Data Mining of Occupant Behavior in Office Buildings

Original
Data Mining of Occupant Behavior in Office Buildings / D'Oca, Simona; Corgnati, STEFANO PAOLO; Hong, Tianzhen. - In: ENERGY PROCEDIA. - ISSN 1876-6102. - 78:(2015), pp. 585-590. [10.1016/j.egypro.2015.11.022]

Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 00 (2015) 000–000

Energy Procedia

www.elsevier.com/locate/procedia

6th International Building Physics Conference, IBPC 2015

Data Mining of Occupant Behavior in Office Buildings

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

[a] Politecnico di Torino – DENERG, Torino, Italy
[b] Lawrence Berkeley National Laboratory, Berkeley, California, USA

Abstract

Literature studies confirm occupant behavior is setting the direction for contemporary researches aiming to bridge the gap between predicted and actual energy performance of sustainable buildings. Using the Knowledge Discovery in Database (KDD) methodology, two data mining learning processes are proposed to extrapolate office occupancy and windows' operation behavioral patterns from a two-years data set of 16 offices in a natural ventilated office building. Clustering procedures, decision tree models and rule induction algorithms are employed to obtain association rules segmenting the building occupants into working user profiles, which can be further implemented as occupant behavior advanced-inputs into building energy simulations.

Keywords: data mining; occupant behavior; office building; window operation; occupancy patterns

1. Introduction

The development of energy-conserving technologies is a necessary but incomplete step toward reduced energy consumption in buildings. Since energy consumption may vary largely due to how occupants interact with system controls and the building envelope, achieving energy conservation emerged as a double challenge, partly technical and partly human. Several studies underlined that huge variability exists in terms of default settings and day-to-day use of control systems and appliances in office buildings, where 'behavior' is central to consumption levels [1-3]. In 2004 Bordass et al. [4] referred to this occurrence as the 'credibility gap', alluding to the loss of credibility when design expectations of energy efficiency and actual building consumption outcomes differ substantially. They stated that credibility gaps arise not so much because occupants preform 'wrong', but because the assumptions often used

* Corresponding author. Tel.: +39-011-090-4507; fax: +39-011-090-4507.
E-mail address: stefano.corgnati@polito.it

are not well enough informed by what really happens in practice. Currently, building simulation tools can only imitate some typical occupant activities in a rigid and pre-defined way (occupancy, use of windows, etc.). Nevertheless, occupant behavior and comfort is stochastic, complex, and multi-disciplinary. New methodologies able to describe more realistic building occupants' behavioral patterns need to be developed.

1.1. Data Mining

Data mining is defined by Cabena et al. [5] as "An interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases and visualization." Data mining techniques are largely diffused into research fields such as marketing, medicine, biology, engineering, medicine, and social sciences to address the issue of patterns extraction from large databases. Application of data mining techniques to building energy consumption and operational data is still in an elementary phase. However, preliminary studies [6-10] showed data mining is a powerful technique in providing insights into energy pattern related to the occupant behavior, facilitating evaluations of building saving potential by improving users' energy profiles as well as driving building energy policy formulation.

1.2. Occupancy

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. Because of the stochastic nature of occupant behavior, the number of people occupying 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 Gunay et al. [11] providing a comprehensive and up-to-date critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices.

1.3. Window opening

The most important issue in between perceived indoor environmental quality and outdoors, in the built environment, is the building envelope [12]. Further, since the building envelope is getting always more thermally efficient, ventilation and air infiltrations due to window openings are increasing their influence with respect to energy use, becoming the most dominant source of thermal loss of the heat balance mechanism [13]. Also, a literature review carried out in 2012 by Fabi et al. [14] of more than 70 scientific papers, indicated that window operation was not only influenced by perceived thermal condition, but it was also seen as a response of sensed indoor air quality, external (outdoor temperature, solar radiation, wind speed, rain) and internal (indoor temperature) environmental conditions, as well as contextual factors (window type, time of the day, season of the year) and personal and cultural preferences. As a consequence, window opening is not only one of the most relevant tools that allow occupants to bring about desired indoor thermal and air quality conditions, by moving air through the building. It also represents one the behavioral actions, which mostly can make design expectations of energy efficiency and actual building consumption outcomes differ substantially.

