A Self-adaptive Energy Management Strategy for Plug-in Hybrid Electric Vehicle based on Deep Q Learning

Runnan Zou¹, Yuan Zou¹*, Yanrui Dong² and Likang Fan³

¹National Engineering Lab for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081
²Beijing University of Technology, Beijing, 100124
³School of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081

* zouyuanbit@163.vip.com

Abstract. With the development of energy management, deep learning-based algorithm has become a widely concerned strategy. The presetting of neural network is deemed as a key of effectiveness of the method. For the purpose of improving fuel economy of plug-in hybrid electric vehicle (PHEV) based on the deep Q learning, an auto-adaptive energy management strategy is proposed in this paper. In order to obtain an optimal learning rate which is one of the key hyper parameter for deep Q network, deep Q learning (DQL) with normalized advantage function (NAF) and genetic algorithm (GA) is combined together. The improvement of optimized learning rate is verified by comparing optimized learning rate with different other learning rates. Simulation results proves the optimized learning rate achieves the best improves fuel economy of PHEV compared with other sets of learning rate. The result indicates the effectiveness of GA in finding an optimal hyper parameter and the effectiveness GA-NAF-DQL in fuel saving in PHEV.

1. Introduction

With the development of plug-in hybrid electric vehicle (PHEV), its advantage on lowering emission and saving fuel has attracted worldwide attention [1]. PHEV is combined by internal combustion engine (ICE), power battery and relatively simple transmission structure. The power distribution of PHEV brings a more complicated problem of energy split strategy between ICE and battery. Traditionally, the energy management strategy (EMS) is divided into two categories, rule-based and optimization-based [2].

The rule-based EMS is acquired by experience of skilled engineers based on experiments and analysis of each assembly. With the advantage of high executability on the online controller, this method has been utilized by Nashat Jalil, Jiankun Peng and Mohamed Elhedi Hmidi [3-5]. The rule-based energy management is capable to be implemented on PHEV while the optimization output is far from the optimal from time to time. Most energy management strategy with predefined rules that were not adaptive to complicated driving conditions online [6]. The optimization-based strategy contains dynamic programming, genetic algorithm, neural network and etc. DP was carried out in EMS in a parallel hybrid vehicle with the method of cloud-computation [7]. A cloud-computing based platform was proposed to facilitate DP which cost heavy computation burden. Genetic algorithm was applied on optimizing vital parameters [8,9,10]. However, the relatively long computational time of DP and GA was the barrier for real-time application. Therefore, DP and GA were applied in offline training.
while DP was deemed as benchmark for comparison.

Recent years, with great computational improvement compared to DP and self-adaptiveness compared with rule-based strategy, more deep learning method was applied in PHEV energy management control [11-15]. The characteristic of self-learning helps energy management adapt different working condition. Compared with DP and GA, deep learning method has higher adaptivity of different working condition and faster computational speed. With the improvement of algorithm based on deep learning, hyper parameters which were inputted into neural network and govern the function of the neural network showed crucial influence toward neural network training and they were difficult to be tuned by standard optimization method [16]. Grid research was the common method for hyper parameter optimization [17,18]. However, there were still blank on the optimization of deep reinforcement learning on the energy management of PHEV.

The purpose of this paper is investigating a novel hyper parameter optimization method for a deep Q network with normalized advantage function (NAF-DQL). Targeting at improving fuel economy of PHEV, the hyper parameter of NAF-DQL is need to be tuned to an optimal solution. The optimization object is learning rate of NAF-DQL and the optimization method is genetic algorithm. Learning rate is encoded as binary code and then inputted into the genetic algorithm. With inheritance and mutation, genetic algorithm outputs the gene suits the “environment” most. With this chromosome, NAF-DQL can obtain the best fuel economy.

The fuel economy is improved by genetic algorithm after tuning the hyper parameter. Moreover, the time cost for convergence decreased under the optimized learning rate. With self-adaptive hyper parameter tuning, NAF-DQL training is improved on aspects of fuel economy and time consumption.

2. PHEV modelling and NAF-DQL network

2.1. Vehicle configuration and PHEV model

A tracked plug-in hybrid electric vehicle which is designed for complicated working is condition is shown in Figure 1.

