IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 June 2024, accepted 22 July 2024, date of publication 29 July 2024, date of current version 14 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3434994

RESEARCH ARTICLE

A HDP Online Energy Management Strategy for Series Hybrid Loaders in a Model Predictive Control Framework

JICHAO LIU[1], YANYAN LIANG[1], KA XUE[1], AND ZHENG CHEN[2]
[1]New Energy Research Institute, Jiangsu XCMG Research Institute Company Ltd., Xuzhou 221000, China
[2]School of Materials and Physics, China University of Mining and Technology, Xuzhou 221116, China

Corresponding author: Jichao Liu (m18501046354@163.com)

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China under Grant 62103415.

ABSTRACT A model predictive control-based heuristic dynamic programming (HDP-MPC) online strategy is proposed in this paper, to further improve the fuel economy for the series hybrid loaders. The back propagation neural network (BPNN) is utilized to establish the dynamics model for series hybrid loader, to effectively characterize its real working process. Next, a HDP-MPC approximate optimal algorithm is proposed. And its convergence and the controlled system stability are separately proved. Besides, an online energy management strategy based on HDP-MPC is designed to implement the energy management controller. Finally, the BPNN dynamics model and online HDP-MPC strategy are simulated and validated. Through the experimental results analysis, contrasted with the analytic model, the established dynamics model can more effectively reflect the actual motion process of the series hybrid loader with higher accuracy. Moreover, compared with the fuzzy rule-based and adaptive equivalent consumption minimization strategy strategies, the designed HDP-MPC strategy can respectively achieve fuel-savings of 28.65% and 10.38%, supplying a new idea for online energy management of the series hybrid loaders.

INDEX TERMS Energy management, fuel optimal control, neural networks, predictive control.

I. INTRODUCTION

Currently, driven by the national dual carbon policy of China [1], [2], series hybrid loaders (SHLs) have already become one of the significant approaches to achieve green sustainable development for loaders [3], [4], [5]. For fully exploiting the potential of energy saving and emission reduction [2], [6] of SHLs, it is extremely necessary to choose a reasonable energy management strategy (EMS) to achieve the power optimization distribution of multiple energy sources for SHLs [7].

For new energy vehicles, many scholars have proposed energy optimization methods represented by rule-based (RB), offline global optimization control, and instantaneous energy consumption control strategies [8], [9]. Among them, the RB strategies [10] are simple and easy to implement, but they excessively depend on expert experience, limiting their fuel-saving robustness. Although the offline global optimization strategies can utilize the prior knowledge to achieve a global optimal solution for fuel consumption, they are usually used as a benchmark for evaluating other strategies due to the heavy computational burden and the whole driving condition requirements [11]. Thus, for realizing the real-time optimization of energy consumption, the instantaneous optimization strategies represented by equivalent consumption minimization strategy (ECMS) [12] were gradually proposed. However, different from on-road vehicles, SHLs always work in low speed and high torque conditions, whose working conditions and environment are complex and changeable, putting forward higher requirements on the robustness of optimization methods. This makes the existing energy consumption methods difficult to directly apply to SHLs.

In recent years, many researchers have presented model predictive control (MPC) and intelligent control [13],

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Valencia-Palomo.

[14], [15], [16] optimization methods, providing a new research idea for the highly nonlinear energy consumption optimization problem of the SHLs. Wherein, the optimization goal of the MPC strategies is to achieve the local optimal control by minimizing the local energy consumption of the prediction horizon [17]. In this process, firstly, the driving condition information of the prediction horizon needs to be predicted for MPC EMSs. Secondly, on the basis of the predicted information, a rational optimization algorithm [18] is selected to solve the optimal control sequence in the prediction time domain. In [19], a MPC EMS based on Markov speed prediction model was developed, which utilized dynamic programming (DP) algorithm to calculate the optimal power allocation in the prediction time domain, achieving the fuel-savings of 29.7% compared to the RB strategy. Xie et al. [20] designed a stochastic MPC EMS based on the Pontryagin minimum principle (PMP), also using Markov chains to achieve online prediction of vehicle speed. Meanwhile, the energy optimization effects of the vehicle in different prediction domains were compared and analyzed. Aiming at the fuel economy problem of hybrid crawler bulldozer, Wang et al. [21] presented a super capacitor energy state tracking MPC EMS based on Markov prediction. Its fuel economy under typical working conditions was nearly 6 % higher than that of the RB strategy. In [22], a MPC EMS based on power demand Markov model was given for fuel cell hybrid loaders, verifying the effectiveness of using Markov model to predict energy demand under comprehensive operating conditions. However, as the vehicle moves forward, the state transition probability matrices of the Markov chain increase rapidly, leading to a heavy calculation burden and poor real-time performance of the strategy. For this reason, the MPC strategies based on neutral network (NN) were presented. Xiang et al. [23] introduced a dual-mode power split EMS based on vehicle speed prediction and nonlinear model predictive control, using radial basis NN to realize dynamic prediction of vehicle speed. In [24], integrating the Markov model and back propagation NN (BPNN), a speed prediction method was proposed. Then, a MPC EMS was designed, realizing a better fuel-saving effect than ECMS under the same working conditions. In [25], a MPC adaptive control strategy based on deep NN was developed. Although this strategy consumed 7.79% more fuel than the DP strategy, it saved fuel about 20% compared with the RB strategy. However, during the actual working environment, the sudden change of working conditions of the SHLs can affect the model accuracy, reducing the fuel-saving effects of the MPC strategies.

Different from MPC EMSs, the intelligent control strategies are to realize energy consumption optimization control by the idea of reinforcement learning (RL). In this process, the optimized effect of the strategy is continuously regulated and makes it constantly approach the optimal solution of fuel consumption problem. Yang et al. [26] proposed a real-time intelligent EMS based on RL and MPC. Compared with Q-learning and traditional MPC, it significantly strengthened the real-time performance of the strategy and decreased the fuel consumption. Considering the global optimization of battery SoC and operating mode in a fixed path, Wang et al. [27] designed the adaptive optimal intelligent control and deep Q-learning intelligent control strategies, respectively. Finally, the fuel economy and real-time performance of them were verified and analyzed comparatively through hardware in loop (HIL) experiment. Integrating engine start-stop and shifting control, Tang et al. [28] applied a double deep RL algorithm to develop a novel EMS. In this process, for realizing the collaborative learning of power allocation and shift strategies, a deep Q-network and a deep deterministic policy gradient were used to control the gear-shifting and the engine throttle opening, respectively. In [29], a deep RL-based multi-objective control strategy was presented, achieving the fine adjustment of motor speed, engine power control, and transmission ratio control.

In addition, as a typical representative of intelligent control strategies, the heuristic DP (HDP) combines the ideas of RL and DP, which can realize the approximate optimal control of the controlled object by the mechanism-less model. Li and Görges [30], [31] presented a self-adaptive loop control algorithm based on execution dependent HDP. Moreover, a self-adaptive EMS was designed, achieving energy optimization effects that are very close to DP strategy while ensuring comfort. In [32], for improving the driving comfort and energy economy of the vehicle, a double-layer EMS based on adaptive DP was presented in car-following scenarios, including the speed prediction layer based on HDP and the energy management control layer based on dual heuristic dynamic programming (DHP). To minimize the equivalent fuel consumption further, Wang and Jiao [33] adopted the DHP algorithm to design an online EMS, improving the fuel economy and the adaptability to different driving cycles. Compared with double deep Q-network and HDP EMSs, the DHP strategy can save fuel by 9.58% and 6.06% under the actual driving cycle, respectively. Nevertheless, the optimization effect of intelligent control strategies excessively relies on the integrality of sample data and the learning approaches of parameters, confining their real-time application. Besides, the series hybrid loaders belong to the electromechanical-hydraulic strong coupling system, whose complex and changeable working conditions can easily lead to mismatched fitting data in intelligent control strategy, reducing the applicability of this method.

In summary, the existing optimization methods do not solve the system model mismatch and state instability caused by the time-varying internal parameters of the complicated working conditions of the SHLs. Therefore, taking a SHL as research object, this paper designs a novel online EMS by combining the advantages of the approximate optimal control of HDP and the rolling optimization of MPC.

To realize the above goals, the main contributions of this paper are as follows:

1) Using BPNN to build the dynamics model of SHL and establishing its fuel optimization problem;

2) Designing the HDP algorithm based on model predictive control, proving its convergence and the controlled system stability, and proposing a MPC-based HDP (HDP-MPC) EMS for SHLs;

3) Executing the simulation comparison analysis for the presented HDP-MPC strategy based on an actual working condition.

The rest of the paper is arranged as follows. The dynamics model of SHL and the fuel consumption problem of rolling optimization are introduced in Section II. Section III designs the HDP-MPC algorithm and proves its convergence and the system stability. Then, the MPC-based HDP online EMS is proposed to implement the energy management controller. The comparative analysis of the designed strategy is described in Section IV. Finally, the conclusions and prospects are shown in Section V.

II. DYNAMICS MODEL AND FUEL OPTIMIZATION PROBLEM

In this section, the powertrain structure of the SHL is described in Fig. 1, and the features maps of the engine and motor are displayed in Fig. 2. Besides, the dynamics model and the fuel optimization problem of the SHL are introduced, respectively.

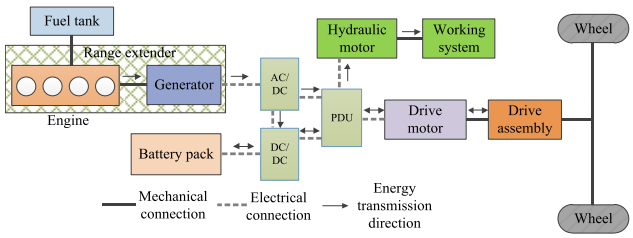

FIGURE 1. The powertrain topology of the SHL.

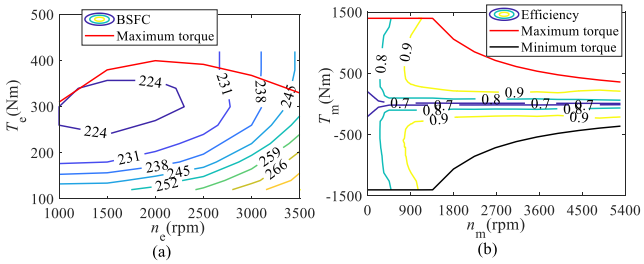

FIGURE 2. The mapping features. (a) The engine map. (b) The motor map.

A. DYNAMICS MODEL OF SHL

Under not considering the impact of complex and variable environments on its motion state, the dynamics model of SHL in the discrete domain is usually represented as

$$\begin{aligned} &x(k+1) \\ &= f(x(k), u(k), k), k \in [k_0, k_f] \\ &\Rightarrow \begin{cases} v(k+1) = v(k) + \dfrac{P_{wh}(k) \cdot 1000}{v(k) \cdot m_V} \\ -\dfrac{1}{2 \cdot m_V} \cdot C_a \cdot \rho_a \cdot A \cdot v(k)^2 \\ -g \cdot (\rho_r \cdot \cos\theta(k) + \sin\theta(k)) \\ n_m(k+1) = n_m(k) + \dfrac{30 \cdot (v(k+1) - v(k)) \cdot i_f}{\pi \cdot r_{wh}} \\ \mathrm{SoC}(k+1) = \mathrm{SoC}(k) - \dfrac{V_b - \sqrt{V_b^2 - 4 \cdot P_b(k) \cdot R_b}}{2 \cdot R_b \cdot C_{Ah}} \end{cases} \end{aligned} \tag{1}$$

where the step of discrete time k is assumed to be 1(s), k_0 and k_f are the start and end moments of the working conditions, respectively. x and u respectively denote the vectors of the state and control. m_v, g, θ, ρ_r, C_D, A, ρ_a, P_{wh},and v express the vehicle mass, gravity acceleration, ramp angle, coefficient of rolling resistance, coefficient of air resistance, frontal area, air density, wheel power, and vehicle velocity, respectively. n_m, i_f, and r_{wh} illustrate the motor speed, total gear ratio of assembly, and wheel radius, respectively. V_b, R_b, and C_{Ah} represent the terminal voltage, inherent resistance, and capacity of the battery pack, respectively.

The total demand power P_d is derived from P_{wh}, which is offered by the engine power P_e and battery power P_b according to the below expression:

$$P_d(k) = P_{wh}(k)/(\eta_m \cdot \eta_t) = P_b(k) + P_e(k) \cdot \eta_g \tag{2}$$

where η_g, η_m,and η_t are the generator efficiency, motor efficiency, and the transmission efficiency, respectively. Because n_m can be calculated by v, the state and control vectors can be represented by $x = [v, \mathrm{SoC}]^T$ and $u = [P_e, P_b]^T$.

FIGURE 3. The working scenario.

