Notes for Paper 2

Title:A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings

Authors:Miller, Clayton, Zoltán Nagy, and Arno Schlueter

Journal:Renewable and Sustainable Energy Reviews

Tags: Upsupervised learning; Visual analytics; Building controls and optimization

Notes

0. Summary

This paper reviewed the application of upsupervised learning in performance analysis for non-residential buildings. Different categories of unsupervised learning thchniques were simply introduced as well as their purposes and applications. Then the author introduced several domains in power performance analysis. Many published papers were involved in each domain, I think the most important thing is that the author introduced many problems which can be handled by unsupervised learning method. Finally, some challenges and opportunities were presented, two main ways to solve the challenges are to enhance the cross-disciplinary collaboration and release specific data sets. This paper tells me what algorithm can I use when I face similiar problems or what problems can the specific algorithm can solve.

1. Problem

1.1 Disadvantages of traditional thchniques

These traditional techniques focus on one building or a small, related collection of buildings, such as a campus. Many require complex characteristic data about each building, such as it geometric dimensions, building materials, the age of type of mechanical systems, and other metadata, to execute the process.

1.2 Why we need to use unsupervised learning?

If one has access to raw data from thousands, or even millions, of buildings, how can analysis be scaled in a reasonable way?

2. Categories of techniques

2.1 clustering

Purpose:
It is used automatically to generate subgroups of similar types of observations.

Application:
This application of clustering is known as load profiling.

Techniques:
Certain types of clustering techniques found in publications in this review are K-means, Principal Component Analysis (PCA), Maximum Likelihood (ML), Hierarchical, Self-Organizing Maps (SOM), Empirical Mode Decomposition (EMC), Fuzzy clustering, Support Vector Machines (SVM), and Ant Colony Clustering.

2.2 novelty dectction

Purpose:
Novelty detection is a category of techniques that finds behavior or observations within a data set that are unique and often do not conform to expected behavior.
Application:
outlier and anomaly detection
Techniques:
Principal Component Analysis (PCA), Generalized Additive Models (GAM), Classification and Regression Trees (CART),Wavelets, Fourier Transforms, Linear Discriminant Analysis (LDA), and Nearest Neighbor methods.

2.3 motif and discord detection

Purpose:
Motif and discord detection is a sub-domain unique to the investigation of time-series data sets. A motif is a subsequence of data that exhibits a pattern that frequently occurs in a data stream. A discord is a subsequence that occurs rarely and is considered anom- alous amongst the rest of the data set.
Application:
Find the motif and discord in electrical consumption data.

Techniques:

Symbolic Aggregate Approximation (SAX), Markov Models, Maximum Likelihood, Dirichlet Process Gaussian, Mixture Models (DPGMM), and Incremental Summarization and Pattern Characterization (ISPC).

2.4 rule extraction

Purpose:

automatically to find relationships between variables within a data set.

Applications:

find shared values of the variables that appear most frequently in the dataset.

Techniques:

The general field of creating Data Association Rules (DAC) is the most common form of this technique. Other methods found in this review include Markov Models and Support Vector Machines (SVM).

2.5 visual analytics

Purpose:

humans and machines cooperate using their respective distinct capabilities for the most useful results

Applications:

finding patterns in energy data and those on more general energy management with focuses on performance indicators.

3 Domains

3.1 smart meters

a network of energy meters, most often focused on the electrical power measurement of a whole building.

Load profiling

Load profiling is the process of grouping temporal subsequences of measured energy data for the purpose of characterizing the typical behavior of an individual customer.
It involves time-series clustering and feature extraction.

Customer classification

These methods often employ load profile clustering as a first step but differentiate themselves in using those features to classify accounts, or buildings, that fit within various categories.

Disaggregation

Disaggregation attempts decompose a measurement signal from a high level reading to the individual loads being measured.

3.2 portfolio analysis

A large of buildings are analyzed for the purpose of managing or optimiz- ing the group as a whole.

Charcterization

evaluate and visualize the range of behaviors and performance of the group

Classification

assigning individual buildings to subgroups of relative performance for the purpose of benchmarking or decision-making

Targeting

identify specific buildings or measures to be imple- mented in a portfolio to improve performance

3.3 operations and controls optimization

Occupancy detection

Occupancy detection using unsupervised techniques infers human presence in a non-residential building without a labeled ground truth dataset or as part of a semi-supervised approach using a subset of labeled data. This occupancy detection is then used for analysis or as inputs for control of systems.

Controls

creating a state of the best operation and energy performance for a building system such as heating, cooling, ventilation or lighting.

Energy management

help users (facilities management professionals or consultants) to understand how the building is consuming energy.

3.4 Anomaly detection

focuses on the detection and diagnostics of problems occurring within a building, its subsystems, and components.

Whole building

Whole building anomaly detection uses the electricity or heating and cooling energy supply in coming to a building to determine sub- sequences of poor performance.

Subsystem

Subsystem anomaly detection focuses on the use of a broader data set to detect and diagnose faults from a lower level.

Components

Component level anomaly detection is a bottom-up fault detection approach that focuses on determining faults in individual equipment.

4. Key challenges

need for better collaboration

different sets of terminology are used to describe similar concepts.
optimize in both the theoretical (accuracy or efficiency of the algorithm) and practical domains (how much energy was saved).

lack of open benchmarking data sets

lack of easy reproducibility.
lack of clarity regarding which is the optimal technique

Opportunities

reproducible research

release of specific data sets.
time series benchmarking data sets should be used to evaluate whether a new proposed algorithm is more beneficial as compared to previous work.

cross-disciplinary cooperation

interdisciplinary collaboration