these, we can infer which exhibit the visitor is viewing (Section 5.2), and employ a model to predict which exhibit the visitor is likely to view next on the basis of this information (Section 5.3).

5.1 Classification-based Inference of Walking and Hovering

To infer walking and hovering behaviour from positional (x, y) coordinates, we employ a window-based approach. We first derive indicative features from a window comprising the previous ω sensor observations, and then provide these features to a purpose-trained classifier for inference. The output of this binary classifier is a label which indicates whether a visitor's activity is walking or hovering.

Prior to deriving the features, we smooth the noisy sensor observations $\langle t, x, y \rangle$ by fitting piecewise cubic splines to the $\langle t, x \rangle$ and $\langle t, y \rangle$ trajectories [Lundgren, 2007], and evaluating these splines at the original timestamps (similarly to Section 4.3). Using the resultant smoothed sensor observations, we compute the following feature set of size $2\omega + 7$ that pertains to (non-directional) velocity and acceleration:

- $\omega - 1$ velocities (each of them calculated as the length of one of the $\omega - 1$ velocity vectors, which in turn are derived from the ω smoothed positional coordinates from within the window)
- Minimum and maximum of the $\omega - 1$ velocities
- Mean and median of the $\omega - 1$ velocities
- Standard deviation of the $\omega - 1$ velocities

²For simplicity of notation, we use (x, y) instead of (x', y') in the remainder of the paper to denote the measured noisy coordinates.

- $\omega - 2$ accelerations (each of them calculated as the length of one of the $\omega - 2$ acceleration vectors, which in turn are derived from the $\omega - 1$ velocity vectors)
- Minimum and maximum of the $\omega - 2$ accelerations
- Mean and median of the $\omega - 2$ accelerations
- Standard deviation of the $\omega - 2$ accelerations

In our experiments (Section 6), we use support vector machines (SVM) to train the classifiers. We employ C-SVC SVMs with an RBF kernel from LIBSVM [Chang and Lin, 2001], using features derived from the previous five (x, y) observations ($\omega = 5$).

5.2 Probability-based Inference of Exhibits

In this section, we describe how we infer the exhibits most likely viewed by the visitor while hovering.

After inferring a visitor’s activity (i. e., walking or hovering) for each sensor observation $\langle t, x, y \rangle$, we extract from the complete (x, y) sequence the sub-sequences of (x, y) coordinates that correspond to hovering behaviour. For each sub-sequence of hovering-labelled (x, y) coordinates, we then calculate the following exhibit scores:

$$\text{score}(i) = \prod_{(x,y)} P(i|x,y) \quad \text{for all exhibits } i \quad (1)$$

where $P(i|x,y)$ is the probability of a visitor viewing exhibit i while hovering at position (x, y) (Section 3). To smooth out possible errors introduced in the classification step (Section 5.1), we delete walking labels that separate two consecutive sub-sequences of hovering labels for which the same exhibit has the highest score. We also remove hovering-labelled sub-sequences of length 1 (the exhibit scores of any affected sub-sequences of hovering labels are recomputed). Finally, all scores are normalised to obtain probabilities.

For each sub-sequence of hovering labels, this procedure yields a probability distribution which specifies how likely a visitor is to view each exhibit.

5.3 Model-based Prediction of Exhibits

Once the viewed exhibits are inferred, we can use this information to predict a visitor’s next exhibit for each (x, y) position at which the visitor is hovering.³ However, as seen in the previous section, there is some uncertainty regarding which exhibit the visitor is actually viewing. We therefore use the *Weighted* approach described by Schmidt *et al.* [2009] for predicting the next exhibit from positional information. For each possible next exhibit i , the *Weighted* approach estimates $P_{\text{next}}(i|x,y)$ as the weighted average of the transition probabilities $P_{j,i}$ from each possible current exhibit j to exhibit i . The weights are the probabilities $P(j|x,y)$ of viewing exhibit j when standing within the square at position (x, y) (Section 3).

$$\hat{P}_{\text{next}}(i|x,y) = \sum_{j=1}^M \{ P(j|x,y) \times P_{j,i} \}$$

³Predictions of a visitor’s next exhibits can be combined with predictions of the personally interesting exhibits to generate recommendations of exhibits that may be overlooked if the predicted next exhibits are actually visited.

