

车辆混合动力控制器策略研究

邹润楠

2020 年 6 月

中图分类号：U463.2

UDC 分类号：629

车辆混合动力控制器策略研究

作者姓名	<u>邹润楠</u>
学院名称	<u>机械与车辆学院</u>
指导教师	<u>邹渊教授</u>
答辩委员会主席	<u>闫清东教授</u>
申请学位	<u>工学硕士</u>
学科专业	<u>机械工程</u>
学位授予单位	<u>北京理工大学</u>
论文答辩日期	<u>2020年6月</u>

Research on Strategy of Hybrid Vehicle Controller

Candidate Name:	<u>Runnan Zou</u>
School or Department:	<u>School of Mechanical Engineering</u>
Faculty Mentor:	<u>Prof. Yuan Zou</u>
Chair, Thesis Committee:	<u>Prof. Qingdong Yan</u>
Degree Applied:	<u>Master of Science</u>
Major:	<u>Mechanical Engineering</u>
Degree by:	<u>Beijing Institute of Technology</u>
The Date of Defense:	<u>June, 2020</u>

车辆混合动力控制器策略研究

北京理工大学

研究成果声明

本人郑重声明：所提交的学位论文是我本人在指导教师的指导下进行的研究工作获得的研究成果。尽我所知，文中除特别标注和致谢的地方外，学位论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京理工大学或其它教育机构的学位或证书所使用过的材料。与我一同工作的合作者对此研究工作所做的任何贡献均已在学位论文中作了明确的说明并表示了谢意。

特此申明。

签 名：_____ 日期：_____

关于学位论文使用权的说明

本人完全了解北京理工大学有关保管、使用学位论文的规定，其中包括：①学校有权保管、并向有关部门送交学位论文的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存学位论文；③学校可允许学位论文被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换学位论文；⑤学校可以公布学位论文的全部或部分内容（保密学位论文在解密后遵守此规定）。

签 名：_____ 日期：_____

导师签名：_____ 日期：_____

摘要

发展兼具传统燃油汽车和纯电动汽车优势的混合动力汽车是我国实现汽车工业跨越式发展、构建环境友好型社会的有效途径。混合动力车辆能量管理策略的研发和整车测试是制造混合动力车辆的关键步骤。目标于解决混合动力车辆能量管理问题，本文以串联混合动力履带车为研究和优化目标，提出了基于 NAF-DQL+PR（混合经验权值优先回放及归一化优势函数的深度强化学习）的混合动力车辆能量管理策略和基于 GA-DQL（基于深度强化学习的遗传算法）的深度强化学习能量管理策略超参数优化方法，设计了面向于混合动力车辆能量管理算法实现的混合动力源实时调试软件，通过搭建硬件在环仿真平台验证了能量管理策略及超参数优化方法的节能性、高效性和最优性。

建立了串联混合动力履带车前向模型，在 python 及 MATLAB/Simulink 环境中完成对车辆发动机、发电机、动力电池、高压功率模块及车辆动力学的数学模型构建并将该模型引入 dSPACE 实时仿真平台。该模型精度达到仿真要求，能支持后续针对能量管理策略的训练和硬件在环验证研究。

提出了基于 NAF-DQL+PR 的混合动力履带车能量管理策略。针对深度强化学习能量管理策略训练时间长、燃油经济性不高的问题，引入经验权值优先回顾模型(PR)，通过赋予历史经验数据优先级和建立 Sumtree 数据存储与抽取结构，实现神经网络对历史经验数据带权值的重复训练。通过构建取值更新权重，实现神经网络带权重更新。将经验权值优先回顾模型与 NAF-DQL 算法有机结合成 NAF-DQL+PR 算法，搭建了基于 NAF-DQL+PR 的能量管理策略。为验证该算法的节能性及高效性，将其训练结果与动态规划(DP)、NAF-DQL 及 DQL 进行发动机工作点、动力电池 SOC、燃油消耗及训练耗时等方面对比，实验结果表明 NAF-DQL+PR 算法节能性仅次于 DP 但训练效率远高于其他算法，验证了 NAF-DQL+PR 算法的节能性和高效性。

提出了基于 GA-DQL 的深度强化学习能量管理算法超参数优化方法。针对深度强化学习算法中学习率对训练效果影响较大且难以通过经验设置的问题，引入遗传算法，通过二进制化学习率为“基因型”、初始化种群、计算适应度、交叉组合和变异等方式对学习率进行优化。为验证 GA-DQL 算法的有效性，将优化过与未优化的 NAF-DQL+PR 算法训练结果进行对比，结果表明优化过的 NAF-DQL+PR 算法具有更高燃油经济性和训练高效性，验证了基于 GA-DQL 的深度强化学习能量管理算法超参数

