CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Learning occupancy in single person offices with mixtures of multi-lag Markov chains
Authors	Manna, Carlo;Fay, Damien;Brown, Kenneth N.;Wilson, Nic
Publication date	2013-11
Original Citation	MANNA, C., FAY, D., BROWN, K. N. & WILSON, N. 2013. Learning occupancy in single person offices with mixtures of multi-lag Markov chains. In: Proceedings 25th International Conference on Tools with Artificial Intelligence ICTAI 2013. Washington DC, USA, 4-6 Nov. Los Alamitos, California: IEEE Computer Society, pp. 151-158.
Type of publication	Conference item
Link to publisher's version	10.1109/ICTAI.2013.32
Rights	© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Download date	2026-10-30 13:11:26
Item downloaded from	https://hdl.handle.net/10468/1393

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Learning Occupancy in Single Person Offices with Mixtures of Multi-lag Markov Chains

Carlo Manna, Damien Fay, Kenneth N. Brown, Nic Wilson
Cork Constraint Computation Centre
Dept. of Computer Science
University College Cork
Cork, Ireland

***Abstract*—The problem of real-time occupancy forecasting for single person offices is critical for energy efficient buildings which use predictive control techniques. Due to the highly uncertain nature of occupancy dynamics, the modeling and prediction of occupancy is a challenging problem. This paper proposes an algorithm for learning and predicting single occupant presence in office buildings, by considering the occupant behaviour as an ensemble of multiple Markov models at different time lags. This model has been tested using real occupancy data collected from PIR sensors installed in three different buildings and compared with state of the art methods, reducing the error rate by on average 5% over the best comparator method.**

***Keywords*-Occupancy prediction; building control; Markov chains**

I. Introduction

The operation of commercial and residential buildings consumes 40% of total US primary energy use, and of that, Heating, Ventilation and Air Conditioning (HVAC) systems consume at least 46% [1]. Poor design and inefficient operation of HVAC systems causes significant energy waste [2]. Although it is possible to improve HVAC system design for energy efficiency, it requires substantial investment to retrofit an existing building with improved HVAC equipment. In contrast, improving the HVAC control algorithms is far more cost effective, as long as expensive additional sensors are not required. A number of recent papers have focused on improved HVAC control algorithms, based on Model Predictive Control (MPC) [3], [4], [5], [6], [7], [8], [9] to find the optimal period of pre-heating or pre-cooling for the room. If advanced control methods are to be effective, they must use reliable predictions of occupancy, otherwise energy will be wasted of occupants dissatisfied. Predicting occupancy, however, is complex due to its temporal and spatial stochastic nature. General methods such as recurrent neural networks or logistic regression do not take into account domain specific knowledge and are generally oriented to predict occupancy just one step ahead. This is problematic for HVAC control, as MPC requires an n-step ahead prediction (typically ranging from several minutes to a few hours).

Methods based on inhomogeneous Markov chains, assume that the duration an occupant spends in a given location depends both on the time spent in the previous state (i.e. location) and on the time of day. This class of methods are generally capable of predicting occupancy with reasonable accuracy, although they have some downsides. The time that an occupant spends in a given location might depend on the time spent in a number of previous states (i.e. locations) and not only in the last state, as is the case for a first order Markov chain. This problem could be overcome by considering an n-order Markov chain. This however adds complexity in terms of computation time, and could also make the model too constrained to be effective by over-specifying the detail of location sequences.

In this paper, we propose a model that takes into account multiple previous occupant states but is much more flexible than methods based on single Markov chains. In particular, we combine several single-state inhomogeneous Markov chains independently, each with a different time delay. This allows us to take into account the occupants' differing previous states. We then use the combined Markov model to predict n steps ahead. We propose a flexible evaluation metric which accumulates future steps into blocks, where the block size varies according to the needs of the control system that acts on the prediction. The approach is validated using real occupancy data collected over a 6 month period from a number of PIR sensors deployed in single and multiple occupant offices, and compared with other approaches from the literature. For single step and n-step ahead experiments, we show that the combined model reduces the error rate of the best comparator method by an average of 5%. For the flexible metric, we show that the combined model reduces the error rate again by 5% over the closest comparator.