In this calculation, the transition probabilities $P_{j,i}$ are derived from the information provided by the Transition Model in Section 3 by setting to zero the columns of the transition matrix that pertain to the already viewed exhibits, and normalising each row of the matrix to 1.⁴

6 Evaluation

This section presents our data collection method and datasets, and describes our experiments and results.

6.1 Data Collection and Datasets

Our dataset of real-world exhibit tours was obtained at the Marine Life Exhibition at Melbourne Museum. It consists of a (manually collected) record of the exhibits viewed by 44 visitors, and the viewing times at the exhibits. On average, each visitor viewed 7.2 of the $M = 22$ exhibits. The data for the viewing model described in Section 3 were obtained separately, by manually annotating a grid-based map to record the positions of visitors to the exhibition.

These data were used together with the method from Section 4 to generate 1000 simulated visits, where each visit comprises time-stamped sequences of (typically noisy) (x, y) coordinates at different noise levels — each element consisting of $\langle t, x, y \rangle$. These 1000 simulated visits are the basis for our evaluation. When generating the visits, we assumed a constant walking speed of $v_w = 3$ km/h and a hovering speed of $v_h = 1$ km/h. Also, we used a sampling rate of one observation per second.

Current range-based positioning systems are often based on processing radio signals, e. g., WiFi and ultra-wide band (UWB). WiFi-based technology typically achieves accuracy levels of 2 to 3.5 metres [Bahl and Padmanabhan, 2000; Lassabe *et al.*, 2009], while future UWB-based systems are expected to achieve accuracy levels of up to 0.15 metres [Hazas *et al.*, 2004]. We therefore considered accuracy levels of $\nu = 0$ to 4.5 metres when generating the visits.

6.2 Experiments and Results

To evaluate our models, we applied *bootstrapping* [Mooney and Duval, 1993] as follows. The 1000 generated visits were split into a training set of 100 visits and a test set of 900 visits. 200 *bootstrap* samples were then generated from the test set, with each bootstrap sample being constructed by sampling from the 900 visits with replacement (200 is the recommended upper bound on the number of samples for bootstrapping [Mooney and Duval, 1993]). The training set remained the same for all samples. Our results are averaged over the bootstrap samples.⁵

We conducted three experiments with these training and test sets: (1) walking/hovering classification; (2) inferring exhibits from positional hovering coordinates; and (3) predicting the next exhibit. All performance differences between models were found to be statistically significant with

⁴Our observations indicate that visitors rarely return to previously viewed exhibits. Hence, we focus on unseen exhibits.

⁵We employed bootstrapping, because only the test data varies for this technique, compared to cross validation which conflates the variation in the training and test data.

Table 1: Inference models and their experimental conditions

Models	Time & (x, y)	Walk/Hover	Exhibits	
			Previous	Current
TL_{all}	sequence of $\langle t, x, y \rangle$	Inferred	Inferred	Inferred
TLA_{all}	sequence of $\langle t, x, y \rangle$	Given	Inferred	Inferred
$Exh_{prev} TLA_{curr}$	sequence of $\langle t, x, y \rangle$	Given	Given	Inferred
<i>Schmidt et al.</i>	<i>one $\langle x, y \rangle$ per exhibit</i>	<i>N/A</i>	<i>Given</i>	<i>Inferred</i>
Exh_{all}	sequence of $\langle t, x, y \rangle$	Given	Given	Given

