Energy and Buildings 88 (2015) 395–408

Contents lists available at ScienceDirect

Energy and Buildings

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

# Occupancy schedules learning process through a data mining framework

Simona D'Oca [a,b], Tianzhen Hong [a,*]

[a] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[b] Polytechnic of Turin, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

ARTICLE INFO

*Article history:*
Received 20 August 2014
Received in revised form
19 November 2014
Accepted 22 November 2014
Available online 17 December 2014

*Keywords:*
Occupant behavior
Data mining
Occupancy schedule
Behavioral pattern
Office building
Building simulation

ABSTRACT

Building occupancy is a paramount factor in building energy simulations. Specifically, lighting, plug loads, HVAC equipment utilization, fresh air requirements and internal heat gain or loss greatly depends on the level of occupancy within a building. Developing the appropriate methodologies to describe and reproduce the intricate network responsible for human-building interactions are needed. Extrapolation of patterns from big data streams is a powerful analysis technique which will allow for a better understanding of energy usage in buildings. A three-step data mining framework is applied to discover occupancy patterns in office spaces. First, a data set of 16 offices with 10 min interval occupancy data, over a two year period is mined through a decision tree model which predicts the occupancy presence. Then a rule induction algorithm is used to learn a pruned set of rules on the results from the decision tree model. Finally, a cluster analysis is employed in order to obtain consistent patterns of occupancy schedules. The identified occupancy rules and schedules are representative as four archetypal working profiles that can be used as input to current building energy modeling programs, such as EnergyPlus or IDA-ICE, to investigate impact of occupant presence on design, operation and energy use in office buildings.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

One of the paramount efforts engineers, architects and policy-makers are currently facing is the need to deliver highly efficient buildings. In the roadmap toward net-zero energy buildings, office buildings play an important role as they represent approximately 17% of the energy used in the U.S. commercial building sector [1].

Several efforts have been made to accelerate the uptake of energy efficiency technologies in office buildings. While the driving factors of building energy performance such as climate, building envelope and building equipment are well recognized, the description of factors such as operation and maintenance, occupant behavior, and indoor environmental conditions are still over-simplified. Often building occupancy schedules are based upon generalized assumptions that hinge on standards, energy codes or rely on the experience of energy modelers. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Standard 90.1-2004 [2] provides standardized occupancy factors for different building types which can be used to design occupancy when actual schedules are unknown (Fig. 1). A daily profile, handled differently for weekend and weekdays, is composed of hourly values, each of which corresponds to a fraction of the occupancy peak load.

Nevertheless the stochastic nature of occupant behavior, the number of people that occupy a space and the duration occupied, is a non-trivial aspect to characterize. Literature studies have focused on the impact of occupancy presence scenarios on energy use in office buildings, with Burak Gunay et al. [3] providing a comprehensive and up-to-date critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices. In 2013 a study conducted by Duarte et al. [4] analyzed the occupancy sensors of a large commercial multi-tenant office building and showed up to 46% variation in occupancy patterns for the time of day, day of the week, holidays and months, when compared with the standardized occupancy schedules in ASHRAE Standard 90.1-2004 [2]. The discrepancy presented by Duarte et al. [4] may lead to the incorrect design of office building equipment and to system inefficiencies. Chang and Hong [5] demonstrated the stochastic nature of occupancy profiles was one of the driving factors behind the discrepancy between the measured and simulated energy consumption in buildings. Based on statistical analysis of measured lighting-switch data, Chang and Hong [5] proved the frequencies of occupants leaving their cubicles and the corresponding durations of absence had significant impact on the total energy use and

* Corresponding author. Tel.: +1 510 4867082; fax: +1 510 4864089.
*E-mail addresses:* thong@lbl.gov, hongtz68@yahoo.com (T. Hong).

http://dx.doi.org/10.1016/j.enbuild.2014.11.065
0378-7788/© 2014 Elsevier B.V. All rights reserved.

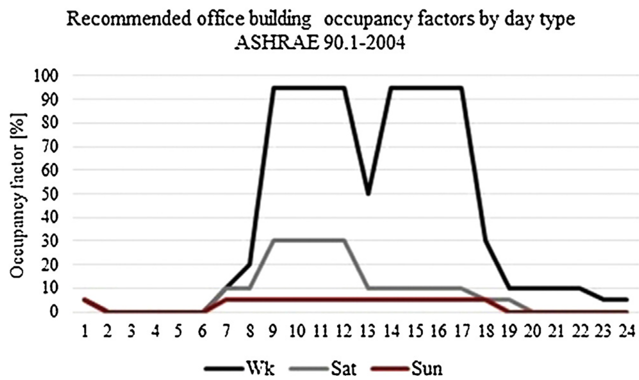


**Fig. 1.** Recommended office building occupancy factors [%] by day type, ASHRAE Standard 90.1-2004.

operational controls of the office building. Results from EnergyPlus simulations to evaluate the impact of occupant behavior on energy use of private offices with single occupancy [6], demonstrated that occupants with wasteful work-style consumed up to 90% more energy than standard users, while austerity work-style occupants used half of the energy of the standard occupants. Moreover, real-time estimation of occupancy in commercial buildings is largely treated with the aim to achieve better dynamic modeling results. However, it is a challenging task to develop reliable mathematical models of occupant presence due to the *stochastic* nature of human behavior [7].

