WIND POWER FORECASTING WITH AN LSTM BASED RNN

Shuangling Yang, Martin Illum, Mads Agerlin Kähler, Hans Christian B. Mikkelsen

s210306, s184286, s184206, s184294

ABSTRACT

Wind power forecasting is crucial to enhance the stability of power integration in power grids. This paper presents how a LSTM neural network can be applied to forecast the power generation from a wind farm. The data used to train the model is provided by Metodyn China. The data consists of numerical weather predictions and actual measurements. It is found that a more simple model structure generates the best results, and that the model can effectively capture the dynamic and trend of power generation. It was further concluded that the feature engineering and data preparation coarsely affected the model's output accuracy, while hyperparameter tuning was of less significance.

***Index Terms*—** LSTM, recurrent neural network, times series forecasting, wind power forecasting

1. INTRODUCTION

The world is ramping up its transition from fossil fuels to more sustainable solutions, and the use of wind turbines for power generation is an important tool in this process. Due to its dependency on wind, wind power generation is not as dispatchable as conventional power systems. For this reason, it is beneficial to predict the expected yield to plan for the need to dispatch traditional power generation systems.

This paper addresses wind power forecasting of three wind turbine farms located in China. The forecast will be based on time series data provided from Meteodyn China, which owns the wind turbine farms, and Numerical weather prediction (NWP). Thus, there is a need for specifying the length of the sequence that will be put into the model.

Wind power forecasting problems can made in the short, medium and long-term ahead. Short-term wind power forecasting is used for improving the overall efficiency of the power grid. Medium- and long-term forecasting are used for planing maintenance of the wind turbine farms, unit commitment and maintenance outages of thermal generators and to schedule grid maintenance and energy storage operations [1]. The short-term time interval is in the range of one hour to several hours, the medium-term time interval is in the range of several hours to a week and the long-term time interval is at least a week [2].

In wind power forecasting is there two main types methods [1]: physical and statistical. The physical methods have on-line access to weather conditions, such as wind speed, wind direction and air pressure etc. for forecasting the wind power. Numerical weather prediction (NWP) is a renowned physical model, which also will be used in this paper to forecast wind power. The statistical methods trains on historic data to predict wind power. Therefore, physical methods has advantages in long-term prediction while statistical method does well in short-term prediction [3].

Over the recent years the use of conventional machine learning and deep learning methods have gained a lot focus in the context of wind power forecasting. Conventional machine learning such as Support Vector Machine (SVM) have been widely used to forecast wind speed [4]. Among the usage of deep learning methods for time series data the standard is the Long short-term Memory Network (LSTM), which will be explored in the Section 1.1.

The paper is organized as follows. Section 2 presents data and feature engineering, Section 3 reveals which methods will be used in the paper. Results are illustrated in Section 4. Conclusion, discussion and future work are presented in respectively Section 5, 6 and 7.

1.1. Previous Work

Due to the growing interest in utilizing wind power, despite its limited dispatchability, a significant amount of research has been made with regard to forecasting the expected yield of wind turbine farms. A short-term wind power forecasting model based on an LSTM network is presented in [5]. They show with simulations that, when compared to a back propagation neural network and SVM model, their LSTM prediction model has higher prediction accuracy and greater potential for engineering applications.

Another paper uses an LSTM model to predict the value of the four features: wind speed, direction, generated active power and theoretical power and the results are compared to the statistical moving average technique (MA) and a multilayer perceptron (MLP) model [6]. The three different methods are compared using the metrics mean absolute error (MAE). They found out that their LSTM model performed the best, had the lowest MSE, in all of the four features.

2. DATASET

This dataset is provided by Meteodyn China, and in its raw form, consists of a provided NWP, as well as actual measurement data for 3 wind farms referred to as cases 1, 2, and 3. Due to limited information regarding the geographical location of the wind farms, and varying data quality for each of the cases, only data from case 3 will be used. For this project, we consider the attributes from the NWP, as well as the measured *Park Power [KW]* attribute, resulting in 116270 observations with 19 attributes. These attributes include the time of the observation, and the wind speed, wind direction, and temperature at that time. The observations in the dataset spans over a time interval of 3 years and 9 months with an interval of 15 minutes. The target attribute for the network's prediction is the total power output of the farm, or the attribute *Park Power [KW]*. The aim is to predict this value in the future 6, 12 and 24 hours based on the NWP data.

