An Adaptive Network-Based Reinforcement Learning Method for MPPT Control of PMSG Wind Energy Conversion Systems

Chun Wei, *Student Member, IEEE*, Zhe Zhang, *Member, IEEE*, Wei Qiao, *Senior Member, IEEE*, and Liyan Qu, *Member, IEEE*

***Abstract*—This paper proposes an artificial neural network (ANN)-based reinforcement learning (RL) maximum power point tracking (MPPT) algorithm for permanent-magnet synchronous generator (PMSG)-based variable-speed wind energy conversion systems (WECSs). The proposed MPPT algorithm first learns the optimal relationship between the rotor speed and electrical power of the PMSG through a combination of the ANNs and the Q-learning method. The MPPT algorithm is switched from the online RL to the optimal relation-based online MPPT when the maximum power point is learned. The proposed online learning algorithm enables the WECS to behave like an intelligent agent with memory to learn from its own experience, thus improving the learning efficiency. The online RL process can be reactivated any time when the actual optimal relationship deviates from the learned one due to the aging of the system or a change in the environment. Simulation and experimental results are provided to validate the proposed ANN-based RL MPPT control algorithm for a 5-MW PMSG-based WECS and a small emulated PMSG-based WECS, respectively.**

***Index Terms*—Artificial neural network (ANN), maximum power point tracking (MPPT), permanent-magnet synchronous generator (PMSG), Q-learning, reinforcement learning (RL), wind energy conversion system (WECS).**

I. Introduction

THE permanent magnet synchronous generator (PMSG)-based variable-speed direct-drive (no gearbox) wind energy conversion systems (WECSs) are becoming more popular for large wind turbines with higher power ratings [1]. Compared with other types of WECSs, the PMSG-based direct-drive WECSs have been found to be more superior owning to their advantages of higher efficiency, higher power density, lower maintenance costs, and better fault-ride through and grid support capability [2]. These advantages make them more promising for large-scale and offshore applications. Much research effort recently has gone into the development of advanced converter topologies, control algorithms, and grid-interface configurations to make the PMSG-based WECSs more suitable for real-world applications [3]–[6].

An effective maximum power point tracking (MPPT) algorithm is essential for modern WECSs to improve the energy capture efficiency from wind. Turbine power profile (TPP), optimal tip speed ratio (TSR), and optimal relationship-based (ORB) controls are three main MPPT methods widely used in large WECSs [7]–[13]. In these methods, the wind turbine characteristics are provided by the manufacturers to obtain the optimal power versus wind speed curve, optimal TSR, and optimal torque or power versus generator speed curve. The accuracy of these wind turbine characteristic curves, however, is largely influenced by the operating environments of the WECSs. For example, the output power may deviate from the normal power curve when there is dirt, bugs, or ice on the blades [14], [15]. In addition, the predetermined optimal curves for the MPPT control will inevitably lose their accuracy caused by the system aging effect, such as mechanical wear and tear, erosion of the turbine blades, and related factors [16], [17], which makes the wind power generation less efficient over time. Therefore, an intelligent online learning algorithm is desired for adapting the MPPT controls for the WECSs, especially those installed in the harsh and less-accessible offshore environment.

The aforementioned three MPPT methods have no online learning capability. Among them, the TPP and TSR are wind speed measurement-based methods, which require costly anemometers to acquire the wind speed information, making them less attractive than the wind speed sensorless ORB method. A perturbation and observation (P&O, also known as hill-climb search) MPPT algorithm is widely used in small WECSs since it does not need field tests and is independent of wind turbine characteristics. However, P&O is not suitable for MPPT control of large WECSs because of the slow response of the WECSs caused by the large system inertia. Some researches combined the P&O and ORB methods, in which the controllers of the WECSs first search the optimal relations between system parameters using a P&O algorithm and then switch to the ORB control once the optimal relations have been found to achieve fast and efficient MPPT control. For example, in [18], the optimal relationship between output power and dc voltage of the power converter was obtained using the P&O method; in [19], a P&O method was used to obtain the optimal relationship of the rectified dc voltage and dc-side current of the diode bridge rectifier in a PMSG-based WECS. In these methods, the P&O MPPT algorithm is executed online when the wind speed is

Manuscript received September 28, 2015; revised December 14, 2015; accepted December 24, 2015. Date of publication January 04, 2016; date of current version June 24, 2016. This work was supported in part by the U.S. National Science Foundation under CAREER Award ECCS-0954938. Recommended for publication by Associate Editor B. Singh.

C. Wei, W. Qiao, and L. Qu are with the Power and Energy Systems Laboratory, Department of Electrical and Computer Engineering, University of Nebraska−Lincoln, Lincoln, NE 68588-0511 USA (e-mail: cwei@huskers.unl.edu; wqiao3@unl.edu; lqu2@unl.edu).

Z. Zhang is with Nexteer Automotive, Saginaw, MI 48601-9494 USA (e-mail: zhe.zhang@nexteer.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPEL.2016.2514370

0885-8993 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

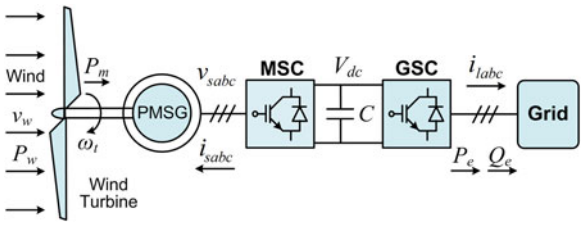


Fig. 1. Configuration of a direct-drive PMSG WECS connected to a power grid.

stable. However, the duration of the stable wind speed varies in the real application and the wind speed may change before the optimal relation is learned. In the P&O MPPT control, the controller of the WECS does not learn from the experience, i.e., it does not use the knowledge of the wind speed conditions it has experienced previously. This decreases the efficiency of the MPPT control.

This paper proposes a new intelligent MPPT algorithm for PMSG-based WECSs based on the reinforcement learning (RL) [20]. The proposed MPPT algorithm first learns online the optimal relationship of the electrical power and rotor speed of the PMSG using the Q-learning method. Once the optimal relationship is learned, the proposed method is switched from the online RL to the ORB algorithm for online MPPT control. When the system ages, the online learning algorithm can be executed again to learn the new optimal relationship. In the proposed MPPT algorithm, artificial neural networks (ANNs) are combined with the Q-learning method to eliminate the need of a large lookup table in the traditional RL used in [21] whose size is usually selected through offline simulation studies. This innovation not only saves the computational cost, but also makes the whole MPPT algorithm easier to design and implement. In addition, ANNs generalize the learning experience and even provide predictions for the knowledge never experienced before; thus, shortening the online learning time and improving the learning efficiency. The proposed RL algorithm enables the WECS to behave like an intelligent agent with memory that learns to act from its own experience without the need of the knowledge of wind turbine parameters or wind speed information. The proposed MPPT algorithm is validated by simulation results for a 5-MW WECS and experimental results for a 200-W emulated WECS equipped with a PMSG.

II. Direct-Drive PMSG-Based WECS

The configuration of a direct-drive PMSG-based WECS is shown in Fig. 1, where the wind turbine is connected to the PMSG directly. The electrical power generated by the PMSG is transmitted to a power grid via a variable-frequency converter, which consists of a machine-side converter (MSC) and a grid-side converter (GSC). The MSC controls the PMSG to extract the maximum power from the wind and/or comply with a command from the wind power plant control center. The GSC maintains a constant dc-link voltage and controls the reactive power that the WECS exchanges with the grid [22].

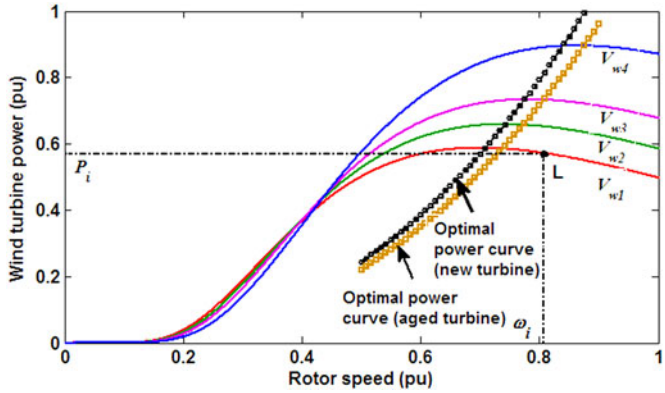


Fig. 2. Typical wind turbine power-rotor speed characteristic curves for different wind speeds.

