Detecting Occupancy Patterns of Smart Buildings for Efficient Energy Management

Mehdi Nourbahskh, Weiren Wang, Vivek Agrawal, Ceara Byrne

ABSTRACT

The assessment of a building's occupancy pattern is an important factor in contextually managing energy requirements. In 2011, the US building sector contributed 41% to the total energy consumption, of which a majority is wasted due to HVAC standards that do not correspond to a building's specific usage. To de-incentivize building energy loss, the US Green Building Council (USGBC) initiated the certification standard, Leadership in Energy and Environmental Design (LEED), however their standards address the problem reactively by focusing on the individual controllability of a room's thermal comfort, as opposed to proactively addressing how buildings respond to usage patterns. The goal of our work is to identify patterns in building occupancies and build prediction models to intelligently distribute the building's energy consumption based on temporal, spatial and seasonal variations. We describe a suite of Machine Learning methods such as Regression Analysis and Gaussian Mixture Models (GMM) / Expectation Maximization (EM) to learn the building occupancy trends. Using Polynomial Regression Models, we achieved the R^2 (Adjusted) value of about 40% based on 'Hour of the Day' occupancy details. Based on the Kolmogorov-Smirnov test comparing our Models for summer and winter months, we obtained a *p* value of 1.07 e-81 showing that the seasonal patterns are statistically significant. Lastly, we discuss the limitations of our work and future work for Project 2.

INTRODUCTION / PROBLEM STATEMENT

In 2011, the US building sector accounted for 41% of the total energy consumed, outperforming the transportation and industrial sectors by 28% and 31%, respectively.[4] In an effort to mediate this, designers and engineers model, simulate, and evaluate the potential energy consumption of buildings, leveraging inputs such as patterns of occupancy, which predicts how occupants intend to use the building, whether it be residential, commercial, or mixed. More heating and cooling ventilation is required for spaces with more tenants; however, currently there is no accurate standard for determining a building's potential pattern of occupancy, nor how these patterns might change over time.

Additionally, occupancy patterns are used in operating building control systems so that they respond proportionally to the number of occupants in buildings. Prior studies show that occupancy pattern detection can reduce the energy consumption of lighting by 50%[9] and air conditioning by 20%[6]. The prediction of occupancy pattern is important in HVAC control systems because the process of bring up an appropriate temperature is time consuming. The objective of this study is developing an AI-based model that predicts the occupancy pattern of buildings. The prediction model can be used to dynamically control HVAC systems for a given day of a season.

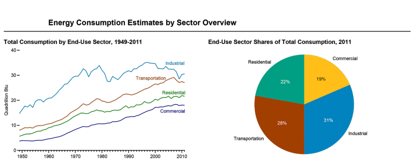

Figure 1: Energy Consumption Estimates by sector[4]

RELATED WORK

Existing building standards for HVAC are established by the American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE) and are developed for architects, structural, and civil engineers to abide by when designing a new building. In particular, ASHRAE 90.1 2004 is the energy standard for all larger commercial and residential buildings. The USGBC's LEED certification for new and existing construction provides a set of energy guidelines and best practices; however they reactively focus on providing individual thermal comfort as opposed to addressing the innate problem of thermal movement through a structure.

While the MERL dataset has been used to study how social networks move through a building by examining the spatial and temporal accumulators[2], none have researched the significance of meteorological or seasonal data on those spatial and temporal values. Several alternative studies have used occupancy patterns to improve the energy consumption of buildings, such as the ARIMAX model developed by the National Research Council in Canada to forecast the power demand of a building.[10] To increase the accuracy of the model, researchers used building occupancy as a significant independent variable. Moreover, researchers in the Lawrence Berkeley National Laboratory defined five occupancy patterns using measured light-switch data.[1] They concluded that such detailed models can replace the fixed or predefined models in energy simulation software. Additionally, Dong and Andrews developed an event-based pattern recognition algorithm to model and predict the behavior of occupants by incorporating temperature, relative humidity, acoustics, lighting, and CO_2 into their models.[3]

Researchers have used various methods to model the occupancy pattern of buildings. Erickson et al. developed an agent-based model that simulated occupancies by modeling the behavior of individuals.[7] They also proposed a multivariate Gaussian model that calculates the probability of a particular state at a given time. To model the temporal dynamics of occupancy, Erickson et al. used an inhomogeneous Markov Chain method, by which they could achieve 42% annual energy savings.[5] Erickson et al. expanded their study by comparing the performance of three Markov Chain algorithms, Closest Distance Markov Chain, Moving Window Markov Chain, and Blended Markov Chain. The results show that the most elegant and the best performing model was the Blended Markov Chain.[8]

The models developed in most of these studies use datasets that are time constrained (e.g., 79 days[10], 110 days[1], 122 days[3]) Thus, because of the seasonal variation of energy consumption, those models are not entirely generalizable. Our research, however, will model the occupancy pattern over the long term (730 days) using temporal, physical, and seasonal variance. The temporal variance represents the occupancy pattern change during a day. The physical variance shows the occupancy pattern change for a specific location and the seasonal variance shows the occupancy pattern change during the seasons.