For SHL under the complex and actual working conditions, its dynamic process has obvious uncertainty. Therefore, it is very difficult for the analytical model (1) to effectively characterize its actual motion process. Here, the digging stones working condition of a SHL is selected as an example to describe this situation, and the working scenario is shown in Fig. 3. Accordingly, the detailed dynamic data information of 70 cycles is described in Fig. 4, including the vehicle speed, the engine power, the battery power, and SoC. It can be clearly seen that the change process of vehicle parameters is highly

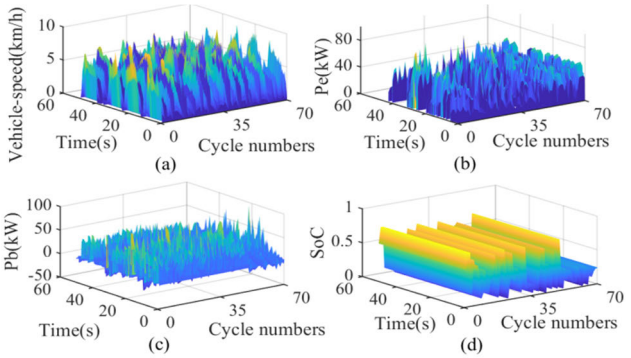

FIGURE 4. The collected information of the working data for SHL. (a) v. (b) P_e. (c) P_b. (d) SoC.

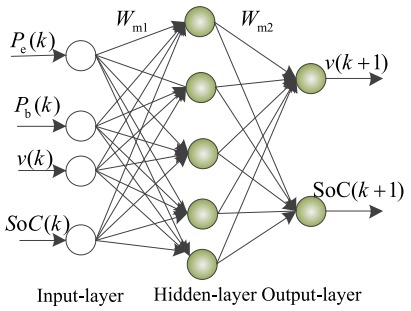

FIGURE 5. The BPNN framework of the vehicle dynamics model.

nonlinear in the same working cycle. In addition, in different working cycles, the change process of the same parameter is diverse. In other words, the dynamic process of the SHL is a highly nonlinear system with strong uncertainty. BPNN has strong fitting ability for the input-output relationships of nonlinear systems with high uncertainty [34]. Therefore, the dynamics model of the SHL is established by using the BPNN. Based on the (3) in [35], $x = [v, \mathrm{SoC}]^{\mathrm{T}}$,and $u = [P_e, P_b]^{\mathrm{T}}$, the BPNN structure is selected as 4-5-2, as shown in Fig. 5. Here, $\boldsymbol{W}_{m1}$ and $\boldsymbol{W}_{m2}$ represent the weight matrixes of the input-to-hidden layer and the hidden-to-output layer, respectively. Namely,

$$\mathbf{W}_{m1} = \begin{bmatrix} W_{m1\_11} & \cdots & W_{m1\_14} \\ \vdots & \ddots & \vdots \\ W_{m1\_51} & \cdots & W_{m1\_54} \end{bmatrix},$$
$$\mathbf{W}_{m2} = \begin{bmatrix} W_{m2\_11} & \cdots & W_{m2\_15} \\ W_{m2\_21} & \cdots & W_{m2\_25} \end{bmatrix}. \tag{3}$$

Moreover, the activation functions of the hidden-layer and the output-layer are selected as tansig and purelin functions, respectively. Then, the BPNN-based dynamics model of SHL can be expressed as

$$x(k+1) = f(x(k),u(k),k), k \in [k_0, k_f]$$
$$\Rightarrow \begin{cases} \boldsymbol{m}_{h1}(k) = \boldsymbol{W}_{m1}(k) \times [u(k)^{\mathrm{T}}, x(k)^{\mathrm{T}}]^{\mathrm{T}} \\ \boldsymbol{m}_{h2}(k) = \dfrac{1 - e^{-\boldsymbol{m}_{h1}(k)}}{1 + e^{-\boldsymbol{m}_{h1}(k)}} \\ \boldsymbol{x}(k+1) = \boldsymbol{W}_{m2}(k) \times \boldsymbol{m}_{h2}(k) \end{cases} \tag{4}$$

After determining the structure and activation functions of the BPNN model, the collected data can be used to train $\boldsymbol{W}_{m1}$ and $\boldsymbol{W}_{m2}$ until the model error E_m meets the following requirements.

$$E_m = \frac{1}{2}\{[\tilde{v}(k) - v(k)]^2 + [\tilde{\mathrm{SoC}}(k) - \mathrm{SoC}(k)]^2\} \le \varepsilon_m \tag{5}$$

where $[\tilde{v}, \tilde{\mathrm{SoC}}]^{\mathrm{T}}$ denotes the actual state value, and ε_m expresses the set target error. The detailed training process refers to our previous work [35], which is not repeated here. To determine the final $\boldsymbol{W}_{m1}$ and $\boldsymbol{W}_{m2}$, the data in Fig. 4 and the (7) in [35] are utilized to train (4). After finishing the training process, the dynamics model of the SHL is obtained.

B. FUEL OPTIMIZATION PROBLEM OF SHL

According to the built dynamics model of the SHL, its fuel optimization problem is depicted in this subsection.

Firstly, in fact, although the condition information of the SHL is hardly to be accurately obtained in advance during each working cycle, its working route is stationary. Namely, the fuel optimization problem of the SHL is practically a finite-time optimization problem. Hence, in the discrete-time domain, its fuel optimization problem can be written as:

$$\begin{cases} J(x(k)) = \sum_{j=k}^{k_f} \gamma^{j-k} Q_f(x(j), u(j), j), k \in [k_0, k_f - 1] \\ \text{s.t.} x(k+1) = f(x(k), u(k), k) \\ v(k) \in [v_{min}, v_{max}] \\ n_m(k) \in [n_{min}, n_{max}] \\ \mathrm{SoC}(k) \in [\mathrm{SoC}_{min}, \mathrm{SoC}_{max}] \\ P_e(k) \in [P_{e\_min}, P_{e\_max}] \\ P_b(k) \in [P_{b\_min}, P_{b\_max}] \end{cases} \tag{6}$$

where $\gamma \in (0, 1]$ is a discount factor. The subscripts represent the maximum and minimum values of each variable. The transient fuel consumption Q_f can be solved by the following formula:

$$Q_f(x(k), u(k), k) = P_e(k) \cdot \dot{m}_e(T_e, n_e) \cdot \Delta t/(3600 \cdot \rho_f) \tag{7}$$

herein, $\dot{m}_e$ indicates the fuel consumption rate of the engine, and its map is characterized by Fig. 2(a). ρ_f is the fuel density. Here, for simplifying the expression, $Q_f(j)$ is used to replace $Q_f(x(j), u(j), j)$.

In terms of the power-coupling hybrid system, the speed and torque of engine can be changeable under the same P_e. While determining P_d for the SHL, it can be allocated to P_e and P_b under the constraint of (2) by different power distribution methods. Diverse combinations of P_e and P_b can generate different fuel consumption for the same v. Therefore, the optimal distribution solutions of P_e and P_b need to be determined, to realize the min(J) in (6).

However, for SHL with complex and variable working conditions, its future working information is hard to be accurately acquired beforehand. Hence, the MPC method is adopted

to solve the fuel optimization problem for SHL. Its crucial idea is that transforms the global optimization into local optimization problems to solve in every prediction window, as described in Fig. 2 in [36]. Accordingly, based on (6), the rolling optimization problem of the fuel consumption can be expressed as:

$$J(x(\tau)) = \sum_{j=\tau}^{k+N-1} \gamma^{j-\tau} Q_{\mathrm{f}}(j) + \gamma^{k+N-\tau} \cdot F(x(k+N)),$$
$$\tau \in [k, k+N-1]$$
$$= Q_{\mathrm{f}}(\tau) + \gamma \cdot J(x(\tau+1)) \tag{8}$$

where N is the prediction window size. $F(x(k+N)) \geq 0$ is the terminal cost from $k+N$ to ∞.

Correspondingly, the optimal value function $J^*(x(\tau))$ satisfies the following expression:

$$J^*(x(\tau)) = \min_{u(x(\tau))} \left\{ Q_{\mathrm{f}}(\tau) + \gamma \cdot J^*(x(\tau+1)) \right\},$$
$$\tau \in [k, k+N-1] \tag{9}$$

The optimal control action meets:

$$u^*(x(\tau)) = \operatorname*{argmin}_{u(x(\tau))} \left\{ Q_{\mathrm{f}}(\tau) + \gamma \cdot J^*(x(\tau+1)) \right\},$$
$$\tau \in [k, k+N-1] \tag{10}$$

In this case, the solution process of (9) in every prediction window is synchronized with (10). This process can be realized by HDP algorithm. For this reason, this paper proposes a MPC-based HDP approximate optimal algorithm to solve (9), whose details are described in the next section.

III. ONLINE ENERGY MANAGEMENT STRATEGY BASED ON HDP-MPC

In this section, a novel HDP-MPC algorithm is presented, as shown in Fig. 6. Next, its convergence and the control system stability are proved, respectively. Furthermore, an online EMS based on HDP-MPC is proposed to implement the energy management controller.

A. THE FRAMEWORK OF HDP-MPC ALGORITHM

As shown in Fig. 6, firstly, the rolling prediction idea of MPC is used to predict the corresponding unknown state set $X = \{x(k), \hat{x}(k+1), \ldots, \hat{x}(k+N-1)\}$ of the system current state $x(k)$ in the prediction horizon $[k, k+N-1]$. Here, $\hat{x}$ is the predicted state, and $x(\tau)$ belongs to any quantity in set X. Secondly, the HDP algorithm is used to solve (9) in the current prediction window, which includes two critic networks (CNs), one actor network (AN), and a model network (MN). Wherein, based on the predictive state $x(\tau)$, the AN and CN are utilized to evaluate the approximate value $\hat{u}(x(\tau))$ and $\hat{J}$, respectively. It is worth noting that the parameters of CN1 and CN2 with the same structure are synchronously updated. The MN is used to describe the dynamic process of the controlled subject. Specifically, the MN is the BPNN-based dynamics model established in Section II. Correspondingly, the optimization principle of the HDP-MPC algorithm is as follows (supposing the current moment is kth(s)).

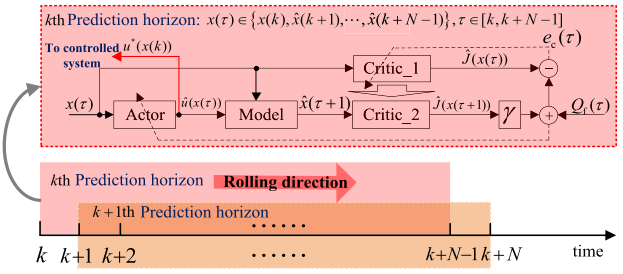

FIGURE 6. The architecture of HDP-MPC algorithm.

Firstly, in $[k, k+N-1]$, according to (9) and (10), u can be seen as the function of x, implying that $\hat{u}(x(\tau))$ can be estimated by the AN based on $x(\tau)$. Then, the dynamics model of the SHL is used to generate the next moment state $\hat{x}(\tau+1)$ based on $x(\tau)$ and $\hat{u}(x(\tau))$. Next, $\hat{J}(x(\tau))$ and $\hat{J}(x(\tau+1))$ can be estimated by CN1 and CN2 through $x(\tau)$ and $\hat{x}(\tau+1)$, respectively. Meanwhile, $Q_{\mathrm{f}}(\tau)$ can be calculated by (7).Lastly, for obtaining $J^*(x(\tau))$, based on time-step τ, the internal network parameters of AN and CN are continuously adjusted by utilizing $Q_{\mathrm{f}}(\tau) + \gamma \cdot \hat{J}(x(\tau+1))$ and $e_{\mathrm{c}}(\tau)$ until they satisfy the set conditions. Namely, in this process, the approximate optimal solution and parameter adjustment of each state in state set are executed until $\sum_{j=k}^{k+N-1} \gamma^{j-k} Q_{\mathrm{f}}(j) + \gamma^{k+N-\tau} \cdot F(x(k+N))$ of this prediction horizon reaches a minimum value. Specifically, to obtain $u^*(x(k))$, a value function J^i is defined and it assumes $J^0(x(\tau)) = 0$. According to the value iteration idea, for $i = 0, 1, \cdots$, the control action can be iteratively updated by the below expression in $\tau \in [k, k+N-1]$

$$u^i(x(\tau)) = \operatorname*{argmin}_{u(x(\tau))} \left\{ Q_{\mathrm{f}}(x(\tau), u^i(x(\tau))) + \gamma \cdot J^i(x(\tau+1)) \right\},$$
$$\tau \in [k, k+N-1] \tag{11}$$

Accordingly, the value function is iteratively updated by:

$$J^{i+1}(x(\tau)) = Q_{\mathrm{f}}(x(\tau), u^i(x(\tau))) + \gamma \cdot J^i(x(\tau+1)),$$
$$\tau \in [k, k+N-1] \tag{12}$$

Based on the above analysis, $J^*(x(\tau))$ and $u^*(x(\tau))$ can be approximated by updating (11) and (12). Meanwhile, $\hat{u}(x(k))$ is regarded as the optimal value $u^*(x(k))$ at k moment. After the system executes $u^*(x(k))$, a new state $x(k+1)$ generates, and then the above optimization process is repeated until the system stops running.

