Applied Energy 222 (2018) 799–811

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

# Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus

Jingda Wu, Hongwen He*, Jiankun Peng*, Yuecheng Li, Zhanjiang Li

*National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China*
*Collaborative Innovation Center of Electric Vehicles in Beijing, Beijing Institute of Technology, Beijing 100081, China*

## HIGHLIGHTS

- A continuous reinforcement learning based energy management of HEB is proposed.
- The discrete action value matrix of Q learning is replaced by continuous neural network.
- Simulation results show that the fuel economy of DQL algorithm is 5.6% better than Q learning.

## ARTICLE INFO

*Keywords:*
Energy management strategy
Continuous reinforcement learning
Deep Q learning
Dynamic programming
Hybrid electric bus

## ABSTRACT

Reinforcement learning is a new research hotspot in the artificial intelligence community. Q learning as a famous reinforcement learning algorithm can achieve satisfactory control performance without need to clarify the complex internal factors in controlled objects. However, discretization state is necessary which limits the application of Q learning in energy management for hybrid electric bus (HEB). In this paper the deep Q learning (DQL) is adopted for energy management issue and the strategy is proposed and verified. Firstly, the system modeling of bus configuration are described. Then, the energy management strategy based on deep Q learning is put forward. Deep neural network is employed and well trained to approximate the action value function (Q function). Furthermore, the Q learning strategy based on the same model is mentioned and applied to compare with deep Q learning. Finally, a part of trained decision network is analyzed separately to verify the effectiveness and rationality of the DQL-based strategy. The training results indicate that DQL-based strategy makes a better performance than that of Q learning in training time consuming and convergence rate. Results also demonstrate the fuel economy of proposed strategy under the unknown driving condition achieves 89% of dynamic programming-based method. In addition, the technique can finally learn to the target state of charge under different initial conditions. The main contribution of this study is to explore a novel reinforcement learning methodology into energy management for HEB which solve the curse of state variable dimensionality, and the techniques can be adopted to solve similar problems.

## 1. Introduction

With the standing development of economy and industry, the amount of vehicles increases rapidly, which has aroused concern of energy unsustainability and environment pollution [1,2]. Hybrid electric bus (HEB) become suitable solution for public transit by utilizing the advantages of both low energy consumption and long endurance mileage [3,4].

In recent years, many configurations of HEBs have been designed and implemented, which can be classified into three categories: series, parallel and power-split [5]. The power split configuration is widely welcomed because of decoupling control for engine rotating speed and vehicle velocity [6,7]. However, limited to the reduction ratio of a single planetary gear, the output torque cannot meet the driving demands of bus. Therefore, the dual-planetary gear hybrid system which add a gear to increase the torque is developed for HEB. In this paper, a power split HEB composed of two planetary gears, engine, integrated starter generator (ISG) and driving motor is chosen as research object.

To maximize fuel economy and keep battery state of charge (SOC) stable, the energy management strategy, which can coordinate the power distribution between engine and motor, is the key factor. Numerous energy management strategies have been proposed and applied to hybrid electric vehicles (HEVs) in these years, which can be

* Corresponding authors at: National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China.
*E-mail addresses:* pengjk87@gmail.com (J. Peng), hwhebit@bit.edu.cn (H. He).

https://doi.org/10.1016/j.apenergy.2018.03.104
Received 4 January 2018; Received in revised form 6 March 2018; Accepted 26 March 2018

generally divided into two categories, namely rule-based and optimization-based [8]. Rule-based strategies are widely adopted in practical application because of reliability. And optimization-based methods usually regulate the control variables by minimizing a predefined cost function within feasible constraints. Usually the optimization-based method can be further divided into: offline and online. Typical offline methods include: dynamic programming(DP) [9–11], convex programming [12,13], particle swarm optimization (PSO) [14]. And online optimization methods such as: Equivalent consumption minimization strategy (ECMS) [15,16], model predictive control (MPC), [17–20]. DP provides a global optimal solution and is usually used as the offline benchmark to explore the fuel economy potential [21]. And if energy management issues are approximately considered as convex, a global optimized solution can be rapidly accomplished by convex programming with any initialization [12]. PSO as a meta-heuristic approach is very suitable for energy management optimization of HEV that has some degree of noise or irregularity [14]. In terms of ECMS, it becomes a more readily implementable local optimization approach. Based on the instantaneous minimization of a cost index, it does not require prior knowledge of driving pattern [16]. MPC is a popular strategy that has been widely adopted in energy management as an effective means of dealing with multivariable constrained control problem. MPC minimizes a series of cost functions over the prediction horizon. And if with the proper prediction method, the MPC-based strategy can achieve the performance similar to DP in specific working conditions [22].

