International Conference on Applied Energy 2021
Nov. 29 - Dec. 5, 2021, Thailand/Virtual
Paper ID: 648

Deep Reinforcement Learning Based Energy Management Strategy for Hybrid Vehicles in Consideration of Engine Start-up Conditions

Zemin Eitan Liu[1], Lubing Xu[1], Yanfei Li[1*], Bin Shuai[2], Shijin Shuai[1]

1 State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China (e-mail: zm-liu18@mails.tsinghua.edu.cn)

2 School of Mechanical Engineering, University of Birmingham, Birmingham, B15 2TT, UK

*Corresponding author: Dr Yanfei Li, liyf2018@tsinghua.edu.cn

ABSTRACT

As one of the most important supervisory control functionalities, the energy management strategy (EMS) of a hybrid electrified vehicle (HEV) optimizes the use of onboard energy resources for energy conservation and emission mitigation. Engine Start-up proved to have great contribution to fuel consumption and emission. A deep reinforcement learning based EMS is proposed for a power-split HEV to reduce the energy consumption and emission by recognizing start-up conditions and decreasing the start-up frequency. The EMSs based on Proximal Policy Optimization (PPO) and Twin-delayed Deep Deterministic Policy Gradient (TD3) are also compared in transient working condition frequency. Simulation study is conducted to demonstrate the advantage of the proposed energy management method. The EMS considering fuel consumption minimization and irrational actions avoidance is optimized by running the vehicle model under the WLTC condition repeatedly. PPO can get 9.02% lower fuel consumption, 25.6% lower start-up times and 8.2% transient working condition percentage than TD3. PPO is more suitable in the EMS domain.

Keywords: Energy Conservation and Emission Reduction; Energy Management Strategy; Deep Reinforcement Learning; Engine Start-up Conditions; Hybrid Electric Vehicle

NONMENCLATURE

Abbreviations

EMS	Energy Management Strategy
PPO	Proximal Policy Optimization
TD3	Twin-delayed DDPG

1. INTRODUCTION

In recent years, new vehicle power-trains are researched and developed by both industry and academia. The ever-increasingly strict legislations and regulations require the continuous reduction of fuel consumption [1] and pollutant emission levels [2, 3]. While electric vehicles (EVs) can be a solution, the deficiency in driving ranges, charging infrastructure, unaffordable battery cost limit the EV spreading nowadays. Hybrid Electric Vehicles (HEVs) are now a promising solution which combine the benefits of conventional combustion engines and novel electric motors. The energy management strategy (EMS) plays the most critical role for HEV CO2 emission control as a supervised strategy and much effort is made to determine the distribution of power from each source (engine or motor) simultaneously satisfying the driver's demand, minimizing energy consumption/emissions, and maintaining the battery's state-of-charge (SOC).

Recently, learning-based EMS has emerged which is based on reinforcement learning (RL) and can be applied model-free which can simplify the model establishment [4, 5]. Compared to traditional RL, which can hardly handle high dimensional problems, the combination of RL with

Selection and peer-review under responsibility of the scientific committee of the 13th Int. Conf. on Applied Energy (ICAE2021).

neural networks called deep reinforcement learning (DRL) can be used to establish EMS suitably.

There are several groups who made research of the DRL-EMS with different algorithms. Qi used Deep Q-Learning (DQL) and Dueling DQL to establish an EMS for a power-split PHEV. The action vector needs to be divided discrete to apply the algorithm [6]. Inuzuka established an EMS with Proximal Policy Optimization (PPO) and used a rule-based controller to filter the irrational control actions [7]. Lian used Deep Deterministic Policy Gradient (DDPG) and combined it with transfer learning [8, 9].

As far as the authors know, there is no research considering the engine start-up conditions which is coupled with the fuel consumption and emission worsen conditions [10]. Furthermore, most of the research applied deterministic algorithms to obtain the optimal strategy. However, the exploration ability of deterministic algorithms is poor and it will increase the fluctuation of control signals. Moreover, engine working point jump is another severe problem in DRL-based EMS, the engine transient conditions of engine can deteriorate fuel consumption and emissions, as well as the system lifetime [11]. This is the unique problem in the domain of HEV EMS, so the applicability of different DRL algorithms should be evaluated.

