

Research Article

Reinforcement Learning Based Novel Adaptive Learning Framework for Smart Grid Prediction

Tian Li,^{1,2} Yongqian Li,^{1,2} and Baogang Li^{1,2}

¹Department of Computer Science, North China Electric Power University, Baoding, Hebei 071000, China

²Department of Electronic and Communication Engineering, North China Electric Power University, Baoding, Hebei 071000, China

Correspondence should be addressed to Tian Li; ncepulitian@126.com

Received 15 December 2016; Revised 25 February 2017; Accepted 23 March 2017; Published 18 April 2017

Academic Editor: Hung-Yuan Chung

Copyright © 2017 Tian Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Smart grid is a potential infrastructure to supply electricity demand for end users in a safe and reliable manner. With the rapid increase of the share of renewable energy and controllable loads in smart grid, the operation uncertainty of smart grid has increased briskly during recent years. The forecast is responsible for the safety and economic operation of the smart grid. However, most existing forecast methods cannot account for the smart grid due to the disabilities to adapt to the varying operational conditions. In this paper, reinforcement learning is firstly exploited to develop an online learning framework for the smart grid. With the capability of multitime scale resolution, wavelet neural network has been adopted in the online learning framework to yield reinforcement learning and wavelet neural network (RLWNN) based adaptive learning scheme. The simulations on two typical prediction problems in smart grid, including wind power prediction and load forecast, validate the effectiveness and the scalability of the proposed RLWNN based learning framework and algorithm.

1. Introduction

Accurate prediction is vital to the safety and economic operation of many engineering systems. For instances, smart grid, one of most complex systems in the world, requires accurate power source and power load forecast techniques to maintain the real-time power balance to supply electricity to customers in a safe and reliable manner. Especially with the high-level penetration of renewable energy at the source side and controllable loads, such as electric vehicle, at the demand side, the operation uncertainty of the smart grid increased rapidly due to the intermittent character of renewable resources and the flexibility of controllable loads. Almost all aspects of operation and planning in smart grid call for advanced prediction techniques which are capable of adapting to the various operation conditions of smart grid and the scalability to other engineering systems.

To achieve a more accurate prediction, many forecasting methods have been proposed for the smart grid. Among them, the neural network (NN) based prediction method has been acknowledged as one of the most effective methods.

Ding et al. [1] showed that neural network-based models outperform time series models, and the authors proposed the corresponding methodology to guarantee the generalization ability of NN model for load forecasting with available data. Tamimi and Egbert [2] introduced a short-term power load forecasting method by integrating the fuzzy logic expert system with artificial neural networks to realize more practical short-term power load forecast results. A robust forecast method based on the cascade neural network is described in [3] to accomplish an efficient short-term load forecasting. However, the methods in [1–3] and references therein lack the capability to adapt to various operation conditions of the smart grid due to the uncertainties introduced by renewables and controllable loads. Besides, they cannot be generalized to other practical systems such as transportation system.

An approach to realize the adaptability of the prediction technique is proposed in [4] by integrating wavelet transform, fuzzy system, and adaptive genetic algorithm with a generalized neural network. Using wavelet multiresolution decomposition by autocorrelation shell representation and multilayer perceptions neural networks modeling of wavelet

coefficients, an effective adaptive method is illustrated in [5]. Similarly, El Desouky and Elkateb [6] proposed a hybrid adaptive forecast method by incorporating artificial neural network and autoregressive integrated moving average model in a promising approach. However, the proposed adaptive prediction methods in [4–6] and references therein also cannot be generalized to other complex engineering systems which are necessary in modern society.

To provide the generalization capability, reinforcement learning (RL) [7], an adaptive method of machine learning and optimal control theory, can be incorporated into an adaptive prediction framework. Mnih et al. [8] proposed an RL based Deep-Q network that allowed a strong self-learning ability of the agent to play Atari 2600 games to win against a human player. This provided us with new insights into realizing adaptive control of large-scale engineering systems with strong structure and parameter uncertainties. Motivated by these works, we tried to apply the RL framework to provide a forecast method with generalization ability and self-learning capability. Definitely, most existing forecast techniques referenced above, including BP neural network [9] and genetic algorithm optimized BP method (GABP) [10], used a set of fixed weight parameters in the model that prevented their adaptation to practical engineering system with varying and complex operation environment. Undoubtedly, there are many potential factors influencing the prediction results, and it is not possible to allow for all these factors in a forecast model. Although the selection of key factors can provide relative accurate prediction results, these models cannot describe some hidden features in the data itself. In this regard, a neural network with multitime scale resolution ability is introduced to mine the hidden relationships existing in the collected data. Wavelet neural network (WNN) is a representative of such neural networks. We implemented the WNN into the RL based adaptive forecast framework to yield reinforcement learning and wavelet neural network (RLWNN) based adaptive forecast approach with multitime scale resolution ability.

