Hindawi
Journal of Sensors
Volume 2019, Article ID 7129872, 12 pages
https://doi.org/10.1155/2019/7129872

Research Article

Estimating Occupancy from Measurements and Knowledge Using the Bayesian Network for Energy Management

Manar Amayri ⬤,[1] **Stephane Ploix**,[2] **Hussain Kazmi**,[3] **Quoc-Dung Ngo**,[4] **and E. L. Abed E. L. Safadi**[5]

[1]*G-SCOP Laboratory/Grenoble Institute of Technology, 1, 46 Avenue Felix Viallet, 38031 Grenoble, France*
[2]*G-SCOP Laboratory/Grenoble Institute of Technology, France*
[3]*KU Leuven, Department of Electrical Engineering (ESAT), Belgium*
[4]*People's Security Academy of Hanoi, Vietnam*
[5]*Grenoble Alpes University (UGA), France*

Correspondence should be addressed to Manar Amayri; eng.amayri@gmail.com

Received 7 June 2018; Accepted 5 December 2018; Published 10 April 2019

Guest Editor: Anatoliy Sachenko

Copyright © 2019 Manar Amayri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A general approach is proposed to determine occupant behavior (occupancy and activity) in offices and residential buildings in order to use these estimates for improved energy management. Occupant behavior is modelled with a Bayesian network in an unsupervised manner. This algorithm makes use of domain knowledge gathered via questionnaires and recorded sensor data for motion detection, power, and hot water consumption as well as indoor CO_2 concentration. Different case studies have been investigated with diversity according to their context (available sensors, occupancy or activity feedback, complexity of the environment, etc.). Furthermore, experiments integrating occupancy estimation and hot water production control show that energy efficiency can be increased by roughly 5% over known optimal control techniques and more than 25% over rule-based control while maintaining the same occupant comfort.

1. Introduction

Buildings and appliances are steadily becoming more energy efficient due to regulatory changes such as building standards. As a result, energy consumption related to human activity is relatively much bigger than before. In addition, demand side management programs in both electric and heat grids can lead to variable tariffs that building occupants have to take into account in their everyday life. These complex interactions have the potential to substantially affect the end energy consumption and influence grid operation. Existing building automation systems can compute relevant set points for the HVAC systems according to predefined occupant calendars, but this is often not an adequate replacement for a dedicated building manager.

Energy management systems, or energy managers, provide two key capabilities to building managers and occupants: automation and recommendations for behavioral change. Because of thermal inertia and diurnal and seasonal demand patterns, these energy managers have to be able to anticipate the upcoming day, at least, i.e., they have to be proactive, rather than reactive, to deal with current conditions.

Occupancy estimation, a key piece to automating energy management, can be done via cameras and a posteriori labeling. The use of supervised learning algorithms on site can be ruled out for most practical cases because of the requirements to obtain relevant data in the form of video stream as it raises privacy concerns. Even in settings where local learning from gathered data might make this possible, annotation, i.e., occupancy labeling from videos, is complicated and noisy. Gathering sufficient data to learn occupancy patterns reliably from this data is a challenging problem.

This paper tackles the aforementioned issues and proposes an occupancy estimation approach based on human knowledge and involving the occupants through a questionnaire. It focuses on developing features for the estimation of

www.manaraa.com

human number and activities which cannot be measured directly, with the help of nonintrusive sensors. Algorithms from machine learning have been adapted to a residential setting which is a sensor test bed with a large number of ENOCEAN sensors (self-powered wireless standard for building automation and smart homes), with regular/event-based data reporting. It proposes an occupancy and activity estimation approach based on the human knowledge and gathering feedback from the occupants in the studied area.

In addition, a novel algorithm has been presented to optimize the energy demand for hot water production in residential buildings using these occupancy estimates. Optimizing energy consumed by hot water systems is likewise an important consideration for energy managers as it accounts for between 10 and 25% of end building energy consumption (Pérez-Lombard et al. [1]). With innovations in facade technology, this draw is expected to further gain in importance. By including occupancy estimates for control of hot water vessels, this work extends the state of the art where efficiency improvements are usually derived from thermodynamic or consumption-based optimization [2].

In [3], the authors present a model of the Bayesian network for an office case study, while in this work, we present comparisons between residential and office case studies besides an application of using occupancy estimation for energy-efficient hot water production.

In the paper [4], a method of estimation occupancy has been proposed using a supervised algorithm (decision tree) with interactive learning concept, which is used to send a question to the end user to collect the required training database. Only office context has been presented and analyzed, while the current paper has proposed a method to estimate occupancy and activities in different contexts using an unsupervised tool Bayesian network (BN); it depends on the sensor data, questionnaire, and knowledge to build the conditional probability tables. An office and 2 residential contexts have been presented and analyzed.

Section 3 presents a state of the art about occupancy and activity estimation. Section 4 discusses the problem statement of occupancy and activity estimation in nonsupervised learning algorithms (i.e., Bayesian networks). Section 5 discusses different case studies for occupancy estimations using Bayesian network (BN) knowledge. Section 6 proposes structures and process for BN. Section 7 analyzes the resulting occupancy and activity estimations for residential contexts. Section 8 presents an experiment integrating occupancy estimation and hot water production control.

2. State of the Art

In the building physics community, there is a growing interest in occupant behavior because of its importance in determining end energy demand and reducing energy waste in buildings [5, 6]. Occupant behavior can be studied from building physics to human biology, through sociology and psychology in order to model and assess thermal and visual comfort as well as other aspects of indoor environmental quality.

Reference [7] uses neural networks for the agents to learn occupant behavior from recorded data to ensure comfort. After a learning phase, agents know which actions increase occupant comfort while operating in different environmental conditions. In a recent study, Reference [8] models various occupancy profiles by calibration to predict the use of fans, heating systems, and windows. The hidden Markov model (HMM) and its variants have been extensively studied and utilized quite often to detect activities in the past. Oliver et al. [9] utilize an extension, *layered HMMs* in their model SEER, to detect various activities like desk work, phone conversations, and presence. The layered structure of their model makes it feasible to decouple different levels of analysis for training and inference. Each level in the hierarchy can be trained independently, with different feature vectors and time granularities. Once the system has been trained, inference can be carried out at any level of the hierarchy. One benefit of such a model is that each layer can be trained individually in isolation, and therefore, the lowest layer which is most sensitive to environmental noise can be retrained without affecting the upper layers.

Different researches in occupancy estimation have also been analyzed. Methods started by using single feature classifiers which estimate two classes (presence and absence) to multisensor and multifeature models. The main approach, which is applied in many commercial buildings, is the use of passive infrared (PIR) sensors for occupancy estimation. However, motion detectors fail to detect a presence when occupants remain relatively still, which is a normal behavior during activities like working on a computer or regular desk work. This leads to consider the use of PIR sensors alone for estimating occupancy is not sufficient. Combining of PIR sensors with other sensors can be useful as discussed in [10]. It uses motion sensors and magnetic reed switches for occupancy counting in order to increase the efficiency of the HVAC systems. However, video cameras have been used a lot in activity recognition and occupancy detection [11]. While most of these can estimate occupancy patterns reliably, they cannot be employed in most office buildings for reasons like privacy and cost concerns. Jaehoon Jung et al. [12] show a normalized human height estimation algorithm using uncalibrated cameras. The algorithm tracks moving objects and carries out tracking-based automatic camera calibration.