A. Wind Turbine Aerodynamic Model

The mechanical power that a wind turbine extracts from the wind is given by [7]

$$P_m = \frac{1}{2}\rho A_r v_w^3 C_p(\lambda, \beta) \tag{1}$$

where ρ is the air density; A_r is the area swept by the blades; v_w is the wind speed; and C_p is the power coefficient, which is determined by the TSR $\lambda = \omega_t R/v_w$ and the blade pitch angle β which is always constant during MPPT control, where ω_t is the turbine rotating speed. There is an optimal value $\lambda_{\rm opt}$ at which the turbine extracts the maximum power from wind given by

$$P_{\max} = \frac{1}{2}\rho A_r \frac{R^3 C_{p\max}}{\lambda_{\rm opt}^3}\omega_t^3 = K\omega_t^3 \tag{2}$$

where $C_{p\max}$ is the maximum power coefficient at $\lambda_{\rm opt}$ and K is the optimal parameter. Fig. 2 shows the typical wind turbine power-rotor speed characteristic curves for different wind speeds, where the optimal power curve (new turbine) describes the maximum power points (MPPs) at different wind speed conditions when the wind turbine is new; while the optimal power curve (aged turbine) represents the MPPs of an aged wind turbine after about five-year operation and declines 9% from that of the new turbine [16].

B. Modeling of the PMSG

The dynamic equations of a three-phase PMSG without saliency can be written in a synchronously rotating dq reference frame as

$$v_{\rm sd} = -R_s i_{\rm sd} + \frac{d\psi_d}{dt} - \omega_e \psi_q \tag{3}$$

$$v_{\rm sq} = -R_s i_{\rm sq} + \frac{d\psi_q}{dt} + \omega_e \psi_d \tag{4}$$

where $v_{\rm sd}$ and $v_{\rm sq}$ are the d-axis and q-axis stator terminal voltages, respectively; $i_{\rm sd}$ and $i_{\rm sq}$ are the d-axis and q-axis stator currents, respectively; R_s is the resistance of the stator windings; ω_e is the electrical angular velocity of the rotor; and ψ_d and ψ_q are the d-axis and q-axis flux linkages of the PMSG, respectively,

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

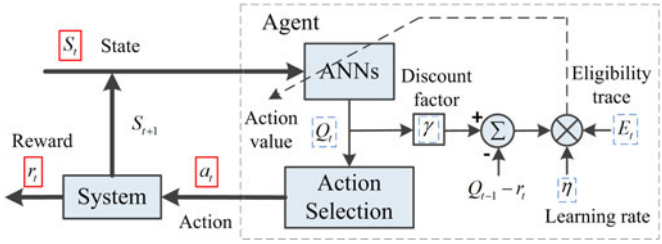


Fig. 3. Schematic diagram of the ANN-based RL algorithm.

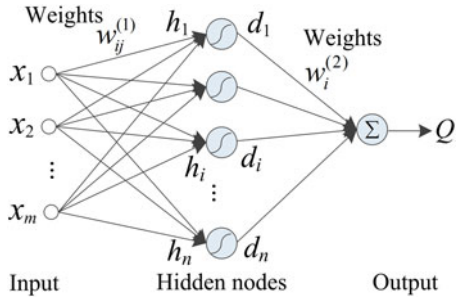


Fig. 4. Schematic diagram of the feedforward ANN for the action value estimation.

given by

$$\psi_d = -L_d i_{\mathrm{sd}} + \psi_m \tag{5}$$

$$\psi_q = -L_q i_{\mathrm{sq}} \tag{6}$$

where L_d and L_q are the *d*-axis and *q*-axis inductances of the PMSG, respectively; and ψ_m is the flux linkage generated by the permanent magnets.

III. ANN-Based RL

In the RL, an agent is able to learn from its own experience by directly interacting with the environment through state *s*, action *a*, and reward *r* [20]. The goal of the agent is to find an optimal action selection policy to maximize the total reward it receives over the future. This total reward is defined by a state value function or an action value function, which estimates how good it is for the agent to be in a given state or to perform a given action in a given state, respectively [23]. Therefore, the estimation of the value function is the core of the RL method. Q-learning is a form of widely used model-free RL method. However, in the traditional Q-learning method, a lookup table is required to store the values for each state-action pair, which becomes impractical for problems with a large state space. Moreover, the determination of the space in the lookup table to store the values of each discrete state usually needs simulation studies, which complicates the design process of the learning algorithm. In many RL problems, a large number of states encountered will never have been experienced before. To save the computational cost, simplify the learning algorithm, and improve the learning efficiency, it is expected that the RL can generalize new inexperienced state values from the previously learned state values [20]. Therefore, much research work has gone into the area of RL with generalization methods [24]. ANNs provide a powerful technique to generalize the value functions of the states and predict the value functions of the states which have never been experienced before. This paper combines ANNs with the Q-learning to improve the generalization performance of the online RL algorithm. Fig. 3 shows the schematic diagram of the proposed ANN-based RL method. For the application of the method in the MPPT control of a WECS, the system in Fig. 3 represents the WECS and the agent represents the controller of the WECS. The measured electrical power and rotor speed form the state, which is used as the input of the ANNs to generate the estimated action values for the selection of the proper action (i.e., the rotor speed control reference). The WECS then enters a new state, and a reward is given to tell whether the previous rotor speed control reference is good or not. After that the weights of the ANNs are updated according to the reward and eligibility traces to make a better estimation for the action values. The definition of state, action, and reward will be provided in detail in Section IV.

A. Q-Learning

The goal of the Q-learning is to learn the value of each action *a* taken from the action apace *A* in each state *s*, which is defined to be the predicted total discounted reward received by the agent over the future as a result of taking that action from the action apace *A*. The one-step Q-learning is defined by [25]

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \eta[r_{t+1} + \gamma \max_{a \in A} Q_{t+1}(s_{t+1}, a) - Q_t(s_t, a_t)]. \tag{7}$$

At each time step *t*, the agent first observes its current state s_t and selects an action a_t to perform. At the same time, the action value $Q_t(s_t, a_t)$ is recorded. Then, the subsequent state s_{t+1} is observed with an immediate reward r_{t+1} and the maximum action value corresponding to s_{t+1}, $\max_{a \varepsilon A} Q_{t+1}(s_{t+1}, a)$, is picked out from a lookup table that stores the action values for each state. The recorded $Q_t(s_t, a_t)$ is then updated to be $Q_{t+1}(s_t, a_t)$ in the next time step according to (7). The parameter $\eta \in (0, 1]$ is the learning rate, which determines how far the previously estimated $Q_t(s_t, a_t)$ is adjusted toward the newly estimated action value $r_{t+1} + \gamma \max_{a \varepsilon A} Q_{t+1}(s_{t+1}, a)$. A large learning rate is selected to enable the agent to learn fast in this paper. $\gamma \in [0, 1)$ is the discount factor that determines the current value of the rewards received by the agent in the future. If the discount factor is small, the agent only cares about the reward received immediately instead of the rewards received in the future. Therefore, a large discount factor is chosen in this paper to make the agent more "farsighted."

B. ANN-Based Q-Learning

In this paper, a multilayer feedforward ANN [26] is used to represent the action value for each action. The schematic diagram of the ANN is shown in Fig. 4. In each time step *t*, the action value can be expressed as

$$Q_t = \sum_{i=1}^{n} w_{t,i}^{(2)} d_{t,i} \tag{8}$$

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

where

$$d_{t,i} = \frac{1}{1+e^{-h_{t,i}}} \tag{9}$$

$$h_{t,i} = \sum_{j=1}^{m} w_{t,\mathrm{ij}}^{(1)} x_{t,j} \tag{10}$$

where *n* and *m* are the total numbers of hidden nodes and inputs of the ANN, respectively; $w_{t,\mathrm{ij}}^{(1)}$ and $w_{t,i}^{(2)}$ $(i = 1, \ldots, n; j = 1, \ldots, m)$ are the connecting weights between the input and hidden layers and between the hidden and output layers of the ANN, respectively; $h_{t,i}$ and $d_{t,i}$ are the input and output of the *i*th hidden node of the ANN, respectively; $x_{t,j} \in S_t$ $(j = 1, \ldots, m)$ is the *j*th input of the ANN, which is also the *j*th element of the state S_t shown in Fig. 3.