B. THE CONVERGENCE AND CONTROLLED SYSTEM STABILITY PROOF OF HDP-MPC ALGORITHM

1) THE CONVERGENCE PROOF OF HDP-MPC ALGORITHM

In this subsection, we prove that the proposed HDP-MPC algorithm can converge to the optimal solution in each prediction time domain. Therein, u^i and J^i are defined by (11) and (12), respectively.

Lemma 1: Let v^i is an arbitrary admissible control sequence, corresponding value function Λ^{i+1} is iterated by:

$$\Lambda^{i+1}(x(\tau)) = Q_f(x(\tau), v^i(x(\tau))) + \gamma \cdot \Lambda^i(x(\tau+1)), \quad \tau \in [k, k+N-1] \tag{13}$$

If $J^0(x(\tau)) = \Lambda^0(x(\tau)) = 0$, then the following inequality holds.

$$J^i(x(\tau)) \le \Lambda^i(x(\tau)), \forall i \ge 0 \tag{14}$$

Proof: First of all, as known from (11), $u^i(x(\tau))$ is got by minimizing $\{Q_f(\tau) + \gamma \cdot J^i(x(\tau+1))\}$. Then, based on $J^0(x(\tau)) = \Lambda^0(x(\tau)) = 0$, the proof is completed by means of mathematical induction.

Lemma 2: Defining the value function J^i as (12), there exists an upper bound $\bar{Z}(x(\tau))$ satisfying:

$$\forall i : 0 \le J^i(x(\tau)) \le J^*(x(\tau)) \le \bar{Z}(x(\tau)) \tag{15}$$

Proof: Assuming $J^0(x(\tau)) = Z^0(x(\tau)) = 0$, a new iterative function Z^i is defined by:

$$Z^{i+1}(x(\tau)) = Q_f(x(\tau), \mu(x(\tau))) + \gamma \cdot Z^i(x(\tau+1)), \quad \tau \in [k, k+N-1] \tag{16}$$

where $Z^i(x(k+N)) = F(x(k+N))$, $i \ge 1$, μ is an arbitrary control strategy that makes the state satisfies the constraints $\|x(\tau)\| < \infty (\forall \tau = k, k+1, \cdots, k+N)$.

For $i \le k+N-1-\tau$, it can be had

$$\begin{aligned} &Z^{i+1}(x(\tau)) \\ &= Q_f(x(\tau), \mu(x(\tau))) + \gamma \cdot Z^i(x(\tau+1)) \\ &= \sum_{n=0}^{n=1} \gamma^n Q_f(x(\tau+n), \mu(x(\tau+n))) + \gamma^2 Z^{i-1}(x(\tau+2)) \\ &\cdots \\ &= \sum_{n=0}^{n=i} \gamma^n Q_f(x(\tau+n), \mu(x(\tau+n))) + \gamma^{i+1} Z^0(x(\tau+i+1)) \end{aligned} \tag{17}$$

And because $Z^0(x(\tau)) = 0$, (17) can be further expressed as:

$$Z^{i+1}(x(\tau)) = \sum_{n=0}^{n=i} \gamma^n Q_f(x(\tau+n), \mu(x(\tau+n))) \tag{18}$$

Moreover, for $i > k+N-1-\tau$, the following equation can be obtained:

$$\begin{aligned} Z^{i+1}(x(\tau)) &= \sum_{n=0}^{n=k+N-1-\tau} \gamma^n Q_f(x(\tau+n), \mu(x(\tau+n))) \\ &\quad + \gamma^{k+N-\tau} F(x(k+N)) \\ &= \bar{Z}(x(\tau)) \end{aligned} \tag{19}$$

According to (18) and (19), the below formula holds.

$$\forall i : Z^{i+1}(x(\tau)) \le \bar{Z}(x(\tau)), \tag{20}$$

Let $v^i = \mu$, from Lemma 1, it can be acquired:

$$J^i(x(\tau)) \le Z^i(x(\tau)) \tag{21}$$

Here, let $\mu = u^*$, it can be obtained based on (21):

$$J^i(x(\tau)) \le J^*(x(\tau)) \tag{22}$$

Besides, considering u^* is the optimal control action, it can be got based on (20):

$$J^*(x(\tau)) \le \bar{Z}(x(\tau)) \tag{23}$$

Therefore, combining (22) and (23), a conclusion can be drawn that (15) holds for $\forall i$.

Lemma 3: If $J^0(x(\tau)) = 0$, then the below inequality holds

$$\forall i : J^{i+1}(x(\tau)) \ge J^i(x(\tau)) \tag{24}$$

Proof: Let $v^i = u^{i+1}$, according to Lemma 1, it can be known that

$$\begin{aligned} \Lambda^{i+1}(x(\tau)) &= Q_f(x(\tau), u^{i+1}(x(\tau))) \\ &\quad + \gamma \cdot \Lambda^i(x(\tau+1)), \ \tau \in [k, k+N-1] \end{aligned} \tag{25}$$

Next, the following formula will be proved by mathematical induction.

$$\Lambda^i(x(\tau)) \le J^{i+1}(x(\tau)) \tag{26}$$

When $i = 0$,, it can be had

$$J^1(x(\tau)) - \Lambda^0(x(\tau)) = Q_f(x(\tau), u^0(x(\tau))) \ge 0, \tag{27}$$

Suppose $\Lambda^{i-1}(x(\tau)) \le J^i(x(\tau))$ holds, from (12) and (25), it can be got:

$$\begin{aligned} &\Lambda^i(x(\tau)) - J^{i+1}(x(\tau)) \\ &= \gamma \cdot \left(\Lambda^{i-1}(x(\tau+1)) - J^i(x(\tau+1))\right) \le 0 \end{aligned} \tag{28}$$

Thus, the proof of (26) is completed. Combining (14) from Lemma 1 and (26), this conclusion of Lemma 3 can be obtained.

Lemma 4: In $[k, k+N-1]$, if $J^0(x(\tau)) = 0$, then the below inequality holds

$$\forall i \ge N : J^{i+1}(x(\tau)) = J^i(x(\tau)) \tag{29}$$

Proof: For $i \ge 1$, let $v^i = u^{i-1}$, it can be acquired based on Lemma 1:

$$\begin{aligned} &\Lambda^{i+1}(x(\tau)) \\ &= \begin{cases} Q_f(x(\tau), v^0(x(\tau))) + \gamma \cdot \Lambda^0(x(\tau+1)), i = 0 \\ Q_f(x(\tau), u^{i-1}(x(\tau))) + \gamma \cdot \Lambda^i(x(\tau+1)), i \ge 1 \end{cases} \end{aligned} \tag{30}$$

According to (12) and (30), for $i \ge N$, it can be shown:

$$\begin{aligned} &\Lambda^{i+1}(x(\tau)) - J^i(x(\tau)) \\ &= \gamma \cdot \left[\Lambda^i(x(\tau+1)) - J^{i-1}(x(\tau+1))\right] \\ &= \gamma^2 \cdot \left[\Lambda^{i-1}(x(\tau+2)) - J^{i-2}(x(\tau+2))\right] \\ &\cdots \\ &= \gamma^{k+N-1-\tau} \cdot \left[\Lambda^{i-N+\tau-k+2}(x(k+N-1)) \right. \\ &\quad \left. - J^{i-N+\tau-k+1}(x(k+N-1))\right] \end{aligned} \tag{31}$$

Considering that when $i \geq N$, it can be obtained

$$\begin{aligned}&\Lambda^{i-N+\tau-k+2}(x(k+N-1))\\&= Q_{\mathrm{f}}(x(k+N-1), u^{i-N+\tau-k}(x(k+N-1))) + \gamma\\&\quad \cdot F(x(k+N))\\&= J^{i-N+\tau-k+1}(x(k+N-1))\end{aligned} \tag{32}$$

Combining (31) and (32), the below equation can be obtained

$$\Lambda^{i+1}(x(\tau)) = J^{i}(x(\tau)), i \geq N \tag{33}$$

Meanwhile, according to (14) from Lemma 1, which is given by

$$J^{i+1}(x(\tau)) \leq J^{i}(x(\tau)), i \geq N \tag{34}$$

Therefore, combining (34) and Lemma 3, the proof of Lemma 4 is finished.

Theorem 1: If $J^0(x(\tau)) = 0$, then $J^i = J^*$and$u^i = u^*$ when $i \geq N$.

*Proof:*From Lemma 2 and Lemma 3, it can be seen that the limit of J^i exists, that is, when $i \to \infty$, $J^i \to J^\infty$. Based on (12), it can be known

$$\begin{aligned}J^{\infty}(x(\tau)) = &\sum_{j=\tau}^{k+N-1} \gamma^{j-\tau} Q_{\mathrm{f}}(x(j), u^{\infty}(x(j)))\\&+ \gamma^{k+N-\tau} \cdot F(x(k+N)), \tau \in [k, k+N-1]\end{aligned} \tag{35}$$

It can be seen from Lemma 2 that $J^i(x(\tau))$ is bounded. Thus, $J^\infty(x(\tau))$ is also bounded. This means $\|x(j)\| < \infty, j \in [\tau, k+N]$ in the constrains of u^∞. Let $\mu = u^\infty$, it can be got from Lemma 2:

$$J^*(x(\tau)) \leq J^{\infty}(x(\tau)) \tag{36}$$

Meanwhile, according to Lemma 2, it can be obtained by

$$J^{\infty}(x(\tau)) \leq J^*(x(\tau)) \tag{37}$$

Hence, based on (36) and (37), it can be obtained that $J^\infty(x(\tau)) = J^*(x(\tau))$. Then, combining Lemma 4, the following equality holds.

$$J^{i}(x(\tau)) = J^{\infty}(x(\tau)) = J^*(x(\tau)), i \geq N \tag{38}$$

Considering (11) and the definition of u^* in (10), it can be easily got that $u^i = u^*$ when $i \geq N$. So far, the Theorem 1 is proved.

2) STABILITY PROOF OF THE CONTROL SYSTEM BASED ON HDP-MPC

The optimal value function of (9) can be rewritten on the basis of (8):

$$\begin{aligned}J^*(x(\tau)) = &\sum_{j=\tau}^{k+N-1} \gamma^{j-\tau} Q_{\mathrm{f}}(x(j), u^*(x(j)))\\&+ \gamma^{k+N-\tau} \cdot F(x(k+N)),\ \tau \in [k, k+N-1]\end{aligned} \tag{39}$$

It can be inferred from Lemma 2 that $J^*(x(\tau))$ is bounded. And then, we make the following assumption.

Assumption 1: There exists a first-order continuous derivable function $F(\cdot)$, for $\forall\tau \in [(n-1)N, nN-1]$ and $n = 1, 2, 3 \cdots$, satisfying

$$J^*(x(\tau)) \leq F(x(\tau)) \tag{40}$$

Theorem 2: If Assumption 1 holds, then the controlled system (1) based on the HDP-MPC method is asymptotically stable in the mean sense.

Proof: In the nth prediction time domain $[(n-1)N, nN-1]$, defining the Lyapunov function:

$$J(\tau) = J^*(x(\tau)) \tag{41}$$

where $\tau \in [(n-1)N, nN-1]$, here, for simplification, γ is set to 1. Based on (39), it can be had

$$J(\tau) = Q_{\mathrm{f}}(x(\tau), u^*(x(\tau))) + J(\tau+1) \tag{42}$$

According to the definition of the optimal value function $J^*(x(\tau))$, it can be concluded that $J(\tau) \geq 0$.Then, $\Delta J(\tau) = J(\tau+1) - J(\tau) < 0$ will be proved when $Q_{\mathrm{f}}(x(\tau), u^*(x(\tau))) \neq 0$.

1) Case 1: For $(n-1)N \leq \tau < \tau+1 \leq nN-1$, it can be obtained based on (42)

$$\Delta J(\tau) = J(\tau+1) - J(\tau) = -Q_{\mathrm{f}}(x(\tau), u^*(x(\tau))) \tag{43}$$

2) Case 2: For $\tau = nN-1$, according to Assumption 1 and (41), the below inequality satisfies.

$$\begin{aligned}J(\tau+1) &= J^*(x(\tau+1))\\&\leq F(x(\tau+1))\\&= J(\tau) - Q_{\mathrm{f}}(x(\tau), u^*(x(\tau)))\end{aligned} \tag{44}$$

Therefore, it can be concluded that

$$\Delta J(\tau) = J(\tau+1) - J(\tau) = -Q_{\mathrm{f}}(x(\tau), u^*(x(\tau))) \tag{45}$$

Based on (43) and (45), $\Delta J(\tau) < 0, \forall\tau = 0, 1, 2, \cdots, \infty$ always holds when $Q_{\mathrm{f}}(x(\tau), u^*(x(\tau))) \neq 0$. According to the Lyapunov stability theorem, $J(\tau)$ is a Lyapunov function, so Theorem 2 is proved.

