CEE 498DS Project 11: Building Energy Predictions - Project Report

This manuscript (permalink) was automatically generated from cathyxinchangli/cee498ds-project11@e850b87 on December 4, 2020.

Authors

- **Xinchang 'Cathy' Li**
  XXXX-XXXX-XXXX-XXXX · cathyxinchangli
  Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

- **Benjamin Smakic**
  XXXX-XXXX-XXXX-XXXX · mkbenja
  Department of Aeronautical & Vehicle Engineering, Royal Institute of Technology, KTH; Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

- **Zhiyi Yang**
  XXXX-XXXX-XXXX-XXXX · zhiyiy2
  Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

- **Mingyu Sun**
  XXXX-XXXX-XXXX-XXXX · TBD
  Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

Abstract

This is the abstract. Testing to see if it will show up.

By utilizing modern electric meters, it possible to collect and store enormous amount of data about household energy consumption. This data can be used to predict energy consumption and help energy providers manage energy (electricity) output and plan for energy peaks/lows.

Literature review

“Forecasting Residential Energy Consumption: Single Household Perspective”

In the paper “Forecasting Residential Energy Consumption: Single Household Perspective” (Zhang, Grolinger & Capretz 2019), the authors attempt to predict energy consumption in residential households, with focus on single households.

According to the authors it is more difficult to predict single household energy (electricity) consumption, compared with e.g. workplace energy consumption. The reason is that single households often differ in energy consumption patterns while workplace patters tend to be more similar. Also, if big workplace buildings or multi-family residential buildings are analysed, any anomalies tend to cancel each other out (with a big enough dataset).

Data set

The data set used originates from an electricity provider in London, Ontario, Canada. It consists of hourly smart-meter readings of electricity usage (kWh) of 15 households between 2014 and 2016.

Firstly, the data set used might not be sufficiently broad. Tracking only 15 households will most likely not capture a variety of electricity consumption patterns. However, it was deemed enough in this case. Furthermore, the data set comes from one city with a certain climate, which means that different environmental prerequisites are not considered. Perhaps using other cities from a different part of the world would lay a foundation for a more advanced ML-algorithm (Machine Learning algorithm). As it is now, the ML-algorithm might be inaccurate for other parts of the world. In addition to this, Zhang, Grolinger & Capretz state that residents of London, Ontario, Canada tend to heat their homes gas heating systems, which affects the electricity consumption drastically.

Secondly, the data set used is pre-processed in different ways. Any missing readings of electricity consumption is replaced with the average value of the previous reading and the next reading, missing weather condition is replaced with the weather condition of the previous hour etc. This is perfectly good way of replacing missing data. However, the authors do not give an explanation as to why this method was chosen, if there are any consequences and if there are other methods of making the data set complete.

Exploratory Data Analysis

The EDA performed by the authors is illustrated in the form of electricity consumption graphs and heat-maps. The patterns show that most households live regular lives (the authors do not define what

“regular lives” mean, though it can be understood by the context). However, there are some exceptions where irregularities occur (e.g. empty homes during the summer when consumption otherwise is the highest), which significantly reduce the precision. This could be improved by adding vacancy detection which could be implemented in the ML-algorithm (which raise privacy concerns). Zhang, Grolinger & Capretz realize that the top three most important variables (i.e. the variables that correlate the most with the output energy consumption) are “temperature”, “hour of the day” and “peak index”. Peak index is a variable that captures important energy usage peaks, such as peak hours, days, seasons etc.

Prediction model and results

The authors used Support Vector Regression (SVR) to predict the energy consumption. SVR is a supervised machine learning algorithm, which means that it is compares an input with an output and is trained by comparing predicted results with true results. It was chosen due to time and computational hardware constraints. No other evaluations or comparison of other machine learning algorithms were made by the authors, so it is difficult to understand why SVM is faster and require less processing power. Zhang, Grolinger & Capretz present the results for home #1 in figure 1. They managed to predict electricity consumption well. The most inaccurate parts are peaks that arise due to random variations. Furthermore, the authors present a table with results for all 15 homes. According to them, time-based splitting is used to check parameter stability over time, which makes the algorithm more accurate. In this case however random sampling performs better in cases where some residential customers have irregular and uncertain patterns. These uncertain patterns make time-based splitting more inaccurate over time. Therefore, both methods are employed. Lastly, mean absolute percentage of error (MAPE) was utilized to measure the performance of the algorithm, which is a widely used performance metric.

Figure 1: Observed electricity consumption compared to predicted electricity consumption for house #1 (of 15) (Zhang et al.). Insert figure Figure 2: Performance results of the predication model for all homes. Insert figure ### Conclusion The biggest strength of the paper is the execution of the chosen methods to achieve desired results. It is a relatively successful attempt at predicting single households, which tend to be more unpredictable compared to multi-family or corporate residential buildings. The biggest weakness is the justification for the chosen methods. The authors do an excellent job of utilizing the chosen methods, but there is little thought put in to why these methods where chosen or why they did certain things. By including a more extensive evaluation and justification for method choices, the target audience and other researchers in the same field can understand better and continue the research. However, it makes the paper longer and more complex, which can be negative for the readers and the target audience.

References