The main contributions of this paper are: 1) to obtain an optimal strategy in consideration of the engine start-up; 2) to compare the applicability of a stochastic algorithm and a deterministic algorithm in the EMS domain.

2. MATERIAL AND METHODS

The method used in this paper is based on i-EMS framework which is shown in Fig. 1.

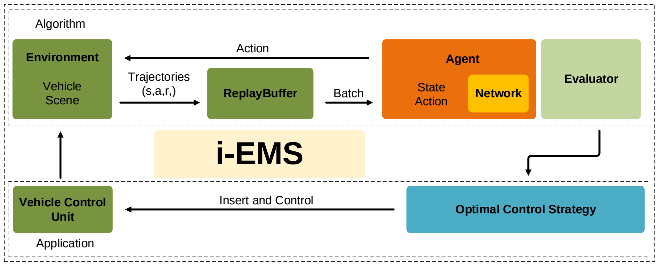


Fig. 1. The framework of i-EMS

2.1 Environment

A vehicle model is built that does not taken the mass inertia and transient process into account to train and evaluate EMS. The hybrid vehicle configuration used is the second generation of Prius THS system as seen in Fig. 2 [12] and the parameters of the vehicle are listed in Table 1.

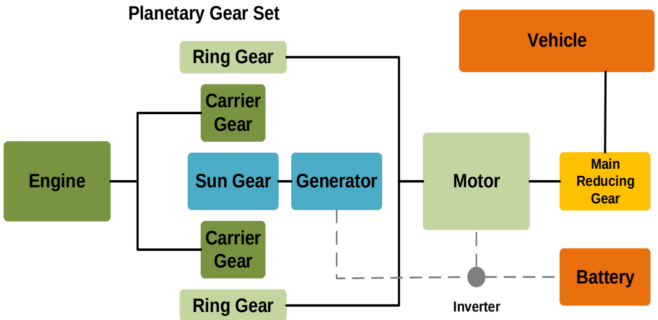


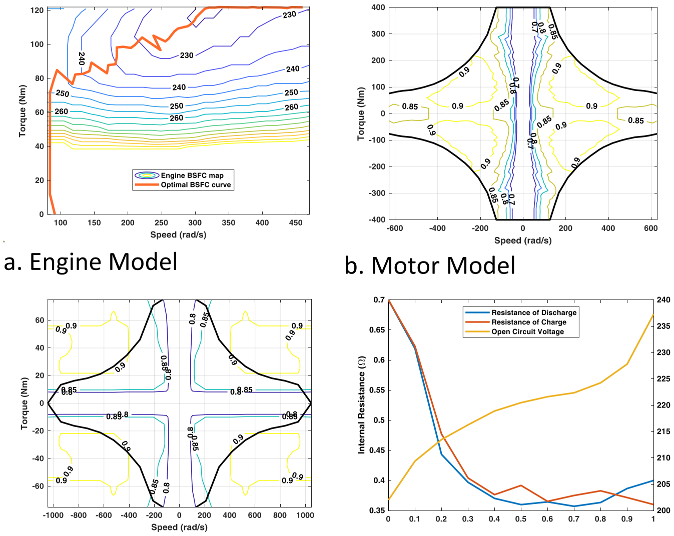
Fig. 2. The configuration of Prius THS

The vehicle dynamics include the driving resistance caused by rolling friction and aerodynamic drag. The ICE fuel consumption is modelled by a map obtained from a power-train test bench as in Fig. 3. With the speed and torque of the engine being chosen, the fuel consumption can be determined by the fuel map, the speed and torque of motor and generator can also be calculated by the gear ratio of the planetary gears[13].

Table 1. Parameters of the vehicle

	Parameters	Value
Vehicle	Curb weight	1449 kg
	Rolling resistance coefficient	0.013
	Air resistance coefficient	0.26
	Frontal area	2.23 m2
Traction motor	Maximum power	50 kW
	Maximum torque	400 Nm
	Maximum speed	6000 rpm
Generator	Maximum power	37.8 kW
	Maximum torque	75 Nm
	Maximum speed	1000 rpm
Engine	Maximum power	56 kW
	Maximum torque	120 Nm
	Maximum speed	4500 rpm
Battery	Capacity	1.54 kWh
	Voltage	237 V
Transmission	Gear ratio from ring gear to wheel	3.93
	Characteristic parameter	2.6