C. DESIGN OF THE CONTROLLER

In order to obtain u^* each time for energy management controller, the network parameters of AN and CN need to be continuously adjusted. In addition, for realizing this process, the specific structures of the AN and CN need to be known in advance. In this paper, BPNN with strong nonlinear fitting and self-learning abilities is adopted to establish the CN and AN in predictive horizon $[k, k+N-1]$. Here, to simplify expression, $J(\tau)$ and $u(\tau)$ are used to replace $J(x(\tau))$ and $u(x(\tau))$, respectively.

1) DESIGN OF CRITIC NETWORK

Firstly, based on the input-output relationship of the critic network in Fig.6, a BPNN with 2-5-1 structure is selected to describe the internal structure of the CN, shown in Fig.7 (a). Considering the nonlinearity of information processing for CN, tansig and purelin functions are chosen as the transfer functions for the hidden-layer and output-layer, respectively. Meanwhile, the thresholds for two layers are set to 0. Accordingly, in $\tau \in [k, k+N-1]$, the transfer relationship of the CN can be represented as

$$\begin{cases} \boldsymbol{c}_{\mathrm{h1}}(\tau) = \boldsymbol{W}^{i}_{\mathrm{c1}}(\tau) \times x(\tau) \\ \boldsymbol{c}_{\mathrm{h2}}(\tau) = (1 - e^{-\boldsymbol{c}_{\mathrm{h1}}(\tau)}) \Big/ (1 + e^{-\boldsymbol{c}_{\mathrm{h1}}(\tau)}) \\ \hat{J}^{i}(\tau) = \boldsymbol{W}^{i}_{\mathrm{c2}}(\tau) \times \boldsymbol{c}_{\mathrm{h2}}(\tau) \end{cases} \tag{46}$$

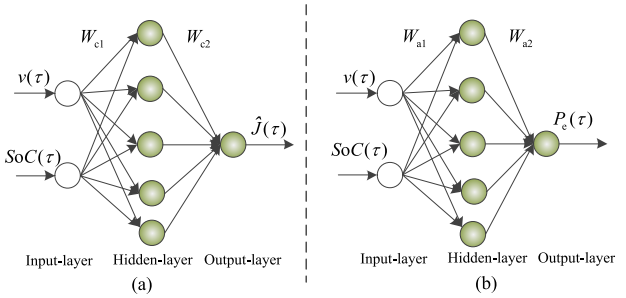

FIGURE 7. BPNN structures for the control network. (a) CN. (b) AN.

where $\boldsymbol{W}_{\mathrm{c1}}$ and $\boldsymbol{W}_{\mathrm{c2}}$ are the weight matrixes of the input-to hidden and the hidden-to-output layers, respectively. i is the iterations. Since the CN learning target is to minimize $e_{\mathrm{c}}(\tau) = \hat{J}^{i}(\tau) - Q_{\mathrm{f}}(\tau) - \gamma \hat{J}^{i}(\tau+1)$, that is

$$E_{\mathrm{c}}(\tau) = \frac{1}{2} e_{\mathrm{c}}^{2}(\tau) \le \varepsilon_{\mathrm{c}} \tag{47}$$

Herein, ε_{c} is a set error, and E_{c} can be utilized to update $\boldsymbol{W}_{\mathrm{c1}}$ and $\boldsymbol{W}_{\mathrm{c2}}$. Considering the real-time training and computational complexity, the gradient descent method is selected to train $\boldsymbol{W}_{\mathrm{c1}}$ and $\boldsymbol{W}_{\mathrm{c2}}$. The specific updating processes are as follows:

$$\begin{cases} \Delta \boldsymbol{W}^{i}_{\mathrm{c1}}(\tau) = l_{\mathrm{c}}(\tau)[-\dfrac{\partial E_{\mathrm{c}}(\tau)}{\partial \boldsymbol{W}^{i}_{\mathrm{c1}}}] = l_{\mathrm{c}}(\tau) \\ [-\dfrac{\partial E_{\mathrm{c}}(\tau)}{\partial \hat{J}^{i}(\tau)} \dfrac{\partial \hat{J}^{i}(\tau)}{\partial \boldsymbol{c}_{\mathrm{h2}}(\tau)} \dfrac{\partial \boldsymbol{c}_{\mathrm{h2}}(\tau)}{\partial \boldsymbol{c}_{\mathrm{h1}}(\tau)} \dfrac{\partial \boldsymbol{c}_{\mathrm{h1}}(\tau)}{\partial \boldsymbol{W}^{i}_{\mathrm{c1}}(\tau)}] \\ = -\dfrac{1}{2} l_{\mathrm{c}}(\tau) \cdot e_{\mathrm{c}}(\tau) \cdot \left\{ \boldsymbol{W}^{\mathrm{T}i}_{\mathrm{c2}}(\tau) \otimes [1 - \boldsymbol{c}_{\mathrm{h2}}(\tau) \otimes \boldsymbol{c}_{\mathrm{h2}}(\tau)] \right\} \\ \quad \times x^{\mathrm{T}}(\tau) \\ \boldsymbol{W}^{i+1}_{\mathrm{c1}}(\tau) = \boldsymbol{W}^{i}_{\mathrm{c1}}(\tau) + \Delta \boldsymbol{W}^{i}_{\mathrm{c1}}(\tau) \end{cases} \tag{48}$$

$$\begin{cases} \Delta \boldsymbol{W}^{i}_{\mathrm{c2}}(\tau) = l_{\mathrm{c}}(\tau)[-\dfrac{\partial E_{\mathrm{c}}(\tau)}{\partial \boldsymbol{W}^{i}_{\mathrm{c2}}(\tau)}] \\ = l_{\mathrm{c}}(\tau)[-\dfrac{\partial E_{\mathrm{c}}(\tau)}{\partial e_{\mathrm{c}}(\tau)} \dfrac{\partial e_{\mathrm{c}}(\tau)}{\partial \hat{J}^{i}(\tau)} \dfrac{\partial \hat{J}^{i}(\tau)}{\partial \boldsymbol{W}^{i}_{\mathrm{c2}}(\tau)}] \\ = -l_{\mathrm{c}}(\tau) \cdot e_{\mathrm{c}}(\tau) \cdot \boldsymbol{c}^{\mathrm{T}}_{\mathrm{h2}}(\tau) \\ \boldsymbol{W}^{i+1}_{\mathrm{c2}}(\tau) = \boldsymbol{W}^{i}_{\mathrm{c2}}(\tau) + \Delta \boldsymbol{W}^{i}_{\mathrm{c2}}(\tau) \end{cases} \tag{49}$$

where, $l_{\mathrm{c}} \in (0, 1]$ denotes the learning factor. $\times$ and $\otimes$ respectively represent the product and Kronecker product of the matrices.

2) DESIGN OF ACTOR NETWORK

Then, the AN is analyzed. $Q_{\mathrm{f}}(\tau) + \gamma \hat{J}^{i}(\tau+1)$ reaching a stable value is the AN learning target. It can be seen from (2), for the require power $P_{\mathrm{d}}(\tau)$ of SHL at the τth(s), $P_{\mathrm{b}}(\tau)$ and $P_{\mathrm{e}}(\tau)$ are not completely independent. In other words, while determining $P_{\mathrm{e}}(\tau)$, $P_{\mathrm{b}}(\tau)$ can be achieved via utilizing

$$P_{\mathrm{b}}(\tau) = P_{\mathrm{d}}(\tau) - P_{\mathrm{e}}(\tau) \cdot \eta_{\mathrm{g}} \tag{50}$$

Besides, Q_{f} is only related to T_{e} and n_{e} from (7). Therefore, $\hat{u}(\tau)$ can be regarded as $[P_{\mathrm{e}}(\tau), P_{\mathrm{d}}(\tau) - P_{\mathrm{e}}(\tau) \cdot \eta_{\mathrm{g}}]^{\mathrm{T}}$.Herein, the AN is used directly to obtain $P_{\mathrm{e}}(\tau)$. Specifically, a BPNN with 2-5-1 architecture is chosen to depict the internal structure of the AN, as described in Fig.7 (b). Similarly, the activation functions of its hidden and output layers still choose tansig and purelin functions, respectively. Then, in $\tau \in [k, k+N-1]$, the forward calculation process of the AN can be written as

$$\begin{cases} \boldsymbol{a}_{\mathrm{h1}}(\tau) = \boldsymbol{W}^{i}_{\mathrm{a1}}(\tau) \times x(\tau) \\ \boldsymbol{a}_{\mathrm{h2}}(\tau) = (1 - e^{-\boldsymbol{a}_{\mathrm{h1}}(\tau)}) \Big/ (1 + e^{-\boldsymbol{a}_{\mathrm{h1}}(\tau)}) \\ P^{i}_{\mathrm{e}}(\tau) = \boldsymbol{W}^{i}_{\mathrm{a2}}(\tau) \times \boldsymbol{a}_{\mathrm{h2}}(\tau) \end{cases} \tag{51}$$

where $\boldsymbol{W}_{\mathrm{a1}}$ and $\boldsymbol{W}_{\mathrm{a2}}$ are the weight matrixes of the input-to hidden layer and the hidden-to-output layer, respectively. Herein, the gradient ascent method is used to adjust $\boldsymbol{W}_{\mathrm{a1}}$ and $\boldsymbol{W}_{\mathrm{a2}}$ by using $Q_{\mathrm{f}}(\tau) + \gamma \hat{J}^{i}(\tau+1)$. The specific training processes are as below:

$$\begin{cases} \Delta \boldsymbol{W}^{i}_{\mathrm{a1}}(\tau) = l_{\mathrm{a}}(\tau) \cdot [\dfrac{\partial \left[Q_{\mathrm{f}}(\tau) + \gamma \hat{J}^{i}(\tau+1) \right]}{\partial \boldsymbol{W}^{i}_{\mathrm{a1}}(\tau)}] \\ = l_{\mathrm{a}}(\tau) \cdot [\dfrac{\partial \left[Q_{\mathrm{f}}(\tau) + \gamma \hat{J}^{i}(\tau+1) \right]}{\partial u^{i}(\tau)} \\ \dfrac{\partial u^{i}(\tau)}{\partial \boldsymbol{a}_{\mathrm{h2}}(\tau)} \dfrac{\partial \boldsymbol{a}_{\mathrm{h2}}(\tau)}{\partial \boldsymbol{a}_{\mathrm{h1}}(\tau)} \dfrac{\partial \boldsymbol{a}_{\mathrm{h1}}(\tau)}{\partial \boldsymbol{W}^{i}_{\mathrm{a1}}(\tau)}] \\ = \dfrac{1}{2} l_{\mathrm{a}}(\tau) \cdot \boldsymbol{W}^{\mathrm{T}i}_{\mathrm{a2}}(\tau) \times [\dfrac{\partial Q_{\mathrm{f}}(\tau)}{\partial u^{i}(\tau)} + \dfrac{\gamma}{4} \cdot \boldsymbol{W}_{\mathrm{m2}}(\tau) \\ \quad \times \{\boldsymbol{W}_{\mathrm{m1u}}(\tau) \\ \otimes [1 - \boldsymbol{m}_{\mathrm{h2}}(\tau) \otimes \boldsymbol{m}_{\mathrm{h2}}(\tau), 1 - \boldsymbol{m}_{\mathrm{h2}}(\tau) \otimes \boldsymbol{m}_{\mathrm{h2}}(\tau)]\} \\ \times \boldsymbol{W}^{\mathrm{T}i}_{\mathrm{c1}}(\tau) \times \left\{ \boldsymbol{W}^{\mathrm{T}i}_{\mathrm{c2}}(\tau) \otimes [1 - \boldsymbol{c}_{\mathrm{h2}}(\tau+1) \otimes \boldsymbol{c}_{\mathrm{h2}}(\tau+1)] \right\}] \\ \otimes [1 - \boldsymbol{a}_{\mathrm{h2}}(\tau) \otimes \boldsymbol{a}_{\mathrm{h2}}(\tau)] \times x^{\mathrm{T}}(\tau) \\ \boldsymbol{W}^{i+1}_{\mathrm{a1}}(\tau) = \boldsymbol{W}^{i}_{\mathrm{a1}}(\tau) + \Delta \boldsymbol{W}^{i}_{\mathrm{a1}}(\tau) \end{cases} \tag{52}$$