2. Knowledge Discovery in Database (KDD) 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 Knowledge Discovery in Databases (KDD) [15]. KDD involves the application of the following six steps:

1. *Data selection:* Creating a target data set, or a subset of variables, on which discovery is to be performed.

2. *Data cleaning and preprocessing:* Removal of noise or outliers, strategies for handling missing data fields.
3. *Data transformation:* Finding useful features to represent the data depending on the goal of the task.
4. *Data mining:* Matching a particular data mining method for searching patterns in the data.
5. *Data interpretation and evaluation:* Deciding appropriate parameters and interpreting mined patterns.
6. *Knowledge extraction:* Consolidating the discovered knowledge to be used for further applications

This study uses the KDD data mining process to extrapolate information on window opening and occupancy schedule patterns from measured building data. Systematic data-mining methodologies (*Cluster Analysis, Association Rules, Decision Tree, Rule Induction*) have been applied, along with the open source software Rapid Miner Studio [16], to identify two frameworks generating advanced modeling inputs of energy related behaviors.

3. Results: the Data Mining Frameworks

An office building having 16 private offices, located in Frankfurt am Main [17] is used as the case study to develop two data mining frameworks. In this study, we used a two complete years dataset with:

a) 10-minute window opening and occupancy interval data;
b) 10-minute indoor and outdoor environmental parameters interval data.

3.1. Occupancy Schedule Learning Framework

A three-step data mining occupancy schedule learning framework is proposed to provide insight into patterns characterizing the probability of an office being occupied at a specific season of the year, day of the week and time of the day (figure 2). Final goal is to identify archetypal user profiles for which different energy conservation strategies, as well as building design recommendations, may be appropriate.

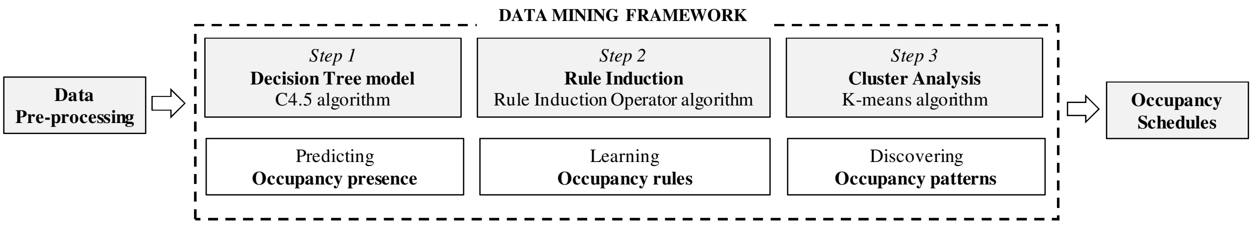

Fig. 2. Proposed occupancy schedule learning framework

In step 1, the data set is mined though a decision tree model, by using the C 4.5 algorithm, which predicts the occupancy presence. Raw data were transformed into more significant pre-processed data representing invariant attributes of the data set and mined though a decision tree model with the goal to predict the value of a label attribute (occupancy) based on predictor attributes (Season, Day of the week, Time of the day, Occupancy State and Window Change) of the data set.

In step 2, a rule induction algorithm is used to mine a pruned set of rules on the results from the decision tree model. The predicted answers of the tree-like graph model are probabilities of the office being vacant/occupied based on the condition defined by the path from the root tree to the final leaf (Figure 3).

