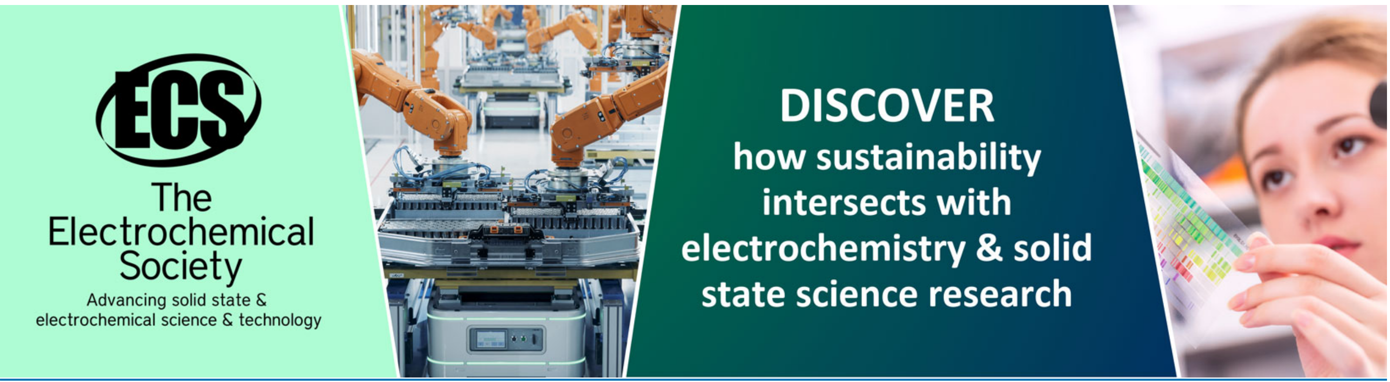
IOP Conference Series: Earth and Environmental Science

Short-Term Wind Speed Prediction based on Deep Learning

To cite this article: Jingchun Chu *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **233** 052007



IOP Conf. Series: Earth and Environmental Science 233 (2019) 052007 doi:10.1088/1755-1315/233/5/052007

Short-Term Wind Speed Prediction based on Deep Learning

Jingchun Chu[1,2], Ling Yuan[1,2], Wenliang Wang[1,2], Lei Pan[1,2], and Jie Wei[1]

[1]Guodian United Power Technology Company Ltd, Beijing, P.R. China 100039

[2]Wind Power Equipment and Control State Key Laboratory, Baoding Hebei 071000

Abstract Wind speed forecasting has great significance to the improvement of wind turbine intelligent control technology and the stable operation of power system. In this paper, the Long Short-term Memory (LSTM) mode with deep learning ability combined with the fuzzy-rough set theory has been proposed to do short-term wind speed prediction. Fuzzy rough sets can reduce input and spatial characteristics. The main factors affecting wind speed were found as input of the prediction model of LSTM neural network. Deep learning conforms to the trend of big data. It has strong generalization ability on massive data learning. The experimental results show that the Fuzzy rough set Long Short-term Memory (FRS-LSTM) model has higher prediction accuracy than traditional neural network.

1.Introduction

The short-term prediction of the wind speed can optimize the peak regulation, reactive power and voltage control[1,3]. In addition, precise prediction can also optimize the control strategy, improve the power generation efficiency, ensure the power quality and reduce the operating cost of the power system [4].

The wind speed prediction method can be divided into two categories: physical method[5-6] and statistical method[7,11]. In these methods, the neural network with its superior nonlinear fitting and generalization ability becomes the most widely used method. The traditional neural network has two disadvantages:(1) stringent requirements for input variables and training samples, too much or too little data inputs will affect the training effect. (2) When the dimension of the feature is too large and it's difficult to effectively extract high-quality features, the neural network can hardly obtain good results.

For the first disadvantage, compared with the current mainstream approach of Principal Component Analysis (PCA)[16], fuzzy-rough set can effectively avoid the loss of important information during the process of dimension reduction. It can discover implied knowledge and reveal the underlying laws by analyzing the incomplete information and reasoning data[17-21].

Deep learning can be a good solution to the second disadvantage. It does not rely on high-quality features, can learn the massive data and has strong generalization ability[12-15]. LSTM, as an excellent variant of the RNN model, solves the problem of gradient disappearance in the process of gradient back propagation and is suitable for dealing with the problems of high correlation with time series[13].

In summary, this paper presents a short-term wind speed prediction model based on FRS-LSTM. Firstly, the model is dimensionally reduced by using the fuzzy rough set, and the main factors affecting wind speed are found, which are used as the input of the prediction model of LSTM neural network. Secondly, LSTM neural network is used for training and learning. Finally, the prediction is made by using the model.The experimental results show that the FRS-LSTM prediction model has higher prediction accuracy than traditional neural network.

2.The Fuzzy Rough Set

Rough set theory can be a good way to reduce the loss of information by using the exact sets, such as upper approximation set, lower approximation set, etc, to describe the uncertainty and ambiguity of knowledge [17,19,21]. The fuzzy sets are used to replace the exact sets and the corresponding upper and lower approximate sets are extended to the fuzzy upper and lower approximate sets respectively. Defined as:

$$u_{p_i x} = \inf_{x} \max\{1 - m_{F_i}(x), m_X(x)\}, \forall i \tag{1}$$

$$u_{p_s x} = \sup_{x} \max\{m_{F_i}(x), m_X(x)\}, \forall i \tag{2}$$

The fuzzy positive domain of the fuzzy equivalence class is defined as:

$$u_{POS_P}(F_i) = \sup_{X \in \frac{U}{Q}} u_{p_i x}(F_i) \tag{3}$$

The membership degree of the object x in the domain U is defined as:

$$u_{POS_P}(x) = \sup_{F_i \in \frac{U}{P}} \min(u_{F_i x}(x), u_{POS_P}(F_i)) \tag{4}$$

For the degree of dependence of the expression condition attribute P on the decision attribute Q, the fuzzy dependency function is defined as:

$$\gamma_P(Q) = \frac{|u_{POS_P}(x)|}{|U|} = \frac{\sum_{x \in U} u_{POS_P}(x)}{|U|} \tag{5}$$

The importance of the $p \subseteq P$ about Q is defined as:

$$\sigma_{PQ}(p) = \gamma_{P+\{p\}}(Q) - \gamma_P(Q) \tag{6}$$

The algorithm pseudo-code is shown in Fig 1. First, choose an empty set R as the initial set. Then, add attributes to the R collection and determine the dependency changes after adding the attributes. If the dependency is increased, it means that after adding the attribute. The corresponding classification capacity increases, R and C have the same categorization ability. At this point, R is the result of the reduction of c attribute, the pseudo-code for the fast reduction algorithm.

$$
\begin{array}{l}
\text{I.}R \leftarrow \{\ \} \\
\text{II.}do \\
\text{III.}T \leftarrow R \\
IV.\forall x \in (C - R) \\
V.if\ \gamma_{R \cup \{x\}}(D) > \gamma_T(D) \\
V\text{I.}T \leftarrow R \cup \{x\} \\
V\text{II.}R \leftarrow T \\
V\text{III.}until\ \gamma_R(D) = \gamma_C(D) \\
\text{I}X.returnR
\end{array}
$$

Fig. 1 Pseudo code for Quick Restore Algorithm

3. The LSTM Neural Network

RNN is prone to gradient disappearance or gradient explosion problems when learning long-term dependencies. As an improved version of RNN, LSTM can learn long-term dependency information and avoid the problem of gradient disappearance[12-13]. In the neural nodes of RNN's hidden layer, LSTM adds a structure called Memory cell to remember past information, and adds three gates (input, forget, and output) to control the use of historical information. Figure 2 is the LSTM neuron structure diagram. The middle part of the diagram is one Memory Block. Each block contains a number of Memory Cells controlled by a set of Gate. Memory Cells in the same Memory Block are under the control of the same set of Gate. Control information flow in and flow out through the Input Gate and the Output Gate. Decide whether to reset the network through the Forget Gate. All the actions are jointly controlled by the output of previous layer, the output of the hidden layer on the last moment, the memory unit information[22-23]. Assuming that the input sequence is $(x_1, x_1, \ldots, x_r)$, the hidden layer state is $(h_1, h_2, \ldots, h_r)$, then at time t:

