Markov decision process of optimal energy management for plug-in hybrid electric vehicle and its solution via policy iteration

Jie Fan*, Yang Ou, Peng Wang, Lei Xu, Zhe Li, HongGuo Zhu, Zhou Zhou.
China Automotive Engineering Research Institute Co., Ltd., Chongqing 401122, China

*Corresponding author: Jie Fan. E-mail address: 384954890@qq.com.

Abstract. This paper proposes a Markov decision process based optimal energy management strategy for plug-in hybrid electric vehicle with a hybrid energy storage system, which mainly consists of an ultracapacitor, a power battery pack and a DC/DC converter. Firstly, the mathematical model is built for the overall system. Secondly, based on modified worldwide harmonized light duty test cycle, the power transition probability matrix is constructed to describe the Markov property of driver’s power demand in natural driving. Then by combining the stochastic property of driver’s power demand and deterministic state transition of the hybrid powertrain system, the energy management optimization problem is formulated into a typical Markov decision process. Finally, the derived optimization problem is solved by policy iteration and the corresponding optimal control map is generated. The optimality of the solution is investigated and discussed in detail.

1. Introduction

Pure electric vehicles (PEVs), which are propelled by power battery only, have achieved great success in China's market [1]. Although PEV has advantages considering its simple powertrain and easy control, it cannot ensure the power battery always to work in high-efficiency area. In addition, when the driver operates emergency braking or acceleration, the battery pack may suffer from overheat problem. In order to overcome the shortcomings mentioned above, a hybrid energy storage system (HESS) combining power battery and ultracapacitor has been proposed to design a plug-in hybrid electric vehicle (PHEV). The complementarity of this HESS lies in the fact that ultracapacitor possesses instant high power property while power battery has relatively high energy density. Their combination can ensure both quick large power output at highly dynamic instants and long driving distance at smooth conditions [2]. However, adding ultracapacitor to existing PEV complicates the powertrain structure and a new energy management strategy is necessary to coordinate the working between power battery and ultracapacitor.

In this paper, a Markov decision process based energy management strategy for the HESS used in plug-in HEV is proposed and its optimal control map is obtained via policy iteration. Not only the mathematical model of HESS is built in detail, but also the stochastic property of driver’s power demand is incorporated into the problem formulation.

The remainder of this paper is organized as follows. In Section 2, the structure of the HESS and its mathematical model are introduced in detail, together with the transition probability matrix (TPM) based modeling of driver’s power demand. In Section 3, the optimal energy management problem is
formulated into a Markov decision process and one of its typical solutions, policy iteration, is used to solve the problem. In Section 4, simulations are conducted to verify the effectiveness of the proposed method and finally Section 5 concludes the whole paper.

2. Modeling of HESS and driver’s power demand

Figure 1 demonstrates the structure of the investigated HESS for the PHEV. The power battery is connected in series with a DC/DC converter and then connected in parallel with an ultracapacitor. The DC/DC converter used here is to balance the voltage of ultracapacitor and the external load. In order to drive the AC motor, a DC/AC converter is connected after the ultracapacitor and the DC/DC converter. Then the motor passes the propulsion through transmission and differential to the wheels to drive the vehicle. The driver’s power demand is passed to the vehicle controller and by communicating with power battery, converters, ultracapacitor and motor, the vehicle controller executes energy management strategy to realize power distribution between different power sources. The basic parameters of the investigated PHEV are shown in Table 1.

![Figure 1 Structure of HESS](image)

Table 1 Parameters of the investigated PHEV

| Parameters                  | Value  |
|-----------------------------|--------|
| Cubic mass                  | 16500kg|
| Rolling resistance coefficient | 0.011  |
| Frontal area                | 6.6m²  |
| Air drag coefficient        | 0.55   |

