Applied Energy 171 (2016) 372–382

Contents lists available at ScienceDirect

Applied Energy

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

# Reinforcement learning-based real-time energy management for a hybrid tracked vehicle

Yuan Zou *, Teng Liu, Dexing Liu, Fengchun Sun

*Collaborative Innovation Center of Electric Vehicles in Beijing and National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, No. 5 South Zhongguancun Street, Haidian District, Beijing 100081, China*

## HIGHLIGHTS

- A systematic control-oriented model for the HTV was built.
- A Markov chain model learns power transition probability recursively.
- The Kullback–Leibler divergence rate determines the transition probability update.
- Reinforcement learning (RL) was applied to optimize the control strategy.
- The strategy improves fuel efficiency and works real time.

## ARTICLE INFO

*Article history:*
Received 6 February 2016
Received in revised form 17 March 2016
Accepted 18 March 2016

*Keywords:*
Hybrid tracked vehicle
Markov chain
Kullback–Leibler divergence rate
Reinforcement learning
Energy management
Control strategy

## ABSTRACT

To realize the optimal energy allocation between the engine-generator and battery of a hybrid tracked vehicle (HTV), a reinforcement learning-based real-time energy-management strategy was proposed. A systematic control-oriented model for the HTV was built and validated through the test bench, including the battery pack, the engine-generator set (EGS), and the power request. To use effectively the statistical information of power request online, a Markov chain-based real-time power request recursive algorithm for learning transition probabilities was derived and validated. The Kullback–Leibler (KL) divergence rate was adopted to determine when the transition probability matrix and the optimal control strategy update in real time. Reinforcement learning (RL) was applied to compare quantitatively the effects of different forgetting factors and KL divergence rates on reducing fuel consumption. RL has also been used to optimize the control strategy for HTV, compared to preliminary and dynamic programming-based control strategies. The real-time and robust performance of the proposed online energy management strategy was verified under two driving schedules collected in the field test. The simulation results indicate the proposed RL-based energy management strategy can significantly improve fuel efficiency and can be applied in real time.



## 1. Introduction

As a promising solution to global warming and air pollution, hybrid electric vehicles (HEVs) are becoming increasingly popular. They have energy storage systems (ESSs) to reduce emissions and fuel consumption. Generally, two types of ESS are used in HEVs: gasoline and electricity. Energy management systems (EMS) play a crucial role in affecting performance, cost effectiveness, and adapatability of HEVs by controlling and distributing energy among multiple ESSs. An optimal energy management strategy can either improve fuel economy or reduce emissions for a HEV. To improve the online efficiency performance of HEVs, a highly efficient and real-time energy management strategy is necessary [1].

### 1.1. Literature review

Numerous researchers around the world have conducted research focused on the energy management strategies of HEVs [2]. Generally, energy management strategies for HEVs can be classified into two major types: rule based and optimization based [3,4]. With humans' early engineering experience, rule-based strategies are simple and widely used for different types of HEVs. For example, Jalil proposed a rule-based energy management strategy by setting thresholds for power split between the engine and battery [5]. Reported fuel economy has improved by 6% in the

* Corresponding author.
*E-mail address:* zouyuan@bit.edu.cn (Y. Zou).

http://dx.doi.org/10.1016/j.apenergy.2016.03.082
0306-2619/

highway cycle and 11% in the urban cycle for a series HEV. Trovão presented an integrated rule-based meta-heuristic optimization approach for a multilevel EMS in a multi-source electric vehicle [6]. Hoffman designed a new rule-based energy management strategy based on a combination of the rule-based and the equivalent consumption minimization strategies (ECMS). Compared to the dynamic programming (DP) based strategy, Hoffman's design requires significantly less computation time with a similar result to DP [7]. However, the performance from any rule-based strategy is generally sub-optimal and highly dependent on proper design of the rules, so efforts have increasingly focused on improving the optimization-based strategy, which theoretically guarantees optimality.

The DP-based energy management strategy can determine the best fuel economy once the driving cycle is given [8–10]. Tsai [8] used the DP algorithm to search the energy management strategy with different design criteria for an extended-range electric vehicle, and a multi-mode switch strategy was extracted from the DP results. However, the real-time and robust performance of this strategy cannot be guaranteed. Koot [10] generated and stored electrical energy only at the most suitable moments and demonstrated a 2% fuel reduction by applying a DP algorithm in a HEV. However, Serrao declared, because of the high computation load and adverse computation direction, the DP algorithm is impossible to use for real-time control [11].

