Automation in Construction 49 (2015) 1–17

Contents lists available at ScienceDirect

Automation in Construction

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

Automated daily pattern filtering of measured building performance data

Clayton Miller *, Zoltán Nagy, Arno Schlueter

Architecture & Building Systems, Institute of Technology in Architecture (ITA), ETH Zürich, Switzerland

ARTICLE INFO

Article history:
Received 2 April 2014
Received in revised form 6 August 2014
Accepted 11 September 2014
Available online 7 October 2014

Keywords:
Building efficiency
Building performance analysis
Knowledge discovery
Clustering
Temporal data mining
Occupancy patterns
Diversity factors

ABSTRACT

The amount of sensor data generated by modern building systems is growing rapidly. Automatically discovering the structure of diurnal patterns in this data supports implementation of building commissioning, fault detection and retrofit analysis techniques. Additionally, these data are crucial to informing design professionals about the efficacy of their assumptions and strategies used in performance prediction simulation models. In this paper, we introduce *DayFilter*, a day-typing process that uses Symbolic Aggregate approXimation (SAX), motif and discord extraction, and clustering to detect the underlying structure of building performance data. Discords, or infrequent daily patterns, are filtered and tagged for deeper, detailed analysis of potential energy savings opportunities. Motifs, or the most frequent patterns, are detected and further aggregated using k-means clustering. This procedure is designed for application on whole building and sub-system metrics from hierarchical building and energy management systems (BMS/EMS). The process transforms quantitative raw data into qualitative subgroups based on daily performance similarity and visualizes them using expressive techniques. We apply *DayFilter* on 474 days of example data from an international school campus in a tropical climate and 407 days of data from an office building from a temperate European climate. Discords are filtered resulting in 17 and 22 patterns found. Selected discords are investigated and many correlate with specific failures and energy savings detected by the on-site operations staff. Six and ten motif candidates are detected in the two case studies. These motifs are then further aggregated to five and six performance clusters that reflect the typical operational behavior of those projects. We discuss the influence of the parameter choices and provide initial parameter settings for the *DayFilter* process.

1. Introduction

Performance and energy data generation in the built environment is rapidly growing [1]. Modern building controls and management systems are improving in their ability to acquire and store measured data as the technology improves. This phenomenon results in vast portfolios of collected data from heterogeneous buildings. Fig. 1 illustrates a general example of various types of measured data from a conventional commercial building. Whole building performance is influenced by layers of complex measurement systems. Aggregated performance metrics or sensors are often measured or calculated at each level of this hierarchy in order to condense the exponential detailed sensor data downstream.

In addition to the increase in building performance data, there is a growing awareness of the gap in performance between building design and operations [2–6]. Multiple studies have documented and validated this phenomenon, with the most extreme mismatch finding measured energy consumption at 5 times predicted consumption for a commercial building [2]. A framework for investigating this gap emphasizes more robustly leveraging measurement data and ensuring that research in this field aligns with actual building engineering practices [3].

From the conventional operations and management side, this performance gap is generally addressed through the use of various performance analysis techniques. The literature describes two major categories of building analysis: top-down, whole building techniques and bottom-up, device-focused diagnostics [7,8].

Top-down approaches such as Energy Information Systems (EIS) are designed to qualify the building's overall performance health. They leverage the whole building and sub-systems level data to show how well a building performs compared to its peers (benchmarking) or simple tracking metrics. Despite their high-level usefulness, these techniques have a limited amount of insight and ignore much of the detailed digital data created in recently built or renovated high performance buildings [9]. In addition, they often aren't able to leverage higher frequency, sub-hourly measurements.

Bottom-up, component level approaches such as commissioning and automated fault detection and diagnostics (AFDD) are more effective at detecting the root cause of performance problems. A review of AFDD approaches for building systems diagnostics describes three general categories: Qualitative Model-based, Quantitative Model-based, and Process History Based [10]. The first two categories often require an understanding of the impact of each detailed data stream in order

* Corresponding author. Tel.: +41 446332991.
E-mail address: miller.clayton@arch.ethz.ch (C. Miller).