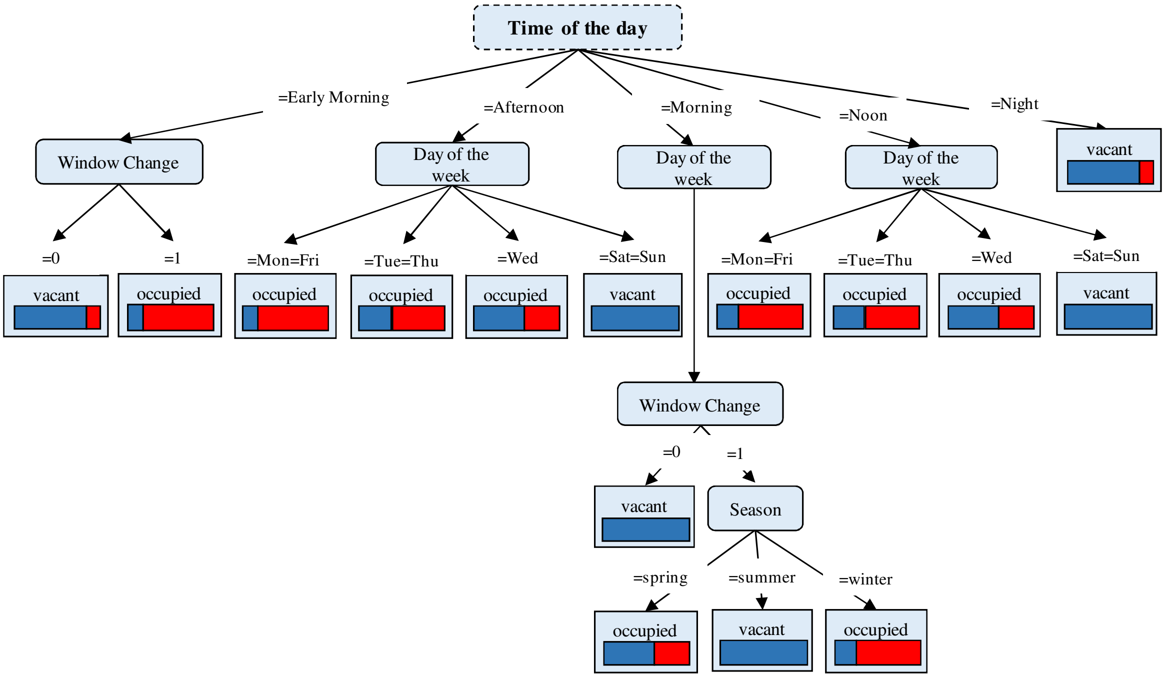

Fig. 3. Decision tree model for the prediction of the office being vacant/occupied

In step 3, a cluster analysis by employing the k-means algorithm is performed in order to disaggregate the occupancy presence into working day schedules (Monday-Friday). The clustering of typical office working activities such as working stable from the office, moving (arriving or leaving) from the office and taking breaks highlighted that the average breakdown of hours spent every day by the monitored users in the single office space may differ broadly, as well as a overtime work and significant shifting in the time the occupancy pattern activity/presence/ intermediate absence was occurring. All such patterns are omitted by using fixed deterministic occupancy schedules. More detailed outcomes occupancy schedule learning framework are presented in D'Oca et Hong [18].

3.2. Window opening Framework

A three step framework is proposed as an improvement of the notion of window opening behavioral patterns not only as merely statistical relevant clusters based on the frequency of openings, but also incorporating the driver-response conditioning dimension with typical window opening habits (figure 4).

