IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 28, 2018, accepted October 18, 2018, date of publication October 23, 2018, date of current version December 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2877206

Adaptive Critic Based Optimal Neurocontrol of a Distributed Microwave Heating System Using Diagonal Recurrent Network

TONG LIU[1,2], SHAN LIANG[1,2], (Member, IEEE), QINGYU XIONG[1,3], AND KAI WANG[1,2]
[1]Key Laboratory of Dependable Service Computing in Cyber Physical Society, Ministry of Education, Chongqing University, Chongqing 400044, China
[2]College of Automation, Chongqing University, Chongqing 400044, China
[3]College of Software Engineering, Chongqing University, Chongqing 401331, China

Corresponding author: Shan Liang (lightsun@cqu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61771077.

ABSTRACT The dynamic capture ability of diagonal recurrent neural network (DRNN) makes them a suitable candidate for implementing real-time nonlinear adaptive controllers to handle the nonlinearity and uncertainty of high-power distributed microwave heating system (HPDMHS). In conventional DRNN-based adaptive control, the diagonal recurrent neural controller (DRNC) is trained online with one step cost and control law are not always optimal. To improve this, this paper couples a simple direct adaptive neural control with adaptive critic design (ACD) technique to achieve the optimal temperature tracking in HPDMHS. After transforming the original optimal temperature tracking control problem into an error regulation problem, the desired control is obtained by a regular DRNC, while the error regulation control is solve by ACD technique using DRNN. Simulation results demonstrate the superiority of the proposed DRNC-ACD method over conventional adaptive control in temperature tracking for HPDMHS.

INDEX TERMS High-power distributed microwave heating system, diagonal recurrent neural network, optimal temperature tracking control, adaptive critic design, direct adaptive control.

I. INTRODUCTION

Microwave heating has been widely applied in food and chemical industries as an attractive alternative to conventional heating methods, as it enjoys the superiorities of brief startup periods, reduced power consumption and shorten processing time [1]. Recent years, to improve the overall efficiency and productivity, numerous studies have been focused on the controller designs in microwave heating systems (MHS) [2]–[4]. Akkari *et al.* [2] proposed a global linearizing control based on a gray-box model, however an undershoot is obvious in the middle stage of microwave thawing process. Yuan *et al.* [4] further presented a sliding mode control strategy based on a thermodynamic model. Besides, considering input saturation, internal uncertainty and external disturbance, Zhong *et al.* [3] proposed a H_∞ guaranteed cost tracking control method based on a one-dimensional cavity model. Results demonstrated that the problem of thermal runaway and local overheating in heated medium can be well prevented. However, the mechanism models mentioned above are characterized by their high nonlinearity, which directly contributes to the high complexity in controller design. Especially in industrial applications, MHS usually integrates multiple microwave sources by means of power synthesis to form the high-power distributed microwave heating system (HPDMHS) to achieve the large scale synchronous heating purpose. In such systems, each independent microwave source is capable of automatically adjusting its power according to the environment, which contributes to the more complex coupling of multi-physical fields and stronger interaction among system states. Although in our previous work, a temperature prediction model was developed for distributed microwave heating system using deep learning and isolation forest [5], the control problem has not been considered. The direct consequence of the nonlinearity and complexity of HPDMHS is the wide spread system uncertainties that prevent the problem solving from using the conventional model-based control approaches. This contributes to the development of more sophisticated techniques which are more suitable to handle such complex system.

Numerous sophisticated control techniques have been arisen recent years to handle such system nonlinearities and complexities, such as extremum seeking control [6], [7], robust control [8], iterative learning control [9] and neural control [10]–[12]. They have been successfully applied in lighting systems, mechanical systems and power systems. Among them, neural networks are found to be the most promising tool in modeling and control of complex real systems with high degree of nonlinearity. This motivates us employ neural networks as an identifier and controller for HPDMHS. Precisely, feedforward neural networks (FNNs) is a static network without the tapped delays. In contrast, diagonal recurrent neural network (DRNN) that consists of both feedforward and feedback connections in network structure allows them to store information for later use, which provides it the ability to capture the dynamics of the plant [13]. Ku and Lee [14] has pointed that DRNN is more effective and efficient for implementing real time nonlinear adaptive controllers than FNNs. From the view of control system design, Kumar *et al.* [15] used two distinct neural networks to control systems adaptively by direct and indirect control. Subsequently, Ku and Lee [13] proposed indirect control architecture by using DRNN. The control architecture consists of a diagonal recurrent neural controller (DRNC) and a diagonal recurrent neural identifier (DRNI). In this regard, an unknown system is identified by the DRNI, which provides the information about the system to the DRNC. The DRNC generates a control signal to drive the unknown system such that the error between actual system output and desired output is minimized. Since the design of DRNC requires rarely information of the plant and the control performance is completely determined by learning algorithm, such control scheme is easy implementation and simple adjustment.

Ku *et al.* [16] further applied this DRNN-based adaptive control in a nuclear reactor system. The use of dynamic back propagation with adaptive learning rate guarantees convergence of whole system. But the plant output does not track the reference model very well in low power regions. Gao *et al.* [17], has employed DRNN for control water tank temperature. The advantage is that it does not require intensive computation at each control instant and has potential applications to the situation where a process is sufficiently nonlinear and real-time implementation is critical. In [18], a robust DRNN based state observer is designed for nonlinear systems, but its control performance has not been tested. In [19], a new weight update algorithm based on stochastic automation has been proposed for DRNN control system design. Chow and Fang [20] further developed two-dimensional theory based online learning algorithm for recurrent neural network control of nonlinear systems. Reference [21] has presented a direct and indirect adaptive control approach for nonlinear systems based on a fuzzy recurrent neural network. Reference [22] has designed an internal model controller for pneumatic manipulators, where a FNN and a DRNN are employed as neural controller and identifier, respectively. Reference [23] used DRNN for control of variable flow heating system and gradient descend is derived for online training. In [24], a sigmoid diagonal recurrent neural network (SDNN) is proposed for identification and control of nonlinear systems, where its new network structure is based on adaptation sigmoid weight of hidden layer neurons. Reference [25] has employed two type of neural networks to control a single-link flexible arm. They have shown that the DRNN-based control achieve a higher precision of tip motion tracking than that of in FNN-based control. In [26], a chaotic DRNN with logistic mapping was proposed to improve the convergent speed and generalization performance of conventional DRNN. Its corresponding adaptive control scheme is further presented to control the chaos. Reference [27] proposed an adaptive control for nonlinear dynamic systems based on DRNN. A new learning algorithm based on Lyapunov stability criterion is developed to update parameters of DRNC. Simulation results show that the proposed control algorithm outperforms the conventional DRNN-based control in terms of accuracy and robustness.

From the literature survey it was found that the application of DRNN as a controller has held great intuitive appeal and has attracted considerable attention in recent years. On the one hand, some papers are focusing on the development of new learning algorithm for DRNC, since the use of traditional gradient descend has the tendency of stucking at local minimal. On the other hand, some papers combine DRNC with various optimization techniques to improve the control performance. However, most of these researches simply use the traditional direct or indirect control scheme. Form the view of cybernetics, two prominent deficiencies are presented in such control scheme: one is that DRNC usually meets the difficulty of many-to-one inverse mappings; the other is that since the DRNC is online trained with only one step error or cost, the control law is not always optimal.