Estimating activities undertaken by building occupants is a more challenging task than estimating occupancy; however, some of the previously mentioned sensing techniques can be extended for this purpose. For instance, the use of pressure and PIR sensors to determine presence/absence in single desk offices and tag activities has been discussed in [13]. However, for various applications like activity recognition or context analysis within a larger office space, information regarding the presence or absence of people is not sufficient and an estimation of the number of people occupying the space is essential.

Various learning models have been used by researchers in the field of activity recognition and occupancy analysis. This generally involves creating a probabilistic or statistical activity model that is augmented with training data. The task of
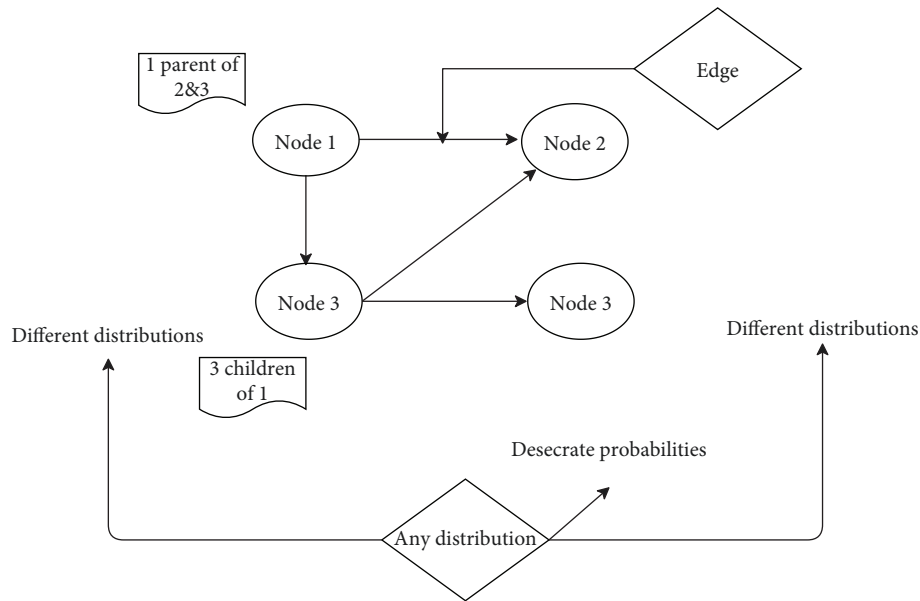
FIGURE 1: Elements of a Bayesian network: different probability distributions.

the model is to recognize patterns that differentiate various classes in the training data and apply this knowledge for the prediction/classification of test data. This allows for data-driven methods, i.e., where domain-specific knowledge is not necessary. Reference [14] classifies these data-driven approaches as follows:

(1) *Generative Modeling.* Here, the target is to create a model that uses training data samples to form a description of the complete input feature space. Examples include the Naive Bayes classifier which assigns a probability to every event or the hidden Markov model which additionally models temporal dependencies of events. These methods can suffer from generalization issues in high-dimensional feature spaces because of limited training data [15]

(2) *Discriminative Modeling.* Rather than representing the entire input space, discriminative modeling only focuses on classification. The primary objective of such a model is to find a decision boundary (or boundaries). Examples include the KNN (K-nearest neighbor) classifier, decision trees, and SVMs [16]

(3) *Heuristic-Based Modeling.* Such models use a combination of both generative and discriminate models along with some heuristic information [17]

In this paper, an occupancy modelling framework is proposed which fully exploits information available from low-cost, nonintrusive environmental sensors and knowledge. Without loss of generality, this information is then combined with a control framework to minimize energy consumed for hot water production. While this allows for meaningful energy efficiency results in residential buildings, the sensing requirements still pose certain limitations regarding occupant privacy.

3. Proposed General Process for Occupants and Activity Estimation

3.1. Bayesian Network (BN). A Bayesian network (BN) is a directed acyclic graphical model for reasoning with uncertainty among the random variables (see Figure 1). Moreover, the variables have been shown as nodes in both cases continuous or discrete, while the direct connections between the nodes have been shown by the edges. A probability function has been built for each node which should take a specific number of values for the input node variables and gives as output; the probability distribution of the variable has been shown by the node. This probability depends on the causal relationships between the variables (cause/effect relation).

A Bayesian network (BN) is a knowledge model of knowledge using conditional probabilities and evidence to derive resulting probabilities.

A tool has been developed using Python 3 to deal with the structure of the building and the input of conditional probabilities. It is an adaptation of the *libpgm* module [18]. In this paper, all conditional probabilities are determined, i.e., they were depending on observations and analyzing office occupant questionnaire.

4. Case Study

4.1. Experimental Setup. In this section, the different case studies are described which were used to evaluate the performance of the proposed algorithm. In the first instance, an office context is proposed because it can be considered as a simple and initial single-zone environment with lots of number of sensors. In addition, two residential buildings have been investigated which vary according to sensors availability, user feedback, and complexity of the environment.

Besides, a practical example for each case study will be presented later in Section 7.

(1) *Case Study 1.* This is an office. The system of the sensor network contains the following:

  (i) An ambiance sensing network measures luminance, temperature, humidity, motions, CO_2 concentration, power consumption, door and window position, and microphone. ENOCEAN protocol has been used [4]

  (ii) There are 2 video cameras to record the real number of occupants and activities. These 2 cameras are installed only for the validation period

  (iii) There is a web application with a centralized database for getting back continuously the data from different sources

(2) *Case Study 2.* This is an apartment context, which is presented with a multizone application including lots of number of sensors and different types of activities. The installation of the sensor network contains the following:

  (i) Temperature sensor for each room

  (ii) Motion sensor for each room

  (iii) Window contact sensor for each room

  (iv) Door contact sensor for each room

  (v) Fridge, oven power consumption, microwave, espresso, steam cooker, kettle, and dishwasher sensors in the kitchen

  (vi) Power consumption sensors for each room connected to different appliances

  (vii) Humidity sensor for each room

  (viii) Luminosity sensor for each room

The apartment consists of two bedrooms, a living room, a kitchen, a working office, a bathroom, and a toilet. The doors and windows are equipped with contact position sensors, which provide a binary number related to the position of the door or window, i.e., 1 for open and 0 for closed. Only the contact sensors give values as a binary number while the rest of the sensors give their respective variables to the extent of the intensity, i.e., the temperature in the kitchen.

(3) *Case Study 3.* This is a multizone house with a different number of activities while a limited number of sensors installed only in the common room. This case study is part of a big project where an energy management system is developed in order to improve building energy efficiency. By implementing the occupancy detection model, the control system can be improved. The complexity of this context model has been increased due to as follows: as in real application settings, getting feedback from occupants during design and validation phase was impossible.