![Figure 1. The topology of PHEV chassis](image)

The modelling of PHEV contains three parts: engine generator set (EGS), battery and power demand. The EGS and battery is linked by AC-DC which transfers alternating current to direct current. Without considering the internal resistance drop and the loss of torque, the relationship among the output voltage of generator output voltage $U_g$, output current of generator $I_g$, output torque of engine $T_{eng}$ and magnetic torque $T_g$ is:

\[
\begin{align*}
U_g &= K_e n_g - K_s n_g I_g \\
T_g &= K_s I_g - K_s I_g^2 \\
\frac{T_{eng}}{I_g} - T_g &= 2\pi \left( \frac{J_e}{r_e + J_g} \right) \frac{dn_g}{dt} \\
n_g &= i_{eg} n_{eng}
\end{align*}
\]

Where $K_e$ represents the coefficient of the electromotive force, $K_s = 3PL^2/\pi$ is the equivalent impedance coefficient of generator, $L^2$ is the generator armature synchronous inductor. $J_e$ and $J_g$ are
the moment of inertia of the engine and generator respectively. \( i_{\text{eg}} \) is the main reduction ratio between the engine and generator set. The modeling of battery utilizes internal resistance model as follows:

\[
\begin{align*}
U_b &= V_{\text{bat}} - I_b R_{\text{bat int}} \\
P_b &= V_{\text{bat}} I_b - I_b^2 R_{\text{bat int}} \\
S\text{OC} &= \frac{V_{\text{bat}} - \sqrt{V_{\text{bat}}^2 - 4 R_{\text{bat int}} P_b(t)}}{2 C_b R_{\text{bat int}}} \\
I_b &= \frac{V_{\text{bat}} - \sqrt{V_{\text{bat}}^2 - 4 R_{\text{bat int}} P_b}}{2 R_{\text{bat int}}}
\end{align*}
\] (2)

Where \( V_{\text{bat}} \) denotes the open circuit voltage; \( I_b \) is the battery output current; \( R_{\text{bat int}} \) is the battery internal resistance; \( C_b \) is the total volume of the battery pack. The PHEV is considered as a rigid body. As a closed loop, the PHEV system receives order from VCU and sends feedback as vehicle state towards VCU the next moment. The propulsion power that need to be generated contains heading power and steering power.

\[
P_{\text{demand}} = (F_r + F_i + F_u + F_s) v_{\text{ave}} + M_j \omega
\] (3)

2.2. NAF-DQL network structure

NAF-DQL is an optimization based on deep Q learning. With the normalized advantage function (NAF), deep Q network can output an optimal continuous solution after training in several episodes. The structure of NAF-DQL is shown in Figure 2.

![Figure 2. The structure of NAF-DQL](image)

For the goal of energy management is to minimize the total fuel cost, the reward function of NAF-DQL is built as follows:

\[
R(s,a) = -\left[ \int_0^T \alpha f_{\text{fuel}}(t) dt + \beta [S\text{OC}(t) - S\text{OC}(t_0)]^2 \right]
\] (4)

In the network, the input is the state of vehicle where is defined as state of charge (SOC), rotation speed of generator and power demand.

\[
\text{state} = [n_s, S\text{OC}, P_{\text{demand}}]^T
\] (5)

The state value is inputted into the input layer where the number of input layer node is 3. The data that acquired by input layer is transferred to hidden layer H1. The data is passed by H1 towards hidden layer H2 by the Sigmoid activation function as follows:

\[
\begin{align*}
\theta_j^{(1)}(t) &= \sum_{i=1}^{N_{\text{ht}}} \theta_j^{(1)}(t) x_i(t) + b_j^{(1)}, j = 1,2, \cdots, N_{\text{ht}} \\
b_j^{(1)}(t) &= \text{Sigmoid}\left( g_j^{(1)}(t) \right) = \frac{1}{1 + e^{-g_j^{(1)}(t)}}
\end{align*}
\] (6)
\( N_{h1} \) denotes the neuron number of hidden layer \( H_1 \). The second hidden layer \( H_2 \) is the same as the first hidden layer. The normalized advantage function is established based on \( V(s) \), \( \mu \), and \( L \) that outputted by output layer. The advantage function is defined as follows:

\[
A(s,a|\theta^t) = \frac{1}{2} (a - \mu(s|\theta^t))^T P(s|\theta^t) (a - \mu(s|\theta^t))
\]  