To handle the above mentioned problems, this paper steps through the components of a new diagonal recurrent neural adaptive control scheme designed specifically to provide highly accurate temperature tracking results for HPDMHS. While using DRNN as the foundation, the traditional direct neural control method is integrated with an adaptive critic design (ACD) technique. The critic method removes the learning process from simply one-step cost, but learns to minimize the sum of all discounted costs, or cost-to-go, which provides the optimal and robust feedback control laws [28]. At the first step of our new design, DRNC is employed to generate desired control signal using distal supervised learning approach. This employs the conventional direct adaptive control frame, where the optimal manner is difficult to achieve. Thus we transform the original optimal tracking control problem into an error regulation problem, which has divided the problem into two parts. The ACD technique is then introduced with DRNN to employ as an optimal regulator for the tracking error dynamics. Through combining the desired control (generated by DRNC) with regulation control (generated by ACD), the final optimal control law can be achieved through the hybrid DRNC-ACD strategy. The main contributions of this paper are summarized as follows.

1) Propose to couple an direct adaptive DRNN control with ACD technique to achieve the optimal tracking control. Compared with [13] and [14], the tracking

error dynamics can be regulated by the added critic and actor DRNN. Compared with [29] and [30], DRNN is employed to replace the conventional FNN as a more robust identifier and controller.

2) The proposed control strategy is applied to HPDMHS, where the system dynamics is completely unknown.
3) The temperature tracking control performance of the proposed method is tested and compared with the conventional DRNN adaptive control in HPDMHS.

The rest of the paper is organized as follow. Sect. II gives the brief introduction of HPDMHS and its identification model. In Sect. III, some preliminaries about DRNN, DRNN-based adaptive control and ACD are introduced. Sect. IV gives the hybrid DRNC-ACD construction for the temperature tracking control in HPDMHS. The effectiveness of the proposed method is tested and verified comparing with conventional DRNN-based control method in Sect. V. Finally, Sect. VI concludes the paper with remarks on the future work.

II. MODELING OF HPDMHS BASED DRNN

A. CONFIGURATION OF HPDMHS

The Schematic of HPDMHS is shown in Fig. 1. The system is mainly constructed of four parts, multi-microwave source subsystem (MSS), multimode cavity (MC), medium transmission subsystem (MTS), and data acquisition subsystem (DAS). The MC has a rectangular shape with an inlet opening for the raw materials and an outlet discharge to receive the heated materials. The materials are distributed evenly on the MTS, of which the speed of conveyor belt can be manipulated by a motor driver. The MSS integrates five microwave sources and each one is capable of continuous adjustment ranging from 0-3kW. The microwave power is then directed into the MC through waveguide for heating purpose. As a DAS, an optical fiber thermometer (FTS-I201, Optsensor, China) is employed at the outlet of the MC to measure the temperature of exported material, while PLC is employed to establish the required connection between device and computer through RS-232/RS-485 converter by the integrated Modbus protocol. The microwave powers, temperature measurement and conveyor speed are then real time recorded until the heating process is finished, thus, a set of input/output data sample ready to be used in system identification to develop the desired model can be obtained.

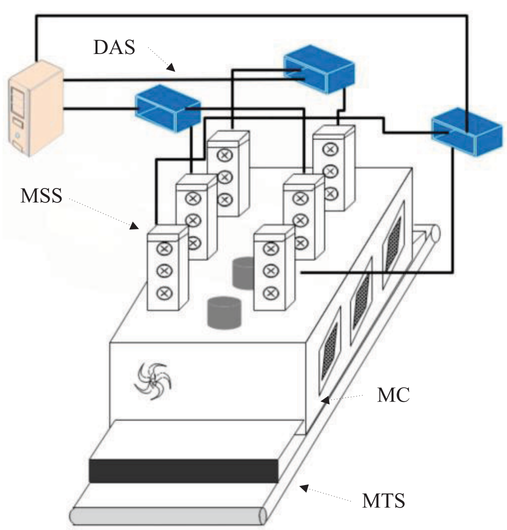

FIGURE 1. Schematic of experimental HPDMHS.

B. SYSTEM IDENTIFICATION BASED ON DRNN

System identification is based on applying a suitable input signal to the system to be identified in order to excite its dynamics and consequently to observe its response to the applied input signal. In this context, a data collecting experiment is conducted to obtain the required input/output data set by a real-time microwave rice drying experiment. In this data set, the input signals to HPDMHS are five microwave powers and conveyor speed. Turning up the output power can enhance the electromagnetic field, while turn down the conveyor speed can prolong the microwave radiation time and thus raising the temperature, vice versa. The output signal to HPDMHS is rice temperature that can be measured by an optical fiber thermometer. In general, the DRNN based identification is a combination of past output and input values from the system and the input signal, respectively. Thus, the mathematical model of HPDMHS can be expressed as follow:

$$y(k+1) = F(y(k), u_{p1}(k), u_{p2}(k), u_{p3}(k), u_{p4}(k), u_{p5}(k), u_v(k)) \quad (1)$$

where $u_{p1}(k)$ to $u_{p5}(k)$ denote the incident powers of five microwave sources, $u_v(k)$ is the conveyor speed which is converted by motor frequency, $y(k)$ and $y(k+1)$ represent the current and next moment sample temperature, respectively. The function $F(\cdot)$ can be identified by neural networks due to its universal approximation ability. In our work, a DRNN with the structure of 7-15-1 is constructed to identify the system. After a sufficiently long training process, it is able to learn the system dynamics. Thus, a DRNI for HPDMHS is built.

III. PRELIMINARIES

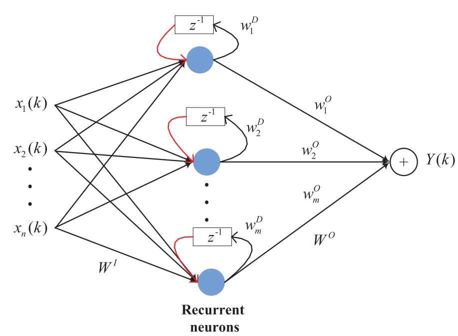

FIGURE 2. Structure of DRNN.

A. DRNN

Fig. 2 shows the structure of DRNN. Input weight vector connected external inputs to hidden layer is denoted as W^I. Neuron presented in the hidden layer is known as recurrent neuron that has an internal feedback loop to itself. The diagonal weights are represented as red curves and in case of DRNN they provide a weighted unit delay output of the recurrent neuron as an input to the same recurrent neuron. Because of this, DRNN can store information for later use and is better at dealing with time-varying input and output. The diagonal weights are represented as $W^D = \{w_1^D, w_2^D, \ldots, w_m^D\}$, where m denotes the number of recurrent neurons in the hidden layer. The output weight is denoted by $W^O = \{w_1^O, w_2^O, \ldots, w_m^O\}$, while the input vector is $X = \{x_1(k), x_2(k), \ldots, x_n(k)\}$, where n denotes the input dimension. Then the mathematical model for DRNN is shown below

$$y(k) = \sum_m W_m^O H_m(k) \tag{2}$$

$$H_m(k) = \varphi(S_m(k)) \tag{3}$$

$$S_m(k) = W_m^D H_m(k-1) + \sum_n W_{nm}^I x_n(k) \tag{4}$$

where for each time step k, W^I denotes the input weights, $S_m(k)$ is the sum of inputs of mth recurrent neuron, $H_m(k)$ is the output of mth recurrent neuron, and $y(k)$ is the output of DRNN. It should be noted that activation function $\varphi(\cdot)$ chosen in the recurrent neuron is usually nonlinear tangent hyperbolic function which enables the output have both positive and negative value, while the activation in the output neuron is considered to be linear so that there is no restriction on its value.

