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. Production of carbon dioxide and other greenhouse gases is heating up the atmosphere and this could be very dangerous for human life. 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 for our growing human population and increasing number of commercial buildings. In addition to creating new buildings, there is an interest in renovating existing buildings into a more energy-efficient 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 thousands of buildings in order to predict how much energy would these buildings consume in the future. Buildings are labeled by their primary use (e.g. office, educational building), square footage and year they were built, and we have historic data on the usage of these buildings across the areas such as chilled water, electricity, hot water, and steam meters. In addition to building usage data, we also have a data on weather features such as temperature, humidity and wind, that could be linked to the building usage data given building site id.

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. These changes in temperature can cause severe droughts, floods and rising sea level which would make our planet less optimal place for human life and life in general. With burning of fossil fuels for energy generation being the activity with the biggest greenhouse footprint, there is a need to both find alternate ways of generating this energy and the need to optimize its consumption. 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 the 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 thousands 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 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, sea level pressure, wind direction and wind speed.

I would like to encode all the features using numeric encoding (normalize them) except for the meter, primary use, year built and timestamp feature. For meter and primary use I would like to use one-hot encoding, for year built I would split it into categories spanning one decade time periods and do numeric encoding on the categories. Finally, depending on the algorithm I use I would do further encoding on the timestamp feature. This is described in more detail in learning algorithm subsection.

1.2 Evaluation criteria

Since we would like to predict energy consumption of a building which is a number, one natural evaluation criteria would be to use root mean squared error (RMSE) metric. However, in the case of predicting energy consumption, one might have a preference of not underestimating usage more than overestimating it in which case root mean squared log error (RMSLE) would be a good metric choice, defined as:

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

where g_i is our estimated value and y_i is the target value. My plan would be to use RMSLE as a metric in this project.

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 Long Short-Term Memory Neural Networks.

For linear regression, I would like to use all the input features as described in the data subsection except for timestamp feature. I could either do polynomial encoding or I could break this one into 3 additional features: month, day of the month and hour in the day in order to make the data look more linear to make this method effective.

LSTM Neural Network is a technique that attracted my attention the most out of all techniques I read about. Their internal structure, based on nodes forming a directed graph over temporal sequence, could be a good choice to model this data based on timestamp sequence.

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.