The models of motor and generator are the corresponding efficiency maps from bench experiments respectively. The Ni-MH battery is modelled by an equivalent circuit model ignoring the temperature change and battery aging as in Github of Lian [14] as seen in Fig. 3.



a. Engine Model b. Motor Model

c. Generator Model d. Battery Model

Fig. 3. The model of engine, motor, generator and battery

In this paper the engine start-up conditions are recognized as when the engine speed was zero in the last state and is positive in the current state and there will be a Boolean variable named *Start* in the vehicle model:

$$Start = (n_{t-1} == 0) \cap (n_t > 0) \tag{1}$$

Generally, the fuel consumption of start-up condition will be idling for 10 seconds and the emission of start-up condition will be idling for 30 seconds [15, 16]. In this paper, only fuel consumption of start-up conditions is considered. The punishment is set in the environment (i.e. set in the vehicle model) as (2) and the agent will learn to recognize and avoid frequent start-up conditions by itself through training.

$$\dot{m}_f = \dot{m}_{f,steady} + 0.9 \times Start \tag{2}$$

where $\dot{m}_f$ is the fuel consumption output of the vehicle model; $\dot{m}_{f,steady}$ is the steady fuel consumption obtained from the engine map.

The scene used to train the agent is the WLTC driving cycle because it has low, medium, high and extra high four parts to represent variable driving conditions.

2.2 Agent

The EMS is represented by a neural network, PPO [17] and TD3 [18] are adopted and compared in this paper to update parameters of the EMS as a stochastic DRL algorithms and a deterministic DRL algorithms, respectively. The process of update is based on the trajectories (state, action and reward) generated by iterative interactions between the Agent and the Environment.

With the speed and acceleration of the vehicle and the radius of wheel, the required torque and rotational speed of the wheel can be calculated. According to the THS dynamic model, the control signal selected in this paper are the speed and torque of the engine. Thus the state vector and action vector are defined as (3) and (4) in this paper.

$$s_t = (v_t,\ a_t,\ SOC_t,\ n_{t-1}) \tag{3}$$

$$a_t = (n_{t,engine},\ T_{t,engine}) \tag{4}$$

where v_t is the speed of the vehicle and a_t is the acceleration of the vehicle in current state, they are determined by the driving cycle. n_{t-1} is the engine speed in last state used to recognize the engine start-up conditions.

Reward includes the fuel consumption, the variation of SOC and the punishments for irrational actions, as in (5):

$$r_t = -\ (\alpha \cdot \dot{m}_{fuel} + \beta \cdot (SOC - 0.65)^2 + \gamma \cdot Inf) \tag{5}$$

where α，β，γ are the coefficients of fuel consumption, variation of SOC and the punishments, respectively.

The specific definition and setting of the variable *Inf* can be seen in other paper of the authors, it will not be explained here.

3. RESULTS

3.1 The strategy performance

The SOC trajectory and engine operation points of PPO are shown in Fig. 4. In the low speed part of the cycle, most of the power is supplied by battery and SOC declined continuously, the engine barely works and only works at idling speed to sustain SOC. In the medium speed part, fuel and electricity drives the vehicle simultaneously. The SOC increases slightly and engine works at the points below 14kW. In the high speed part and extra high speed part, the engine working point is optimized to balance the SOC and lower the fuel consumption. The speed and torque of engine vary in a wide range to adapt different conditions. The engine and the battery work in coordination to drive the car meanwhile optimize the fuel consumption and SOC variance.