B. DRNN-BASED ADAPTIVE CONTROL

Weight vectors W^I, W^D and W^O denote the parameters of DRNC. These parameters are required to be adjusted during online training. This will make DRNC yield desired control signals which causes the output of DRNI to follow the desired reference model output. The gradient descend is used to derive the weight adjustment rule. The conventional DRNN-based adaptive control scheme is shown in Fig. 3. $u(k)$ denotes DRNC output. $y(k)$ and $y_r(k)$ denote the DRNI and reference model output, respectively. The TD component represents the tapped delay whose output is a unit delay value of its input, hence $y(k-1)$ and $u(k-1)$ are outputs of TD components, which are introduced to the inputs of DRNC. In general, the unknown plant is first offline identified by DRNI. Then its weights are frozen. The DRNC is used to drive the DRNI so that the error $e(k) = y_r(k) - y(k)$ between the actual output and the desired output is minimized. For convenience of analysis, the weight vectors of DRNN are generalized expressed as W^G. $E(k)$ is defined as the objective function which can be expressed as

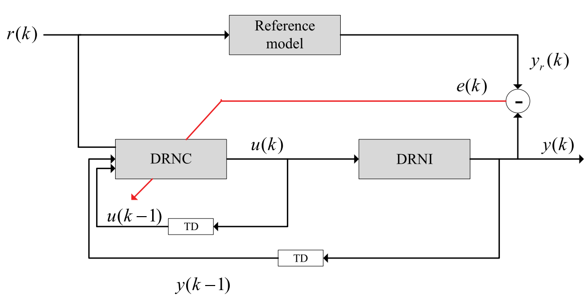

FIGURE 3. Block diagram of DRNN-based adaptive control.

$$E(k) = \frac{1}{2}(e(k))^2 \tag{5}$$

Then using the backpropagation, the update rule is as follow:

$$\begin{aligned} W^G(k+1) &= W^G(k) + \eta \Delta W^G(k) \\ &\quad + \alpha(W^G(k) - W^G(k-1)) \\ \Delta W^G(k) &= \frac{\partial E(k)}{\partial W^G(k)} = \frac{\partial E(k)}{\partial e(k)} \frac{\partial e(k)}{\partial W^G(k)} \end{aligned} \tag{6}$$

where η represents learning rate, α denotes momentum factor which enables to speed up the learning without causing instability in the system. $\Delta W^G(k)$ denotes the required adjustment to be made in the weights. It should be noted that $\frac{\partial y(k)}{\partial u(k)}$ is required when solving $\frac{\partial e(k)}{\partial W^G(k)}$ by chain rule. This term is called jacobian/sensitivity of the plant, which can be directly solved through backpropagation by DRNI.

C. ACD

Optimal control theory has been an active area of nonlinear system for several decades. However, the development of nonlinear optimal control technique is mainly hampered by solving Hamilton-Jacobi-Bellman (HJB) equation. The "curse of dimensionality" is almost inextricable in solving the HJB equation by traditional dynamic programming (DP) [31]. To apply the optimization idea of DP successfully for nonlinear system, ACD was proposed in 1997 [32]. It employs an actor-critic way commonly used in reinforcement leaning to search the optimal policy. Critic methods remove the learning process one step from control network, so that desired control action information is not necessary. The critic network learns to approximate the cost-to-go (the function J of the Bellman's equation in DP) and uses the output of an action network as one of its inputs [33].

The family of ACDs has been classified into several main schemes including heuristic dynamic programming (HDP), dual heuristic dynamic programming (DHP), action-dependent HDP (ADHDP), ADDHP, globalized DHP (GDHP) and ADGDHP [28]. In this paper, the simple and powerful HDP [34] approach is adopted for the design of an optimal regulator of error dynamic system.

Considering a discrete-time nonlinear system that can be expressed as

$$y(k+1) = f(y(k)) + g(y(k))u(k) \tag{7}$$

where $y(k)$ represents the state vector of the system, $u(k)$ denotes the control input, $f(\cdot)$ and $g(\cdot)$ are system functions.

For optimal tracking control problem, the control objective is to find the optimal control $u^*(k)$ so that the nonlinear system (7) can track a reference trajectory $y_r(k)$ in an optimal manner [29], [30]. The tracking error can be defined as follow:

$$e(k) = y(k) - y_r(k) \tag{8}$$

Assume that the desired trajectory satisfies the following form:

$$y_r(k+1) = f(y_r(k)) + g(y_r(k))u_r(k) \tag{9}$$

where $u_r(k)$ denotes the desired control input that can be obtained by

$$u_r(k) = g^{-1}(y_r(k))(y_r(k+1) - f(y_r(k))) \tag{10}$$

Then, the regulation control $u_e(k)$ is defined by

$$u_e(k) = u(k) - u_r(k) \tag{11}$$

Considering (7)-(11), the tracking error $e(k+1)$ can be expressed as

$$\begin{aligned} e(k+1) &= y(k+1) - y_r(k+1) \\ &= f(e(k) + y_r(k)) - y_r(k+1) \\ &\quad + g(e(k) + y_r(k))(u_e(k) + u_r(k)) \\ &= f(e(k) + y_r(k)) - y_r(k+1) \\ &\quad + g(e(k) + y_r(k))u_e(k) \\ &\quad + g(e(k) + y_r(k)) \\ &\quad \cdot g^{-1}(y_r(k))(y_r(k+1) - f(y_r(k))) \end{aligned} \tag{12}$$

From (12), we can see that it is possible to use $u_e(k)$ and $e(k)$ to obtain the tracking error $e(k+1)$. For convenience of analysis, (12) can be rewritten as

$$e(k+1) = f_e(k) + g_e(k)u_e(k) \tag{13}$$

where

$$\begin{aligned} f_e(k) &= f(e(k) + y_r(k)) \\ &\quad + g(e(k) + y_r(k))g^{-1}(y_r(k))(y_r(k+1) - f(y_r(k))) \\ &\quad - y_r(k+1) \end{aligned}$$

and $g_e(k) = g(e(k) + y_r(k))$.

To this end, our objective is to find the control sequence $u_e(\cdot) = [u_e(k), u_e(k+1), \ldots]$ to minimizes the following cost-to-go:

$$J(e(k)) = \sum_{i=k}^{\infty} \beta^{i-k} U(e(i), u_e(i)) \tag{14}$$

where β is the discount factor with $0 < \beta \leq 1$, and U denotes utility function that is a measure of one-step cost of control. This can be selected based on minimum-fuel considerations, minimum energy, minimum risk, etc [35]. In our problem, the utility function U is based on the minimum temperature tracking error and microwave energy, which is designed as follow:

$$U(e(i), u_e(i)) = e(i)^T Pe(i) + u_e(i)^T Qu_e(i) \tag{15}$$

where P and Q are symmetric positive definite matrices with appropriate dimensions.

Now, from (11) it is observed that the tracking control $u(k)$ consists of a predetermined desired control $u_r(k)$ corresponding to the reference trajectory $y_r(k)$ and an error regulation control $u_e(k)$ corresponding to the tracking error $e(k)$. Before proceeding, the following theorem is needed.

Theorem 1: Let $u_e(k)$ be an admissible control such that the tracking error system (13) is asymptotically stable. Then the error system function $f_e(k)$ is bounded satisfying

$$\begin{aligned} \|f_e(k)\|^2 \leq \frac{1}{2}(\zeta\lambda_{\min}(P)\|e(k)\|^2 \\ + (\zeta\lambda_{\min}(Q) - 2g_M^2)\|u_e(k)\|^2) \end{aligned} \tag{16}$$

where $\lambda_{\min}(P)$ is the minimum eigenvalue of P, $\lambda_{\min}(Q)$ is the minimum eigenvalue of Q. $\zeta > \frac{2g_M^2}{\lambda_{\min}(Q)}$ is a known positive constant.