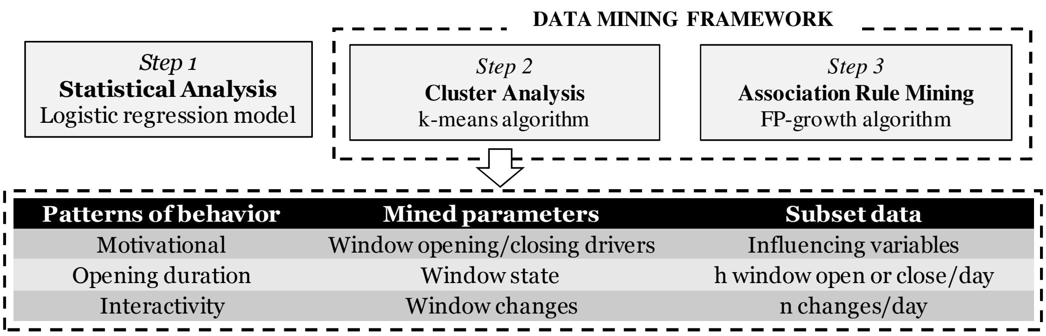

Fig. 4. Proposed window opening learning framework

In step 1, a statistical analysis technique (logistic regression) was applied to the given data set to discover the factors (variables and coefficients) influencing window opening and closing behavior. Results indicated indoor air temperature, outdoor air temperature, time of the day (office arriving time and early morning) and occupancy presence are the top drivers for window opening. Indoor air temperature, time of the day (office leaving time and evening), occupancy presence and outdoor air temperature emerged as top drivers for window closing. Three clusters of motivational drivers (*thermal-driven, thermal/time-driven, time-driven*) emerged for the window opening action, while two motivational drivers (*thermal-driven, time-driven*) were clustered in the given data set (figure 5).

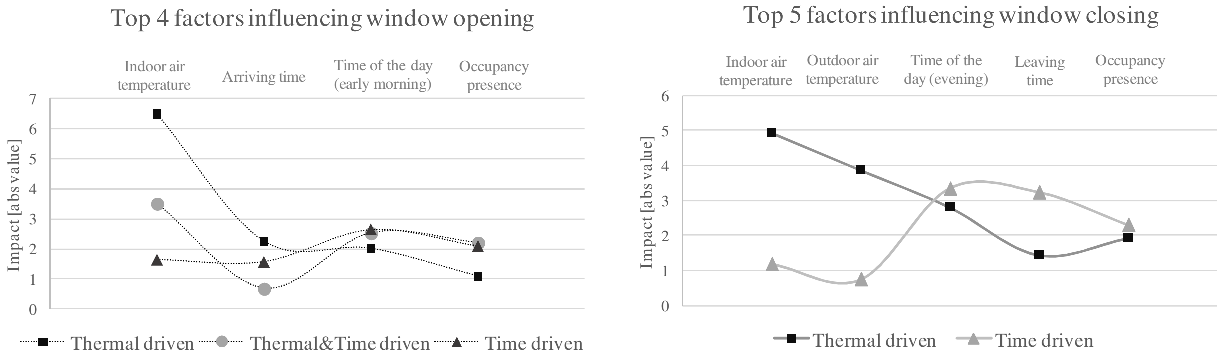

Fig. 5. Clustered top influencing factors for window opening and closing

In step 2, clustering procedures were employed to obtain behavioral patterns, which are not only statistical relevant driver-response conditioning clusters, but also incorporate the motivational dimension with typical habits. Goal was to estimate why (motivational pattern), for how long (opening duration pattern) and how often (interactivity pattern) windows were operated by users in offices of the same building.

In step 3, the clustered patterns constitute a base for association rules segmenting the building occupants into typical office user profiles. Two typical window opening profiles emerged: user α (80% of offices assigned) mainly *physical environmental driven* and user β (80% of offices assigned) mainly *contextual driven (table 1).*

Table1. Window opening office user profiles emerged from the data-mining framework

Patterns of behavior	**User α (80%)**	**User β (20%)**
Motivational	physical environmental driven	contextual driven
Opening duration	short periods	long periods
Interactivity	infrequently	frequently

The results indicated that office users interact with windows principally driven by thermal discomfort (indoor air temperature) but also behave accordingly to daily routine (time of the day) and habits (arriving and leaving time). From the association rule, it emerged that when interacting with windows to restore the indoor environmental quality, users mainly driven by *physical environmental parameters* – such as indoor and outdoor temperature – had less impact on natural ventilation than users driven by *contextual factors and habits* – such as time of arriving and leaving the office – opening windows for shorter periods of time and usually interacting less frequently. More detailed outcomes of the proposed window opening learning framework are presented in D'Oca and Hong [19].

4. Discussion

Scalability of solutions is one of the most critical point in pattern mining. Results of the mined occupancy schedules and window opening rules are circumstantial to the case study building and do not represent the complete set of patterns that can be derived within the given data set. Nevertheless, they characterize the most compact, physical meaningful and high quality set of patterns that can be derived with satisfactory performance. Moreover, the proposed data mining frameworks are generic enough to be possibly applied to new behavioral data sets providing solutions to direct specific operation and maintenance strategies. The future applications of the proposed method to discern occupant behavior advanced inputs (obAIMs) to be implemented into a building energy modeling programs would strongly support the investigation of the impact of typical working occupancy patterns on design and operation of office appliances and control systems, both at a building and block of building level.