Moreover, building the occupancy model depends mostly on analyzing historical data.

The unsupervised learning algorithm is a practical paradigm for the residential and commercial building environment, where supervised learning is mostly not accepted by the users because of privacy issues.

5. Methodology

As previously mentioned, three cases of study were used to exemplify the capability of the proposed approach to estimate occupancy and activity. In this section, the adaptations to the methodology as mandated by different sensory inputs are highlighted [19].

5.1. Case Study 1: Occupancy Estimation in Office Building. In the case of the office building, collecting occupancy and activity feedback was easy due to the presence of video cameras and microphone. Besides, the possibility to interview occupants during the design and validation periods of the estimation models provided additional information.

This office has been equipped with 30 sensors to evaluate the occupancy and activity estimation of the trained Bayesian network. To estimate the number of occupants, its relationship with the sensed variables in the office environment has to be discovered. To do so, the office environment can be represented as a set of state variables, $A_t = [A_1, A_2, \ldots, A_m]_t$. This set of state variables A at any instance of time t must be indicative of occupancy. A state variable can be termed as a feature and therefore the set of features as feature vector. Similarly, m-dimensional space that contains all possible values of such a feature vector is the feature space.

The underlying approach for the experiments is to formulate the classification problem as a mapping from a feature vector to some feature space that comprises several classes of occupancy or activity. Therefore, the success of such an approach depends on how good the selected features are. In this case, features are attributes from multiple sensors accumulated over a time interval. The choice of interval duration is highly context dependent and has to be done according to the required granularity. However, some features do not allow this duration to be arbitrarily small. As an example, it has been observed that CO_2 levels do not rise immediately, and one of the factors affecting this delay is the ventilation of the space being observed. The results presented in this paper are based on an interval of $T_s = 30$ minutes (which is referred hereafter as 1 quantum). Features are the information extracted from raw data, i.e., acoustic pressure from a microphone, time slot, occupancy from power consumption, door or window position, motion counting, day type, and indoor temperature.

One quantitative measurement of the usefulness of a feature is *information gain*, which depends on the concept of *entropy* [20]. Information gain is helpful to distinguish among a large set of features, the most worthwhile to consider for occupancy estimation. A supervised learning approach has been used. Occupancy counting was manually

annotated using a video feed from two cameras strategically positioned in an office. A decision tree algorithm has been used because it results in human readable rules and can be adjusted using expert knowledge and yield estimation rules (if then) from the decision tree structure. It is likely that from the large set of features considered in [20], some may not be useful to achieve our target of occupancy classification. After removing less important features according to the information gain formulation, the following main features were considered:

(1) *Motion Counter*. The PIR sensor in use is a binary sensor that reports a value of 1 whenever it senses some motion. These impulses have to be computed for each time quantum

(2) *Acoustic Pressure from Microphone*. The RMS amplitude feature was defined from the recording signal

(3) *Occupancy from Power Consumption*. Four sensors measuring power consumption have been connected to occupant laptops in the office

Some important questions have to be answered while building the BN structure:

(1) What type of knowledge is relevant to collect for filing the conditional probability tables of the BN?

(2) What variables within these domains should be modeled?

(3) What are the relationships between these variables?

(4) What probabilities describe these relations?

Instead of modelling occupancy levels as the exact number of occupants, three levels of occupancy have been chosen: low, medium, and high number of occupants. This reflects a compromise for simplicity as increasing the number of levels increases the complexity of the table of conditional probability of a Bayesian network structure. Reference [20] indicates a correlation between the power consumption and the number of occupants in the office. For the sake of simplicity, power consumption has also been discretized into three levels: low, medium, and high or L, M, and H, respectively. This discretization yields a probability table with 9 values (Figure 2).

To collect the required information for conditional probability tables of the Bayesian network, two questions were posed to the office occupants to determine the probability table for power consumption:

(1) The approximate schedule of occupants arriving and leaving the office

(2) The average time of laptop usage during working hours

According to user answers, the conditional probability tables are filled directly or can be calculated. It is important to note that the calendar of occupants could be used instead
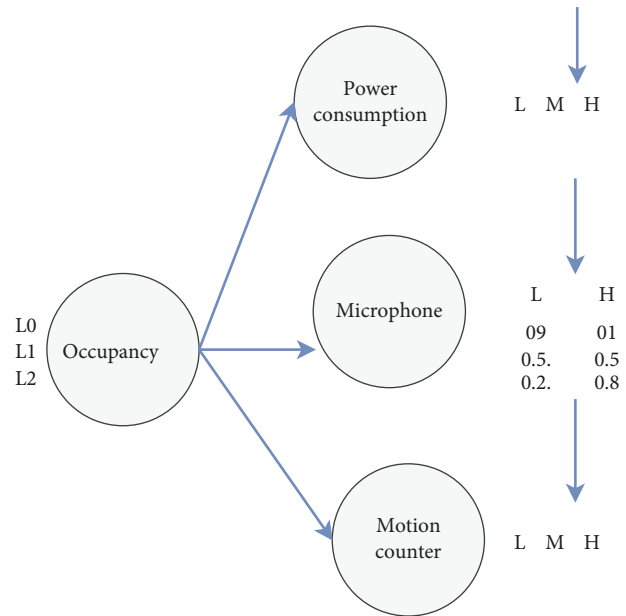

FIGURE 2: Bayesian network (BN) structure of an office H358.

of the question, in order to facilitate the configuration of the BN structure.

The recorded signal from the microphone can be discretized into two levels: low acoustic pressure and high acoustic pressure or L and H, respectively. This gives a table of probabilities with 6 values Figure 2.

The following questions were proposed to determine the probability table for the microphone:

(1) The kind of activities in the office (computer work, presentation, and Skype meeting)

(2) The frequency of discussions between the colleagues during the working day

The probability tables for motion counter have been suggested according to the general knowledge for three different cases: low motion, medium motion, and high motion or L, M, and H, respectively.