$$\begin{cases} \Delta \boldsymbol{W}^{i}_{\mathrm{a2}}(\tau) = l_{\mathrm{a}}(\tau) \cdot [\dfrac{\partial \left[Q_{\mathrm{f}}(\tau) + \gamma \hat{J}^{i}(\tau+1) \right]}{\partial \boldsymbol{W}^{i}_{\mathrm{a2}}(\tau)}] \\ = l_{\mathrm{a}}(\tau) \cdot [\dfrac{\partial \left[Q_{\mathrm{f}}(\tau) + \gamma \hat{J}^{i}(\tau+1) \right]}{\partial u^{i}(\tau)} \dfrac{\partial u^{i}(\tau)}{\partial \boldsymbol{W}^{i}_{\mathrm{a2}}(\tau)}] \\ = l_{\mathrm{a}}(\tau) \cdot (\dfrac{\partial Q_{\mathrm{f}}(\tau)}{\partial u^{i}(\tau)} + \dfrac{\gamma}{4} \cdot \boldsymbol{W}_{\mathrm{m2}}(\tau) \times \{\boldsymbol{W}_{\mathrm{m1u}}(\tau) \\ \otimes [1 - \boldsymbol{m}_{\mathrm{h2}}(\tau) \otimes \boldsymbol{m}_{\mathrm{h2}}(\tau), 1 - \boldsymbol{m}_{\mathrm{h2}}(\tau) \otimes \boldsymbol{m}_{\mathrm{h2}}(\tau)]\} \\ \times \boldsymbol{W}^{\mathrm{T}i}_{\mathrm{c1}}(\tau) \\ \times \left\{ \boldsymbol{W}^{\mathrm{T}i}_{\mathrm{c2}}(\tau) \otimes [1 - \boldsymbol{c}_{\mathrm{h2}}(\tau+1) \otimes \boldsymbol{c}_{\mathrm{h2}}(\tau+1)] \right\}) \times \boldsymbol{a}^{\mathrm{T}}_{\mathrm{h2}}(\tau) \\ \boldsymbol{W}^{i+1}_{\mathrm{a2}}(\tau) = \boldsymbol{W}^{i}_{\mathrm{a2}}(\tau) + \Delta \boldsymbol{W}^{i}_{\mathrm{a2}}(\tau) \end{cases} \tag{53}$$

where, $l_{\mathrm{a}} \in (0, 1]$ denotes the learning factor. $\boldsymbol{W}_{\mathrm{m1}u} = \boldsymbol{W}_{\mathrm{m1}}(:,1{:}2)$.

After designing the structures of the CN and AN, the energy management controller based on HDP-MPC will be realized by designing the online EMS.

D. ONLINE EMS BASED ON HDP-MPC

As known from (1), the state information of working condition includes the vehicle speed v and SoC. For solving MPC-based HDP strategy, firstly, it is very necessary to predict the working condition state information in prediction horizon in advance, which will be elaborated in detail in this subsection. For speed online prediction, a double-layers speed prediction model based on BPNN and long short-term memory (BPNN-LSTM) have been designed. Please refer to our previous work [37] for specific details, which can not be elaborated here. For SoC online prediction, considering the powerful nonlinear fitting ability of BPNN, thus it is used to build the SoC prediction model.

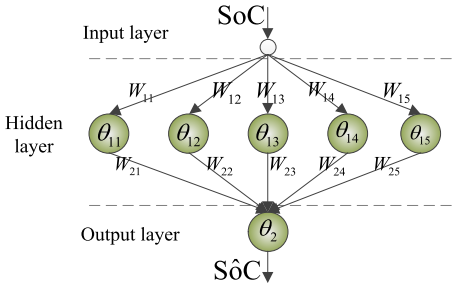

FIGURE 8. BPNN structure for SoC prediction model.

Because the relation of input and output parameters meets (3) in [35], a BPNN with 1-5-1 structure is chosen, as described in Fig. 8. Correspondingly, the model can be represented as

$$\begin{cases} \boldsymbol{H}_{\mathrm{in}} = \mathrm{SoC}(k) \times \boldsymbol{W}_1 + \boldsymbol{\theta}_1 \\ \boldsymbol{H}_{\mathrm{ou}} = g\left(\boldsymbol{H}_{\mathrm{in}}\right) \\ \hat{\mathrm{SoC}}(k+1) = h(\boldsymbol{H}_{\mathrm{ou}} \times \boldsymbol{W}_2 + \boldsymbol{\theta}_2) \end{cases} \tag{54}$$

here, $\mathrm{SoC}(k)$ is the real value of current moment SoC, and $\hat{\mathrm{SoC}}(k+1)$ denotes the predictive value of next moment. They are regarded as the input and output variables for BPNN, respectively. $\boldsymbol{\theta}_1 = [\theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}]$ and $\boldsymbol{\theta}_2 = \theta_2$ respectively denote the thresholds of the hidden and output layers. $\boldsymbol{W}_1 = [W_{11}, W_{12}, W_{13}, W_{14}, W_{15}]$ and $\boldsymbol{W}_2 = [W_{21}, W_{22}, W_{23}, W_{24}, W_{25}]^{\mathrm{T}}$ are the weights of input-to-hidden and hidden-to-output layers, respectively. The tansig and purelin functions are selected as the hidden-layer transfer function g and the output-layer transfer function h, respectively. After that, employing the collected SoC data in Fig.4 (d) to train and update the thresholds and weights, the predictive model of SoC can be obtained. By means of the speed and SoC prediction models, the condition information in the prediction window can be predicted and obtained.

Then, the HDP algorithm is used to achieve the online rolling optimization of fuel consumption in $[k, k+N-1]$, and an online EMS based on HDP-MPC is proposed. In the process of designing the energy management controller based on HDP-MPC, firstly, the MPC method is used to predict other unknown state information within the prediction window $[k, k+N-1]$ corresponding to the current state $[v(k), \mathrm{SoC}(k)]$, determining the state set $\{[v(k), \mathrm{SoC}(k)], [\hat{v}(k+1), \hat{\mathrm{SoC}}(k+1)], \cdots, [\hat{v}(k+N-1), \hat{\mathrm{SoC}}(k+N-1)]\}$.

Specifically, in terms of speed prediction online, based on velocity speed $v(k)$, acceleration $a(k)$, slope angle $\theta(k)$, and pedal degree $p(k)$ of the current moment, the designed BPNN-LSTM prediction model is used to predict speed sequence $\left\{\hat{v}(k+1), \hat{v}(k+2), \cdots, \hat{v}(k+N-1)\right\}$ in $[k, k+N-1]$. Then, $P_{\mathrm{d}}(\tau)$ of each time step τ can be calculated by (1) and (2), here, $\tau \in [k, k+N-1]$. Besides, for SoC prediction, based on $\mathrm{SoC}(k)$ of the current moment, the established BPNN prediction model of SoC is utilized to predict SoC sequence $\left\{\hat{\mathrm{SoC}}(k+1), \hat{\mathrm{SoC}}(k+2), \cdots, \hat{\mathrm{SoC}}(k+N-1)\right\}$ in $[k, k+N-1]$. Based on the above analysis, the architecture for online EMS implementation of HDP-MPC controller is displayed in Fig. 9. Correspondingly, the concrete implement process is described in Fig. 10 (supposing the present moment is kth (s)). After executing the algorithm in Fig.10, $\left[P^{*}_{\mathrm{e}}(k), P^{*}_{\mathrm{b}}(k)\right]$ is sent to the SHL, producing the new state $[v(k+1), \mathrm{SoC}(k+1)]$ of SHL at $(k+1)$ th (s), and entering the next round of MPC adjustment.

As the vehicle continues to move forward, the above process will be executed rolling until the vehicle stops driving, and the corresponding detailed algorithm is shown in **Algorithm 1.**

IV. RESULTS AND DISCUSSION

In this section, the BPNN-based dynamics model of the SHL and the designed online EMS based on HDP-MPC are experimented and analyzed, respectively.

A. EXPERIMENTAL SETTINGS

1) TRAINING DATA SETTINGS

The data shown in Fig. 4 are used to train the BPNN-based dynamics model of SHL offline in this section. Besides, the

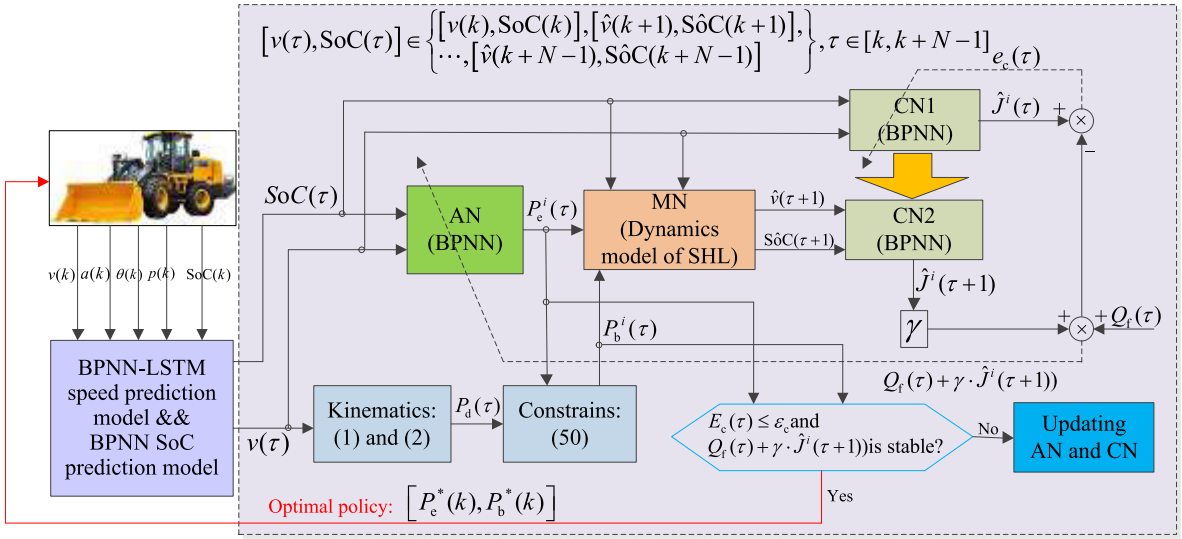

FIGURE 9. The online EMS implementation framework based on HDP-MPC.

training and test data account for 70 % and 30 %, respectively. In order to improve the calculation precision, all the sample data are normalized to [−1, 1] based on (25) in [36].

TABLE 1. The parameters for SHL.

Type	Description
Vehicle body	$m_v=17500\mathrm{kg}; A=10\mathrm{m}^2; C_D=0.7; r_{wh}=0.75\mathrm{m};$ $\eta_t=0.92; \rho_r=0.03;$
Engine	3500rpm@400Nm
Motor	5300rpm@1400Nm
Generator	3600rpm@700Nm
Battery	$C_{Ah}=100\mathrm{Ah}; V_b=600\mathrm{V}; R_b=0.19\mathrm{ohm}; \eta_{b\_dis}=\eta_{b\_ch}=0.95$
Other	$g=9.81\mathrm{m/s}^2; \rho_a=1.2\mathrm{kg/m}^3; \mathrm{SoC}_{min}=0.3; \mathrm{SoC}_{max}=0.8;$

2) PARAMETERS INITIALIZATION

α and η of (7) in [35], and ε_m for MN are set to 0.95, 0.05, and 0.0005, respectively. γ, ε_c, l_c, and l_a of **Algorithm 1** are set to 1, 0.001, 0.05 and 0.001, respectively.

3) VEHICLE SETTINGS

For fully validating the performance of the proposed online EMS, the HIL bench is utilized to establish the vehicle simulation platform, as shown in Fig.11. The vehicle parameters for SHL are shown in TABLE 1.

4) ERROR ESTIMATE METHODS

In this paper, the four errors, that is, absolute error *AE*, average-absolute error *MAE*, root-average-square error *RMSE*, and average-relative error *MRE*, are utilized to estimate the effectiveness of the dynamics model and the designed HDP-MPC strategy. (26) in [36] gives the elaborate expression for four errors.

B. ANALYSIS OF BPNN DYNAMICS MODEL OF SHL

When the BPNN dynamics model of the SHL is trained and completed, $\boldsymbol{W}_{m1}$ and $\boldsymbol{W}_{m2}$ in (4) can be determined, as shown in TABLE 2. For verifying the effect of the established dynamics model for SHL, the analytical model and BPNN model of the SHL are simulated and compared. Considering the strong periodicity of the SHL operation, this paper takes the working scenario of Fig. 3 as the simulation environment and performs one cycle simulation. The change process of actual and model states of the SHL is described in Fig. 12. Among them, AE between the actual and model states of the SHL is shown in Fig. 13, and the corresponding *MAE*, *RMSE*, and *MRE* are listed in TABLE 3. It can be seen from Fig. 13 that the AE changing range of BPNN model is smaller than that of analytical model. In addition, as known from TABLE 3, the corresponding *MAE*, *RMSE*, and *MRE* of the BPNN model and actual states for SHL are also less than the analytical model, respectively. Based on the above analysis, the BPNN model can more effectively reflect the actual motion process of the SHL with higher accuracy. Namely, the vehicle motion process evolved by it is closer to the actual motion process of the SHL.

