International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 248

DEEP REINFORCEMENT LEARNING BASED ENERGY MANAGEMENT OF HYBRID ELECTRIC VEHICLES WITH EXPERT KNOWLEDGE

Renzong Lian[1], Jiankun Peng[1*], Yuankai Wu[1], Huachun Tan[2], Hongwen He[1], Jingda Wu[1]

1 School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

2 School of Transportation, Southeast University, Nanjing 211102, China

ABSTRACT

Reinforcement learning for energy management of hybrid electric vehicles has become a research hotspot. In this paper, a deep reinforcement learning (DRL) based energy management strategy (EMS) combined with expert knowledge is proposed, and an improved framework of deep deterministic policy gradient is adopted. In order to realize a reasonable tradeoff in the EMS, a multi-objective function of the fuel consumption and the battery charge-sustaining is established. In terms of action space of DRL, simplified action space, i.e. the optimal brake specific fuel consumption (BSFC) curve, is applied to the engine, thereby improving the sampling efficiency of DRL. The simulation results demonstrate that the expert knowledge can improve fuel economy and speed up convergence efficiency of the DRL based EMSs.

Keywords: Energy management strategy, Hybrid electric vehicle, Weight assignment, Expert knowledge, Deep deterministic policy gradient.

NONMENCLATURE

Abbreviations	
HEVs	Hybrid electric vehicles
DDPG	Deep deterministic policy gradient
DRL	Deep reinforcement learning
DP	Dynamic programming
EMSs	Energy management strategies
SOC	State of charge
BSFC	Brake Specific Fuel Consumption
Symbols	
n_g, n_{out}, n_e	Speed of generator, ring gear, engine
T_g, T_{out}, T_e	Torque of generator, ring gear, engine
fuel	Fuel consumption
V_{oc}	Open-circuit voltage
P_{bat}	The output power of battery
R_{int}	The internal resistance of charge and discharge
V_{fuel}	Fuel consumption
$SOC(t)$	The state of charge at time t
SOC_0	The charge-sustaining reference value of SOC

1. INTRODUCTION

The electrification of the automobile has recently become a trendy topic. However, the development of electric vehicles encounters bottlenecks because of the technical challenges on the power battery and fuel cell. Given that, hybrid electric vehicles (HEVs) are regarded as an important role in reducing emissions within the current infrastructure [1]. In general, HEVs contain two or multiple power sources, and therefore the energy management system is an indispensable component for HEVs. Through appropriate strategies, it is able to efficiently operate multiple power sources, thereby reducing fuel consumption and greenhouse gas emissions.

Existing energy management strategies (EMSs) are less adaptable to complex driving schedule or take too much time to optimize. Reinforcement learning (RL) algorithms have offered an alternative solution for the challenging control problem under both virtual and real-world environments [2]. In the field of energy management of HEVs, related works in [3-5] have also shown that RL, such as Q learning, deep Q learning and DDPG, have a strong learning ability and adaptability under complex driving cycles, and consume less

Selection and peer-review under responsibility of the scientific committee of the 11th Int. Conf. on Applied Energy (ICAE2019).

computational resources. These studies show that RL approach is a potential solution for EMSs. Energy management of HEVs usually targets to minimize fuel consumption and keep battery charge-sustaining However, existing researches of DRL based EMSs almost focus on optimizing the fuel consumption and do not take into account the influence of weight assignment between energy saving and battery charge-sustaining in the objective function. So far no more systematic and in-depth research has been conducted. In addition, all the above RL algorithms are model-free and learn optimal EMS solution in a "trial-and-error" manner. Model-free RL relies on a mass of real samples from environment, which often suffers from low sampling efficiency in order to achieve better performance [2].