II. Related work

Learning occupancy profiles from sensor readings is becoming common—see, for example, Georgia Tech Aware Home [10], Intelligent Room [11], Adaptive Home [12] and MavHome [13]. [12] developed a prototype system that controls basic residential comfort systems such as air heating, lighting, ventilation and water heating. This system studied

the actions taken by occupants, such as adjusting thermostats or choosing light configurations, and attempted to infer patterns to predict these actions. The information received from the sensors was used to learn occupancy patterns, but no domain knowledge was considered, nor statistical evaluation to make optimal predictions. [14] used Bayesian methods to simulate occupant presence using sensory data (infrared, motion, or carbon dioxide). [15] describes an algorithm for simulation of occupant presence, to be used as an input for future occupant behavior models within building simulation tools. It considered occupant presence as an inhomogeneous Markov chain interrupted with periods of absence, thereby coming up with a time series of the state of presence for every occupant for every respective zone/building. The approach requires a large amount of historic data for each individual occupant to find the optimal model. In practical situations, the available data may be limited, at least to begin with, and the occupant's mode of behaviour may change, thus requiring an effective restart with a small data set, and so methods that are able to predict occupancy with limited data are required. Further, the model assumed occupancy is based only on the previous state, and ignores temporary absences. In [14], the authors developed a network of passive infrared occupancy sensors and analyzed sensory data using tools based on Bayesian probability theory (specifically, belief networks) to determine occupancy. The most important contribution of that paper was the building of a probability distribution over the number of occupants and their locations in a building, given past and present sensor measurements and any other contextual information that existed. Such an approach can be computationally intensive; also this was strictly an occupancy detection model and did not take into account the prediction of occupancy.

III. Occupancy Models and Algorithms

In this section we give a brief overview of some previous methods used to predict occupant presence and for time sequence analysis and forecasting in general. These methods are also used as comparators for our proposed approach described in the next section.

A. First-order Markov Chain

The first-order Markov chain technique is a well-established stochastic method that has been extensively studied and applied in a number of different applications [16]. Each state of the system is dependent only on the previous state together with the probabilities in a state transition matrix. The transitions can be time-dependent, so it is particularly suited to modeling a number of real-world situations and has been used for occupancy simulation and prediction [15].

B. Input-Output Hidden Markov Chain

The IOHMM is based on filtering the raw PIR measurements and associating with them three states. This is shown in Figure 1 (a); S_1: *Out*, S_2: *In and Moving*, and S_3: *In and Still*. When in one of these states it causes an *emission* which is the PIR reading; i.e. the observed data. The whole model is placed into bins 5 minutes wide giving 288 bins per day and the emission from the PIR in each bin is the number of minutes during which motion was observed. The time in state is determined by a transition matrix, but key to the IOHMM is that the transition matrix is dynamic and a function of an external variable, in this case the time of day. Specifically, we define the observation to be $x_t \in [0, 5]$ (the sum in minutes of the observed motion in a 5 minute interval) where t is the time of day. In addition, define the time of day to be an integer, $u_t \in [0, 1, \ldots, 288]$ and the transition between states may then be expressed as:

$$A(i, j, t) = P(q_t = S_j | q_{t-1} = S_i, u_t = u, B) \qquad (1)$$

where $A(i, j, t)$ is the probability of moving from state i to state j *at time* t, q_t is the state the system is in at time t and B is a matrix of emission probabilities with:

$$B(i, k) = P(x_t = k | q_t = S_i) \qquad (2)$$

where $B(i, k)$ is the probability of a PIR reading of k given that the system is in state i. The set $\{B(i, k), A(i, j, t)\}$ must be estimated from the data and this is an involved task [17] due to the time dependence of the process; the vast majority of the literature is devoted to homogeneous Markov Chains, the Markov chain here is inhomogeneous, and the asymptotic properties of inhomogeneous chains are not well understood [18].[1] In summary, an Expectation Maximisation (EM) algorithm is applied. The forwards-backwards algorithm is used to give an estimate of the state transitions in the training set (given the time dependent transition matrices); i.e. this is the E-step [19]. Given the state transitions, an estimate of the emission probabilities and the time dependent transition matrices is created using counts of the events (the M-step). The iteration is found to converge quite rapidly (we found 10 iterations were sufficient).

Figure 1 (b) shows two quantities derived from $A(i, j, t)$ for an occupant in the computer science department at UCC.[2] The two quantities are $a_{out-->in} = A(0, 1, t) + A(0, 1, t)$ and $a_{in-->out} = (A(1, 0, t) + A(2, 0, t))/2$, the probability of going from *Out* to *In* (moving or still) and the probability of going from *In* to *Out*. As can be seen, $a_{in-->out}$ is approximately 1 from midnight to about 7am indicating that the system will always transition to the out state at these times. Between 7 and 9am $a_{out-->in}$ rises and peaks around 9am indicating that this occupant arrives between 8am and 9am; note that for the other occupants we found a similar pattern but the times can be moved

[1] The inference procedure is beyond the scope of this paper and we refer the interested reader to [19].

[2] For privacy reasons results from a test occupant are presented (one of the authors in fact).