$$i_t = sigmoid(w_{hi}h_{t-1} + w_{xi}x_t) \tag{7}$$

$$f_t = sigmoid(w_{hf}h_{t-1} + w_{xi}x_t) \tag{8}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tanh(w_{hc}h_{t-1} + w_{xc}x_t) \tag{9}$$

$$o_t = sigmoid(w_{ho}h_{t-1} + w_{hx}x_t + w_{co}c_t) \tag{10}$$

$$h_t = o_t \otimes \tanh(c_t) \tag{11}$$

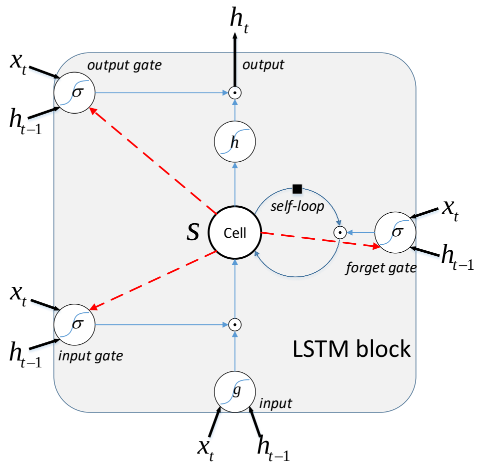


Fig. 2 Structure diagram of LSTM

4. FRS-LSTM Model for Wind Speed Forecasting

4.1 FRS-LSTM Model

The prediction model is composed of two parts: fuzzy-rough set factor reduction (FRS) and neural network prediction (LSTM). The space dimension of input information is simplified by using FRS sensitivity to noise. Then take them as the input for the prediction part of the LSTM neural network. Properly compress the data variables, use LSTM to do the training and learning. The approximate implicit input-output nonlinear relation is extracted to obtain the wind speed prediction effect.

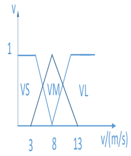
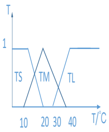
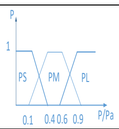
4.2 Reduce attribute

Fuzzy rough set theory is used to reduce attributes and remove redundant information[20,21].

First, the initial decision table was determined with the predicted wind speed v (t + 1) as the decision attribute and 35 influencing factors as the condition attribute.

Then select the appropriate fuzzy membership function and fuzzy the attributes according to the physical attributes of each attribute. Table 1 introduces several selections of fuzzy membership functions.

Table 1 Fuzzy membership function of wind speed, temperature and air pressure

category	fuzzy membership functions	description
speed		The wind speed is divided into low wind speed, wind speed, high wind speed three states $U/v=\{V_S,V_M,V_L\}$
temperature		Normalized treatment.
Pressure		$P_i'=\frac{p_i-p_{\min}}{p_{\max}-p_{\min}}, i=1,2,\ldots,n$

Finally, simply the 35 condition attributes according to the attribute membership functions defined above.

In this paper, we use the python language to complete the construction of the LSTM neural network which is based on the neural network learning framework. Here's the simplified model: 3 layers of LSTM, 108 nodes per layer, uses the rolling prediction method, selects the data of the first 10 days as the training samples to predict the 144 wind speed of the 11th day, selects the data of the 2nd day to the 11th day as the training samples to predict the 144 wind speed of the 12th day, and so on[23].

