Energy 160 (2018) 544–555

Contents lists available at ScienceDirect

Energy

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

Online Markov Chain-based energy management for a hybrid tracked vehicle with speedy Q-learning

Teng Liu [a, b], Bo Wang [c, *], Chenglang Yang [c]

[a] State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130025, China
[b] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Ontario N2L3G1, Canada
[c] School of Mathematics and Statistics, Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, No. 5 South Zhongguancun Street, Haidian District, Beijing 100081, China

ARTICLE INFO

Article history:
Received 26 June 2017
Received in revised form
21 June 2018
Accepted 7 July 2018
Available online 11 July 2018

Keywords:
Hybrid tracked vehicle
Markov chain
Induced matrix norm
Onboard learning algorithm
Reinforcement learning
Speedy Q-learning

ABSTRACT

This brief proposes a real-time energy management approach for a hybrid tracked vehicle to adapt to different driving conditions. To characterize different route segments online, an onboard learning algorithm for Markov Chain models is employed to generate transition probability matrices of power demand. The induced matrix norm is presented as an initialization criterion to quantify differences between multiple transition probability matrices and to determine when to update them at specific road segment. Since a series of control policies are available onboard for the hybrid tracked vehicle, the induced matrix norm is also employed to choose an appropriate control policy that matches the current driving condition best. To accelerate the convergence rate in Markov Chain-based control policy computation, a reinforcement learning-enabled energy management strategy is derived by using speedy Q-learning algorithm. Simulation is carried out on two driving cycles. And results indicate that the proposed energy management strategy can greatly improve the fuel economy and be employed in real-time when compared with the stochastic dynamic programming and conventional RL approaches.

1. Introduction

Hybrid electric vehicles (HEVs) seem to be the most promising solution to overcome the increasing energy crisis and environmental pollution in recent decades [1]. Two types of energy sources, electricity and gasoline, have been placed in a HEV to make it possible to improve fuel economy and reduce exhaust emissions [2]. The energy management strategies are the critical technology for HEV to achieve the best performance and energy efficiency through power-split control [3]. One major difficulty to achieve this goal is how to adapt to multiple driving cycles. Along with the development of HEV, an effective and real-time energy management strategy is necessary for HEV to accommodate different driving conditions.

1.1. Literature review

Currently, the energy management strategies for HEV are mainly optimization-enabled strategies considering necessary physical constraints [4], such as restraints on state of charge in battery, torque and rotational speed of engine and output power of battery and engine. Optimization-based energy management strategies can be further divided into global optimization and real-time optimization cases. Since the complete knowledge of the driving cycle is predefined, dynamic programming (DP) algorithm is employed to make a globally optimal control decision. Ref. [5] leveraged DP to optimize the fuel economy for a velocity coupling HEV system with eleven modes. Serrao etc. [6] compared the DP with other two methods to demonstrate its optimality. Additionally, Pontryagain's Minimum Principle (PMP) technique is also adopted to improve the energy efficiency of the propulsion system via the global optimal control. A piecewise linear approximation strategy is combined with the PMP to derive the optimal control for plug-in HEV in Ref. [7]. Zhang etc. [8] applied PMP to optimize the control strategy for a dual-motor-driven electric bus under three different driving cycles. Convex programming (CP) is another global optimization method that derives the energy management strategy based on the convex modeling and rapid solution search. In Ref. [9], CP is used to implement a framework of simultaneous optimal energy storage systems sizing and energy management. Hu etc. [10]

* Corresponding author.
E-mail address: wangbo89630@bit.edu.cn (B. Wang).

https://doi.org/10.1016/j.energy.2018.07.022
0360-5442/

presented a high-efficiency CP framework to construct the swiftly adapting charging/power management controls to wind intermittency. However, the global optimization strategies can only be feasible in off-line simulation since the driving cycle is generally unknown in practical application.

Besides, the stochastic dynamic programming (SDP) algorithm is also developed to search the optimal energy management strategy for HEV through taking the random characteristics of the vehicle speed and drivers' behaviors into account. Ref. [11] employed SDP to address the energy management for a series hybrid tracked vehicle based on the Markov chain driver model. Xi etc. [12] co-optimizes the use of energy storage for multiple applications with SDP while accounting for market and system uncertainty. In real-time optimization, equivalent consumption minimization strategy (ECMS) [13] and model predictive control (MPC) [14] are two most representative optimization-based approaches. The ECMS explores the precise co-state value to achieve the local optimization, which strongly depends on the validity of velocity predictions [15]. For MPC, the controller presents an energy management strategy via DP, genetic algorithm, quadratic programming, or nonlinear programming. For example, the information provided by the onboard navigation system is utilized in the MPC framework [16]. An adaptive approach is developed based MPC to consider the load torque estimation and prediction in energy management problem [17]. Genetic algorithm and MPC is combined in Ref. [18] to minimize the energy consumption. Furthermore, a multi-layer perception is presented based on MPC and it is proved to guarantee globally-bounded closed-loop stability [19]. Nevertheless, the performance of the MPC control is highly influenced by the future information, such as prospective speed or power prediction [20].