Given those inherent problems, human expertise of HEVs is considered and applied to the DRL algorithm in a prior knowledge form. Through incorporating prior knowledge into the multi-objective function, the weight assignment between energy saving and battery charge-sustaining is studied comprehensively. Furthermore, the optimal brake specific fuel consumption (BSFC) curve is introduced into the action space of engine, which can reduce the search space and enhance the sampling efficiency of model-free algorithm [6]. In order to avoid the dimension curse, the continuous state and action representation method, namely deep deterministic policy gradient (DDPG), is used in this research.

The research encompasses three perspectives that contribute to relevant research: 1) the DDPG-based EMS combined with expert knowledge is proposed; 2) a multi-objective reward function of EMS is established, and make a comprehensive study on the weight assignment of the multi-objective function; 3) the action space of DRL, namely the working range of engine, is significantly reduced by the optimal BSFC curve.

2. PRIUS POWERTRAIN MODEL

The Prius model is the second generation of the Toyota hybrid system, and the powertrain architecture is shown in Fig. 1. The powertrain configuration is mainly composed of a gasoline engine, a driving motor, and a generator, and equipped with a small capacity lithium battery, which can be used to drive the driving motor and the generator.

The vehicle dynamic is modeled by longitudinal force balance equation, and the driving force of the vehicle is mainly provided by the engine and motor. The core power-split component of Prius is a planetary gear. Through this structure, Prius can realize the power coupling and adapt to different driving cycles. The engine and generator are respectively connected with the planet carrier and the sun gear of the planetary gear, and the ring gear is not only linked with the motor but also fixed with the output shaft. After passing through the main reducer, the power is finally transmitted to the wheels.

The engine, generator and motor are modeled by their efficiency maps from bench experiments. The Li-ion battery is modeled by an internal resistance model. As one of the energy sources, it supplies power to the motor and generator.

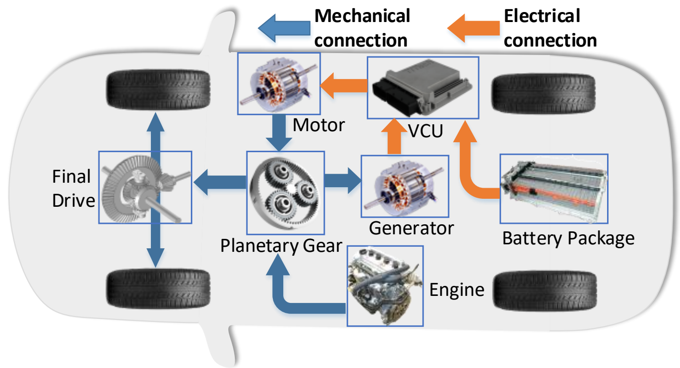

Fig. 1 The architecture of Prius powertrain

3. METHODOLOGY

3.1 Expert knowledge of HEVs

In this section, there are two kinds of expert knowledge: the optimal BSFC curve of engine and the internal resistance of battery are introduced into the DRL based EMSs. They will act as the constraints to guide the EMSs towards global optimization.

In order to achieve better fuel economy and reduce gas emissions, the engine should operate in the low-BSFC region [7]. This rule has been always incorporated into EMSs by expertise engineers. In order to make use of this expert knowledge, the action variable of DDPG based EMSs, namely the engine power, is set according to the optimal BSFC curve of engine (the red curve in engine map shown in Fig. 2). Therefore, the engine will work along the optimal BSFC curve of engine, rather than along the global engine map. The optimal BSFC curve of an engine is the minimum fuel consumption rate at a given engine power, which is used as the prior knowledge in this research. Since this method simplifies the action search space of engine, it enables the EMS controller to search optimal solution in a smaller space, and thereby reducing the computation time.

In addition to the optimal BSFC curve of engine, the internal resistance of battery is also an important influence factor. According to the characteristics of the battery resistance, the internal resistance is in a small range when the value of SOC ranges from 0.4 to 0.85, which ensures that the battery is in a high efficiency range. When the value of SOC is 0.6, the internal resistance of the battery is the minimum during charge-discharge process. In this research, the original value of SOC is set as 0.65.