TABLE 2. BPNN parameters of dynamics model for SHL.

Parameters	Values
$\boldsymbol{W}_{m1}$	$\begin{bmatrix} -0.771 & -0.608 & 0.084 & 0.887 \\ 0.498 & -1.449 & 0.043 & 0.858 \\ -0.093 & 0.588 & 0.024 & 0.245 \\ 0.550 & -0.104 & 0.051 & 0.395 \\ -0.992 & -0.939 & 0.147 & 1.441 \end{bmatrix}$
$\boldsymbol{W}_{m2}$	$\begin{bmatrix} -0.369 & 0.046 & -0.146 & 1.358 & -0.189 \\ -0.277 & -0.029 & 1.457 & -0.158 & -0.171 \end{bmatrix}$

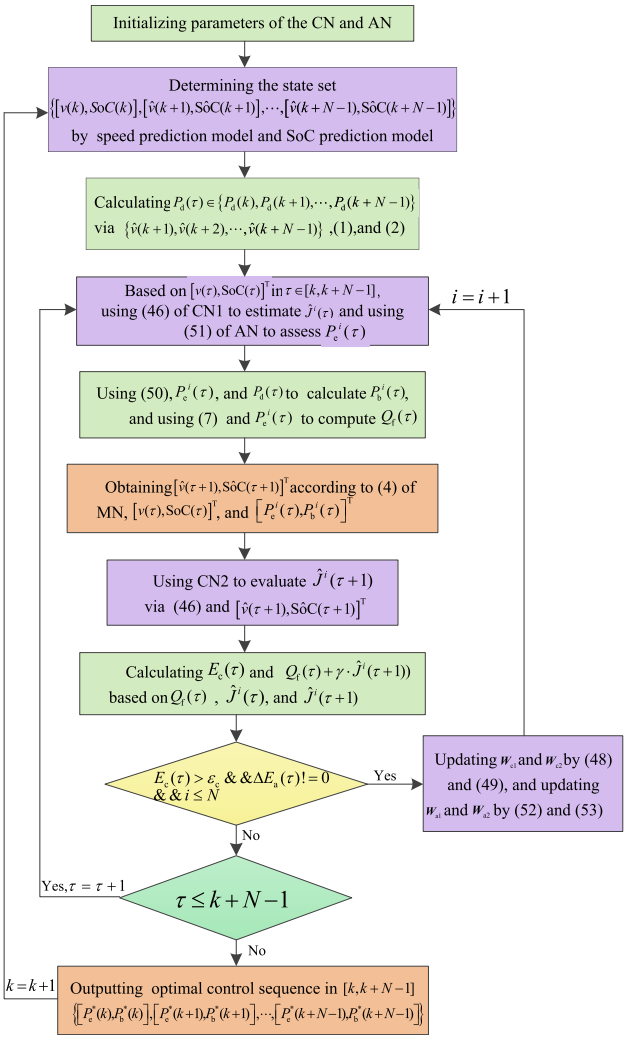

FIGURE 10. The flowchart of HDP-MPC EMS.

TABLE 3. Errors between model state and actual state.

Model state	*MAE*	*RMSE*	*MRE*
BPNN-speed	0.0593	0.0035	0.2469
Analytical-speed	0.1948	0.0121	0.3406
BPNN-SoC	0.0106	0.0047	0.0013
Analytical-SoC	0.0173	0.0089	0.0021

C. PERFORMANCE ANALYSIS OF ONLINE EMS BASED ON HDP-MPC

1) INITIAL VALUE DETERMINATION OF THE CONTROLLER

Before using the HDP-MPC controller to find the optimal strategy, the initial weights of the CN and AN need to be determined in advance. In order to make the CN and AN converge as soon as possible online, the data in Fig. 4 are used to train $\boldsymbol{W}_{c1}$, $\boldsymbol{W}_{c2}$, $\boldsymbol{W}_{a1}$, and $\boldsymbol{W}_{a2}$ offline. The training process is shown in Fig. 14. It can be seen that E_c gradually declines and stabilizes, and $Q_f + \gamma \cdot \hat{J}^i$ gradually increases and stabilizes as the training iterations increase, which is consistent with the conclusion of Lemma 3 and Lemma 4 in Section III. Specifically, when the training iterations reach to 491, $E_c = 0.00098 \leq \varepsilon_c$ and $Q_f + \gamma \cdot \hat{J}^i$ can be stabilized at 6.93395. Correspondingly, the final initial weights are shown in TABLE 4.

Algorithm 1 Online EMS Algorithm Based on HDP-MPC

// Step 1: Initializing parameters:
γ, $\boldsymbol{W}_{c1}$, $\boldsymbol{W}_{c2}$, l_c, $\boldsymbol{W}_{a1}$, $\boldsymbol{W}_{a2}$, ε_c, l_a, k and N;
for $k = k_0 : 1 : k_f - 1$// *Loop control until the vehicle stops*
// Step 2: Condition information prediction:
Obtaining $[v(k), a(k), \theta(k), p(k), \mathrm{SoC}(k)]$; // *The current moment condition information*
Using the speed prediction model:
$[v(k), a(k), \theta(k), p(k)] \Rightarrow \{\hat{v}(k+1), \hat{v}(k+2), \cdots, \hat{v}(k+N-1)\}$;
Using the SoC prediction model:
$\mathrm{SoC}(k) \Rightarrow \{\hat{\mathrm{SoC}}(k+1), \hat{\mathrm{SoC}}(k+2), \cdots, \hat{\mathrm{SoC}}(k+N-1)\}$
Determining the state set in $[k, k+N-1]$:
$\{[v(k), \mathrm{SoC}(k)], [\hat{v}(k+1), \hat{\mathrm{SoC}}(k+1)], \cdots, [\hat{v}(k+N-1), \hat{\mathrm{SoC}}(k+N-1)]\}$
// Step 3: Calculating the demand power:
Running (1) and (2); // *Getting*
$P_d(\tau) \in \{P_d(k), P_d(k+1), \cdots, P_d(k+N-1)\}$
for $\tau = k : 1 : k+N-1$
$i = 1$;
// Step 4: Evaluating $\hat{J}^i(\tau)$ *and* $\hat{u}^i(\tau)$***:***
Producing (46); // *Utilizing* $[v(\tau), \mathrm{SoC}(\tau)]^{\mathrm{T}}$ *to evaluate* $\hat{J}^i(\tau)$
Executing (51); // *Employing* $[v(\tau), \mathrm{SoC}(\tau)]^{\mathrm{T}}$ *n to evaluate* $P_e^i(\tau)$
Executing (50); // *Utilizing* $P_e^i(\tau)$ *and* $P_d(\tau)$ *to calculate* $P_b^i(\tau)$
$\hat{u}^i(\tau) = \left[P_e^i(\tau), P_b^i(\tau)\right]^{\mathrm{T}}$;
// Step 5: Calculating $\left[\hat{v}(\tau+1), \hat{\mathrm{SoC}}\mathrm{n}(\tau+1)\right]^{\mathrm{T}}$ *and* $Q_f(\tau)$***:***
Running (4); // *Obtaining* $\left[\hat{v}(\tau+1), \hat{\mathrm{SoC}}(\tau+1)\right]^{\mathrm{T}}$
Running (7); // *Getting* $Q_f(\tau)$
// Step 6: Evaluating $\hat{J}^i(\tau+1)$**n :**
Producing (46); // *Utilizing* $\left[\hat{v}(\tau+1), \hat{\mathrm{SoC}}(\tau+1)\right]^{\mathrm{T}}$ *to evaluate* $\hat{J}^i(\tau+1)$
// Step 7: Calculating $E_c(\tau)$ *and* $Q_f(\tau) + \gamma \cdot \hat{J}^i(\tau+1)$**n :**
Producing (47); // *Calculating* $E_c(\tau)$
$E_a^i(\tau) = Q_f(\tau) + \gamma \cdot \hat{J}^i(\tau+1)$; // *Calculating* $Q_f(\tau) + \gamma \cdot \hat{J}^i(\tau+1)$)
$\Delta E_a(\tau) = \left|E_a^{i+1}(\tau) - E_a^i(\tau)\right|$; // $\Delta E_a(\tau)$ denotes the changing rate of $E_a^i(\tau)$
// Step 8: Optimal strategy judgment:
while $E_c(\tau) > \varepsilon_c \&\& \Delta E_a(\tau) != 0 \&\& i \leq N$
// Step 9: Updating weights of the CN and AN:
Running (48) and (49); // *Updating* $\boldsymbol{W}_{c1}$ *and* $\boldsymbol{W}_{c2}$
Running (52) and (53); // *Updating* $\boldsymbol{W}_{a1}$ *n and* $\boldsymbol{W}_{a2}$
//Then Executing Step 4∼ Step 7.
$i = i + 1$;
end while
// Step 10: Optimal strategy outputting:
while $\tau > k + N - 1$// *Output n optimal policy in* $[k, k+N-1]$
Outputting
$\{[P_e^*(k), P_b^*(k)], [P_e^*(k+1), Pn_b^*(k+1)], \cdots, [P_e^*(k+N-1), P_b^*(k+N-1)]\}$;
$[P_e^*(k), P_b^*(k)] \rightarrow$ SHL // *Only executing the first value*
end while
end for
end for

2) DETERMINATION OF PREDICTION WINDOW SIZE

After the initial weights of the HDP-MPC controller are determined, the controller can be utilized to solve the optimal strategies according to **Algorithm 1**. However, different prediction window size N can affect the convergence speed of the CN and AN in the controller and fuel consumption. Therefore, it is necessary to choose a reasonable N to ensure the real-time performance and fuel economy of the HDP-MPC strategy.

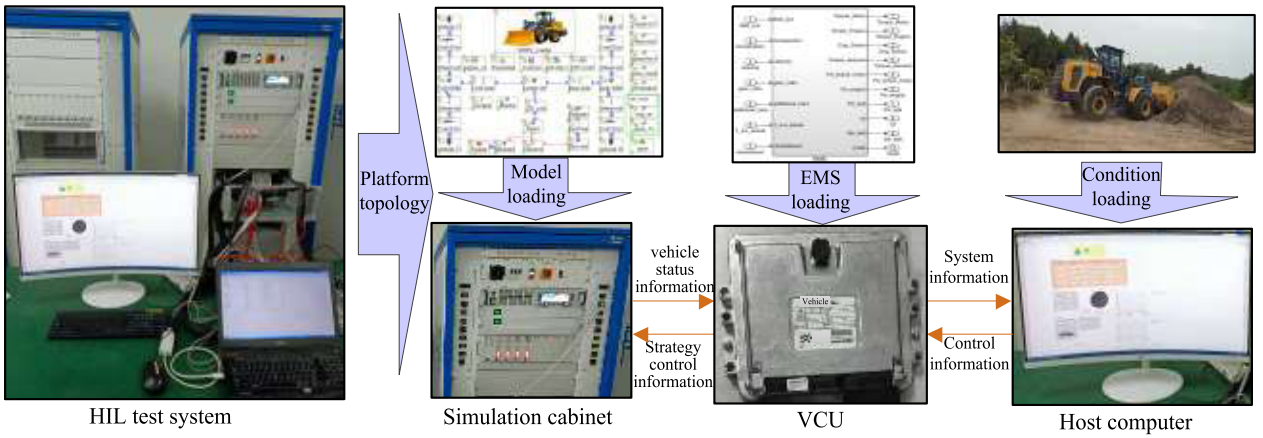

FIGURE 11. Vehicle simulation platform.

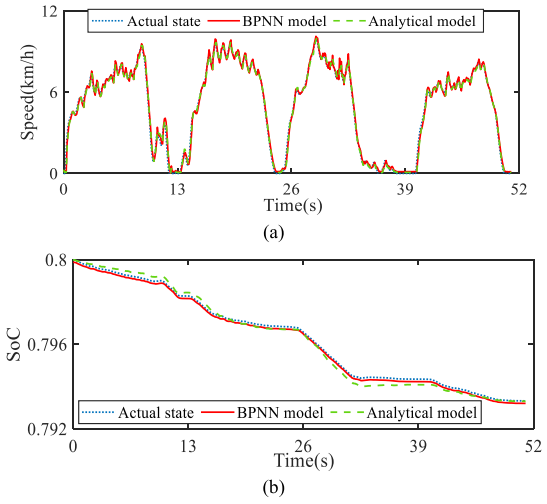

FIGURE 12. The actual and model states of the SHL. (a) Speed. (b) SoC.