Some stochastic models of the occupancy level of single offices have been proposed in the last decade within the scientific community [8–10]. Wang et al. [8] examined the statistical properties of occupancy in single person offices of a large office building in San Francisco and found that, while vacancy intervals could be treated as a constant over the day, occupancy intervals were more complex due to their varied distribution in time. Tabak and de Vries [9] proposed a model to predict the occurrence and the frequency of intermediate break activities during an office working day (i.e. walking to a printer/mailbox or using the bathroom). For each intermediate activity, a probabilistic formula was presented for use in office occupancy schedule designs. More recently, Sun et al. [10] developed a stochastic model, using a binomial distribution to represent the total number of occupants working overtime and an exponential distribution to represent the duration of overtime periods. Moreover, Stoppel and Leite [11] presented a probabilistic occupancy model simulating annual building occupancy rates based on frequency, duration and seasonality of occupants' long vacancy activities that can be further implemented into a building simulation model.

Additionally, there has been a gowning interest in agent-based models (ABM) to simulate patterns of human individual action and presence at the building level. Most notably, the Markov Chain method, provides a simulation approach to capture the movement process per occupant in the time and space dimensions of building models. The earliest ABM of occupant presence using a Markov Chain was proposed in 2005 by Yamaguchi et al. [12] in the development of a district energy system simulation model. The working state of each occupant of a group of commercial was simulated based on appliances energy consumption data, where the times of arrival, lunch break and departure were selected on a 5 min interval with a random distribution by using the inverse function method (IFM).

One of the first agent-based models of occupancy in single office was provided in 2008 by Page et al. [13]. The model predictions used a Markov Chain to create random occupancy profiles (i.e. time of arrival and departure, periods of intermediate absence and presence, as well as periods of long absence from the space) based on and validated by sensor data, and were later used as an input to occupant behavior models within building simulation tools [13]. Wang et al. [14] handled occupancy as the straightforward result of occupant movement processes which occurred among the spaces inside and outside a building. By using the Markov Chain method, the model generated the location for each occupant and the zone-level occupancy for a whole office building type. Additionally, Virote and Neves-Silva [15] used the Markov Chain method to relate behavior in an office space catalogued by data logger measurements to occupant presence in the office building.

More recently in 2014, Dong and Lam [16] developed a real-time predictive control model for building heating and cooling systems based upon the occupancy behavior pattern detection in coordination with local weather forecasting, using advanced machine learning methods including Adaptive Gaussian Process, Hidden Markov Model, Episode Discovery and Semi-Markov Model.

Currently, more granular real-time measurements of the occupant presence, movement and interaction with system controls (thermostats, lighting) and building envelope action (windows, shades) are streamed. Sensor networks enable multidisciplinary and integrated layers of big data source collection, providing reliable information on occupancy recognition and scheduling, in addition to building performance and operation.

State-of-the-art data mining methods provide a powerful analysis technique to extrapolate useful and understandable occupancy patterns from big data streams.

For clarity, data mining is defined in 2001 by Hand et al. [17] as: "The analysis of large observation datasets to find unsuspected relationships and to summarize the data in novel ways so that owners can fully understand and make use of the data." Cabena et al. [18] provided another definition as: "An interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases and visualization to address the issue of information extraction from large databases." In many applications, it is difficult to extrapolate useful information from monitored building data due to large data scattering. Instead patterns of data discovered through data mining techniques may present applicable solutions at high levels of abstraction. Data mining of frequent patterns has been a focused theme in data mining research for over a decade with a comprehensive review provided by Han et al. [19]. Although, data mining techniques are largely applied to research fields such as marketing, medicine, biology, engineering, medicine, and social sciences, the application of a data mining framework to building energy consumption and operational data, is still in elementary phases. One highly effective technique of data mining for obtaining information on human-building interaction is the use of patterns correlating repetitive behaviors and actions to typical user profiles [20–22]. In this context, between 2011 and 2012 Yu et al. [23–26] tested several systematic data mining methodologies for identifying and improving occupant behavior in buildings. The results showed that the analysis methodology was powerful in providing insights into energy patterns related to the occupant behavior, facilitating evaluation of building saving potential by improving users' energy profiles as well as driving building energy policy formulation [23–26].

## 2. Methodology

Traditional methods of turning data into useful knowledge require data cleaning, analysis and interpretation. However, such manual data analysis often becomes impractical, slow and expensive as data volume grows exponentially. In view of these facts, researchers in the field of machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition and data visualization, have focused their attention on the

متن کامل مقاله

دریافت فوری

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات