Towards automated inference of occupant behavioral dynamics using plug-load energy data

Andrew J. Sonta[1a], Perry E. Simmons[1b], Rishee K. Jain[1c]

[1] Urban Informatics Lab, Department of Civil & Environmental Engineering,
Stanford University, 473 Via Ortega, Stanford, CA 94305
a) asonta@stanford.edu; b) perry.simmons@stanford.edu; c) rishee.jain@stanford.edu

ABSTRACT

When building systems and occupants use energy, they create data — much of it unstructured and characterized as long running time series. Energy data captured at the plug level offers an opportunity not only to analyze highly granular building activities, but also to infer information about the behavior of occupants. Previous work examining occupant behavior typically seeks to understand how individual occupant schedules can be better modeled to improve the efficiency of building system operations, and therefore they treat individual actions as entirely self-contained. However, individual behavior — including that which draws power in buildings — is highly influenced by the inherent spatial and social network structures of occupants. Therefore, understanding the underlying spatio-social occupant network in a commercial building is integral to driving more energy efficient operations. Doing so is challenging since network relationships are highly complex and difficult to directly measure using traditional methods (e.g. surveys) and will require a deep understanding of occupant behavioral patterns. In this paper, we propose an automated methodology for inferring occupant behavioral patterns by classifying raw plug load data ascribed to individual occupants into *energy use states*. Our method utilizes a Gaussian Mixture Model to model occupant energy use variability and probabilistically classify energy use data into one of three generalized states. We present preliminary results of our classification algorithm using empirical data from a fifty-person commercial office building in San Francisco, California.
Keywords: Energy use, plug load, energy clustering, occupant dynamics

INTRODUCTION

Commercial buildings are responsible for about 20% of total US energy consumption, much of which is driven by human behavior. Occupants drive the use of systems that consume energy in buildings, from heating and ventilation to miscellaneous plug loads. Building systems are often designed with energy efficiency in mind, yet system efficiency optimizations typically occur with rudimentary information about occupant activity in the space, such as generalized occupancy schedules. Truly integrated building design and management would respond dynamically to occupant behavior. Occupant dynamics within buildings are highly complex, however, and fully understanding them requires reconciling spatial, temporal, and social dimensions of behavior. The complex challenge of understanding occupancy while improving building design and management warrants working toward a more robust understanding of occupant energy use dynamics in buildings.

Occupants and their plug loads create data that can be analyzed to better understand relationships — both among the occupants themselves, and between occupants and building systems. One useful tool in understanding such relationships is the graph, or network, defined as a set of nodes and edges. Considering occupants as nodes in a complex network offers a unique vantage point for discerning occupants' energy use relationships. We define energy use relationships as the similarities and dissimilarities of behaviors that directly influence energy consumption in a building (Chen et al. 2012; Gulbinas and Taylor 2014). Graphs of occupants — which describe the characteristics of each node as well as the relationships between pairs of occupants — have been shown to be useful in analyzing the complex dimensions of occupant energy use (Sonta et al. 2017).

Building an energy use relationship network of occupants in a building requires either a precise mathematical definition of how relationships form within a building, or granular measurements of the complex occupant activities that drive energy use in a building. With devices like plug load sensors, granular measurements are becoming increasingly available, yet the data can be highly unstructured. In this paper, we present a methodology for analyzing temporally and spatially granular plug load data to characterize the states of occupants' energy use behavior. Ultimately, such an analysis will be useful in characterizing relationships of occupants and *inferring* a spatio-social network of building occupant relationships.

RELATED WORK

Data-driven energy efficiency in buildings

Many studies have noted the importance of occupant behavior in affecting building energy performance. Previous work looking at commercial buildings has found that total energy use can be expressed as a combination of a baseline amount of energy consumption and a *human-driven* amount of energy consumption (Taherian et al. 2010). This human element is difficult to characterize, but recent research has highlighted three key dimensions of occupant dynamics in buildings: spatial, temporal, and social (Anderson et al. 2014; Gulbinas et al. 2015; Gulbinas and Taylor 2014). Recent research in addressing one or more dimension has utilized agent-based modeling combined with whole-building simulation tools (Azar and Menassa 2012); tools for improving control of building zones (Jazizadeh et al. 2014), as well as predicting occupant energy use to identify recurring patterns (Khosrowpour et al. 2016); and analysis of the relationship between network dynamics and the effectiveness of eco-feedback tools (Anderson and Lee 2016; Gulbinas et al. 2014; Siero et al. 1996).

Such analysis has much merit in advancing our understanding of the relationship between the human element and building systems, but it also underscores the importance of using data-driven methods to make important connections between occupant behavior and building energy use. Recent data-driven work has found that spatially granular occupancy sensors can yield significant energy savings (Agarwal et al. 2010). Further, recent work has found that sensors and their data streams can be useful in detecting activities within buildings (Milenkovic and Amft 2013).

Network theory in the built environment

One useful tool for representing granular data streams generated by many interconnected nodes is a graph, or network of nodes. Analysis of networks in the built environment has led researchers to determine that relationships between occupants highly impact energy consumption decision making and the usefulness of eco-feedback tools (Chen et al. 2012; Gulbinas et al. 2014; Khashe et al. 2016). Additionally, network models have been found to be useful in detecting situations in which occupant energy use is unexpected, based on the structure of the nodes and edges in the model (Sonta et al. 2017). This area of research has noted the importance of the edge weightings in these models (Chen et al. 2012), yet there has been little research that works toward understanding how graphs in the built environment can be constructed using granular and high-fidelity data.

Constructing a network of occupants in a building — and determining the edge weightings between nodes that describe their energy use relationship — requires a granular understanding of the occupant activity within the building. Within a building, information about occupants can be collected in the form of energy consumption at the desk level. This data can be used to abstract information about the occupants, such as the type of activities they are likely to be performing in an office building (Milenkovic and Amft 2013). For the purposes of building a graph of occupants, determining occupants' energy use states from granular plug load data will ultimately enable the inference of a graph of occupants that is built solely on plug load energy use data.

METHODOLOGY

In this section, we describe our methodology for classifying granular plug load data for each occupant into *energy use states*. We define *energy use states* as ranges of energy use for each occupant that describe one of three situations:

1) "off" — an occupant does not have their office equipment on and is not at his or her desk
2) "away" — an occupant's office equipment is on but not actively drawing power (e.g., equipment is "asleep")
3) "online" — an occupant's office equipment is on and actively drawing power

In the following subsections, we describe (1) the dataset, (2) the data cleansing process, (3) our classification algorithm, and (4) the role of classification in further analysis.

Data collection and cleansing

We collected plug load energy use data from the SF Department of the Environment office in central San Francisco. Energy use data was recorded using Enmetric power strips, which report the energy consumption in watt-hours over fifteen minute intervals. Our data reports the energy consumption for 52 occupants over fifteen minute intervals for the entire year 2015. Figure 1 shows a histogram of energy use values for a single occupant over the course of a full year. It is clear from this figure that this occupant's energy use can be roughly classified into one of two states, described by peaks in the histogram: one corresponding to a very low amount of energy use, and one corresponding to a range of higher energy use values.

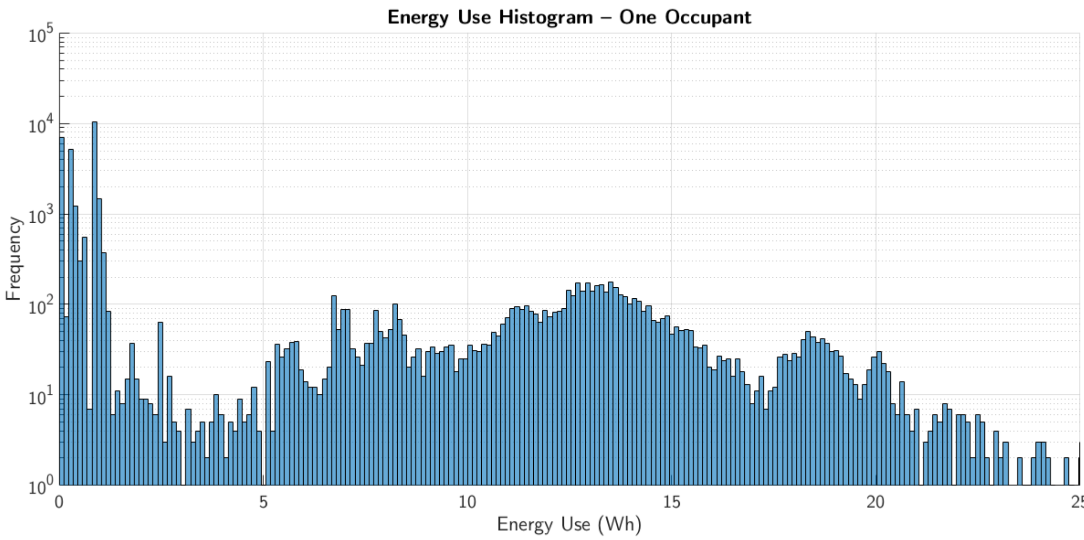

Figure 1: Histogram of a single occupant's energy use readings over one year

We define a "clean" energy time-step to be an interval in which the fifteen-minute energy use in watt-hours is recorded for all occupants. An accurate recording for every occupant in the network is necessary to ensure the accurate inference of relationships between occupants in the network. When an Enmetric power strip loses connectivity, it continues to record energy-use data and reports the cumulative energy-use over the period of lost connectivity in the time-step when connectivity is restored. This results in erroneously high readings for that time step. To avoid erroneous measurements in the analyzed data, we restricted our analysis to those occupants whose plug load monitors had strong connections throughout the period of study.

We also restricted our analysis to a period of 41 days with a single lost-connectivity event and otherwise clean data. The lost-connectivity event spanned three, fifteen minute intervals and occurred for every occupant outlet. We interpolated energy use in the missing time steps by assuming an equal distribution of the cumulative energy use across the missing time steps.

Energy state classification

Through exploration and visualization of the clean data we observed how energy use for individual occupants "jumped" between energy-use magnitudes at different points throughout a day. We hypothesize these sudden transitions in energy use represent important changes in occupant behavior. The observation that energy use for most occupants fluctuated between one of three *energy use states* motived the use of a classification algorithm that could classify each occupant's energy use values into one of three states (Figure 2).

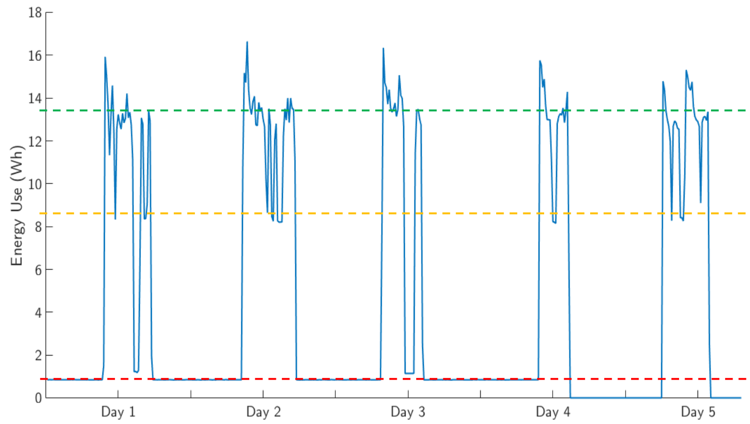

Figure 2: Energy use readings for a single occupant over one work week, with annotations of possible energy use states (green – "online", yellow – "away", red – "off").

Our methodology captures these important shifts in occupant energy states. Figure 3 describes our overall clustering algorithm. To account for possible shifts in individual occupants' baseline energy use (such as shifts due to different configuration of desktop appliances), we classify the data into *energy use states* separately for each day. We classified energy use data points into one of three energy states ("online", "away", "off") by applying a two-step classification algorithm to each day of data for all 49 occupants in the "clean" dataset. For every occupant, we extracted 41 day-vectors each containing 96 indices representing energy consumption values.

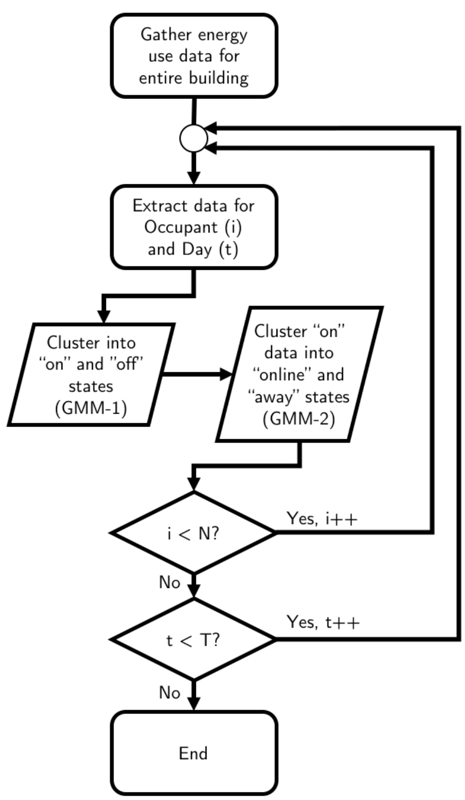

Figure 3: Classification algorithm flow diagram

Following previous work on time series heart rate data that used a Gaussian Mixture Model (GMM) to model heart rate variability (Costa et al. 2012), we use a GMM to model energy use variability. Our GMM was used to classify the daily time series energy use values for each occupant. For each vector of daily energy use values, we first classified each 15-minute energy consumption value in as either “off” or “on” using a 2-component GMM clustering algorithm. The “on” data was then extracted and further classified as either “online” or “away” using a 2-component GMM based on our hypothesis the observed changes in energy use stem from the presence or absence of an occupant performing an energy intensive task.

RESULTS

Classifying the raw data into energy states allows us to construct interpretable energy use models which capture significant changes in energy consumption for each occupant power strips. Figure 4 shows an example of the transformation from raw occupant energy data to an energy state model for one day’s worth of data for one occupant.

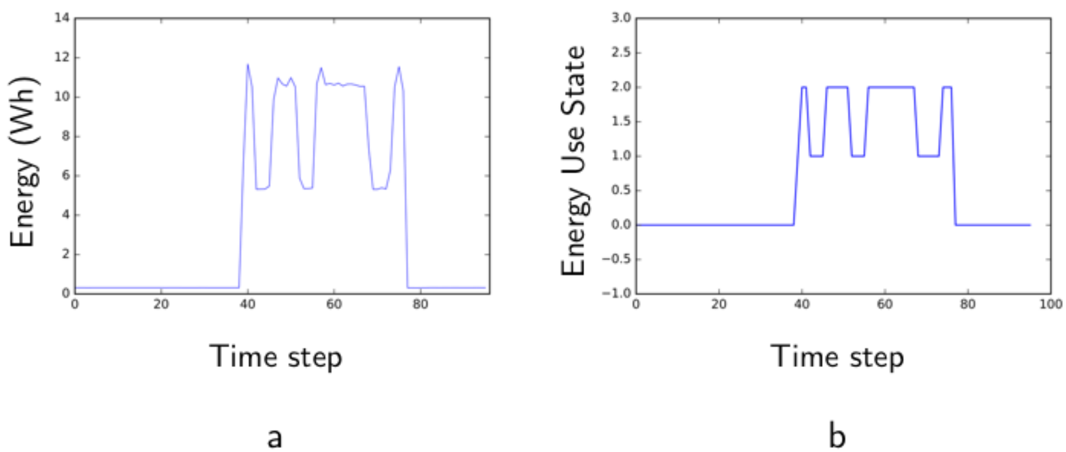

Figure 4: a) raw energy readings, and b) classified energy use states for single occupant over one day