(7)

The deep Q network is updated each episode by the loss between two networks. With the update method of gradient descent algorithm, the evaluation network is updated each episodes and the cloned to target network at set intervals. The update function is following:

\[
\begin{align*}
    w_{out}^{(l)}(t+1) &= w_{out}^{(l)}(t) + \lambda(t) \Delta w_{out}^{(l)}(t) \\
    \Delta w_{out}^{(l)}(t) &= -\frac{\partial L(w)}{\partial w_{out}^{(l)}(t)}
\end{align*}
\]  

(8)

\( \lambda(t) \) denotes the learning rate of the evaluation network which is the hyper parameter being optimized by GA. \( \partial L(w)/\partial w_{ij}^{(k)}(t) \) is the partial derivative of the loss function to weight matrix.

3. Genetic algorithm based optimization method

Genetic algorithm is a kind of heuristic search algorithm which is utilized here for the optimization of hyper parameter of NAF-DQL. With the basic idea of evolution, genetic algorithm is built for the optimization of learning rate for NAF-DQL. The combination of GA and NAF-DQL brings a self-adaptive energy management strategy for PHEV. The most important hyper parameter in NAF-DQL is learning rate. In order to adapt GA, learning rate is dispersed and encoded into a chromosome. The learning rate after discretization is between 0.1 and 0.0001. The scope of learning rate is selected based on experience. The process of genetic algorithm is consist of evolution, calculating fitness, selection, reproduction, crossover and mutation. The whole process is shown in Figure 3.

![Figure 3. The training process of GA-NAF-DQL](image)

Here, for the convenience of calculation, the binary encoding mode is utilized. In order to obtain the best learning rate for train NAF-DQL to an optimal solution, the fitness of chromosome is the reward of strategy. In each episode of training, chromosomes are decoded to decimal and inputted into NAF-DQL. After the convergence of NAF-DQL, the reward is obtained and denotes to the fitness. Chromosomes are selected by the judgement of fitness. Chromosomes with high fitness which means low reward are more likely to be chose to the next episode. Crossover and mutation are both established for the goal of generating new chromosome and expanding the area of search.

4. Result and discussion

The presented GA-NAF-DQL algorithm is evaluated in this section. The proposed strategy is aiming at optimizing fuel economy of PHEV by finding the most suitable hyper parameter. Therefore, NAF-DQL with the most optimal learning rate is compared with other NAF-DQL with unoptimized learning rate. A driving cycle is employed for training. The velocity and power demand are shown in Figure 4.
4.1. Result and comparison between different learning rates

After the training of GA-NAF-DQL, the best DNA acquired is 0.00020645. The optimized learning rate is inputted into NAF-DQL to verify the improvement of the optimized learning rate.

The training reward, loss, SOC and working points of the PHEV are shown in Figure 4. As it is presented in Figure 4(c), the reward of NAF-DQL converges to -120 gradually within 10 episodes. The loss between two networks converges to 0 after 50 steps of training. The change route of SOC is presented in Figure 4(e), SOC drops from 0.7 to 0.58 within 600 steps of training. With no rapid charge or discharge of SOC, the battery helps ICE work in the fuel economic area as it is shown in Figure 4(f). Most of working points are located in the medium fuel economic zone while some of them reach the high fuel economic zone as it is shown in Figure 4(f). The simulation result of reward, loss, SOC and working points indicates the learning rate of 0.00020645 guarantees the battery and ICE to work together meanwhile the battery discharges smoothly and ICE works in the economic area.

![Figure 4. The training result of learning rate at 0.00020645](image)

The optimized learning rate is compared with other three sets of learning rate on aspects of working points, reward and SOC that shown in Figure 5. Compared with other learning rates, more working points of learning rate at 0.00020645 locates in the fuel economic zone. Working points of learning rate at 0.01 and 0.001 locate at low fuel economic area. The working points distribution of learning rate at 0.0001 and 0.00020645 are similar but more working points of optimized learning rate are located at high fuel economic zone which leads to the result of best fuel economy compared with others.

