Analyzing the Influence of Weather on Electricity Demand

Jiazhou Liu
New York University
NEW York, United States
jl7147@nyu.edu

Abstract

This application aims at analyzing the relationship between electricity demand and weather conditions with big data tools under Spark framework. The goal is to predict how electricity demand of an area will fluctuate, as weather changes and how price can be affected accordingly. The analytical approaches of this application consist of making assumptions based on dataset observations and verifying those assumptions by cleaning and organizing data for demonstration of clear patterns, as well as building up learning models in order to make predictions.

I. Introduction

The application is based on the hourly energy demand generation and weather dataset from Kaggle. The core aspect of the application is to establish a model, which is capable of predicting the electricity demand, based on different weather conditions through the days. The approaches of analysis include using core Spark API, Spark SQL and MLlib to discover certain patterns inside data, based on the assumption that weather condition of an area has significant influence over electricity demand. The methodology of this application is closely related to machine learning, and classic models such as Linear Regression, Decision Tree Regression, Random Forest Regression etc. are applied during the analytical process.

II. Motivation

In recent years, forecasting in electricity market is identified as one of the most significant area of data science, since forecasting electricity demand has huge economical potentials. One of them is helping the government to predict local electricity market, so that building renewable based electrical infrastructure becomes viable, which facilitates the development of environmentally friendly electricity generation. In addition, understanding the influence of weather over electricity demand can aid the government to be prepared under extreme weather conditions, which is critical to the safety of modern society. Therefore, I believe conducting such research is highly crucial to promote environment protection and sustainable development of our generations.

III. Related Work

Analyzing the Impact of Weather Variables on Monthly Electricity Demand by Shanti Majithia, Ching Lai Hor and Simon Watson is closely related to the focus of this paper, and its abstract is provided as the background:

The electricity industry is significantly affected by weather conditions both in terms of the operation of the network infrastructure and electricity consumption. Following privatization and deregulation, the electricity industry in the UK has become fragmented and central planning has largely disappeared. In order to maximize profits, the margin of supply has declined and the network has been run closer to capacity in certain areas. Careful planning is required to manage future electricity demand within the framework of this leaner electricity network. There is evidence that the climate in the UK is changing with a possible 30C average annual temperature increase by 2080. This paper investigates the impact of weather variables on monthly electricity demand in England and Wales. A multiple regression model is developed to forecast monthly electricity demand based on weather variables, gross domestic product and population growth.

At the beginning, the authors introduce the idea that the climate in the United Kingdom is showing long-term changes. Following the idea, they suggest that such changes will lead to more extreme weather conditions. By analyzing the weather data for the past 13 years, they successfully build a regression model which is able to accurately predict electricity demand under both normal and extreme weather conditions, which is similar to the goal of this application: evaluating the relationship between electricity demand and weather conditions and trying to make predictions based on weather parameters.

Later in the paper, they compare the performance of their multiple regression model with artificial neural network, socio-economic model, and Box and Jenkin's model, and conclude that the forecasting accuracy of their model can achieve equal or better level. In addition to building the model, they find a strong inverse relationship between electricity demand and temperature: in the winter, there is a significant lighting and heating load which coincides with the lower temperatures. Conversely, in the summer, demand tends to be lower, however, above a critical temperature, consumption tend to rise again due to an increase in cooling and air-conditioning loads.

During the conclusion section, they claim that climate conditions could have significant influence over electricity demand, especially humidity as a variable. Moreover, they believe that their model is robust enough to predict electrical demand in the future, given reliable estimates of the relevant climate-related and socio-economic parameters.

IV. Datasets

A. Energy Demand Dataset - electricity demand, generation, and prices of Spain

Size: 6.3MB

Schema:

time: Date	datetime localized to CET
total load forecast: Float	forecasted electricity demand in MW
total load actual: Float	actual electricity demand in MW
price day ahead: Float	forecasted electricity price in EUR/MWh
price actual: Float	actual electricity price in EUR/MWh

(only used fields are provided due to large number of fields)

B. Weather Features Dataset - Weather features temperature, humidity, etc. of Spain

Size: 19.9MB

Schema:

dt iso: Date	datetime localized to CET
city name: String	name of city
temp: Float	temperature in K
temp min: Float	minimum temperature in K
temp max: Float	maximum temperature in K
pressure: Integer	pressure in hPa
humidity: Integer	humidity in percentage
wind speed: Integer	wind speed in m/s
wind deg: Integer	wind direction in degree
rain 3h: Float	rain in last 3 hours in mm
snow 3h: Float	snow in last 3 hours in mm
clouds all: Integer	cloud cover in percentage
weather main: String	short description of current weather
weather description: String	long description of current weather