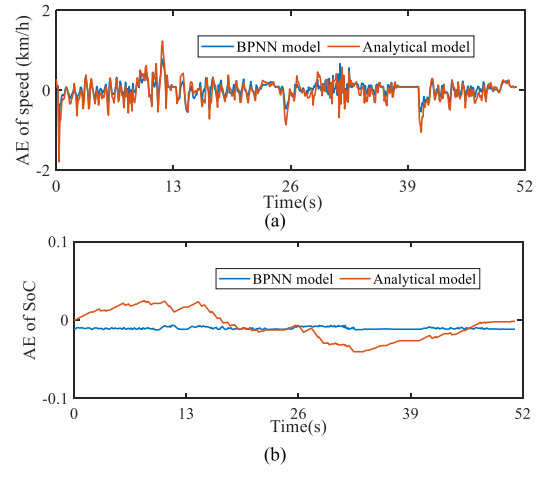

FIGURE 13. AE between the actual and model states of the SHL. (a) Speed. (b) SoC.

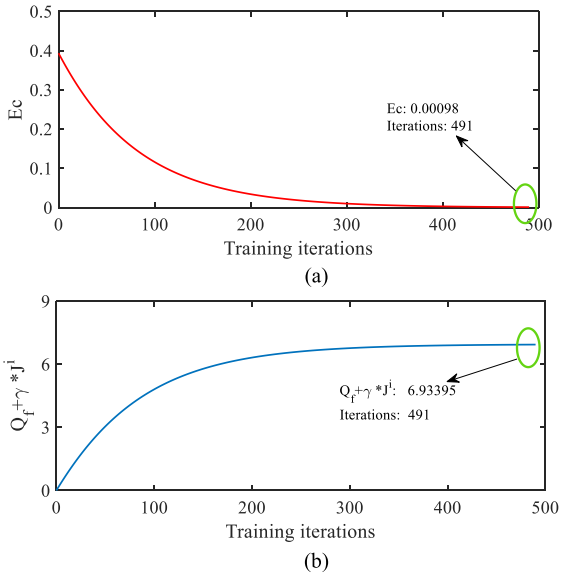

FIGURE 14. The training process of the CN and AN. (a) E_c of the CN. (b) $Q_f + \gamma \cdot \hat{J}^i$ of the AN.

TABLE 4. Initial weights of the CN and AN.

Parameters	Values
W_{c1}	$\begin{bmatrix} 1.774 & -0.535 \\ 1.874 & -1.147 \\ 0.387 & 1.491 \\ 1.360 & -2.049 \\ 1.354 & -0.561 \end{bmatrix}$
W_{c2}	$[1.930 \quad 3.024 \quad -0.868 \quad 3.513 \quad 1.566]$
W_{a1}	$\begin{bmatrix} -0.968 & 0.938 \\ -0.417 & 0.621 \\ 0.830 & -0.537 \\ -0.593 & 1.139 \\ 0.404 & -0.522 \end{bmatrix}$
W_{a2}	$[1.370 \quad 0.825 \quad -0.558 \quad 1.589 \quad 0.006]$

As a result, for obtaining a proper N to make E_c and $Q_f + \gamma \cdot \hat{J}^i$ meet the constraints for **Algorithm 1**, the convergences of the CN and AN under $N \in \{10, 20, 50, 100\}$ are respectively analyzed, as described in Fig. 15. Compared with

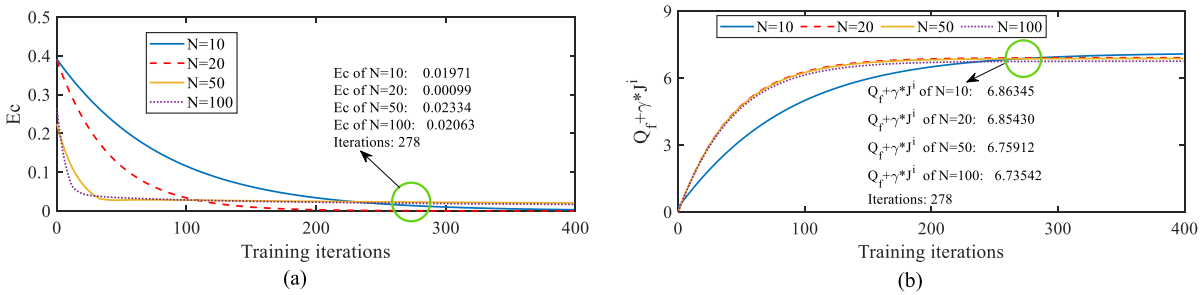

FIGURE 15. The convergence process of the CN and AN under different N. (a) E_c. (b) $Q_f + \gamma \cdot \hat{J}^i$.

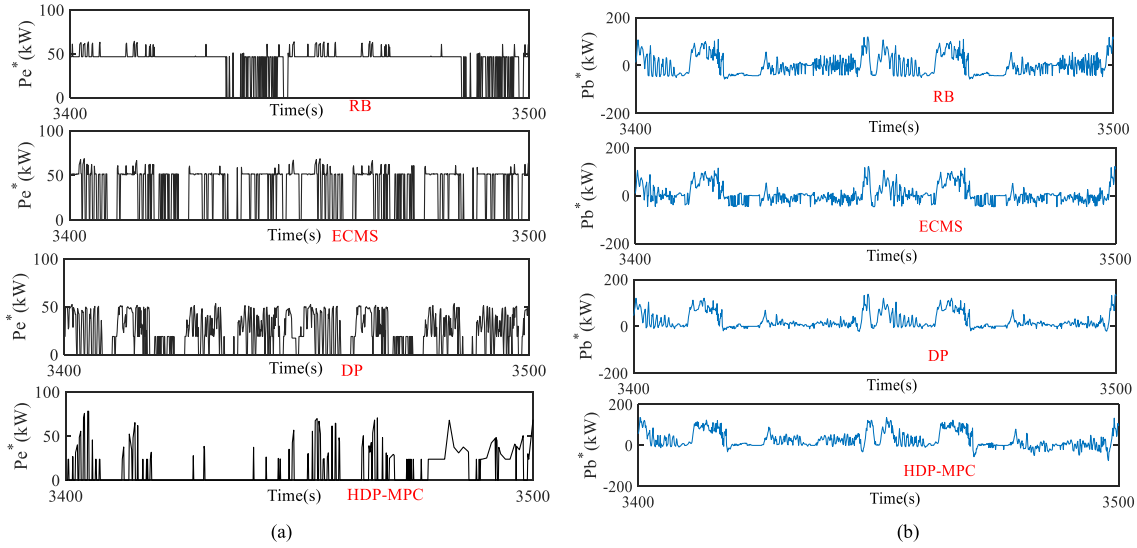

FIGURE 16. The allocated P_e^* and P_b^* of four EMSs. (a) P_e^*. (b) P_b^*.

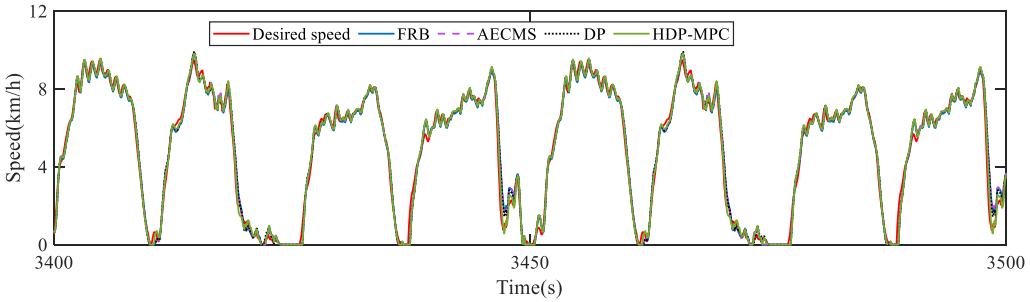

FIGURE 17. Vehicle speed curves of four EMSs.

$N \in \{10, 50, 100\}$, it is thus clear that E_c of $N = 20$ can satisfy $0.00099 \leq \varepsilon_c$ while the training iterations reach to 278. The corresponding $Q_f + \gamma \cdot \hat{J}^i$ can stabilize at 6.85430. Besides, it can be seen from Fig. 15(b) that the stable $Q_f + \gamma \cdot \hat{J}^i$ will decrease with the increasing of N, namely, the fuel consumption will reduce as N increases. Despite all this, to ensure the real-time performance of the controller execution, N is set to 20 in this paper.

3) PERFORMANCE ANALYSIS OF EMS

After determining the relevant parameters of the HDP-MPC controller, the **Algorithm 1** can be utilized to achieve real-time optimization of fuel consumption for SHL. In order to validate the effects of designed HDP-MPC strategy, the performances of fuzzy RB (FRB), adaptive ECMS (AECMS), DP and HDP-MPC EMSs are compared and analyzed under the same experimental conditions. In this paper, taking the stone operation scenario shown in Fig. 3 as the simulation environment, the energy consumption experiment of 70 cycles are carried out. In terms of the capacity and characteristics of lithium iron phosphate battery used in the studied SHL, the starting value SoC_{max} and termination value SoC_{min} of the battery SoC are separately set to 0.8 and 0.3. Considering too much cycles in the whole simulation, here,

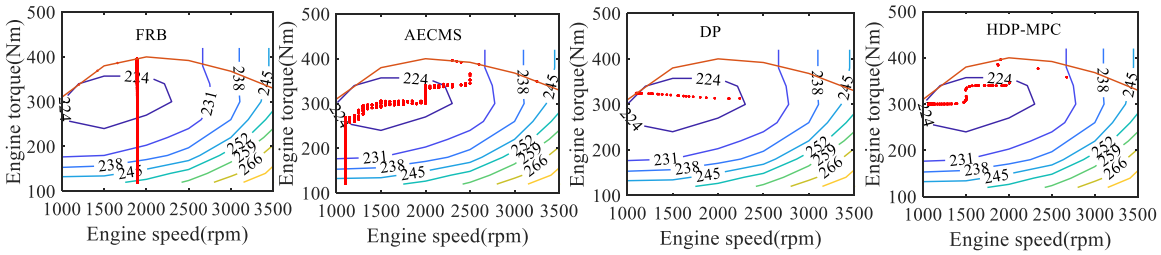

FIGURE 18. Working point distribution of the engine for four EMSs.

only the corresponding P_e^* and P_b^* curves of the four EMSs are displayed under two cycles, as shown in Fig. 16.

First of all, while the SHL performs the allocated P_e^* and P_b^* of the four strategies, the speed tracking effects of two cycles are indicated in Fig. 17. It follows that the four strategies can achieve the effective tracking of the target speed.

And then, as known from Fig.16, because of the different optimization methods, the four strategies make the changing profiles of P_e^* and P_b^* different. Correspondingly, the distribution of engine operation points is also different, as shown in Fig. 18. Compared with the FRB and AECMS EMSs, the operation points of HDP-MPC are more concentrated in the high-efficiency region, which are very close to those of DP strategy based on global optimization.

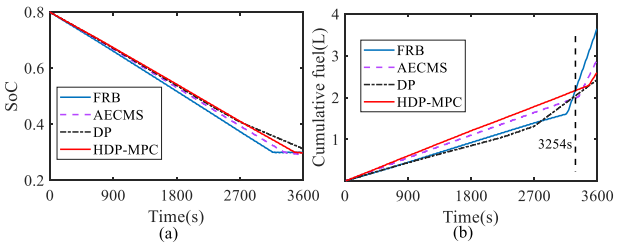

FIGURE 19. SoC and fuel profiles of four EMSs. (a) SoC. (b) fuel.

Accordingly, Fig.19 gives the changing curves of SoC and fuel consumption for the four EMSs. From here we see that although the SoC of the four EMSs can reach to the expected 0.3, their fuel optimization effects are also significantly different. TABLE 5 lists the fuel consumption of the four EMSs.

TABLE 5. The fuel consumption for the four EMSs.

EMSs	FRB	AECMS	DP	HDP-MPC
Initial SoC	0.8	0.8	0.8	0.8
Final SoC	0.3	0.3	0.3	0.3
Fuel	**3.63**	**2.89**	**2.42**	**2.59**
Simulation time of one step (s)	**0.0039**	**0.0048**	**600**	**0.0059**

Specifically, as known from Fig. 19, before 3254s, contrasted to the proposed HDP-MPC method, because the FRB and AECMS EMSs are more tend to use electricity, their fuel consumption is less than that of HDP-MPC strategy. Nevertheless, as the vehicle continues to move forward, the optimization effect of fuel consumption of the HDP-MPC strategy is gradually superior to that of FRB and AECMS EMSs because of its local optimization solution method. Compared with FRB and AECMS, it can be seen from TABLE 5 that the fuel consumption of the designed HDP-MPC strategy is reduced by 28.65 % and 10.38 %, respectively. In addition, in order to analyze the gap between the local approximate optimization strategy and the global optimization strategy in energy consumption optimization, the HDP-MPC EMS is also comparatively analyzed with the DP-based global optimization strategy. It can be clear that from Fig. 19 and TABLE 5, on the basis of both SoC falling to 0.3, thanks to global information, the fuel consumption of the DP is 6.56 % lower than the proposed HDP-MPC strategy. However, its optimization time is much longer than the online HDP-MPC strategy.