APPROACH

The goal of this first project was to construct an AI-based model that would predict the occupancy patterns in a smart building. Leveraging human motion data collected from Mitsubishi Electric Research Labs (MERL), we investigated the temporal, physical, and seasonal models of occupancy patterns.[11] Further analysis of the dataset required us to re-establish our goal and research questions, and thus we reframed the model to determine peaks in occupancy flow at floor level entrances and exits in order to predict when an HVAC system should turn on each day of the week.

Figure 2 shows the study's methodology. First, we preprocessed the data and ran the Heat Map analysis to determine if any pattern existed. Then, we ran the Regression Analysis to estimate any relationship among variables. Based on the results of the Heat Map, we clustered the sensors and ran Expectation Maximization. Finally, we evaluated the results using Kolmogorov-Smirnov test.

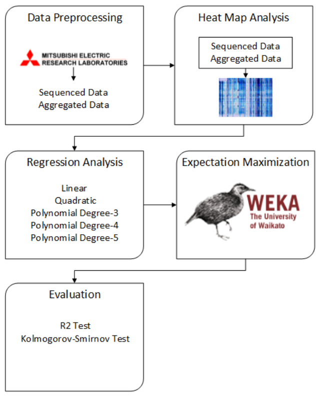

Figure 2: Research Methodology

The Dataset

The Mitsubishi Electric Research Labs (MERL) dataset represents the movement of people over the course of two years, recorded by over 200 wireless motion sensors. Individual wireless sensors detect the movement of people using passive infrared motion detectors. The data formats are various; for example, one includes four parts: Sensor ID, Start Timestamp, End Timestamp and Trigger Count, while another includes the Sunrise, Sunset Time, and Temperature.

Preprocessing the Data

We took several approaches to fully understanding the dataset. First, through our regression analysis, we noted that people rarely stay in the building during the holidays and weekends and that the occupancy is stationary at that time. We extracted the data for each workday and eliminated the holidays and weekends data. The raw data are fragmented because they recorded every sensor triggered in milliseconds. We then put the data in sequence and aggregated to every five minutes.

Heat Map

To start, we consolidated the frequency of sensor activations in the MERL dataset over all available monthly intervals in order to create a Heat Map. The vertical axis delineates the motion sensors and horizontal axis shows the months within the two year period that the data were collected. The chart shows the frequency of activations per sensor throughout the month and how these activations changed over the course of the two years. The output was color coded in 10 step increments, based off of the mean activations throughout the two years, or approximately 11,000 triggers. This was done to identify occupancy patterns and potential seasonal patterns.

Regression Models

Regression Analysis is the most common statistical process to establish elationships and evaluate the correlation between different variables. In order to identify patterns based on the MERL dataset, we merged the sensor trigger information with the location map for the sensors. We further segregated the interlinked location-based occupancy data into clusters or "wings" based on the consolidation of trigger frequencies of co-located sensors.

Additionally, we created subsets of the location-occupancy counts summarized based on "Hour of the Day", "Day of the Week" and "Month of the Year". We then ran regression models (Linear, Quadratic and Polynomial Regressions of Degrees, 3, 4 and 5) on the co-located subsets of data. We evaluated values for Ordinary (unadjusted) R-squared and R-squared adjusted for the number of coefficients for each Regression model:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}, \quad R^2_{adj} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SST}$$

Figure 3: R^2 *and* R^2 *Adjusted*

Here, *SSE* is the Sum of Squared Error, *SSR* is the Sum of Squared Regression, *SST* is the Sum of Squared Total, n is the Number of Observations, and p is the Number of Regression Coefficients (including the intercept). Because R^2 increases with added predictor variables in the regression model, the adjusted R^2 accommodates for the number of predictor variables in the model.