The main contributions of this paper are as follows: (1) a reinforcement learning based adaptive learning framework has been proposed to enable the learning capability to the prediction method; (2) wavelet neural network has been implemented to the adaptive learning framework to realize a multitime scale resolution; (3) wind power prediction and power load forecast are utilized as benchmark problems to validate the effectiveness of the proposed RLWNN based adaptive learning framework.

The rest of this paper is organized as follows. In Section 2, the architecture of the RLWNN based adaptive learning framework is introduced along with fundamental concepts of reinforcement learning, wavelet neural network, and adaptive critic mechanism. Section 3 depicts the RLWNN based data forecasting method and its implementation methodology. Case studies were performed on wind power and power load prediction in Section 4 to demonstrate the effectiveness of the proposed method. Section 5 concludes the paper and illustrates future research directions.

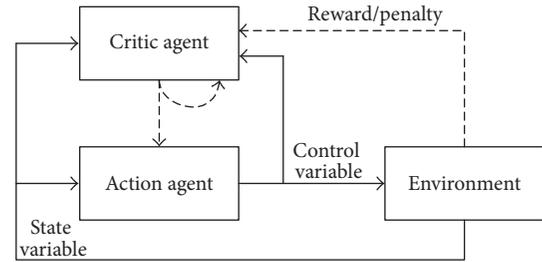


FIGURE 1: Adaptive critic-action structure based on reinforcement learning.

2. Reinforcement Learning Based Adaptive Learning Framework

2.1. Reinforcement Learning Based Adaptive Critic Structure. Typically, an adaptive forecast process is essentially a feedback control. The algorithm controls the tuning of key weights according to the operation condition and real-time prediction results. From the perspective of control, reinforcement learning, one of the main branches of machine learning, is closely related to both optimal control and adaptive control. RL is a method for solving optimization problems that involves an actor or agent that interacts with its environment and modifies its actions, or control policies, based on the stimuli received in response to its action [11]. To be specific, RL refers to a class of methods that enable the design of adaptive controllers that learn online, in real-time, and improve solutions to user-prescribed optimal control problems [12, 13]. One famous RL algorithm is the so-called actor-critic (AC) design method [13] as illustrated in Figure 1.

The learning mechanism performed by the AC structure includes two steps, policy evaluation and policy improvement, respectively. The policy evaluation step is performed by observing the results or consequences from the implementation of current actions. These results are evaluated with a performance index, also known as a value function, which quantifies how close the current action is to the optimal one. Based on the assessment of performance, the control policy can be modified or further improved to generate a new strategy that yields a value that is improved relative to the previous one. By implementing continual policy evaluation and policy improvement, the agent can improve its control policy to optimize a long-term performance index.

From the mathematical perspective, RL lays the foundation on the dynamic programming which cannot be applied to large-scale continuous action spaces or state spaces due to “curse of dimensionality.” The technique of value function approximation (VFA), such as using NN to represent the value function or policy in RL, is typically adopted to address this issue, and it generalizes the Bellman equation to large-scale, continuous action spaces, and state spaces of a nonlinear system. As shown in Figure 1, the basic structure of RL based adaptive critic-action is comprised of three fundamental concepts, including environment (also called the controlled object), critic agent, and action agent. The control strategy generated by the action agent is implemented on the

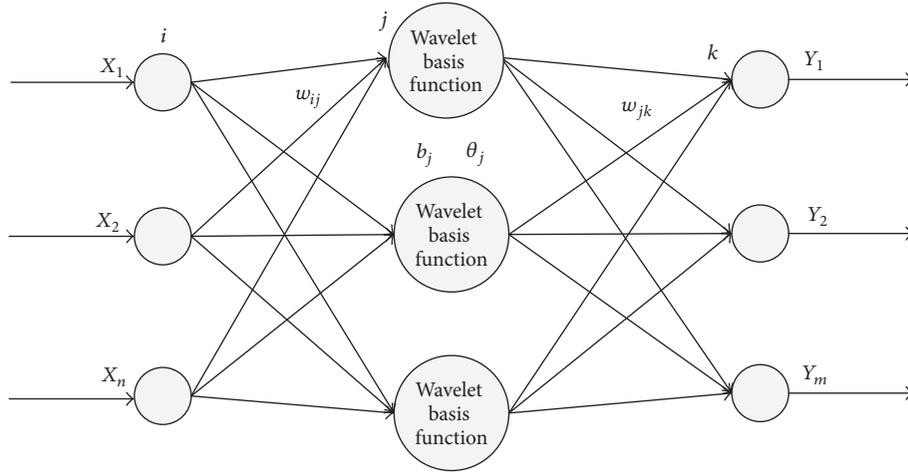


FIGURE 2: Structure of standard wavelet neural network.