Based on the above analysis, under insuring real-time performance and velocity following effect, although the online EMS based on HDP-MPC has slightly higher fuel consumption than DP strategy, it possesses excellent online adjustment ability. Besides, compared with the FRB and AECMS EMSs, the HDP-MPC EMS can further reduce the fuel consumption of the vehicle, offering a novel solution method for online energy management of SHL.

V. CONCLUSION

This paper proposed a HDP online EMS based on predictive control to further improve the fuel economy and working condition adaptability of the SHLs. Specifically, the main work of this paper is as follows: (1) A MPC-based HDP approximate optimal algorithm was proposed, and its convergence and the controlled system stability were proved, respectively. (2) The energy management controller for SHL was designed by using the proposed algorithm. Moreover, an online EMS based on HDP-MPC was proposed to perform the energy management controller. (3) The BPNN dynamics model and online HDP-MPC strategy were validated under the stone working scenario. By the experimental results analysis, the BPNN model can more effectively reflect the actual motion process of the SHL with higher accuracy, contrasted with analytic model. Although the fuel consumption of HDP-MPC EMS is slightly 6.56% higher than DP strategy, it can optimize the fuel consumption of the vehicle real-time. In addition, compared with the FRB and AECMS, the proposed HDP-MPC strategy can further reduce the fuel consumption, offering a new idea for online energy consumption optimization of SHL.

In addition, there are some random disturbance factors that affect the robustness of the system, such as ambient temperature, tire pressure, and slope. In the future, fully considering random disturbance factors, the further research will be conducted based on the proposed system architecture of this paper, designing a robust HDP-MPC strategy to solve the approximate optimal control problem of the nonlinear system with random disturbance.

ACKNOWLEDGMENT

The authors really thank the anonymous reviewers for their careful reading the manuscript, comments and suggestions.

REFERENCES

[1] S. Iqbal, A. Xin, M. U. Jan, M. A. Abdelbaky, H. U. Rehman, S. Salman, S. A. A. Rizvi, and M. Aurangzeb, "Aggregation of EVs for primary frequency control of an industrial microgrid by implementing grid regulation & charger controller," *IEEE Access*, vol. 8, pp. 141977–141989, 2020.
[2] M. U. Jan, A. Xin, M. A. Abdelbaky, H. U. Rehman, and S. Iqbal, "Adaptive and fuzzy PI controllers design for frequency regulation of isolated microgrid integrated with electric vehicles," *IEEE Access*, vol. 8, pp. 87621–87632, 2020.
[3] Z.-M. Tong, J.-Z. Miao, Y.-S. Li, S.-G. Tong, Q. Zhang, and G.-R. Tan, "Development of electric construction machinery in China: A review of key technologies and future directions," *J. Zhejiang Univ.-Sci. A*, vol. 22, no. 4, pp. 245–264, Apr. 2021.
[4] P. Yadav, V. K. Saini, A. S. Al-Sumaiti, and R. Kumar, "Intelligent energy management strategies for hybrid electric transportation," in *Proc. IEEE IAS Global Conf. Renew. Energy Hydrogen Technol.*, Mar. 2023, pp. 1–7.
[5] Z. E. Liu, Q. Zhou, Y. Li, S. Shuai, and H. Xu, "Safe deep reinforcement learning-based constrained optimal control scheme for HEV energy management," *IEEE Trans. Transport Electrific.*, vol. 9, no. 3, pp. 4278–4293, Sep. 2023.
[6] S. Iqbal, A. Xin, M. U. Jan, S. Salman, A. U. M. Zaki, H. U. Rehman, M. F. Shinwari, and M. A. Abdelbaky, "V2G strategy for primary frequency control of an industrial microgrid considering the charging station operator," *Electronics*, vol. 9, no. 4, p. 549, Mar. 2020.
[7] Q. Wen, F. Wang, M. Cheng, B. Xu, and Z. Sun, "Adaptive equivalent consumption minimization strategy for off-road hydraulic hybrid vehicles: A cycle-to-cycle optimization approach," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2346–2357, Mar. 2022.
[8] J. Liu, Y. Chen, W. Li, F. Shang, and J. Zhan, "Hybrid-trip-model-based energy management of a PHEV with computation-optimized dynamic programming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 338–353, Jan. 2018.
[9] J. Wu, J. Ruan, N. Zhang, and P. D. Walker, "An optimized real-time energy management strategy for the power-split hybrid electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 3, pp. 1194–1202, May 2019.
[10] J. Liu, Y. Liang, Z. Chen, and W. Chen, "Energy management strategies for hybrid loaders: Classification, comparison and prospect," *Energies*, vol. 16, no. 7, p. 3018, Mar. 2023.
[11] Z. Fu, H. Wang, F. Tao, B. Ji, Y. Dong, and S. Song, "Energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles using deep reinforcement learning with action trimming," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7171–7185, Jul. 2022.
[12] I. Shafikhani and J. Åslund, "Analytical solution to equivalent consumption minimization strategy for series hybrid electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2124–2137, Mar. 2021.
[13] M. U. Jan, A. Xin, H. U. Rehman, M. A. Abdelbaky, S. Iqbal, and M. Aurangzeb, "Frequency regulation of an isolated microgrid with electric vehicles and energy storage system integration using adaptive and model predictive controllers," *IEEE Access*, vol. 9, pp. 14958–14970, 2021.
[14] L. Li, S. You, C. Yang, B. Yan, J. Song, and Z. Chen, "Driving-behavior-aware stochastic model predictive control for plug-in hybrid electric buses," *Appl. Energy*, vol. 162, pp. 868–879, Jan. 2016.
[15] J. Han, H. Shu, X. Tang, X. Lin, C. Liu, and X. Hu, "Predictive energy management for plug-in hybrid electric vehicles considering electric motor thermal dynamics," *Energy Convers. Manage.*, vol. 251, Jan. 2022, Art. no. 115022.
[16] Y. Wu, H. Tan, J. Peng, H. Zhang, and H. He, "Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus," *Appl. Energy*, vol. 247, pp. 454–466, Aug. 2019.
[17] X. Zhang, L. Yang, X. Sun, Z. Jin, and M. Xue, "ECMS-MPC energy management strategy for plug-in hybrid electric buses considering motor temperature rise effect," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 1, pp. 210–221, Mar. 2023.
[18] J. Liu, Y. Liang, Z. Chen, and H. Yang, "An ECMS based on model prediction control for series hybrid electric mine trucks," *Energies*, vol. 16, no. 9, p. 3942, May 2023.
[19] S. Xie, T. Liu, H. Li, and Z. Xin, "A study on predictive energy management strategy for a plug-in hybrid electric bus based on Markov chain," *Automot. Eng.*, vol. 40, no. 8, pp. 871–877, Aug. 2018.
[20] S. Xie, X. Hu, Z. Xin, and J. Brighton, "Pontryagin's minimum principle based model predictive control of energy management for a plug-in hybrid electric bus," *Appl. Energy*, vol. 236, pp. 893–905, Feb. 2019.
[21] H. Wang, Y. Huang, A. Khajepour, and Q. Song, "Model predictive control-based energy management strategy for a series hybrid electric tracked vehicle," *Appl. Energy*, vol. 182, pp. 105–114, Nov. 2016.
[22] T. Li, H. Liu, and D. Ding, "Predictive energy management of fuel cell supercapacitor hybrid construction equipment," *Energy*, vol. 149, pp. 718–729, Apr. 2018.
[23] C. Xiang, F. Ding, W. Wang, and W. He, "Energy management of a dual-mode power-split hybrid electric vehicle based on velocity prediction and nonlinear model predictive control," *Appl. Energy*, vol. 189, pp. 640–653, Mar. 2017.
[24] P. Shen, Z. Zhao, X. Zhan, J. Li, and Q. Guo, "Optimal energy management strategy for a plug-in hybrid electric commercial vehicle based on velocity prediction," *Energy*, vol. 155, pp. 838–852, Jul. 2018.
[25] G. Jinquan, H. Hongwen, P. Jiankun, and Z. Nana, "A novel MPC-based adaptive energy management strategy in plug-in hybrid electric vehicles," *Energy*, vol. 175, pp. 378–392, May 2019.
[26] N. Yang, S. Ruan, L. Han, H. Liu, L. Guo, and C. Xiang, "Reinforcement learning-based real-time intelligent energy management for hybrid electric vehicles in a model predictive control framework," *Energy*, vol. 270, May 2023, Art. no. 126971.
[27] Y. Wang, K. Li, X. Zeng, B. Gao, and J. Hong, "Investigation of novel intelligent energy management strategies for connected HEB considering global planning of fixed-route information," *Energy*, vol. 263, Jan. 2023, Art. no. 125744.
[28] X. Tang, J. Chen, H. Pu, T. Liu, and A. Khajepour, "Double deep reinforcement learning-based energy management for a parallel hybrid electric vehicle with engine start–stop strategy," *IEEE Trans. Transport. Electrific.*, vol. 8, no. 1, pp. 1376–1388, Mar. 2022.
[29] J. Chen, H. Shu, X. Tang, T. Liu, and W. Wang, "Deep reinforcement learning-based multi-objective control of hybrid power system combined with road recognition under time-varying environment," *Energy*, vol. 239, Jan. 2022, Art. no. 122123.
[30] G. Li and D. Görges, "Ecological adaptive cruise control and energy management strategy for hybrid electric vehicles based on heuristic dynamic programming," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3526–3535, Sep. 2019.
[31] G. Li and D. Görges, "Fuel-efficient gear shift and power split strategy for parallel HEVs based on heuristic dynamic programming and neural networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9519–9528, Oct. 2019.
[32] W. Tang, Y. Wang, X. Jiao, and L. Ren, "Hierarchical energy management strategy based on adaptive dynamic programming for hybrid electric vehicles in car-following scenarios," *Energy*, vol. 265, Feb. 2023, Art. no. 126264.
[33] Y. Wang and X. Jiao, "Dual heuristic dynamic programming based energy management control for hybrid electric vehicles," *Energies*, vol. 15, no. 9, p. 3235, Apr. 2022.
[34] M. T. Hagan, H. B. Demuth, and B. Mark, *Neural Network Design*. Beijing, China: China Machine Press, 2002.
[35] J. Liu, Y. Chen, J. Zhan, and F. Shang, "Heuristic dynamic programming based online energy management strategy for plug-in hybrid electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4479–4493, May 2019.
[36] J. Liu, Y. Chen, J. Zhan, and F. Shang, "An on-line energy management strategy based on trip condition prediction for commuter plug-in hybrid electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3767–3781, May 2018.
[37] J. Liu, Y. Liang, Z. Chen, H. Li, W. Zhang, and J. Sun, "A double-layer vehicle speed prediction based on BPNN-LSTM for off-road vehicles," *Sensors*, vol. 23, no. 14, p. 6385, Jul. 2023.

JICHAO LIU received the B.S. degree in communication engineering from Anhui Jianzhu University, Hefei, China, in 2010, and the M.S. degree in control engineering from Inner Mongolia University of Technology, Hohhot, China, in 2014, and the Ph.D. degree in control science and engineering from Beijing University of Technology, Beijing, China, in 2019. He is currently a Leader with the New Energy Research Institute, Jiangsu XCMG Research Institute Company Ltd. His research interests include electric drive systems and energy management strategy of hybrid electric vehicles.

YANYAN LIANG received the B.S. degree in vehicle engineering from Shandong Agricultural University, Taian, China, in 2016, and the M.S. degree in vehicle engineering from Jiangsu University, Zhenjiang, China, in 2019. She is currently a Technical Research Engineer with the New Energy Research Institute, Jiangsu XCMG Research Institute Company Ltd. Her research interests include power system matching and energy management of new energy vehicles.

KA XUE received the B.S. and M.S. degrees in mechanical engineering from Wuhan University of Technology, Wuhan, China, in 2010 and 2012, respectively. He is currently an Engineer with the New Energy Research Institute, Jiangsu XCMG Research Institute Company Ltd. His research interests include hybrid system integration and control strategies of construction machinery.

ZHENG CHEN received the Ph.D. degree from the State Key Laboratory of Solidification Processing, Northwestern Polytechnical University, China, in 2009. He is currently a Professor with the School of Materials and Physics, China University of Mining and Technology (CUMT). His research interests include non-equilibrium solidification of undercooled melt, models, and experiments.