Gaussian Mixture Models (GMM) / Expectation Maximization (EM) for Summer and Winter

GMM/EM is one of the classical unsupervised learning algorithms. When it comes to a clustering problem, the GMM is extremely efficient if the model is composed of several Gaussian models, each of which represents a distribution of a feature. GMM is a parametric Probabilistic Density Function (PDF) composed of weighted sums for component Gaussian densities. GMM is often combined with the EM algorithm to compute the cluster center and shape. The EM algorithm includes two steps – Expectation Evaluation & its Maximization. These two steps are repeated and each iteration is guaranteed to increase the log-likelihood and guaranteed to converge to an optimal likelihood function.

The outcome of both the Heat Map and the Regression Models suggested existence of strong Gaussian Distribution characteristics in the underlying dataset. To exploit this data characteristic, we decided on a more focused approach by limiting the experiment to 4 representative sensors based on their strategic location; specifically, proximity to the floor exits and placement in hallways with high foot-fall. We utilized EM on GMM to identify the correlation between the data obtained, based on our experiment with other methods.

RESULTS & EVALUATION

Heat Map

We observed that there were visible patterns representative of seasonal variances and clusters of high occupancy / footfall areas (see Figure 4). In particular, the summer months showed more activity compared with winter months, but it also helped identify which sensors were more active throughout the building.

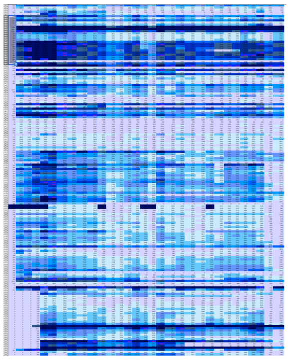

Figure 4: Heat Map Analysis (Top: Months, Left: Sensor #)

Regression Models

We observed that the best-case values for Ordinary R^2 and R^2 Adjusted were observed for higher order Polynomial Regression, and yet were still under 40%. This observation was significant as it only explained less than 50% of the variability of the response data around its mean.

Table 1: The Evaluation of Regression on Occupancy vs. Hour of the

Regression Model	R_Sqr Value	R_Sqr_Adj Value
Linear	0.2404023340	-
Quadratic	0.2460365997	0.2450412684
Polynomial Degree-3	0.3780666542	0.3768342895
Polynomial Degree-4	0.3794027981	0.3777620917
Polynomial Degree-5	0.3874839456	0.3854584296

Day, for Wing 8-C

The results also indicated strong alignment to the Normalized Gaussian Distribution. Based on these observations, we further broke the data down into Weekdays vs. Week ends and a similar pattern emerges as can be seen on the following figures:

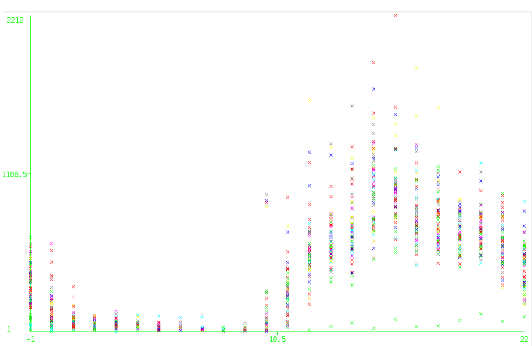

Figure 5: Occupancy Patterns Over a 24-Hour Cycle, Without Weekends

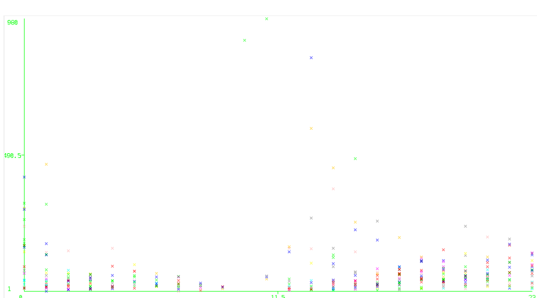

Figure 6: Occupancy Patterns Over a 24-Hour Cycle, Only Weekends

GMM/EM Results:

Figure 7 identifies the correlation between the summer and winter. The distribution between the two seasons looks similar; both distributions start at 0.3, peak around 0.6, and end at 1. To statistically compare these two samples, we ran Kolmogorov-Smirnov test (see Evaluation).

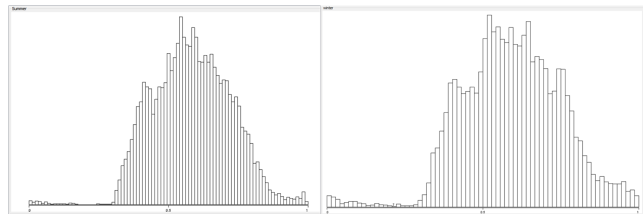

Figure 7: The Motion Distribution over two seasons: Summer (L) and Winter (R)

We used Weka to run EMs and the results show two clusters for winter and summer. The mean of the winter is around 2 PM with a high standard deviation (4.4). The means of the two clusters for winter are 12 and 4 PM with a standard deviation around 2.8 for both clusters.

