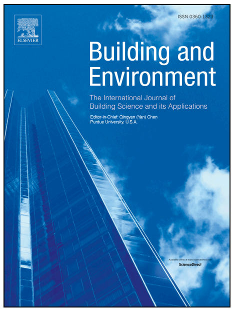
Accepted Manuscript



Making sense of building data: New analysis methods for understanding indoor climate

Angela Geronazzo, Gail Brager, Sanyogita Manu

PII: S0360-1323(17)30537-1

DOI: 10.1016/j.buildenv.2017.11.030

Reference: BAE 5183

To appear in: *Building and Environment*

Received Date: 16 August 2017

Revised Date: 8 November 2017

Accepted Date: 20 November 2017

Please cite this article as: Geronazzo A, Brager G, Manu S, Making sense of building data: New analysis methods for understanding indoor climate, *Building and Environment* (2017), doi: 10.1016/j.buildenv.2017.11.030.

Making sense of building data: new analysis methods for understanding indoor climate

Angela Geronazzo[1 *], Gail Brager[2], Sanyogita Manu[3]
[*]corresponding author, angela.geronazzo@polimi.it
(1) Dipartimento di Informatica, Elettronica e Bioingegneria, Politecnico di Milano, Milano, Italy
(2) Center for the Built Environment, University of California Berkeley, CA 74720, USA
(3) Centre for Advanced Research in Building Science and Energy, CEPT University, Navrangpura, Ahmedabad 380009, India

Keywords: Passive buildings, field data, decision trees, comfort, data representation

Abstract

This work presents a novel approach for exploratory analysis of indoor environment measurements. Building monitoring is a complex task constrained by technological and environmental limits. Monitoring processes can be long, employ multiple sensors and equipment, and may involve multiple buildings across extended periods. Once the data is collected, it's not always straightforward how to make sense of it. When faced with complex heterogeneous amounts of data, early exploratory analysis is helpful to make sense of the broad patterns, highlight data weaknesses and features, and identify directions for more in-depth analysis. Inaccuracy, incompleteness and inconsistence are common characteristics of monitored field data, which then often needs intensive pre-processing to address data quality issues and improve data analysis results. This paper describes techniques from computer science, such as decision tree induction and others, that are not normally applied to field measurements of indoor environments in buildings, and as an example applies them to a complex set of data from climate-responsive spaces in India. The findings illustrate the power of these techniques for using simplified measures as proxies to extract more sophisticated information about building operation, and for quickly identifying patterns in how different design and operational characteristics affect thermal comfort.

1. Introduction

The objective of this paper is to bridge the fields of computer science and building science, and suggest new methods for exploring field-based datasets in order to make sense of its complexities, and quickly identify patterns of performance that would help focus subsequent, more detailed statistical analysis. In particular, the methodology uses decision tree induction, and shows how to choose a set of feature attributes relevant for thermal performance analysis.

Field data of building performance can often be messy, suffering in part from issues inherent to sensors such as data incompleteness and inconsistency, presence of erroneous records, and inaccurate values. It thus requires extensive data pre-processing for the subsequent analysis to be effective and meaningful. Pre-processing of monitored data aims to improve data quality and assess its "fitness for use", which is a measure of data quality that assesses if the dataset is suitable for the purpose of the analysis [1]. Data quality comprises multiple factors including accuracy, completeness, consistency, timeliness, believability and interpretability. As a first step, data-cleaning tasks are typically applied to the datasets to isolate missing values, smooth noisy points while identifying outliers, and correct inconsistencies in the data. The subsequent step requires making sense of the dataset in order to proactively assess its fitness for use with reference to the overall research objectives. This process is particularly crucial for complex field datasets where the monitoring process might involve multiple buildings and span across multiple years. Moreover, building monitoring can be affected by a variety of limitations including technological issues connected to the deployed sensors, constraints imposed by the building owners and occupants, etc. Once the monitoring is finished, and before analysis begins, it is necessary to check if and how these limitations might have affected the dataset, and eventually the analysis results.

The exploratory methods presented in this paper utilized a dataset developed as part of the U.S.-India Joint Center for Building Energy Research & Development (CBERD) [2]. This dataset was collected during a complex field study involving multiple buildings located in different climate zones in India, where each adopted a different set of climate responsive strategies. The broad purpose of the field study was to investigate the performance of climate responsive buildings in terms of their indoor thermal environments, the potential impact on occupant thermal comfort, and the effectiveness of combining both natural ventilation and low-energy mechanical systems in "mixed mode" buildings. The monitoring process included instrumenting each building with multiple sensors, collecting data about air temperature, relative humidity, surface temperatures for selected façades, as well as microclimate information from a weather station close to the site. Each building was monitored for one year and the monitoring process was specifically tailored to each building's specific passive strategies and design features. The monitoring of the whole set of buildings