CS230

Predicting increased electricity consumption during severe weather events using Deep Neural Networks

Vasu G.Patel
Civil and Environmental Engineering Department
Stanford University
vgpatel1@stanford.edu

Amoh K. Benjamin
Civil and Environmental Engineering Department
Stanford University
amohk@stanford.edu

Abstract

The unprecedented levels of climatic changes has led to increased unpredictability of the users' electricity consumption, causing Energy Management Systems (EMS) to make an unplanned shutdowns of electricity grids. Accurate energy/electricity consumption prediction is an essential component in ensuring reliability of the grid and providing steady electricity output during severe weather events. *Machine learning (ML)* and *Deep Learning (DL)* methods is recognized as the best suited approach for understanding co-relations between weather conditions and electricity consumption. However, there has not been the avalanche of the usage of ML methods for accurately predicting increased electricity consumption. Our project focuses on using the ML, DL methods to predict the increased electricity consumption from the baseline consumption during extreme weather conditions.

1 Introduction

Energy usage over the years have been increasing rapidly due to the economic growth and human development. The high proportion of energy consumed by buildings leads to major environmental problems causing climate change, air pollution, thermal pollution, among others, which deploys a severe impact on the existence of mankind[1]. With all these, extreme weather conditions are expected to increasingly and critically influence the efficiency and economy of all energy systems The demand to understand the energy consumption due to different factors have enthralled researchers, making researchers look into machine learning techniques[2,3,4].

The majority of studies estimating electricity demand, particularly residential cooling and heating loads, have focused over the long-term changes considering climate change projections [5]. The findings emphasize that climate change increases cooling loads and reduces heating loads, which can possibly lead to a net increase in electricity demand. A recent study focusing on California concluded that under high temperature (climate change scenario of Representative Concentration Pathway (RCP) 8.5), residential electricity demand can increase between 47 and 87% (depending on the electrification levels) during 2020 and 2060[6].

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

The focus of this work is to provide a good prediction of the increased electricity consumption as a result of severe weather events which will be useful for future forecasting of electricity consumption and support operational planning in power systems.

2 Related work

Several studies have been conducted to predict energy demand previously. In the past, statistical techniques were used mainly to predict energy demand. Munz at al. predicted a time series of irregular patterns using K-means clustering [7]. Kandananond used different forecasting methods - auto regressive integrated moving average (ARIMA), Artificial Neural Network (ANN) and multiple linear regression (MLR) - to predict energy consumption[8].

However, due to the irregular patterns of energy demand, statistical techniques have limited performance performance and many model of prediction using machine learning methods have been investigated. Dong et al. predicted the demand of building energy using SVM with consumption and weather information [9]. Ekici and Aksoy predicted the building energy needs with properties of buildings without weather conditions [10]. In this project, we provide the increased electricity demand during extreme weather conditions. We use monthly electricity consumption data and various weather features as an inputs to train our deep neural networks.

3 Dataset and Features

For training our model, we have considered California's weather [11] and electricity [12] data from July 2015 to Nov 2021. Input data consists of daily weather/climatic conditions being recorded by several stations across California. Daily weather data contains total 13 features, namely: temperature [average (TAVG), min (TMIN), max (TMAX)], wind [total wind movement (WDMV), peak gust wind speed (WSFG), average wind speed (AWND), direction of peak wind gust (WDFG), peak gust time (PGTM)], precipitation (PRCP), snowfall (SNOW), snow depth (SNWD), water [water equivalent of snow on the ground (WESD), water equivalent of snowfall (WESF)].Output data consists of daily electricity consumption (in MWh) from all the sectors including residential, commercial and industrial.

3.1 Data Pre-Processing and Feature Selection

Weather and Electricity data is required to be pre-processed before it can be fed into the model for generating relevant results.

3.1.1 Weather Data Pre-Processing

Daily weather data consisted recordings (feature values) from several stations which might not necessarily have a non-zero value for each weather feature. Hence, we carried out a data pre-processing step by taking average of all non-zero feature recordings across all stations in the California for a given day to come up with a single value for each feature for a given day. Following this, out of total 13 input weather features, we carried out a feature selection to narrow down our weather input parameters to come up with a final set of features namely, 'AWND','PRCP','TAVG','TMIN','TMAX'. These weather features showed the maximum impact on the daily electricity consumption. We observed that some values weren't numbers so we replaced these values with zero.

