Energy Engineering
DOI:10.32604/EE.2020.011619
Article

Deep Learning Approach with Optimizatized Hidden-Layers Topology for Short-Term Wind Power Forecasting

Xing Deng[1,2] and Haijian Shao[1,2,*]

[1]School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, 212003, China
[2]School of Automation, Key Laboratory of Measurement and Control for CSE, Ministry of Education, Southeast University, Nanjing, 210096, China
[*]Corresponding Author: Haijian Shao. Email: jsj_shj@just.edu.cn
Received: 20 May 2020; Accepted: 13 August 2020

Abstract: Recurrent neural networks (RNNs) as one of the representative deep learning methods, has restricted its generalization ability because of its indigestion hidden-layer information presentation. In order to properly handle of hidden-layer information, directly reduce the risk of over-fitting caused by too many neuron nodes, as well as realize the goal of streamlining the number of hidden layer neurons, and then improve the generalization ability of RNNs, the hidden-layer information of RNNs is precisely analyzed by using the unsupervised clustering methods, such as Kmeans, Kmeans++ and Iterative self-organizing data analysis (Isodata), to divide the similarity of raw data points, and maps the hidden-layer information into the feature space where data separation is easily implemented. Experiments based on dataset from the National Renewable Energy Laboratory (NREL) is proposed to demonstrate the performance of the proposed approaches, the average forecasting errors of which is respectively increased by 2.1%, 7.6%, 10.26% with respect to 6-steps, 12-steps and 18-steps in four seasons over the ones that achieved using the traditional deep learning approaches.

Keywords: Deep learning; hidden-layer information; cluster analysis; model architecture optimization

1 Introduction

Energy is an essential material basis for human social progress and economic development. Traditional energy such as coal, oil and gases has not only limited quantity but also brings potential pollution to the human survival environment. In the past decades, the social development is affected by the oil crisis and various climate change factors, so the development of renewable energy gradually becomes the consensus of the international community. Wind energy as a renewable energy is produced by the air flow acting and its distribution is randomness, instantaneity and seasonal in short-term, which has the inhibitory effect for the accurate wind speed forecasting. Accurate short-term wind speed forecasting (STWSF) can reduce the economic losses caused by grid integration at all levels of the transmission and distribution in grid. In order to ensure the safety of wind farm system. Especially offshore, the demand for organized operation and maintenance strategies is relatively high, such as the decomposition-based hybrid time series forecasting methods could provide the comprehensive review of wind power forecasting. Faults diagnosis is an effective approach to provide the reliable strategy to ensure that the wind turbine can perform the high availability and low energy consumption.

Deep learning approach such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), is a computational model to handle the time series, which has been widely used in dermatologist-level classification [1], health care [2], power conversion systems [3,4], load forecasting [5,6], fault-tolerant predictive power control [7], GPU computing [8] and natural language processing [9]. However, the improper analysis of hidden-layers information still seriously hampers the performance of deep learning. For static images with very high practical value, Esteva et al. [1] designed a deep neural network with multiple hidden layers to analyze the dermatologist-level classification of skin cancer images. Experiments show that deep learning can effectively assist dermatologists to improve the accuracy of cancer detection. The final classification accuracy could be essentially further improved if the architecture designs of hidden-layer in CNN, such as pooling and information presentation are properly considered. The hierarchical representations of the hidden-layer information had been extracted via a deep learning method by Young et al. [9] to analyze the numerous data processing tasks. Experiments indicated that multiple hidden-layers benefit the efficiency improvement of the deep learning. However, the model's performance was still limited especially since the information of multilayer hidden-layer is not quantitatively processed. Even for data with higher dimensions, the deep learning method can still effectively improve the prediction accuracy of secondary structure, accessible surface area (ASA) with real representing local and nonlocal structure [10], etc.