controlled object. For instances, in power load forecast, this is the data of the power load itself. The critic agent is used to evaluate the control effects of the action agent. The reward or penalty generated for different time stages will influence the value function of the critic agent. Function approximation methods are used to provide a universal learning ability.

The data prediction process in engineering system can typically be modeled as a discrete-time nonlinear control system:

$$x(t+1) = f(x(t), u(t)), \quad (1)$$

where $x \in R^m$ denotes system states and $u \in R^n$ control variable. A cost function for system (1) can be defined as

$$J(x(i), i) = \sum_{k=i}^{\infty} \alpha^{k-i} r(x(k), u(k), k), \quad (2)$$

where $r(x_k, u_k)$ is the reward function and α ($0 < \alpha < 1$) denotes discount factor, which is utilized to reflect the impacts of current and future reward on cost function. J is the cost-to-go function with respect to $x(i)$ and the optimization goal is to select control sequence $u(k)$ to minimize cost function (2).

Based on Bellman optimality principle, the optimal cost function satisfies

$$\begin{aligned} J^*(x(t), t) \\ = \min_{u(t)} \{r(x(t), u(t), t) + \alpha J^*(x(t+1), t+1)\}. \end{aligned} \quad (3)$$

Corresponding optimal control $u^*(t)$ satisfies

$$\begin{aligned} u^*(t) \\ = \arg \min_{u(t)} \{r(x(t), u(t), t) + \alpha J^*(x(t+1), t+1)\}. \end{aligned} \quad (4)$$

Above all, from the control perspective, the essence of adaptive prediction is to obtain the optimal control $u^*(t)$ for minimizing the performance index J such as prediction error.

2.2. Wavelet Neural Network with Multitime Scale Resolution.

Wavelet neural network is a type of neural network that allows the multitime scale resolution. WNN is based on the wavelet transformation that provides superior time and the frequency domain location capabilities and the adaptive adjudication ability. These models are widely used in function approximation, differential equations, and other nonlinear system analyses. Wavelet-based function is often utilized to perform the kernel function of the neural network, as illustrated in Figure 2.

As for the wavelet neural network, the output of the hidden neural nets can be determined by

$$h(j) = h_j \left(\frac{\left(\sum_{i=1}^k \omega_{ij} x_i - b_j \right)}{a_j} \right), \quad j = 1, 2, \dots, l, \quad (5)$$

where x_i denotes the input data; $h(j)$ is the output of j th hidden net; w_{ij} is the weights between the input layer and hidden layer; b_j represents the expansion factor of wavelet base function h_j while a_j represents the scale factor of h_j , which can be calculated by

$$h_j = \cos \left(\frac{\gamma \left(\sum_{i=1}^k \omega_{ij} x_i - b_j \right)}{a_j} \right) e^{-\left(\sum_{i=1}^k \omega_{ij} x_i \right)^2 / 2}. \quad (6)$$

2.3. RLWNN Based Adaptive Learning Framework. In this section, we combined the multitime scale resolution capability of wavelet neural network with the adaptive critic structure of reinforcement learning to propose a RLWNN based adaptive learning framework, as illustrated in Figure 3.

As in Figure 3, the PID neural network and the wavelet neural network are, respectively, adopted as the action agent and critic agent in Figure 1 and termed as action network and critic network in the proposed RLWNN based adaptive learning framework. The critic network aims to estimate the defined cost function J which satisfies the Bellman optimality equation and to make the estimated cost function closely