优化方法的有效性。

研制了基于 LabVIEW 平台的混合动力源实时调试软件。基于 CAN 通讯协议，使用事件触发和定时触发，针对混合动力源发动机、发电机、动力电池及高压功率模块建立实时监控模块及实时绘图模块，针对 100 个参数建立实时参数修改模块。该软件有助于加速混合动力车辆能量管理策略实车调试，有助于缩短研发周期。

搭建了串联混合动力履带车硬件在环仿真平台。以快速原型开发器 MotoTron 为整车控制器，将第二章建立的串联混合动力履带车前向模型作为被控系统嵌入 dSPACE，连接上位机、整车控制器及被控系统基于以太网和 CAN 通讯协议建立硬件在环仿真平台。从第五章优化过的能量管理策略中提取神经网络参数移植入整车控制器对被控系统进行控制。进行 NAF-DQL+PR 及 NAF-DQL 硬件在环仿真实验，二者对比结果表明，NAF-DQL+PR 具有更高燃油经济性表现且对陌生工况有更好迁移性。通过使用基于 dSPACE 和 MotoTron 的硬件在环实时仿真平台对混合动力源实时调试软件做硬件在环测试，结果表明该软件具有实时性，可实时反映混合动力源运行状态及对混合动力源参数进行修改，验证该软件的性能。

关键词：混合动力履带车辆；能量管理；深度 Q 学习算法；经验权值优先回放；NAF-DQL+PR 算法；硬件在环仿真

Abstract

Developing hybrid electric vehicle which combines advantage of both electric vehicle and internal combustion engine vehicle is an effective way for China to realize the great improvement of vehicle industry and build an environmentally friendly society. The development of energy management strategy for hybrid electric vehicle and the test of the strategy on vehicle are the key step for manufacturing hybrid electric vehicle. Aiming at solving the energy management problem of hybrid electric vehicle, this thesis selects tracked plug-in hybrid electric vehicle as research and optimization object. An energy management strategy based on NAF-DQL+PR, a hyper parameter optimization method for deep reinforcement learning based energy management and a real-time debugging software for the realization of energy management strategy are proposed. The hardware in loop simulation is built in order to verify the fuel economy, effectiveness and optimality of NAF-DQL+PR and the software.

The control-oriented model of tracked plug-in hybrid electric vehicle is established. The mathematical modelling of engine, generator, battery, AC-DC and vehicle dynamic model is established in environment of python and MATLAB/Simulink. This model which is transplanted into dSPACE real-time simulation platform reaches the precision of simulation and it can support the training and hardware in loop simulation for energy management strategy.

An energy management strategy based on NAF-DQL+PR for tracked plug-in hybrid electric vehicle is established. Aiming at the problem of deep reinforcement learning on long training time and low fuel economy, the module of prioritized replay is introduced. By endowing priority to history training data and building the structure of Sumtree for storing and extraction, the retraining of neural network on history training data is realized. With the building of importance-sample weight, with priority update of neural network is realized. The combination of prioritized replay and NAF-DQL puts forward the energy management strategy of NAF-DQL+PR. In order to verify the effectiveness of the strategy, the training result is compared with results from dynamic programming (DP), NAF-DQL and DQL in aspects of engine working point, SOC, fuel consumption and training time. the result of

comparison indicated that the fuel economy of NAF-DQL+PR is only lower than that of DP while the training efficiency is much higher than others. The fuel economy and effectiveness of NAF-DQL+PR is verified.

A hyper parameter optimization method for deep reinforcement learning based energy management is established. Aiming at the problem of learning rate affecting training result a lot and the difficulty of finding a suitable learning rate by experience, genetic algorithm is introduced. By encoding learning rate to binary gene, initializing population, calculating fitness, crossover and mutation to make learning rate evolved itself. In order to verify the effectiveness of the optimization, NAF-DQL+PR that has been optimized is compared with un-optimized. Result indicates the optimized NAF-DQL+PR has higher fuel economy and training efficiency. This result verifies the effectiveness of hyper parameter optimization method for deep reinforcement learning based energy management.

A real-time debugging software for hybrid power source is carried out. Based on the CAN protocol, event trigger and timing trigger, a monitor module and real-time drawing module for engine, generator, battery and AC-DC in hybrid source are built. A real-time adjustment module for 100 parameters is built. This software is able to accelerate debugging process of hybrid vehicle and reduce research and development time.

