May 2018

Go with the Flow: Recurrent Networks for Wind Time Series Multi-Step Forecasting

Jaume MANERO [a,1] Javier BÉJAR [a] and Ulises CORTÉS [a,b]
[a] *Universitat Politècnica de Catalunya - BarcelonaTECH, Barcelona, Spain*
[b] *Barcelona Supercomputing Center, Barcelona, Spain*

Abstract. One of the ways of reducing the effects of Climate Change is to rely on renewable energy sources. Their intermittent nature makes necessary to obtain a mid-long term accurate forecasting. Wind Energy prediction is based on the ability to forecast wind speed. This has been a problem approached using different methods based on the statistical properties of the wind time series.

Wind Time series are non-linear and non-stationary, making their forecasting very challenging. Deep neural networks have shown their success recently for problems involving sequences with non-linear behavior. In this work, we perform experiments comparing the capability of different neural network architectures for multi-step forecasting obtaining a 12 hours ahead prediction using data from the National Renewable Energy Laboratory's WIND dataset[2].

Keywords. Time series, Recurrent Neural Networks, Multi-Step prediction, Seq2Seq

1. Introduction

Wind Power Generation is a critical contributor to the electrical supply systems in many countries. This penetration will see a steep increase in the next few years due to the renewable push needed to fulfill the targets established by the agreements at the Climate Change Paris Conference. Thus, developing more reliable techniques for the integration of wind power is critical, and forecasting the energy generation output is a key task.

Wind time series are complex and difficult to forecast, and many methods have been proposed, from the short-term accurate persistence, or the whole family of linear time series models (AR, ARIMA, etc.), some non-linear statistical methods, and finally Machine Learning methods [3]. In this paper, we will test a set of Neural Network architectures applied to wind forecasting and show some preliminary results.

2. The Wind Energy Generation Forecasting Task

Wind Energy is generated by the action of wind on the blades of the turbines. The power generated is mainly dependent on the airspeed. There is some discussion about if it is

[1] Corresponding Author: Jaume Manero, UPC, Barcelona, Spain; E-mail: jaume.manero@upc.edu.
[2] We thank the US National Renewable Energy Laboratory (NREL) for the use of their wind energy datasets.

better to forecast wind speed or power generated, with no consensus. In the commercial applications, the final goal is to find the best possible power prediction, to perform this, independently of the approach, it is required to discover the internal patterns of the wind.

Wind turbines have sensors that offer information in real time, generating streams of data. Typically, a wind turbine time series will be a set of observations several years long containing variables related to the wind and other meteorological dimensions. All these observations can be generated at different heights. As the wind at 100 meters high is the one that moves the blades, it is probably the measure with the most top relevance.

Two properties make wind time series challenging to predict, namely, non-stationarity and non-linearity. There have been different analysis of wind time series behavior in the literature, and it usually depends on the geographical location and weather conditions. There are parts of the series with linear characteristics that are easier to predict, but the usual behavior of a wind time series has both properties.

Additionally to these difficulties, one of the main interests in this domain is to be able to obtain mid/long-term predictions. This means that it is not the next step of the series the real goal, but a distant horizon. This increases uncertainty in the task.

3. The WIND dataset

An especially challenging problem in this domain is the availability of the data. Usually, real data from industry are proprietary and are not available. To address this problem, the data used in our experiments comes from the National Renewable Energy Laboratory (NREL) [2]. These data offer production and meteorological data (wind speed, wind direction, temperature, humidity and energy) synthesized from Meteorological global models for over 120,000 sites in the US. All this information is stored in 5 minutes intervals for six years. We will use in this paper a limited number of sites, but given the large geographical diversity of the sites, it is the aim in the future to test the performance of different prediction methods according to the sites different profiles.

4. Methodology and experiments

Given the complex nature of wind forecasting, the goals are to test different methodologies to gain some insight about their adequacy for the task, based on their accuracy.

We will address two dimensions in our experiments. First, the methodology for the multi-step prediction, so we can obtain mid-term forecasts. Second, the methodology for computing the prediction, that can be defined as a regression problem.

Several approaches are possible for multi-step forecast based on how the future time steps are generated [6]. The simplest being the *recursive* approach, where a data window $[t_{i-k}, t_{i-1}]$ and prediction for time t_i are used for predicting t_{i+1} iteratively. According to the literature, this is not very good for mid/long term prediction because of the compound error effect of reusing predicted data. It also has the handicap that only the target series can be used as input and other available exogenous variables cannot be used.

The *direct* approach obtains predictions by computing a regression for each of the time points on the future horizon. This has the advantage of not reusing predicted data but is more expensive given that multiple models have to be obtained. Another approach is to use multiple regression or sequence to sequence prediction [5]. This means that all the future time steps are obtained at the same time without reusing predictions and with a unique model. We will use these two techniques in our experiments.

For solving the regression problem, we have chosen regression support vector machines (SVR) [1] and two neural network methods, multi-layer perceptron (MLP) and recurrent networks (RNN) [4]. The SVR will be used as a baseline and only for direct prediction. MLP and RNN will be used for direct prediction and multiple regression.

