Advanced Engineering Informatics 37 (2018) 1–13

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Understanding building occupant activities at scale: An integrated knowledge-based and data-driven approach

Andrew J. Sonta, Perry E. Simmons, Rishee K. Jain*

Urban Informatics Lab, Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Room 269B, Stanford, CA 94305, United States

ARTICLE INFO

Keywords:
Building energy
Data-driven
Gaussian mixture model
Knowledge-based
Occupant activities
Occupant dynamics
Space utilization
Zero-training

ABSTRACT

Buildings are our homes and our workplaces. They directly affect our well-being, and they impact the natural global environment primarily through the energy they consume. Understanding the behavior of occupants in buildings has vital implications for improving the energy efficiency of building systems and for providing knowledge to designers about how occupants will utilize the spaces they create. However, current methods for inferring building occupant activity patterns are limited in two primary areas: First, they lack adaptability to new spaces and scalability to larger spaces due to the time and cost intensity of collecting ground truth data for training the embedded algorithms. Second, they do not incorporate explicit knowledge about occupant dynamics in their implementation, limiting their ability to uncover deep insights about activity patterns in the data. In this paper, we develop a methodology for classifying occupant activity patterns from plug load sensor data at the desk level. Our method makes us of a common unsupervised learning algorithm—the Gaussian mixture model—and, in addition, it incorporates explicit knowledge about occupant presence and absence in order to preserve adaptability and effectiveness. We validate our method using a pilot study in an academic office building and demonstrate its potential for scalability through a case study of an open-office building in San Francisco, CA. Our method offers key insights into spatially and temporally granular occupancy states and space utilization that could not otherwise be obtained.

1. Introduction

Buildings are integral to our daily lives. People spend an estimated 87% of their time indoors [1], and researchers have shown that buildings directly affect our well-being [2]. Moreover, buildings worldwide account for over 19% of energy-related CO_2 emissions and 51% of global energy consumption [3], making them an integral part of our sustainable energy future. Fundamentally, buildings consume this energy to provide their occupants with services, including thermal and visual comfort, access to water, and power for electronic devices. As a result, understanding the relationship between buildings and their occupants is central to designing buildings that enhance occupant well-being, improve service delivery, and reduce energy usage.

We define occupant dynamics as the complex interactions between buildings and humans, encompassing occupant presence, occupant behavior (*i.e.*, the specific actions that occupants take in buildings, such as working at a workstation, taking a break, or even interacting with lighting or heating, ventilation, and air conditioning (HVAC) controls), occupant activity states (*i.e.*, abstracted and categorized information about occupant behavior), and the impact occupant behavior has on building operations. These dynamics are challenging to understand due to the increasing complexity of our building systems and the socio-technical complexities of occupant behavior. Even gaining a clear picture of the spatial and temporal activity patterns of occupants within a building is a non-trivial task [4]. While new types of sensors have facilitated more data-driven approaches to understanding occupant-building dynamics, they suffer from a few key limitations. Sensors designed to directly detect occupancy often mischaracterize the spaces they are sensing due to the complexity associated with various building spaces [5]. New statistical and data mining techniques that have been proposed to infer occupancy patterns from emerging high-fidelity data streams such as light levels [6], energy use [7], sound [8], and video [9] typically require a significant amount of ground truth training data that is cumbersome and often cost prohibitive to collect, thereby limiting their applicability and feasibility at scales beyond small pilot studies. Conversely, knowledge-based approaches to understanding occupant dynamics in buildings (*e.g.*, surveys, on-site engineering audits) can yield insights on occupant dynamics [10,11] but suffer from common reliability and scalability issues associated with indirect collection instruments [12].

* Corresponding author.
E-mail addresses: asonta@stanford.edu (A.J. Sonta), perry.simmons@stanford.edu (P.E. Simmons), rishee.jain@stanford.edu (R.K. Jain).

https://doi.org/10.1016/j.aei.2018.04.009
Received 22 September 2017; Received in revised form 15 April 2018; Accepted 16 April 2018

In various aspects of building design, construction, and management involving human activity, researchers have shown that combining expert knowledge about buildings with automated computing techniques can vastly improve the effectiveness of the embedded methods. In the context of augmented reality within buildings, researchers have shown that integrating explicit engineering knowledge about building layout and operator movement into the automated augmented reality framework can improve the accuracy of the overall system [13]. In construction management, the process of extracting meaningful information about the activities of construction workers from raw cellphone data can be enhanced by incorporating explicit engineering knowledge about the necessary levels of detail required for improving the effectiveness of construction activity simulations [14]. These studies and others like them emphasize the point that automated methods can be made more accurate and effective by integrating knowledge about the specific domain in the design of the overall methodology.