5.2. Case Study 2: Occupancy Estimation in Residential Building. Similar to case study 1, three occupancy levels have been chosen: low, medium, and high number of occupants. Furthermore, similar observations about correlation between occupancy and power consumption hold. The structure of the Bayesian network for occupancy estimation in the apartment has been developed taking into account the interaction between different rooms in the apartment. It is composed of 6 Bayesian networks, which can be seen in Figure 3 as follows:

(1) Five Bayesian networks, one for each room (two bedrooms, one kitchen, one common, and one office) in order to estimate the absence/presence in each considered room. Power consumption and motion detector sensors have been used in each room for presence estimation, while in the master bedroom,
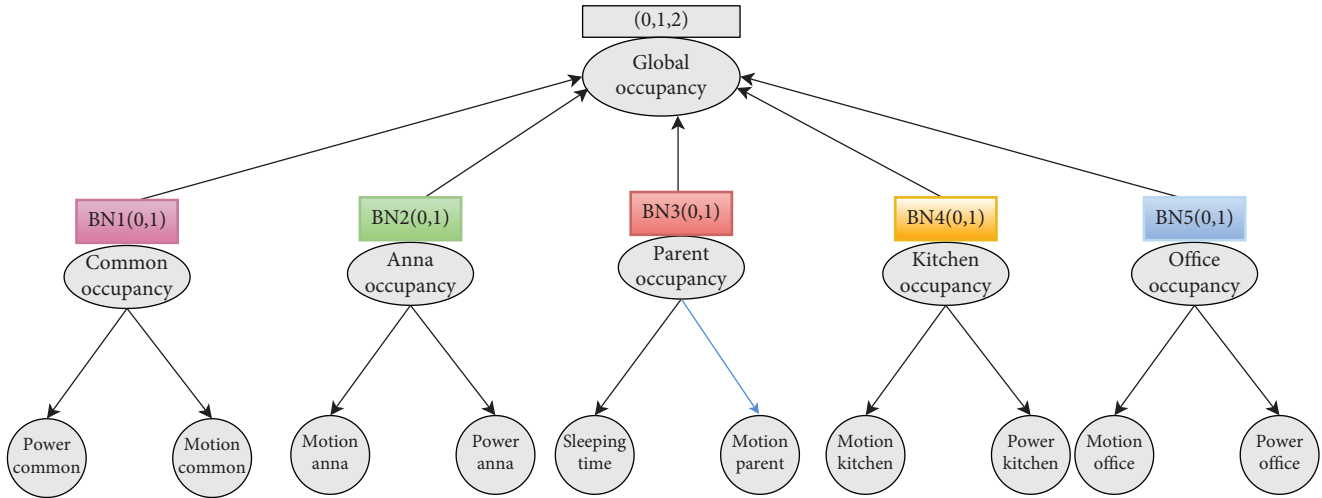
FIGURE 3: Bayesian network structure case study (apartment).

the time slot has been integrated because the presence in this room is related mostly to a sleeping time

(2) One additional global Bayesian network for the whole apartment to adapt the estimation results from the previous 5 Bayesian networks, because of the interactions between the rooms

Six conditional probability tables have been defined for each Bayesian network besides the global one. The global Bayesian networks required a table of 100 values, which include all the combinations of the absence/presence in each room in the apartment to estimate 3 levels of occupancy: low, medium, and high in the whole apartment.

For data collection, multiple sensors are installed in the apartment to sense different variables. Practically, it is challenging for the residential application to interact with the occupants in real time. Therefore, in this case study, the required knowledge has been extracted from questionnaires and analyzing historical data.

The following questions have been proposed to determine the probability tables:

(1) The average time for staying at the apartment during a weekday and weekend

(2) The frequency of inviting short- or long-term guests

(3) The frequency of leaving the apartment for visiting friends or traveling

According to the user answers, the conditional probabilities are either calculated or filled directly in the tables.

As presented in the Bayesian network structure (see Figure 2), power consumption and motion detection sensors are the most relevant features for occupancy estimation. This has been extrapolated from previous work in the office environment due to information gain concept.

5.3. Case Study 2: Activity Estimation in Residential Building. Tracking user activities in daily life and estimating occupant activities are important in different areas (i.e., mirroring

TABLE 1: Activity types.

Activities (1) with interactions	Activities (2) without interactions
Cleaning	Working
Talking	Sleeping
Doing sport	Cooking
Watching TV	Family meal

service and also to determine best time for interacting). In this section, activities are proposed as a human state per room for each time quantum. In order to classify occupant activities, two categories have been defined: (1) activities with possibility to interact and interrupt the occupants (hereafter referred to as activities (1)) and (2) when it is inconvenient for the occupants to be prompted for interaction (activities (2)) (see Table 1). In this case, the interaction time with the occupants could be recognized and utilized in more advanced applications as explained earlier. The proposed system in this case consists of interconnected Bayesian networks which can be seen in Figure 4. Within the available sensors in the apartment, 4 Bayesian networks have been built for each activity (family meal, cooking, sleeping, and working) and a global one to adapt an estimation result, similar to what has been analyzed in occupancy estimation.

(1) Family meal is related to time of the day, the kitchen motion detector, and the cooking activity

(2) Cooking is related mostly to cooking time and use of the oven

(3) Sleeping is mostly related to the slot time and motion detector in a parent room

(4) Working is usually done in the office with consuming power and motion detector in the office

5.4. Case Study 3: Absence/Presence Estimation. This case study is similar to case study 2 in the sense that it also deals with the residential built environment. However, because of
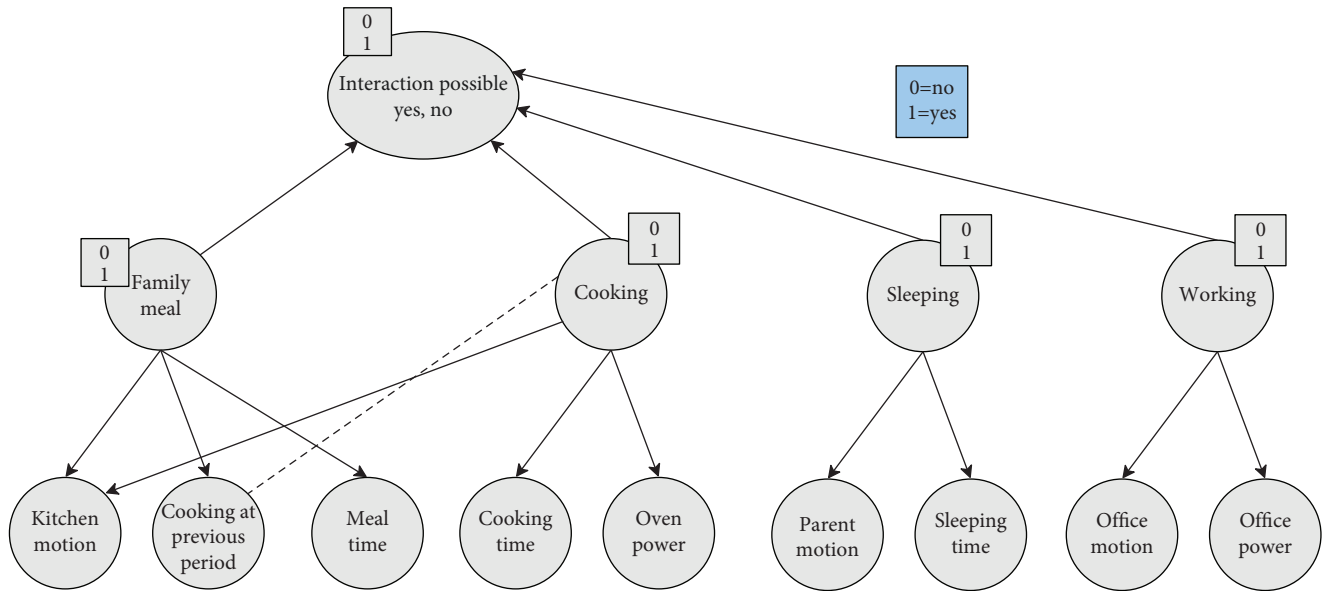
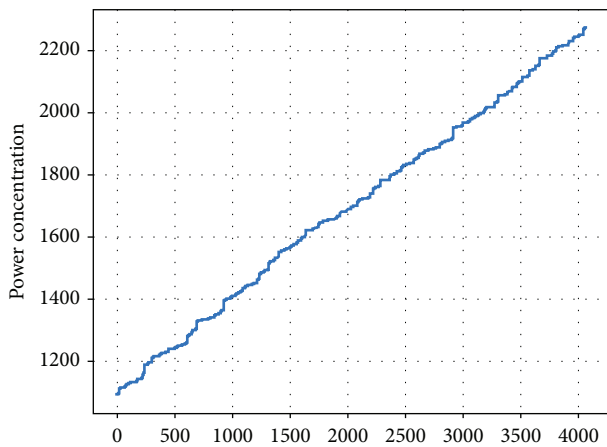FIGURE 4: Bayesian network structure case study (activities).

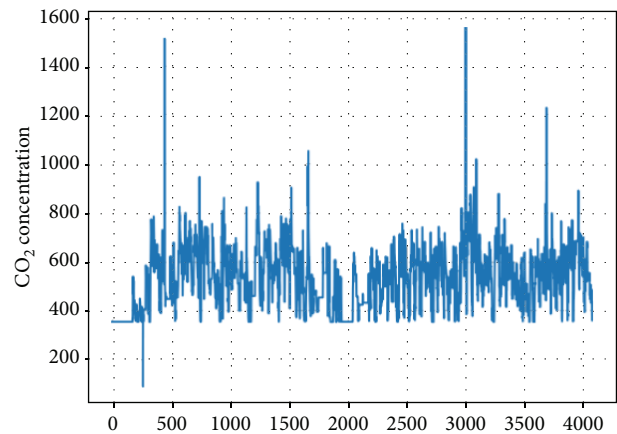FIGURE 5: Accumulated power consumption (kWh) from 01/06/2016 to 01/12/2016.

FIGURE 6: CO_2 concentration from 01/06/2016 to 01/12/2016.

limited sensing, it presents a more challenging estimation problem than case study 2. It is a multizone and multiactivity house with limited number of sensors located only in the living room. Here, the focus is on binary estimation of occupancy and not on activity estimation. The target from this application is to provide a presence profile over the day to improve the control part of the energy management system.

Historical consumption has been extracted from a database, in the absence of feedback from occupants. A Bayesian network structure of occupancy estimation has been built, with two levels for occupancy: absence/presence. According to [20], the most relevant sensors for occupancy estimation are as follows:

(i) Total energy consumption from smart meter sensors

(ii) CO_2 concentration

(iii) Total water flow consumption

Water flow sensors give the volume consumed during each quantum of time (i.e., 30 minutes). In order to collect the required information for conditional probability tables, historical data has been analyzed for power consumption, water flow consumption, and CO_2 concentration. Figures 5 and 6 present, respectively, the historical data during six months.

Figure 5 shows kWh data which has been extracted from the smart meter. It reads the energy consumed every 30 minutes. Power consumption has been calculated to be used in the occupant model. Similar to power consumption, water flow consumption has been extracted from the accumulated consumption in (L/h) (the liters of water consumed each hour).

A Bayesian network structure and probability tables have been built using these main sensors, which can be seen in Figure 7. Due to the goal of estimating absence/presence of occupants, two different cases have been considered: low
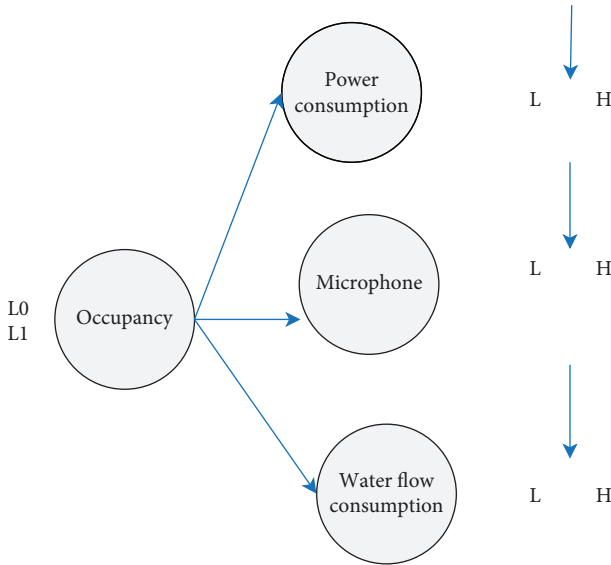
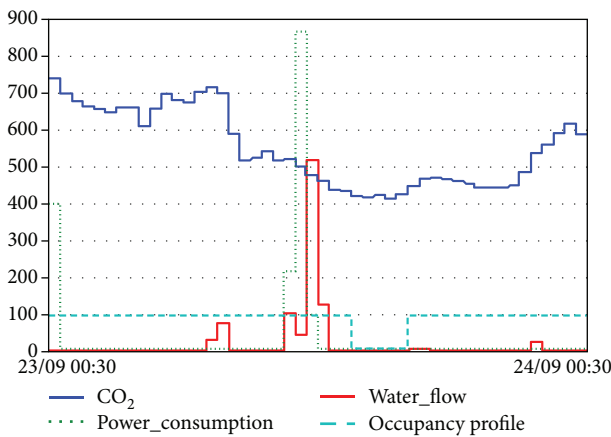FIGURE 7: Bayesian network structure case for Enervalis.

FIGURE 8: Generate occupancy profile from historical data of CO_2, water consumption, power consumption, and expert knowledge during one day and time quantum = 30 minutes.

consuming and high consuming or L and H, respectively. It gives a probability table with 6 values. In a similar way, the probability tables of water consumption and CO_2 concentration have been defined.

The process for calculating the conditional probabilities in case of using historical data starts by analyzing the data each day, which can be seen in Figure 8 to choose the trust days (days which do not exhibit a mixed presence profile) and remove the rest. The presence profile has been generated according to the values of data sensors. Expert knowledge is used in case of no consistency in the occupancy profile, before computing the conditional probability, i.e., the conditional probability value of consuming water during occupant presence, from a trust day in Figure 8, is equal to 13/43 or the number of samples where water was consumed divided by the number of samples of occupant presence. Different trust days have been considered to calculate the conditional probability table.

TABLE 2: Estimation results from the Bayesian network.