Taking into account the wind power forecasting, the hybrid forecasting method [11–15] usually treated as the soft computing method can combine with different processing approaches to improve the forecasting accuracy compared with the physical and statistical methods. The short-term wind power forecasting process based on RNNs is established through the repeated iterations of recursive results obtained by preciously steps. The hidden-layers of RNNs can compress the input sequence or feature information into a vector, and acting as the input of the arbitrarily time, and then propagates the error information in reverse order over time to adjust the network architecture. Analogously, the processing of the hidden-layers is similar to "Black Box", that means the topology and information representative cannot be handled properly. This undoubtedly increases the difficulty of RNNs network parameter adjustment, hidden-layer topology optimization and high-precision modeling [16–19]. In addition, the improper representation of hidden-layers information results the generation of the limited information in coverage that may decrease over time. Draye et al. [13] used the clustering analysis to divide the hidden-layer information in the training network into several sets composed by similar objects, and the associated information is entered into the corresponding neurons by means of fixed weights and presetting metric. Experiments indicated that this method can effectively control the flow of hidden-layer information, and bring a remarkable effect on optimizing neural network structure to improve the computing performance. However, due to the improper processing of hidden-layer information, its iterative pattern appears after merging, leads to the untimely updating of information. There are also many kinds of information modes in hidden-layer because of the diversity of input variables, the predictive accuracy and processing efficiency of deep learning can be further improved if hidden-layer information can be properly handled.

The rest of this paper is organized as follows, the proposed approaches, such as the hidden-layer information analysis in RNN and the model evaluation indicator are considered in Section 2, and the experimens based on the hidden-layer information analysis is given to demonstrate the performance of the proposed approaches in Section 3. This paper and further work is concluded in Section 4.

2 Proposed Approaches

The motivation and objectives of this paper is the formulation of the statistical analysis of the hidden-layer information in RNNs, streamline the number of network neurons and optimize the network's architecture. The processing flow framework of this paper is given in Fig. 1.

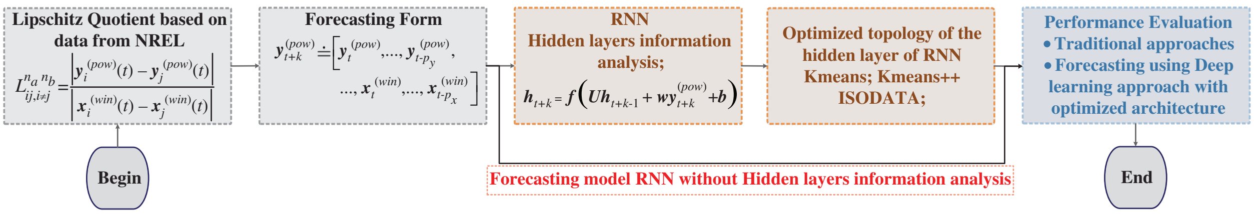


Figure 1: Processing flow framework of the proposed deep learning approach

Firstly, Lipschitz quotient is used to estimate the time dealy that used as the forecasting model order, and organize the available inputs into an proper used for forecasting. Secondly, three unsupervised algorithms, i.e., Kmeans, Kmeans++ and Isodata, are used to analyze the distribution probability of the hidden-layer information in RNNs, and obtain the approximate number of the hidden-layers neurons for forecasting. Finally, experiments based on the proposed approaches is evaluated based on the data from NREL, and compared with the deep learning approaches in previous works.

2.1 Hidden-Layer Information Analysis

The time-frequency analysis method wavelet analysis as well as the Lipschitz quotient are respectively used to avoid the local transient that may propagates over time and estimate the model order based on the proposed approach in [20]. The dataset utilized in this paper come from NREL. Every sample comes with two fields: wind speed (M/S, 80M) $\boldsymbol{x}_t^{(win)}$ and netpower (MW) $\boldsymbol{y}_t^{(pow)}$ at the t-th time instance, respectively. Let time sequence $\boldsymbol{y}_{t+k}^{(pow)} \mathrm{B} \left[\boldsymbol{y}_t^{(pow)}, \ldots, \boldsymbol{y}_{t-p_y}^{(pow)}, \ldots, \boldsymbol{x}_t^{(win)}, \ldots, \boldsymbol{x}_{t-p_x}^{(win)}\right]$ be the inputs, where p_y and p_x are positive integers defined by Lipschitz quotients and treated as time-lags associated with the wind power $\boldsymbol{y}_t^{(pow)}$ and the wind speed $\boldsymbol{x}_t^{(win)}$. Then the information at the hidden-layers refer to the precious inputs and current one can be represented as follows:

$$\boldsymbol{h}_{t+k} = \boldsymbol{f}\left(\boldsymbol{U}\boldsymbol{h}_{t+k-1} + \boldsymbol{w}\boldsymbol{y}_{t+k}^{(pow)} + \boldsymbol{b}\right) \tag{1}$$

where $\boldsymbol{f}$ is the nonlinear transformation associated to Hyperbolic tangent (tanh) function etc, and $k \in N^+$ is a factor to measure the proceeding k-ahead forecasting intervals, such as short-term, medium-term or long-term forecasts. Machine learning methods are mainly divided into two categories: supervised learning and unsupervised learning. The former usually by acquiring the characteristics of the training sample to establish the model used to describe the given data and its corresponding output. Clustering analysis as representative methods for latter typically used to divide the similarity of raw data points according to preset metrics has been extensively used in various similarity analysis, which can establish the model directly without any prior knowledge such as Kmeans. Even the expression of information is abstract, especially manual labeling of samples with large volume is more difficult due to lack of samples' prior knowledge. In particular, not all data are independent and identically distributed (i.i.d), and subjective labeling is prone to make errors. Because of the complex distribution of hidden-layer information in deep learning, unsupervised learning can effectively reduce the possibility of samples being analyzed incorrectly and improve the accuracy of information analysis. For the given observation set $\boldsymbol{S}$,

$$\boldsymbol{S} = \{\boldsymbol{S_1}, \boldsymbol{S_2}, \ldots, \boldsymbol{S_T}\} = \{\boldsymbol{h_{1+k}}, \boldsymbol{h_{2+k}}, \ldots, \boldsymbol{h_{T+k}}\}, \boldsymbol{k} \in \boldsymbol{N}^+ \tag{2}$$

Kmeans as the most widely used unsupervised algorithm for the hidden-layer information analysis, the primary schedule of which is to minimize the within-cluster sum of squares (WCSS) according to

$$\arg\min_{\boldsymbol{S}} \sum\nolimits_{t=1,\ldots,T} \sum\nolimits_{k\in N^+} \left\|\boldsymbol{h_{t+k}} - \mu_{\boldsymbol{t+k}}\right\|^2 = \arg\min_{\boldsymbol{S}} \sum\nolimits_{t=1,\ldots,T} |\boldsymbol{S}_i| \mathrm{Var}\boldsymbol{S}_i \tag{3}$$

where $\mu_{\boldsymbol{t+k}}$ is the mean of data in $\boldsymbol{S}$ constructed by hidden-layers information. The basic principle of Kmeans is to allocate the observed data to the nearest clustering according to the distance by minimizing WCSS, but the global optimal solution cannot be guaranteed because the Kmeans usually converges to a local optimal solution. The signals' spectrum distribution is more complex because of the complexity of data distribution, such as the instantaneous and intermittent nature of wind power, which is still linear separable even if the original data is mapped into a higher dimensional feature space. The wind power time series with arbitrary length can be effectively processed in RNNs since both of the current inputs and previous one could be connected between neurons, and the short-term dependence between different states will be sufficiently reflected in short-term wind power forecasting.

2.2 Model Performance Measurement

In order to guarantee the accuracy of clustering analysis, ISODATA based on soft assignment known as the dynamical clustering or iterative self-organizing data analysis, can used to approximate the samples' essence through gradual evolution in the process of fuzzy clustering. ISODATA can effectively overcome the shortcoming of other clustering methods, such as the initial classification is not easy to extract the essential attributes of data, and increase the probability of getting better clustering results. The aforementioned clustering method will be used to analyze the hidden-layers information and evaluate their similarity in deep learning. The final forecasting result is validated by the performance criteria, such as Root mean square error (RMSE), to evaluate the forecasted errors in forecasting modeling. RMSE is defined as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1,\ldots,n} \left(\boldsymbol{y}_{tf}^{(pow)} - \hat{\boldsymbol{y}}_{tr}^{(pow)}\right)^2}{\sum_{t=1,\ldots,n} \left(\boldsymbol{y}_{tf}^{(pow)}\right)^2}} \tag{4}$$

where $\boldsymbol{y}_{tf}^{(pow)}$ and $\hat{\boldsymbol{y}}_{tr}^{(pow)}$ are the observation vector and the forecasting vector, respectively. The aforementioned clustering method based on the performance criteria in deep learning will be used to analyze the hidden-layers information and evaluate their similarity in experiments.

3 Experiments

The data used for the experimental evaluation come from the NREL, with the interval from January 1, 2004 to December 31, 2004. The data preprocessing method, Lipschitz quotient and wavelet analysis are respectively used to improve the quality of the utilized data, estimate the model order and avoid the meteorological time series' local transient feature that may propagated over time. In order to promote the precise analysis of the hidden-layer information, Kmeans++ which is generally similar Kmeans has been added and as the benchmark methods for comparison in this experiment. The only significant difference between them is that the initial centroids of the Kmeans++ are not randomly selected, but implemented by fitness proportionate selection through adding the initialization procedure with higher probabilities that may be chosen as the next centroid. All the clustering results in four season (spring, summer, autumn and winter) achieved by four unsupervised algorithms, i.e., Kmeans, Kmeans++ and ISODATA have been shown in Fig. 2. The corresponding standard deviation of the centroids distances of the used four clustering methods has been listed in Tab. 1.

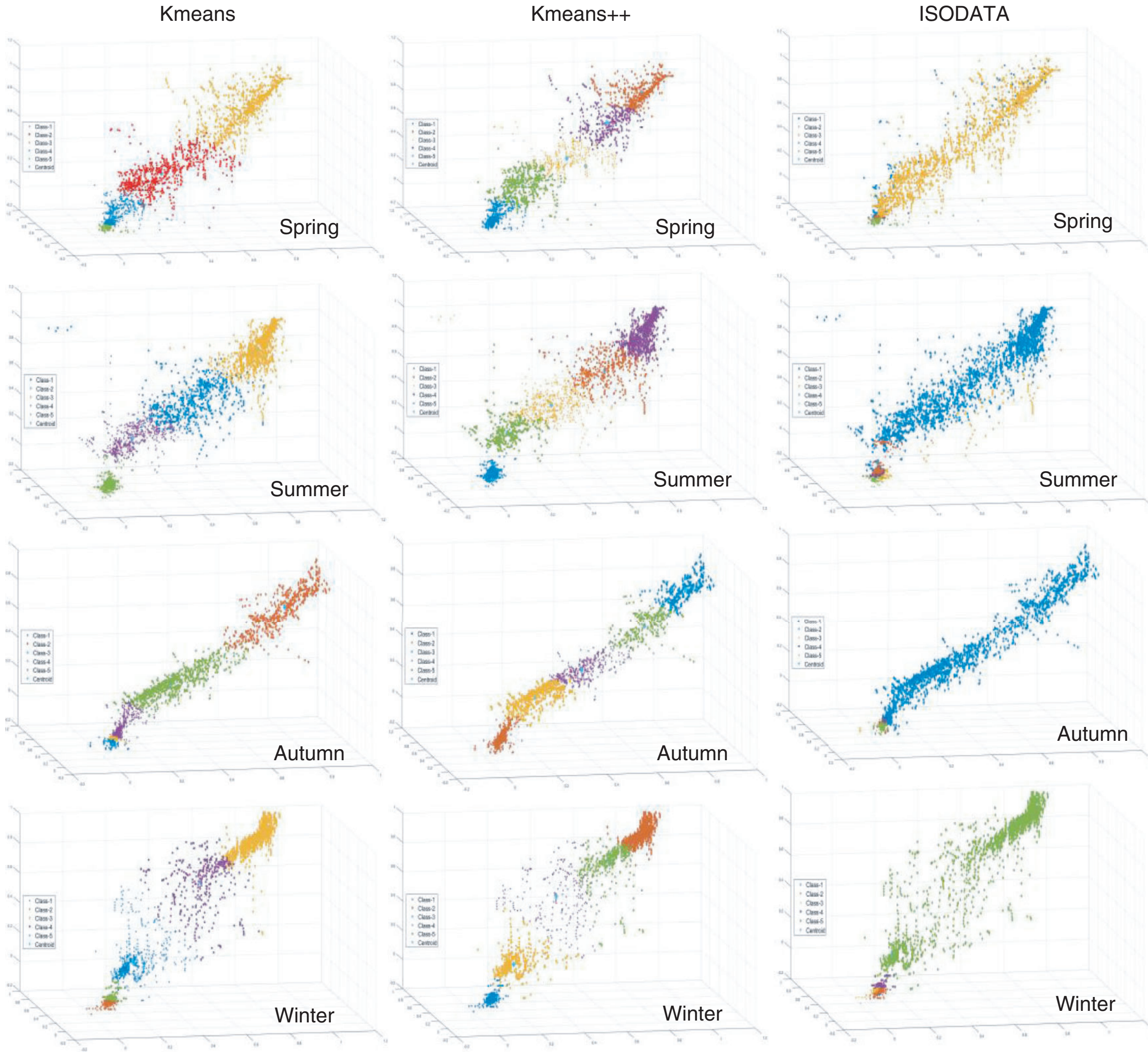


Figure 2: The clustering result of the hidden-layers information

The three columns in Fig. 2 shows the results of different clustering results for the same data in the four seasons. Essentially, the clustering results obtained by Kmeans and Kmeans++ are similar with each other even under different cluster centers. Although the clustering analysis and similarity assessment results derived based on a given European distance measurement, and still can ensure convergence to a certain cluster center, however, there are two disadvantages of the outlined methods: 1) there is no guarantee that if the cluster center are best in the given data, and 2) there is no guarantee of consistency in the judgment of different sample categories, such as the similar results for wind energy data analysis in spring and autumn, while the results of the relatively dramatic summer and winter clustering results are different. In addition, the K-value of the Kmeans and Kmeans++ should be pre-setted first. Because random initial center selection has an effect on the calculation results, so this paper uses multiple experiments to average the analysis. The ISODATA algorithm is based on the Kmeans, to increase the cluster results of two operations: "splitting" and "merging". That is, the two cluster centers