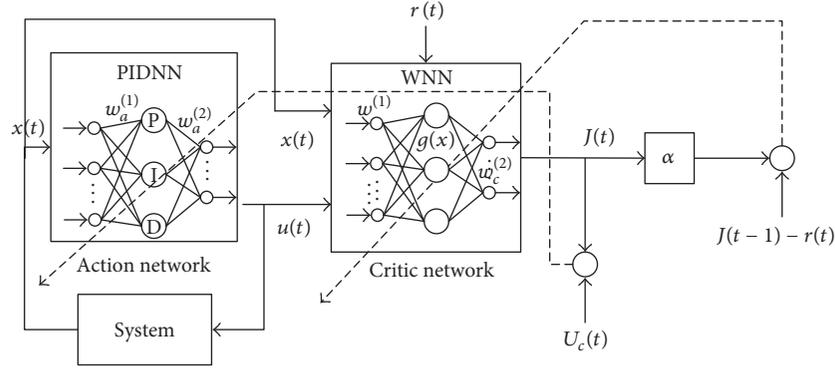


FIGURE 3: RLWNN based adaptive learning framework.

approximate the real one. This target can be realized by minimizing a loss function $E_c(t)$ by training the critic network and updating its weights; that is,

$$E_c(t) = \frac{1}{2} e_c^2(t), \quad (7)$$

$$e_c(t) = \alpha J(t) - [J(t-1) - r(t)].$$

The weights tuning method of critic network can be determined based on (stochastic) gradient descent algorithm as in

$$w_c(t+1) = w_c(t) - l_c(t) \frac{\partial E_c(t)}{\partial w_c(t)}, \quad (8)$$

where $w_c(t)$ denotes the weights at the time slot t and $l_c(t) > 0$ is the learning rate of critic network, which should decrease with the learning process and keep in a range at the end of the adaptive learning process.

Action network is used to estimate the optimal control policy u , and it can be the increment of the adjustable parameter of a traditional prediction model. The purpose of adjusting action network is to optimize the cost function, and this can be realized by minimizing the following loss function $E_a(t)$ to train action network: that is,

$$E_a(t) = \frac{1}{2} e_a^2(t) \quad (9)$$

$$e_a(t) = J(t) - U_c(t),$$

where $U_c(t)$ is the goal of cost function, which is usually set to zero.

Different from critic network, the action network in the proposed RLWNN based adaptive learning framework aims to minimize the output of critic network. The training process of action network can be realized by optimizing $J(t)$, and the updating algorithm can choose gradient descent algorithm; that is,

$$w_a(t+1) = w_a(t) - l_a(t) \frac{\partial E_a(t)}{\partial w_a(t)}, \quad (10)$$

where $w_a(t)$ is the weight of action network at time slot t , $l_a(t) > 0$ is the learning rate of action network, and $l_a(t) > 0$

should decrease during the learning process and also keep in a range at the end of the adaptive learning process.

Through the training of action network and critic network, the weights of the proposed reinforcement learning based adaptive learning framework can be determined online with the collected operational data or historical data $x(t)$. As a result, the RLWNN method provides an online performance optimization capability for the smart grid prediction field.

3. WNN Based Forecasting Method with RLWNN Adaptive Framework

For the convenience of implementation, the detailed procedure and algorithm of the RLWNN based adaptive learning framework are depicted in Figure 4. It is worth mentioning that an original prediction method is assumed to be utilized before introducing the proposed RLWNN adaptive framework. To be specific, the wavelet neural network is adopted as the original prediction method in Figure 4, and it can be easily substituted by any other existed prediction techniques. The procedure can be elaborated as follows.

Step 1 (enable prediction ability for the original data forecast model (WNN in this paper)). (1) Initialize randomly weights w_{ij}^0 between the input layer and the hidden layer of WNN, weights w_{jk}^0 between the hidden layer and the output layer, and the scale factor b_j^0 of wavelet basis function.

(2) Set network structure parameter of original WNN prediction model, including neuron number N of input layer, neuron number N_h of hidden layer, and neuron number M of output layer.

(3) Set original WNN prediction network fit error reference e and maximum iteration number N_d .

(4) Utilize historical data to realize supervised learning to yield a set of parameters $w_{ij}^k, w_{jk}^k, b_j^k, q_j^k$ with higher fitting ability.

(5) Predict data with obtained parameters $w_{ij}^k, w_{jk}^k, b_j^k, q_j^k$, input history data $X_1^0, X_2^0, \dots, X_n^0$, and yield forecast value $Y_1^0, Y_2^0, \dots, Y_m^0$ for the consequent m time slots.