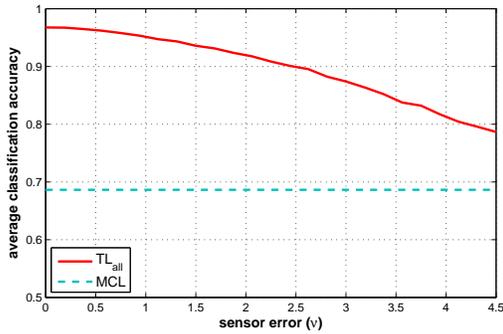


Figure 2: Average walking/hovering classification accuracy against sensor error

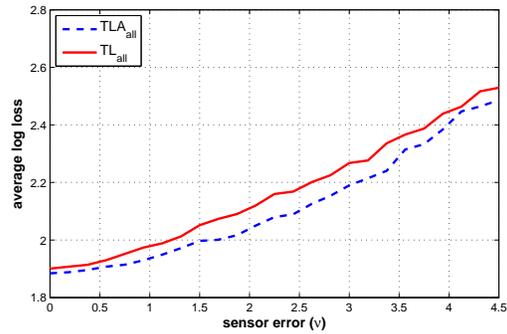


Figure 3: Average log loss of actually viewed exhibits against sensor error

$p \ll 0.001$ (evaluated using two-tailed paired t-tests on the bootstrap samples).

Table 1 summarises the models used in our experiments, indicating the inferred versus given information (only the first two models, i. e., those with grey background, are used in our first two experiments). The top model TL_{all} (*Time-Location for all observations*) is the most realistic, as its information is akin to that obtained from sensor readings (i. e., a sequence of time-stamped (x, y) coordinates). The models then become progressively less realistic, starting with TLA_{all} (*Time-Location-Action for all observations*), where the walking/hovering labels are considered given, up to Exh_{all} , where the walking/hovering labels, previous exhibits and current exhibit are given. To contextualise our work, Table 1 also shows Schmidt *et al.*'s model [Schmidt *et al.*, 2009] (typeset in italics), but its results are excluded from our evaluation, as it does not model trajectories or temporal information.

Walking/hovering classification. To evaluate the performance of our walking/hovering classification method (Section 5.1), we gave as input sequences of times and positions $(\langle t_n, x_n, y_n \rangle; n = 1, 2, \dots)$. For each walking/hovering classification, we considered the five positional observations made within the last four seconds ($\omega = 5$). As visitors hover slightly less than 69% of the time, and walk between exhibits for the rest of the time, we under-sampled the hovering portion of the training data to balance the classes.⁶

⁶We under-sampled the larger class, rather than over-sampling the smaller class, in order to retain the variance of the latter class. We also experimented with unbalanced data, but the performance

Figure 2 depicts classification accuracy as a function of sensor error, where the majority class baseline (MCL) assumes that a person is always hovering (the results are averaged over the 22 exhibits of the Marine Life Exhibition). Our results show that for no sensor error, our SVM classifier (TL_{all}) is able to infer whether a visitor is walking or hovering with approximately 97% accuracy. Classification accuracy decreases to about 88% as the sensor error increases to 2.75 metres (the middle of the range for WiFi technology).

Inferring exhibits from positional hovering coordinates.

To evaluate the performance of our mechanism for inferring the sequence of visited exhibits, we gave as input sequences of times and positions $(\langle t_n, x_n, y_n \rangle; n = 1, 2, \dots)$ and walking/hovering labels (one label for each element in a sequence). The probabilities of viewed exhibits were calculated once for given (known) walking/hovering labels, and once for labels inferred using the SVM classifier (Section 5.1). The inferences were made as described in Section 5.2, and resulted in a probability distribution of the exhibit being viewed by a visitor for each sub-sequence of hovering labels.

Figure 3 depicts the average *log loss* (negative log of the probability of the actually viewed exhibit), averaged over the 22 exhibits, as a function of sensor error. The figure compares the performance obtained when the walking/hovering labels are inferred (TL_{all}) with that obtained when the labels are given (TLA_{all}). It is worth noting that the comparison was done for the timestamps where the inferred and given hovering labels overlap, but the exhibit probabilities used for the was inferior to that obtained with the balanced data.