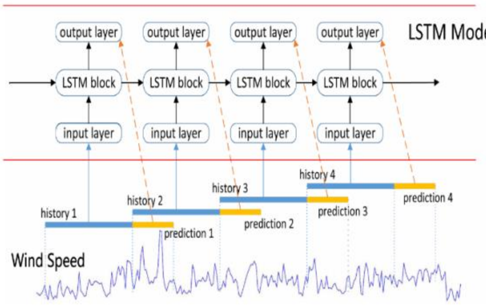
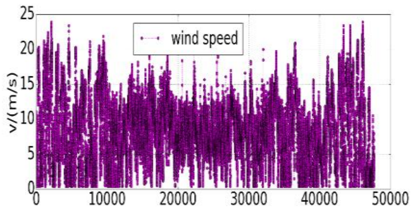


Fig. 3 LSTM prediction model

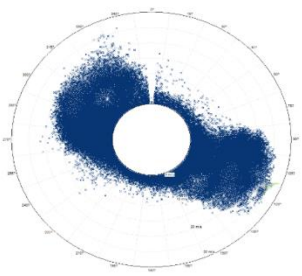
The input of neural network is simplified attribute set after reduction of fuzzy rough set. The output is the predicted wind speed at the moment of v(t + 1), and the specific model of LSTM is shown in figure 3.

The original data was obtained from a wind farm in Xinjiang on January 1, 2013, solstice, 2016. The wind speed sampling height is 50m. The sampling interval is ten minutes. The specific acquisition parameters are wind speed, air density, temperature, wind direction, air humidity, air pressure and so on. Wind speed time series are shown in Figure 4-a, 4-b:

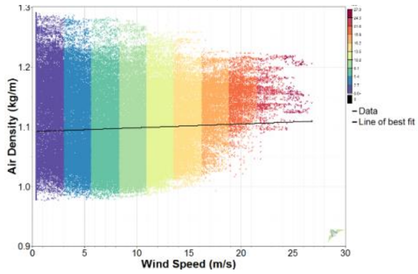
Air density is mainly concentrated to 1.0-1.3, Figure 4-c shows the relationship between wind speed and air density.



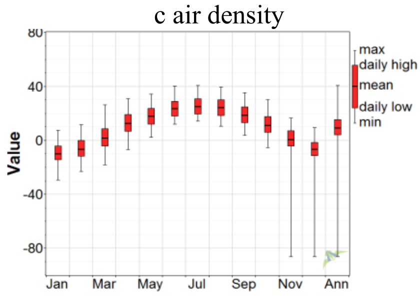
a speed



b rose



c air density



d temperature

Fig. 4 Wind parameters

5. The Experiment

By setting the threshold value, the attributes are added in turn to calculate the dependencies, and the condition attributes are retained or not based on the increase and decrease of the dependencies. The final set of attributes has been defined as:

$$R=\{v(t),v(t-1),v(t-2),v(t-3),v(t-3),T(t),T(t-1)\} \tag{12}$$

The attributes in the set represent the wind speed and temperature at all times. From the final simplified attribute set, we can see that air density, humidity and wind direction have little effect on the wind speed. Removing these redundant parameters can optimize the training speed and the accuracy of the neural network.

The data is extracted from the training parameters determined by the final attribute set. They are used as input information of LSTM neural network. The main steps are as follows:

(1) The activation function of LSTM module is defined as Relu. The Rectified Linear Unit is $f_x=\max(0,x)$.Compared with the common sigmoid and tan functions, Relu activation function has two main advantages. First, gradient is not saturated. So in the reverse propagation process, gradient dispersion problem has been reduced. Second, calculation speed is faster. In the forward propagation process, only need to set up the threshold value to speed up the calculation of forward propagation, so as to greatly improve the convergence rate.

(2) To determine the activation function of fully connected artificial neural network, select the default linear activation function in keras.

(3) The cut-off rate of nodes in each layer is set to 0.2 to prevent over fitting in the wind speed prediction training.

(4) Determine the error evaluation index. Select the average absolute error (MAE) and the average

absolute percentage error (MAPE), the expression as shown in (12),(13).

$$E_{MAE} = \frac{1}{N}\sum_{i=1}^{N}|v_i - v_i'| \tag{13}$$

$$E_{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|v_i - v_i'|}{v_i} \tag{14}$$

(5) Determine the iterative update method of the weight parameter to select the optimizer. Using the RMS prop algorithm, avoid the disappearance of the learning rate by setting a certain ratio between the historical update and the gradient squared. Set the learning rate as 0.001.

