Sustainable Cities and Society xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Hidden Markov Models for indirect classification of occupant behaviour

J. Liisberg[a,*], J.K. Møller[a], H. Bloem[c], J. Cipriano[b], G. Mor[b], H. Madsen[a]

[a] DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark
[b] CIMNE-UPC, Terrassa, Spain
[c] European Commission, JRC – IET, RE-Unit, TP 450, I-21020 Ispra, Italy

ARTICLE INFO

Article history:
Received 16 February 2016
Received in revised form 5 July 2016
Accepted 6 July 2016
Available online xxx

Keywords:
Occupant behaviour
Indirect classification
Hidden Markov Models

ABSTRACT

Even for similar residential buildings, a huge variability in the energy consumption can be observed. This variability is mainly due to the different behaviours of the occupants and this impacts the thermal (temperature setting, window opening, etc.) as well as the electrical (appliances, TV, computer, etc.) consumption.

It is very seldom to find direct observations of occupant presence and behaviour in residential buildings. However, given the increasing use of smart metering, the opportunity and potential for indirect observation and classification of occupants' behaviour is possible. This paper focuses on the use of Hidden Markov Models (HMMs) to create methods for indirect observations and characterisation of occupant behaviour.

By applying homogeneous HMMs on the electricity consumption of fourteen apartments, three states describing the data were found suitable. The most likely sequence of states was determined (global decoding). From reconstruction of the states, dependencies like ambient air temperature were investigated. Combined with an occupant survey, this was used to classify/interpret the states as (1) absent or asleep, (2) home, medium consumption and (3) home, high consumption. From the global decoding, the average probability profiles with respect to time of day were investigated, and four distinct patterns of occupant behaviour were observed. Based on the initial results of the homogeneous HMMs and with the observed dependencies, time dependent HMMs (inhomogeneous HMMs) were developed, which improved forecasting. For both the homogeneous and inhomogeneous HMMs, indications of common parameters were observed, which suggests further development of the HMMs as population models.

1. Introduction

In building-design optimisation, energy diagnosis, performance evaluation and building energy simulations, the impact of the occupants' behaviour is often under-recognised and over-simplified. The influences of occupant behaviour are complex and stochastic. In recent years, the importance of occupant behaviour has been recognised, and many new approaches have been developed to model the effect of occupant behaviour. To achieve an overview of the approaches in modelling occupant behaviour in buildings, a small literature study has been carried out.

In building-related models for occupant behaviour, there are two main focus areas, (1) occupant presence and movement and (2) occupant interaction with indoor climate (adjusting a thermostat, opening a window for ventilation, turning on lights or closing blinds). Studies related to these areas are typically related to either residential or commercial buildings (Andersen, Iversen, Madsen, & Rode, 2014; D'Oca, Fabi, Corgnati, & Andersen, 2014).

The following is a cursory review of the papers in the literature study.

In modelling occupant presence and movement in office buildings, both homogeneous (Wang, Yan, & Jiang, 2011) and time-inhomogeneous (Andersen et al., 2014; Page, Robinson, Morel, & Scartezzini, 2008) Markov chains have been used. The models are used as input for building energy simulations. In Andersen et al. (2014) a comparison of the performance between homogeneous and time-inhomogeneous models was carried out, and the inhomogeneous model was found to be superior.

In the work on modelling overtime schedules in office buildings, Sun, Yan, Hong, and Guo (2014) uses a binomial distribution

* Corresponding author.
E-mail address: jarl@dtu.dk (J. Liisberg).

http://dx.doi.org/10.1016/j.scs.2016.07.001
2210-6707/

Nomenclature

AIC	Akaike information criterion
BIC	Bayesian information criterion
HMM	Hidden Markov Model
CRPS	Continuous Rank Probability Score
cdf	cumulative distribution function
pdf, $p(x)$	probability mass or density function
m	number of states
t, s	a time stamp in discrete time
T	maximum of t, i.e. $t \in \{1, \ldots, T\}$
$\mathbb{N}$	the natural numbers
$\mathbb{R}$	the real numbers
$i, j, k \in \mathbb{Z}$	integers
C_t	the state of a Markov chain at time t
X_t	the state of the random process $\{X_t\}$ at time t
x_t	the observation of the random process $\{X_t\}$ at time t
$\boldsymbol{A}, \boldsymbol{\Gamma}$	matrices
$\boldsymbol{a}, \boldsymbol{\theta}$	row vectors
$\boldsymbol{a}'$	a column vector

to represent the number of occupants working overtime and an exponential distribution to describe the duration of overtime. The overtime model is used to generate overtime schedules as input to building energy simulations.

Based on seven years of measuring window opening and closing behaviour, three modelling methods for prediction of actions on windows were developed (Haldi & Robinson, 2009). The methods are logistic probability distributions, Markov chains and continuous-time random processes.