Proof: We define the following positive definite Lyapunov function:

$$V_L(k) = e^T(k)e(k) + \zeta J(e(k)) \tag{17}$$

where $J(e(k)) = \sum_{i=k}^{\infty} \beta^{i-k} U(e(i), u_e(i))$ is defined in (14). Since we are working in discrete domain, the difference of the Lyapunov function is given by

$$\Delta V_L(k) = V_L(k+1) - V_L(k) \tag{18}$$

It can be rewritten as

$$\Delta V_L(k) = e^T(k+1)e(k+1) - e^T(k)e(k) + \zeta\Delta J(e(k)) \tag{19}$$

where $\Delta J(e(k)) = J(e(k+1)) - J(e(k))$. Using (13) and (14), we can get

$$\begin{aligned} \Delta V_L(k) = (f_e(k) + g_e(k)u_e(k))^T(f_e(k) + g_e(k)u_e(k)) \\ - e^T(k)e(k) - \zeta(e^T(k)Pe(k) \\ + u_e^T(k)Qu_e(k)) \end{aligned} \tag{20}$$

After applying the Cauchy-Schwarz inequality, we can get

$$\begin{aligned} (f_e(k) + g_e(k)u_e(k))^T(f_e(k) + g_e(k)u_e(k)) \\ \leq 2\|f_e(k)\|^2 + 2g_M^2\|u_e(k)\|^2 \end{aligned} \tag{21}$$

$$\begin{aligned} e^T(k)Pe(k) + u_e^T(k)Qu_e(k) \\ \leq \lambda_{\min}(P)\|e(k)\|^2 + \lambda_{\min}(Q)\|u_e(k)\|^2 \end{aligned} \tag{22}$$

Thus,

$$\begin{aligned} \Delta V_L(k) &\leq 2\|f_e(k)\|^2 + 2g_M^2\|u_e(k)\|^2 - e^T(k)e(k) \\ &\quad - \zeta(\lambda_{\min}(P)\|e(k)\|^2 + \lambda_{\min}(Q)\|u_e(k)\|^2) \\ \Delta V_L(k) &\leq 2\|f_e(k)\|^2 - (\zeta\lambda_{\min}(Q) - 2g_M^2)\|u_e(k)\|^2 \\ &\quad - \zeta\lambda_{\min}(P)\|e(k)\|^2 - \|e(k)\|^2 \end{aligned} \tag{23}$$

Considering the goal of the tracking error system (13) being asymptotically stable $\Delta V_L(k) < 0$, we require

$$\begin{aligned} \|f_e(k)\|^2 \leq \frac{1}{2}(\zeta\lambda_{\min}(P)\|e(k)\|^2 \\ +(\zeta\lambda_{\min}(Q) - 2g_M^2)\|u_e(k)\|^2) \end{aligned} \tag{24}$$

Therefore, if the bound in (24) is true, we can get $\Delta V_L(k) < 0$, implying the asymptotic stability of (13). ■

Remark 1: Theorem 1 shows that if $f_e(k)$ is bounded satisfying (16), then, for the error system (13), there must exist an admissible control $u_e(k)$ not only stabilizes the system (13) but also guarantees (14) is finite.

In this sense it can be said that the optimal tracking control problem of (7) is transformed into an optimal regulation problem of (13). Noted that cost-to-go can be rewritten in a recursive form which is as follow:

$$\begin{aligned} J(e(k)) &= e(i)^T Pe(i) + u_e(i)^T Qu_e(i) \\ &\quad + \sum_{i=k+1}^{\infty} \beta^{i-k-1} U(e(i), u_e(i)) \\ &= e(i)^T Pe(i) + u_e(i)^T Qu_e(i) + \beta J(e(k+1)) \end{aligned} \tag{25}$$

According to Bellman's optimality principle, the optimal cost-to-go $J^*(e(k))$ satisfies the discrete-time HJB equation

$$J^*(e(k)) = \min_{u_e} \left\{ e(k)^T Pe(k) + u_e(k)^T Qu_e(k) + \beta J^*(e(k+1)) \right\} \tag{26}$$

It is observed that the optimal cost-to-go $J^*(e(k))$ is obtained by an optimal regulation control $u_e{}^*(k)$, which is able to minimize $J(e(k))$ as $J^*(e(k))$. Then the optimal regulation control $u_e{}^*(k)$ is formulated as

$$\begin{aligned} u_e{}^*(k) &= \arg \min_{u_e(k)} \left\{ e(k)^T Pe(k) + u_e(k)^T Qu_e(k) \right. \\ &\quad \left. + \beta J^*(e(k+1)) \right\} \\ &= -\frac{\beta}{2} g^T(e(k) + y_r(k)) \frac{\partial J^*(e(k+1))}{\partial e(k+1)} \end{aligned} \tag{27}$$

where $u_e{}^*(k)$ is the optimal regulation control associated to the optimal cost-to-go $J^*(e(k+1))$. However, it is impossible to obtain $e(k+1)$ as well as $J^*(e(k+1))$ at the current time step. Thus, the approximate optimal solution of HJB equation is perused instead of the analytical solution and HDP is employed to approximately solve this kind of optimal control problems.

HDP successively adapts two neural networks to determine the optimal cost-to-go and control laws, namely, an action network (which dispenses the regulation control signals) and a critic network (which approximates the optimal cost-to-go). The adaptation process starts with an arbitrarily chosen control by the action network; the critic network then guides the action network toward the optimal solution at each time step. In HDP, action-critic connections are mediated by a model network approximating dynamics of the plant. In this case, the DRNI is used as model network.

IV. DRNC-ACD CONSTRUCTION FOR TEMPERATURE TRACKING CONTROL IN HPDMHS

In this part, we implement the DRNC-ACD construction for temperature tracking control in HPDMHS. In (13), the original tracking control problem has been converted to the error regulation problem. It should be noted that the desired control $u_r(k) = [u_{rp1}(k), u_{rp2}(k), u_{rp3}(k), u_{rp4}(k), u_{rp5}(k), u_{rv}(k)]$ can be obtained by setting $y(k) = y_r(k)$ and $u(k) = u_r(k)$ in the original system (7), i.e.,

$$\begin{aligned} y_r(k+1) &= F(y_r(k), u_r(k)) \\ u_r(k) &= F^{-1}(y_r(k+1), y_r(k)) \end{aligned} \tag{28}$$

where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. This inspires us to employ the conventional DRNN-based direct adaptive control to learn the inverse dynamics of the system. The DRNC is established with the aid of the trained DRNI as shown in Fig. 4, where $y_r(k)$ is defined as

$$y_r(k) = \begin{cases} Rk + y_{ini}, & 0 \le k \le \dfrac{y_{\max} - y_{ini}}{R} \\ y_{\max}, & k > \dfrac{y_{\max} - y_{ini}}{R} \end{cases} \tag{29}$$

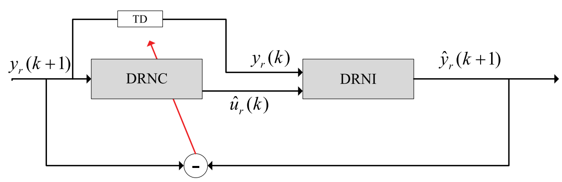

FIGURE 4. Block diagram of DRNN-based direct adaptive control.

where R denotes the temperature rising rate, y_{ini} is the initial temperature, while y_{max} denotes steady temperature. $\hat{y}_r(k+1)$ is the estimated reference trajectory, $\hat{u}_r(k)$ denotes the estimated desired control. Using the defined error function

$$\begin{aligned} e_g(k+1) &= \hat{y}_r(k+1) - y_r(k+1) \\ E_g(k+1) &= \frac{1}{2} e_g{}^T(k+1) e_g(k+1) \end{aligned} \tag{30}$$

The gradient-based adaptation rule of DRNC is same with Eq. (6). After a sufficiently long training process, it is assumed that the DRNC has learned the inverse dynamics of DRNI and desired control law $u_r(k)$ can be obtained.