A hardware in loop simulation for tracked plug-in hybrid electric vehicle is established. The model built in Chapter 2 is introduced into dSPACE as controlled object. With the MotoTron as vehicle control unit, computer, vehicle control unit and controlled object are connected together by Ethernet and CAN to build the hardware in loop simulation platform. The neural network that acquired from the optimized strategy in Chapter 5 is transplanted to vehicle control unit. By comparing NAF-DQL+PR and NAF-DQL in hardware in loop simulation, NAF-DQL+PR is proved to have higher fuel economy and better adaptivity when facing unfamiliar working condition. The real-time debugging software is tested in the hardware in loop simulation with the platform of dSPACE and MotoTron. The simulation result indicated the real-time of the software. The software is capable of monitoring the status of the hybrid power source and adjusting parameters in real time. The effectiveness of the software is verified.

Key Words: Tracked plug-in hybrid electric vehicle; energy management; deep Q-learning algorithm; prioritized replay; NAF-DQL+PR algorithm; hardware-in-the-loop simulation; real-time debugging software

目录

摘要	I
Abstract	III
第 1 章 绪论	1
1.1 研究背景与意义.....	1
1.2 混合动力车辆硬件技术发展现状.....	3
1.2.1 插电式混合动力车辆发展现状.....	3
1.2.2 增程式混合动力车辆发展现状.....	4
1.3 混合动力车辆能量管理策略研究现状.....	5
1.3.1 基于规则的能量管理策略.....	5
1.3.2 基于优化的能量管理策略.....	6
1.3.3 基于学习的能量管理策略.....	8
1.4 当前存在的问题及论文主要内容.....	9
1.4.1 混合动力能量管理存在的问题.....	9
1.4.2 混合动力车辆调试技术存在的问题.....	9
1.4.3 论文主要研究内容.....	10
第 2 章 串联式混合动力履带车建模研究	12
2.1 发动机-发电机组及高压功率模块模型构建	14
2.2 动力电池组模型构建.....	15
2.3 履带车辆动力学模型.....	17
2.4 仿真模型最优问题构建.....	18
2.5 本章小结.....	20
第 3 章 串联混合动力履带车能量管理策略研究	21
3.1 强化学习及深度强化学习理论概述.....	21
3.2 基于 NAF-DQL+PR 的能量管理策略研究	22
3.2.1 基于深度强化学习的能量管理方法构建.....	22
3.2.2 经验权值优先回放模型构建.....	25
3.2.3 基于经验权值优先回放的深度神经网络更新方法构建.....	26
3.3 算法仿真及结果分析.....	29

3.3.1 离线仿真工况.....	29
3.3.2 深度强化学习网络参数设定及训练环境.....	29
3.3.3 深度强化学习训练结果及对比.....	30
3.4 本章小结.....	33
第 4 章 基于深度强化学习的混合动力能量管理策略超参数优化.....	35
4.1 基于遗传算法的深度强化学习能量管理算法超参数优化.....	35
4.2 超参数优化算法验证.....	38
4.2.1 离线仿真工况.....	38
4.2.2 超参数算法验证参数设定及环境.....	38
4.2.3 GA-DQL 算法验证及对比	38
4.3 本章小结.....	41
第 5 章 基于深度强化学习的混合动力能量管理策略硬件在环仿真验证研究.....	42
5.1 混合动力源调试程序开发.....	42
5.1.1 系统总体设计.....	43
5.1.2 软件设计.....	44
5.2 串联混合动力履带车硬件在环仿真平台.....	47
5.2.1 实时仿真被控系统搭建.....	47
5.2.2 整车控制器开发及搭建.....	48
5.2.3 硬件在环仿真平台架构.....	49
5.3 基于 NAF-DQL+PR 的混合动力能量管理策略硬件在环验证	49
5.4 混合动力源实时调试程序试验调试.....	51
5.5 本章小结.....	52
结论	54
全文总结.....	54
创新点.....	55
工作展望.....	56
参考文献	57
攻读学位期间发表论文与研究成果清单	63
学术论文.....	63