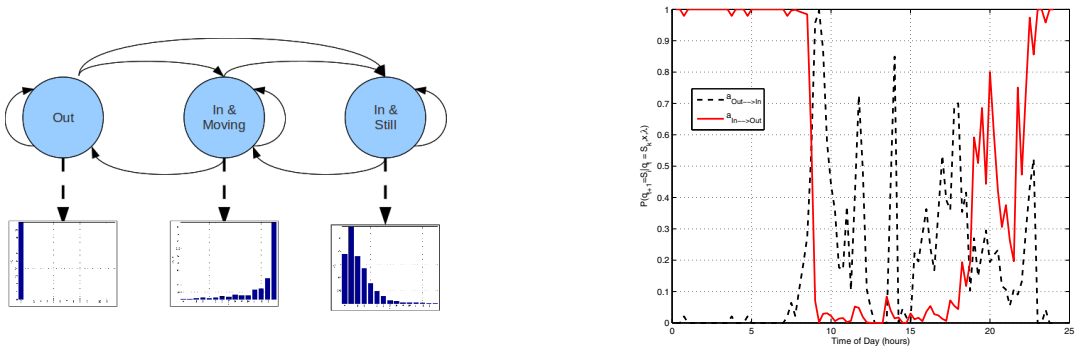


Figure 1. An Input Output Hidden Markov Model for Occupancy (a) overview (b) $A(i, j, t)$ for an example user.

onwards and the rise in $a_{out-->in}$ can be more and less pronounced. The occupant generally has a break around 2:30pm and leaves the office at 6pm; there are peaks in both $a_{out-->in}$ and $a_{in-->out}$ later in the evening with $a_{in-->out} > a_{out-->in}$; this produces short in periods late in the evening which can be attributed to the cleaners.

The IOHMM is a generative model, i.e. the $P(q_t = S_i|u_t, q_{t-1} = S_j)$ is estimated as part of the modelling process. This allows us to sample from the process and thereby generate ensembles of different occupancy patterns. Figure 2 (a) shows the average of 100 samples for the test occupant and the actual average occupancy. As can be seen, there is wide agreement, but the IOHMM has difficulty zoning in on lunchtime and with the cleaning staff late at night. Finally, Figure 2 (b) shows an actual occupancy pattern and the corresponding (ten) ensembles. The average of these ensembles corresponds well with the actual.

C. NARX Neural Network

Artificial neural networks (ANNs) have been successfully applied to a number of time series prediction and modeling tasks, including financial time series prediction, biomedical time series modeling, communication network traffic prediction and several others [16]. Occupancy sequences are represented by a binary time series where only discrete values of a sequence are available. To build the prediction model, recurrent neural networks are more appropriate, since they solve the difficulties of the temporal relationships of inputs between observations at different time steps, by maintaining internal states that have memory. Nonlinear Autoregressive with eXogenous input (NARX) [20] are a well-established class of recurrent neural network that have been successfully applied to modeling time series.

In particular, NARX are a class of discrete-time nonlinear systems that can be mathematically represented as:

$$x(t+1) = f[x(t), ..., x(t-d_x+1); u(t), ..., u(t-d_u+1)] \quad (3)$$

where $x(t)$ and $u(t)$ denote, respectively, the variable of interest and the externally determined input variable of the model at discrete time step t, while $d_x \leq 1$ and $d_u \leq 1$, with $d_u \leq d_x$, are respectively, the input-memory and output-memory orders.

Because of the ability of such a model to take into account both the previous states of an occupant (as variable of interest) and the time of day (as external input), it was selected for the comparison test during our experiments.

IV. The Proposed Method

In this section we describe the proposed Mixtures of Multi-lag Markov chains (MMLM) for the occupancy prediction problem. We want to predict the dichotomic variable $x(t)$ (x has only two values i.e: in or $away$), at time t, for each $t = 1, \ldots, T$. Here, T is the number of the time slots we have in each day. For example, if we have a sample time of 5 minutes, and we aim to forecast the occupancy between 8am to 8pm, we have 12 time slots for each hour, for a total of 12 x 12 = 144 time slots per day.

First of all, at each time t, we are interested in predicting a sequence $X(t)$ of values for the variable x such that $X(t) = \{x(t), x(t+1), \ldots, x(t+M)\}$. This can be reduced to predicting the value for the variable x, $(t+m)$-steps ahead with $m = 0, \ldots, M$ individually, then composing the individual predictions to get the sequence $X(t)$.

