Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 122 (2017) 229–234

Energy Procedia

www.elsevier.com/locate/procedia

CISBAT 2017 International Conference – Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale, CISBAT 2017 6-8 September 2017, Lausanne, Switzerland

Unsupervised load shape clustering for urban building performance assessment

Jimeno A. Fonseca[a,b*], Clayton Miller[a,c], Arno Schlueter[a,b]

[a]*Future Cities Laboratory (FCL), Singapore-ETH Centre (SEC), 1 Create Way, Singapore 13892, Singapore*
[b]*Architecture and Building Systems (A/S), ETH Zurich, John-von-Neumann-Weg 9, Zurich 8093, Switzerland*
[c]*Building and Urban Data Science (BUDS) Lab, National University of Singapore (NUS), 4 Architecture Drive, 117566 Singapore*

Abstract

This paper presents a method to automatically cluster typical days of energy consumption in one or several buildings. The method is based on an optimized version of the Symbolic Aggregate approXimation (SAX) method. SAX is a data mining technique for clustering time series with recent applications in building fault detection and building performance assessment. The number of clusters and accuracy of SAX highly depends on two highly sensitive input variables, i.e., the word size and the alphabet size. We propose the use of the genetic algorithm NSGA-II to optimize the number of words and alphabet size of SAX subjected to three fitness objectives, i.e., maximize data accuracy and compression and minimize complexity. In addition, we propose the use of MAVT as selection method of the optimal solution. The methodology is applied to measured energy consumption data of three representative buildings on a university campus in Singapore. Potential future uses of the approach include advanced studies in fault detection and calibration of urban building performance models.

Peer-review under responsibility of the scientific committee of the CISBAT 2017 International Conference – Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale

Keywords: Building performance; Data mining; Daily Profile Extraction;

* Corresponding author.
E-mail address: fonseca@arch.ethz.ch

1876-6102
Peer-review under responsibility of the scientific committee of the CISBAT 2017 International Conference – Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale
10.1016/j.egypro.2017.07.350

1 Introduction

Grouping similar behavior in time-series data is a well-established process to characterize the way buildings use energy. The practice of daily load profile clustering is a specific technique to group building performance data. This procedure begins with the collection and cleaning of raw, sub-hourly sensor data from electrical meters or utility smart meters. Sub-hourly building performance data are chunked into 24-hour sub-sections. These diurnal patterns are then clustered by comparing the measurements for all samples at each hour or sub-hourly mark of the day and using distance metrics to group profiles that are roughly similar. Several unsupervised clustering techniques have been tested in their application to performance measurement data from buildings. A seminal review of clustering techniques load pattern grouping provides an overview of phases of analysis, clustering techniques, and an in-depth discussion of quantified validation and comparison of several tested techniques [1]. Other studies focused on the comparison of clustering algorithms and distance metrics as applied specifically to load profiles are cited in [2], [3]. Research outlining the application of load profiling in non-residential buildings includes its use for improving load forecasting [4] and utility grid analytics [5]. One key approach of [6] defined as the *DayFilter* process, was recently developed as a specific diurnal load profile clustering and filtering process for building energy consumption. The method utilizes the Symbolic Aggregate approXimation (SAX) clustering technique. In contrast to other clustering techniques, SAX has the advantage of speed and ease of use on large time-series data sets [7] [8]. However, SAX requires several inputs such as the size of the generated word (sub-sequence window size) and alphabet size (magnitude breakpoints) as input variables and manual selection of these values is onerous when applying the process to large groups of buildings.

This research built upon previous work by refining the use of measured load shape profiling in the domain of urban-scale building performance simulation. The approach consists of automating the SAX clustering technique with the use of multi-objective optimization for selection of the word and alphabet size in SAX. For demonstration purposes, the method is applied to hourly energy consumption of three typical buildings of the Nanyang Technological University (NTU) Campus of Singapore.

1.1 Data collection and processing

Sub-hourly metered data of the year 2015 is gathered from the BMS of each building and cleaned with a time-of-week and temperature (TWOT) model *[9]*. TWOT served to remove outliers and fill in gaps in the data. Fig. *1* illustrates typical raw data for three of the most common use types in the area of study.

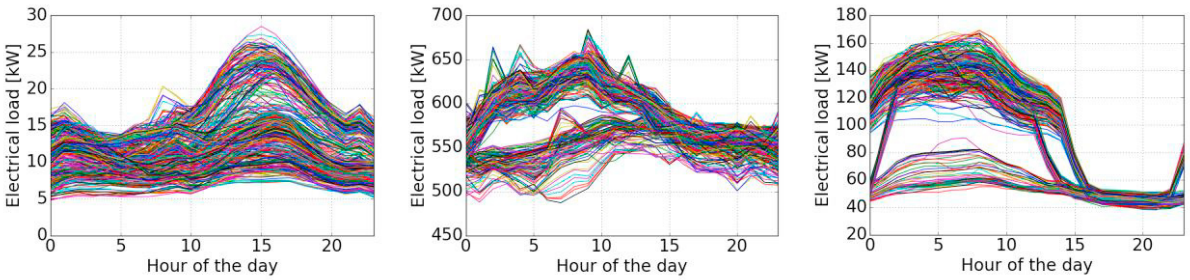

Fig. 1: Raw daily load profiles from three typical building use types. A. dorm, b. laboratory, c. office

1.2 Clustering

The SAX process converts time-series data into a set of strings that can then be grouped together for the purposes of clustering, motif and discord analysis, and finding anomalous behavior. Fig. 2 illustrates the SAX process as implemented on two days of data. Within the context of daily load profile, each day is divided into 24-hour periods starting at midnight. Each day is then further divided into segments based on the input parameter *w,* or window size. This parameter dictates how many sub-sequences are contained within a day period, and thus, how much data is averaged to create each letter of a string and how many letters long the string will be. The magnitude of the measured