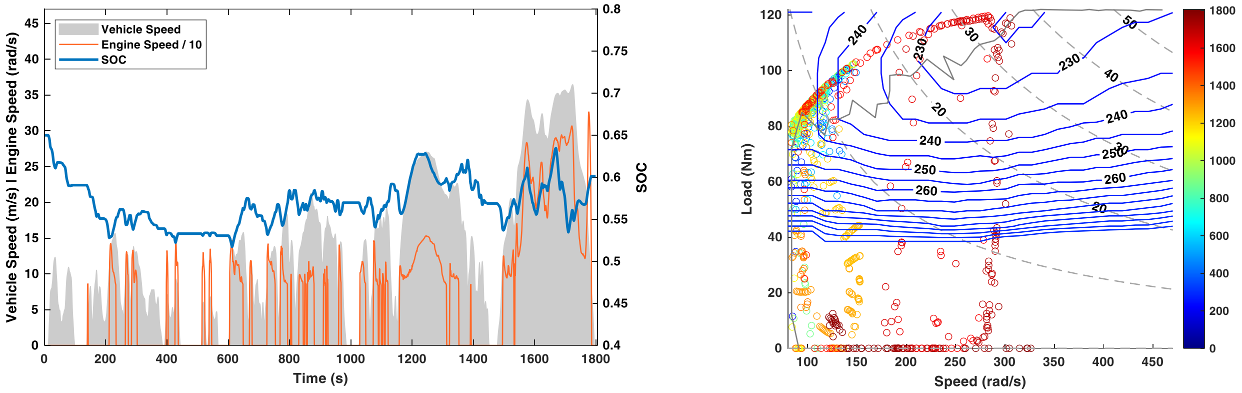


Fig. 4. SOC trajectory and engine operation points of PPO

Compared with the results including fuel consumption with 3.224L/100km and start-up times with 46, which did not consider the engine start-up conditions in PPO training process, the strategy in consideration of start-up punishment can obtain 6.73% lower fuel consumption and 37.25% less start-up times. This demonstrated that DRL-based EMS has the ability to recognize the fuel deterioration of start-up conditions and optimize the total fuel consumption.

The agents using TD3 and PPO are trained for similar time and the SOC trajectory and engine operation points of TD3 are shown in Fig. 5. It can be seen that the SOC trajectory and engine working mode are similar which demonstrates that the optimal strategy obtained from different algorithms will be similar. However, the engine operation points are quite different from each other, TD3 is more likely to work in high speed conditions and there are more transient working conditions.

3.2 Comparison of PPO and TD3

The fuel economy, start-up frequency, transient working condition percentage of both algorithms are compared as seen in Table. 3. The transient working conditions are defined by setting a threshold for torque derivative. When the absolute value of the derivative is less than 20 Nm/sec, then the operating condition is considered to be steady [11]. It can be seen that the strategy obtained by PPO can get less fuel consumption, less start-up times and less engine transient working condition percentage.

TD3 is based on DDPG which is a deterministic policy optimization algorithm, and this kind of algorithm will output a deterministic action under one state and has a poor exploration ability. Whereas PPO is a kind of stochastic policy optimization algorithm which will output a distribution of action and has a better exploration ability. At the same time, TD3 is much more sensitive to the setting of hyperparameter. Hence, the strategy obtained by PPO has better performance than that obtained by TD3. Stochastic policy optimization

Table. 3. Statistic data of strategies from PPO and TD3

Algorithm	Fuel Consumption/100km (L)	Start-up Times (n)	Transient Working Condition Percentage (%)
PPO	3.007	29	10.8%
TD3	3.305	39	19.0%

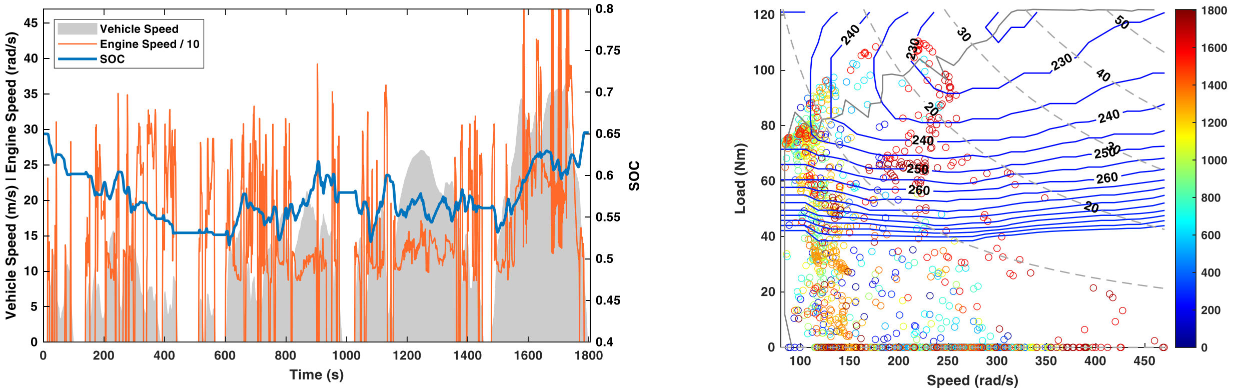


Fig. 5. SOC trajectory and engine operation points of TD3

algorithm can achieve less transient working condition percentage so that it is more suitable in this kind of EMS control problem.