Based on the one-step Q-learning in (7), the weight matrix W_t of the ANN can be updated according to

$$\Delta W_t = \eta[r_{t+1} + \gamma \max_{a \in A} Q_{t+1}(s_{t+1}, a) - Q_t] \nabla_w Q_t \tag{11}$$

where W_t represents either connecting weight matrix $W_t^{(1)}$ or $W_t^{(2)}$; $Q_t$is the simplified form of $Q_t(s_t, a_t)$; $\nabla_w Q_t = \partial Q_t / \partial W_t$ is a vector of the gradient of Q_t with respect to W_t calculated as follows by using the back-propagation method [26]:

$$\frac{\partial Q_t}{\partial w_{t,i}^{(2)}} = d_{t,i} \tag{12}$$

$$\frac{\partial Q_t}{\partial w_{t,\mathrm{ij}}^{(1)}} = w_{t,i}^{(2)} d_{t,i}(1 - d_{t,i}) x_{t,j}. \tag{13}$$

To make the training process faster, the Q-learning is combined with the eligibility trace theory, which is one of the basic mechanisms of RL [20], and the updating rule of the weight matrix becomes [24]

$$\Delta W_t = \eta[r_{t+1} + \gamma \max_{a \in A} Q_{t+1}(s_{t+1}, a) - Q_t] \sum_{k=0}^{t} (\lambda\gamma)^{t-k} \nabla_w Q_k \tag{14}$$

whereλ is the trace decay parameter. A larger λ means that more credit in the current reward will be given to the previous states and actions. Therefore, a large λ is used to make the algorithm converge fast in this paper. Rather than using the greedy term $\max_{a \varepsilon A} Q_{t+1}(s_{t+1}, a)$, the $Q_{t+1}(s_{t+1}, a_t)$ associated with the selected action a_t is used for updating the weights and the weight update rule becomes

$$\Delta W_t = \eta[r_{t+1} + \gamma Q_{t+1}(s_{t+1}, a_t) - Q_t] \sum_{k=0}^{t} (\lambda\gamma)^{t-k} \nabla_w Q_k. \tag{15}$$

During the online learning process, each weight $w_{t,\mathrm{ij}}^{(1)}$ and $w_{t,i}^{(2)}$ $(i = 1, \ldots, n; j = 1, \ldots, m)$ in the weight matrixes $W_t^{(1)}$and$W_t^{(2)}$ of the ANN holds an eligibility trace $e_{t,\mathrm{ij}}^{(1)}$ and $e_{t,i}^{(2)}$, respectively, and the weight matrices are updated according to

$$W_{t+1} = W_t + \eta[r_{t+1} + \gamma Q_{t+1}(s_{t+1}, a_t) - Q_t] E_t \tag{16}$$

where E_t represents the eligibility trace matrix $E_t^{(1)}$ or $E_t^{(2)}$.

For the ANN that outputs the action value for the selected action a_t, the eligibility trace is updated by

$$E_t = \nabla_w Q_t + \lambda\gamma E_{t-1} \tag{17}$$

and for all other ANNs that output the action values for other unselected actions, their eligibility traces are updated by

$$E_t = \lambda\gamma E_{t-1}. \tag{18}$$

IV. RL Application for MPPT Control of WECSs

This section defines the state space *S*, action space *A*, reward *r*, and the action selection policy for applying the ANN-based RL algorithm to the MPPT control of PMSG-based WECSs. The controller of the WECS observes the state of the system, i.e., the operating point of the WECS, and chooses a control action from the action space. The WECS then enters a new operating point and receives a reward to update the action values. During this process, the wind speed measurement is not required. To find the MPP and extract more energy from the wind, an action with a higher value will be chosen with a greater possibility.

A. State Space

Fig. 2 shows that for each wind speed, there is only one optimum rotor speed at which the maximum wind power is extracted. In addition, there is only one rotor speed-turbine power curve for a specific wind speed. As a result, a rotor speed–power pair (e.g., the point L in Fig. 2) is sufficient to represent a state, i.e., an operating point of the WECS based on which an action could be chosen. For example, if the WECS detects that the operating point locates at L (ω_i, P_j) at wind speed v_{w1} by measuring its electrical output power and rotor speed, it will choose an action to move its operating point to the left to get closer to the MPP and gain more energy instead of moving to the right or remaining at the point L. In real applications, the optimal relationship of the generator rotor speed ω_r and electrical output power P_e is used, from which the state space is generated as follows:

$$S = \{(\omega_r, P_e)\}. \tag{19}$$

Therefore, the measured rotor speed ω_r and electrical power P_e are the two inputs to the ANNs [i.e., $m = 2$ in (10) in this paper] to obtain the action values.

B. Action Space

The action space for a PMSG-based WECS that operates in the speed-control mode can be expressed by

$$A = \{a | + \Delta\omega_r, 0, -\Delta\omega_r\} \tag{20}$$

where $\Delta\omega_r$ is the change of the rotor speed control command; and 0 means that no modification is made and the previous rotor speed control command is used. Three ANNs are required to estimate the values of these three actions. The rotor speed control command can, thus, be defined as

$$\omega_{r,t+1}^{*} = \omega_{r,t}^{*} + a \tag{21}$$

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

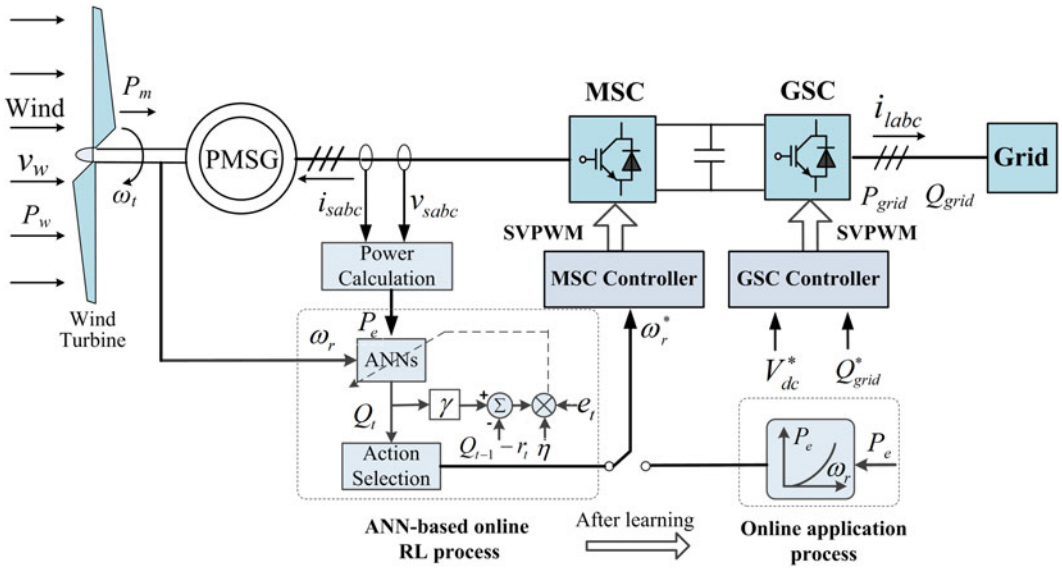


Fig. 5. Overall control scheme of the PMSG-based WECS containing the proposed MPPT algorithm.

where $\omega^*_{r,t}$ and $\omega^*_{r,t+1}$ are the previous and new rotor speed control commands, respectively.

C. Reward

After taking an action, the WECS receives a reward to evaluate the selected action. To find the MPP faster, the WECS receives a positive reward when it reaches the MPP and a negative reward if it locates in other operating points. Fig. 2 shows that a power curve around the MPP is relatively flat. When the operating point reaches that flat region, it is hard to detect the change of the output power. In other words, if a rotor speed "increase" or "decrease," action does not lead to the variation of the output power under stable wind speed condition, the WECS reaches the MPP. Therefore, the reward function is defined as

$$r_{t+1} = \begin{cases} +1, & \text{if } |P_{e,t+1} - P_{e,t}| \le \delta_1 \text{ and } a_{t-1} = \pm\Delta\omega_r \\ -1, & \text{otherwise} \end{cases} \tag{22}$$

where $P_{e,t+1}$and $P_{e,t}$ are two values of the electrical output power of the WECS in two successive time steps; δ_1 is a small positive number; and a_{t-1} is the action taken in the last step.

D. Action Selection Policy

The action selection policy is based on the Boltzmann exploration [20]. It chooses an action a_i in a state s with a probability $p(s, a_i)$ according the action values as follows:

$$p(s, a_i) = \frac{e^{Q(s,a_i)/\tau}}{\sum_{a_i} e^{Q(s,a_i)/\tau}} \tag{23}$$

where τ is a positive parameter called temperature, which controls the randomness of the exploration. A higher temperature causes the action selection to be more random, whereas a low temperature causes a higher value action to be selected with a greater chance. In this paper, a large value of τ is first used to encourage the agent to try all the actions regardless of their action values, and then the value τ is decreased to a small value to enable the agent to use the experience it has learned.