In a field study of the thermal comfort of office occupants (Haldi & Robinson, 2008), logistic regression was used to predict the probability of occupants' actions.

In a simulation study of an adaptive automation system for the visual comfort of office occupants (Gunay, O'Brien, Beausoleil-Morrison, & Huchuk, 2014), the models for predicting occupants' turning light on/off and opening/closing blinds are based on Markovian state transition probabilities.

For air-conditioning in residences, Tanimoto and Hagishima (2005) identifies on/off state transition probability functions dependent on indoor and outdoor temperature. These functions are requisite for applying a Markov model to a cooling schedule.

A methodology to predict residential occupants' time-dependent activities is presented in Wilke, Haldi, Scartezzini, and Robinson (2013). Using a time-use survey, the model is calibrated based on three time-dependent quantities: (1) the probability of being home, (2) the conditional probability of starting an activity while at home, and (3) the probability distribution function for the duration of the activity. Transitions between activity types are modelled as an inhomogeneous Markov process.

Studies in building energy simulations (D'Oca et al., 2014; Fabi, Andersen, Corgnati, & Olesen, 2013) have investigated the impact of changing from standardised occupant behaviour profiles to a probabilistic approach in simulating these profiles. D'Oca et al. (2014) showed a large increase in energy consumption, with this approach.

Based on data mining using cluster analysis, Yu, Fung, Haghighat, Yoshino, and Morofsky (2011) examines the influences of occupant behaviour on building energy consumption. A methodology for identifying energy-inefficient behaviour in residential buildings was developed.

Yu, Haghighat, and Fung (2016) provides an overview of recent studies undertaking predictive and descriptive tasks in the building field. This is done by using data-mining techniques to extract hidden but useful knowledge. For occupant behaviour, a key issue is to understand the interactions between occupant behaviour and other influencing factors.

From this literature study, different approaches seem highly problem-specific. Many use Markov chains/processes in the description of the transition between presence, non-presence, movement between rooms and transitions between activities. This indicates that Markov chains/processes are highly useful for modelling occupant behaviour in a wide range of settings. With the idea to extract hidden knowledge from data, and using Markov chains to model occupant behaviour, this has spurred us to look at methods to observe occupant behaviour in an indirect manner e.g. Hidden Markov Models (HMMs).

When measuring the electricity consumption in similar residential buildings, the variability in the consumption is often very large. This is mainly due to the diversity of occupant behaviour. The occupants not only impact the electricity consumption, but also the general energy consumption (Aerts, Minnen, Glorieux, Wouters, & Descamps, 2014; Andersen, 2012; Zhao, Lasternas, Lam, Yun, & Loftness, 2014). Due to privacy concerns, and the cumbersome work of obtaining direct observations of occupant behaviour, indirect means of classifying occupant behaviour are needed. Several models have been developed for simulation purposes using data-mining approaches (Zhao et al., 2014). Based on a time-use survey, it is suggested that occupant behaviour in residential buildings could be classified according to the following three states: (1) at home and awake, (2) sleeping, or (3) absent (Aerts et al., 2014). Given the increasing use of smart metering by the utilities, the potential of using these metering data for indirect classification of residential occupant behaviour is now possible. Applying a homogeneous Hidden Markov Model (HMM) to electricity consumption data from a residence results in a number of states that could be interpreted in a similar manner (Zucchini & MacDonald, 2009).

The focus of the study presented in this paper is to investigate the applications of HMMs on frequent observations of electricity consumption in residences. The study seeks to test the hypothesis that, by applying HMMs on observations of electricity consumption, we can:

1 Classify the states of the HMM, i.e. of the occupant(s) in accordance to occupant behaviour.
2 Identify possible covariates/explanatory variables.
3 Forecast and simulate future energy consumption.

(1), (2) and (3) can be solved by both homogeneous and time-inhomogeneous models. It is suggested that to improve the capabilities for forecasting and simulation, covariates/explanatory variables and time-inhomogeneous Markov chains, are needed (Andersen et al., 2014).

The study also seeks to investigate whether the HMMs for each residence can be collected in population models (Madsen & Thyregod, 2011) to forecast or simulate groups of residences.

The aim is to present a modelling framework for HMMs on frequent observations of electricity consumption, and then apply this framework to several residential apartments. Focus will be on interpreting the states of the HMMs to validate the models and suggest further development of these models.

The outcome is an initial framework for using HMMs on frequent observations of electricity consumption and proposals for further model development. This study is a further elaboration of some of the results in Bloem et al. (2015).

Please cite this article in press as: Liisberg, J., et al. Hidden Markov Models for indirect classification of occupant behaviour. *Sustainable Cities and Society* (2016), http://dx.doi.org/10.1016/j.scs.2016.07.001