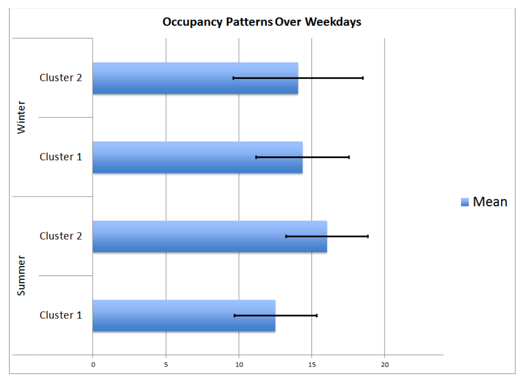

Figure 8: The Mean and Standard Deviation Distribution over Summer and Winter

Kolmogorov-Smirnov Test

The K–S test, a nonparametric test of the equality of continuous probability distributions, used to compare the PDFs for the occupancy data from two seasons, summer (May - July, 2006) and Winter (Nov 2006 - Jan 2007) and evaluate the correlation between seasonal patterns derived by the GMM EM model.

We observed the following results as part of our experiment:

- h value: 1
- p value: 1.0739 e-81

The very small value for p indicates that we can reject the null hypothesis and the results are statistically significant.

DISCUSSION

The observations made on the data during the Preprocessing Phase required our team to revise our research question so that it focused on the temporal variance with the objective of building a predictive model to help building managers identify the optimum time to turn on and off HVAC equipment. In order to do so, we restricted our dataset to model a small subset of sensors and analyze time-based variances. By including more representative samples, such as sensor data or meteorological data, we expect to improve our model.

Current results show that the model of predicting the occupancy pattern is complicated. To improve the prediction, we intend to look into neural networks, deep learning models, and Hidden Markov Models. Another issue remains of how to correctly map sensor activities to the occupancy patterns. Finding a method to avoid sensor activities that prevent multiple counts for the same people would improve the result.

Potential avenues for further research consist of extending this research problem to the optimization of building evacuation plans. Understanding and predicting the "escape behavior" of occupants has implications within firefighter mediation and building safety in the event of a catastrophe.

REFERENCES

1. Chang, W.-K. and Hong, T. 2013. Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data. Build. Simul., 6, 1, 23-32.
2. Connolly, C. I., Burns, J.B. , and Bui, H. H. 2008. Recovering social networks from massive track datasets. IEEE Workshop on Applications of Computer Vision. Copper Mountain, CO, IEEE, 2008.
3. Dong, B. and Andrews, B. 2009. Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. In 11th International IBPSA Conference, Glasgow, Scotland, 1444-1451.
4. EIA 2011. Annual nergy review, US Department of Energy, Washington, DC, 38.
5. Erickson, V. L., Carreira-Perpinan, M. A. and Cerpa, A. E. 2011. OBSERVE: Occupancy-based system for efficient reduction of HVAC energy. In 10th International Conference on Information Processing in Sensor Networks (IPSN), Chicago, IL, 258-269.
6. Erickson, V. L. and Cerpa, A. E. Occupancy based demand response HVAC control strategy. In Proceedings of the Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, Zurich, Switzerland, ACM,7-12.
7. Erickson, V. L., Lin, Y., Kamthe, A., Brahme, R., Surana, A., Cerpa, A. E., Sohn, M. D. and Narayanan, S. 2009. Energy efficient building environment control strategies using real-time occupancy measurements. In Proceedings of the Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, Berkeley, California, ACM, 19-24.
8. Erickson, V.L., Miguel, A., Carreira-Perpi, M.A. and Cerpa, A. E. 2014. Occupancy modeling and prediction for building energy management. ACM Trans. Sen. Netw.10, 3, 1-28.
9. Harle, R. K. and Hopper, A. 2008. The potential for location-aware power management. In Proceedings of the Proceedings of the 10th International Conference on Ubiquitous Computing, Seoul, Korea, ACM,302-3011.
10. Newsham, G. R. and Birt, B. J. 2010. Building-level occupancy data to improve ARIMA-based electricity use forecasts. In Proceedings of the Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, Zurich, Switzerland, ACM, 13-18.
11. Wren, C., Ivanov, Y., Leigh, D. and Westhues, J. 2007. The MERL motion detector dataset: 2007 workshop on massive datasets. Technical report.