2.1. Battery pack model

The first-order RC model is adopted to describe the battery’s dynamic characteristics as it is a good balance between simplicity and accuracy. The model consists of an ohmic resistance $R_o$, a RC network (which incorporates a polarization resistance $R_p$ and a polarization capacitor $C_p$) and an open circuit voltage (OCV). Its discretized state equation and output equation are as follows [3]:

$$\begin{align*}
U_p(k+1) &= \exp\left(-\frac{\Delta t}{T_c}\right)U_p(k) + \left(1 - \exp\left(-\frac{\Delta t}{T_c}\right)\right)R_o \left(i_L(k) - \frac{\Delta t}{3600Q}\right) \\
U_b(k) &= \text{OCV}(SOC) - U_p(k) - i_L(k)R_o
\end{align*}$$

Eq.(1)

where SOC represents the battery’s state of charge, $i_L$ denotes the incitation current (positive for discharge and negative for charge), $T_c$ is the time constant for the RC network and $T_c=R_pC_p$. $U_p$ represents the voltage across the RC network and $U_b$ denotes the terminal voltage. $\Delta t$ is the sampling interval and $k$ is the time index. The relationship between battery model parameters and SOC are shown in Table.2 [4].
Table 2 Relationship between battery model parameters and SOC

| Parameters | Relationship |
|------------|--------------|
| OCV        | OCV=13.91SOC⁶-35.62SOC⁵+29.61SOC⁴-5.232SOC³-3.959SOC²+2.023SOC+3.502 |
| R_o        | R_o=0.01483SOC²-0.02518SOC+0.1036 |
| R_p        | R_p=-1.212e⁰.03383SOC+1.258 |
| T_c        | T_c=2.151e².132SOC+27.2 |

The above model is for the battery cell. In order to meet the power demand of the investigated PHEV, 27 battery cells are first connected in parallel to form a battery module and then 138 such battery modules are connected in series to form the final battery pack.

2.2. Ultracapacitor model

The ultracapacitor is modeled as an ideal capacitor (maximum voltage \( V_{\text{max}}=576 \text{V} \) and capacity \( C_u=13.83 \text{F} \)) in series with a constant resistance \( R_c=0.0756 \Omega \). Its discretized state equation and output equation are as follows:

\[
SOV(k+1) = SOV(k) - i_c(k) \Delta t / C_u \\
U_c(k) = SOV(k) \times V_{\text{max}} - R_c i_c
\]

where \( SOV \) represents the state of voltage, \( i_c \) denotes the load current and \( U_c \) is its terminal voltage.

2.3. DC/DC model

For the DC/DC converter, only its operational efficiency is considered here. The DC/DC converter’s efficiency \( \eta \) is determined by its output current and output power. Its efficiency map is shown in Figure 2 [5].

![Efficiency map of DC/DC converter](image)

2.4. Power demand model

Worldwide harmonized light duty test cycle (WLTC) is used to simulate the driver’s speed profile. However, because the simulated vehicle is a bus, in most cases its maximum speed will not exceed 80km/h, thus only the first three parts of WLTC are selected and they are then multiplied by scaling factor 0.75 to derive the final speed profile, as shown in Figure 3(a). Following Eq.(3) is used to transform the speed to the required power.

\[
P = \frac{u_a}{\eta_T} (Mgf \cos \alpha + Mg \sin \alpha + \frac{C_D A u_a^2}{21.15} + \delta M \frac{du_a}{dt})
\]

where \( u_a \) represents the vehicle speed. \( M \) is the vehicle’s cubic mass. \( \eta_T \) denotes the transmission efficiency. \( f \) is the rolling resistance coefficient. \( \alpha \) is the road inclination. \( C_D \) is the air drag coefficient. \( A \) is the vehicle’s frontal area and \( \delta \) is the conversion ratio of vehicle rotating mass.

According to previous researches, driver’s power demand has Markov property and its stochastic characteristic can be modeled by TPM. In this paper, we discretize power demand with interval of 20kW and use neighborhood method to construct its TPM according to following Eq.(4):

\[
P_{i,j} = \frac{N_{i,j}}{N_i} \quad N_i = \sum_{j=1}^{N} N_{i,j}
\]

Eq.(4)
where $p_{i,j}$ represents the $i$th column $j$th row element in TPM. $N_{i,j}$ represents the total times that the driver’s power demand transits from $i$th level to $j$th level and $N_i$ denotes the total times that power demand transits from $i$th level.

