HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

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

Occupancy driven building performance assessment

Dimosthenis Ioannidis[a,b], Pantelis Tropios[a], Stelios Krinidis[a,*], George Stavropoulos[a], Dimitrios Tzovaras[a], Spiridon Likothanasis[b]

[a] Information Technologies Institute, Centre for Research and Technology Hellas, 6th Km Charilaou-Thermi, Thermi-Thessaloniki, 57001, Greece

[b] Pattern Recognition Laboratory Computer Engineering and Informatics, University of Patras, Rio, Patras, Greece

HIGHLIGHTS

- Big data analysis of building measurements.
- Building occupancy-based knowledge mining for the building performance assessment.
- Key performance indicators for evaluating the building performance and operation.
- Visual analytics techniques for assessing building performance.
- Heterogeneous building information correlation.

ARTICLE INFO

Article history:
Published online 16 November 2016

Keywords:
Building performance assessment
Big data analytics
Visual analytics
Human presence
Occupancy extraction
Building occupancy visualization

ABSTRACT

In this paper, we focus on the building performance assessment using big data and visual analytics techniques driven by building occupancy. Building occupancy is a paramount factor in building performance, specifically lighting, plug loads and HVAC equipment utilization. Extrapolation of patterns from big data sets, which consist of building information, energy consumption, environmental measurements and namely occupancy information, is a powerful analysis technique to extract useful semantic information about building performance. To this end, visual analytics techniques are exploited to visualize them in a compact and comprehensive way taking into account properties of human cognition, perception and sense making. Visual Analytics facilitates the detailed spatiotemporal analysis building performance in terms of occupancy comfort, building performance and energy consumption and exploits innovative data mining techniques and mechanisms to allow analysts to detect patterns and crucial point that are difficult to be detected otherwise, thus assisting them to further optimize the building's operation. The presented tool has been tested on real data information acquired from a building located at southern Europe demonstrating its effectiveness and its usability for building managers.

Peer review under responsibility of Qassim University.

* *Corresponding author. Fax: +30 2310 474 128.*
E-mail address: krinidis@iti.gr (S. Krinidis).

http://dx.doi.org/10.1016/j.jides.2016.10.008

1. Introduction

Building performance is a fundamental task for energy management. It is a multi-parametric task, where each parameter affects different building's aspects in different level. In order to estimate and assess performance of an actual building, models require inputs such as building size, shape, orientation, construction materials, heating, ventilation, and air-conditioning (HVAC) system size and type, interior and exterior lighting, appliance loads, lighting and cooling loads and other physical parameters. Many of these inputs depend on the level of occupancy within a building, thus occupant behavior in buildings moved in the focus of researchers.

Occupancy behaviours and trajectories are either random, business processes driven (*e.g.* printing, using specific appliances, *etc.*) or comfort-related (*e.g.* adjusting thermostats, opening windows for ventilation, turning on/off lights, *etc.*). Either way, they may change the indoor environment in terms of temperature, humidity, luminance *etc.* which influences the overall building performance. Nevertheless the stochastic nature of occupants' behavior, the number of people that occupy a space, their trajectories through building spaces and the duration occupied is an important aspect. Occupants' locations within the building vary throughout the day and this distribution can be valuable information when evaluating demand control strategies. Therefore, there is a need for more detailed measurements of the occupants' presence, movement [1] and interaction with system controls (thermostats, lighting), equipment (PCs, printers, *etc.*) and building envelope action (windows, blinds) [2]. To this end, assessment tools able to correlate, combine and analyze the occupancy with all other building information by using knowledge mining techniques in conjunction with advanced visualization mechanisms are necessary to evaluate buildings' performance.

Knowledge Mining is a process which analyzes the large volumes of data from the various perspectives and summarizes it into useful information; knowledge mining has become an essential component in various fields of research. It is a computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. A number of studies have been conducted to identify the associations and correlations between measured data. Traditionally, researchers utilized knowledge mining techniques such as:

(a) **Anomaly detection/outlier detection** [3,4] also known as outlier detection is the search for items or events which do not conform to an expected pattern. The patterns thus detected are called anomalies and often translate to critical and actionable information in several application domains. Three broad categories of anomaly detection techniques exist. **Unsupervised anomaly detection** techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set. **Supervised anomaly detection** techniques require a data set that has been labelled as "normal" and "abnormal" and involves training a classifier (the key difference to many other statistical classification problems is the inherent unbalanced nature of outlier detection). **Semi-supervised anomaly detection** techniques construct a model representing normal behaviour from a given normal training data set, and then testing the likelihood of a test instance to be generated by the model.

(b) **Association rule learning** [3,4] is a popular and well researched method for discovering interesting **relations between variables** in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness. The term of "trend analysis" is also used to express the identification of common patterns within a large database.

(c) **Clustering analysis** is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory knowledge mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bio-informatics. Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them.

(d) **Classification analysis**: In machine learning and statistics, classification is the problem of identifying to which of a set of categories (sub-populations) a list of observations belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.

(e) **Regression analysis** [5,6] is a statistical process for estimating the relationships among variables. It includes many techniques for modelling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables.

In the field of Building performance, knowledge mining approaches are used in building automation to identify usage scenarios [7], or to estimate the energy consumption in residential buildings [8], and tropical regions [9]. Characterization of electric energy consumers was acquired using knowledge mining [10]. It was also used to analyze data collected from simulations [11], or wireless sensor networks [12]. Most of these studies focus on the energy consumption of buildings, but few evaluate occupant related aspects of building performance or the geometrical information of the buildings. In the presented work the big amount of data (geometrical information of the building, equipment and space energy consumption and detailed space occupancy information) will be correlated in the spatio–temporal domain providing fundamental information about the building's operation and performance.

Visual analytics focuses on analytical reasoning using interactive visualizations. Schneiderman et al. [13] proposed a