3.1.2 Electricity Data Pre-Processing

Daily electricity consumption (in MWh) is a real valued number. To simplify our model complexity and improve the accuracy of model predictions, we compute the normalized Delta electricity demand using following equation:

$$Delta = \frac{DailyDemand - Mean}{StandardDeviation}$$

where mean and Standard Deviation are computed for entire electricity dataset from July 2015 to Nov 2021. We choose data normalization using this method to make our model robust for its applicability to other locations, duration (any given year) and variability in electricity consumption values.

Our Deep Neural Network intends to predict the output using SoftMax multi-class classification to have the required level of accuracy. Therefore we further carry out electricity data pre-processing by classifying the computed delta into five and ten classes.

Class	Delta	Lower Demand (MWh)	Upper Demand (MWh)
0	-1.99117014	534646	659828.6
1	-0.871252018	659828.6	785011.2
2	0.248666104	785011.2	910193.8
3	1.368584226	910193.8	1035376.4
4	2.488502348	1035376.4	1160559

4 Methods and Experiments

We used Deep Neural Network for training our dataset into ten class or five class classification as explained in below sub sections. Following data pre-processing, total data-set is split into train and test in the ratio of 95:5. These datasets are clean and normalized to obtain x train, y train, x test and y test, as illustrated in Fig 1. In addition, we carry out data augmentation using Synthetic Minority Oversampling Technique (SMOTE) to improve the distribution of dataset, reduce the variance and obtain higher accuracy.

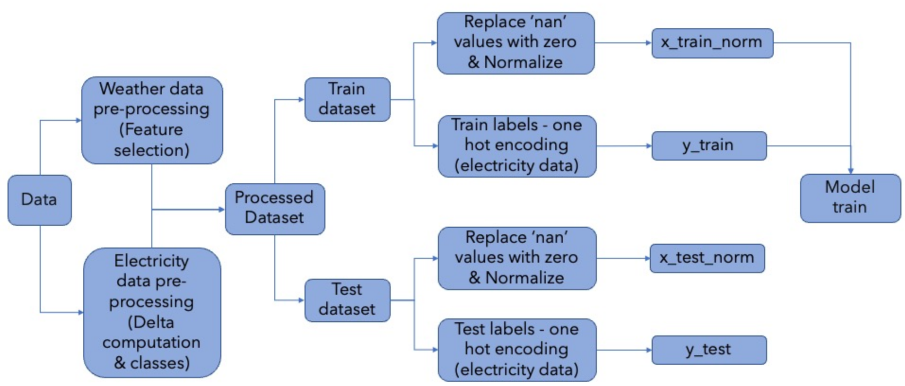


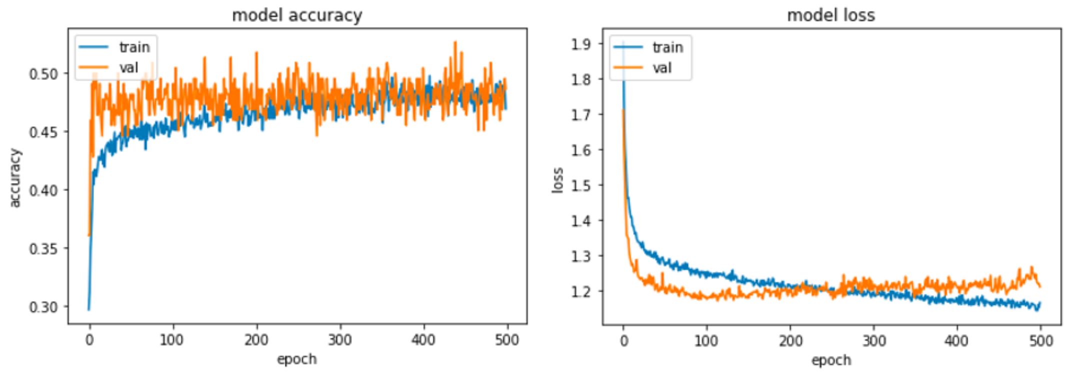
Figure 1: Data pre-processing Model pipeline

4.1 Model Architecture

We implement deep neural network on original dataset of 2336 data points considering both ten class and five class classification. Our Deep Neural Network architecture consists of five hidden layers with ReLu activation function and a final SoftMax output function providing the probabilities for each classes. Each hidden layer consist of Dropout regularization with 0.2 probability to avoid over-fitting of data during training. Hyper parameter tuning was performed and we found the optimum learning rate is 0.0008 and five hidden layers produces the best accuracy, increasing number of hidden layers beyond five does not significantly improve model performance and even decreases model performance after 7 hidden layers. Number of neurons in the hidden layers were either 128, 64 or 32. For compiling our model we use categorical cross entropy as our loss function (for multi-class classification problem), Adams optimizer, and accuracy as our metric.

