Automation in Construction 94 (2018) 233–243

Contents lists available at ScienceDirect

Automation in Construction

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

Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings

Wei Wang[a,b], Jiayu Chen[a,c,*], Tianzhen Hong[b,**]

[a] Department of Architecture and Civil Engineering, City University of Hong Kong, Y6621, AC1, Tat Chee Ave, Kowloon, Hong Kong
[b] Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[c] Architecture and Civil Engineering Research Centre, City University of Hong Kong Shenzhen Research Institute, Shenzhen, Guang Dong 518057, China

ARTICLE INFO

Keywords:
Occupancy prediction
Environmental sensing
Wi-Fi sensing
Data fusion
Machine learning

ABSTRACT

Occupancy information is crucial to building facility design, operation, and energy efficiency. Many studies propose the use of environmental sensors (such as carbon dioxide, air temperature, and relative humidity sensors) and radio-frequency sensors (Wi-Fi networks) to monitor, assess, and predict occupancy information for buildings. As many methods have been developed and a variety of sensory data sources are available, establishing a proper selection of model and data source is critical to the successful implementation of occupancy prediction systems. This study compared three popular machine learning algorithms, including k-nearest neighbors (kNN), support vector machine (SVM), and artificial neural network (ANN), combined with three data sources, including environmental data, Wi-Fi data, and fused data, to optimize the occupancy models' performance in various scenarios. Three error measurement metrics, the mean average error (MAE), mean average percentage error (MAPE), and root mean squared error (RMSE), have been employed to compare the models' accuracies. Examined with an on-site experiment, the results suggest that the ANN-based model with fused data has the best performance, while the SVM model is more suitable with Wi-Fi data. The results also indicate that, comparing with independent data sources, the fused data set does not necessarily improve model accuracy but shows a better robustness for occupancy prediction.

1. Introduction

Occupancy information refers to building occupants' presence, movement, and behaviors that significantly impact a building's energy consumption [1–3]. Similar to traditional approaches of improving envelopes, shapes, and materials design of a building, understanding and utilizing occupancy information also can optimize a building's energy efficiency [4,5]. With occupancy information, both centralized and decentralized facilities, such as air-conditioning systems, can optimize their operations and avoid energy waste [6]. Wang and Shao conducted a 30-day experiment in a library and applied mining approaches to model the occupancy patterns; the results suggest that 26.1% of the building's energy cost could be saved through occupancy-based facility control [7]. Therefore, researchers suggest that accurate, high-resolution occupancy information can not only improve the performances of heating, ventilation and air conditioning systems (HVAC) but also guide the design of building energy modeling [8–11].

To acquire occupancy information, researchers propose employing environmental sensors, such as light sensors, CO_2 sensors, temperature sensors, and relative humidity (RH) sensors, using methodologies developed to assess occupancy based on the variation of indoor environment parameters [12–14]. At the same time, as Wi-Fi networks are widely installed in modern buildings and play an important role in communication between occupants and building services, some researchers suggest using Wi-Fi signals as a tool for occupancy detection [7,15]. Recently, researchers compared Wi-Fi based detection with environment-based detection approaches and suggested that Wi-Fi detection is more accurate. For instance, Ouf et al. investigated both Wi-Fi connections and CO_2 concentration-based occupancy assessment using Pearson's correlations and their results suggest that Wi-Fi counting is more accurate and reliable [16]. Similarly, Wang et al. employed Wi-Fi probes to actively interpret Wi-Fi signal requests and responses, and also achieved higher detection accuracy than CO_2-based detections [17]. Inspired by Ouf et al.'s findings, this study proposes the fusion of Wi-Fi connections and environmental parameters to further improve the accuracy of occupancy prediction. Due to the heterogeneous nature

* Correspondence to: J. Chen, Department of Architecture and Civil Engineering, City University of Hong Kong, Y6621, AC1, Tat Chee Ave, Kowloon, Hong Kong.
** Corresponding author.
E-mail addresses: jiaychen@cityu.edu.hk (J. Chen), thong@lbl.gov (T. Hong).

https://doi.org/10.1016/j.autcon.2018.07.007
Received 15 March 2018; Received in revised form 22 June 2018; Accepted 10 July 2018
0926-5805/

of data sources, the study utilized popular machine learning algorithms to assess occupancy through various features. With the results from an on-site experiment, this study compares the performance of these algorithms, specifically ANN, kNN, and SVM, with the combination of fused and independent data sets. The conclusions of this work can provide guidelines for selecting the appropriate machine learning algorithms and data sets in the practical implementation of occupancy sensing systems.

This work is organized as follows: Section 2 reviews and discusses related works; Section 3 includes the methodology of the feature-based machine learning techniques for occupancy modeling; Section 4 introduces the experiment, ground truth, and assessment metrics; Sections 5 and 6 show the results and discuss occupancy profile features; and Section 7 presents the conclusions and research contributions of this work.

2. Literature review

2.1. Occupancy detection and prediction with sensors

Buildings consume energy to provide a comfortable thermal environment and air quality that satisfies residents' needs, so understanding occupants' needs and behaviors is the key to promoting building energy efficiency while maintaining service quality [18–20]. The International Energy Agency (IEA) Energy in Buildings and Communities (EBC) Programme Annex 66 regards occupant behavior as one of the most significant considerations for building and system design, and regards occupancy modeling and prediction as the premise of a responsible building management system [21]. Annex 66 highlighted major issues regarding occupancy sensing, modeling, and evaluation to encourage developers to promote building energy efficiency from the perspective of occupants.