are merged into one when the value of the center of two clusters is less than a certain threshold. The cluster will be splitted into 2 clusters when the standard difference of a cluster is less than a certain threshold or the number of cluster samples exceeds a certain threshold. The cluster analysis results obtained by ISODATA in Fig. 2, the cluster center distance of which compared with two aforementioned methods is relatively closer, and the same type of data are divided into more subgroups. In fact, ISODATA is acturally sensitive to the noise of the data, so the data category gap is relative small, so according to three outlined cluster analysis results of the hidden layer information, too many hidden nodes can easily cause over-fitting in the foreasting.

Table 1: Standard deviation of the centroids distances in clustering methods

Methods	Seasons	1	2	3	4	5	6
Kmeans	Spring	0.0018	0.3889	0.3402	0.3239	0.3229	0.3274
	Summer	0.0015	0.3714	0.3403	0.3165	0.3202	0.3236
	Autumn	0.0036	0.3477	0.3264	0.3119	0.3044	0.3021
	Winter	0.0086	0.4323	0.3638	0.3490	0.3507	0.3427
Kmeans++	Spring	0.0018	0.4394	0.4030	0.3081	0.3199	0.2870
	Summer	0.0015	0.4886	0.2944	0.2982	0.2971	0.3510
	Autumn	0.0036	0.1130	0.3805	0.2839	0.2767	0.3124
	Winter	0.0086	0.3096	0.3751	0.3391	0.3127	0.3242
ISODATA	Spring	0.0004	0.0024	0.0033	0.0075	0.0038	0.2912
	Summer	0.0010	0.0174	0.0026	0.0068	0.0049	0.0041
	Autumn	0.0019	0.0017	0.0032	0.0022	0.0019	0.0037
	Winter	0.0011	0.4686	0.0058	0.0025	0.0585	0.0146

The standard deviation of the centroids distances listed in Tab. 1 indicates that the overall deviation under the Euclidean measure corresponding to last two columns is less than 0.01. Approximate number of the hidden-layers can cover all the information needed for forecasting modeling, based on the aforementioned analysis, this indicates that 6 could be considered as the number of hidden layer neurons, and the input signal can ensure that its characteristics can fully stimulate all modes needed for RNNs. Both Levenberg-Marquardt (LM) and Quasi-Newton methods (BFGS) are used in RNNs. The $k = (6, 12, 18)$ steps (corresponding to 1, 2 and 3 hours) ahead of actual wind power forecasting are provided in this section, and the training (80% of total data in each season) and testing (20% of total data in each season) subsets are utilized in this experiment. The computer configuration is 3.1 GHz CPU and 8G RAM, with MATLAB 8.6.0.267246 (R2015b). The final wind power forecasting error for each season is tabulated in Tab. 2.