3.2 DRL-based energy management strategy

DRL can be used to solve the Markov decision process, which consists of agents and environments. In Fig. 2, it shows the interaction between the agent, i.e. the energy management strategy, and environment, i.e. the vehicle and driving environment, for HEV energy management. In this research, DDPG is introduced into EMSs to learn the optimal policy of the agent, and the state variables and action variables are set as follows:

$$\begin{cases} State = \{SOC, velocity, acceleration\} \\ Action = \{engine\ power\} \\ Reward = -\{\alpha[fuel(t)] + \beta[SOC_0 - SOC(t)]^2\} \end{cases} \tag{1}$$

The reward, namely the multi-objective reward function, of the DDPG based EMSs consists of two parts, including the instantaneous fuel consumption of the engine and the battery charge-sustaining. Based on the expert knowledge of battery above, the charge-sustaining reference value of SOC, namely SOC_0, is selected as 0.6 according to the minimum internal resistance of charge and discharge. Besides, the multi-objective function must satisfy the constraints of upper and lower bounds of the battery internal resistance.

In Eq. (1), α represents the weight of fuel consumption, and β represents the weight of battery charge-sustaining. A key challenge of the multi-objective function is the weight assignment between fuel consumption and battery charge-sustaining, i.e. the configuration between α and β. Different weights between them present different results.

DDPG is an actor-critic, model-free algorithm in the field of DRL, which can operate over continuous state and action spaces. Therefore, there is no need to discretize the action spaces. The critic and actor are represented by deep neural networks, which means that DDPG uses multilayer perceptron to learn in large state and action spaces. The critic network is learned by Bellman equation as follow, and the actor is updated by applying the chain rule to the expected return in regard to the actor parameters [8]:

$$\begin{cases} y_t = r(s_t, a_t) + \gamma Q'\left(s_{t+1}, u'\left(s_{t+1}|\theta^{u'}\right)|\theta^{Q'}\right) \\ L(\theta^Q) = E[(Q(s_t, a_t|\theta^Q) - y_t)^2] \\ \nabla_\theta L(\theta^Q) = E[(r + \gamma Q'\left(s_{t+1}, a_{t+1}|\theta^{Q'}\right) \\ \qquad\qquad - Q(s_t, a_t|\theta^Q))\nabla_\theta Q(s_t, a_t|\theta^Q)] \\ \nabla_{\theta^\mu} J \approx E\left[\nabla_{\theta^\mu} Q(s, a|\theta^Q)|_{s=s_t, a=u(s_t|\theta^\mu)}\right] \\ \qquad = E[\nabla_a Q(s, a|\theta^Q)|_{s=s_t, a=u(s_t)} \nabla_{\theta^\mu} u(s|\theta^u)|_{s=s_t}] \end{cases} \tag{2}$$

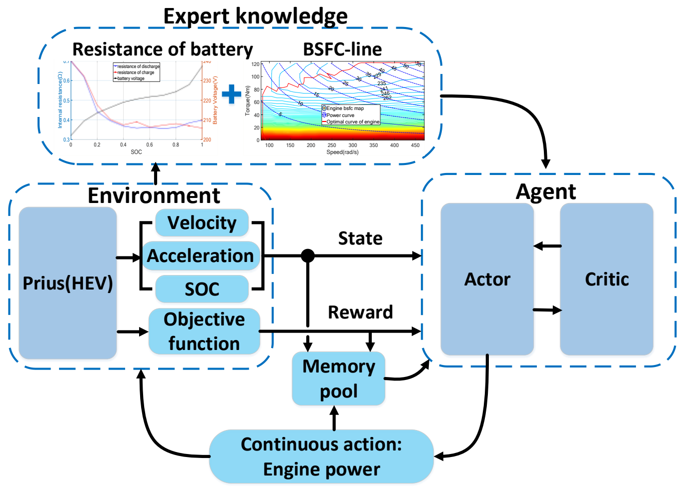

Fig. 2 Agent-Environment interactions of DDPG

4. *R*ESULT

In this section, a mixed standard driving cycle (NEDC, LA92, China city) is selected as the driving cycle to test our energy management strategy. For comparison, dynamic programming (DP) is used as a benchmark.