2.1. Data Preparation

Constructing an accurate and efficient model requires that the data that can effectively reflect the real information we wish for the model to learn from. To achieve this, we first cleaned the data to correct or remove incorrect data. Then we conducted a preliminary data exploration to visualize and understand the correlation between variables more intuitively.

An examination of the target variable *Park Power [KW]* was performed, where it was found that more than four percent of the data exceed the maximum theoretical wind energy at a specific wind speed, which could be calculated based on the power curve provided by Meteodyn China. For these coincidences, we cut off the original data to theoretical values. We also discarded the sequences of observations lacking 96 consecutive timesteps, because our model uses 96 observations as a window (24 hours).

The wind rose diagram, shown in figure 1, plots the distribution of wind speed and direction, showing that the wind is mainly from the southwest. In addition, the wind direction has a long-tail distribution, while the wind speed is normally distributed, which is in line with our initial expectations.

Finally, we plotted the correlation diagram in figure 2 to investigate the correlation between the target variable and the features, and to understand the covariation trend between the independent variable. It is shown that the wind speed in all heights are the variables which are mostly correlated with the target variable, although the.

2.2. Feature Engineering

In order to create and select the most useful predictor variables for the model, we combined existing features as well as applied transformations, creating new derived features with stronger predictive capabilities.

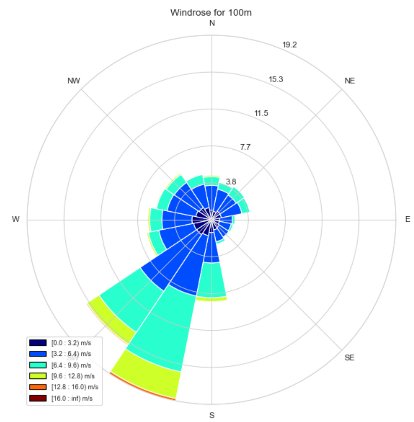

Fig. 1. Windrose for measurements at 100m height

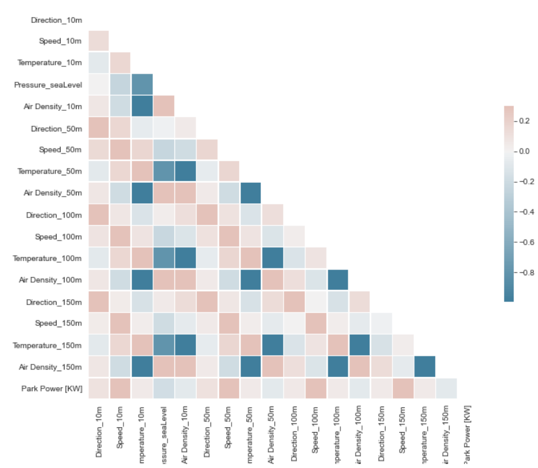

Fig. 2. Correlation of original attributes

To make the date attribute (day, month) in the raw data show its periodic quality during the training process, we applied periodic functions (sin and cos) to the attribute. With this transformation, the seasonal trend will be more easily interpreted by the model. The wind direction given in degrees (0° to 360°) and wind speed given in (m/s) were transformed into converted into two-dimensional wind vectors. This was done to make sure wind direction does not impact the model at low or zero wind speeds, as well to transform the data away from the degree format. In addition, we computed the change in wind direction and wind speed between timesteps, to provide the model with further context about the current state and development of the weather.

3. METHOD

3.1. RNN Model Structure

The primary layer of our model is a unidirectional recurrent layer, arranged in a many-to-one network. All the features from each element in the input sequence are passed to the recurrent layer, and the last output is passed through a fully-connected layer with an output layer consisting of a single node. This value is the target attribute at the for the timestamp following that of the last element in the sequence.

3.1.1. LSTM

LSTM (Long Short Term Memory) units are a further development of standard RNN units. It employs four gates which control the flow of information flow of the unit, thereby improving its ability to retain long-term information, as well as improving its response to training. Specifically, it to some degree addresses the vanishing gradients problem, where propagation through time causes the gradient to achieve increasingly small values, thereby having a significantly diminished affect when training against information with longer delays between dependencies [7].