Four layers are designed in RNNs, i.e., 1-2-1 for input-hidden-output layer. Input layer contains 9 neurons associated to 9 features determined by feature selection method [20], and the output layer contains only one neuron refer to wind power. The number of the neurons in the first hidden layer is estimated by $H_n = 2N_{in} - \alpha,\ \alpha \in [0, 6]$, where N_{in} is the number of nodes in the input layer, and the second hidden layer is set as 6 as the analysis of the clustering methods. The maximum epochs, learning rate and convergence goal are 100, 0.01 and 0.001, respectively. Tansig and purelin are chosen as the activation function for the hidden layer and the output layer in the neural network.

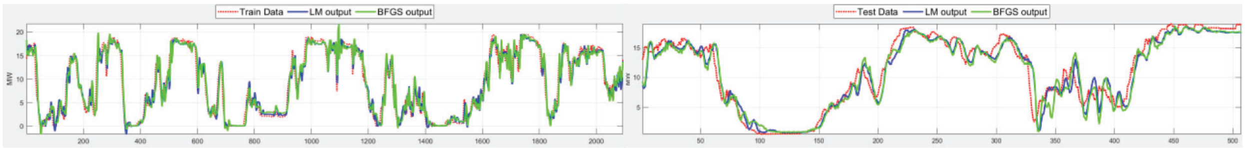
The average costing times of the proposed approach based on the optimal RNNs comparing with traditional RNNs are reduced by 36%, 35% and 35% refer to 6-steps, 12-steps and 18-steps ahead forecasting results in four seasons, respectively. For the average forecasting results RMSE with respect to

6-steps, 12-steps and 18-steps ahead between the optimal and traditional RNNs is reduced by 9%, 22% and 24%, respectively. Both the computing time and average forecasting results indicate that the proposed approach can significantly improve the generation ability of the RNNs. Because the precise and reliable analysis of hidden-layer information determines the reasonable number of hidden layer nodes, thus reducing the complexity of the connection relation of nonlinear weights between two layers of hidden neurons, implementing the purpose of optimizing network structure and improving the efficiency of predictive modeling, as well as enhancing the generalization ability of RNNs.

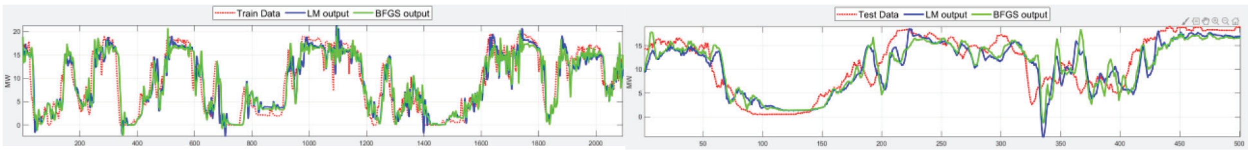
Table 2: Wind power forecasting error

Seasons		Spring			Summer	
FS	6	12	18	6	12	18
ET	2618.05	2628.08	2590.00	2562.37	2559.97	2568.85
PAET	1647.43	1650.92	1698.90	1659.93	1706.99	1655.99
RMSE1	0.1664	0.2330	0.2930	0.1155	0.2428	0.2936
RMSE1PA	0.1432	0.2509	0.3071	0.1151	0.2419	0.2428
RMSE2	0.1799	0.5499	0.4844	0.1126	0.3089	0.4464
RMSE2PA	0.1474	0.2606	0.3188	0.1330	0.2918	0.3446
RMSE3	0.1847	0.2477	0.3544	0.1292	0.2692	0.3233
RMSE3PA	0.1506	0.2623	0.3046	0.0805	0.1999	0.2506
RMSE4	0.1799	0.4832	0.4844	0.1126	0.3089	0.4464
RMSE4PA	0.1608	0.2623	0.3046	0.0805	0.1999	0.2506
Seasons		Autumn			Winter	
FS	6	12	18	6	12	18
ET	2577.15	2564.74	2588.12	2725.27	2668.14	2629.49
PAET	1692.22	1679.37	1662.40	1664.47	1719.20	1676.69
RMSE1	0.1703	0.3463	0.3924	0.1899	0.2815	0.3313
RMSE1PA	0.1701	0.3155	0.3817	0.1473	0.2047	0.2531
RMSE2	0.1558	0.3598	0.5350	0.1762	0.4302	0.5804
RMSE2PA	0.1497	0.3085	0.4911	0.1577	0.2320	0.2763
RMSE3	0.1982	0.4297	0.4267	0.2303	0.3063	0.3363
RMSE3PA	0.1621	0.3532	0.3674	0.2175	0.2840	0.2331
RMSE4	0.1558	0.3598	0.5351	0.1762	0.4302	0.5804
RMSE4PA	0.1434	0.3232	0.4474	0.1476	0.3806	0.4236