Next, we employs HDP to solve the optimal regulation control law $u_e^*(k) = [u_{ep1}^*(k), u_{ep2}^*(k), u_{ep3}^*(k), u_{ep4}^*(k), u_{ep5}^*(k), u_{ev}^*(k)]$. In the HDP control scheme, there are three neural networks, which are DRNI (act as model network), critic DRNN and actor DRNN. Combining the DRNC (dispenses desired control) with HDP (generates optimal error regulation control), the structure of DRNC-ACD is shown in Fig. 5. The blue line means weight transmission, where two critic DRNNs show the time difference during the algorithmic procedure, and they actually are the same network.

Additionally, it should be noted that before implementing the HDP algorithm, the training of both DRNC and DRNI should be completed. Then, the corresponding weights are frozen.

The error function $E_c(k)$ for critic DRNN is defined as

$$\begin{aligned} E_c(k) &= \frac{1}{2} e_c(k)^T e_c(k) \\ e_c(k) &= \hat{J}(e(k)) - [U(e(k), \hat{u}_e(k)) + \hat{J}(e(k+1))] \end{aligned} \tag{31}$$

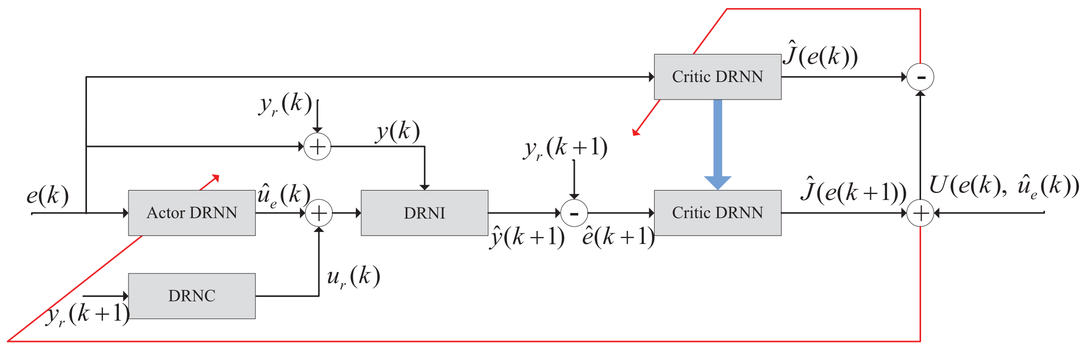

FIGURE 5. **Block diagram of DRNC-ACD control.**

The optimal cost-to-go $J^*(e(k))$ can be gradually approximated as the error function $E_c(k)$ is minimized over time. The adaption rule for critic DRNN is gradient-based backpropagation which can be seen in (6).

The target value of desired regulation control $u_e(k)$ for actor DRNN can be obtained by (27). Thus, the weights of actor DRNN are updated to minimize the following objective error function $E_a(k)$:

$$E_a(k) = \frac{1}{2} e_a(k)^T e_a(k)$$
$$e_a(k) = \hat{u}_e(k) - u_e(k) \quad (32)$$

Again, using the gradient-descend method, the actor DRNN can be trained to generate the optimal regulation control signal $u^*{}_e(k)$. The updating rule for actor DRNN is the same with critic DRNN.

Additionally, it should be noted from (27) that the control coefficient matrix $g(y(k))$ is required for computing the target value of regulation control . However, it is very difficult to know the system dynamics of HPDMHS, thus we can not obtain $g(y(k))$ directly. In this paper, we establish the system model by a DRNI which can provide a platform for estimating the value of control coefficient matrix. Reconsidering (7), for DRNI we can get

$$g(k) = \frac{\partial \hat{y}(k+1)}{\partial u(k)} \quad (33)$$

This term is the jacobian of plant output to its input, which can be directly solved through backpropagation of DRNI.

The training procedure consists of two training cycles: one for critic DRNN and other for actor DRNN as illustrated in Fig. 6. At the first time step k, the critic DRNN's training is carried out first with small initialized random value. Its incremental optimization is implemented by (31). After the training of critic DRNN becomes convergent, its weights are frozen. The training of actor DRNN is continued by (32) until the convergence of actor DRNN has been achieved.

The actor DRNN's weights are now frozen. The system operation condition is changed, where the time step k equals to $k + 1$, the critic DRNN's cycle starts again. In this way, the HDP algorithm iterates these two operations while from time to time changing the time step. At every time step, the actor DRNN's weights are updated based on the improved control law $u_e(k)$, and the critic DRNN's weights are updated based on the improved cost-to-go function $J(e(k))$. The algorithmic procedure is finished until the maximum time step has been reached and the optimal control signals are obtained over the horizon.

V. SIMULATION RESULTS

In this section, both the proposed control method and the conventional DRNN-based adaptive control method will be implemented into the HPDMHS illustrated in Sect. II. Their control performance will be tested and compared throughly.

The objective of the designed controller is to maneuver the plant so that it will follow the reference response which is given in (29), where the temperature rising rate $R = 0.125$, the initial temperature $T_{ini} = 35°C$, steady temperature $T_{max} = 60°C$. In HPDMHS, due to the physical limitations of magnetrons, the incident power for each microwave source only can be adjusted between maximum value and minimum value $0W \le u_{pi} \le 3000W (i = 1, \ldots, 5)$

To begin with, a DRNI is established by a DRNN with the structure of 7-15-1. With the initial weights being chosen randomly in $[-0.1, 0.1]$, we train the DRNI for 100 steps using 700 data samples by gradient descend algorithm. The training result is shown in Fig. 7. It is clear that after training, the temperature error is within 1°C, so the DRNI successfully learns the dynamics of the HPDMHS. Next, both controllers are implemented on the DRNI.

A. CONVENTIONAL DRNN-BASED ADAPTIVE CONTROL

With the aid of the trained DRNI, a DRNN with structure 8-10-6 is employed to construct the DRNC in the conventional adaptive control system (as shown in Fig. 3). The learning rate decides the speed at which the parameters of DRNC will adjust during the control process. A high value of η may lead to instability and result in out-of-bound output, while a small value contributes to a slow learning process as little improvement will be made in parameters from one