![Figure 5. Working points and SOC of 4 sets learning rate](image)

The curve of SOC is presented in Figure 5(b). The SOC curve of the optimized learning rate is more smooth compared with other sets for the reason of more working points of the optimized learning rate in high fuel economic zone. This result indicates the optimized learning rate is more battery-protective compared with others.

5. Conclusion

An optimization method for hyper parameter in NAF-DQL in energy management of PHEV is put forward in this paper. A model consists of vehicle forward dynamic model and components model is built. The optimization algorithm based on GA and NAF-DQL is proposed. The evolved NAF-DQL reaches the best fuel economy in the energy management of PHEV.
Some conclusions and remarks are drawn as follows:

1. Based on the genetic algorithm, the framework of GA-NAF-DQL was carried out for the optimization of hyper parameter in energy management in PHEV. Function of selection, crossover and mutation were provided in GA-NAF-DQL.

2. The improvement of the optimized learning rate was investigated by comparing with other sets of learning rate. The optimized learning rate achieved the best fuel economy compared with other sets.

3. The improvement of the optimized learning rate acted the best on protecting battery on smooth charge and discharge compared with other sets.

References

[1] Liu C, Guo J, Zhang W and Liu K 2019 Research on AECMS energy management strategy of PHEV ITNEC 2019.

[2] Xie S, Hu X, Xin Z and Brighton J 2019 Pontryagin’s minimum principle based model predictive control of energy management for a plug-in hybrid electric bus Applied Energy 236 893–905.

[3] Jalil N, Kheir N A and Salman M 1997 A rule-based energy management strategy for a series hybrid vehicle 1997.

[4] Peng J, Fan H, He H and Pan D 2015 A rule-based energy management strategy for a plug-in hybrid school bus based on a controller area network bus Energies 8(6) 5122-5142.

[5] Hmidi M E, Salem I B and Amraoui L E 2019 Canonical particle swarm optimization algorithm based a hybrid vehicle New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins

[6] Hu Y, Li W, Xu K, Zahid T, Qin F and Li C 2018 Energy management strategy for a hybrid. electric vehicle based on deep reinforcement learning Applied Sciences 8(2) 187.

[7] Fan J, Zou Y, Kong Z and Heide L 2019 Cloud-computation based optimal driving for a parallel hybrid electric vehicle Journal of Beijing Institute of Technology 1

[8] Panday A and Bansal H O 2015 Optimal fuel control of series-parallel input split hybrid electric vehicle using genetic algorithm based control strategy ICEEE, IEEE 1-5

[9] Zhu T, Li B, Zong C and Wu Y 2017 Series hybrid electric vehicle power system optimization based on genetic algorithm IOP Conference 235 12-13

[10] Liu T, Yu H, Guo H and Zou Y 2018 Online energy management for multimode plug-in hybrid electric vehicles IEEE Transactions on Industrial Informatics 1-1

[11] Hu C, Yang C, Shi X, Zou R, Liu L, Chen H 2020 Investigation of rotating stall in radial vanefess diffusers with asymmetric inflow Aerospace Science and Technology 96 105546:1-13

[12] Hu Y, Li W, Xu H and Xu G 2015 An online learning control strategy for hybrid electric vehicle. based on fuzzy Q-learning Energies 8(10) 11167-11186

[13] Zou Y, Liu T, Liu D, Liu D and Sun F 2016 Reinforcement learning-based real-time energy management for a hybrid tracked vehicle Applied Energy 171(June) 372-382

[14] Fang Y, Song C, Xia B and Song Q 2015 An energy management strategy for hybrid electric bus based on reinforcement learning Proceedings of the 2015 27th Chinese Control and Decision Conference, CCDC 2015, IEEE 2015 4973-4977

[15] Xie S, Hu X, Qi S and Lang K 2018 An artificial neural network-enhanced energy management strategy for plug-in hybrid electric vehicles Energy 163(NOV.15) 837-848

[16] Liessner R, Schmitt J, Diernemann A and Baker B 2019 Hyperparameter optimization for deep reinforcement learning in vehicle energy management ICAART 2019

[17] Mania H, Guy A, Recht B 2018 Simple random search provides a competitive approach to reinforcement learning CoRR abs/1803.07055

[18] Barsce J C, Palombarini J A and Martinez E C 2018 Towards autonomous reinforcement learning: automatic setting of hyper-parameters using Bayesian optimization CoRR abs/1805.04748