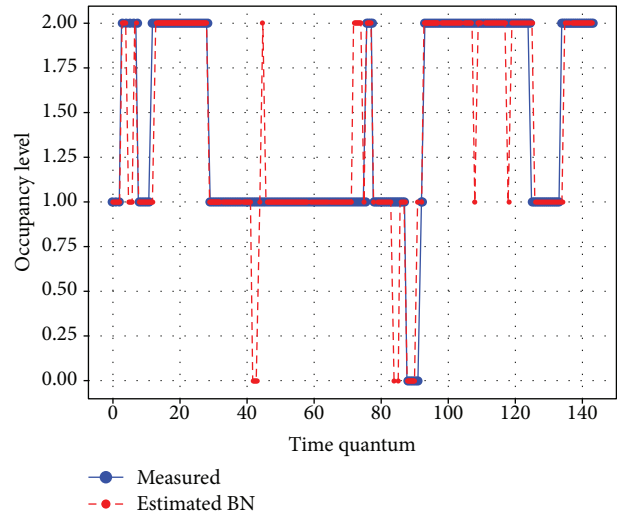| Classes	Average error of BN	Accuracy of BN	Number of supports
Class 1	97.2%	0.001	400
Class 2	84%	0.21	170
Class 3	79.6%	0.53	54
Average	91.3%	0.09	624

FIGURE 9: Occupancy estimation from a Bayesian network, time quantum = 30 minutes in Rue Cuvier apartment.

Where sufficient historic data is available, it is straightforward to determine these days.

6. Results

6.1. Case Study 1. A data set covering 10 days from 04/05/2015 to 13/05/2015 has been used for building a Bayesian network structure. Figure 2 shows the results obtained from the Bayesian network for three levels and 3 main features. Both actual and estimated occupancy profiles have been plotted in a graph with the number of occupants and time relations (quantum time was 30 minutes). The accuracy achieved from Bayesian network was 91% (number of correctly estimated points divided by the total number of points), and the average error was 0.09 persons. Table 2 represents the average error values for each class of estimation, while "support" in Table 2 indicates the number of events (sensor data each quantum time) in each class and average support indicates the sum of all events in the 3 classes.

The office H358 was the first initial step for analyzing the knowledge-based approach for occupancy estimation, where all the required information was available in this office.

6.2. Case Study 2. A data set covering the year 2016 has been used for building a Bayesian network structure. Data for validation were not easy to obtain from the occupants; only 3 days to test the estimation model have been collected. Figure 9 shows the results obtained for 3 days from
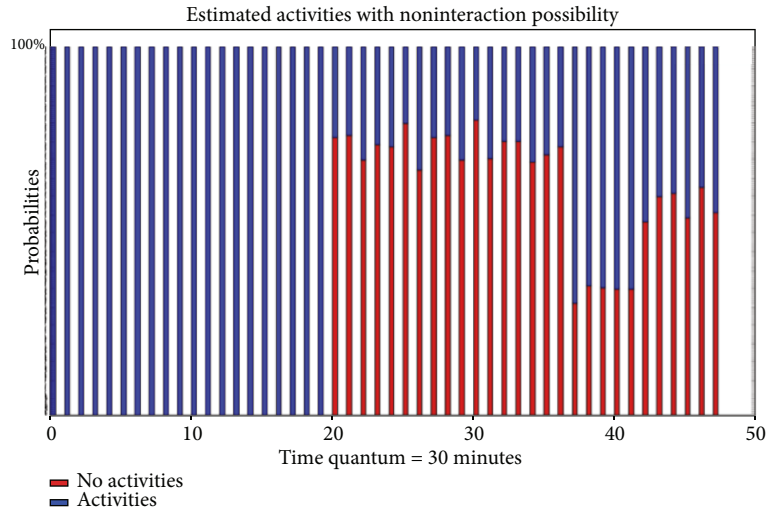
Figure 10: Activities distribution every 30 minutes during one year. It gives the probability with possible interaction with the user (no activities) in red color of each bar and the probability with no possible interaction (of activities) in blue color for each time quantum.
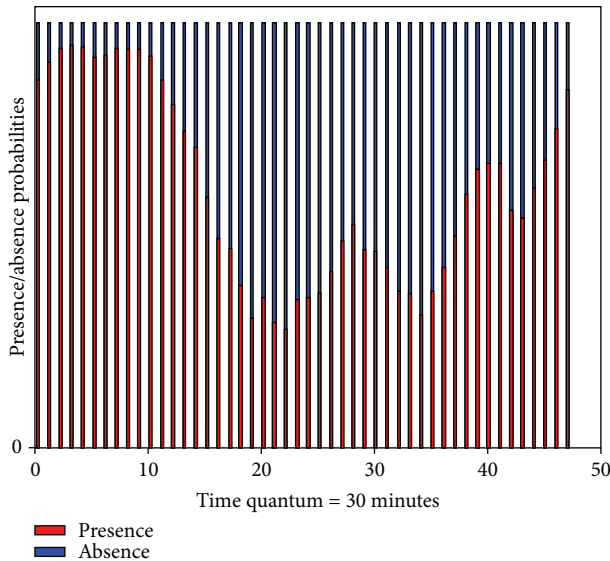

Figure 11: Occupancy estimation profile during one year over 48 half hours. It gives the probability of absence in blue color and the probability of presence in red color for each time quantum.

Table 3: Knowledge-based occupancy model applications.

Case study	Multizone	Multiactivity	Sensing and feedback	Validation
Case study 1	No	No	+ + +	Yes
Case study 2	Yes	Yes	+ +	Yes
Case study 3	Yes	Yes	-	No

estimation of 4 activities (sleeping, cooking, eating, and working) during 3 days.

The current framework provides an excellent trade-off for activity recognition in a residential environment while not requiring prohibitively expensive sensor data. In doing so, occupancy and activity estimation is not limited to the use of cameras anymore, thereby marking an advance on most current research [11] which usually requires cameras to cover the whole observation area.

6.3. Case Study 3. A data set covering six months from 01/06/2016 has been used to build the Bayesian network with two levels of occupancy, absence/presence. Finally, the occupancy presence profile over the day has been generated, to be integrated into the control system, which can be seen in Figure 11. It gives the probability of absence in blue color and the probability of presence in red color for each time quantum, i.e., at time 00:00, the probability of presence is 90% (red part of the bar).

Validating results directly is one of the most challenging aspects of real applications. In this case, it was because of the difficulty in contacting the end users. Consequently, building occupants could not be involved in the energy system process with frequent interruptions. In addition, only limited sensing is available because of privacy concerns. Thus, both occupancy estimation and validation are challenging tasks. The solution, in this case, was to collect enough data to build the estimation model.