参与科研项目	63
致谢	64

第1章 绪论

1.1 研究背景与意义

作为汽车工业大国，我国汽车销量连续十几年居世界第一，从 2005 年的 576 万辆增长到 2018 年的 2808 万辆，占世界销量的 30%^[1]。2018 年我国汽车工业及相关行业占国内生产总值 8.9%，汽车制造业对行业上下游影响力巨大。2019 年，我国汽车制造 2572.1 万辆，售出 2576.9 万辆，进口汽车 108.6 万辆。随着我国国民生活水平提高，汽车在百姓家庭中日益普遍。我国汽车保有总量及人均汽车保有量日益攀升。由此带来的能源和环境问题日益凸显。多个大城市提出限制油车购买及出行政策，借以解决汽车普及带来的环境问题。如图 1.1 所示。随着我国经济高速发展，汽车保有量和汽油消耗量持续攀升，北京市空气中 pm2.5 含量在 2014 年达到 $84.8 \mu\text{g}/\text{m}^3$ 启动治理措施，空气质量逐渐改善。

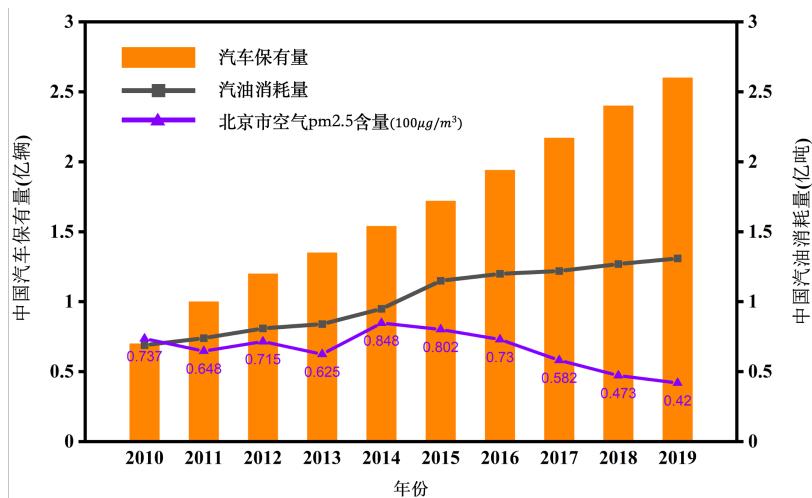


图 1.1 2010 年至 2019 年我国汽车保有量、汽油消耗量及北京市 pm2.5 含量($100\mu\text{g}/\text{m}^3$)

我国汽车市场以汽油和柴油车辆为主，城市小轿车、城市公交车及货运车辆占据绝大部分市场。由图 1.1 可见，以上三种车辆总量逐年上升。我国 2019 年石油表观消耗量 6.6 亿吨，对外依存度 70%。由于汽车大量消耗汽油、柴油且我国石油储存量及人均石油储量不及世界平均水平，我国严重依赖石油进口，国际油价波动严重影响我国国民日常生活。与此同时，由于汽车驾驶带来的环境问题日益严重，成为多个大城市主要治理目标。《中国环境空气质量管理评估报告(2018)》指出随着中国政府对新能源汽车战略的推进并提出世界最严格排放标准，京津冀等主要地区空气污染状况日益

好转。此外，由于我国汽车工业发展较国外开始晚，传统汽车行业被国外企业占领，我国汽车消费市场自主品牌占有率低于 40%，如图 1.2 所示。因此，为减少汽车尾气排放对环境的影响，为减少我国对进口石油的依赖，为提升我国汽车影响力，发展新能源汽车，研究新能源汽车技术，制造我国自主开发的新能源汽车势在必行。