Based on the instantaneous Hamiltonian function, Pontryagain's minimum principle (PMP) makes real-time control possible [12,13]. The core technology in PMP explores the accurate value of a parameter called co-state. Serrao [11] made a comparative analysis of the co-state and revealed the essential equivalence between PMP and DP. Liu and Sharma also elaborated that the co-state in PMP is just the derivative of the cost function in DP with respect to the state variable [14,15]. Kim and Xu [16,17] showed the optimization performance from PMP can be very close to that of DP through calculating the appropriate co-state. However, the iterative calcalations keep PMP from being implemented online directly. Based on the same theoretical background of PMP, ECMS was proposed to associate current electricity usage with future fuel consumption [18–20]. Rizzoni [19,20] developed a new adaptive strategy by adding an on-the-fly algorithm into the ECMS framework to estimate the equivalent co-state according to driving conditions.

To make online optimization feasible, many advanced intelligent algorithms, such as stochastic dynamic programming (SDP) [21,22], game theory (GT) [23], and reinforcement learning (RL) [24,25], have been proposed to settle the energy management problem for multiple types of HEVs. Lin [21] optimized the power management problem for a parallel HEV through an SDP algorithm, but the high computational burden makes it difficult to implement online. Liu [26,27] compared the performance of RL and DP as well as RL and SDP, and the simulation results showed the computational time of RL is much less than that of PMP while the control performance from the RL alogorithm is much closer to that of the DP algorithm. However, the transition probability matrix in the RL algorithm cannot be immediately updated online; thus, the robustness of this strategy cannot be guaranteed for different driving conditions.

### 1.2. Motivation and innovation

The purpose of this study is to propose a real-time and robust energy management approach via Markov chain-based recursive algorithm and RL to enhance the energy management efficiency and performance for a hybrid tracked vehicle (HTV). Three perspectives are contributed in this paper. First, to use effectively the statistical information of the online driving event, a Markov chain-based real-time power-request recursive algorithm for learning transition probabilities has been derived. Second, a Kullback–Leibler (KL) divergence rate technique is developed for deciding when to update the transition probability matrix (TPM) and the optimal control strategy in real time. The RL algorithm-based online updating strategy for TPM is developed in a systematic way. RL is applied to validate quantitatively the influences on fuel economy with regard to forgetting factors and KL divergence rates. The proposed RL-based energy management strategy has been evaluated through comparison to the stationary online and DP-based control strategies. The real-time and robust performance of the energy management strategy is validated under three different driving schedules collected in the field test.

### 1.3. Organization of the paper

The remainder of this paper is organized as follows. Section 2 describes the configuration of the HTV and the systematic modeling approach. The real-time recursive algorithm for learning transition probability and the Kullback–Leibler (KL) divergence rate are illustrated in Section 3. The verification and evaluation of the proposed energy management strategy are reported in Section 4, and Section 5 concludes this paper.

## 2. HTV modeling and optimal control problem formulation

### 2.1. Vehicle configuration

The HTV architecture is shown in Fig. 1. The power sources come from two parts: the engine-generator set (EGS) and the battery pack. The EGS consists of a 300 kW diesel engine and a 270 kW permanent magnet generator. The 50 Ah lithium–ion battery pack has 470 V rated voltage [27]. Detailed modeling of the EGS and battery are illustrated in Section 2.2. The essential parameters of the HTV are listed in Table 1 [28].

### 2.2. Modeling of the HTV

A systematic control-oriented model is established for the HTV to evaluate the control performance of different energy management strategies. Because this study is focused largely on the performance of the control strategy, it is assumed that the EMs are only power-conversion devices with the same average efficiency [27,28]. The EGS, battery, and power-request models are described as follows.

#### 2.2.1. Modeling the EGS

The equivalent electric circuit illustrated in Fig. 2 comprises the diesel engine, permanent magnet generator, and three-phase full-wave rectifier. The engine outputs 300 kW rated power at the speed of 3100 r/min and 2200 Nm rated output torque within the speed range from 650 r/min to 2100 r/min. The generator outputs 270 kW rated power within the speed range from 2500 r/min to 3100 r/min and 960 Nm rated torque within the speed range from 0 to 2500 r/min [27]. The dynamics of the EGS are described by the following equation [29]:

$$\begin{cases} K_e I_g - K_x I_g^2 = T_g \\ U_g = K_e n_g - K_x n_g I_g \\ 0.1047\left(\frac{J_e}{i_{eg}^2} + J_g\right)\frac{dn_g}{dt} = \frac{T_{en}}{i_{eg}} - T_g \\ n_{en} = n_g / i_{eg} \end{cases} \tag{1}$$

where $K_e$ is the coefficient of the electromotive force; $K_x n_g$ is the electromotive force, in which $K_x = 3PL^g/\pi$; $L^g$ is the synchronous