(only used fields are provided due to large number of fields)

Links to both datasets:

https://www.kaggle.com/nicholasjhana/energy-consumption-generation-prices-and-weather#weather_features.csv

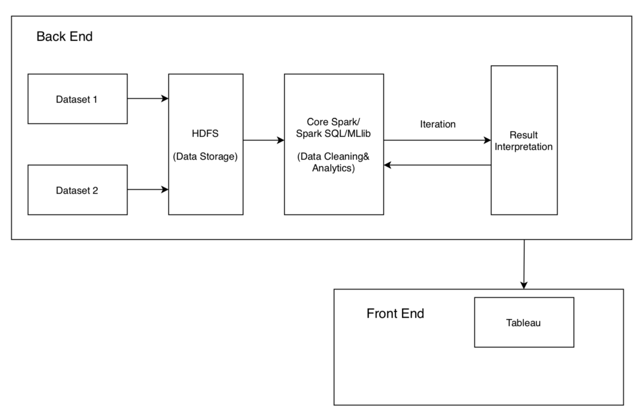
V. Description of Analytic

The analysis mainly includes three stages: creating chart to discover relationship between certain weathers and electricity demands, calculating correlation coefficient between weather parameters and electricity demand, as well as that between electricity demand and pricing, and building learning models to predict electricity demand based on weather parameters of strong correlations.

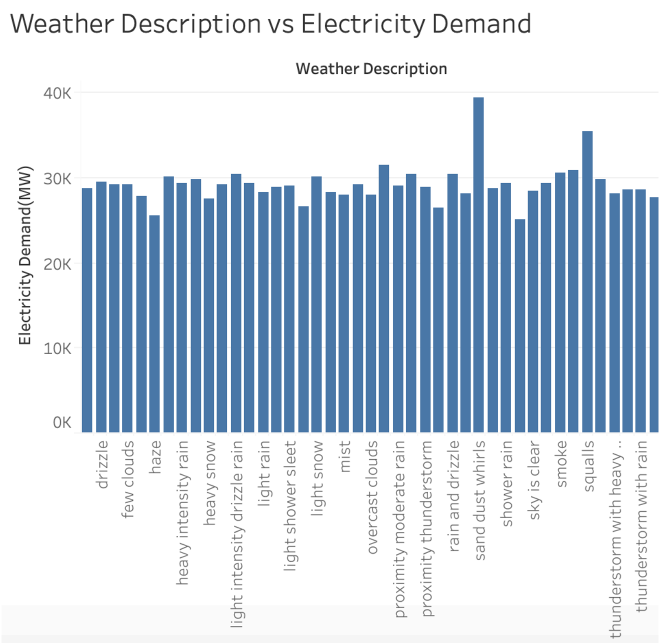
The findings derived from the analysis demonstrate that certain weather parameters are closely related to the electricity demand, which agrees with the assumption made before the analysis, and it's necessary for the government to consider weather elements while building electricity infrastructure, as well as providing electricity for local area. Detailed analytical procedures and findings will be discussed in the analysis section.