We constructed an energy-state model for every occupant for every day of our analysis. Our classification methodology allows us to compare energy use states for each individual in the analysis. Such comparisons can lead to insights about behavioral correlations between individuals in the building. Figure 5 shows heat maps of both the raw energy use data and the classified energy use states for each occupant in our study over the course of one day. This figure illustrates interesting behavioral correlations that are not easily recognized with the raw data. For example, the data points indicated by the blue circles (occupants 41-49 in the middle of the day), show that the classification algorithm allows us to determine when certain individuals are at the same energy use state at the same time. In this cause, each member of the cluster of occupants is in the “away” state, possibly indicating that they all took a break at the same time. In a complete network, this would indicate that these occupants are more likely to have significant energy use relationships. One possible opportunity from this analysis — to improve the design and management of the space — is to ensure that occupants with similar behavioral patterns are sitting in the same lighting or HVAC zone within the building.

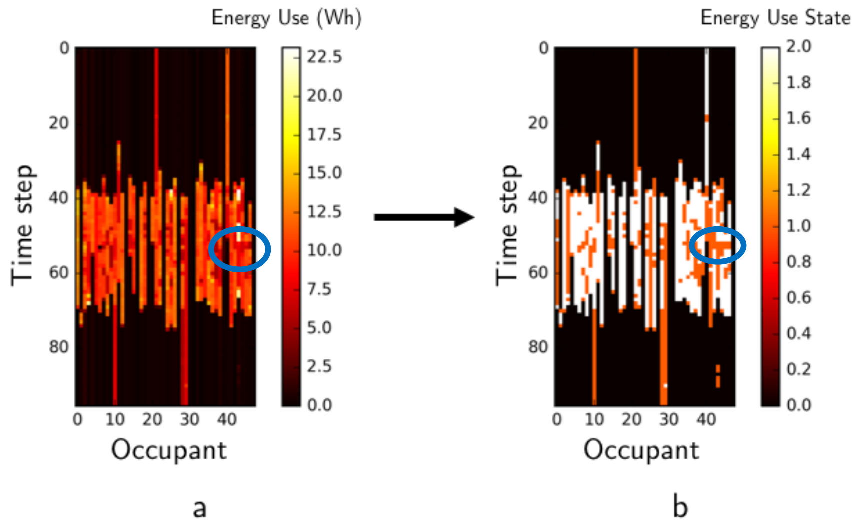

Figure 5: Heat map of occupants' a) raw energy readings, and b) classified energy use states

CONCLUSIONS & FUTURE WORK

The purpose of this study was to propose and test a classification algorithm that takes as an input raw, granular data describing occupants' energy use behavior at their desks, and abstracts out useful information about their state of energy use. Our work contributes to the literature by extending the use of Gaussian Mixture Models (GMM) to high-resolution energy usage data. Future work aims to validate our algorithm by conducting controlled experiments and collecting annotated plug-load data.

Additionally, our proposed classification method will be useful in future work that considers the problem of discerning energy use relationships among occupants in a building, because it allows for comparison of states of energy use between difference occupants, rather than comparison of the raw values themselves. In determining energy use relationships among occupants, it is their *relative behavior* that matters most. Understanding the complex dynamics of occupancy and occupant-driven energy use in buildings will ultimately help us make smarter decisions about how we design and manage our buildings.

