Predicting energy consumption of a building based on historic usage rates and weather data

Tamara Šumarac
Department of Physics
Harvard University
Cambridge, MA 02138
tsumarac@g.harvard.edu

Abstract

One of the biggest problems facing the world today is global warming. With electricity and heat production being one of the biggest contributors to the greenhouse gases, there is a demand in constructing a more energy-efficient buildings and retrofitting the old ones. An estimate of how much energy would improved building consume relative to the old one can incentivize investors to pursue these renovations. For this purpose it is useful to have an accurate prediction of how much energy would a building consume in future given how much that same building was consuming in the past. In this project I would like to use the data [1] of historic usage rates and observed weather over one year period across hundreds of buildings in order to predict how much energy would these buildings consume in the future.

1 Introduction

Emission of greenhouse gasses into the atmosphere is known as the main cause of global warming. Through human activity, specifically burning fossil fuels, deforestation and farming, large amounts of these gasses are released into the atmosphere every year, causing our planet to warm up at a rapid speed. Energy generated by fossil fuels is mostly used for electricity and heating, and for the purposes of optimizing this energy consumption there is a demand in constructing a more energy-efficient buildings and retrofitting already built ones. To motivate investors to pursue renovations needed to optimize energy consumption of existing buildings, an accurate estimate of how much energy would improved building consume relative to the old one is important. One can make these estimates using programs based on physical principles that take as input climate data and many building features data such as geometry, materials used for constructing, positioning of the building and its orientation. These programs have been greatly utilised, however they can become very costly to handle when one wants to define large number of retrofits [2]. An alternate approach to estimating energy consumption would be to make these estimates using historical usage data. In this method, building features, building historical usage data and climate data are used together to predict building's energy consumption by applying a learning process. In this project I would like to use the latter approach to estimate buildings energy consumption. In subsections to follow I will describe the data I will use for the project along with the objective function and learning algorithm that I plan to apply to it.

1.1 Data

In this project I would like to use the data [1] of historic usage rates and observed weather over one year period across hundreds of buildings in order to predict how much energy would these buildings consume in the next two year period across the areas such as chilled water, electricity, hot water and

steam meters.
The input features for the learning can be split into three categories: consumption features, building features and weather features, where the output of this model, target data, would be the energy consumption in units of kWh. Consumption features will include meter feature indicating which area are we learning the consumption of (chilled water, electricity, hot water or steam meters) and timestamp feature that will indicate at which point in time was the meter reading taken. Building features will include primary use feature, an indicator of the primary category of activities for the building, square feet, year built and floor count. Weather features will include air temperature, cloud coverage, dew temperature, precipitation depth, sea level pressure, wind direction and wind speed.

1.2 Evaluation criteria

Since we would like to predict energy consumption of a building, which is a number, I will use mean squared error (MSE) as a natural evaluation metric:

$$\frac{1}{n}\sum_{i=1}^{n}(g_i - y_i)^2 \tag{1}$$

where g_i is our estimated value and y_i is the target value.

1.3 Learning algorithm

There is a vast literature of different approaches used for building energy consumption predictions. In review [3] authors have discussed the performance of Artificial Neural Networks, Support Vector Machine and hybrid methods for load forecasting. In [2], in addition to previously mentioned techniques, authors have also discussed application of Gaussian distribution regression models and hierarchical clustering for this problem. In [4] authors have investigated the effectiveness of Long Short-Term Memory (LSTM) technique and its variations in prediction of electricity consumption data.
For this project I would like to build a model for this data using two different methods: Linear regression and LSTM Neural Networks.

2 Methods

2.1 Data processing and data selection

For this project I did two types of data processing and selection: one for the features data and one for the target data.
A significant proportion of features data was missing in the downloaded data files. In order to have the largest number of data for the training, I filled these missing elements for all the input features. For the weather data features, I filled the missing data by taking the average of the closest non missing values of that feature, whereas for the building data I filled the missing data for the feature with the mean of that feature for that building's primary use group.
For the target data, given that consumption data has large daily and weekly variations, and given that this data should be what we predict, I chose, as a first step, not to include the data of the buildings that have missing values for more than a day. Also, as a first step, I ended up working only with buildings whose primary use is educational, which ended up being 118 buildings that I used for learning in total. As a next step in this project, I would extend this to different types of buildings using one-hot encoding for the primary use feature.