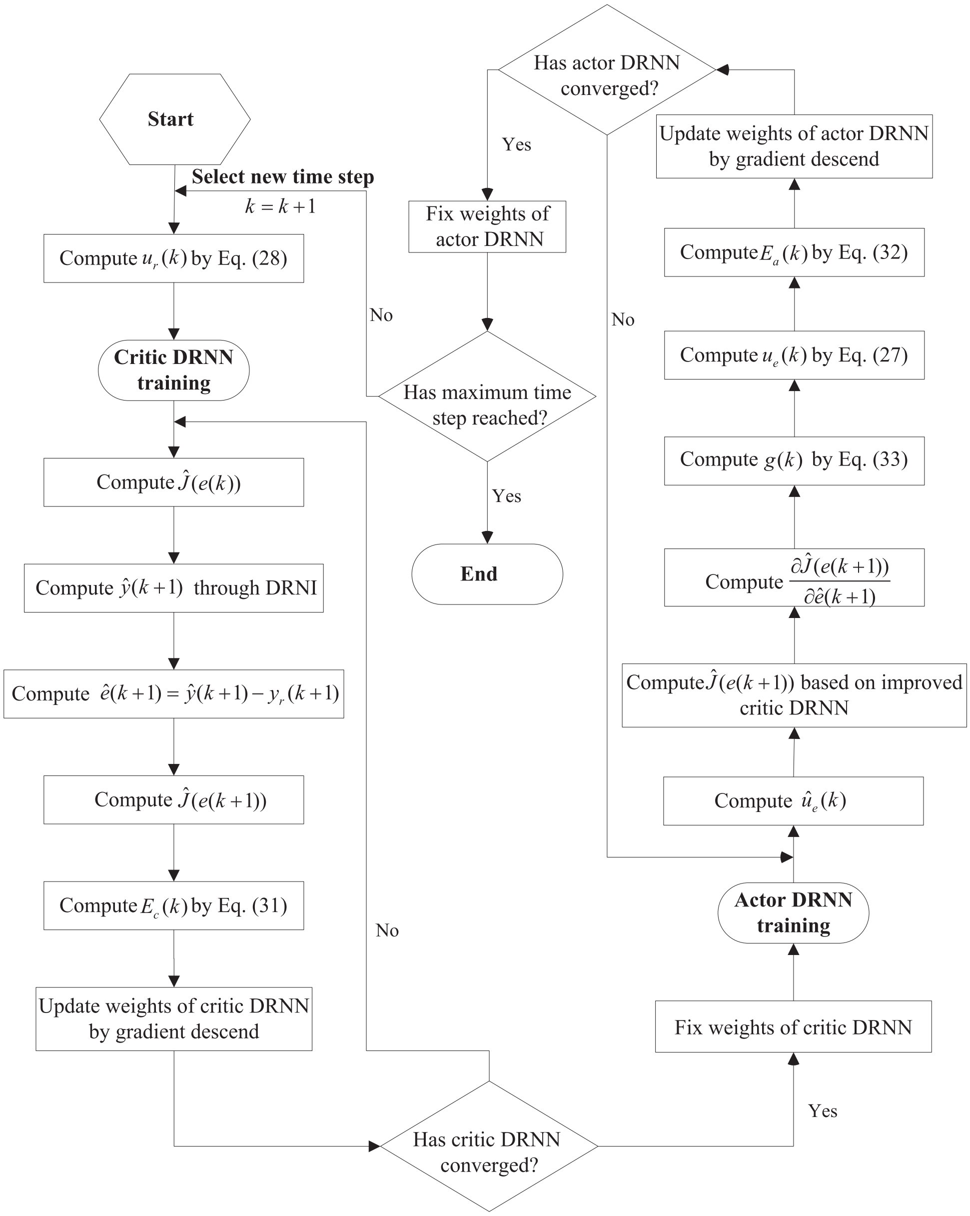

FIGURE 6. Training procedure for HDP algorithm.

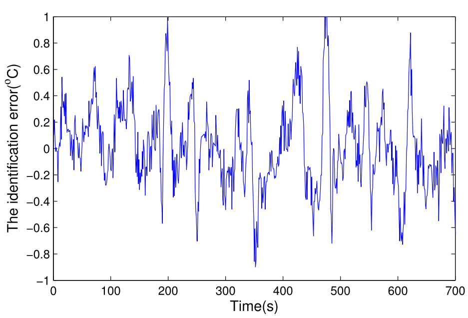

FIGURE 7. The system identification error.

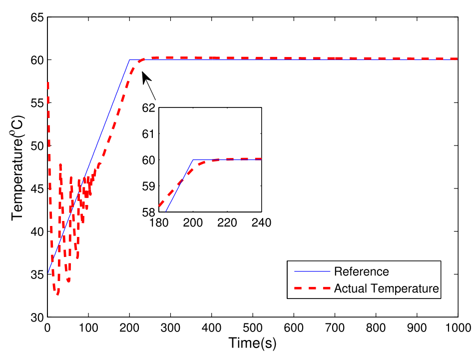

FIGURE 8. Temperature tracking performance based on DRNN-based adaptive control algorithm.

iteration to the next. In our study, an adaptive learning rate is chosen where its value decreases along the time step $\eta = a \times e^{(-b \times k)}$, where a, b are constant, k denotes the current time step. Besides, a momentum term α is added in the update equation to increase the rate of learning and yet avoiding the danger of instability. In simulation, parameter a is taken to be 0.8, b is 0.001 and momentum term α is set to 0.005.

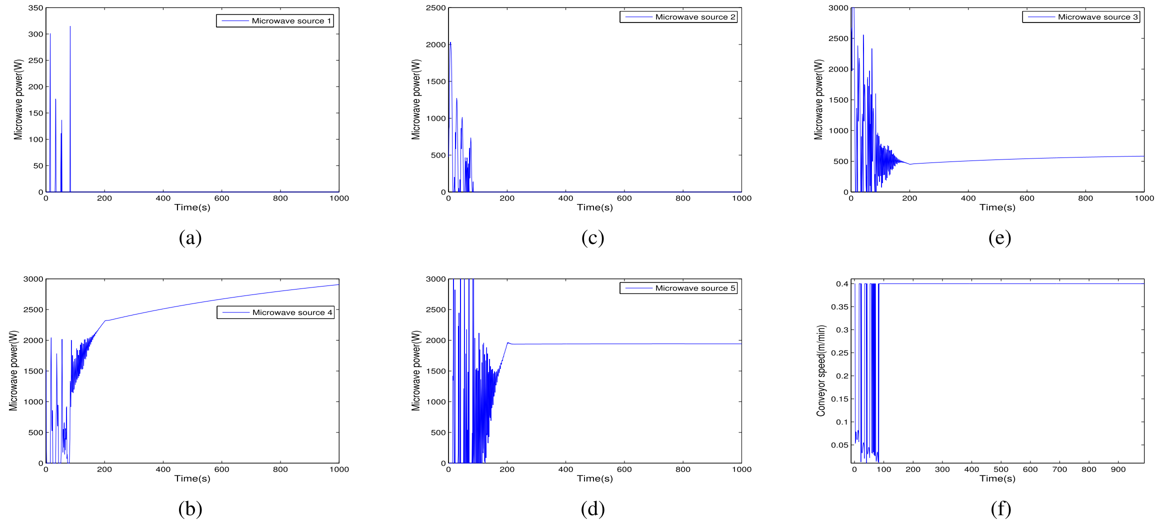

FIGURE 9. System inputs variations based on adaptive control. (a) Mcrowave power 1. (b) Mcrowave power 4. (c) Mcrowave power 2. (d) Mcrowave power 5. (e) Mcrowave power 3. (f) Conveyor speed.

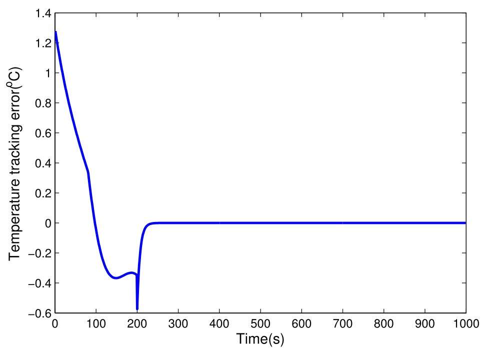

FIGURE 10. Training error of DRNC.

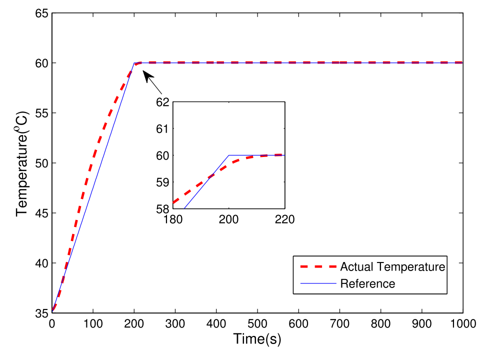

FIGURE 11. Temperature tracking performance based on DRNC-ACD control algorithm.

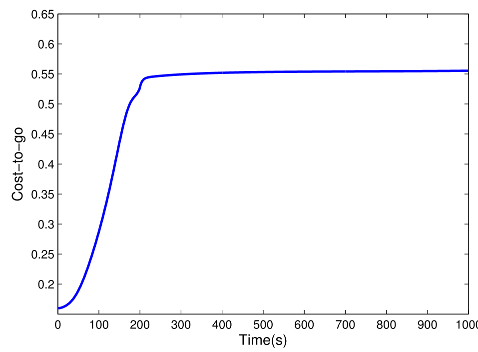

FIGURE 12. Cost-to-go of critic DRNN.

