Mid-year Progress Report

Topic: Vetting the Energy and Security of Smart Buildings with Data Science

Zhen Li, MyVan Vo, Dinuka Gallaba

Date: 3/1/2019

Introduction

The Cyber- Physical systems has wide range of applications in civil, transportation and energy infrastructure. While infrastructure are becoming smarter the vulnerability to cyber-attacks has been increased. Our goal in this research project is to investigate the statistical and data mining techniques that can be implemented in buildings with “smart” sensor and control systems.

Project Process

Our initial approach in this project was to identify and clean the data set that need to be analyses and used in future predictions. We collect the data from the ReNEWW House at Purdue. We were able to categorize the data in to two sections, energy and water consumption/ production (daily) and living conditions (temperature and humidity). Since all the data was collected with respect to time, we address the problem with time series analysis. When we analyze the data from energy and water consumption we were able to observe some anomalies. We have decided to investigate on anomaly detection methods using data science approach (Figure 1)[1]. At the same time we decided to work on some data mining techniques to monitor energy consumption of the building [2]. For the temporal data we used Autoregressive Integrated Moving Average (ARIMA)[3] analysis. We found a review article that address energy management using data science that gave us good insight to carry-on our project.

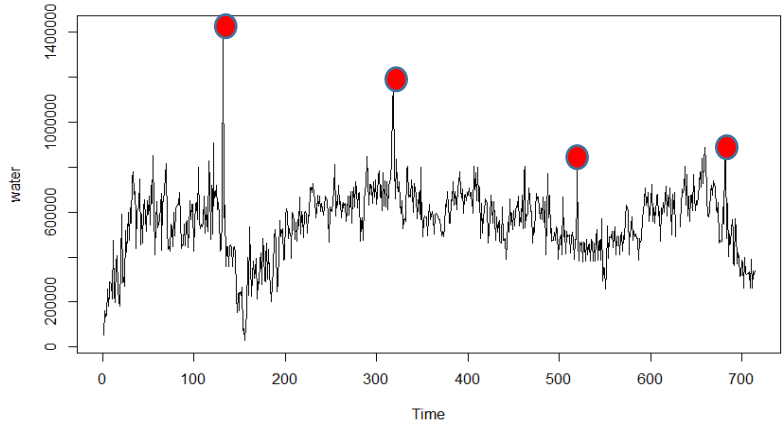

Figure 01 : Anomaly in water Consumption

Meetings

Two conference calls and expected to meet at Purdue University before the JSM Conference to prepare the presentation/poster.

Conferences

Abstract Submitted

1. Poster Presentation: 2019 Joint Statistical Meetings (JSM) at Denver Colorado

   (July 27 –August 1) Abstract Number 307118.

2. Symposium on Data Science and Statistics in Bellevue, Washington, May 29 - June 1

References

[1]. Zucker, G., Habib, U., Blöchle, M., Wendt, A., Schaat, S. and Siafara, L.C., 2015, September. Building energy management and data analytics. In 2015 international symposium on smart electric distribution systems and technologies (EDST) (pp. 462-467). IEEE.

[2]. Capozzoli, A., Piscitelli, M.S. and Brandi, S., 2017. Mining typical load profiles in buildings to support energy management in the smart city context. Energy Procedia, 134, pp.865-874.

[3] Molina-Solana, M., Ros, M., Ruiz, M.D., Gómez-Romero, J. and Martín-Bautista, M.J., 2017. Data science for building energy management: A review. Renewable and Sustainable Energy Reviews, 70, pp.598-609.

Attachment: Conference Abstract

Vetting the Energy and Security of Smart Buildings with Data Science

Cyber Physical system has various applications in civil, transportation and energy infrastructure. With the development of sensor technology and communication, all of the above infrastructures become more intelligent and self-sustained. Meanwhile those facilities have also become vulnerable to cyber attacks due to the gaps in the cyber security and physical protection. Our goal is to use a smart building as a platform to develop data mining technologies in intelligent buildings and to apply the technologies to common Cyber Physical Infrastructures (CPIs), which will have significant influence on the system security. The chosen residence for the case study is a research level residence at Purdue University. The installed sensor network provides continuous data regarding the utility consumption like water and energy in real or near-real time. With the accumulated data over a period of two years, the researchers used time dependent data mining techniques to figure the relationships between and within different aspects of building performances. Firstly, time series analysis and domain analysis like Discrete Fourier Transformation were used for having an overview of the data profile. Next, data was divided in different clusters based on the time. By analyzing the data with the statistical methods like Markov model and autoregressive integrated moving average methods for each cluster, it is concluded our approach could effectively predict the energy consumption. Moreover, some correlation of utility usage, weather and human activity were generated based on the mathematical and statistical analysis. Further study will be focused on more data types and larger data size to achieve the goal that secure the CPIs with data science.