V. Proposed MPPT Algorithm for PMSG-Based WECSs

The proposed intelligent MPPT algorithm consists of two processes: an online learning process and an online application process. The overall control scheme of a PMSG-based WECS containing the proposed MPPT algorithm is shown in Fig. 5. In the online learning process, the ANN-based RL algorithm is used to search for the MPPs to learn the optimal relationship between ω_r and P_e. In the application process, the WECS is controlled by the ORB method using the optimal relationship learned. When the optimal power curve changes as a result of system aging or an environmental change, the online learning algorithm can be executed again to obtain a new optimal power curve.

A. Wind Speed Evaluation

Each time before the rotor speed command is changed in the online learning process, the wind condition is evaluated. The learning algorithm will not be executed unless the wind speed is stable. The wind speed evaluation criterion is given by [19]

$$\delta_2 = \frac{|P_{e,t+1} - P_{e,t}|}{P_{e,t}} \tag{24}$$

where δ_2 is the relative variation of two successive samples of the output electrical power. If δ_2 is smaller than a predefined threshold (e.g., 5%–10%) for a few seconds, which depends on the time constant of the WECS (e.g., 5–10 s for MW-scale WECS and can be shorter for smaller WECSs), the wind speed is assumed to be stable and the online learning process will be executed.

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

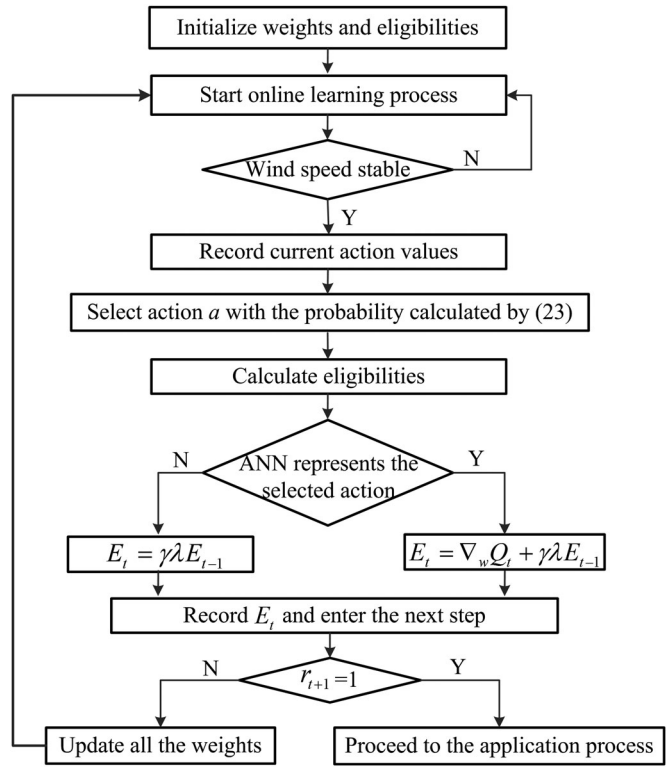


Fig. 6. Flowchart of the proposed ANN-based RL MPPT control.

B. ANN-Based RL Online Learning Process

After initialization, the MPPT algorithm starts to learn when the wind speed is stable. In each cycle of learning, $\omega_{r,t}$ and $P_{e,t}$ are measured and used as the two inputs to the three ANNs to calculate the three action values in the current sampling period individually. Then, an action a_t is selected from the action space defined by (20) using the Boltzmann exploration policy. The rotor speed control command is then set according to (21). The eligibility of each ANN is calculated according to (17) or (18), which depends on whether the ANN represents the action that has been selected. In the next sampling period, a new state s_{t+1} will be observed and the reward r_{t+1} will be calculated by (22). If the reward r_{t+1} becomes "+1," which means the MPP is found, the online learning algorithm stops and the MPPT algorithm switches to the online application process. If the reward r_{t+1} is "−1," the controller of the WECS continues to learn and the weights of the three ANNs are then updated according to (16). The flowchart of the proposed ANN-based RL MPPT control is shown in Fig. 6.

C. ORB-Based Online Application Process

When the MPP is found, the optimum relationship of ω_r andP_e is calculated by

$$P_e = K_{\text{opt}}\omega_r^3 \tag{25}$$

where K_{opt} is the optimal parameter to be determined. Then, in the application process, for the WECS in the speed control mode, the rotor speed command is obtained by [10]

$$\omega_r^* = \sqrt[3]{\frac{P_e}{K_{\text{opt}}}}. \tag{26}$$

Therefore, the WECS can be controlled by the fast and high-efficiency ORB MPPT method in the application process.

TABLE I
PARAMETERS OF THE PMSG-BASED WECS USED IN SIMULATION STUDY

Specifications	Values
Rated power	5 MW
Rated voltage	3 kV
Rated frequency	10 Hz
Stator winding resistance	0.001 p.u.
Stator leakage reactance	0.04 p.u.
Unsaturated reactance	$L_d = L_q = 0.4$ p.u.
Magnetic strength	1.07 p.u.

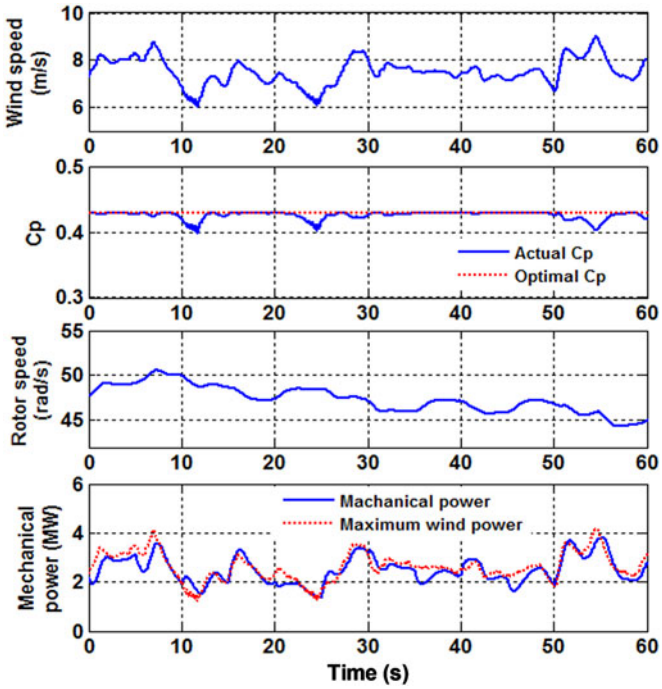


Fig. 7. Simulation results of the WECS controlled by the conventional P&O MPPT method.

VI. SIMULATION RESULTS

The proposed ANN-based RL MPPT control algorithm is validated by simulation studies in PSCAD/EMTDC for a 5-MW PMSG-based WECS. The schematic of the overall control scheme of the WECS is shown in Fig. 5, in which the rotor speed of the PMSG is controlled by the vector control method [22]. The parameters of the PMSG are listed in Table I.

The WECS is first operated in a randomly varying wind condition using the conventional P&O MPPT method. Fig. 7 shows the wind speed profile used in the simulation and the results of C_p, rotor speed, maximum wind power, and input mechanical power of the PMSG. The results show that the controller of the PMSG is always searching for the MPP. Although the actual value of C_p tracks the maximum value well with only a

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

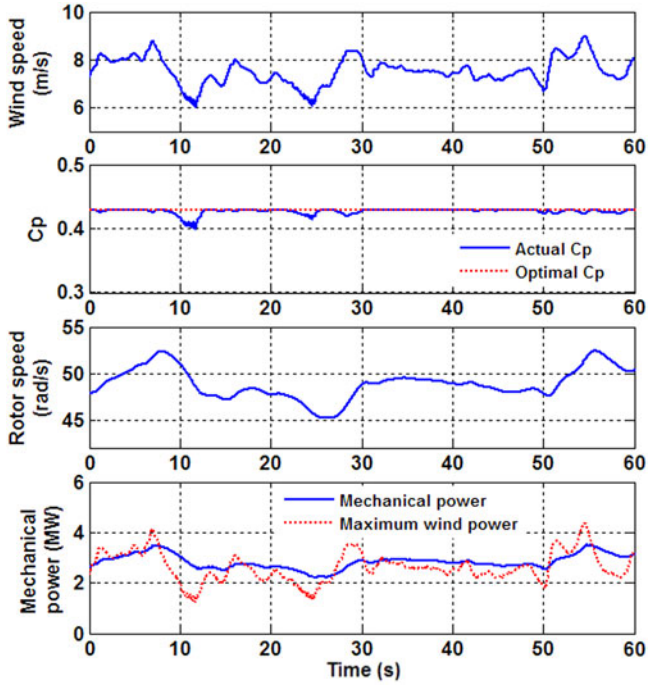


Fig. 8. Simulation results of the WECS in the online application process of the ANN-based RL MPPT control.

small deviation, the rotor speed of the PMSG changes with a fixed value in each fixed time step. As a result, an undershoot or overshoot appears in the mechanical power of the PMSG each time the rotor speed changes, which makes the PMSG mechanical power fluctuating and unsmooth and may be undesirable in the WECS operation. This issue is caused by the nonminimum phase characteristic of the conventional P&O method [27].