Further, as stated previously, we expect the location of an occupant to depend on the length of time he or she has spent in their current location. To take this into account, we should consider n previous states of the occupant. Rather than a vector of the previous states in a single model, which may overspecify the history, we will construct n different single-state Markov models which differ in the time lag between the history state and the current time. Let $\hat{\psi}_n(x_{t+m}|x_{t-n})$, for $n = 1, \ldots, N$, be the single model that for each t is able to predict the value for the variable x, $(t+m)$-steps ahead, using only the previous state at time $t-n$. The transition probability matrix for each model can be evaluated statistically using an initial training set. Moreover, in each model the transition probability matrix will depend on t, as

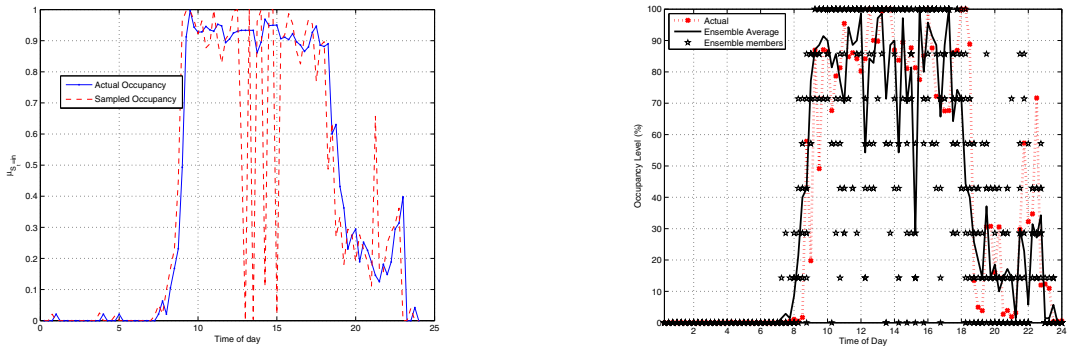


Figure 2. IOHMM results: (a) The average occupancy versus observed for a user. (b) A ten ensemble prediction.

for any occupant, the transition from one state to an another could be influenced by the time of the day.

Finally, in order to forecast the value for the variable x, $(t+m)$-steps ahead, we combine the single models together. In particular, for each t and for each m, let $\hat{\psi}(x_{t+m}|x_{t-1}, x_{t-2}, \ldots, x_{t-N})$ be the predicted value for the variable x at $(t+m)$-steps ahead, based on the knowledge of N previous states assumed by x, such that:

$$\hat{\psi}(x_{t+m}|x_{t-1}, x_{t-2}, ...x_{t-N}) = \sum_{n=1}^{N} \pi_n(t)\hat{\psi}_n(x_{t+m}|x_{t-n}) \tag{4}$$

in which each single model $\hat{\psi}_n(x_{t+m}|x_{t-n})$ incorporates only one previous state at time $t-n$, for $n = 1, \ldots, N$. Notice that for $n = 1$ such a model is a first-order Markov chain, but with $n > 1$ there is no conceptual relation between these models and any Markov chain of n-order. A sketch of the model is shown in Figure 3.

The individual coefficients $\pi_n(t)$ are functions of the time instant t. The idea behind this choice is that different model components can predict with different performance in different regions of the input space, and the coefficients determine which components are dominant in which region. This can be seen in Figure 4, which uses $n = 3$ and which shows the accuracy of each model for different time slots of the day. While the three models are more or less equally accurate at the start and end of the day, in the middle of the day, from approximately time slots 20 to 130, the three models have significantly different accuracy depending on the time slot.

The coefficients $\pi_n(t)$ must satisfy the usual constraints $0 \leq \pi_n(t) \leq 1$ and $\sum_{n=1}^{N} \pi_n(t) = 1$. The coefficients are trained and updated during each iteration in order to assign more weight to the more accurate models and penalizing the less accurate. The procedure to update the coefficients $\pi_n(t)$ is as follows:

$$\pi_n(t) = \frac{\rho_n(t)}{\sum_{n=1}^{N} \rho_n(t)} \tag{5}$$

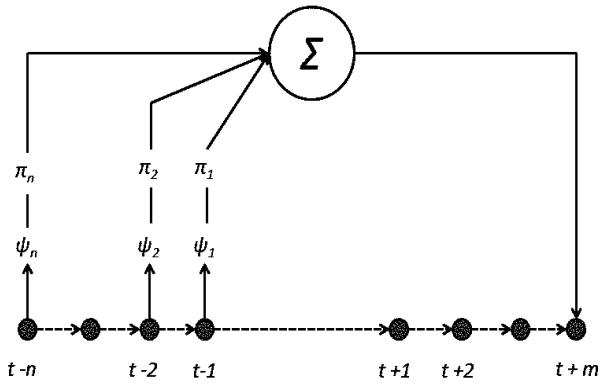


Figure 3. The proposed MMLM: each single model ψ_n uses only one previous state to predict the occupant state at $t+m$. Then, all models are combined together.

where $\rho_n(t)$ are called the raw coefficients. At the start of the procedure the raw coefficients $\rho_n(t)$ are all equal. Then, in each subsequent iteration they are updated so that $\rho_n(t+T) = \rho_n(t)+1$ if the corresponding model $\psi_n(t|t-n)$ at time t classifies correctly, otherwise $\rho_n(t+T) = \rho_n(t)$. Although there are many other strategies that could be used to train the coefficients $\pi_n(t)$, this simple one is very effective mainly because it avoids weighting too heavily models that are slightly more accurate than the others (such as the strategies that increase/decrease the weight coefficients exponentially).