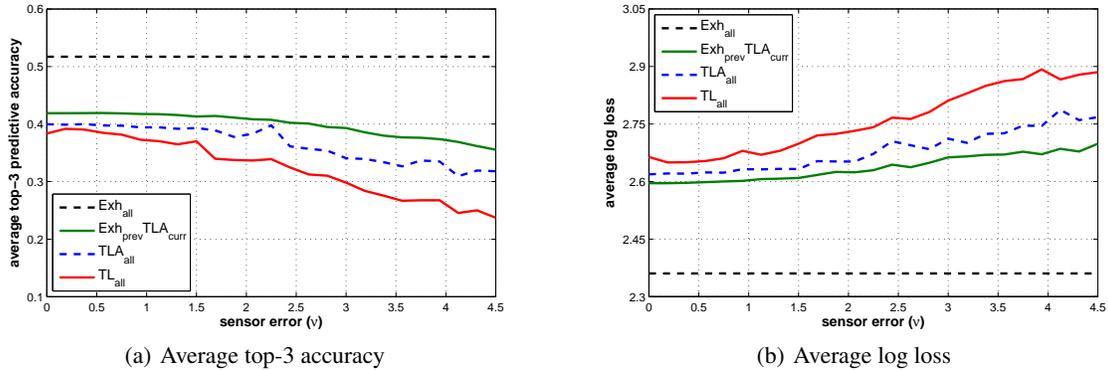


Figure 4: Predictive performance of the four models against sensor error

comparison were calculated for all the inferred or given hovering labels in each continuous sub-sequence of hovering labels. This explains the (expected) slight drop in performance for inferred hovering labels, since, as seen in the first experiment, the inferred labels are sometimes wrong. Also, as expected, performance deteriorates as sensor error increases.

Predicting the next exhibit. This experiment determines the effect of different assumptions regarding available information on predictive accuracy. We consider our four models from Table 1, whose information ranges from time-stamped positional sensor logs (TL_{all}) to sequences of viewed exhibits (Exh_{all}). In line with Schmidt *et al.* [2009], for all four models, the next exhibit was predicted using the transition matrix learned from the 44 tours observed at the Marine Life Exhibition (Section 3). For Exh_{all}, we used the transition matrix directly (the transition probabilities for previously visited exhibits were set to zero), while for the other models, we used the *Weighted* approach described in Section 5.3.

Figures 4(a) and 4(b) show, respectively, the average top-3 accuracy and average log loss for various levels of sensor error for the four models described in Table 1 (the results are averaged over the 22 exhibits). For this experiment, *log loss* is defined as the negative log of the probability with which the exhibit actually viewed *next* is predicted, and *top-3 accuracy* measures how often the exhibit actually viewed next is one of the three exhibits predicted with the highest probability. We employ top-3 rather than top-1 accuracy because the top probabilities are often quite similar due to the physical layout of the exhibition. As seen in the figures, the higher the uncertainty about a visitor’s behaviour and the higher the sensor error, the lower the accuracy and the higher the log loss (statistically significant). Note that Exh_{all} is invariant to sensor noise, as all the information is assumed given (Table 1). Interestingly, the differences in performance between the three lower-information models (TL_{all}, TLA_{all} and Exh_{prev}TLA_{curr}) are relatively small, and their performance profiles are quite flat up to $\nu = 1.5$ metres, diverging slightly from there on. The creditable performance up to $\nu = 1.5$ metres means that one can expect acceptable predictive performance from sensor-based systems.

7 Conclusions

This paper offered a realistic model of sensor-based information, significantly extending the work of Schmidt *et al.* [2009]. Our framework enables us to study the impact of different assumptions regarding sensor noise and available sensor information on inferential performance regarding viewed exhibits. The accuracy of these inferences in turn affects the performance of user models, viz models of visitors’ interests and of exhibits they are likely to visit. As expected, predictive performance deteriorates for every experimental parameter that is inferred (rather than given), and also as sensor error increases. However, interestingly, performance remains quite stable for sensor noise up to 1.5 metres, which is an encouraging result for real-world systems.

Our inferential and predictive models in combination support the generation of recommendations of exhibits that may be of interest but are likely to be missed. Our models may also be used to influence the strength of recommendations as a function of the reliability of the information on which the recommendations are based. An additional application of our results is in guiding the layout of sensing devices in a museum, e. g., it may be advantageous to place more devices in locations where the inferences are more uncertain.