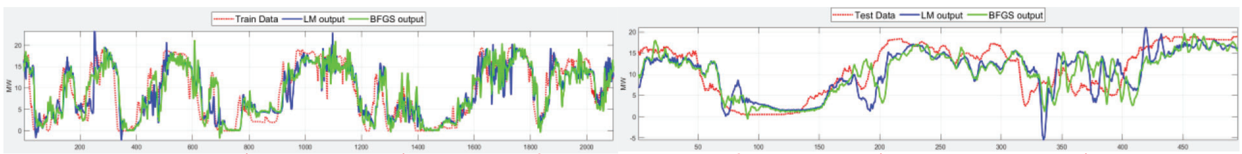
Note: FS: Forecasting Steps; ET: Average elapsed time in seconds by the approach; PAET: Average elapsed time in seconds by the proposed approach in this paper; RMSE1: The RMSE using the LM training method based on the training sample; RMSE1 and RMSE2: The forecasting RMSE using the RNNs with LM training method based on the training samples and testing samples, respectively; RMSE1PA and RMSE2PA: The forecasting RMSE using the optimal RNNs with LM training method based on the training samples and testing samples, respectively; RMSE3 and RMSE4: The forecasting RMSE using the RNNs with BFGS training method based on the training samples and testing samples, respectively; RMSE3PA and RMSE4PA: The forecasting RMSE using the optimal RNNs with BFGS training method based on the training samples and testing samples, respectively.



1 hours (6-steps ahead) wind power forecasting results of spring, 2004 (10 Minutes per point)



2 hours (12-steps ahead) wind power forecasting results of spring, 2004 (10 Minutes per point)



3 hours (18-steps ahead) wind power forecasting results of spring, 2004 (10 Minutes per point)

Figure 3: 1–3 hours (6, 12, 18-steps ahead) wind power forecasting results in Spring, 2004

4 Conclusion

The hidden-layer information of RNNs based on three unsupervised cluster algorithms are applied to optimize the RNNs architecture and improve the wins power's forecasting accuracy in this paper. The proposed approach firstly utilized the wavelet analysis and Lipschitz quotient to avoid the local transient that may propagates over time, and estimate the model order, respectively. Secondly, three unsupervised cluster algorithms such as Kmeans, Kmeans++ and Isodata are used to analyze the distribution probability of the hidden-layer information and optimize the architecture of RNNs. Since that the proper handling of hidden-layer information can directly reduce the risk of over-fitting that usually caused by too many neuron nodes, so the number of hidden layer neurons is streamlining and therefore the RNNs generalization ability is improved. Finally, the performance of the proposed approaches is evaluated based on the real data from NREL, and compared with the results derived by traditional RNNs. The experimental evaluation indicates that the analysis of the RNNs hidden-layer information can effectively stream the number of the hidden layer neurons, optimize the architecture and significantly improve the generalization ability of RNNs.

Acknowledgement: The authors acknowledge two anonymous reviewers providing valuable comments and helpful suggestions to improve the manuscript.

Funding Statement: This research is supported by National Natural Science Foundation of China (Nos. 61902158, 61806087).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M. et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542(7639),* 115–130. DOI 10.1038/nature21056.
2. Hinton, G. (2018). Deep learning—a technology with the potential to transform health care. *JAMA, 320(11),* 1101–1102. DOI 10.1001/jama.2018.11100.
3. Yaramasu, V., Dekka, A., Durán, M. J., Kouro, S., Wu, B. (2017). PMSG-based wind energy conversion systems: survey on power converters and controls. *IET Electric Power Applications, 11(6),* 956–968. DOI 10.1049/iet-epa.2016.0799.