4.1 Weight assignment of multi-objective function

As stated in the section 3, the multi-objective function of our model contains two parts: the fuel consumption and the cost of battery charge-sustaining derived from the prior knowledge, and there must be a tradeoff between them. Hence, the weight assignment between α and β in Eq. (1) is studied in this section. For comparison purposes, α is fixed to 1, β is changed from 0.3×350 to 10×350 gradually, and 350 is determined by 100 times the maximum instantaneous fuel consumption. Moreover, the influence of terminal SOC is also taken into account in this research.

Fig. 3 shows the comparison of the SOC trajectories under different weight settings. It reveals that the higher the β, i.e. the higher weight of battery charge-sustaining, the better charge-sustaining effect the SOC trajectory shows. In order to achieve better sustainability, the engine needs to run frequently to charge the battery, which leads to a relatively higher fuel consumption. The multi-objective function with the lowest weight of β

(0.3x350) has failed to maintain SOC above the lower bound. The results in Table 1 reveal the relationship between the fuel consumption and battery charge-sustaining, and it can be seen that the system achieves the optimal balance between them with the weight of 0.6x350. Therefore, 0.6x350 is selected as the weight of battery charge-sustaining in this research.

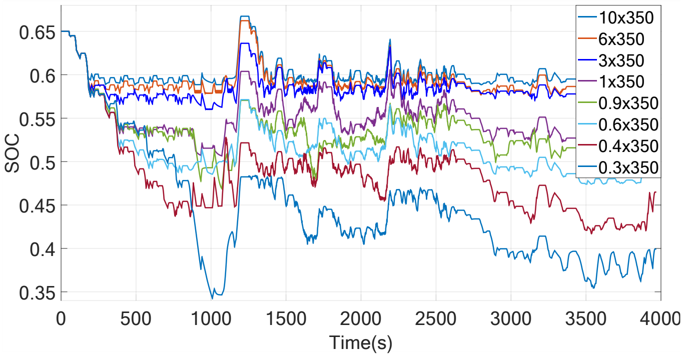

Fig. 3 The SOC trajectories of different weight settings

Table 1 The comparison of different weight settings

β	V_{fuel}(L/100Km)	Terminal SOC	V_{fuel} of DP (L/100Km)	Fuel economy
0.3 × 350	3.755	0.403	3.359	89.4%
0.4 × 350	3.704	0.466	3.428	92.5%
0.6 × 350	3.721	0.509	3.474	93.4%
0.9 × 350	3.758	0.531	3.503	93.2%
1 × 350	3.795	0.541	3.506	92.4%
3 × 350	3.925	0.590	3.563	90.8%
6 × 350	3.942	0.603	3.585	90.9%
10 × 350	3.958	0.611	3.602	91.0%

4.2 The working range of engine based on expert knowledge

In order to verify the value of prior knowledge, we conduct an experimental comparison between a DDPG model with/without prior knowledge. For the model without prior knowledge, the action variables are directly set as the engine speed and torque, which means that the agent need to explore optimal engine working point in the whole engine map. As shown in Fig. 4, it is clear that the working points of engine of our model are distributed in the area of low fuel consumption rate along the optimal BSFC curve of engine. On the contrary, the working points of engine without expert knowledge are scattered throughout the engine map. Within a limited number of training episodes, the algorithm with prior knowledge has the ability to get better fuel consumption than the algorithm without prior knowledge.

Moreover, by introducing prior knowledge into the energy management strategy, the algorithm is able to converge faster than the one without prior knowledge. In Fig. 5, the algorithm combined with the optimal BSFC curve starts to converge from the 50th episode, but the algorithm without the optimal BSFC curve does not converge until 180 episodes. The results demonstrate the effectiveness and advantage of this approach.