(6) Add in Batch Normalization. In the training of deep neural networks, the input distribution of each layer will change along with the change of layer parameters. This is called the Internal Covariate Shift. It will bring additional parameter conversion costs. However, the problem can be solved effectively and the training speed can be improved by adding in Batch Normalization.

(7) Determine time-steps. There is a strong correlation between the front wind speed and the rear wind speed in the repeated tests.

6. Experiment Results

By changing the number of layers, the size of the layers, the sequence length or any one of the other adjustable hyper parameters for this model could to performed an exhaustive training on this model and achieve better results. Fig.5 is one of the LSTM model. Fig 6 is training and evaluation models.

So, the data itself has a natural continuity in time. Although the convergence speed of small time steps is fast, the accuracy is not high. Excessive time steps make the model training harder and it may reduce the accuracy of the model because of the loss of the continuity. So, we should find an optimal time steps.

Layer (type)	Output Shape	Param #
lstm_16 (LSTM)	(None, None, 20)	2320
dropout_16 (Dropout)	(None, None, 20)	0
lstm_17 (LSTM)	(None, None, 15)	2160
dropout_17 (Dropout)	(None, None, 15)	0
lstm_18 (LSTM)	(None, 20)	2880
dropout_18 (Dropout)	(None, 20)	0
dense_6 (Dense)	(None, 4)	84

Fig. 5 One of the LSTM model

```
Train on 55858 samples, validate on 6207 samples
Epoch 1/50
55858/55858 [==============================] - 26s - loss: 0.8714 - val_loss: 0.8091
Epoch 2/50
55858/55858 [==============================] - 25s - loss: 0.7636 - val_loss: 0.7331
Epoch 3/50

...
```

Fig. 6 Training and evaluation models

The comparison diagram of training loss and verification loss is drawn to better understand whether the network is over fitting or not. Figure 7 is the comparison diagram of verification loss and loss.

Fig. 7 loss and val-loss

Forecast wind speed using BP, fr-bp, LSTM and fr-lstm.

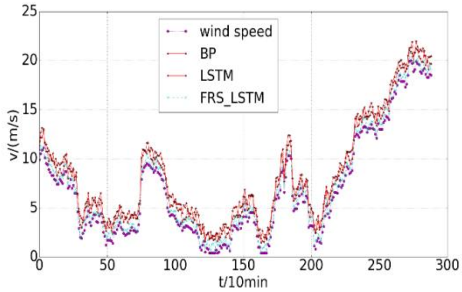


Fig. 8 Forecasting results

The actual wind speed value and two estimated values are compared in Figure 8.

Comprehensive comparison found:

1) It can be seen from Table 2 that using LSTM is better than BP neural network.

2) FRS-LSTM has the best effect by using FRS and has the optional training speed.

Table 2 Comparison of prediction results

method	E_{MAE}	E_{MAPE}	E_{MAX}	PCT/%	
				$E_{AE}>1$	$E_{APE}>1$
BP	0.8321	0.0860	3.4026	35.03	28.18
LSTM	0.6410	0.0672	2.1181	23.90	21.34
FRS-LSTM	0.5432	0.0516	1.9837	15.11	19.89

7.Conclusion

In this paper, the concept of deep learning is introduced into the wind turbine. LSTM neural network realizes the prediction of wind speed through the learning of various parameters. It can provide important support for the smooth operation of power system and the optimization of control strategy. The fuzzy rough set theory is used to reduce many factors that affect wind speed. It simplifies the input of the neural network prediction model and improves the accuracy and speed. Compared with the traditional neural network prediction method, MAE and MAPE in FRS-LSTM wind speed forecasting model have decreased and the accuracy has been improved greatly.

References

[1] WangQ. (2016) Mechanism of the Imapct of DFIGs on Power System Stability[D]. North China Electric Power University(Beijing) 51-59.

[2] Zhang ZY. (2016) Effect of Large Scale Wind Power Integration on Power System AGC Control Parameters[D]. North China Electric Power University(Beijing) 28-32.