In the current setting, the proposed MMLM has many advantages compared with a first order Markov chain or an nth-order Markov chain. As the 1st order Markov chain only considers one previous state, the time spent in a state able to be considered as input is simply the sample time.

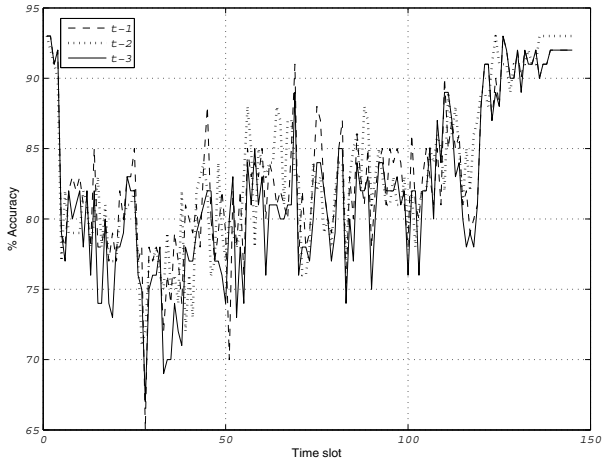


Figure 4. Accuracy vs. time for three models with different time lag ($t-1$, $t-2$, $t-3$). Notice how the accuracy changes with time for each models. We have different dominant models for different regions of the input space.

With MMLM, however, it is possible to take into account all locations of an occupant in a given interval, without any sample time limitation, taking into account multiple previous states without having the complexity of a Markov chain of order n.

V. Experiments

In this section we describe the experimental testing to validate our approach. First we describe the real-world data sets used and then we present prediction results on this set.

A. Data set

Occupancy data was collected from 10 different users in their offices by means of a PIR sensor on the desk of each occupant. The data collection period covers approximately 6 months, from February up to August 2012. The acquired data needed to be pre-processed before the experiments could begin. Problems either with the sensors, or with the communication between these sensors and the server used to store the data, caused some loss of data, reducing the amount of usable data from 6 months to approximately 3-4 months (depending on the user). PIR sensors only acquire changes of the state of the monitored variable; this means that the time and date are recorded when: *(i)* the first time the sensor detects motion in the office when it previously considered it to be empty (from the unoccupied state to the occupied state), and *(ii)* when the sensor has not detected any movement when it previously considered it occupied (switch from the occupied state to the unoccupied state). Furthermore, as the sensor can only detect two states of presence, occupied or vacant, it cannot distinguish whether the zone is multiply occupied or not. The presence of guests in an office space can depend both on whether the owner of that office is present (in which case there is no reason to distinguish between the owner or the guest), or the guest presence is temporary (so only continuous presence of at least 1 minute has been accounted for). Hence, the first step in processing the raw data was to check which days of acquisition had suffered from the aforementioned technical problems, and only data that are completely intact have been selected. This data was then cleaned of all periods of absence lasting less than one minute (this usually corresponds to a sensor that stops recording the presence of an occupant because she/he is too still for her/his movement to be detected). We then constructed a time series of data with a 1 min time step for each user. Finally, in order to use the data set for the validation test, we downsampled the time series to a sample time of 1 sample each 5 minutes, considering such sample time a reasonable choice to study the occupancy behaviour. The data was then split into two subsets: the first was used for the extraction of information and the training of the occupancy models, while the second one was been used for the test validations.

B. Results

As noted above, the experiments were carried out on data sets from ten different users. The participants are expected to have different behaviour profiles because of their different jobs (lecturers, researchers and PhD students). In each of the following tests, the Mixture of Multi-Lag Markov chain approach (MMLM) is compared with the methods previously described (i.e. inhomogeneous Markov chain (MC), Input-Output Hidden Markov chain (IOHMM) and the Narx Neural Network (NARX)).

Throughout the experiments in this section, the sample time used is 1 sample each 5 minutes and the day time interval was between 8am and 8pm. Hence we have 144 data points to be predicted for each day. The total number of data points can depend on the user, ranging from approximately 3,000 predicted data points (around 20 days) up to more than 12,000 predicted data points (about 90 days). The same training and test sets were used for each algorithm, ranging from 1 week (for the users with fewer data points) up to 20 days (for the user with more data points). Finally, throughout the test, we set $n = 6$ for MMLM and $d_u = d_x = 6$ for the NARX model.