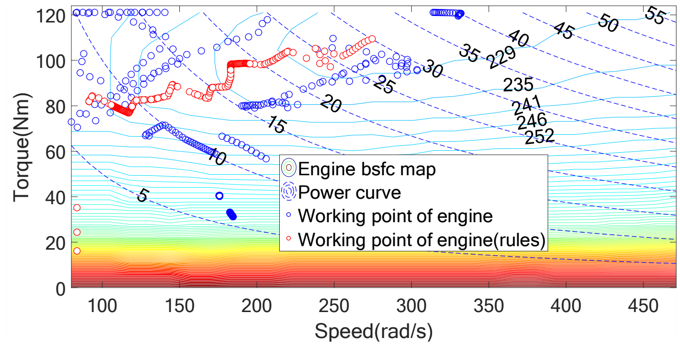

Fig. 4 The working points of engine

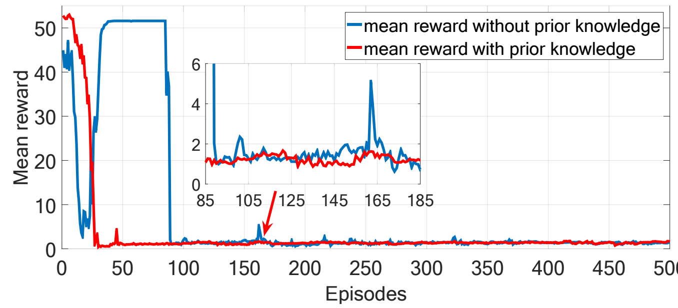

Fig. 5 The convergence speed

CONCLUSION

In this paper, we investigated a DRL based EMS of HEVs by incorporating expert knowledge. Two kinds of expert knowledge are considered: BSFC curve of HEV engine and battery internal resistance. The experimental results show that the proposed method is able to improve the performance of EMSs. By allocating weights between fuel consumption and battery charge-sustaining properly, fuel consumption is significantly reduced by up to 4%. Moreover, the simplified action space improves the convergence efficiency by 72%. In our future work, a universal multi-objective reward function will be explored, so as to adapt to different kinds of hybrid electric vehicles.

ACKNOWLEDGEMENT

This work was supported by National Key R&D Program of China (Grand No.2018YFB0105900), National Natural Science Foundation of China (Grant No.51705020 & No.61620106002) and the Project funded by China Postdoctoral Science Foundation (Grant No. 2016M600933). The authors would also like to thank

Dr. Qin Li and Dr. Hailong Zhang for their corrections and helpful suggestions.

REFERENCE

[1] CM Martinez, Hu X, Cao D et al. Energy Management in Plug-in Hybrid Electric Vehicles: Recent Progress and a Connected Vehicles Perspective[J]. IEEE Transactions on Vehicular Technology Connected Vehicles Series 2017; 66: 1-16.

[2] Schulman J, Levine S, Abbeel P, et al. "Trust region policy optimization," in ICML, 2015.

[3] Tan H, Zhang H, Peng J, et al. Energy management of hybrid electric bus based on deep reinforcement learning in continuous state and action space[J]. Energy Conversion and Management, 2019, 195: 548-560.

[4] Wu J, He H, Peng J, et al. Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus[J]. Applied Energy 2018; 222: 799-811.

[5] Wu Y, Tan H, Peng J, et al. Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus. Applied Energy 2019; 247: 454-466.

[6] Parr R, Russell S J. Reinforcement learning with hierarchies of machines. In Advances in Neural Information Processing Systems 1998; 10: 1043–1049.

[7] Zhang X, Li C-T, Kum D, et al. Configuration Analysis of Power-Split Hybrid Vehicles with a Single Planetary Gear[J]. IEEE Transactions on Vehicular Technology 2012; 61: 3544-3552.

[8] Lillicrap T P, Hunt J J, Pritzel A, et al. "Continuous Control with Deep Reinforcement Learning", in ICLR, 2016.