[3] Yu Q,Sun H D,Tang Y,et al. (2013) Impact on Angle Stability of Power System With Doubly Fed Induction Generators Connected to Grid[J]. Power System Technology(12):3399-3405.

[4] Zhao H S,Liu X J,Li C. (2013) Active power control strategy of wind farm based on dynamic

wind speed information of unit[J]. Electric Power Science and Engineering, (01):9-15.
[5] Wang H, Liu D, Wang J L. (2015) Ultra-Short-Term Wind Speed Prediction Based on Spectral Clustering and Optimized Extreme Learning Machine[J]. Power System Technology,(05):1307-1314.
[6] Han S, Yang Y P, Liu Y Q. (2008) Application study of three methods in wind speed prediction[J]. Journal of North China Electric Power University,(03):57-61.
[7] Zhou J Q, Li Y, Qu W D,et al. (2015) A Study on BP Neural Network Ultra-Short Term Wind Speed Forecast Based on Historical and Spatial Wind Speed Data[J].Power System and Clean Energy, (01):109-112.
[8] Liu X J,Cen T Y,Zheng W S,et al. (2014) Neural Network Wind Speed Prediction Based on Fuzzy Rough Set and Improved Clustering[J]. Proceedings of the CSEE, (19):3162-3169.
[9] Yin D Y, Sheng Y F, Jiang M J, et al. (2014) Short-term wind speed forecasting using Elman neural network based on rough set theory and principal components analysis[J]. Power System Protection and Control (11):46-51.
[10] Zhu X Y, Zhang J H, Fu S P,et al. (2013) Optimization of SVM Parameters Based on Improved PSO and Its Application in Wind Speed Prediction[J]. ElectricPower, (11): 105-108.
[11] Zeng J, Zhang H. (2009) A Wind Speed Forecasting Model Based on Least Squares Support Vector Machine[J]. Power System Technology, (18):144-147.
[12] Lecun Y, Bengio Y，Hinton G．(2015) Deep learning[J]．Nature 521(7553): 436-444.
[13] Pulver A,Lyu S. (2017) LSTM with Working Memory[J]. 339-348.
[14] Zhao R, Yan R, Wang J, et al. (2017) Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks[J]. Sensors, 17(2).
[15] Liu L, Chen W, Cao G. (2016) Prediction of neonatal amplitude-integrated EEG based on LSTM method[C]// IEEE International Conference on Bioinformatics and Biomedicine. IEEE Computer Society, 2016: 497-500.
[16] Tang Y B, Gui W H, Peng T, et al. (2014) Transformer Fault Diagnosis Model Based on PCA and KICA Feature Extraction[J]. High Voltage Engineering, 2014, (02): 557-563.
[17] LI C．(2016).Research on Multi - granular Fuzzy Rough Set[J]. J．of Math, 2016, (01): 124-134.
[18] Deng D Y, XU X Y, Pei M H. (2015) F-fuzzy rough sets and its reducts[J]．Journal of Zhejiang Normal University(Nat． Sci. (01): 58-66.
[19] Wang X F,Wang S H, Du H Z,et al. (2015).Fault diagnosis of chemical industry process based on FRS and SVM[J]. Control and Decision, (02): 353-356.
[20] Xu J Q. (2011) Representation and Application of Fuzzy Rough Set[J]．Fuzzy Systems and Mathematics, 2011, (01): 146-152.
[21] Huang Z H ,H B Q. (2005) Research on the Theory of Fuzzy Rough Set Theory[J]．Fuzzy Systems and Mathematics, 2005, (04): 125-134.
[22] Zhao Z, Chen W, Wu X, et al. (2017) LSTM network: a deep learning approach for short-term traffic forecast[J]．Iet Intelligent Transport Systems, 11(2): 68-75.
[23] Yang X Z, Ke Y Y, Liang X,et al. (2016).A Study on the Prediction of Generator Set Pollutant Emissions Based on LSTM[J]. Electrical Automation, (05): 22-25.
[24] Woo W C, Wong W K, Shi X J,et al. (2016) A deep-learning method for precipitation nowcasting[C]// Wmo Wwrp，International Symposium on Nowcasting and Very-Short-Range Forecast.