Building and Environment 107 (2016) 1–9

Contents lists available at ScienceDirect

Building and Environment

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

Development of an occupancy prediction model using indoor environmental data based on machine learning techniques

Seung Ho Ryu [a, b], Hyeun Jun Moon [b, *]

[a] *EaaS Project, Energy Convergence & Data TF, Convergence Laboratory, Institute of Convergence Technology, KT Corporation, Seoul, 06763, Republic of Korea*

[b] *Department of Architectural Engineering, Dankook University, Yongin, 16890, Republic of Korea*

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form
31 May 2016
Accepted 30 June 2016
Available online 4 July 2016

Keywords:
Occupant behavior
Occupancy prediction model
Indoor environmental data
Decision tree model
Hidden Markov model

ABSTRACT

Occupant presence and behavior in buildings have significant impact on space heating, cooling and ventilation demand, energy consumption of lighting and appliances, and building controls. For this reason, there is a growing interest on modeling occupant behavior, especially occupancy information. An occupancy prediction model based on an indirect approach using indoor environmental data is important due to privacy concerns and inaccurate measurements associated with the direct approach using cameras and motion sensors. However, such an indirect-approach-based occupancy prediction model has not yet fully discussed in building simulation domain. To tackle these issues, this study aims to develop an indoor environmental data-driven model for occupancy prediction using machine learning techniques.

The experiments in the Building Integrated Control Test-bed (BICT) at Dankook University was conducted to collect the ground truth occupancy profiles, indoor and outdoor CO_2 concentrations and electricity consumptions of lighting systems and appliances for a data mining study. The results show that the proposed indoor environmental data-driven models for occupancy prediction using the decision tree and hidden Markov model (HMM) algorithms are well suited to account for occupancy detection at the current state and occupancy prediction at the future state, respectively.

1. Introduction

A fundamental objective of energy efficient buildings is to facilitate a comfortable environment for occupants while maintaining minimal energy consumption. Information regarding occupancy in buildings is a key component to achieve this task. Occupant presence and behavior in buildings have significant impact on space heating, cooling and ventilation demand, energy consumption of lighting and appliances, and building controls [1–6]. The following studies support this argument. Hong and Lin [7] presented the simulation study results for a typical single-occupancy office compared to the standard or reference work style. According to the study, an austere work style consumes up to 50% less energy, while a wasteful work style consumes up to 90% more energy. Masoso and Grobler [8] presented the results of an energy audit study on six randomly selected commercial buildings in South Africa. The results show that more than 50% of energy is used during unoccupied weekday hours rather than during official working hours. 19–28% of the building energy is consumed during unoccupied weekend hours.

For this reason, there is a growing interest on modeling occupant behavior, especially occupancy information. Numerous methods have been developed to study occupant behavior at both individual and group levels. These methods can be categorized as "direct approach" and "indirect approach" [9].

The direct approach is a method based on direct detection of occupants using direct positioning technologies such as passive infrared (PIR) motion detector, video camera and radio-frequency identification (RFID). These technologies were used to learn occupant behavior and assist in building system controls. Lymberopoulos et al. [10] proposed a system, BehaviorScope, for interpreting occupant's activity patterns using camera networks. The presented system can only detect occupant's activity patterns; it cannot predict occupant's activity patterns. Shih [11] also proposed a monitoring system enabling continuous occupancy detection and tracking based on image-based depth sensors and

* Corresponding author.
E-mail address: honjunmoon@gmail.com (H.J. Moon).

http://dx.doi.org/10.1016/j.buildenv.2016.06.039
0360-1323/

programmable pan-tilt-zoom (PTZ) cameras using a support vector machine (SVM) model. The results show that the proposed SVM-based observation measurement provides more reliable tracking performance. The presented system cannot predict occupancy as well. Duarte et al. [12] presented the analysis of occupancy sensor data for a large commercial, multi-tenant office building using data mining techniques to derive occupancy diversity factors for private offices, open offices, hallways, conference rooms, break rooms, and restrooms in order to better set energy simulation parameters. Erickson et al. [13] proposed an occupancy prediction model based on real-time occupancy data obtained from a wireless sensor network, Smart Camera Occupancy Position Estimation System (SCOPES) using the Markov Chain model.