Acknowledgements

This research was supported in part by grant DP0770931 from the Australian Research Council. The authors thank Daniel F. Schmidt for his involvement at early stages of this research, Liz Sonenberg and Carolyn Meehan for fruitful discussions and their support, and David Abramson, Jeff Tan and Blair Bethwaite for their assistance with the computer cluster.

References

- [Bahl and Padmanabhan, 2000] Paramvir Bahl and Venkata N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of the 19th Annual Joint IEEE Conference on Computer Communications (INFOCOM-00)*, pages 775–784, 2000.
- [Bohnert and Zukerman, 2009] Fabian Bohnert and Ingrid Zukerman. Non-intrusive personalisation of the museum

- experience. In *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization (UMAP-09)*, pages 197–209, 2009.
- [Bohnert *et al.*, 2008] Fabian Bohnert, Ingrid Zukerman, Shlomo Berkovsky, Timothy Baldwin, and Liz Sonenberg. Using interest and transition models to predict visitor locations in museums. *AI Communications*, 21(2-3):195–202, 2008.
- [Chang and Lin, 2001] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [Cheverst *et al.*, 2002] 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.
- [Dijkstra, 1959] Edsger W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [Hatala and Wakkary, 2005] 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.
- [Hazas *et al.*, 2004] Mike Hazas, James Scott, and John Krumm. Location-aware computing comes of age. *IEEE Computer*, 37(2):95–97, 2004.
- [Lassabe *et al.*, 2009] Frederic Lassabe, Philippe Canalda, Pascal Chatonnay, and François Spies. Indoor Wi-Fi positioning: Techniques and systems. *Annals of Telecommunications*, 64:651–664, 2009.
- [Lundgren, 2007] Jonas Lundgren. SPLINEFIT, 2007. Software available at <http://www.mathworks.com/matlabcentral/fileexchange/13812-fit-a-spline-to-noisy-data>.
- [Mooney and Duval, 1993] Christopher Z. Mooney and Robert D. Duval. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Sage Publications, Newbury Park, CA, USA, 1993.
- [Petrelli and Not, 2005] 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.
- [Philipose *et al.*, 2004] Matthai Philipose, Kenneth P. Fishkin, Mike Perkowitz, Donald J. Patterson, Dieter Fox, Henry Kautz, and Dirk Hahnel. Inferring activities from interactions with objects. *IEEE Pervasive Computing*, 3(4):50–57, 2004.
- [Schmidt *et al.*, 2009] Daniel F. Schmidt, Ingrid Zukerman, and David W. Albrecht. Assessing the impact of measurement uncertainty on user models in spatial domains. In *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization (UMAP-09)*, pages 210–222, 2009.
- [Stock *et al.*, 2007] Oliviero Stock, Massimo Zancanaro, Paolo Busetta, Charles Callaway, Antonio Krüger, Michael Kruppa, Tsvika 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.
- [Véron and Levasseur, 1983] Eliséo Véron and Martine Levasseur. *Ethnographie de l'Exposition*. Bibliothèque Publique d'Information, Centre Georges Pompidou, Paris, France, 1983.
- [Wang *et al.*, 2009] Yiwen Wang, Lora Aroyo, Natalia Stash, Rody Sambeek, Yuri Schuurmans, Guus Schreiber, and Peter Gorgels. Cultivating personalized museum tours online and on-site. *Interdisciplinary Science Reviews*, 34(2):141–156, 2009.
- [Zancanaro *et al.*, 2007] Massimo Zancanaro, Tsvika Kuflik, Zvi Boger, Dina Goren-Bar, and Dan Goldwasser. Analyzing museum visitors' behavior patterns. In *Proceedings of the 11th International Conference on User Modeling (UM-07)*, pages 238–246, 2007.
- [Zhao and Guibas, 2004] Feng Zhao and Leonidas Guibas. *Wireless Sensor Networks: An Information Processing Approach*. Morgan Kaufmann, 2004.