The simulation results of the proposed method in the ORB online application process are shown in Fig. 8 for the same varying wind condition used in Fig. 7. It can be seen that a fast MPPT is achieved by using the ORB control expressed by (26). During the wind variation, the actual value of C_p tracks the maximum value well. Compared with the fast variation of the wind speed, the rotor speed and the input mechanical power of the PMSG are smoothed because of the large system inertia. Therefore, the proposed MPPT algorithm enables the PMSG-based WECS with large system inertia to output smoother maximum power with less fluctuation than the conventional P&O MPPT method under the same varying wind speed conditions. This is desired in the WECS operation.

Fig. 9 compares the electrical energies produced by the WECS controlled by the two different MPPT algorithms in the 1-min simulation. The proposed MPPT algorithm enables the WECS to produce a total energy of about 158 MJ, which is 6% more than 149 MJ energy produced by the WECS with the conventional P&O MPPT method.

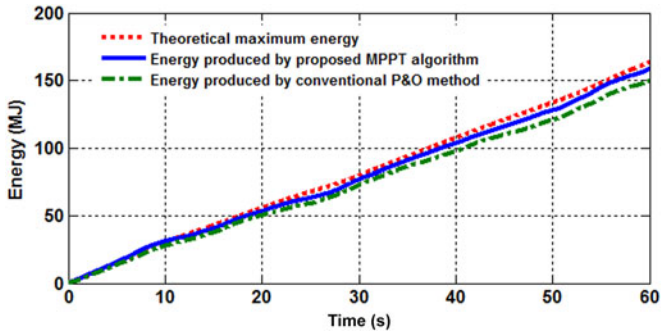


Fig. 9. Comparison of the energies produced by the PMSG equipped with the proposed and conventional MPPT control algorithms in the 1-min simulation.

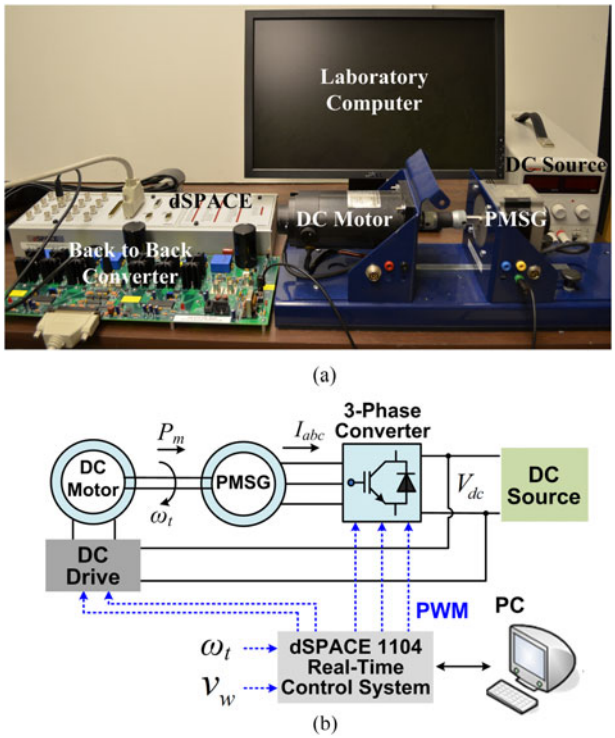


Fig. 10. Test stand of the PMSG-based WECS emulator: (a) the setup and (b) the schematic.

VII. Experimental Results

A. Test Stand Setup

An experimental test stand, as shown in Fig. 10(a), is designed to validate the proposed ANN-based RL MPPT control algorithm. The schematic of the overall test stand of the PMSG-based WECS emulator is shown in Fig. 10(b), which includes a 200-W salient-pole PMSG connected back to back with a 200-W dc machine. The dc machine is powered by a dc source through a dc drive and works as a prime mover motor to emulate the dynamics of a wind turbine to drive the PMSG directly. The GSC and power grid in Fig. 2 is replaced by a dc source, which maintains the dc bus voltage and absorbs the power generated by the PMSG via a three-phase converter that works as the MSC. The specifications of the dc motor and the PMSG are listed in Table II. The vector control scheme [22] is applied to control the rotor speed of the PMSG. The overall control scheme is implemented on a dSPACE 1104 real-time control board. All of the experimental results are recorded by using the ControlDesk installed on a laboratory computer, which is connected with the

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

TABLE II
PARAMETERS OF THE DC MOTOR AND PMSG FOR EXPERIMENTS

Specifications	DC Machine	PMSG
Maximum Speed	3500 r/min	3000 r/min
Rated Power	200 W	200 W
EMF Constant	0.0087 V/r/min	0.0095 V/r/min
Stator Resistance	0.39 Omega;	0.23 Omega;
Inductance(s)	Armature Inductance 0.67 mH	$L_d = 0.275$ mH $L_q = 0.364$ mH
# of Pole Pairs	N/A	4

TABLE III
PARAMETERS OF THE ANN-BASED RL MPPT ALGORITHM USED IN EXPERIMENTS

Parameters	Values
η in (7)	0.95
γ in (7)	0.75
λ in (14)	0.75
$\Delta\omega_r$ in (20)	5 rad/s
δ_1 in (22)	0.8 W
τ in (23)	Decrease from 1 to 0.05 in 20 sample steps

dSPACE system. The parameters of the ANN-based RL MPPT control algorithm used in the experimental studies are listed in Table III.

B. Experimental Results

Experiments are conducted on the test stand to verify the convergence and effectiveness of the proposed ANN-based RL MPPT control algorithm under constant, step-change, and time-varying wind speeds. The proposed MPPT algorithm is also compared with the conventional P&O MPPT algorithm by experimental results to show its superiority. The proposed MPPT algorithm is first tested under a constant wind speed of 7 m/s to verify its convergence to the optimal value. The power coefficiency C_p of the emulated wind turbine as well as the rotor speed and output electrical power of the PMSG are shown in Fig. 11. The controller of the PMSG starts to search for the MPP from the intial rotor speed of 185 rad/s and then decreases the rotor speed to explore. A negetive reward is given for choosing this wrong action. After that the controller changes its searching direction and increases the rotor speed of the PMSG. From about 14 to 20 s, the controller stops searching and stays at about 235 rad/s. The controller tries to explore again after this stop and reaches the MPP (245 rad/s, 66 W) at about 30 s.

The online learning process of the proposed MPPT control algorithm is executed under a step-change wind speed profile to show the difference of the ANN-based RL from the conventional P&O algorithm. In this test, the wind speed changes between 7 and 8 m/s alternatively every 30 s. The values of C_p of the emulated wind turbine and the rotor speed and output electrical power of the PMSG using the proposed and conventional MPPT control algorithms are shown in Figs. 12 and 13, respectively. The PMSG is very "naive" before 90 s by using the proposed RL algorithm, as shown in Fig. 12, because it is gathering experience during this period. The maximum C_p before

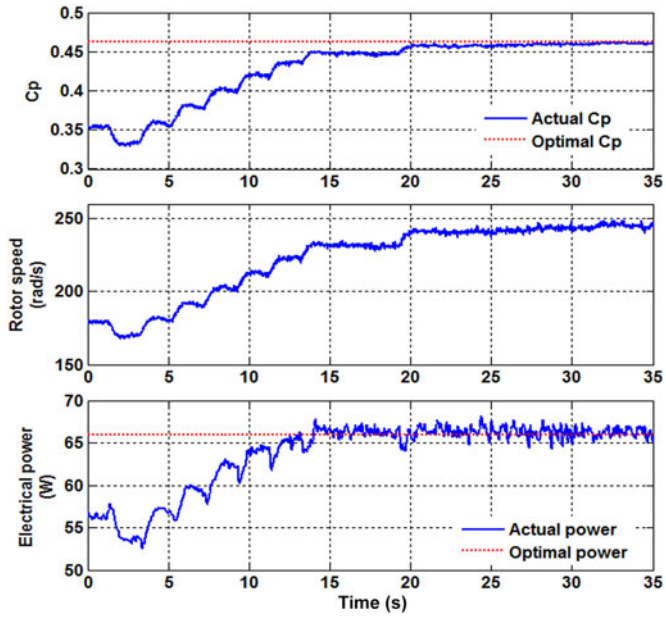


Fig. 11. Performance of the proposed ANN-based RL MPPT control algorithm under constant wind speed.