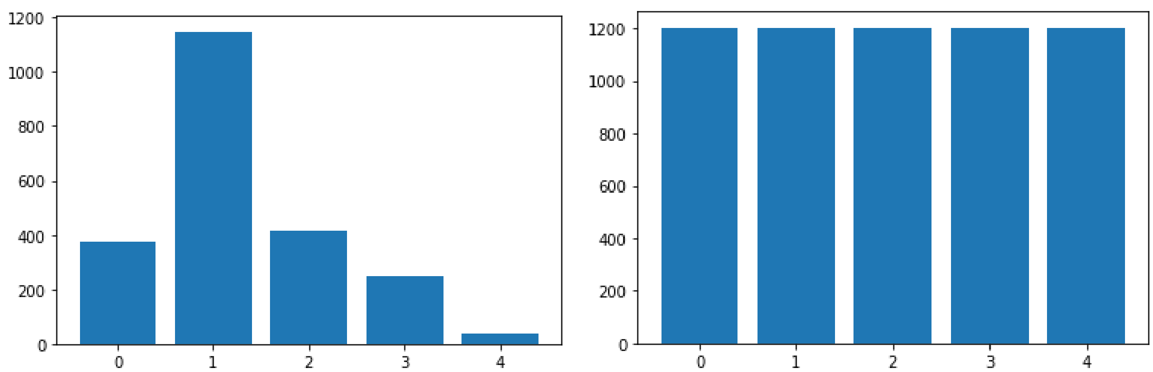
4.2 Hurdles with Ten class classification

Training and testing on ten class classification leads to very low accuracy (training accuracy: 48% and validation accuracy: 49%) and high variance is observed because of very small dataset to train on. From the Fig. 2 it is found that validation accuracy is higher than training accuracy and the validation



(a) (b)

Figure 2: Ten class classification accuracy and loss



(a) (b)

Figure 3: Data distribution for each class without (a) and with (b) SMOTE implementation

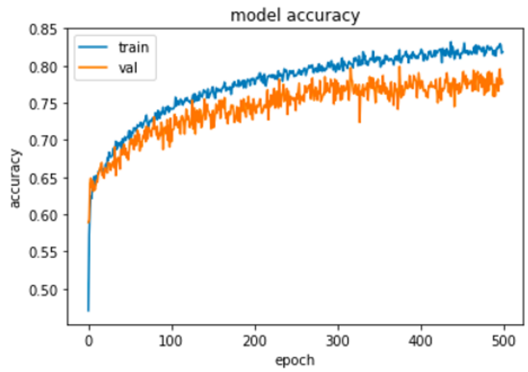
loss is initially lower than training and it goes higher than training later after 250 epochs. Hence, we decided to change our classification to five classes and implement SMOTE to increase our dataset and generate evenly distributed data.

4.3 Data Augmentation using SMOTE

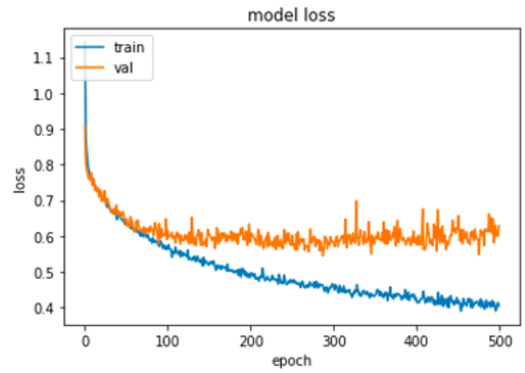
Due to limited availability of input and output database in the same frequency (hourly vs daily) and for the same duration (i.e. range of years), we could collect at maximum of common 2336 data points for California's weather and electricity consumption starting July 2015 to Nov 2021. This dataset had unequal distribution of electricity consumption classes (Fig. 3 (a)). Hence, data augmentation is implemented using SMOTE to generate uniformly distributed classes (Fig. 3 (b)). Using this technique it generates equal distribution of data points for each classes and the dataset increased from 2336 to 6015. Implementing SMOTE helped has a positive effect in improving model performance (accuracy) and reducing the over-fitting (i.e. improved generalization).