2.2 Feature encoding

For this report I did two different types of encoding and I compared how these different encodings affect the goodness of the fit.
First encoding is what I called simple encoding in my code [5]. For this encoding I replaced the timestamp feature with hour, month and year features. After this I normalized all the input features, including these and this was the input for my learning algorithm.
Second encoding, that I called polynomial encoding, was inspired by Moon et al [6]. Given that

energy consumption has clear daily variations, if we encode day as a numeric feature, then even though 0000hrs follows right after 2359hrs, they will numerically look like they are far apart. This encoding solves this problem by substituting day feature with day_sin and day_cos feature, with frequency that corresponds to a 24h. I used this type of encoding for day, week and year feature with frequency of 24h, 7 days and 365 days respectively in order to reflect the cyclical nature of electricity consumption. I also added additional feature called is_weekday to handle lower amplitudes of energy consumption over weekend relative to weekdays. Wind_direction feature in units of degrees has similar issue as described above, where 0 and 359 degrees are numerically far away but physically they are close. Thus I substituted wind_speed and wind_direction features by combining them together into wind_x and wind_y features. For year_built feature instead of making it numeric, I grouped it into categories of 40 year period given that histogram of year_built feature showed 3 peaks across 40 year period. After these encodings, I normalized all the input features before sending it to the learning algorithm.

2.3 Learning algorithm

For this report I used linear regression with Adams optimization as my learning algorithm. Two hyperparameters that I tuned were learning rate and batch size. I divided all the data for training, evaluation and test in roughly 70%-20%-10% ratio.

3 Results

For the milestone report I wanted to compare how two different encodings, as described in feature encoding section, affect the predictions of the linear regression model. The comparison of prediction is given in Figure 1 below.

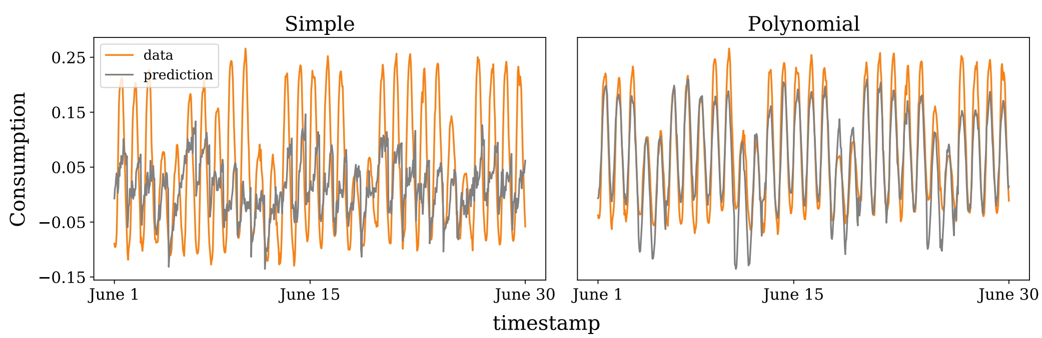

Figure 1: Linear regression model fit for the month of June 2016 of averaged electricity consumption of 118 educational buildings for two different types encodings as described in feature encoding section. The values of electricity consumption presented here are normalized w.r.t. training data mean and standard deviation values.

Test data presented in the figure corresponds to the electricity consumption data for the month of June 2016 averaged over 118 educational buildings. As expected, the more involved encoding, that takes into account cyclical nature of electricity consumption, gave better prediction of the test data compared to encoding that just normalizes input features. The polynomial encoding was able to capture hourly variations well, as well as the fact that over weekend, electricity consumption was lower compared to weekday consumption.

Linear regression model with polynomial encoding performed well on this test data in terms of capturing daily and weekly variation frequencies. It is, however, not performing that well when it comes to the variation amplitudes. This is especially visible for the weekend predictions where the model overestimates the electricity consumption, while during the weekdays this model underestimates the value of the amplitude variation. With linear regression model, we didn't have any memory feature of what the consumption value was in the past, and I believe that this type of information could improve the fit. The next step for the project would be to incorporate this type of memory feature through LSTM neural network and to see if the architecture that this type of network provides can improve energy consumption fit.

References

[1] URL: `https://www.kaggle.com/c/ashrae-energy-prediction`.

[2] Saleh Seyedzadeh et al. "Machine learning for estimation of building energy consumption and performance: a review". In: *Visualization in Engineering* 6 (2018).

[3] Mohammad Azhar Mat Daut et al. "Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review". In: *Renewable and Sustainable Energy Reviews* 70 (2017), pp. 1108–1118.

[4] Daniel L. Marino, K. Amarasinghe, and M. Manic. "Building energy load forecasting using Deep Neural Networks". In: *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society* (2016), pp. 7046–7051.

[5] URL: `https://github.com/TamaraSumarac/6.862_MachineLearningProject`.

[6] Jihoon Moon et al. "A comparative analysis of artificial neural network architectures for building energy consumption forecasting". In: *International Journal of Distributed Sensor Networks* 15.9 (2019).