SMARTINTERNZ INTERNSHIP PROJECT REPORT

Name - Aman Bahuguna

Project ID - SPS_PRO_298

Project Title : Predicting The Energy Output Of Wind Turbine Based On Weather Conditions Watson Auto AI

Predicting The Energy Output Of Wind Turbine Based On Weather Conditions Watson Auto AI

Abstract :

Wind energy plays an increasing role in the supply of energy world-wide. The energy output of a wind farm is highly dependent on the weather conditions present at its site. If the output can be predicted more accurately, energy suppliers can coordinate the collaborative production of different energy sources more efficiently to avoid costly overproduction. In this paper, we predict energy prediction based on weather data and analyse the important parameters as well as their correlation on the energy output.

Keywords:

Python, Python for Data Analysis, Machine Learning, IBM Cloud, IBM Watson

1. INTRODUCTION

- Our aim is to map weather data to energy production. We wish to show that even data that is publicly available for weather stations close to wind farms can be used to give a good prediction of the energy output. Furthermore, we examine the impact of different weather conditions on the energy output of wind farms. We are building an IBM Watson AutoAI Machine Learning technique to predict the energy output of wind turbine. The model is deployed on IBM cloud to get scoring end point which can be used as API in mobile app or web app building. We are developing a web application which is built using node red service. We make use of the scoring end point to give user input values to the deployed model. The model prediction is then showcased on User Interface to predict the energy output of wind turbine.

2. DATASET

- The main goal of this paper is to use public data to check the feasibility of wind energy prediction by using an industrial-strength off-the-shelf non-linear modeling and feature selection tool. In our study, we investigate and predict the energy production of the wind farm Woolnorth in Tasmania, Australia based on publicly available data. The energy production data is made publicly available by the Australian Energy Market Operator (AEMO) in real time to assist in maintaining the security of the power system. 2 For the creation of our models and the prediction, we associate the wind farm with the Australian weather station ID091245, located at Cape Grim, Tasmania. Its data is available for free for a running observation time window of 72 hours. 3 3.1 Data We collected both the weather and energy production data for the time window from September 2010 to July 2011. The output of the farm is available with a rate of one measurement every five minutes, and the weather data with a rate of one measurement every 30 minutes. The wind farm's production capacity is split into two sites, which complicated the generation of models. The site "Studland Bay" has a maximum output of 75 MW, and "Bluff Point" has a maximum output of 65 MW and is located 50km south of the first site. The weather station is located on the first site. For wind coming from the west (which is the prevailing wind direction), the difference in location is negligible. But if the wind comes from the north, there is an immediate energy and wind increase, plus another energy increase 1–2 2 Australian Landscape Guardians: AEMO Non-Scheduled Generation Data: www.