For practical implementations of occupancy detection and prediction, researchers have proposed various models using various environmental sensors. Wang et al. proposed the use of CO_2 concentration to estimate occupancy dynamics and to allow demand-orientated ventilation [22,23]. Dodier et al. applied passive infrared sensors in two private offices and operated building systems with Bayesian probability theory [24].

In addition to CO_2 concentration-based method, the energy consumption of devices in buildings can also be a good indicator of occupancy. Diaz and Jimenez measured the power consumption of computers to imply occupancy information, comparing it to the CO_2 concentration method [25]. Further, in Ryu and Moon's study [26], the energy consumption of lighting and appliances was applied in machine learning techniques for occupancy prediction. Gul and Patidar proposed a study to understand energy consumption and occupancy in a multipurpose academic building and pointed out that occupancy patterns could help the management team to redesign control strategies for optimum energy performance in their buildings [27]. Amayri et al. also used four power consumption sensors to estimate occupancy and compared it with CO_2 concentration and other sensors by applying rule-based estimators [28]. Diao et al. [29] proposed the identification and classification of occupants' activities with direct energy consumption outcomes and energy time use data through k-modes clustering, probability neural networks, and inhomogeneous Markov chain modeling based on American Time Use Survey (ATUS) data. Yu et al. [30] employed k-Means clustering analysis to examine the effects of different behavior patterns on energy consumption and revealed that four identified behavior clusters have significant attributes with respect to some end-use loads, such as HVAC load.

Extending occupancy models based on single sensing sources, researchers proposed integrating multiple environmental parameters, such as indoor air temperature, RH, air pressure, CO_2 concentration, volatile organic compounds rate, and noise level, for better precision and accuracy. Pedersen et al. utilized sensors to collect all these parameters in an open test room and a three-room dorm and reported a detection accuracy as high as 98% for dichotomous occupancy status (occupied or vacant) [31]. Szczurek et al. concluded that combining CO_2 concentration, temperature, and RH can sufficiently detect occupancy within 60-minute data segmentation [32]. Yang et al. incorporated sound and motion sensing with environmental sensors and showed that accuracy can reach 88.74% and 86.5% in two respective in-lab experiments [33]. Yang and Becerik-Gerber integrated light, sound, motion, CO_2, temperature, RH, passive infrared sensors, and door switch sensing results, and applied an autoregressive moving average model, neural network, Markov chain, and logit regression to model occupancy profiles [34].

Recently, radio frequency networks have been presented as an alternative occupancy detection approach for their convenience and flexibility. For example, Li et al. used the radio frequency identification (RFID) system to measure and monitor occupants' spatial dynamics [10]. As Wi-Fi networks are available in most commercial buildings, researchers tend to explore the possibility of using Wi-Fi connections to count the number of occupants in network-covered spaces. Chen et al. demonstrated that Wi-Fi connection and disconnection events have a positive relationship with energy load variation [35]. Balaji et al. concluded that Wi-Fi connections could identify occupancy profiles with an 83% accuracy [36]. Wang and Chen utilized a Wi-Fi based indoor positioning system to obtain the meshed spatial distribution of occupants [37,38]. They simulated occupancy's impacts on building air-conditioning control systems and reported a 22% energy saving with a demand-based facility control. Later, they proposed an occupancy-based multi-zone outdoor air control model and reported a 23.6% ventilation energy saving during working days [39]. Implementing occupancy-driven lighting control based on Wi-Fi connections, Zou et al. demonstrated a 93.09% and an 80.27% of energy saving when compared with static scheduling and an infra-red-based lighting control schemes [40]. This study intends to combine both Wi-Fi network and environmental sensors to examine whether a fused data source is able to further improve the performance of occupancy prediction models.

2.2. Models for occupancy prediction

As occupancy information is a periodic time series, researchers have developed numerous time-dependent and non-linear models to analyze sensing results. These models can be categorized as Markov chain, artificial neural network (ANN), k-nearest neighbors (kNN), support vector machine (SVM), Classification and Regression Trees (CART), extreme learning machine (ELM), linear discriminant functions (LDA), and their variations. Peng et al. applied kNN and multi-learning processes to predict occupancy profiles based on historical records of office areas [41]. Szczurek et al. utilized a kNN algorithm with LDA and studied air temperature, RH, and CO_2 to classify occupancy patterns [32]. Lin et al. estimated the number of people in a crowd with SVM classification and visualized occupancy as head-like contour [42]. Ryu and Moon developed new ANN input features, including the first order shift, moving average, rate of change of CO_2 concentration, and indoor/outdoor CO_2 ratios to predict occupancy variations [26]. Jiang et al. applied ELM and CO_2 concentration to assess occupancy in an office room [43]. Yang and Becerik-Gerber installed light, sound, motion, CO_2, temperature, RH, passive infra-red, and door-switch sensors and cameras in three typical offices to collect cross-referenced experiment results; they then applied auto-regressive moving average, ANN, Markov chain, and logit regression algorithms to cross-validate occupancy profiles [34]. Similarly, Candanedo and Feldheim also evaluated a cross-platform data source (temperature, humidity, light, and CO_2) to predict occupancy with different statistical classification models (LDA, CART, and random forest) [12], obtaining an accuracy level of approximately 97% using only two of the environmental parameters with the LDA model for daily analysis. Chen et al. also proposed a fusion framework that increases estimation accuracy by 5–14% with these