In this section, three applications have been discussed (see Table 3), and the Bayesian networks with different
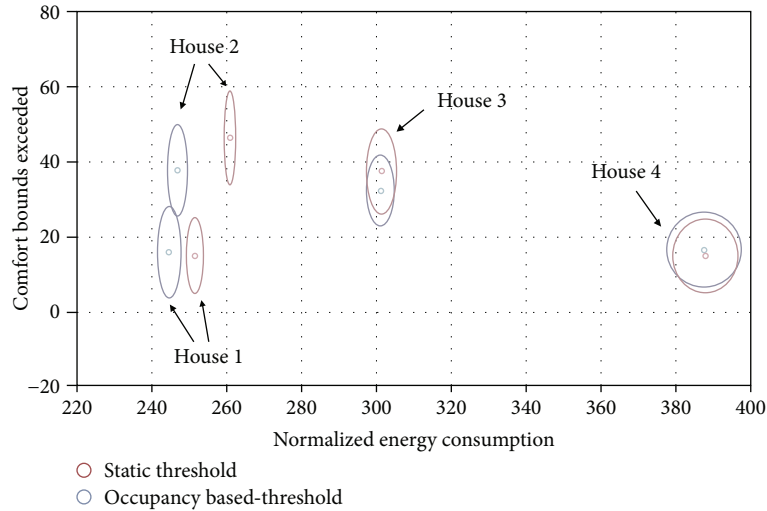
11/06/2016 to 13/06/2016 which were holidays and maximum number of occupants is 4 while minimum is 1. Both actual and estimated occupancy profiles have been plotted in a graph with the number of occupants and relation to time (time quantum was 30 minutes). The accuracy achieved from the Bayesian network was 82% (number of correctly estimated points divided by the total number of points), and the average error is 0.17 persons.

During one year, an average daily activity profile has been extracted over 48 half hours, which can be seen in Figure 10. It gives the probability with possible interaction with the user (no activities) in red color of each bar and the probability with no possible interaction (of activities) in blue color for each time quantum (i.e., half an hour). An average error of 0.24 per activity has been achieved by

FIGURE 12: Impact of occupancy estimation on the trade-off between occupant discomfort (number of hot water draws where consumption temperate fell below 45°C) and energy consumption (number of annual reheat cycles) for four simulated houses.

structures have been applied as the solution to the problem of occupancy and activity estimation. The different structures are due to differences in available sensors, knowledge, and the complexity of the studied area.

7. Energy-Efficient Hot Water Production Using Occupancy Estimates

Typically, hot water production systems consume energy due to regular reheat cycles following a naive rule-based controller.

Here t refers to the control action of the heating element at time t. When it is set to 1, it reheats the storage vessel and remains idle when the mode is 0. This decision is made based on temperature sensor information from the storage vessel T_s which is compared against a threshold $T_{tg} - \Delta T$, where T_{tg} is the temperature and the vessel is reheated to during a reheat cycle, and ΔT is the temperature and the storage vessel is allowed to fall relative to T_{tg} before a reheat cycle is initiated. This behavior can be optimized by making the reheat behavior dependent on the remaining energy content in the storage vessel and the (predicted) behavior of the human occupant.

Here W_t is the hot water volume left in the vessel, defined here as the amount above 45°C. Since the energy content in the storage vessel is not observable directly, W_t can only be approximated using a model of the storage vessel. This model can be built using either thermodynamics knowledge of the storage vessel or can be learned directly from sensor data [21]. The model then defines the remaining volume of hot water in the storage vessel which is compared against $\Sigma_{k=i}^{i+n} W_k$, the amount of hot water predicted to be consumed over the next n time steps. Furthermore, a safety margin, M, is introduced in the control formulation to account for unexpected draws caused by stochastic human behavior and is usually considered a fixed threshold-based value.

This greedy approach to reheating the storage vessel belongs to the family of just-in-time control strategies where both late heating and early heating are penalized [22].

In doing so, it improves the energy efficiency of hot water production while maintaining occupant comfort. This strategy is demonstrably optimal for climate agnostic heating elements such as electric resistance or gas boilers, given no additional knowledge about human behavior. For heat pumps, where ambient temperature affects the efficiency, it performs suboptimally (since an earlier reheat cycle might benefit from higher ambient temperature) thereby necessitating more sophisticated optimization strategies such as meta-heuristics for planning [23]. In this work, we consider an electric resistance heater simulated with hot water draws from four representative households.

Stochasticity in human behavior means that, over the long run, there are violations of occupant comfort which lead to water consumption temperatures dropping below the specified limits occasionally. This is mostly because predicting hot water demand accurately is extremely difficult, and sometimes the demand exceeds both the demand prediction and the safety margin, M.

We demonstrate that by making M a function of the predicted occupancy, the efficiency-comfort trade-off can be improved upon. This new threshold takes the following form: $M_t = O$, where O_t is the occupancy estimate at time t and is a probability and then replaces in the original control formulation used to determine $u1$ at every time instant. Intuitively, when the occupancy estimate is low, the required safety margin can likewise be made lower and vice versa. This essentially decomposes the occupant behavior into two distinct streams: one represents the expected occupant consumption at any time (which is based on historical trends) and the second represents the possibility for an unexpected draw (which is correlated with the occupancy estimate) [24].

Figure 12 shows that for four simulated houses with different occupancy and demand patterns, 2 houses (houses

1 and 2) gained substantially from using an occupancy estimate. The prediction error for hot water demand W_t was between 25% and 50%. The fixed safety margin, M, as defined above, was varied between 10 liters and 50 liters. For each chosen value of M, this was further modulated with an occupancy estimate O_t for controlling the threshold dynamically. The results of this sensitivity analysis are represented by the ellipses in Figure 12. The comfort bound exceeded index is defined as the amount of occurrences where the occupant drew hot water at a temperature below 45°C (normalized to an entire year). The number of reheat cycles is used as a proxy for the energy consumed to reheat the storage vessel on an annual basis.

It is evident from the figure that including an occupancy estimate in the control problem can reduce the number of reheat cycles by an additional 5% over the control strategy defined with a static threshold. Moreover, this strategy was around 25% more efficient than the rule-based control presented earlier for all the considered houses. These gains are however a strong function of occupant behavior. Occupancy estimates have an important modulating effect on the energy consumption patterns of the vessel. More concretely, for the houses that show negligible efficiency gains (houses 3 and 4), the estimated occupancy values were usually high which meant the modulation did not have much impact on the value of M (the average annual occupancy prediction was above 0.9 for these cases). For houses 1 and 2, on the other hand, the average occupancy prediction was closer to 0.6, which meant that the controller could set back the reheat cycles successfully during these times leading to lower energy consumption or achieve better occupant comfort.

The energy efficiency improvements in hot water production derive from making the reheat cycles subordinate to occupant behavior and presence/absence. In this, a prediction of occupant demand provides most of the savings. However, this can come at the cost of occupant comfort. Using occupancy estimates provides a way to help solve this problem. Alternatively, it can be used to further improve energy efficiency while maintaining the same level of occupant comfort.

8. Conclusion

Knowledge-based approach with nonsupervised learning procedure has been applied using the Bayesian network. The proposed method can generalize well because the structure is not assumed; rather, it is discovered from the questionnaire and can therefore be extended to any room with expert knowledge. Additionally, three applications have been discussed with the Bayesian networks built using sensor data and knowledge coming, respectively, from observation and questionnaire. The structure of these Bayesian networks differs according to the application environment. The level of complexity increased from one application to the other because of missing occupancy and activity feedback, the information available from the questionnaire, or the fewer installed sensors.