landscapeguardians.org.au/data/demo/ (last visited August 31st, 2011) 3 Australian Government, Bureau of Meteorology: weather observations for Cape Grim: www.bom.gov.au/products/IDT60801/IDT60801.94954.shtml (last visited August 31st, 2011) 7 hours later (the time delay depends on the actual wind speed). Similarly, if the wind comes from the south, there will be an increase in the energy production (although no wind is indicated by the weather station) and then, 1–2 hours later, an energy increase accompanied by a measured wind speed increase. 3.2 Data pre-processing To perform data modeling and variable selection on collected data, we had to perform data pre-processing to create a table of weather and energy measurements taken at the same time intervals. Energy output of the farm is measured at the rate of 5 minutes, including the time stamps of 0 and 30 minutes of every hour when the weather is measured. Our approach was to correlate weather measurements with the average energy output of the farm reported in the [0, 25] and [30, 55] minute intervals of every hour. Such averaging makes modeling more difficult, but uses all available energy information. Different time scales used in the weather and energy data were automatically converted to one scale using a DateList function in Wolfram Mathematica 8, which is the scientific computing environment in which DataModeler operates. Because of many missing, erroneous, and duplicate time stamps in the weather data, we obtained 11022 common measurements of weather and averaged energy produced by the farm from October 2010 to June 2011. These samples were used as training data to build regression models. From 18 variables of the weather data at Cape Grim, we excluded two variables prior to modeling: more than 75% of values for the Pressure MSL variable were missing and the Wind Direction variable was non-numeric. As test data we used 1408 common half-hour measurements of weather and averaged energy in July 2011. 3.3 Data Analysis and Model Development As soon as weather and energy data from different sources were put in an appropriate input-output form, we were able to apply a standard data-driven modeling approach to them. A good approach employs iterations among three stages: Data Collection/Reduction, Model Development, and Model Analysis and Variable Selection. In hard problems, many iterations are required to identify a subspace of minimal dimensionality where models of appropriate accuracy and complexity tradeoffs can be built. 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 Year Month Day Hour Minute Temperature ApparentTemperature DewPoint RelativeHumidity WetBulbDepression WindSpeed WindGust WindSpeed2 WindGust2 PressureQNH RainSince9am EnergyOutput 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 CorrelationMatrixPlot of Data Columns Fig. 2. Data variables are heavily correlated (Blue: positively, Red: negatively). Our problem is challenging for several reasons. First, it is hard to predict the total wind energy output of the farm in the half hour following the moment when weather is measured, especially when the weather station is several kilometers away from the farm). Second, public data does not offer any information about the wind farm except for wind energy output. Third, our training data covers the range of weather conditions observed only between October 2010 and June 2011, while the test data contains data from July, implying that our models must have good generalization capabilities as they will be extrapolated to the unseen regions of the data space. And lastly, our most challenging goal is to use all 16 publicly available numeric weather characteristics for energy output prediction, although many of them are heavily correlated (see Figure 2). Multi-collinearity in hard high-dimensional problems is a major hurdle for most regression methods. Symbolic regression via GP is one of the very few methods that does not suffer from multicollinearity and which is capable of naturally selecting variables from the correlated subset for final regression models. Because ensemble-based symbolic regression and robust variable selection methodology are implemented in DataModeler, we settled on a standard model development and variable selection procedures using default settings. The modeling goals of this study are: (1) to identify the minimal subset of driving weather features that are significantly related to the wind energy output of the wind farm, (2) to let genetic programming express these relationships in the form of explicit input-output regression models, and (3) to select model ensembles for improved generalization capabilities of energy predictions and to analyze the quality of produced model ensembles 9 using an unseen test set. Our approach is to achieve these goals using two iterations of symbolic regression modeling. At the first exploratory stage, we run symbolic regression on training data to identify driving weather characteristics significantly related to the energy output. At the second modeling stage, we reduce the training data to the set of selected inputs and run symbolic regress to obtain models, and model ensembles for predicting energy output.

3. Methodology

- Forecasting models for wind power can be divided into two overall groups. The first group is based upon analysis of historical time series of wind, and a second group uses forecasted values from a numerical weather prediction (NWP) model as an input. However, wind power forecasting is generally described in terms of physical methods, traditional statistical or ‘black box’ methods and more recently the so-called learning approaches, artificial intelligence or ‘gray box’ methods. Hybrid methods can involve some aspect of all of these.

4. Future Work :

- Most wind power forecasting models study 'regular' wind conditions. The EU funded project called 'Safewind' aims to improve wind power prediction over challenging and extreme weather periods and at different temporal and spatial scales [61]. Development activities are on-going to reduce error in wind power prediction, to improve regionalized wind power forecasting for on-shore wind farms and to derive methods for wind power prediction for offshore wind farms. It is possible that the use of ensemble and combined weather prediction methods may enhance forecasting.

5. Conclusion :

- One of the ultimate goals of every wind power prediction model is to estimate the wind power output as early and as accurately as possible. Wind power will become more attractive for system and market operators as NWP model accuracy improves and as easier to use forecasting techniques are developed. Wind power prediction tools are invaluable because they enable better dispatch, scheduling and unit commitment of thermal generators, hydro plant and energy storage plant and more competitive market trading as wind power ramps up and down. Overall accurate wind power prediction reduces the financial and technical risk of uncertainty of wind power production for all electricity market participants. When smart grid technology and intelligent load management techniques (such as controlled water and space heating and chilling, and electric vehicle charging) are deployed, integration of wind power will become a more straightforward task. Many aspects of existing grid systems, conventional thermal generation and the management of the power system are circa 70 years old, whereas large-scale adoption of wind energy has only occurred in just the last 15 years. Furthermore, a more diverse generation portfolio mix, which includes energy storage plant, offshore wind, wave and tidal will also make wind power integration less operationally intensive for system operators.