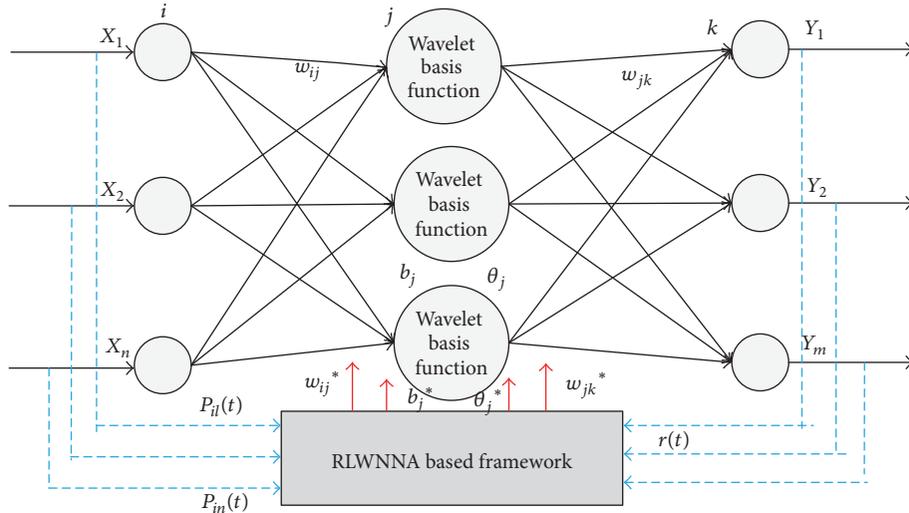


FIGURE 4: WNN based forecasting method with RLWNN adaptive framework.

Step 2 (enable RLWNN based adaptive framework with online optimization ability). (1) Between the adjacent prediction time intervals ij , use the actual data $P_{i1}, P_{i2}, \dots, P_{in}$ and the network state parameters $w_{ij}^k, w_{jk}^k, b_j^k, q_j^k$ as the input parameters of RLWNN based adaptive framework.

(2) Train the critic network to approximate performance index $J(t)$ and action network to approximate optimal control or optimal increment, Δw_{ij}^* , Δw_{jk}^* , Δb_j^* , Δq_j^* . Note that when training one of the networks, the weights of another network are assumed to be fixed.

(3) Substitute $w_{ij}^k, w_{jk}^k, b_j^k, q_j^k$ network with $w_{ij}^*, w_{jk}^*, b_j^*, q_j^*$ to carry out the prediction of the following m time interval, with input data $P_{i1}, P_{i2}, \dots, P_{in}$, and output data $Y_1^*, Y_2^*, \dots, Y_m^*$.

(4) $Y_1^*, Y_2^*, \dots, Y_m^*$ are the optimal prediction values.

(5) End.

One thing that needs to be mentioned is that the WNN procedure is ended with a given stop criterion such as maximum iteration number and acceptable tolerance.

4. Case Study

As indicated previously, the smart grid calls for advanced adaptive prediction methods to handle the uncertainties introduced by the wind power penetration at the source side and controllable loads at the demand side. In this respect, we implemented the proposed adaptive learning framework on two benchmark problems in smart grids, including wind power prediction and load forecast. Note that the wind power data is obtained from an actual wind power plant located in China, while the power load data is collected from the PJM electricity market.

4.1. Wind Power Prediction. Wind power can be quite variable and therefore is a perfect benchmark problem to evaluate

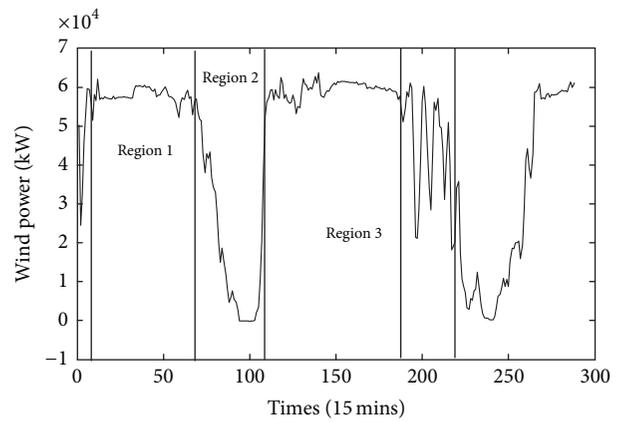


FIGURE 5: Typical profile of wind power output.

the effectiveness of the proposed prediction method. Motivated by this, we first implemented the proposed RLWNN adaptive framework on wind power output forecast.

A typical wind power output curve is illustrated in Figure 5. The wind power output typically has three representative regions with different features. It is better to utilize different forecasting methods for each region, which can be realized by implementing the proposed RLWNN adaptive framework and the corresponding prediction methods.

We used actual wind power data to verify the proposed method. For comparison, the traditional BP method, the GABP model, and the RLWNN based adaptive forecast method were separately implemented. A white noise mechanism was introduced to improve the generalization ability of the forecast model. Three cases are illustrated to evaluate the performance of these techniques.