5 Results and Discussion

Before implementing SMOTE for the five class classification, our model achieves higher accuracy of nearly 48%. Running this model on test data set reveals the following precision, recall and F-1 score:



(a)



(b)

Figure 4: Accuracy and Loss for Five class classification - post-SMOTE implementation

Table 1: Pre-SMOTE Precision, Recall and Accuracy

Class	Precision	Recall	F1-score
0	0.25	0.05	0.09
1	0.73	0.89	0.80
2	0.65	0.55	0.59
3	0.67	0.80	0.73
4	0.00	0.00	0.00
Accuracy			0.68
Macro average	0.46	0.46	0.44
Weighted average	0.63	0.68	0.64

Following the SMOTE implementation (Section 4.3), running the latest dataset on Deep Neural network architecture with five layers for 500 epochs, we were able to achieve accuracy of 85% with a loss of 0.36. In addition to accuracy, we computed Precision, Recall and F-1 score on test dataset which is as shown in table below:

Table 2: Post-SMOTE Precision, Recall and Accuracy

Class	Precision	Recall	F1-score
0	0.68	0.73	0.71
1	0.66	0.60	0.63
2	0.79	0.68	0.73
3	0.62	0.76	0.68
4	0.87	0.83	0.85
Accuracy			0.72
Macro average	0.72	0.72	0.72
Weighted average	0.72	0.72	0.72

6 Conclusion and Future Work

After trying various NN model architecture, five-layered network is the best network for our model with three layers of dropout. SMOTE implementation was helpful in reducing over-fitting and increase accuracy of our model. In the future, we are considering increasing our data size and implementing RNN (using PyTorch lightning) for predicting electricity consumption using time series data. Also, we hope to practice various recommendations to increase our accuracy beyond the 85% to approximately 95%. We can use the nearest neighbours to fill out NAN values for the weather data rather than 0.

7 Contributions

Vasu curated and pre-processed electricity data, developed Deep NN model architecture, implemented SMOTE for increasing the dataset to have uniform distribution. Amoh researched into the various literature related to the topic and obtained the sources for the electricity and weather dataset. I curated the weather data, developed the data pre-processing to feeding into the DNN network and helped developed the DNN architecture. We both worked together on the hyper-parameter tuning, the write up and the PPT creation and the 4-minute video presentation.

References

[1] B. Dandotiya and H. K. Sharma, Climate Change and Its Impact on Terrestrial Ecosystems, Business Science Reference, Mayfield (2012). https://www.igi-global.com/gateway/book/52734.

[2] S. Fathi, R. Srinivasan, A. Fenner, et al., "Machine learning applications in urban building energy performance forecasting: A systematic review," Renewable and Sustainable Energy Reviews 133, 110287 (2020).

[3] G. Serale, M. Fiorentini, and M. Noussan, Development of algorithms for building energy efficiency, Woodhead Publish- ing, https://doi.org/10.1016/B978-0-12-819946-6.00011-4 (2020). https://www.sciencedirect.com/science/article/pii/B9780128199466000114.

[4] L. Li, Sun, Wanming, et al., "Impact of natural and social environmental factors on building energy consumption: Based on bibliometrics," Journal of Building Engineering 37, 102136 (2021).

[5] Guido Franco, Alan H. Sanstad, "Climate change and electricity demand in California", 87, 139–151 (2008). https://link.springer.com/article/10.1007

[6] Pardo, Angel, Meneu, Vicente Valor, Enric, "Temperature and seasonality influences on Spanish electricity load", 55-70, 1, 2002, https://www.sciencedirect.com/science/article/pii/S0140988301000822

[7] G. C. Gerhard Mu nz, Sa Li, "Traffic anomaly detection using k-means clustering," In GI/ITG Workshop MMBnet (2007).

[8] K. Kandananond, "Forecasting electricity demand in thailand with an artificial neural network approach," Energies 4, 1246–1257 (2007).

[9] B. Dong, C. Cao, and S. Lee, "Applying support vector machines to pre- dict building energy consumption in tropical region," Energy and Build- ings (May-2005).

[10] DEkici, B. Bektas, and U. T. Aksoy, "Prediction of building energy con- sumption by using artificial neural networks," Advances in Engineering Software 40, 356–362 (2009)

[11] NOAA Weather data: https://www.ncdc.noaa.gov/cdo-web/search

[12] Daily California's electricity consumption:

https://www.eia.gov/electricity/gridmonitor/dashboard/daily$_generation_mix/regional/REG-CAL$