A systematic exploration of the parameters for all methods was performed. For SVR, different kernels will be used, namely RBF, linear, quadratic and cubic polynomial with a wide range of values for the C parameter and the bandwidth for the RBF kernel.

For the MLP, there were experimented architectures from one to three layers of different sizes with a linear output, sigmoid and ReLU activation functions and different values of dropout on each layer. The direct approach had one output neuron and the multiple regression as many outputs as the prediction horizon.

For the RNN architecture, from one to three layers of LSTM or GRU units with different sizes, with tanh, sigmoid and ReLU activation functions for the output of the recurrent units and different levels of recurrent dropout. For the direct approach, a MLP with linear output was used. For the multiple regression, an encoder-decoder architecture was used [5], where the recurrent layers were used as a first stage, performing the encoder task. The state obtained from the encoder was used as input for another recurrent network, acting as a decoder, which generates a sequence with the length of the prediction horizon. Each time step of the decoder had direct access to the state from the encoder additionally to the state from the previous step.

5. Results

The raw data have a five minutes sampling, to reduce computational cost and assuming a realistic forecasting scenario of hourly predictions, the data was reduced to an hourly sampling by averaging the measures hourly. The data used for the prediction included the wind speed and direction at 100 meters, barometric pressure and air density. The hour and month of the data were added as complementary variables. The training data consisted on the first four years of the series, the fifth year was used as a test set for tunning the model parameters and the sixth year was used as the validation set. The training dataset has a size of 40912 values and the test and validation a size of 10228 values.

Different windows lengths were used as input, ranging from three to 36 previous measures. The forecast was the wind speed from one to 12 hours ahead.

To compare the results the determination coefficient (R^2) was chosen. The data was z-normalized, so the coefficient is equivalent to $1 - MSE$. Figure 1 shows the averaged R^2 for the best 20 results for each architecture for a specific site. The RNN model with direct prediction is consistently the best model followed by multiple regression with MLP. RNN with seq2seq performs similarly to the multiple regression MLP for the short term, but the mid-term predictions decay faster. The MLP and SVR with direct prediction have very similar results far from the rest.

Performing a two-sided Kolmogorov-Smirnov test for equality of distribution for the R^2 values for each time step of the prediction horizon for the best two architectures (RNN direct and MLP multi-regression) all distributions have less than $1e-4$ as p-

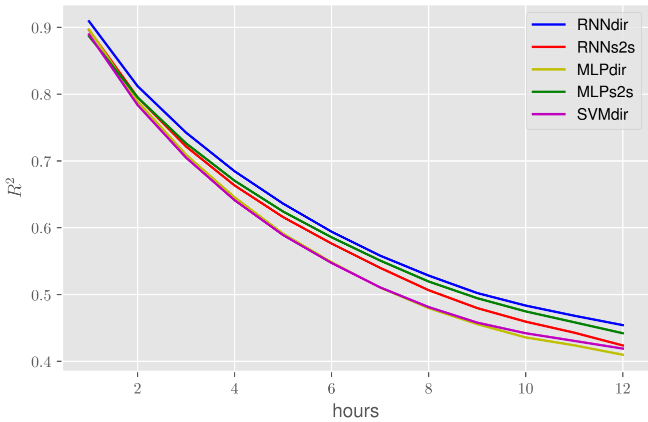

Figure 1. Mean of R^2 for the 20 best experiments for each architecture, 12 hours horizon.

value, indicating that their distributions are different. The difference of the means for the different hourly prediction of the 20 best results is in the range 0.007 to 0.022. The best RNN architectures have two layers of GRU units with ReLU activation functions, drop out of around 0.3 with a window input from 16 to 24 hours.

6. Conclusions and future work

The preliminary results of our experiments show that RNN architectures with direct multi-step prediction can obtain reasonable mid-term predictions of wind speed with consistent accuracy among a limited number of sites. Other approaches show a significant decrease in accuracy the further the horizon of prediction. The experiments also show that the direct approach for multi-step prediction have better results compared to a multiple regression/sequence to sequence approach.

Other multi-step prediction methods have to be explored, combined with different RNN architectures. For instance, given that it is more important the mid-term prediction, a multiple regression focused only on the more distant future seems more interesting. Also, further experiments using more advanced methods for the RNN sequence to sequence architectures have to be explored, like teacher forcing or attention mechanisms.

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, Berlin, Heidelberg, 2006.

[2] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa. The wind integration national dataset (wind) toolkit. *Applied Energy*, 151:355–366, 2015.

[3] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. *The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview, 2^{nd} edition*. ANEMOS.plus, 2011.

[4] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[5] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. *CoRR*, 2014.

[6] S. B. Taieb and A. F. Atiya. A bias and variance analysis for multistep-ahead time series forecasting. *IEEE transactions on neural networks and learning systems*, 27(1):62–76, 2016.