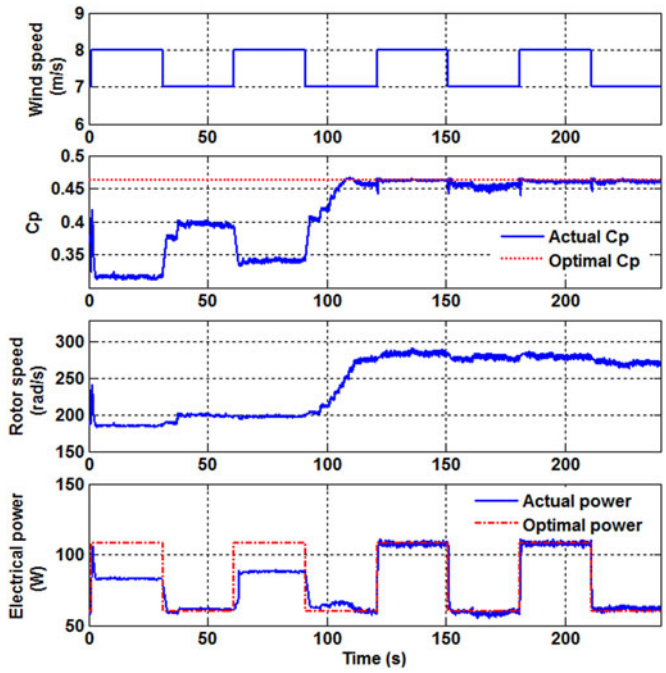


Fig. 12. Performance of the proposed ANN-based RL MPPT control algorithm under step-change wind speeds.

90 s is 0.4, which is below the optimal value 0.46. The rotor speed has no obvious variation in this period mainly because the controller does not know exactly where to go. Also, there is a large difference between the actual and optimal output electrical powers. The first time that C_p reaches the optimal value occurs at approximately 110 s. After that, the PMSG becomes "learned" and continues to learn; each time the PMSG experiences the same wind speed, it behaves better. For example, the performance of the PMSG from 210 to 240 s is better than that from 150 to 180 s under the same wind speed of 7 m/s, which

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

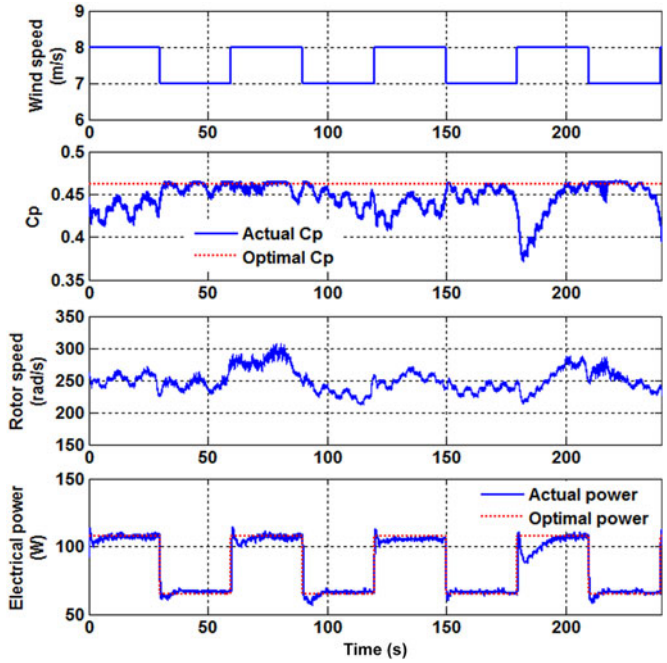


Fig. 13. Performance of the conventional P&O MPPT control algorithm under step-change wind speeds.

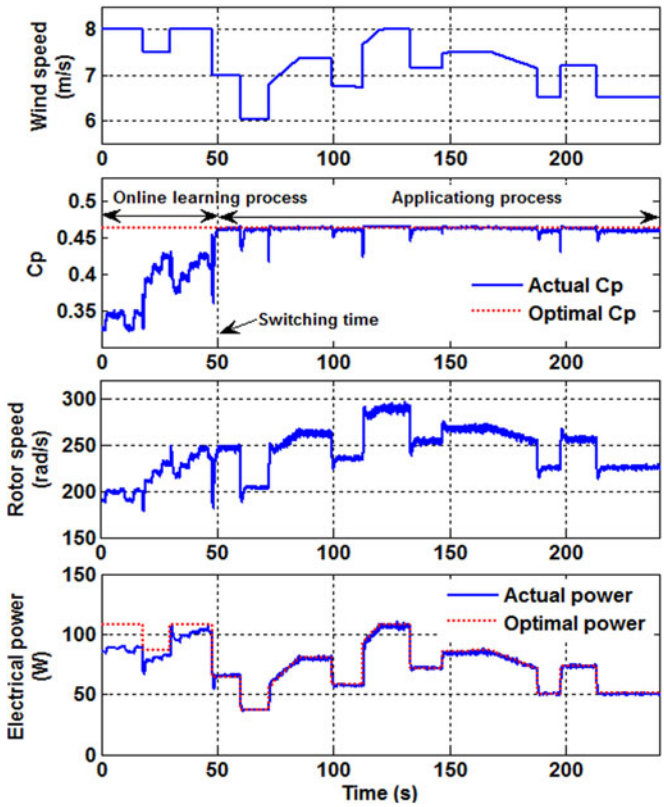


Fig. 14. Transition of the online learning process and the application process in the proposed ANN-based RL MPPT control algorithm.

can been seen from the C_p curve in Fig. 12. Fig. 13 shows the same results as in Fig. 12 when the conventional P&O algorithm is used. Each time the wind speed changes, the controller of the PMSG searches for the MPP without using the previous experience, which reduces the MPPT efficiency and may even lead to a search in a wrong direction (e.g., around 180 s), causing a significant reduction of the output power. This, however, does not happen when the proposed RL algorithm is used as the RL enables the controller of the WECS to learn from its experience, which is the main advantage of the proposed ANN-based RL method over the conventional P&O method. Therefore, by using the knowledge on the wind speed conditions that the controller has learned previously, the proposed ANN-based RL method improves the MPPT efficiency of the WECS.

The proposed entire MPPT control algorithm, including the online learning and application processes as shown in Fig. 6, is tested under a variable wind speed profile. In this test, the wind speed is step changed or ramp changed randomly between 6 and 8 m/s to simulate relatively stable wind speed conditions for the execution of the online learning process. Fig. 14 shows the C_p of the emulated wind turbine and the rotor speed and output electrical power of the PMSG in these two processes. Before 50 s, the PMSG explores and learns by using the proposed ANN-based RL method. During this period, a negative reward is given each time the rotor speed command changes to make the PMSG learn fast. The value of C_p increases from 0.32 from the beginning when the wind speed is 8 m/s and increases more significantly when a similar wind speed is experienced from about 30 through 48 s, which indicates that the PMSG learns from its own experience. At 50 s, the C_p reaches its optimal value of 0.46, at which the optimal relation between the rotor speed and the output electrical power is obtained. At the same time, a positive reward is given to the MPPT controller to trig a transition from the online learning process to the application process. After 50 s, the PMSG is operated in the application process using the learned optimal relation to achieve the fast and effective MPPT control. As a result, the C_p and the output electrical power track their optimal values well each time the wind speed changes. When the learned optimal relation is no longer valid due to factors such as aging of the system, the online RL can be reactivated to obtain the new optimal relation.

The proposed MPPT control algorithm using the learned optimal relation in the application process is compared with the conventional P&O MPPT method using a 2-min field-measured varying wind speed profile. The wind speed changes randomly between 6 and 8 m/s. The C_p of the emulated wind turbine and the rotor speed and output electrical power of the PMSG using the proposed and conventional MPPT control algorithms are shown in Figs. 15 and 16, respectively. During the test, the C_p and output electrical power track their optimal values well by using the proposed MPPT algorithm, as shown in Fig. 15. The largest deviation of the C_p from its optimal value is 0.02 and no obvious deviation between the actual and optimal output electrical powers is found. However, when using the conventional MPPT method, the largest deviation of the C_p from its optimal value is 0.06, and deviations of the actual electrical power from its optimal value are observed from time to time, as shown in Fig. 16.

