Machine Learning for Customer Energy Segmentation and Forecasting

Dominique Smith, Kristen Jaskie, Dr. Andreas Spanias

Abstract — **With smart meters becoming more prevalent, household and commercial energy usage are monitored constantly. Using this data, we track how much energy each house or building is using and segment these into clusters, where energy use is similar. Evolutionary clustering allows us to track changes in customer segments over time. After performing customer segmentation, LSTM neural networks can be used to forecast energy use for each customer segment. This forecast can help in the development of energy efficiency programs, the planning of energy grids, consumer targeting, etc.**

Index terms: machine learning, energy, load forecasting, customer segmentation, neural networks, adaptive clustering

I. Introduction

With smart meter usage, every house and building is sending out data about energy usage every hour, or even every fifteen minutes. This provides an extremely large amount of energy load data measured over numerous time steps. Analyzing this energy data can be useful to economists, utility companies, and regulators to forecast energy loads and trends in energy usage. We would like to be able to use this analysis to answer questions like how does the increase of solar panels and electric cars change energy usage? How does residential energy usage change now that more people have flexible work schedules? [1]

In this paper, we introduce a new forecasting algorithm that first compresses and then clusters the energy load data. We believe this pre-processing will both improve overall forecasting and help in early identification of changing energy usage trends.

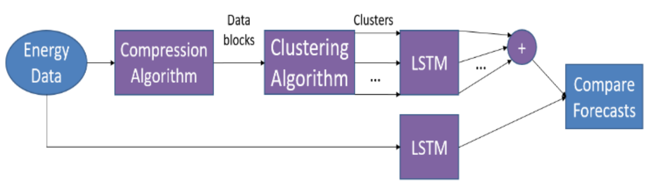

Figure 1: Block diagram for the energy analysis.

To compress the data, we define data blocks as a certain number of weeks, either four weeks for a month or thirteen weeks for a quarter. Vectors are created for each day in the block, so a four-week block will have 28 vectors of size 24x1, one sample per hour. A block then takes the average hourly value from each day of the week. Following the example, the four-week block will be represented by a 24x7 matrix, with each column being the average from the four Sundays, four Mondays, and so on. This creates smoother data with less noise [1].

Once the data is compressed, we feed the compressed data blocks into our clustering algorithm. Since energy usage can change over time and people can move in or out of an area, we propose using an adaptive evolutionary clustering algorithm. [2] proposes an adaptive k-means algorithm that does not need an initial number of clusters and can adapt to allow new data points and reshaping of the clusters. [3] gives an evolutionary clustering algorithm that focuses on tracking clusters. At each time step, the proximities of the data points are calculated, and a forgetting factor is used to estimate what the next state will be, with static clustering occurring at each step. Taking from [2] and [3], we will use a Fast and Adaptive Customer Segmentation clustering algorithm that allows clusters to grow and shrink over time and change the number of clusters [1]. The figure below shows some preliminary customer segmentation using static clustering.

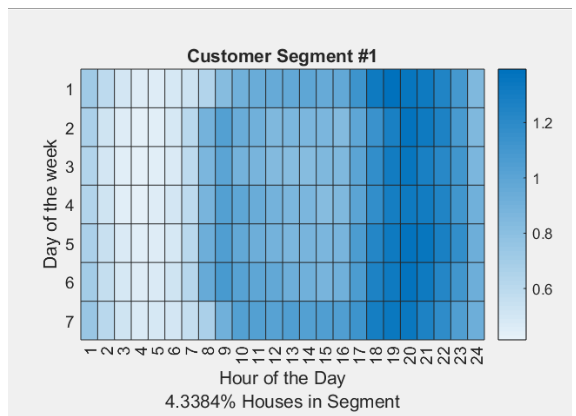

Figure 2: Heatmap of customer segment energy use [4].

Once we have the customer segments, we will use a long short-term memory (LSTM) neural network to predict energy usage for each segment. A LSTM is a type of recurrent neural network great for time-series data because the network uses previous time steps to learn and forecast [5]. [6] uses an altered LSTM for forecasting events at Uber, showing that LSTM's are a good choice to use for forecasting events. We show that using a modified LSTM on compressed clustered data provides improved energy load forecasting.

References

[1] K. Jaskie, J. Marvin, A. Spanias, "Fast and adaptive energy load customer segmentation", Technology ID M19-147P, 2018.

[2] S. K. Bhatia, "Adaptive K-Means Clustering", in 17th Intern. Florida AI Research Society Conf. (FLAIRS), Miami Beach, Florida, May 2004, pp. 695-699.

[3] K. S. Xu, M. Kliger and A. O. Hero, "Adaptive evolutionary clustering," Data Mining and Knowledge Discovery, 2014.

[4] UK Power Networks, *Low carbon london data*, London, England: UK Power Networks, 2014. Accessed on: June 3, 2019. [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

[5] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network" *arXiv preprint arXiv:1808.03314*, 2018.

[6] N. Laptev, J. Yosinski, L. E. Li, S. Smyl, "Time-series extreme event forecasting with neural networks at Uber", *ICML 2017 Time Series Workshop*, 2017.

This project was funded in part by the Nation Science Foundation, award number 1540040.

[7] U. Shanthamallu, A. Spanias, C. Tepedelenlioglu, M. Stanley, "A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications," Proc.s 8th IEEE IISA2017, Larnaca, August 2017.

This project was funded in part by the Nation Science Foundation, award number 1540040.