First, in Table I, we report the results for the 1-step ahead prediction, showing the error rate—that is, the percentage of predictions that did not match the actual occupancy as recorded in the data obtained from the PIR sensors. If *tp* is the number of true positives, *tn* the number of true negatives, *fp* the number of false positives and *fn* the number of false negatives, then the error rate, or inaccuracy, is defined as follows:

$$Inaccuracy = 1 - \frac{tp + tn}{tp + tn + fp + fn} \tag{6}$$

From the table, we can see that both the MMLM and the MC approaches consistently outperform the other two meth-

Table I
1-STEP AHEAD PERCENTAGE OF INACCURATE FORECASTS FOR: THE PROPOSED MIXTURE OF MULTI-LAG MARKOV CHAINS (**MMLM**), INPUT OUTPUT HIDDEN MARKOV MODEL (**IOHMM**), INHOMOGENEOUS MARKOV CHAIN (FIRST ORDER) (**MC**), AND NARX NEURAL NETWORK (**NARX**).

	MMLM	**IOHMM**	**MC**	**NARX**
user 1	6.50	14.09	7.02	17.38
user 2	3.28	31.48	4.34	4.19
user 3	10.43	10.49	11.15	23.65
user 4	1.49	25.91	1.5	1.42
user 5	8.61	33.14	9.89	12.01
user 6	6.16	23.67	5.58	17.24
user 7	8.15	16.28	9.28	26.08
user 8	5.71	15.17	5.27	16.65
user 9	5.52	15.75	6.72	18.05
user 10	8.62	21.83	9.9	18.07

ods, with significantly lower error rates. The MC method, a simple 1^{st} order model, is surprisingly effective, only once reporting an error rate over 10%. The proposed MMLM method, however, consistently improves over MC, with an average drop in the error rate of 0.6 percentage points, or a reduction of 7% in the number of errors.

Table II
1-HOUR AHEAD PERCENTAGE OF INACCURATE FORECASTS (12-STEPS AHEAD) FOR: THE PROPOSED MIXTURE OF MULTI-LAG MARKOV CHAINS (**MMLM**), INPUT OUTPUT HIDDEN MARKOV MODEL (**IOHMM**), INHOMOGENEOUS MARKOV CHAIN (FIRST ORDER) (**MC**), AND NARX NEURAL NETWORK (**NARX**).

	MMLM	**IOHMM**	**MC**	**NARX**
user 1	13.96	17.42	14.53	18.17
user 2	6.45	40.56	6.55	9.17
user 3	18.81	17.13	19.27	24.57
user 4	1.54	18.62	1.58	1.34
user 5	11.53	42.67	12.45	11.97
user 6	14.04	34.39	14.51	17.11
user 7	20.73	29.19	21.79	25.52
user 8	12.57	24.38	15.89	19.58
user 9	13.36	23.25	14.29	21.22
user 10	15.54	25.08	16.03	18.03

However, as stated in the previous sections, forecasting 1-step ahead (i.e. in this case the next 5 minutes of occupancy) is of limited use for HVAC control, where longer term predictions are required. Table II shows the results of predicting occupancy for 1 hour ahead (or 12-steps), where the prediction is updated every 5 minutes. That is, as each new point is revealed from the data set during the experiment, a new 1 hour prediction is performed. The error rates are then computed and averaged over each predicted step ahead. We can see that the overall accuracy drops compared to 1-step ahead prediction, as expected. However, the pattern of results remains the same, and we achieve acceptable accuracy from both the MC and MMLM approaches. Again, MMLM consistently improves on MC, with an average reduction in error rate of 0.8 percentage points, and a reduction of 5% in the number of errors. The corresponding figures for the improvement of MMLM over NARX and IOHMM are 3.8 percentage points and 18%, and 14 percentage points and 47%, respectively.

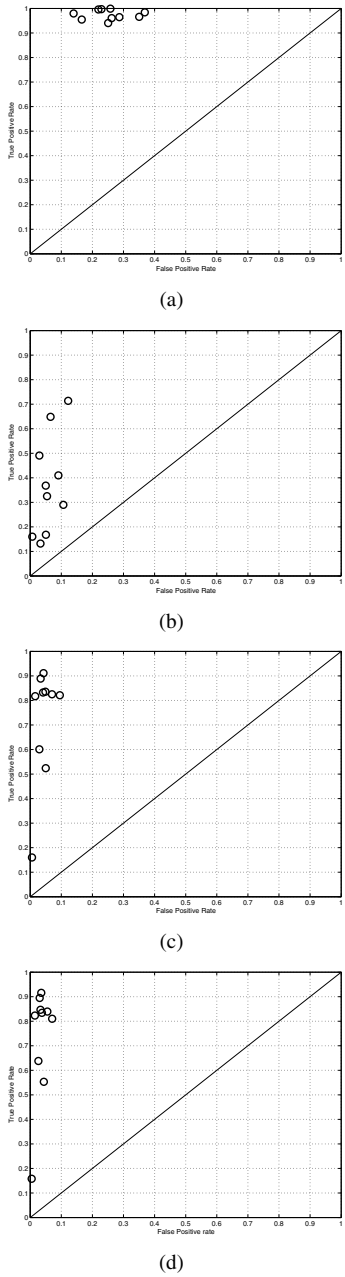


Figure 5. ROC space for the 10 individual users, reported as false positive rate vs. true positive rate for: (a) IOHMM, (b) NARX, (c) MC and (d) MMLM. The ideal point is $(0, 1)$ which corresponds to the value 0 for the false positive rate and 1 for the true positive rate.