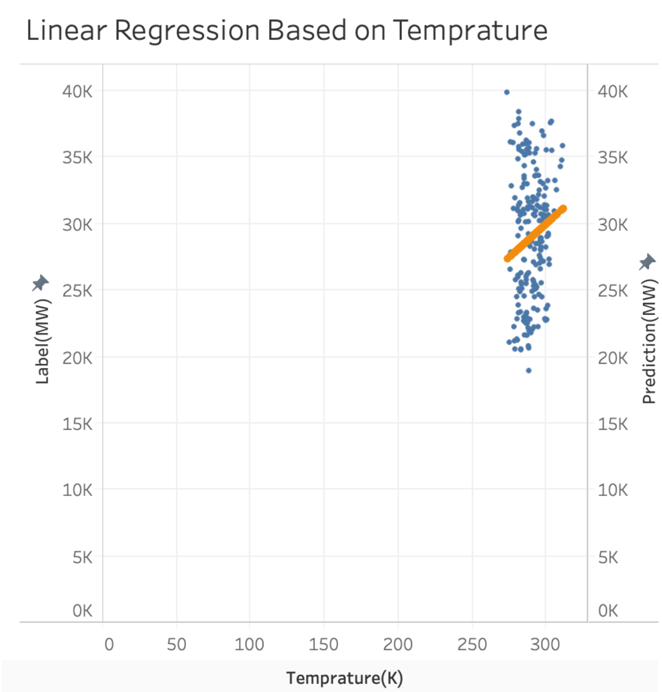
VI. Application Design

The following design diagram shows the fundamental structure of the application and also the workflow of data through the analytic process. The weather and electricity demand datasets are stored in HDFS, and they are later used by core Spark, Spark SQL and MLlib for data analysis. The analytic process takes many iterations; inside each iteration, the result from analysis is interpreted, and based on the interpretation, analytic approaches change to reflect the latest understanding of the datasets. The final outputs of the analysis are visualized and presented in Tableau Dashboard.



The first graph below shows the relationship between weather description and electricity demand by a bar chart, and the second graph below shows the relationship between temperature and electricity demand by scatter dot plot with regression line. Detailed interpretations of the graphs will be discussed in analysis section.





VII. Actuation or Remediation

In terms of the analysis, several actuations or remediations can be derived. Firstly, during weathers such as "sand dust whirls" and "squalls", local government should make sure electricity generation is adequate to handle potential demand bursts. Secondly, when forecasting electricity demand, weather parameters such as temperature and humidity should always be considered as input features. Finally, because price and demand are highly positively correlated, after making predictions about future demands, organizations such as factory, should consider increasing or decreasing events that depend on electricity accordingly to minimize their costs.

VIII. Analysis

(In this section, describe: Your experimental setup (tools, platforms), problems (with data, performance, tools, platforms, etc.). Describe what you learned. Discuss limitations of the application. Make recommendations for others, e.g. best practices.)

The focus of the analysis is to discover the relationship between various weather elements and electricity demands, and the analysis consists of three different stages. The first stage focuses on exploring data with Spark SQL, while the other two stages focus on generalizing patterns with tools of MLlib.

At the first stage, a bar chart is created to visualize the relation between weather description and demand. The chart shows that weathers describes as "sand dust whirls" and "squalls" cause a significant increase of average electricity demand, which means these weathers are highly influential and potentially able to cause shortage of electricity or inflict severe pressure on the electricity network of the local area.

At the second stage, PCC(Pearson correlation coefficient) between each weather parameter and electricity demand is calculated. temp, temp min, temp max, humidity, wind speed are the five parameters with strongest correlations:

Weather Parameter	PCC
humidity	-0.245
temp	0.181
temp min	0.179
temp max	0.170
wind speed	0.126

The PCC between electricity demand and price is also calculated, and it has value 0.436, which suggests demand and price are highly positively correlated.

At the final stage, various learning models including Linear Regression, Decision-Tree Regression, Random Forest Regression, Gradient-Boost-Tree Regression are applied, in purpose of making predictions about electricity demand in terms of the five parameters listed above. Since multiple linear regression is difficult to visualize in practice, temperature is chosen as the feature to create linear regression graph. From the graph, as temperature increases, electricity demand tend to have higher value. However, the data points are dispersive, which means it's hard to generalize data based on linear equation.

RMSE(Root Mean Square Error) for each model is provided in table below as evaluation:

Prediction Model	RMSE
Linear Regression	4423
Decision-Tree Regression	4359
Random Forest Regression	4355
Gradient-Boost-Tree Regression	4322

According to the table, different prediction models tend to produce similar RMSE, which indicates that the prediction accuracy is limited by the correlation between features and labels, despite model choices. Therefore, RMSE can potentially be reduced further if given more related features.

IX. CONCLUSION

This application is useful for individuals or organizations that are interested in finding relationship between weather and electricity demands. It analyzes weather's effect over demand from aspects of both general weather category and specific weather parameters. Although based on the result, it's not enough to produce accurate demand predictions purely based on weather, it's a good starting point for someone to build forecasting model in the future.

X. FUTURE WORK

For future work, the prediction accuracy of current application can be improved, if given other weather features as well as features of other categories. Also, the front-end application could have more interactive features for users. In addition, weather and electricity demand data of other countries can also be added, so that the application becomes more general and useful to a larger scope of users.

ACKNOWLEDGMENT

I would like to express my gratitude to Cloudera for providing CDH(Cloudera Distribution Hadoop) and also Tableau for providing Tableau Desktop as trial.

REFERENCE

[1] Watson, S., Majithia, S., & Hor, C.L. Analyzing the impact of Weather Variables on Monthly Electricity Demand, IEEE 2005

[2] Nicholas Jhana Hourly Energy Demand Generation and Weather, 2019