Two inspiring innovative techniques, named reinforcement learning (RL) and game theory (GT), are also proposed to build an optimal controller for HEVs. RL can derive a model-free and apdative control for energy management problem [21]. And the global optimality of the GT is evaluated in Ref. [22] via comparing with the DP method. Liu etc. [23] proposed a bi-level control framework to combine the predictive learning with RL to formulate the energy management strategy. Ref. [24] presents a GT controller with the cost penalizing fuel consumption, NO_x emissions, battery state of charge deviation, and vehicle operating conditions deviation. Over the new European driving cycle, the GT controller acquires the closest control performance to the existing DP controller. Markov Chain (MC) models are quite well-suited to represent the uncertainty in the driving environment, which can lower both the information required for implementation and the on-board computing burden [25]. Based on the MC models, Liu etc. compared the control performance of RL and SDP as well as two different RL-based algorithms [26], and the results indicated the advantages of RL over the SDP in fuel economy and computational time [27]. However, the issue that the popular Q-learning algorithm overestimates action values under certain conditions is not considered in previous energy management of HEV [28]. Meanwhile, to the best of our knowledge, combining RL algorithm with the on-board learning MC models has not been surveyed, and the existing RL-enabled energy management strategy cannot guarantee adaptive to various driving conditions.

1.2. Motivation and innovation

The main purpose of this brief is to construct a real-time energy management strategy by a collaboration of MC-based onboard learning algorithm and speedy Q-learning (SQL) algorithm. Three primary contributions are presented in this paper. Firstly, an onboard learning algorithm is proposed for MC models to learn the transition probability of power demand in real-time. Secondly, the induced matrix norm (IMN) is served as an initialization criterion for MC models learning. Thus, a set of models representing different segments of power demand can be evolved and the IMN is applied to select control policy that matches the current driving condition best. Finally, the SQL algorithm is developed to evaluate the onboard learning algorithm and avoid selecting overestimated values in control policy computation. In addition, the proposed energy management strategy is compared with the SDP and conventional Q-learning algorithm to estimate its performance in different driving conditions.

1.3. Organization

The remainder of this paper is organized as follows: the induced matrix norm and the recursive algorithm for updating the transition probability matrix are illuminated in Section 2; In Section 3, the onboard learning algorithm for MC models learning and the SQL algorithm are discussed; the comparative research between different energy management strategies are conducted in Section 4; conclusions are given in Section 5.

2. Problem formulation and background

The vehicle being studied is a hybrid tracked vehicle (HTV) with a series topology. The powertrain configuration is sketched in Fig. 1. The main power components consist of a battery pack, an engine-generator set (EGS), and two traction motors. EGS and battery constitute the main power sources to propel the powertrain. For EGS, the rated power of engine is 52 kW at the speed of 6200 rpm. The rated output power of generator is 40 kW within the speed range from 3000 rpm to 3500 rpm. Power split controls between the EGS and battery are the key technologies to realize the fuel efficiency improvement. The elementary parameters of the powertrain are shown in Table 1. The modeling of the EGS and battery is introduced in subsection 2.1. Since the historical vehicle speed is known in real-time, the power demand P_{dem} can be calculated as follows

$$\begin{cases} P_{dem} = (F_r + F_i + F_a)\bar{v} + M\omega \\ F_r = mg \cdot f \\ F_i = ma \\ F_a = (C_D A / 21.15)\bar{v}^2 \end{cases} \tag{1}$$

where F_r, F_i and F_a are the rolling resistance, inertial force and aerodynamic drag, respectively. m is the vehicle mass, g is the gravity acceleration, a is the vehicle acceleration, C_D is the aerodynamic coefficient and A is the fronted area. M is the resisting yaw moment, $\bar{v}$ and ω are the average velocity and rotational speed for the tracked vehicle.

2.1. Optimization objective

The generator speed is selected as the state variable that can be calculated according to the torque equilibrium constraint

$$\begin{cases} \dfrac{dn_g}{dt} = \left(\dfrac{T_e}{i_{e-g}} - T_g\right) \Big/ 0.1047\left(\dfrac{J_e}{i^2{}_{e-g}} + J_g\right) \\ n_e = n_g / i_{e-g} \end{cases} \tag{2}$$

where n_g and n_e are the rotational speeds, T_g and T_e are the torques of the generator and engine, respectively, and T_e is decided by the throttle variable $th(t)$ using the expression $T_e = th^{*}interp(n_g, T_{e,\max})$, wherein *interp* indicates the interpolation function and $T_{e,\max}$ is the