Fig. 17 compares the electrical energies produced by the PMSG controlled by the two different MPPT algorithms in the

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

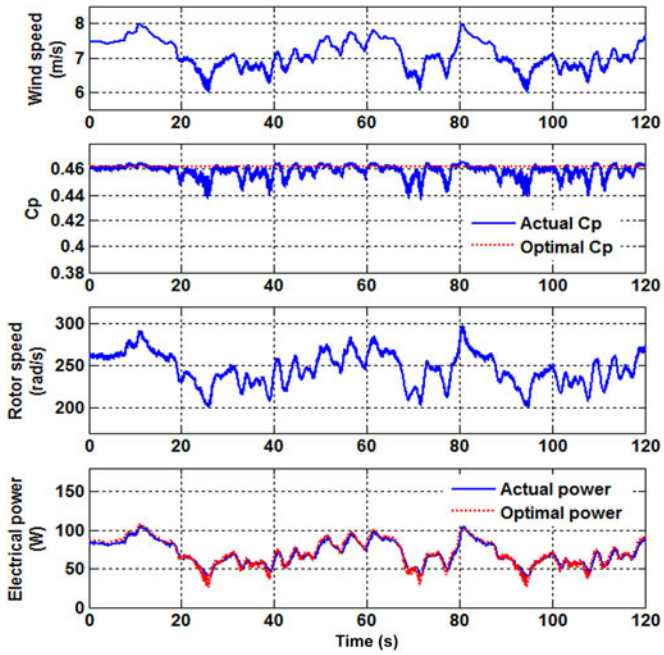


Fig. 15. Performance of the proposed ANN-based RL MPPT control algorithm under variable wind speeds in the application process.

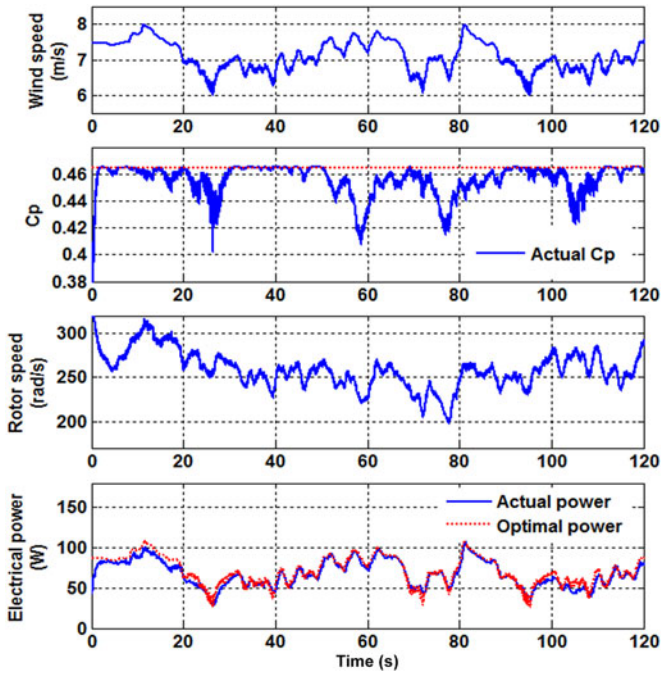


Fig. 16. Performance of the conventional P&O MPPT control algorithm under a variable wind speed condition.

2-min experiments. The energy produced by the PMSG with the proposed MPPT control method tracks the theoretical maximum energy well in the 2-min experiment. The proposed MPPT algorithm enables the PMSG to produce a total energy of about 8.4 kJ in the 2 min, which is 5% more than 8 kJ energy produced by the PMSG with the conventional P&O MPPT method.

It has been found that the power generation capability of a wind turbine will degrade year by year due to the aging of the system, which causes a changing power curve [16]. The MPPT

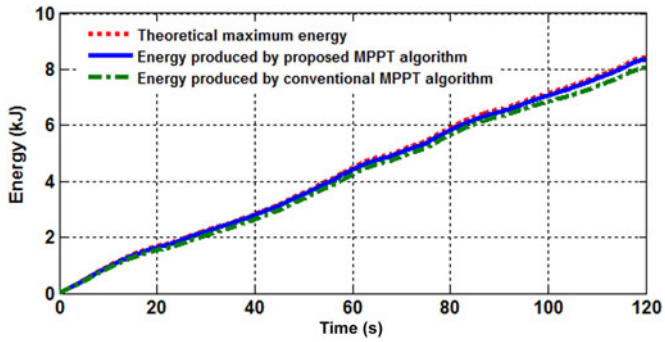


Fig. 17. Comparison of the energies produced by the PMSG equipped with the proposed and conventional MPPT control algorithms in the 2-min experiment.

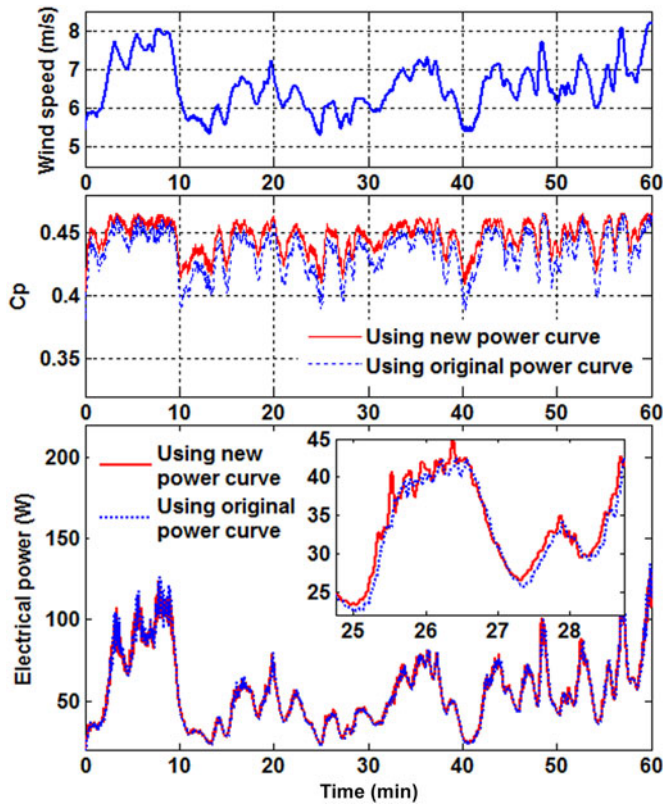


Fig. 18. MPPT performance using the original and newly learned optimal power curves after the WECS ages.

control will not be accurate if a fixed optimal power curve is used after the wind turbine has operated for some years. The proposed MPPT control method enables the controller of the WECS to learn the optimal power curve again when the system ages so the WECS can be controlled with the new optimal power curve to produce more electrical power. To emulate an aged wind turbine after five-year operation, the original wind turbine power-rotor speed characteristic curves are decreased by 9% (see Fig. 2). The online learning process is reactivated and the controller of the emulated wind turbine learned a new optimal power curve. The MPPT performance is then tested using the original and newly learned optimal power curves using a field-measured 1-h varying wind speed profile. The wind speed, C_p of the emulated wind turbine, and output electrical power of the PMSG are shown in Fig. 18. It clearly shows that the C_p as the result of

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

using the original optimal power curve is always smaller than that using the new optimal power curve. In addition, the output power of the PMSG by using the newly learned optimal power curve is larger than that using the original optimal power curve. Therefore, the proposed MPPT control algorithm enables the MPPT control to be adaptive to the aging process of the wind turbine such that it can always generate the maximum wind power.

VIII. Conclusion

This paper has proposed an intelligent ANN-based RL MPPT algorithm for PMSG-based WECSs. ANNs and the Q-learning method have been combined to learn the optimal relationship between the rotor speed and output electrical power of the PMSG. Then, the WECS is controlled based on the learned optimal relationship for MPPT. The online RL algorithm can reactivated to obtain a new optimal relationship when the original one is no longer valid due to factors such as system aging. The proposed online learning algorithm enables the WECS to learn from its own experience, thus improving the learning efficiency. The effectiveness of the proposed MPPT control algorithm has been demonstrated by simulation results for a 5-MW PMSG-based WECS and experimental studies for an emulated 200-W PMSG-based WECS. The proposed method can be extended to other types of WECSs, such as doubly fed induction generator-based WECSs.

References

[1] A. D. Hansen, F. Iov, F. Blaabjerg, and L. H. Hansen, "Review of contemporary wind turbine concepts and their market penetration," *J. Wind Eng.*, vol. 58, no. 4, pp. 1081–1095, Apr. 2011.

