Building and Environment 126 (2017) 318–330

Contents lists available at ScienceDirect

Building and Environment

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

Exploration of the Bayesian Network framework for modelling window control behaviour

Verena M. Barthelmes[a,b,*], Yeonsook Heo[b], Valentina Fabi[a], Stefano P. Corgnati[a]

[a] Department of Energy (DENERG), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
[b] Department of Architecture, University of Cambridge, 1-5 Scroope Terrace, Cambridge, United Kingdom

ARTICLE INFO

Keywords:
Occupant behaviour
Bayesian networks
Window control behaviour
Stochastic modelling

ABSTRACT

Extended literature reviews confirm that the accurate evaluation of energy-related occupant behaviour is a key factor for bridging the gap between predicted and actual energy performance of buildings. One of the key energy-related human behaviours is window control behaviour that has been modelled by different probabilistic modelling approaches. In recent years, Bayesian Networks (BNs) have become a popular representation based on graphical models for modelling stochastic processes with consideration of uncertainty in various fields, from computational biology to complex engineering problems. This study investigates the potential applicability of BNs to capture underlying complicated relationships between various influencing factors and energy-related behavioural actions of occupants in buildings: in particular, window opening/closing behaviour of occupants in residential buildings is investigated. This study addresses five key research questions related to modelling window control behaviour: (A) variable selection for identifying key drivers impacting window control behaviour, (B) correlations between key variables for structuring a statistical model, (C) target definition for finding the most suitable target variable, (D) BN model with capabilities to treat mixed data, and (E) validation of a stochastic BN model. A case study on the basis of measured data in one residential apartment located in Copenhagen, Denmark provides key findings associated with the five research questions through the modelling process of developing the BN model.

1. Introduction

Accounting for uncertainty has become a crucial aspect in the domain of building energy simulation for incorporating human behaviour that impacts building energy performance and comfort expectations. Human behaviour such as occupancy, control of energy systems, occupants' interaction with the building envelope and other comfort criteria settings are considered as key sources of uncertainty in the prediction of building energy use. Indeed, occupant behaviour varies significantly between individuals, which results in large variation of the indoor environmental quality and energy consumptions of the buildings [1–3]. Extended literature reviews and state-of-the-art analyses confirm that an accurate modelling of occupant behaviour is a key factor to bridging the gap between predicted and actual energy performance of buildings [4–8]. Frequently, simulation-based design analysis relies on standard use and operation conditions such as fixed schedules for occupancy levels, light switching, ventilation rates and temperature setting. These assumptions often lead to an oversimplification of the human-related variables creating discrepancies between predicted and real energy use of the building. Thus, in recent years, probabilistic modelling approaches have been applied to capture the stochastic nature of energy-related human behaviour when predicting building energy consumptions in dynamic simulation programs [9].

Occupant's action of window opening/closing has an important impact on building energy use and indoor environmental quality (IEQ) by changing the amount of fresh air to the building. Several studies have been carried out to develop stochastic models for predicting the occupant's interaction with the windows. These models are based on statistical algorithms to predict the probability of a specific condition or event, such as the window state or the window opening/closing action, given a set of environmental or other influential factors. Most popularly used methods include logit analysis, probit analysis, and Markov chain processes. Nicol [10] developed a logit regression model to predict the state of windows in a probabilistic manner as the function of indoor and outdoor temperatures. Andersen et al. [11] also used a logistic regression model based on a more comprehensive set of indoor and outdoor environmental variables to infer the probability of opening and closing a window. The study on the basis of field measurements from 15 Danish

* Corresponding author. Department of Energy (DENERG), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy.
E-mail address: verena.barthelmes@polito.it (V.M. Barthelmes).

http://dx.doi.org/10.1016/j.buildenv.2017.10.011
Received 29 June 2017; Received in revised form 5 October 2017; Accepted 8 October 2017
Available online 10 October 2017

dwellings defined four separate models of occupants' window action behaviour patterns for different ownerships and ventilation types. Logit regression models have been also applied in other studies for modelling window control behaviour [12–14]. Zhang and Barret [15] developed a probit model for predicting window opening/closing actions in a naturally ventilated office building considering the outdoor temperature as the only independent variable. Haldi and Robinson [16] and [17] tested different modelling approaches and demonstrated that a discrete-time Markov process approach, which takes into account real dynamic processes, leads to a higher predictive power compared with the logit regression approach. Modelling approaches based on Markov chain processes are used in Refs. [18] and [19] to predict window states based on their previous states in office buildings and houses, respectively. As these models consider real dynamic processes by providing transition probabilities between the states of a window, they are limited to capture the dynamic effect of changes in indoor and outdoor environmental conditions on window opening and closing actions.

This paper investigates the capabilities of the Bayesian Network framework to model occupant behaviour in the context of thermal comfort and building energy analyses in order to bridge the gap between simulations' outcomes and reality. Bayesian Networks (BNs), or rather graphical belief networks, are widely applicable and have become a popular representation for encoding uncertainty in decision-making processes based on incomplete datasets [20]. In recent years, BNs have been used in many fields, from On-line Analytical Processing (OLAP) [21], cancer prognosis and epidemiology [22], the modelling of dwelling fire development and occupancy escape [23], to speech recognition [24]. In the buildings domain, BNs have been introduced to estimate the effects of the indoor climate on the productivity of occupants [25], to investigate the relationship between indoor environmental parameters, measurements from body sensors and self-reported activities by the occupants [26], to predict occupancy patterns in buildings [27,28], to model energy-related user behaviour for building energy management [29,30], and to predict indoor environmental conditions [31]. So far, these studies based on BN models treat either discrete variables only or continuous variables only.

In comparison to the above-mentioned regression-based models, BN-based approaches are able to flexibly model complex relationships between diverse explanatory variables and window control behaviour by constructing a joint probability distribution over different combinations of the domain variables. Indeed, the BN model permits to easily model joint conditional dependencies of the entire set of variables through a graphical representation of the model structure [32]. The BN model also allows for structuring a variety of explanatory variables and multiple target variables in a hierarchical manner. In addition, BNs are demonstrated to yield good prediction accuracy even with small datasets [33]. They also have capabilities to handle incomplete datasets by using Expectation-Maximization (EM) algorithms [34] in which missing data can be marginalized by integrating over all the possibilities of the missing values. Furthermore, the BN model provides a clear semantic representation of relationships between variables, which facilitates flexibly structuring a model and training it against available data in wider and interdisciplinary research communities.

This paper demonstrates the applicability of the Bayesian Network (BN) framework for predicting window opening/closing behaviour of building occupants based on the measurements in a residential apartment located in Copenhagen, Denmark. In particular, the paper addresses five key research questions related to developing a BN model for predicting window-use patterns. The first set of three research questions addresses general issues relevant to modelling window control behaviour:

A. Which variables are key drivers that determine window control behaviour?
B. What level of correlations resides between variables and should they be captured in the BN model?
C. What is the most suitable target variable of window control behaviour?

Regarding the first question, the Kolmogorov-Smirnov Test (K-S Test) is applied to evaluate which variables are main drivers for window control actions. For the second question, the Kendall Tau correlation coefficient is used to investigate correlations between identified variables and accordingly model them in the BN. The third question (C) investigates different target variables commonly used in the literature (i.e., window opening/closing event and window state) in terms of the modelling accuracy.

The second set of research questions addresses modelling challenges related to the applicability of the BN framework for modelling occupants' window control behaviour:

D. How to handle mixed data in the BN framework?
E. How to validate stochastic BN models?

A key question of this paper addresses how to handle mixed data in the BN framework. Traditional BN approaches to treat either discrete variables or continuous variables are not suited to modelling window control behaviour as datasets typically consist of both continuous variables (e.g., indoor temperature, CO_2 concentration) and non-continuous variables (e.g., binary control actions, time of the day). This study tries to overcome this problem by proposing a modelling procedure that allows for handling mixed data, particularly with use of the bnlearn package [35] in the statistical software R environment [36]. The prediction accuracy of the model is evaluated through a series of methods suitable to validate stochastic models.

2. The Bayesian Network framework

2.1. Bayesian Networks

Bayesian Networks are graphical models that represent probabilistic dependencies between discrete or continuous variables (X_i) [37]. In the models, variables are presented by nodes and their relationships are represented by arcs. The direction of arcs determines a hierarchical structure of nodes. Fig. 1 shows an example of a Bayesian Network that represents the probabilistic dependencies between an occupant's action and a set of variables (VAR) that potentially impact the action.

A network structure is often explained with a family metaphor; if there is an arc starting from one node to another, the former is a parent of a child (the latter). Extending the metaphor, in a directed chain of nodes, one node is an ancestor of another if it appears earlier in the chain, whereas a node is a descendant of another node if it comes later in the chain. For instance, as shown in Fig. 1, as there is an arc from X_1 to X_3, node X_1 is a parent of node X_3. The graphical structure of a Bayesian network, denoted as $G = (V,A)$, is a Directed Acyclic Graph (DAG), where V is the node (or vertex) set and A is the arc (or edge) set. The DAG defines a factorization of the joint probability distribution of $V = \{X_1, X_2, \ldots, X_n\}$, often called the global probability distribution, into a set of local probability distributions, one for each variable [36]. This factorization is based on the assumption that Bayesian Networks have a Markov property [37], which indicates that the state of a random variable X_i depends only on its parents ΠX_i. In general, Bayesian Network modelling requires the assumption of the Markov property.

In principle, BN models flexibly represent different typologies and handle a mix of various data types. Yet, so far, BN models used in most existing studies are limited to either a discrete case or a continuous case. This limitation is mostly due to the fact that, unfortunately, most software available for developing BN models are applicable to either discrete or continuous data and, thus, do not permit yet to handle a mix of continuous and discrete datasets in one BN model, particularly when discrete variables are conditional on continuous variables. Equations (1) and (2) define a joint probability distribution for a discrete case and