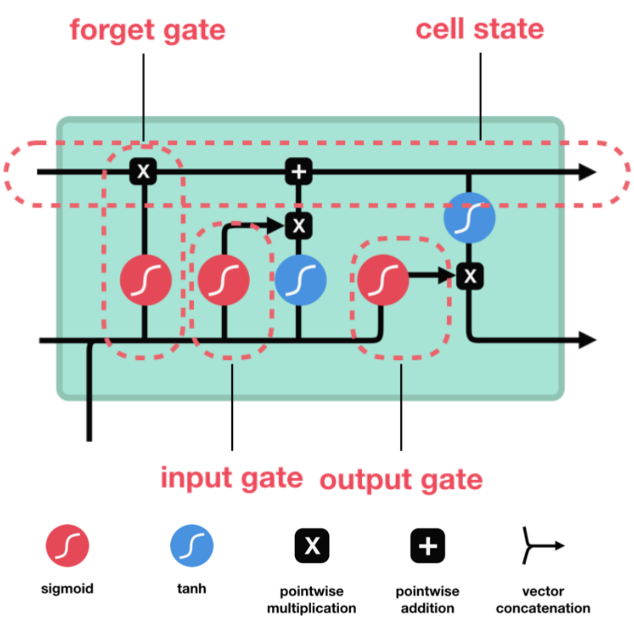

Fig. 3. An LSTM unit and the data flow through its gates [8]

3.2. Data Windowing

The network accepts sequences of input features when predicting the park power. For this, a windowing function for the data was created to generate sequences for each predicted value desired in the output. Formally, consider multi-variate time-series data, and represent the set of all features for all samples as f. The window can then be defined as:

$$f : \{x_0, x_1, x_2, ...x_{s-2}, x_{s-1}, x_s\} \tag{1}$$

where sub-index s represents the number of samples. The variable size of s indicates that the window can be configured for different lengths, altering the length of the sequence provided to the network. Determining the length of the sequence is a compromise between training speed, and to some degree, accuracy of the trained network. We chose a value of 96, equivalent to a period of 1 day. This value indicates that the samples from x_0 to x_{95} are used as past evidence to predict the value for x_{96}, which is given by:

$$(x_0, x_1, ...x_{94}, x_{95}) \rightarrow X_1 \text{ and } x_{96} \rightarrow y_1 \tag{2}$$

In the next step, the window is shifted one element, meaning it now encloses the values from x_1 to x_{96} and predict the value for x_{97}.

3.3. Performance metrics

In this paper, the performance of the models will be evaluated on mean absolute error (MAE) and root mean squared error (RMSE). MAE is known as the average absolute error between the predicted value and actual value. The equation for MAE looks as followed:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |p - \hat{p}| \tag{3}$$

In this equation, p is the actual power output of the wind turbine farm and $\hat{p}$ is the predicted power output of the wind turbine farm. The desired value of MAE is zero, where the model performance would be perfect, exactly matching the measured values. RMSE gives the overall agreement between predicted and actual value avoiding error compensation. RMSE can be written as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p - \hat{p})^2} \tag{4}$$

Where p is once again the actual power output of the wind turbine farm and $\hat{p}$ is the predicted power output of the wind turbine farm. Again, desired value of RMSE is zero, where the model performance would be perfect.

3.4. Model Training & Sensitivity Analysis

Several model configurations were tried during our search for the best performing model structure. The final model consisted of a sole LSTM layer and a linear layer. The model was trained on a training set containing 90% of the observations, and the last 10% was was used as a test set. Below in table 1 The best performing hyperparameters for the model can be seen. Due to computational limits, a grid search was not applied as a method to optimize the hyperparameters.

Hyperparameters	Value
Loss Function	Mean squared loss
LSTM Cells	64
Optimizer	Adam
Epochs	16
Batch Size	8

Table 1. LSTM Hyperparameter values

Instead, a manual sensitivity analysis was performed adjusting the learning rate, batch sizes, hidden unites, LSTM layers, direction/bidirectional, optimizers and attempts to implement regularization. This was done by running multiple models, and performing adjustments based on the output and redoing the process. Therefore it can be assumed that further optimization of the hyperparameters is possible.

4. RESULT

4.1. Model Performance

To evaluate the model, predictions have been made 6, 12 and 24 hours ahead. The test data and 24 hour predictions have been plotted in figure 4. It can clearly be seen that the model picks up the general trends of the *Park_Power_KW* attribute. However, it can also be seen that the magnitude of the predictions is relatively off within certain time intervals.