All of the above research rely on gathering and analysis of occupancy data directly collected in existing buildings. Although direct measurement methods are effective in detecting building occupant behavior, privacy can be a major concern for deploying these technologies in the real world. Moreover, PIR motion detectors only provide the presence or absence of occupant information rather than the number of occupants which is highly useful for building controls [13]. For these reasons, recently, there is a growing interest in non-intrusive “indirect” occupant measurement methods which use energy consumption and environmental sensor data mining. Zhao et al. [14] proposed an indirect practical data mining approach using office appliance power consumption data to learn about occupancy. However, the presented data-driven model can only detect an occupancy at the current state. Arora et al. [15] also proposed an indirect practical data mining approach using non-intrusive measurements such as motion detection, power consumption and CO_2 concentration to estimate the number of occupants in a zone. However, the presented data-driven model can only detect occupancy at the current state. Moreover, they utilized motion sensors as well as indirect measurement methods for modeling.

In summary, on the application for real office buildings, occupancy prediction based on the non-intrusive indirect approach using indoor environmental data including environmental factors and energy consumption factor is more suitable than direct measurement approaches due to privacy concerns and inaccurate measurements associated with the direct approach. However, an occupancy prediction model based on indoor environmental data has not yet fully discussed in building simulation domain. Hence, this study focuses on developing an occupancy prediction model based on indirect sensing data such as energy consumption and environmental data using machine learning techniques. In this study, the combination of the act of living or staying in a particular place and the number of occupants who are in a particular building or room at one time will be articulated in terms of “occupancy.”

2. Methodology

The study to develop models for occupancy prediction is organized into two steps. First step is development of a model to detect occupancy at the current state based on indoor environmental data such as CO_2 concentration, energy consumption of the lighting system and appliances obtained from sensing networks in buildings using machine learning techniques, especially the decision tree algorithm. Second step is development of a model for occupancy prediction at a future state based on indoor environmental data and estimated occupancy data obtained from the developed occupancy detection model in Step 1 using machine learning techniques, especially the hidden Markov model (HMM).

2.1. Experiment setup

To develop models for occupancy detection and prediction based on indoor environment data, experiments were performed in the Building Integrated Control Test-bed (BICT) at Dankook University, Korea. This test-bed is used to integrate sensing and actuating technologies to achieve energy efficiency while providing a comfortable environment for occupants [16]. BICT includes sensors for various environmental parameters such as temperature, humidity, CO_2, illuminance, and motion. In addition to sensing networks, several meters are measuring electricity consumption of the HVAC systems, lighting systems, and appliances. A web-based camera is also included in the sensing networks to track the exact occupancy in the test-bed. Fig. 1 shows the BICT used in this study.

The experiments aimed to collect ground truth occupancy profiles, indoor environmental data and electricity consumptions of lighting systems and appliances in order to perform a data mining study. In machine learning, the term “ground truth” refers to the accuracy of the training set's classification for supervised learning techniques. Data collection in the BICT for this study took place during seven non-consecutive weekdays (2015.10.21, 10.30, 11.03, 11.04, 11.06, 11.11, and 11.13). Occupancy, environmental factors, and electricity consumption were recorded at 1-min intervals during those days. The environmental factors included indoor and outdoor CO_2 concentration. The electricity consumption included the lighting system and appliances. During the experiments for collection of ground truth occupancy profiles, occupancy schedule was not defined and researchers that were involved in the experiments worked and studied freely in the test-bed. Figs. 2 and 3 show the ground truth occupancy profiles, environmental data and metered electricity consumption data obtained in the BICT and used for the data mining study.

2.2. Decision tree algorithm

The Classification and Regression Tree (CART) algorithm, one of the decision tree algorithms, was selected to detect occupancy using indoor environmental and energy consumption data. The decision tree is a branched graphical classification model. Easy interpretation of the graphical representation of classification results is one of the advantages of the decision tree. Decision tree models a set of data into various predefined classes, categorization and generalization of a given dataset. The goal of a decision tree is to create a classification model that predicts the value of a target attribute (response) based on several input attributes (predictors). Each interior node (leaf node) of tree corresponds to one of the predictors and the number of edges (branches) of a nominal

Fig. 1. Building Integrated Control Test-bed (BICT) used in the development of the occupancy prediction model.