4. CONCLUSIONS

In this paper, a deep reinforcement learning based energy management strategy in consideration of the engine start-up conditions for a rough Prius model is established. Meanwhile a comparison of PPO and TD3 is conducted to update the parameter of the strategy. The results showed that this kind of DRL-based EMS can recognize the fuel consumption worsen of start-up conditions in environment and give a satisfied performance. Stochastic policy optimization algorithm can get 8.2% less transient working condition percentage and it is more suitable in EMS control problems.

ACKNOWLEDGEMENT

This work was supported by the State Key Laboratory of Automotive Safety and Energy under Project No. KF2029.

REFERENCE

[1] Wübbeke J, Meissner M, Zenglein M J, et al. Made in China 2025[J]. Mercator Institute for China Studies. Papers on China, 2016, 2: 74.

[2] Mee M. Limits and Measurement Methods for Emissions from Light-duty Vehicles China 6[J]. 2016.

[3] Continental A G. Worldwide Emission Standards and Related Regulations[J]. Continental Automotive GmbH: Regensburg, Germany, 2019.

[4] Zhou Q, Li J, Shuai B, et al. Multi-step reinforcement learning for model-free predictive energy management of an electrified off-highway vehicle[J]. Applied Energy, 2019, 255: 113755.

[5] Shuai B, Zhou Q, Li J, et al. Heuristic action execution for energy efficient charge-sustaining control of connected hybrid vehicles with model-free double Q-learning[J]. Applied energy, 2020, 267: 114900.

[6] Qi X, Luo Y, Wu G, et al. Deep reinforcement learning enabled self-learning control for energy efficient driving[J]. Transportation Research Part C: Emerging Technologies, 2019, 99: 67-81.

[7] S. I, F. X, B. Z, et al. Reinforcement Learning Based on Energy Management Strategy for HEVs[C]. 2019.

[8] Lian R, Tan H, Peng J, et al. Cross-Type Transfer for Deep Reinforcement Learning Based Hybrid Electric Vehicle Energy Management[J]. IEEE Transactions on Vehicular Technology, 2020, 69(8): 8367-8380.

[9] Q. Z, D. Z, B. S, et al. Knowledge Implementation and Transfer With an Adaptive Learning Network for Real-Time Power Management of the Plug-in Hybrid Vehicle[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021: 1-11.

[10] Yang Z, Ge Y, Thomas D, et al. Real driving particle number (PN) emissions from China-6 compliant PFI and GDI hybrid electrical vehicles[J]. Atmospheric Environment, 2019, 199: 70-79.

[11] Siokos K. Low-Pressure EGR in Spark-Ignition Engines: Combustion Effects, System Optimization, Transients and Estimation Algorithms[D]. Clemson University, 2017.

[12] Kim N, Rousseau A, Rask E. Vehicle-level control analysis of 2010 Toyota Prius based on test data[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2012, 226(11): 1483-1494.

[13] Liu Z E, Zhou Q, Li Y, et al. An Intelligent Energy Management Strategy for Hybrid Vehicle with irrational actions using Twin Delayed Deep Deterministic Policy Gradient[J]. IFAC-PapersOnLine, 2021, 54(10): 546-551.

[14] Lian R, Peng J, Wu Y, et al. Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle[J]. Energy (Oxford), 2020, 197: 117297.

[15] Gaines L G T, Rask E, Keller G. Which Is Greener: Idle, or Stop and Restart? Comparing Fuel Use and Emissions for Short Passenger-Car Stops[C]. 2013.

[16] Huang Y, Surawski N C, Organ B, et al. Fuel consumption and emissions performance under real driving: Comparison between hybrid and conventional vehicles[J]. Science of The Total Environment, 2019, 659: 275-282.

[17] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[J]. arXiv preprint arXiv:1707.06347, 2017.

[18] Fujimoto S, Hoof H, Meger D. Addressing function approximation error in actor-critic methods[C]. PMLR, 2018.