Case 1 (wind power located in region 1). The prediction results for this scenario are presented in Figure 6. We can

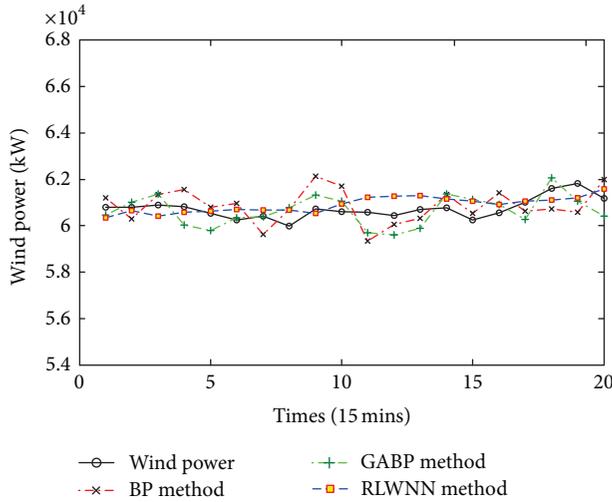


FIGURE 6: Prediction results for wind farm output in region 1.

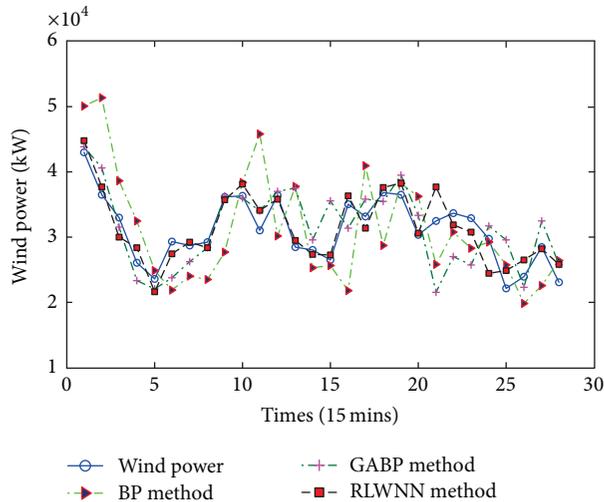


FIGURE 7: Prediction results for wind farm output in region 2 and region 3.

TABLE 1: Mean relative error.

	BP	GABP	RLWNN
Case 1	4.27%	2.09%	2.73%
Case 2	19.75%	12.58%	5.96%
Case 3	—	—	8.61%

conclude from Figure 6 that BP, GABP, and RLWNN have superior prediction ability when wind power output is located in region 1. The average relative errors of these three methods are shown in Table 1.

Case 2 (wind power located in region 2 and region 3). The prediction results under this case are illustrated in Figure 7, and the average relative error of three methods in this case is shown in Table 1.

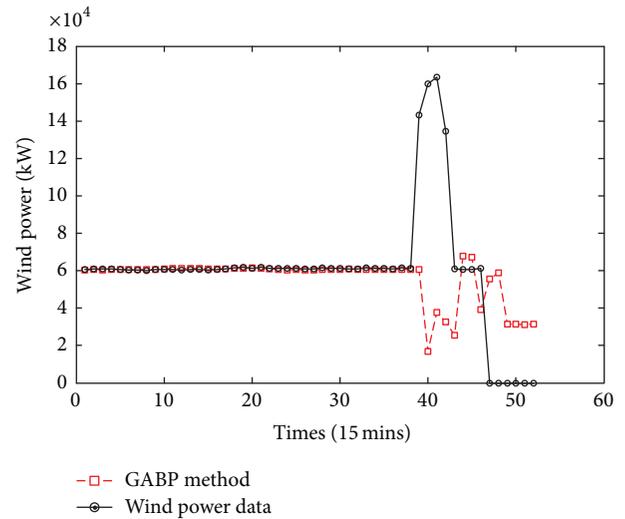


FIGURE 8: Prediction results of GABP when operating point changes.

We can learn from Figure 7 that both GABP and RLWNN adaptive methods have a relative higher prediction accuracy compared with BP method under the circumstance when wind power output is unsmooth, especially oscillation. It should be mentioned that, at the oscillation point and operation point far away from the setting point, the above three methods have relatively poor prediction accuracy.

Case 3 (wind power located in region 1, region 2, and region 3). Figure 8 shows the prediction results of GABP when the wind power output changes frequently, to illustrate the necessity of implementing an adaptive prediction method.