(1) Case study 1, with an office environment, is the first step and perhaps the easiest environment to analyze and evaluate the knowledge-based approach due to availability of all the required sensors, occupant feedback, observation (video camera, microphone, and power consumption), and validation in one zone area. Occupancy estimation using the Bayesian network model knowledge gave excellent performance with an average estimation error for the office context of 0.08 occupants over a ten-day test period

(2) Case study 2, a residential environment with multiple zones and many activities, proved to be more complex due to the interaction between the rooms and the requirement to estimate activities. On the other hand, collecting data from a high number of sensors in the apartment and the cooperation of the user for replying to the questions during the design and validation period facilitated the construction of the required models. An average error of 0.17 for the number of occupants and 0.24 for activity estimation has been achieved

(3) Case study 3 presented a more complex real-world application. The setting was a multizone and multiactivity house with very limited number of sensors and no possibility for obtaining feedback from the occupants. In order to cope with this case, 6 months of historic sensor data have been analyzed to extract the required knowledge for building an occupancy estimation model. Because of lack of communication with the building occupants, only limited validation could be performed in this case

Experiments to integrate occupancy estimation and hot water production control have also been carried out. These conclude that energy efficiency can be effectively increased by up to 5% when occupancy estimates are incorporated into the optimal control formulation.

This paper presented important results in estimating occupancy and reducing energy demand and, in doing so, it contributes to the growing body of literature illustrating the energy-occupancy nexus.

Future directions of the work include using these occupancy estimates to not just improve energy efficiency but also perform active demand side management. This has the potential to reduce energy costs for building users as well as providing support to the broader energy grid.

Data Availability

The data available on request through a data access committee, a centralized database with a web application for retrieving data from different sources continuously is available. To have access of the data, Vesta Company gives the right to this access, https://www.vestasystem.fr.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.
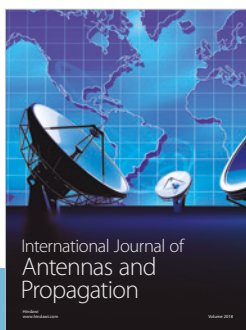
Acknowledgments

References

[1] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, 2008.

[2] H. Kazmi, F. Mehmood, S. Lodeweyckx, and J. Driesen, "Gigawatt-hour scale savings on a budget of zero: deep reinforcement learning based optimal control of hot water systems," *Energy*, vol. 144, pp. 159–168, 2018.

[3] M. Amayri, S. Ploix, and Q.-D. Ngod, "Estimating occupancy from measurement and knowledge with Bayesian networks," in *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, December 2016.

[4] M. Amayri, S. Ploix, P. Reignier, and B. Sanghamitra, "Towards interactive learning for occupancy estimation," in *ICAI16 - International Conference on Artificial Intelligence.*, Las Vegas, NV, USA, 2016.

[5] H. Kazmi, S. D'Oca, C. Delmastro, S. Lodeweyckx, and S. P. Corgnati, "Generalizable occupant-driven optimization model for domestic hot water production in NZEB," *Applied Energy*, vol. 175, pp. 1–15, 2016.

[6] R. V. Andersen, J. Toftum, K. K. Andersen, and B. W. Olesen, "Survey of occupant behaviour and control of indoor environment in Danish dwellings," *Energy and Buildings*, vol. 41, no. 1, pp. 11–16, 2009.

[7] B. Mathieu, "Influence du comportement de loccupant sur la performance energetique du batiment," PhD thesis, Universita de Toulouse, 2014.

[8] J. Langevin, j. Wen, and G. Patric, "Simulating the human-building interaction: development and validation of an agent-based model of office occupant behaviours," in *Proc. Windsor Conference*, Windsor, England, 2014.

[9] N. Oliver, A. Garg, and E. Horvitz, "Layered representations for learning and inferring office activity from multiple sensory channels," *Computer Vision and Image Understanding*, vol. 96, no. 2, pp. 163–180, 2004.

[10] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng, "Occupancy-driven energy management for smart building automation," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building - BuildSys '10*, pp. 1–6, Zurich, Switzerland, November 2010.

[11] M. Milenkovic and O. Amft, "Recognizing energy-related activities using sensors commonly installed in office buildings," *Procedia Computer Science*, vol. 19, pp. 669–677, 2013.

[12] J. Jung, I. Yoon, S. Lee, and J. Paik, "Object detection and tracking-based camera calibration for normalized human height estimation," *Journal of Sensors*, vol. 2016, Article ID 8347841, 9 pages, 2016.

[13] T. A. Nguyen and M. Aiello, "Beyond indoor presence monitoring with simple sensors," in *Proceedings of the 2nd International Conference on Pervasive Embedded Computing and Communication Systems*, pp. 5–14, Rome, Italy, 2012.

[14] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.

[15] P. Gargallo, P. Sturm, and S. Pujades, "An occupancy–depth generative model of multi-view images," in *Computer Vision – ACCV 2007*, Y. Yagi, S. B. Kang, I. S. Kweon, and H. Zha, Eds., vol. 4844 of Lecture Notes in Computer Science, pp. 373–383, Springer, Berlin, Heidelberg, 2007.

[16] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, pp. 28–39, 2016.

[17] J. A. B. Franz Pernkopf, *Efficient Heuristics for Discriminative Structure Learning of Bayesian Network Classifiers*, Mach ne Learn ng, 2010.

[18] A. Anka and A. Panda, "pgmpy: probabilistic graphical models using python," in *Proceedings of the 14th Python in Science Conference*, Austin, TX, USA, 2015.

[19] M. Amayri and S. Ploix, "Decision tree and parametrized classifier for estimating occupancy in energy management," in *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*, pp. 397–402, Thessaloniki, Greece, April 2018.

[20] M. Amayri, A. Arora, S. Ploix, S. Bandhyopadhyay, Q. D. Ngo, and V. R. Badarla, "Estimating occupancy in heterogeneous sensor environment," *Energy and Buildings*, vol. 129, pp. 46–58, 2016.

[21] H. Kazmi, F. Mehmood, and M. Amayri, "14th International Symposium on Pervasive Systems, Algorithms and Networks 2017 11th International Conference on Frontier of Computer Science and Technology 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC)," *Smart Home Futures: Algorithmic Challenges and Opportunities*, vol. 2017, pp. 441–448, 2017.

[22] W. Kubiak, "Minimizing variation of production rates in just-in-time systems: a survey," *European Journal of Operational Research*, vol. 66, no. 3, pp. 259–271, 1993.

[23] S. VoÃ, S. Martello, I. Osman, and C. Roucairo, "Meta-heuristics: Advances and trends in local search paradigms for optimization," in Springer Science & Business Media, 2012.

[24] M. Amayri, H. Kazimi, Q.-D. Ngo, and S. Ploix, "Estimating occupancy in residential context using Bayesian networks for energy management," in *ICMLPC 2018: International Conference on Machine Learning for Prediction and Control*, vol. 12, London, UK, February 2018.