http://dx.doi.org/10.1016/j.autcon.2014.09.004
0926-5805/

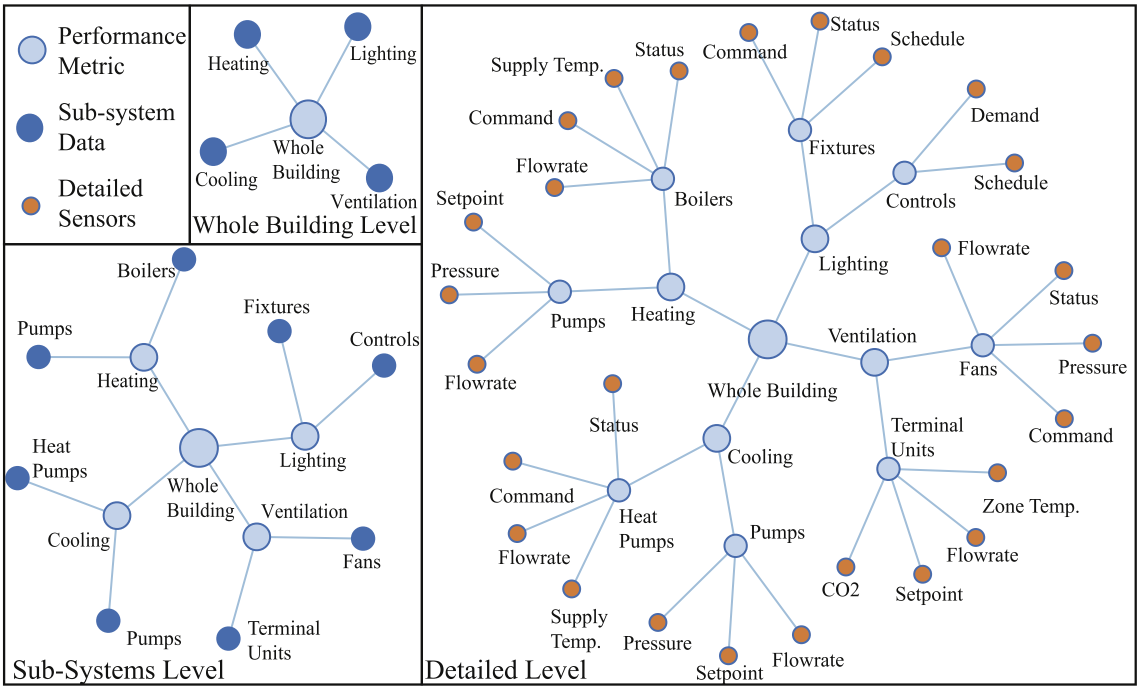

Fig. 1. Example of levels of building performance data complexity.

to set thresholds or parameters for detection of anomalies. Process history based methods rely on large amounts of empirical, measured data to create statistical models or use pattern recognition to find operational anomalies. Only process history based approaches are identified as useful with little a priori knowledge. However, they are implied to be inferior due to weaknesses such as the inability to extrapolate beyond a range of training data, amount of data necessary, and being specific to a particular dataset.

Beyond AFDD and EIS, another very active research topic is the process of calibrating the building simulation model developed in the design phase with measured performance data from the operations phase [11]. The benefits of such a process have long been lauded as key in understanding of the performance gap; first, in identifying the deficiencies in modeling engines and assumptions and, second, in investigating potential performance deviation in operations. This field was one of the first to investigate the use of *day-typing* as a means of parameter reduction of measured data for simulation feedback. Much of the literature in simulation model calibration treats measured raw data preparation and day-typing in a manual way, often ignoring the shape and magnitude patterns and relying on rules-of-thumb regarding schedule creation. These approaches add to the cost, time and lack of automation burden that calibration suffers from in real implementation.

A comprehensive study of building performance tracking was completed by the California Commissioning Collaborative (CACx) and funded by the California Energy Commission (CEC) to characterize the technology, market, and research landscape in the United States. Three of the key tasks in this project focused on establishing the state of the art [12], characterizing available tools and barriers to adoption [8], and establishing standard performance metrics [13]. These reports were accomplished through investigation of the available tools and technologies on the market as well as discussions and surveys with building operators and engineers. The common theme amongst the interviews and case studies was the *lack of time and expertise* on the part of the involved operations professionals. The findings showed that installation time and cost were driven by the need for a controls engineer to develop a full understanding of the building and systems. We interpret these results as a latent need for techniques that take into consideration the people, process, and philosophy aspects of the performance analysis equation [14]. The effort described in this paper addresses this challenge by focusing on automatically finding insight in large, unstructured building performance datasets *as part of an analysis process*.

1.1. Parameter-light exploratory analysis for building performance data

We draw inspiration from other time-series analysis and visualization applications in order to address the progression of data mining in the building industry. One emerging trend is that “data mining algorithms should have as few parameters as possible, ideally none. A parameter-free algorithm prevents us from imposing our prejudices and presumptions on the problem at hand and let the data itself speak to us [15].” This approach is known as *parameter-free* or *parameter-light* data mining. The efficacy of these algorithms has been proven comparable or better than many more complex, traditional time-series data mining approaches [15].

An emerging circumstance in the building industry is the consolidated analysis of multiple buildings or portfolios by third-party experts [16, 17]. The responsibility of managing and mining performance data is shifted from operations staff to data and building science experts who develop specific skills and efficiencies of scale. This scaled analysis and intervention addresses the previously-mentioned time and expertise deficiency and the cost effectiveness of building performance investigations. This scenario requires computation techniques which, on one hand, condense information more effectively than conventional top-down techniques, and on the other hand, requires less a priori knowledge than bottom-up, component-level approaches. Therefore, exploratory visualization and data mining techniques could be designed as part of a process to bridge these gaps. Our research combines traditional AFDD with exploration techniques such as time-series pattern recognition and visualization.

We propose a new context for the process history based methods found in the literature by testing their usefulness not as a full-scale automated fault detection and diagnostics (AFDD) approach, but as an