In this paper, we present a new methodology that integrates knowledge-based and data-driven approaches to understanding occupant activities in buildings with the goal of informing enhanced building design and energy efficient operations. Our method infers activity states for individual occupants using time-series data from low-cost, off-the-shelf plug load sensors. It incorporates explicit domain knowledge about how occupant activities impact plug load data into a common unsupervised learning algorithm—the Gaussian mixture model—to characterize the data into abstracted levels of activity. We design our method to be able to automatically analyze the highly variable data associated with occupant presence separately from the less variable data associated with occupant absence. This design decision in our method allows it to more deeply characterize the data while maintaining adaptability to new spaces, potential for scalability to larger spaces, and high accuracy. We validate and demonstrate that our method is able to determine individual occupancy states with a high-level of accuracy on a small control study, and we demonstrate the merits and applicability of our approach on a case study of a real 47-person open office in San Francisco, CA, USA.

2. Background

Building designers and managers are increasingly utilizing sensors and the data they collect to make decisions about how buildings are designed, built, and operated [15]. These sensors measure properties such as air temperature and humidity, lighting levels, sound, movement, and plug load energy use [16–20]. Each of these types of sensors produces time-series data that provides information about the changing state of the building. In many cases, data produced within a building can be utilized to make decisions that can improve the energy efficiency of that building: for example, a lighting sensor may provide feedback to lighting controls that can dim the overhead lighting if the building is receiving enough light from outdoors. In others, data can be used to understand characteristics of existing buildings so that the design of future buildings can be improved: for example, data describing existing building occupancy can be linked with predictive energy models to increase the accuracy of energy models [21].

This explosion of data has created an opportunity to provide new knowledge to engineers, designers, and building managers. In particular, previously unavailable information about the state of occupancy in buildings—the presence or absence of occupants as well as their activities—can be useful both for efficient building control of existing buildings and for improved space planning of future buildings [22]. Along with other data streams specific to each building system, the detection of occupant activities has been shown to be significant in addressing all forms of energy use in buildings, from lighting control [23–25] to HVAC control [26,27]. In addition, as knowledge about space use becomes more widely available to designers, the integration of design heuristics with occupancy models will be integral to designing spaces that better suit the needs of occupants [28]. In this section, we discuss the state of data-driven decision making in buildings for energy efficient building operations and improved building design, as well as the importance of occupancy and the state of the art for detecting occupant presence and occupant behavior in buildings. We elucidate the need for a robust, adaptable method for determining the activity states of occupants in buildings.

2.1. Data-driven and occupant-driven energy efficiency

Over recent years, the analysis of building energy data with statistical and data mining techniques has been shown to be helpful in improving energy efficient management of building systems. Within buildings, researchers have worked toward achieving a condition in which building systems—such as lighting, heating, and cooling—are provided only as much as they are needed, and only where and when they are needed. Matching these building systems with occupancy information has been shown to lead to significant energy savings [17,18]. Recently in commercial buildings, energy use data collected through power strips installed at the individual outlet level have been used for multiple approaches to save energy in buildings: to show that energy is wasted due to inefficient occupant behavior, such as leaving lights or other systems on during non-occupied hours [29]; to calibrate and improve the accuracy of building energy models in conjunction with other building data sources [30]; and to describe the behavior of occupants and improve schedule modeling in buildings [31].

Many studies have noted the high impact occupant presence and behavior has on building energy use [32–34]. Jia et al. [35] has noted that occupant behavior (as distinct from occupancy) relates to more than just the presence or absence of occupants in buildings—that is, the *activities* of occupants within the building have a large impact on building energy performance. However, this human element, which is responsible for much of building energy use, is often difficult to characterize. One reason is because it is multidimensional, requiring a fundamental understanding of spatial, temporal, and social dimensions of occupant behavior [36]. Understanding each of these dimensions and reconciling their effects on occupant behavior is critical to gaining a broad understanding of occupant behavior and its impact on building energy use. Furthermore, the structure and type of the social network of occupants has been shown to be highly influential when it comes to how occupants behave and adapt to information in buildings [37,38]. Researchers have shown that providing the right information to occupants can lead to changes in behavior that reduce buildings' energy consumption [39–43]. Due to the energy-consumption impact, complexity, and ever-changing nature of occupant dynamics in buildings, there remains a pressing need to better understand them.

2.2. Occupancy data and space utilization

While whole-building data and occupancy data have typically been studied in the context of energy efficient management of existing buildings, they also have the potential to be tremendously useful in providing knowledge to designers in the early stages of building design. Previous research has utilized model-based optimization in the design of buildings [44], and more specifically, in the planning of space layouts in buildings [45–47]. Recent work has conceptualized models that utilize computing in the assessment of the functional properties of designed spaces [48]. Specifically, analyzing designs for their ability to perform their function—for example, the ability for a proposed office space to promote a productive work environment—depends on knowledge from empirically based methods (*e.g.*, surveys) [49].

Architects have traditionally used personal perceptions of how occupants will use the spaces they design in their planning process. Formalized integration of human-centered knowledge into the building design process has previously been focused on perceptions of space [50] and heuristics for improved layouts [51], among others. More recent work has underscored the notion that it is difficult to quantify and