The initial weights of DRNC is chosen close to zeros which is set to $[-0.1, 0.1]$.

The temperature tracking result based on conventional DRNN-based adaptive control is shown is Fig. 8. A significant oscillation can be seen at the temperature rising stage. The reason for this is that arbitrary weights are selected at the beginning of the training, which leads to undesired tracking performance. The weights are then adjusted at each iteration and gradually trend to be convergent. Although temperature can keep steady at 60°C, its rising stage can hardly tack the reference trajectory. The system inputs u_{p1}, u_{p2}, u_{p3}, u_{p4}, u_{p5} and u_v are shown with the phenomenon of rapid switching action in Fig. 9. Because of the constrained system inputs, the switching control phenomenon is shown during the microwave heating system. This rapid regulation can hardly be achieved in a real system, as it will destroy the magnetron and motor driver.

B. DRNC-ACD CONTROL

To improve the control performance, the ACD is employed to combine with DRNC. Firstly, with the aid of DRNI, a DRNC with structure 1-8-6 is trained to generate the desired control $u_r(k)$ using the distal supervised learning approach (as shown in Fig. 4). The learning rate and momentum term are set to be $\eta = 0.1 \times e^{(-0.001 \times k)}$ and $\alpha = 0.001$, respectively. The initial weights of DRNC is chosen randomly in $[-0.1, 0.1]$. We train the DRNC using 1000 data samples, the error between reference and system output is shown in Fig. 10,

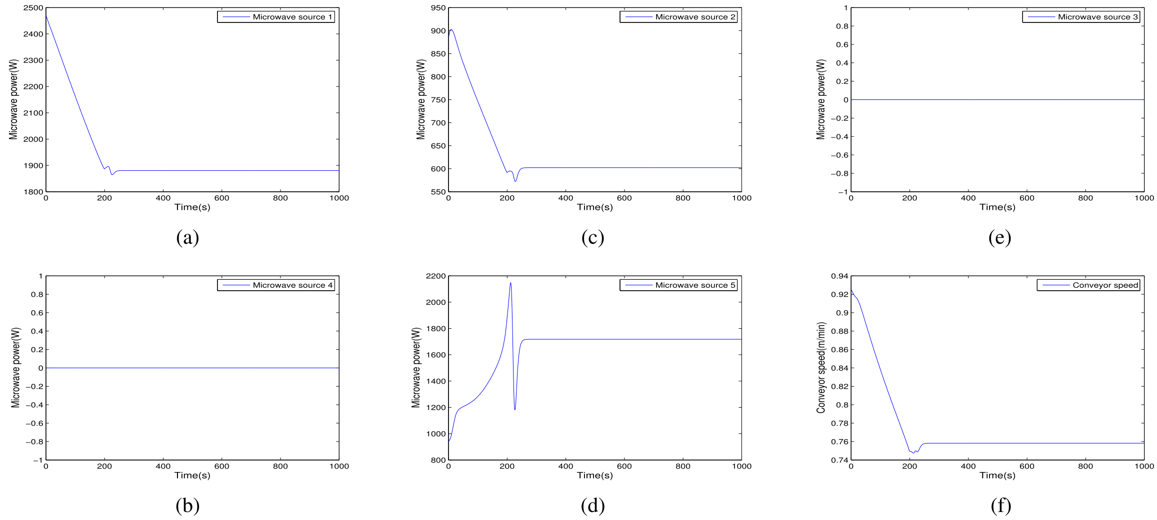

FIGURE 13. System inputs variations based on DRNC-ACD control. (a) Mcrowave power 1. (b) Mcrowave power 4. (c) Mcrowave power 2. (d) Mcrowave power 5. (e) Mcrowave power 3. (f) Conveyor speed.

which is visible that the DRNC has successfully learn the inverse dynamics of the system.

After finishing the training of DRNC, their weights are kept unchanged and HDP algorithm gets start. Two DRNNs are used to construct the actor DRNN and critic DRNN with structures 1-8-6 and 1-4-1, respectively. The initial weights of critic and actor DRNN are all set to random in $[-0.1, 0.1]$. The learning rates and momentum factors for actor DRNN and critic DRNN are both $\eta = 0.1 \times e^{(-0.05\times k)}$ and $\alpha = 0.001$. Using the gradient descent algorithm, we online implement the HDP for 1000 time steps with each time step containing 200 inner-loop training steps for both critic and actor DRNN.

Fig. 11 demonstrates the outperformance achieved by DRNC-ACD strategy, as the tracking performance in temperature rising stage has been significantly improved. The system can track the reference trajectory $y_r(k)$ in an optimal manner. From the view of control system design, the rising time of response curve in Fig. 11 shows an approximate 20% reduction compared with that of in Fig. 8, which again provides the evidence of superior efficiency of DRNC-ACD method over conventional DRNN-based adaptive control method. The curve of cost-to-go is shown in Fig. 12, where it becomes convergent within 200 steps. Fig. 13 presents the real time regulation of five microwave powers and conveyor speed, which is obvious that the switching rate of system inputs is significantly reduced. This is beneficial to magnetron and motor driver.

VI. CONCLUSION

This paper couples the classical DRNC with ACD technique to solve the temperature tracking control problem in microwave heating system. Firstly, the original tracking control problem is transformed into an error regulation problem with theoretical analysis. The DRNC-ACD construction is then designed with four neural networks, namely, DRNC, DRNI, critic DRNN and actor DRNN. During online learning process, DRNC is used to generate the desired control $u_r(k)$, while the critic DRNN evaluates the effect of the regulation control policy $u_e(k)$ by estimating a cost-to-go, and then the actor DRNN adjusts the current policy to minimize the cost estimated by the critic. As this process continues repeatedly, the optimal control policy is approximated gradually. Results show that the proposed controller can achieve a higher tracking control precision with less inputs oscillation than that of in conventional DRNN-based adaptive control method.

Despite the advantages presented by the new design, the proposed controller when equipped with gradient descend algorithm can be prone to stuck at the local minimal, especially when the operational condition is changed significantly. Further works will focus on improving the stability of controller, a more powerful and robust online learning technique for DRNC-ACD will be investigated.

REFERENCES

[1] S. Chandrasekaran, S. Ramanathan, and T. Basak, "Microwave material processing—A review," *AiChE J.*, vol. 58, no. 2, pp. 330–363, 2012.

[2] E. Akkari, S. Chevallier, and L. Boillereaux, "Global linearizing control of MIMO microwave-assisted thawing," *Control Eng. Pract.*, vol. 17, no. 1, pp. 39–47, 2009.

[3] J. Zhong, L. Shan, and Q. Xiong, "Receding horizon H_∞ guaranteed cost tracking control for microwave heating medium with temperature-dependent permittivity," *ISA Trans.*, vol. 73, no. 1, pp. 249–256, Feb. 2018.

[4] Y. Yuan, S. Liang, Q. Xiong, J. Zhong, and Z. Wang, "Thermodynamics model based temperature tracking control in microwave heating," *J. Thermal Sci. Technol.*, vol. 11, no. 1, pp. 4–14, 2016.

[5] K. Wang *et al.*, "Learning to detect local overheating of the high-power microwave heating process with deep learning," *IEEE Access*, vol. 6, pp. 10288–10296, 2018.

[6] C. Yin, S. Dadras, X. Huang, Y. Cheng, and H. Malek, "The design and performance analysis of multivariate fractional-order gradient-based extremum seeking approach," *Appl. Math. Model.*, vol. 62, pp. 680–700, Oct. 2018.