5. Conclusion

The outcomes of the proposed research suggest data mining techniques are proficient in highlighting behavioral patterns associating the driver-response conditioning motivational dimension with typical window opening and occupancy habits, overcoming the lack of personalization of statistical methods. It is often difficult to extrapolate useful building occupants information from big monitored building data set by means statistical analysis. Due to the data scattering at this level, statistical analysis techniques may fail in obtaining reliable mathematical models by over or under fitting the data. Instead, patterns of data discovered through data mining techniques may highlight commonsense knowledge. In this context, data mining techniques have been shown as able to automatically extrapolate valid, novel, potential useful and understandable building occupant patterns from big data streams. Data mining techniques are not intended to substitute or contrast the direct stochastic models or indirect agent-based models already developed for the integration of occupant behaviors into building energy simulations. More likely, the knowledge discovered through data mining techniques aims to overcome the shortcomings of more traditional techniques, specifically when dealing with big data stream, by providing reliable models of energy related behaviors with fast legibility and high replication potential.

References

[1] Hong T, Lin HW. Occupant behavior: impact on energy use of private offices. ASim 2012 - 1st Asia conference of IBPSA; Shanghai, China, November 25-27, 2012.

[2] Haldi F, Robinson D. On the behaviour and adaptation of office occupants. Building and Environment 2008; 43:2163–2177.

[3] Masoso OT, Grobler LJ. The dark side of occupants' behaviour on building energy use. Energy and Buildings 2010; 42:173-177.

[5] Bordass W, Cohen R, Field J. Energy Performance of Non-Domestic Buildings: Closing the Credibility Gap. Building Performance Congress; Frankfurt, April 19-24, 2004.

[6] Cabena P, Hadjinian P, Stadler R, Verhees J, Zanasi A. Discovering data mining: From Concept to Implementation, Prentice-Hall, Inc.; 1998.

[7] Yu Z, Haghighat F, Fung BCM, Yoshino H. A decision tree method for building energy demand modeling, Energy and Buildings 2010; 42: 1637–1646.

[8] Yu Z, Haghighat F, Fung BCM, Yoshino H. Morofsky E. A systematic procedure to study the influence of occupant behavior on building energy consumption, Energy and Buildings 2011; 43:1409–1417.

[9] Yu Z, Haghighat F, Fung BCM, Zhou L. A novel methodology for knowledge discovery through mining associations between building operational data, Energy and Buildings 2012; 47:430–440.

[10] Zhao J, Yun R, Lasternas B, Wang H, Lam KP, Aziz A, Loftness V. Occupant behavior and schedule prediction based on office appliance energy consumption data mining, Proceedings on CISBAT 2013, Lausanne, Switzerland, 2013.

[11] Gunay HB, O'Brien W, Beausoleil-Morrison I. A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices. Building and Environment 2013; 70:31-47.

[12] Raja IA, Nicol JF, McCartney MA, Humphreys M. Thermal comfort: use of controls in naturally ventilated buildings. Energy and Building 2001; 33:235-44.

[13] Wagner A, Gossauer E, Moosmann C, Gropp T, Leonhart R.Thermal comfort and workplace occupant satisfaction – results of field studies in German low energy office buildings. Energy and Building 2007; 39:758–769

[14] Fabi V, Andersen RV, Corgnati SP, Olesen BW. Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and model. Building and Environment 2012; 58:188-198

[15] Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery: An Overview, Advances in Knowledge Discovery and Data Mining, AAAI Press / The MIT Press, Menlo Park, CA; 1996

[16] RapidMiner Studio, http://rapid-i.com/content/view/181/190/.

[17] IEA Annex 53 Task Force. Final report, Total energy use in residential buildings – the modeling and simulation of occupant behavior; 2012.

[18] D'Oca S, Hong T: Occupancy schedules learning process through a data mining framework. Energy and Building 2015; 88:395-408.

[19] D'Oca S, Hong T: A data-mining approach to discover patterns of window opening and closing behavior in offices. Building and Environment 2014; 82:726-739.