We also report in Figure 5 the ROC space for the four methods for the 12-step ahead prediction, in which we can see see that the proposed MMLM method has a very low false positive rate. That is particularly important as a low

false positive rate means the algorithm rarely forecasts the presence of an occupant when she/he is actually away, which allows the control algorithm to avoid wasting energy. The high true positive rate improves occupant comfort.

Although the 1-hour ahead prediction is more appropriate as input to HVAC control systems, the error rate may still be misleading. For example, issuing a false positive for 40 minutes ahead and a false negative for 45 minutes ahead contributes two errors, but from the point of view of the HVAC system, it would receive a correct prediction of 1 occupancy for the 10 minute period starting 40 minutes ahead. More precisely, let 0 denote the event when an occupant is *away* and 1 the case when she/he is *in office*. Consider the three cases where first we have a predicted 1-hour sequence $p_1 = \{111111111111\}$, while the actual sequence is $a_1 = \{000000000000\}$, secondly we have a predicted sequence $p_2 = \{000000111111\}$, while the actual sequence is $a_2 = \{111111000000\}$ and, finally we have a predicted 1-hour sequence $p_3 = \{101010101010\}$, while the actual sequence is $a_3 = \{010101010101\}$. In all cases the inaccuracy is 100%, but there is a significant difference between them from the perspective of HVAC control. Implementing a predictive control strategy for heating in the first case means total energy dissipation. In the second case, in a naive strategy the heater would be turned off for 30 minutes then turned on for the next 30 minutes, resulting in wasted energy for the second 30 minutes, and occupant dissatisfaction for the first 30 minutes. However, in the final case, the predicted sequence is simply shifted one step forward in time, and the effect on energy consumption and occupant dissatisfaction would be marginal.

To evaluate this we define a new flexible metric, in which we attempt to quantify the errors based on block size relevant for the particular control strategy. Let σ_p be the number of predicted events of a single type (i.e. either *away* or *in office*) in a time interval and σ_a the number of the actual occurred events of the same type in the same time interval. Finally τ is the total number of time steps in the interval, and we define the *event rate error*, (ere), as follow:

$$ere = \frac{|\sigma_a - \sigma_p|}{\tau} \tag{7}$$

The value of τ can be varied as appropriate for different application settings, based on the rising time for thermal variables, which may vary with HVAC equipment, the predictive control strategy and the physical environment. In Table III, we compare the proposed MMLM with the other methods, in terms of event rate error for different values of τ. As before, the aim is to keep the error ere as low as possible. Again, the pattern among the different methods is consistent. In particular, the MMLM method improves over the MC method by on average 0.7 percentage points, or reduces the error rate by 5%. Figure 6 shows the percentage improvement in the error for MMLM over each of the other methods. We note that there is little variation in the error rate as we increase the block size.

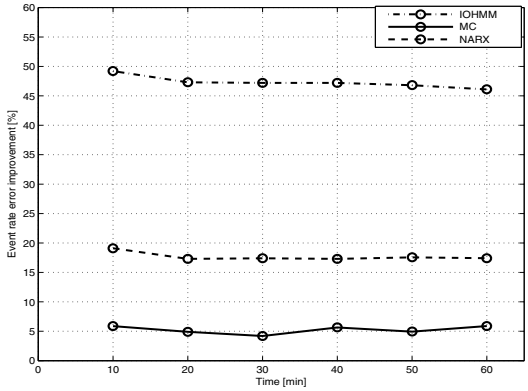


Figure 6. Improved event rate error for the proposed MMLM compared with MC method, NARX and IOHMM.

Table III
AVERAGE EVENT ERROR RATE FOR 12-STEP AHEAD PREDICTION, FOR VARIOUS BLOCK SIZES (τ) FOR: THE PROPOSED MIXTURE OF MULTI-LAG MARKOV CHAINS (**MMLM**), INPUT OUTPUT HIDDEN MARKOV MODEL (**IOHMM**), INHOMOGENEOUS MARKOV CHAIN (FIRST ORDER) (**MC**), AND NARX NEURAL NETWORK (**NARX**).