![Figure 3] Simulation speed profile and the corresponding power TPM. (a)Speed profile. (b)TPM.

Using above method, the constructed TPM is shown in Figure 3(b). It can be seen that elements in diagonal position have relatively higher value, which means that the drive’s power demand seldom changes sharply and in most cases adjacent sampling points have similar power demand.

3. Markov decision process based problem formulation and solution via policy iteration

Due to the stochastic property of drive’s power demand, it is difficult to get the optimal energy management solution for this problem if using traditional deterministic problem formulation and solution methods such as particle swarm optimization, genetic algorithm, etc. However, Markov decision process provides a systematic procedure to model and solve such problems with stochastic properties. In this section, we will transform the proposed optimal energy management problem into standard Markov decision process and introduce one of its solutions, namely policy iteration.

3.1. Problem formulation

The key to define a Markov decision process is to find its corresponding state TPM and reward. In this paper, the combination of driver’s required power, battery SOC and ultracapacitor SOV determines the system state $s$. The battery current is selected as control variable $a$ and its range is set to be between $-154A(-2C)$ and $154A(2C)$. Therefore, the system’s corresponding TPM $p(s’,r|s,a)$ can be derived by merging the deterministic transition of Eq.(1)–Eq.(2) and stochastic transition of Eq.(4), where reward $r$ can be defined as following Eq.(5) as the control target here is to minimize the system’s energy loss.

$$r = P_{\text{batt LOSS}} + P_{\text{cap LOSS}} + P_{\text{DCDC LOSS}} = i_b^2 R_a + \frac{U_b^2}{R_p} + i_c^2 R_c + U_b i_L (1 - \eta) \eta^{-z} \quad \text{Eq.(5)}$$

where $z$ denotes a logical value, which is set to be 1 if the battery is charged and 0 when the battery discharges. According to Markov decision process theory, the solution is to find an optimal map $\pi^*(s)$ so as to maximize the value of state $s$, which is the cumulative discounted reward as following Eq.(6).

$$V(s) = - E \left( \sum_{t=0}^{\infty} \gamma^t r_{t+1} \right) \quad \text{Eq.(6)}$$

where $\gamma$ is a discounting factor ($0 \leq \gamma \leq 1$), which is used to transform the value of future rewards to current step. Because Markov decision process usually maximize the reward while in this problem the reward is defined as the energy loss which should be minimized, a negative sign is added in Eq.(6).

3.2. Policy iteration

Policy iteration is one of effective methods to solve Markov decision process problem. The detailed algorithm of policy iteration is shown in Table.3 [6].
Battery current (A)

Required power is -20kW

Required power is 20kW

Required power is 140kW

Required power is -140kW

Table 3 Policy iteration algorithm

1. Initialization
   \[ V(s) \in \mathbb{R} \text{ and } \pi(s) \in A(s) \text{ arbitrarily for all } s \in S \]
2. Policy Evaluation
   Repeat
   \[ \Delta \leftarrow 0 \]
   For each \( s \in S \):
   \[ v \leftarrow V(s); V(s) \leftarrow \sum_{s',r} p(s',r|s,s) \pi(s)[r + \gamma V(s')]; \Delta \leftarrow \max(\Delta, |v - V(s)|) \]
   Until \( \Delta < \theta \) (a small positive number)
3. Policy Improvement
   \[ Policy-stable \leftarrow true \]
   For each \( s \in S \):
   \[ old \text{ action} \leftarrow \pi(s); \pi(s) \leftarrow \arg\max_a \sum_{s',r} p(s',r|s,s) \pi(s)[r + \gamma V(s')]; \]
   If \( old \text{ action} \neq \pi(s) \), then \( Policy-stable \leftarrow false \)
   If \( Policy-stable \), then stop and return \( V \approx v^* \) and \( \pi \approx \pi^* \); else go to 2.

4. Optimal solution and simulation results

In this section, we will first discuss the reasonability of the derived optimal control map. Then we apply this control policy on the investigated PHEV through simulation and discuss the coordinated working principle between the battery pack and ultracapacitor.