ACKNOWLEDGEMENTS

The authors would like to the San Francisco Department of the Environment for their help in facilitating the gathering and sharing of the plug load data utilized in this study. The material presented is based in part upon work supported by Stanford Graduate Fellowship, Center for Integrated Facility Engineering and the US National Science Foundation under Grant No. 1461549. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Agarwal, Y., Balaji, B., Gupta, R., Lyles, J., Wei, M., and Weng, T. (2010). "Occupancy-driven energy management for smart building automation." *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building - BuildSys '10*, ACM Press, New York, New York, USA, 1.

Anderson, K., and Lee, S. (2016). “An empirically grounded model for simulating normative energy use feedback interventions.” *Applied Energy*, Elsevier Ltd, 173, 272–282.

Anderson, K., Lee, S., and Menassa, C. (2014). “Impact of Social Network Type and Structure on Modeling Normative Energy Use Behavior Interventions.” *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, 28(1), 30–39.

Azar, E., and Menassa, C. C. (2012). “Agent-Based Modeling of Occupants and Their Impact on Energy Use in Commercial Buildings.” *Journal of Computing in Civil Engineering*, 26(4), 506–518.

Chen, J., Taylor, J. E., and Wei, H. H. (2012). “Modeling building occupant network energy consumption decision-making: The interplay between network structure and conservation.” *Energy and Buildings*, 47, 515–524.

Costa, T. et. al. (2012). “Gaussian Mixture Model of Heart Rate Variability.” *PLoS ONE*, (J. Bourdon, ed.), Public Library of Science, 7(5), e37731.

Gulbinas, R., Jain, R. K., and Taylor, J. E. (2014). “BizWatts: A modular socio-technical energy management system for empowering commercial building occupants to conserve energy.” *Applied Energy*, 136, 1076–1084.

Gulbinas, R., Khosrowpour, A., and Taylor, J. (2015). “Segmentation and Classification of Commercial Building Occupants by Energy-Use Efficiency and Predictability.” *IEEE Transactions on Smart Grid*, 6(3), 1414–1424.

Gulbinas, R., and Taylor, J. E. (2014). “Effects of real-time eco-feedback and organizational network dynamics on energy efficient behavior in commercial buildings.” *Energy and Buildings*, 84, 493–500.

Jazizadeh, F., Ghahramani, A., Becerik-Gerber, B., Kichkaylo, T., and Orosz, M. (2014). “User-led decentralized thermal comfort driven HVAC operations for improved efficiency in office buildings.” *Energy and Buildings*, 70, 398–410.

Khashe, S., Heydarian, A., Becerik-Gerber, B., and Wood, W. (2016). “Exploring the effectiveness of social messages on promoting energy conservation behavior in buildings.” *Building and Environment*, 102, 83–94.

Khosrowpour, A., Gulbinas, R., and Taylor, J. E. (2016). “Occupant Workstation Level Energy-use Prediction in Commercial Buildings: Developing and Assessing a New Method to Enable Targeted Energy Efficiency Programs.” *Energy and Buildings*.

Milenkovic, M., and Amft, O. (2013). “An opportunistic activity-sensing approach to save energy in office buildings.” *Proceedings of the the fourth international conference on Future energy systems - e-Energy ’13*, 247.

Siero, F. W., Bakker, A. B., Dekker, G. B., and Van Den Burg, M. T. C. (1996). “Changing Organizational Energy Consumption Behaviour Through Comparative Feedback.” *Journal of Environmental Psychology*, Academic Press, 16(3), 235–246.

Sonta, A. J., Jain, R. K., Gulbinas, R., Moura, J. M. F., and Taylor, J. E. (2017). “OESPG: A Computational Framework for Multidimensional Analysis of Occupant Energy Use Data in Commercial Buildings.” *Journal of Computing in Civil Engineering (in press)*.

Taherian, S., Pias, M., Coulouris, G., and Crowcroft, J. (2010). “Profiling energy use in households and office spaces.” *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*, ACM, 21–30.