τ	**interval**	**MMLM**	**MC**	**IOHMM**	**NARX**
2	10	0.1288	0.1368	0.2536	0.1692
4	20	0.1303	0.1370	0.2475	0.1576
6	30	0.1287	0.1344	0.2438	0.1559
8	40	0.1258	0.1333	0.2386	0.1522
10	50	0.1385	0.1457	0.2604	0.1680
12	60	0.1234	0.1311	0.2291	0.1494

VI. CONCLUSION

We have proposed a Mixtures of Multi-Lag Markov Chains method to predict single person occupancy. The MMLM method allows us to get the benefits of a multi-order markov model, in that it can take account of multiple previous states, but without over-specifying the sequence. The method has been evaluated on real occupancy data collected by PIR sensors deployed in offices. The method has been compared against other state of the art methods. The results show that the MMLM method is able to achieve between 79% and 98% accuracy in 1 hour (12-step) ahead prediction, outperforming the other state of art approaches. In particular, it reduces the average error rate of its nearest competitor by 5%. The method is intended to be used as input to predictive control algorithms for HVAC systems, and we have introduced a flexible error metric designed to be tunable to the particular control algorithm and HVAC equipment. Finally, the runtime complexity is low, and hence is feasible for use as an occupancy prediction method in real time predictive control.

ACKNOWLEDGMENT

This work is supported by Intel Labs Europe and IRC-SET through the Enterprise Partnership Scheme, and also in part by Science Foundation Ireland through SFI Research Cluster ITOBO (grant No. 07.SRC.I1170), and grant No. 08/PI/I1912.

REFERENCES

[1] *Buildings Energy Data Book. Energy Efficiency and Renewable Energy*. U.S. Department of Energy, 2011.

[2] W. Angel. *HVAC Design Sourcebook*. McGraw-Hill Engineering, 2012.

[3] S. Goyal, H. Ingley, and P. Barooah. Zone-level control algorithms based on occupancy information for energy efficient buildings. In *American Control Conference (ACC)*, pages 3063–3068, 2012.

[4] V. Erickson, M. Carreira-Perpinan, and A. Cerpa. OBSERVE: Occupancy-based system for efficient reduction of hvac energy. In *Information Processing in Sensor Networks (IPSN)*, pages 258–269, 2011.

[5] Y. Ma, G. Anderson, and F. Borrelli. A distributed predictive control approach to building temperature regulation. In *American Control Conference (ACC)*, pages 2089–2094, 2011.

[6] A. Aswani, N. Master, J. Taneja, D. E. Culler, and C. Tomlin. Reducing transient and steady state electricity consumption in hvac using learning-based model-predictive control. *Proceedings of the IEEE*, 100(1):240–253, 2012.

[7] J. Siroky, F. Oldewurtel, J. Cigler, and S. Privara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88:3079–3087, 2011.

[8] F. Oldewurtel, A. Parisio, C. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, 45:15–27, 2012.

[9] A. E.-D. Mady, G. Provan, C. Ryan, and K. N. Brown. Stochastic model predictive controller for the integration of building use and temperature regulation. In W. Burgard and D. Roth, editors, *AAAI*. AAAI Press, 2011.

[10] V. Lesser, M. Atighetchi, B. Benyo, B. Horling, V. L. M. Atighetchi, A. Raja, R. Vincent, P. X. A. S. X. Zhang, T. Wagner, P. Xuan, and S.X. Zhang. The intelligent home testbed. In *Proceedings Of The Autonomy Control Software Workshop*, 1999.

[11] M.C. Torrance. Advances in human computer interaction: The intelligent room. In *Working Notes of CHI 95 Research Symposium, Denver, Colorado*, 1995.

[12] M.C. Mozer. The neural network house: An environment that adapts to its inhabitants. In *Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments*, 1998.

[13] D.C. Michael, M. Youngblood, E.O. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja. Mavhome: An agent-based smart home. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, pages 521–524, 2003.

[14] R. Dodier, G. Henze, D. Tiller, and X. Guo. Building occupancy detection through sensor belief networks. *Energy and Buildings*, 38(9):1033–1043, 2006.

[15] J. Page, D. Robinson, N. Morel, and J.-L. Scartezzini. A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings*, 40(2):83 – 98, 2008.

[16] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer,, 2006.

[17] Peter Hall and Efstathia Bura. Nonparametric methods of inference for finite-state, inhomogeneous Markov processes. *Bernoulli*, 10(5):919–938, 2004.

[18] Don Coppersmith and Chai Wah Wu. Conditions for weak ergodicity of inhomogeneous Markov chains. *Statistics and Probability Letters*, 78(17):3082–3085, 2008.

[19] Yan Li and Heung-Yeung Shum. Learning dynamic audio-visual mapping with input-output Hidden Markov models. *IEEE Transactions on Multimedia*, 8(3):542–549, 2006.

[20] T. Lin, B. G. Horne, P. Tino, and C. L. Giles. Learning long-term dependencies in narx recurrent neural networks. *IEEE Transactions on Neural Networks*, 7(6):1329–1351, 1996.