4.1. Optimal control map

Part of the optimal control map derived through policy iteration has been shown in Figure 4. Four scenarios where the driver’s required powers are -20kW, 20kW, -140kW and 140kW have been demonstrated. It can be seen that when required power is -20kW, in most cases the battery current is 0, which means the ultracapacitor will absorb the regenerative energy. However, when SOV is close to 1, which means the ultracapacitor is nearly fully-charged and cannot store more energy, the battery current becomes negative and the regenerative energy is mainly absorbed by the battery. When required power is 20kW, the ultracapacitor is still the main energy resource and the battery only discharges when SOV is very low. From the above two scenarios, it can be concluded that the derived control policy tends to use ultracapacitor as the energy buffer when the required power is relatively low. When required power is -140kW, if SOV is high, which means the ultracapacitor cannot absorb excessive energy any more, the regenerative energy is mainly absorbed by the battery. In addition, it is interesting to note that even when the required power is negative, the battery may still discharge when SOV is relatively low. The reason is that the battery tends to make the ultracapacitor charged so as to ensure the ultracapacitor’s buffer capability. When required power is 140kW, the battery becomes the main resource. If SOV is extremely low, the battery will increase its discharge current not only to satisfy the power demand but also to charge the ultracapacitor. If SOC is extremely low, the battery will decrease its discharge current and excessive power demand will be met by ultracapacitor.

Figure 4 Optimal control map derived through policy iteration
4.2. Simulation and discussion

Simulation has been conducted to verify the effectiveness of the proposed method. The initial SOC of the battery pack is set to be 0.9 and initial SOV of the ultracapacitor is set to be 1. Figure 5 demonstrates the simulation result. It can be seen that the derived control policy can effectively coordinate the co-operation between battery pack and ultracapacitor. When the required positive power is extremely large, the ultracapacitor will assist the battery pack to satisfy driver’s demand. In most cases of regenerative braking, the ultracapacitor takes the responsibility of storing the recyclable kinetic energy. Especially, from Figure 7(c), where two dashed red lines refer to ±30kW, it is noticeable that the magnitude of battery power is mostly around 30kW if it is not very large. The reason is that the DC/DC converter can achieve highest efficiency when it works around 30kW, which can be seen from Figure 2. The optimal policy tries to adjust the battery’s power to be close to 30kW so as to increase DC/DC efficiency. In addition, because DC/DC also has desirable efficiency when both of its power and current are large, the optimal policy tends to make the battery work in high-rate discharging state.

5. Conclusion

Energy management strategy is crucial for ensuring economic performance of hybrid electric vehicle. In this paper, a Markov decision process based energy management strategy for a HESS used in PHEV is proposed and its optimal control map is obtained via policy iteration. Simulations are conducted to verify the effectiveness of the proposed method. Results show that the optimal control policy can make the two power sources (battery pack and ultracapacitor) work coordinately with each other while improving the efficiency of overall system.

Acknowledgements

The authors acknowledge the Innovation Support Program for the Leader in Science and Technology of Chongqing under Grants (010267) for financial support.

References

[1] Zhang X, Zou Y, Fan J and Guo H 2019 Usage pattern analysis of Beijing private electric vehicles based on real-world data Energy 167 1074–85
[2] Song Z, Hofmann H, Li J, Han X and Ouyang M 2015 Optimization for a hybrid energy storage system in electric vehicles using dynamic programming approach Appl. Energy 139 151–62
[3] Fan J, Zou Y, Zhang X and Guo H 2019 A novel State of Health estimation method for Lithium-ion battery in electric vehicles J. Phys. Conf. Ser. 1187
[4] Zou Y, Hu X, Ma H and Li S E 2015 Combined State of Charge and State of Health estimation over lithium-ion battery cell cycle lifespan for electric vehicles J. Power Sources 273 793–803
[5] Ahmed, O.A, Bleijs, J.A.M 2017 An overview of DC–DC converter topologies for fuel cell–ultracapacitor hybrid distribution system Renew. Sust. Energ. Rev 42 609-626
[6] Eduardo F Morales, Julio H Zaragoza 2011 An Introduction to Reinforcement Learning IEEE