[2] M. Liserre, R. Cardenas, M. Molinas, and J. Rodriguez, "Overview of multi-MW wind turbines and wind parks," *IEEE Trans. Ind. Electron.*, vol. 58, no. 4, pp. 1081–1095, Apr. 2011.

[3] J. Wang, D. Xu, B. Wu, and Z. Luo, "A low-cost rectifier topology for variable-speed high-power PMSG wind turbines," *IEEE Trans. Power Electron.*, vol. 26, no. 8, pp. 2192–2200, Aug. 2011.

[4] S. Li, T. A. Haskew, R. P. Swatloski, and W. Gathings, "Optimal and direct-current vector control of direct-driven PMSG wind turbines," *IEEE Trans. Power Electron.*, vol. 27, no. 5, pp. 2325–2337, May 2012.

[5] S. Zhang, K. Tseng, D. M. Vilathgamuwa, T. D. Nguyen, and X. Wang, "Design of a robust grid interface system for PMSG-based wind turbine generators," *IEEE Trans. Ind. Electron.*, vol. 58, no. 1, pp. 316–328, Jan. 2011.

[6] P. Li, Y. Song, D. Li, W. Cai, and K. Zhang, "Control and monitoring for grid-friendly wind turbines: Research overview and suggested approach," *IEEE Trans. Power Electron.*, vol. 30, no. 4, pp. 1979–1986, Apr. 2015.

[7] B. Wu, Y. Lang, N. Zargar, and S. Kouro, *Power Conversion and Control of Wind Energy Systems*. Hoboken, NJ, USA: Wiley, 2011.

[8] Y. Zhao, C. Wei, Z. Zhang, and W. Qiao, "A review on position/speed sensorless control for permanent magnet synchronous machine-based wind energy conversion systems," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 1, no. 4, pp. 203–216, Dec. 2013.

[9] S. Morimoto, H. Nakayama, M. Sanada, and Y. Takeda, "Sensorless output maximization control for variable-speed wind generation system using IPMSG," *IEEE Trans. Ind. Appl.*, vol. 41, no. 1, pp. 60–67, Jan./Feb. 2005.

[10] A. Mirecki, X. Roboam, and F. Richardeau, "Architecture complexity and energy efficiency of small wind turbines," *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 660–670, Feb. 2007.

[11] N. Nayanar, N. Kumaresan, and N. Ammasai Gounden, "A single-sensor-based MPPT controller for wind-driven induction generators supplying DC microgrid," *IEEE Trans. Power Electron.*, vol. 31, no. 2, pp. 1161–1172, Feb. 2016.

[12] Z. M. Dalala, Z. U. Zahid, W. Yu, Y. Cho, and J. Lai, "Design and analysis of an MPPT technique for small-scale wind energy conversion system," *IEEE Trans. Energy Convers.*, vol. 28, no. 3, pp. 756–767, Sep. 2013.

[13] K. Kim, T. L. Van, D. Lee, S. Song, and E. Kim, "Maximum output power tracking control in variable-speed wind turbine systems considering rotor inertial power," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3207–3217, Aug. 2013.

[14] J. Park, J. Lee, K. Oh, and J. Lee, "Development of a novel power curve monitoring method for wind turbines and its field tests," *IEEE Trans. Energy Convers.*, vol. 29, no. 1, pp. 119–128, Mar. 2014.

[15] H. Long, L. Wang, Z. Zhang, Z. Song, and J. Xu, "Data-driven wind turbine power generation performance monitoring," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6627–6635, Oct. 2015.

[16] I. Staffel and R. Green, "How does wind farm performance decline with age?" *Renew. Energy*, vol. 66, pp. 775–786, June 2014.

[17] G. Hughes. (2012). The performance of wind farms in the United Kingdom and Denmark. Renewable Energy Foundation, London, England. [Online]. Available: http://www.ref.org.uk/attachments/article/280/ref.hughes.19.12.12.pdf

[18] Q. Wang and L. Chang, "An intelligent maximum power extraction algorithm for inverter-based variable speed wind turbine systems," *IEEE Trans. Power Electron.*, vol. 19, no. 5, pp. 1242–1249, Sep. 2004.

[19] Y. Xia, K. H. Ahmed, and B. W. Williams, "A new maximum power point tracking technique for permanent magnet synchronous generator based wind energy conversion system," *IEEE Trans. Power Electron.*, vol. 26, no. 12, pp. 3609–3620, Dec. 2011.

[20] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[21] C. Wei, Z. Zhang, W. Qiao, and L. Qu, "Reinforcement learning-based intelligent maximum power point tracking control for wind energy conversion systems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6360–6370, Oct. 2015.

[22] W. Qiao, X. Yang, and X. Gong, "Wind speed and rotor position sensorless control for direct-drive PMG wind turbines," *IEEE Trans. Ind. Appl.*, vol. 48, no. 1, pp. 3–11, Jan./Feb. 2012.

[23] C. Szepesvári, "Algorithms for reinforcement learning," *Synthesis Lectures Artif. Intell. Mach. Learn.*, vol. 4, no. 1, pp. 1–103, 2010.

[24] G. A. Rummery and M. Niranjan, "On-line Q-learning using connectionist systems," Eng. Dept., Cambridge Univ., Cambridge, U.K., Tech. Rep. TR-166, Sep. 1994.

[25] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, pp. 279–292, 1992.

[26] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics Intell. Lab. Syst.*, vol. 39, no. 1, pp. 43–62, Nov. 1997.

[27] M. Nasiri, J. Milimonfared, and S. H. Fathi, "Modeling, analysis and comparison of TSR and OTC methods for MPPT and power smoothing in permanent magnet synchronous generator-based wind turbines," *Energy Convers. Manage.*, vol. 86, pp. 892–900, 2014.

Chun Wei (S'12) received the B.S. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2009, and the M.S. degree in electrical engineering from North China Electric Power University, Beijing, in 2012. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Nebraska−Lincoln, Lincoln, NE, USA.

His research interests include wind energy conversion systems, power electronics, motor drives, and computational intelligence.

Zhe Zhang (S'10–M'15) received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2010, and the Ph.D. degree in electrical engineering from the University of Nebraska−Lincoln, Lincoln, NE, USA, in 2015.

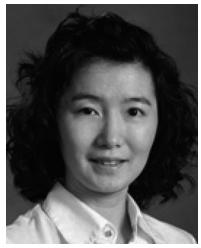
He is currently with Nexteer Automotive, Saginaw, MI, USA, as a Senior Motor Control Engineer. His research interests include control of motor drives, wind energy conversion systems, and power electronics.

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.

Wei Qiao (S'05–M'08–SM'12) received the B.Eng. and M.Eng. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 1997 and 2002, respectively, the M.S. degree in high-performance computation for engineered systems from Singapore-MIT Alliance, Singapore, in 2003, and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA in 2008.

Since August 2008, he has been with the University of Nebraska−Lincoln, Lincoln, NE, USA, where he is currently an Associate Professor at the Department of Electrical and Computer Engineering. He is the Author or Coauthor of three book chapters and more than 170 papers in refereed journals and conference proceedings. His research interests include renewable energy systems, smart grids, condition monitoring, power electronics, electric machines and drives, and new energy conversion devices.

Dr. Qiao is an Editor of the IEEE TRANSACTIONS ON ENERGY CONVERSION, an Associated Editor of the IET Power Electronics and the IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN POWER ELECTRONICS, and the Corresponding Guest Editor of a special section on Condition Monitoring, Diagnosis, Prognosis, and Health Monitoring for Wind Energy Conversion Systems of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS. He was an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS from 2010 to 2013. He received the 2010 U.S. National Science Foundation CAREER Award and the 2010 IEEE Industry Applications Society Andrew W. Smith Outstanding Young Member Award.



Liyan Qu (S'05–M'08) received the B.Eng. (with the highest distinction) and M.Eng. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 1999 and 2002, respectively, and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2007.

From 2007 to 2009, she was an Application Engineer with Ansoft Corporation, Irvine, CA, USA. Since January 2010, she has been with the University of Nebraska−Lincoln, Lincoln, NE, USA, where she is currently an Assistant Professor at the Department of Electrical and Computer Engineering. Her research interests include energy efficiency, renewable energy, numerical analysis and computer-aided design of electric machinery and power electronic devices, dynamics and control of electric machinery, permanent-magnet machines, and magnetic devices.

Authorized licensed use limited to: National Chung Cheng University. Downloaded on March 11,2021 at 09:24:02 UTC from IEEE Xplore. Restrictions apply.