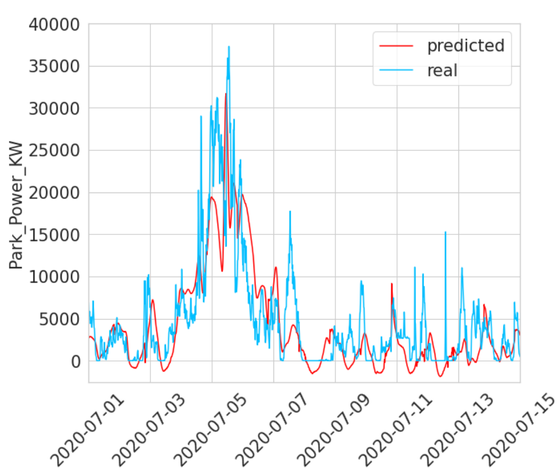

Fig. 4. Model predictions for 15 days of the test set

To further evaluate the LSTM model against our baseline a linear regression model, a table with the performance metrics mean absolute error and mean squared root error.

Model	MAE	RMSE
6 Hour Predictions		
LR	11829.57	13563.26
LSTM	4126.27	6240.98
12 Hour Predictions		
LR	11741.18	*13056.59*
LSTM	4112.32	6213.55
24 Hour Predictions		
LR	11734.20	12319.38
LSTM	4663.13	6557.77

Table 2. Linear Regression and LSTM performances by MAE and RMSE

It is clear the the LSTM model clearly is superior to the baseline, for all the time intervals. The baseline performs similarly for every time step which was to be assumed. It can also be seen that the LSTM model becomes slightly worse for the 24 hours predictions, while the 6 hour and 12 hour prediction are quite similar. In general is the LSTM model roughly 2.8 times better performing for the MAE and 2 times better for the RMSE.

4.2. Limitations of the model

One limitation of the model could be that it is trained on data from solely one wind farm. This means that the model will be very specialized to this wind farm, therefor the model might not perform similarly on data from other wind-farms. This could be due to different geographical parameters influencing the amount of electricity produced. An example could be a ocean wind-farm compared to a land based one. The roughness of the terrain and wind shear has a huge impact on the efficiency of the turbine[9].

5. DISCUSSION

Overall, we found that our work in feature engineering and data preparation coarsely affected the model's output accuracy, while the development of the variations in our sensitivity analysis served primarily to optimize the learning, and achieve an extra but limited improvement in its performance.

Firstly, when training deep learning models, it is extremely important to have a lot of data. The data we got from Myteodyn China was separated in three csv-files, where each of the files was data from a different wind turbine farm. Two of the three files had a lot of missing values for the target variable (*Park Power [KW]*) and because of the way we cleaned and made the windows, a significant portion of the data would be removed for two of the three farms. Therefore was it not possible to make adequate models based on the data for all three wind turbine farms.

The data from the Myteodyn China farms resulted in a usable prediction model, but our limited knowledge of the park power measurement and NWP's origin constitutes a significant limitation in our ability to develop the model, as well as our ability to generalize the model to other wind farms.

While our model's predictions have undoubtedly captured the trend of the observations in our validation set, there are still several intervals where its accuracy is affected by significant peaks that were predicted by the model, but which are not as apparent in the actual measurements for the corresponding period, and vice versa.

A possible significant factor in these predictions is the NWP data, on which our model is heavily dependant. While we believe the NWP to be predicted with an adequately well researched method, we have no way of knowing this is the case, since we were unable to obtain a detailed description of the methodology involved.

Further work with this issue could involve researching the accuracy of the NWP, to determine to what degree erroneous predictions for larger time intervals can be attributed to inaccuracy in the NWP.

In the development of our model, we followed the most promising path in terms of validation loss, with the loss ultimately determining the structure of our network. A few details of our network could perhaps hint at better performance, if we made an effort to research more complex model structures [10].

One such detail is the fact that the regularization technique dropout is not applicable to an LSTM with a single layer, which means the technique is not supported by our model. It is possible, that a model with more than one layers, and equal performance, would improve further with the introduction of the dropout technique.

6. CONCLUSION

Wind power forecasting is crucial to enhancing the stability of power integration. This project developed an LSTM-based prediction model, which provides short-term prediction of wind-based power generation.