[7] C. Yin, S. Dadras, X. Huang, J. Mei, H. Malek, and Y. Cheng, "Energy-saving control strategy for lighting system based on multivariate extremum seeking with Newton algorithm," *Energy Convers. Manage.*, vol. 142, pp. 504–522, Jun. 2017.

[8] C. Yin, X. Huang, Y. Chen, S. Dadras, S.-M. Zhong, and Y. Cheng, "Fractional-order exponential switching technique to enhance sliding mode control," *Appl. Math. Model.*, vol. 44, pp. 705–726, Apr. 2017.

[9] W. He, T. Meng, X. He, and S. S. Ge, "Unified iterative learning control for flexible structures with input constraints," *Automatica*, vol. 86, pp. 326–336, Oct. 2018.

[10] S. Zhang, Y. Dong, Y. Ouyang, Z. Yin, and K. Peng, "Adaptive neural control for robotic manipulators with output constraints and uncertainties," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5554–5564, Nov. 2018.

[11] W. He, Y. Ouyang, and J. Hong, "Vibration control of a flexible robotic manipulator in the presence of input deadzone," *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 48–59, Feb. 2017.

[12] H. Gao, W. He, C. Zhou, and S. Changyin, "Neural network control of a two-link flexible robotic manipulator using assumed mode method," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2018.2818120.

[13] R. Kumar, S. Srivastava, J. R. P. Gupta, and A. Mohindru, "Diagonal recurrent neural network based identification of nonlinear dynamical systems with Lyapunov stability based adaptive learning rates," *Neurocomputing*, vol. 287, no. 26, pp. 102–117, Apr. 2018.

[14] C.-C. Ku and K. Y. Lee, "Diagonal recurrent neural networks for dynamic systems control," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 144–156, Jan. 1995.

[15] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Mar. 1990.

[16] C.-C. Ku, K. Y. Lee, and R. M. Edwards, "Improved nuclear reactor temperature control using diagonal recurrent neural networks," *IEEE Trans. Nucl. Sci.*, vol. 39, no. 6, pp. 2298–2308, Dec. 1992.

[17] F. Gao, F. Wang, and M. Li, "A simple nonlinear controller with diagonal recurrent neural network," *Chem. Eng. Sci.*, vol. 55, no. 7, pp. 1283–1288, Apr. 2000.

[18] R. Zhu, T. Chai, and C. Shao, "Robust nonlinear adaptive observer design using dynamic recurrent neural networks," in *Proc. Amer. Control Conf.*, Albuquerque, NM, USA, Jun. 1997, pp. 1096–1100.

[19] T. A. Condarcure and M. K. Sundareshan, "A learning automaton approach to trajectory learning and control system design using dynamic recurrent neural networks," in *Proc. IEEE Int. Conf. Neural Netw.*, Orlando, FL, USA, Jun./Jul. 1993, pp. 2684–2689.

[20] T. W. S. Chow and Y. Fang, "A recurrent neural-network-based real-time learning control strategy applying to nonlinear systems with unknown dynamics," *IEEE Trans. Ind. Electron.*, vol. 45, no. 1, pp. 151–161, Feb. 1998.

[21] C.-H. Lee and C.-C. Teng, "Identification and control of dynamic systems using recurrent fuzzy neural networks," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 4, pp. 349–366, Aug. 2000.

[22] X. Wang and G. Peng, "Modeling and control for pneumatic manipulator based on dynamic neural network," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Washington, DC, USA, Oct. 2003, pp. 2684–2689.

[23] C.-Y. Liu, K. Li, Z.-Y. Song, X.-L. Song, and H.-F. Wang, "Variable flow heating control system based on neural network decoupling," in *Proc. IEEE Int. Conf. Grey Syst. Intell. Services*, Nanjing, China, Nov. 2007, pp. 860–864.

[24] T. Aboueldahab and M. Fakhreldin, "Adaptive control of dynamic nonlinear systems using Sigmoid Diagonal Recurrent Neural Network," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Istanbul, Turkey, Oct. 2010, pp. 4341–4345.

[25] M. Amiri, M. B. Menhaj, and M. J. Yazdanpanh, "A neural-network-based controller for a single-link flexible manipulator: Comparison of FFNN and DRNN controllers," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Hong Kong, Jun. 2008, pp. 1686–1691.

[26] L. Wang, Z. Meng, Y. Sun, L. Guo, and M. Zhou, "Design and analysis of a novel chaotic diagonal recurrent neural network," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 26, nos. 1–3, pp. 11–23, 2015.

[27] R. Kumar, S. Srivastava, and J. R. P. Gupta, "Diagonal recurrent neural network based adaptive control of nonlinear dynamical systems using Lyapunov stability criterion," *ISA Trans.*, vol. 67, pp. 407–427, Mar. 2017.

[28] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 997–1005, 1997.

[29] C. Qin, H. Zhang, and Y. Luo, "Optimal tracking control of a class of nonlinear discrete-time switched systems using adaptive dynamic programming," *Neural Comput. Appl.*, vol. 24, nos. 3–4, pp. 531–538, 2014.

[30] Y. Huang and D. Liu, "Neural-network-based optimal tracking control scheme for a class of unknown discrete-time nonlinear systems using iterative ADP algorithm," *Neurocomputing*, vol. 125, no. 3, pp. 46–56, Feb. 2014.

[31] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32–50, Aug. 2009.

[32] P. J. Werbos, "Advanced forecasting methods for global crisis warning and models of intelligence," *General Syst. Yearbook*, vol. 22, pp. 25–38, 1977.

[33] J. Si and Y.-T. Wang, "Online learning control by association and reinforcement," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 264–276, 2001.

[34] C. Mu, D. Wang, and H. He, "Novel iterative neural dynamic programming for data-based approximate optimal control design," *Automatica*, vol. 81, pp. 240–252, Jul. 2017.

[35] H. Zhang, D. Yue, and X. Xie, "Distributed model predictive control for hybrid energy resource system with large-scale decomposition coordination approach," *IEEE Access*, vol. 4, pp. 9332–9344, 2017.

TONG LIU received the B.Sc. degree in automation from the College of Automation, Chongqing University, Chongqing, China, in 2016, where he is currently pursuing the Ph.D. degree in control theory and control engineering. Since 2018, he has been a Visiting Ph.D. Student with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K., under the supervision of Prof. Chen. His research interests include nonlinear system identification, neural networks, and adaptive dynamic programming with their applications on microwave heating systems.

SHAN LIANG (M'08) received the M.Sc. degree in control science and engineering from the College of Automation, Chongqing University, Chongqing, China, in 1995, and the Ph.D. degree from the Department of Mechanical Systems Engineering, Kumamoto University, Kumamoto, Japan, in 2004. His current research interests include numerical modeling, electromagnetic theory, nonlinear systems, adaptive control, and sensor networks. He is currently a member of the IEEE Control Systems and the Society of Instrument and Control Engineers, Japan. He is also a Co-Editor of the *International Journal of Sensing, Computing and Control*.

QINGYU XIONG received B.Sc. and M.Sc. degrees from Chongqing University in 1986 and 1991, respectively, and the Ph.D. degree from Kyushu University, Japan, in 2002. He is currently a Professor and a Supervisor for Ph.D. student with Chongqing University. His main research directions include intelligent system and intelligent computation, pervasive computation, and embedded system.

KAI WANG received the B.S. degree in electronics and information engineering from Sichuan University, Chengdu, China, in 2000, and the Ph.D. degree in control theory and control engineering from Chongqing University, Chongqing, China, in 2009. He was a Visiting Scholar with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA, in 2014. He is currently an Associate Professor with the Department of Automation, Chongqing University. His current interests include machine learning, deep learning, and the application of big data.