4. Liu, H., Mi, X. W., Li, Y. F. (2018). Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. *Energy Conversion and Management, 156,* 498–514. DOI 10.1016/j.enconman.2017.11.053.
5. Shi, H., Xu, M., Li, R. (2018). Deep learning for household load forecasting—a novel pooling deep RNNS. *IEEE Transactions on Smart Grid, 9(5),* 5271–5280. DOI 10.1109/TSG.2017.2686012.
6. Kong, W., Dong, Z. Y., Hill, D. J., Luo, F., Xu, Y. (2018). Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions on Power Systems, 33(1),* 1087–1088. DOI 10.1109/TPWRS.2017.2688178.
7. Gonçalves, P. F., Cruz, S. M., Abadi, M. B., Caseiro, L. M., Mendes, A. M. (2017). Fault-tolerant predictive power control of a DFIG for wind energy applications. *IET Electric Power Applications, 11(6),* 969–980. DOI 10.1049/iet-epa.2016.0494.
8. Coelho, I. M., Coelho, V. N., Luz, E. J. D. S., Ochi, L. S., Guimarães, F. G. et al. (2017). A GPU deep learning metaheuristic based model for time series forecasting. *Applied Energy, 201,* 412–418. DOI 10.1016/j.apenergy.2017.01.003.
9. Young, T., Hazarika, D., Poria, S., Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine, 13(3),* 55–75. DOI 10.1109/MCI.2018.2840738.
10. Yang, Y., Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A. et al. (2017). SPIDER2: A package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. In: Zhou Y., Kloczkowski A., Faraggi E., Yang Y. (eds.), *Prediction of protein secondary structure. methods in molecular biology*, vol. 1484. Humana Press, New York, NY. DOI 10.1007/978-1-4939-6406-2_6.
11. Rajalakshmi, M., Rengaraj, R., Bharadwaj, M., Kumar, A., Naren, N. (2018). An ensemble based hand vein pattern authentication system. *Computer Modeling in Engineering & Sciences, 114(2),* 209–220.
12. Shao, H., Deng, X. (2018). AdaBoosting neural network for short-term wind speed forecasting based on seasonal characteristics analysis and lag space estimation. *Computer Modeling in Engineering & Sciences, 114(3),* 277–293.
13. Draye, J. P., Cheron, G., Libert, G., Godaux, E. (1997). Emergence of clusters in the hidden layer of a dynamic recurrent neural network. *Biological Cybernetics, 76(5),* 365–374. DOI 10.1007/s004220050350.
14. Sak, H. C. S. I., Senior, A., Beaufays, F. C. C. O. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. *Fifteenth Annual Conference of the International Speech Communication Association*. Singapore.
15. Bandara, K., Bergmeir, C., Smyl, S. (2017). Forecasting across time series databases using long short-term memory networks on groups of similar series. arXiv preprint arXiv: 1710. 03222.
16. Yulita, I. N., Fanany, M. I., Arymurthy, A. M. (2017). Fuzzy clustering and bidirectional long short-term memory for sleep stages classification. *IEEE International Conference on Soft Computing, Intelligent System and Information Technology*. Bali, Indonesia.
17. Baytas, I. M., Xiao, C., Zhang, X., Wang, F., Jain, A. K. et al. (2017). Patient subtyping via time-aware LSTM networks. *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Halifax, Nova Scotia, Canada.
18. Shwartz-Ziv, R., Tishby, N. (2017). Opening the black box of deep neural networks via information. arXiv preprint arXiv: 1703.00810.
19. National Renewable Energy Laboratory, Grid Modernization, Eastern and Western Data Sets, Wind Power Integration Data Set, https://www.nrel.gov/grid/wind-integration-data.html.
20. Deng, X., Shao, H., Hu, C., Jiang, D., Jiang, Y. (2020). Wind power forecasting methods based on deep learning: a survey. *Computer Modeling in Engineering & Sciences, 122(1),* 273–301. DOI 10.32604/cmes.2020.08768.