Firstly, the visualization-based data analysis was conducted and outliers were removed to reduce the complexity of overall data understanding. Furthermore, more derived features with stronger predictive capabilities were created to improve the accuracy of modeling. In addition, data windowing was implemented, using 96 previous observations (24 hours) as a window to predict the 97th. Finally, a sole LSTM layer and a linear layer was chosen as the final model structure based on different hyperparameter combination experiments.

The final results showed that our model can effectively capture the dynamic trend of the wind energy, except for occasional missed peaks or dips in the park power.

7. FUTURE WORK

With regards to future work, the first step would be to experiment with the RNN model structure, because the initial structure is very simple, with only one LSTM layer. Many of the newer papers, which tackles wind power forecasting, explore the use of the Gated Recurrent Unit (GRU) [11] [12]. The reason that GRU would be interesting to explore is because it mostly achieve better results than LSTM [13] and its computational time is faster than LSTM. The main reason for this is because it have fewer gates and therefore fewer tensor operations [14].

It could also be interesting to experiment with the sequence length (window size). It would be possible that the model would get better results with another sequence length, instead of the chosen one day (96 time steps) sequence length. A half day (48 time steps) or two days (192 time steps) would be logical alternatives. Other papers have used optimization algorithms such as genetic algorithm (GA) to obtain the optimal sequence length [15], which also would be worth exploring.

8. REFERENCES

[1] Aoife M. Foley, Paul G. Leahy, Antonino Marvuglia, and Eamon J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1–8, 1 2012.

[2] Karim Moharm, Mohamed Eltahan, and Ehab Elsaadany, "Wind speed forecast using LSTM and Bi-LSTM algorithms over gabal el-zayt wind farm," *Proceedings - 2020 International Conference on Smart Grids and Energy Systems, SGES 2020*, pp. 922–927, 11 2020.

[3] Ma Lei, Luan Shiyan, Jiang Chuanwen, Liu Hongling, and Zhang Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915–920, 5 2009.

[4] M. A. Mohandes, T. O. Halawani, S. Rehman, and Ahmed A. Hussain, "Support vector machines for wind speed prediction," *Renewable Energy*, vol. 29, no. 6, pp. 939–947, 5 2004.

[5] Xiaoyun Qu, Xiaoning Kang, Zhang Chao, Jiang Shuai, and Xiuda Ma, "Short-term prediction of wind power based on deep Long Short-Term Memory," *Asia-Pacific Power and Energy Engineering Conference, APPEEC*, vol. 2016-December, pp. 1148–1152, 12 2016.

[6] Imre Delgado and Muhammad Fahim, "Wind turbine data analysis and lstm-based prediction in scada system," *Energies*, vol. 14, no. 1, 1 2021.

[7] Schmidhuber Hochreiter, "Long Short-term Memory," *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[8] Michael Phi, "Illustrated Guide to LSTM's and GRU's: A step by step explanation," 2018, [Online; accessed 6-December-2021].

[9] Danish Wind Industry Association, "Roughness and Wind Shear," http://drømstørre.dk/wp-content/wind/miller/windpower%20web/en/tour/wres/shear.htm, 2003, [Online; accessed 22-December-2022].

[10] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals, "Recurrent Neural Network Regularization," *ICLR 2015*, 09 2014.

[11] Yong Lin, Haiing Zhang, JiYan Liu, WenJie Ju, JinYou Wang, and XiaoDong Chen, "Research on short-term wind power Prediction of GRU based on similar days," *Journal of Physics: Conference Series*, vol. 2087, no. 1, pp. 012089, 11 2021.

[12] Qingyun Xie, Lianqing Song, Yongkang He, and Pengju Dang, "Short-term Wind Power Prediction Model Based on GSA Optimized GRU Neural Network," *Proceedings of 2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence, ICIBA 2020*, pp. 581–585, 11 2020.

[13] V. Bharat Kumar, V. Mallikarjuna Nookesh, B. Satya Saketh, S. Syama, and J. Ramprabhakar, "Wind Speed Prediction Using Deep Learning-LSTM and GRU," pp. 602–607, 11 2021.

[14] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1724–1734, 2014.

[15] Farah Shahid, Aneela Zameer, and Muhammad Muneeb, "A novel genetic LSTM model for wind power forecast," *Energy*, vol. 223, 5 2021.