According to Figure 8, one can learn that starting at the 39th point, the operation condition of wind generator began to change, the relative prediction error of GABP was 2.5%, and when the operation point changed, the GABP method with fixed model parameters showed a decrease in prediction ability from the 39th to the 44th point. GABP recovers prediction ability from 44th point to 46th point when the wind power operation condition approaches the normal condition. However, GABP results again become less predictive when the operation condition changed again at the 47th point. Normal prediction capability was not regained when operation condition was restored again as indicated by the prediction results from the 48th to 52nd point. An adaptive method is needed to handle this issue.

Figure 9 shows the prediction results in this case for the proposed RLWNN adaptive framework. According to the results in Figure 9, similar to the BP and GABP prediction methods, RLWNN adaptive method can be used to produce a relatively precise prediction result when the wind power output is smooth. Under the condition of fluctuating wind power, the prediction of RLWNN at the first varying point is relative poor while the consequent prediction points locate in the acceptable range. The region from the 1st to the 16th point is similar to region 2, and the interval from the 17th to 22nd point is similar to region 3 as shown in Figure 5.

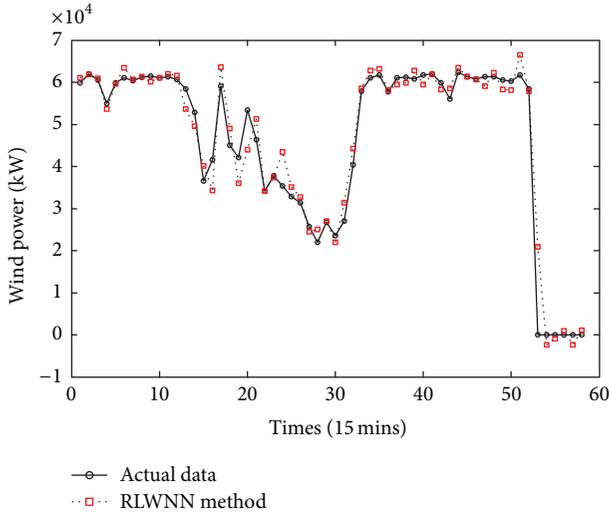


FIGURE 9: Prediction results of RLWNN for operating point changes situation.

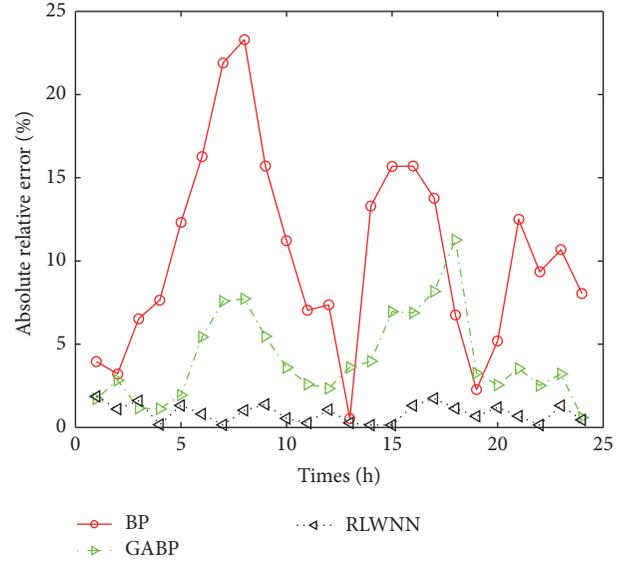


FIGURE 11: Prediction error for power load.

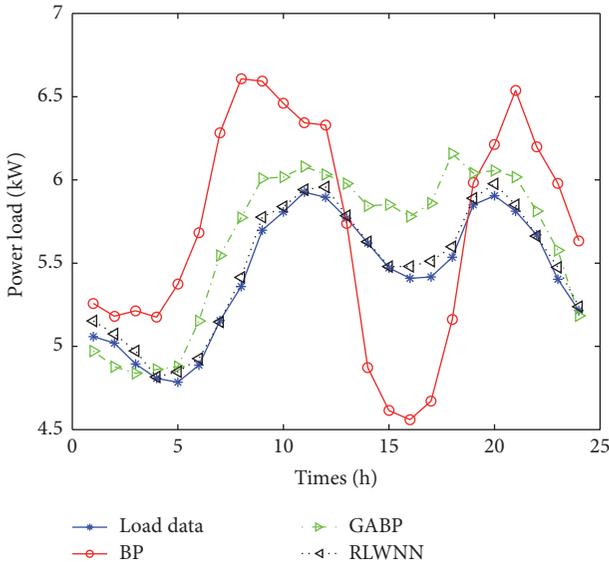


FIGURE 10: Forecast results for power load.

