Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 118 (2015) 1319 – 1324

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Sustainable Design, Engineering and Construction

Learning Energy Consumption and Demand Models through Data Mining for Reverse Engineering

Hariharan Naganathan[a], Wai K. Chong[a*], Nong Ye[b]

[a]*School of Sustainable Engineering and Built Environment, Arizona State University, Tempe, AZ, United States.*
[b]*Professor, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University,Tempe,AZ,United States.*

Abstract

The estimation of energy demand (by power plants) has traditionally relied on historical energy use data for the region(s) that a plant produces for. Regression analysis, artificial neural network and Bayesian theory are the most common approaches for analysing these data. Such data and techniques do not generate reliable results. Consequently, excess energy has to be generated to prevent blackout; causes for energy surge are not easily determined; and potential energy use reduction from energy efficiency solutions is usually not translated into actual energy use reduction. The paper highlights the weaknesses of traditional techniques, and lays out a framework to improve the prediction of energy demand by combining energy use models of equipment, physical systems and buildings, with the proposed data mining algorithms for reverse engineering. The research team first analyses data samples from large complex energy data, and then, presents a set of computationally efficient data mining algorithms for reverse engineering. In order to develop a structural system model for reverse engineering, two focus groups are developed that has direct relation with cause and effect variables. The research findings of this paper includes testing out different sets of reverse engineering algorithms, understand their output patterns and modify algorithms to elevate accuracy of the outputs.

Peer-review under responsibility of organizing committee of the International Conference on Sustainable Design, Engineering and Construction 2015

Keywords: Reverse engineering, Data mining, Energy consumption, Statistical Analysis, Supply-Demand Characteristics

Corresponding author: E-mail address:ochong@asu.edu

1877-7058
Peer-review under responsibility of organizing committee of the International Conference on Sustainable Design, Engineering and Construction 2015
doi:10.1016/j.proeng.2015.11.392

1. Introduction

Rapid global energy demand growth has raised concerns over energy supply shortages, exhaustion of energy resources and negative environmental impacts [1]. Electricity is considered to be a secondary energy source as it is generated by coal, petroleum, natural gas and nuclear fuel [4]. China leads the world in total electricity generation from renewable sources while the United States generates the most electricity from non-hydroelectric renewable sources [5].

Utility companies rely on historical records of base, average and peak electricity loads to project future energy use requirements. A power plant often generates enough energy to cover base or average load, and uses its excess capacities or procures energy from external sources to cover the gaps between peak loads or sudden demand surge, and base/average loads. Excess energy production is a major cause of energy inefficiency.

This research focuses on understanding why traditional analysis approaches may not fulfill the need of modern day data, and why new models are needed. The study also suggests new models that integrate the relational physical system, social and environmental factors of a building with their interoperability and interconnectivity. The research moves away from traditional data mining techniques (such as Bayesian networks, artificial neural networks, regression analysis, and decision trees) towards a new approach for computationally efficient data mining techniques used for reverse engineering.

2. Data collection and analysis

Data was collected from the Energy Information System (EIS) of several buildings at ASU. The data includes electricity, solar, heating load, cooling load, watt/square feet, etc., and are divided into fifteen minutes intervals.

Table 1. List of Data collected

Factors	Units	Data Frequency (min)	Nature of the data	Buildings
Electricity	kWh	15	Physical	119 buildings
Solar	kWh	15	Environmental	
Heating load	BTU	15	Physical	
Outside air temperature	Centigrade	15	Environmental	
Heat index	Centigrade	15	Environmental	
Cooling load	Ton hours	15	Physical	
Watts/ Sq. ft.	Watts	15	Physical	
Occupancy	Persons	60	Social	
Types of occupants	Persons + Types	15	Social	
Building contents	Description	15	Physical	

2.1 Regression analysis and results

The data collected are naturally large and complex as most of them are divided into fifteen minutes interval. A simple regression analysis is conducted to determine the difficulties of adopting traditional techniques on a large data set, and to emphasize the need of new data mining technique for better prediction accuracy. Figure 1 shows the linear relationship (R^2) between human counts in the building and other factors, such as electricity, cooling load, heating load and Watt/Sq. ft. The analysis is only done for the months of March, June and October, 2013. The analysis is also conducted for every hour. The analysis exhibits interesting results with most of the R^2 ranging above 60% in all three months. The Watt/Sq. ft. curve and electricity consumption curve coincide except for June. It is difficult to predict from the results that human occupants have nominal impacts on the building energy consumption since the values range between 20% and 85%. Thus, linear regression fails to provide a clear perspective on the structural relationship between the factors on large and complex data set.

The regression analysis technique is typically used for modeling linear relationships, whereas we expect system models of energy consumption/demand to be nonlinear due to the interconnectivity and interdependencies of system factors. The decision and regression tree technique has difficulty of effectively learning structural relationships of variables from a large data set. The analysis helps to identify the relationship between the human counts and their impacts on electricity user, cooling load, heating load and Watt per square foot. Also, time is an important factor to be considered as these statistical techniques consumes more time for large data sets.