Different from traditional approaches, several novel methods have emerged in energy management of HEVs in recent two years, such as reinforcement learning (RL). As a research hotspot in the field of artificial intelligence, machine learning and automatic control [23], RL is considered as one of the core technologies in designing intelligent systems [24]. Q learning is a famous and effective RL algorithm and has been applied in HEV energy management strategy recently. Literature [25] presented RL-based energy management strategy for a plug-in parallel HEV. [26] applied Q learning into energy management of series hybrid electric tracked vehicle and registered fuel economy better than Stochastic Dynamic Programming algorithm. Data-driven Q learning algorithm is employed in [27] for a power-split plug-in HEV energy management and achieve good fuel saving performance. Literature [28] implemented predictive energy management strategy for a parallel HEV based on Q learning and computational time of the strategy was less than DP-based strategy.

The action value function is a judging system that gives action values based on historical training data and current data, this value reflects the favorable level when choosing an action, and thus agent can choose the optimal action. It is clear that Q learning can give the satisfactory control orders only if action value function has been well trained. The form of action value function in Q learning is a discretized lookup-table matrix whose size is decided by dimensions of state and action variables. However, continuous or multi-dimensional state variables are usually needed in the practical application [29], which lead the iterative computation of this matrix increase sharply and result in intractable for convergence of the training process. In all, discretized action value function posed a challenge in application of Q learning algorithm for complex HEV configuration.

Due to the above problems, researchers use the function approximation like neural network to approximate the Q matrix in RL. Literature [30] proposed a Deep Auto-Encoder model, and the model is suitable for visual perception input signal or small dimension of the state space. Deep fitted Q learning is served for vehicle motion control in [31]. Literature [32] used deep neural network (DNN) to take on the role of action value matrix and presented deep Q learning (DQL) algorithm to solve the Atari 2600 game. Under such approach, artificial intelligence showed a level of comparable competition that of human players, thus DQL method is considered to be a groundbreaking work in the field of Deep Reinforcement Learning.

This paper discusses a DQL-based energy management strategy for a power split HEB, where DNN is employed as the action value function. The powertrain and multiple power components were modeled firstly. After that the DQL algorithm is introduced and implemented to optimize the energy management problem for both minimizing the fuel consumption over a specific time horizon and keeping battery SOC stable. Furthermore, the DQL is compared with Q learning in terms of training difficulty, and the effects of different state variable on the action value function are illustrated in several scenarios. In addition, the proposed energy management technique is compared with DP-based strategy in test scenario to demonstrate its optimality and adaptability on different driving cycles excluding training conditions.

The reminder of this paper is organized as follows. In Section 2, the power split HEB is introduced and modeled. Main ideas and specific method of DP and DQL algorithm are depicted in Section 3; Section 4 provides the training comparison of DQL and Q learning and impact of input state variables. Simulation results of DQL-based energy management strategy are compared with results in Q learning-based method and the optimal solution obtained by DP-based method in Section 5. The last section is a conclusion of this paper.

## 2. Power split hybrid electric bus powertrain model

### 2.1. System configuration

The vehicle model investigated is a 12 m long bus from practical project and the topology of its powertrain is shown in Fig. 1.

The powertrain configuration composed of a diesel engine (140 kW), a driving motor (178 kW) and a ISG (118 kW). Besides, a 26 kWh battery is chosen as the energy storage system. The specific parameters of HEB are listed in Table 1.

### 2.2. The power request model

The core components of power-split HEB configuration are planetary gears: PG1 and PG2. Specific mechanical structure of powertrain is shown in Fig. 2. In this structure, engine, ISG is connected with sun gear and planet carrier of PG1, respectively. The ring gear is fixed with planet carrier of PG2 and simultaneously link with output shaft. The ring gear of PG2 is fixed with vehicle body, which allows PG2 to be used as a single gear delivering the output power of motor.

The speed and torque relationship between components are shown as

$$
\begin{aligned}
&n_g - (1+k_1)n_e + k_1 n_r = 0\\
&T_g\colon T_e\colon T_r = 1\colon -(1+k_1)\colon k_1\\
&n_m = (1+k_2)n_r\\
&T_m = \frac{T_r}{1+k_2}
\end{aligned}
\tag{1}
$$

In the above equations, $n_g$, $n_e$, $n_m$ and $n_r$ are rotating speed of ISG, engine, driving motor and ring gear of PG1; $T_g$, $T_e$, $T_m$ and $T_r$ are torque of ISG, engine, driving motor and ring gear of PG1; $k_1$ and $k_2$ are gear

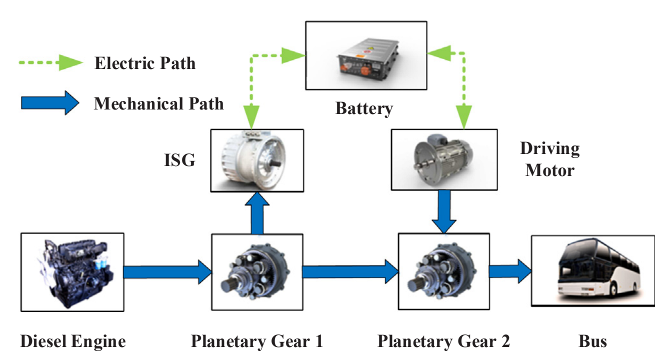


**Fig. 1.** Topology of the HEB powertrain.