4.2. Power Load Prediction. In this section, the electrical power load data are obtained to test the effectiveness of the proposed RLWNN adaptive method. Using BP neural network, GABP network, and RLWNN method to predict the power load, the results and the error are illustrated in Figures 10 and 11, respectively.

We can learn from Figures 10 and 11 that, through a step-by-step improvement, the maximum prediction error can be reduced to 2% of the RLWNN adaptive method compared with 23.6% of BP method and 11.1% of GABP method.

Through the above analysis, the conclusion can be drawn that the BP model is the simplest and easiest technique to be implemented but has a relatively lower prediction accuracy. The accuracy of GABP model is improved compared to the BP model, but the time complexity of GABP is relatively high.

The RLWNN model has the highest prediction accuracy due to its online adaptive learning capability.

5. Conclusion

In this paper, a RLWNN based adaptive learning framework is proposed by combining the self-learning capability of reinforcement learning with the multitime scale resolution capability of the wavelet neural network. The proposed RLWNN adaptive framework is utilized to construct a universal adaptive prediction structure. Wind power prediction and power load prediction results validated the effectiveness of the proposed framework.

It is worth noting that the proposed method is essentially a data-driven method; incorporating some prior knowledge including some key factors, which influence the wind power and power load such as real-time electricity price in the demand-response of the electricity market, will further improve the efficiency of the proposed RLWNN adaptive method. Besides, the RLWNN based adaptive learning framework is not confined in smart grids and can be easily implemented in other engineering systems.

Conflicts of Interest

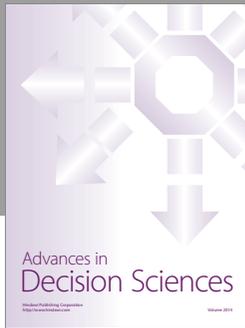
The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported in part by the Fundamental Research Funds for the Central Universities in 2017, the Fundamental Research Funds for the Central Universities (no. 2015MS130), the National Natural Science Foundation of China (nos. 61501185 and 61377088), and Beijing Natural Science Foundation (no. 4164101).

References

- [1] N. Ding, C. Benoit, G. Foggia, Y. Besanger, and F. Wurtz, "Neural network-based model design for short-term load forecast in distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 72–81, 2016.
- [2] M. Tamimi and R. Egbert, "Short term electric load forecasting via fuzzy neural collaboration," *Electric Power Systems Research*, vol. 56, no. 3, pp. 243–248, 2000.
- [3] S. Kouhi and F. Keynia, "A new cascade NN based method to short-term load forecast in deregulated electricity market," *Energy Conversion and Management*, vol. 71, pp. 76–83, 2013.
- [4] D. K. Chaturvedi, A. P. Sinha, and O. P. Malik, "Short term load forecast using fuzzy logic and wavelet transform integrated generalized neural network," *International Journal of Electrical Power and Energy Systems*, vol. 67, pp. 230–237, 2015.
- [5] B.-L. Zhang and Z.-Y. Dong, "An adaptive neural-wavelet model for short term load forecasting," *Electric Power Systems Research*, vol. 59, no. 2, pp. 121–129, 2001.
- [6] A. El Desouky and M. Elkateb, "Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA," *IEE Proceeding—Generation, Transmission and Distribution*, vol. 147, no. 4, pp. 213–217, 2000.
- [7] M. L. Littman, "Reinforcement learning improves behaviour from evaluative feedback," *Nature*, vol. 521, no. 7553, pp. 445–451, 2015.
- [8] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [9] Z. Xiao, S.-J. Ye, B. Zhong, and C.-X. Sun, "BP neural network with rough set for short term load forecasting," *Expert Systems with Applications*, vol. 36, no. 1, pp. 273–279, 2009.
- [10] S. Ding, C. Su, and J. Yu, "An optimizing BP neural network algorithm based on genetic algorithm," *Artificial Intelligence Review*, vol. 36, no. 2, pp. 153–162, 2011.
- [11] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems Magazine*, vol. 32, no. 6, pp. 76–105, 2012.
- [12] F.-Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: an introduction," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 39–47, 2009.
- [13] P. J. Werbos, "A menu of designs for reinforcement learning over time," in *Neural Networks for Control*, pp. 67–95, MIT Press, 1990.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