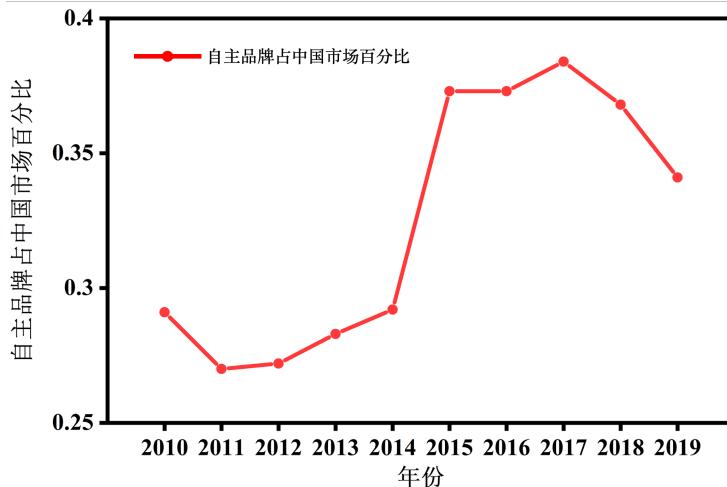


图 1.2 自主品牌占有中国市场份额

当下我国及世界新能源汽车研究主要集中于纯电动汽车，混合动力汽车（包括插电式和增程式混合动力汽车）及燃料电池汽车。其中纯电动汽车我国开展研究较早，通过针对电池、电机、电控系统及整车的研究、设计及制造，我国已形成一批具有自主知识产权的电动汽车产品、公司及上游零部件供应商。我国纯电动汽车产品已实现对海内外的辐射。燃料电池汽车将燃料电池装载于车上作发电装置，一般使用材料为氢气或储氢物质。与纯电动汽车不同，燃料电池汽车需搭载燃料电池反应堆及储氢装置，其优点在于近似于零排放、零污染运行平稳无噪声。混合动力汽车结合了传统燃油车及纯电动汽车的优势，作为传统燃油汽车和纯电动汽车中间的过度构型，混合动力汽车具有传统燃油车所不具备的低排放、高机动性和纯电动汽车所缺少的长续航里程及相对高的安全性。目前我国市场上混合动力汽车以插电式混合动力汽车为主，包括诸如比亚迪唐、比亚迪秦等多个我国自主品牌。

同时，由于国家 2019 年取消及减少了针对纯电动汽车购车优惠，如今纯电动汽车购车成本有所上升，消费者消费意愿下降，发展混合动力汽车，能继续执行国家清洁能源与环境保护的战略，减少我国对进口石油的依赖，并实现提升消费者购买欲望，提振我国汽车消费市场。

混合动力车辆一般由两种以上动力源提供动力，一般混合动力车辆使用动力源为

内燃机及动力电池，部分混合动力车辆使用超级电容及燃料电池做储能装置。于是，控制多个动力源协同工作并在行驶过程中实现最优控制，是设计和研发混动动力车辆的关键。混合动力车辆动力源能量分配是混合动力车辆的研究重点，面向于实际，混合动力车辆动力源的调试也是不可避免的关键环节。设计混合动力源能量管理策略提升混合动力车辆行驶过程中燃油经济性，设计混合动力源调试方法加快混合动力汽车研发和更新，是当前研究混合动力车辆的关键问题。

1.2 混合动力车辆硬件技术发展现状

混合动力车辆系统包含两个及以上能量源，多个能量源之间通过机械和电气的方式进行耦合，能量源不同的结合方式带来不同的混合动力车辆构型，不同构型具有相对的优点和缺点。常见的混合动力车辆构型分类方法有两种：插电式混合动力车辆和增程式混合动力车辆。

1.2.1 插电式混合动力车辆发展现状

插电式混合动力车辆构型如图 1.3 所示。图中所展示为一辆插电式串联混合动力汽车，组件包括发动机、盘式发电机、动力电池组、功率变化耦合装置、电机逆变器及两个永磁同步电机。其中发动机连接盘式发电机直接进行发电，所发三相电经过功率耦合装置 AC-DC 与动力电池所发直流电一并输入至电动机侧，经过电机逆变器将直流电变为三相电驱动永磁同步电机。当车辆运行速度较慢时所需功率较低，发动机将停止工作，动力电池组直接驱动电机。当车辆进入快速行驶阶段时，发动机-发电机组和动力电池组一起驱动电机工作。插电式混合动力车辆主要特点为其电池容量较大，可在充电桩等相关设施上充电。对于较短里程出行，插电式混合动力车辆可直接使用纯电行驶。由于插电式混合动力汽车既可以充电也可以加油，因此具有纯电动车和传统燃油车二者的优势，短途旅行及车速较慢时可使用纯电行驶模式，长途旅行或高速行驶时可以使用传统驱动模式或混动驱动模式，方便驾驶员出行。保证了使用及体验的同时提升燃油经济性，并实现节能减排。由于插电式混合动力车辆需要集成电动车和传统燃油车两套动力系统，因此插电式混合动力车辆的设计及制造技术及针对于插电式混合动力车辆发动机和电池的能量分配问题是当前研究的关键。

目前插电式混合动力车辆应用广泛，以其广泛的适应性在军用民用领域有诸多应用。中国市场上，比亚迪秦包含有一台 1.5L 的发动机和 13kWh 的动力电池组，能实现 70km 的纯电续驶里程，最高车速达到 185km/h，理想情况下百公里油耗仅 1L。保

时捷推出了一款 Panamera E-Hybrid，电池容量 14.1kWh，可实现 50km 纯电续驶里程并可达到最高车速 278km/h。军用方面，爱沙尼亚 Milrem 公司研制了模块化履带地面无人车 THeMIS 将发动机及动力电池分别安装于左右两履带，实现插电式混合动力履带车在复杂工况下的多模态行驶。本文研究目标为插电式混合动力履带车。

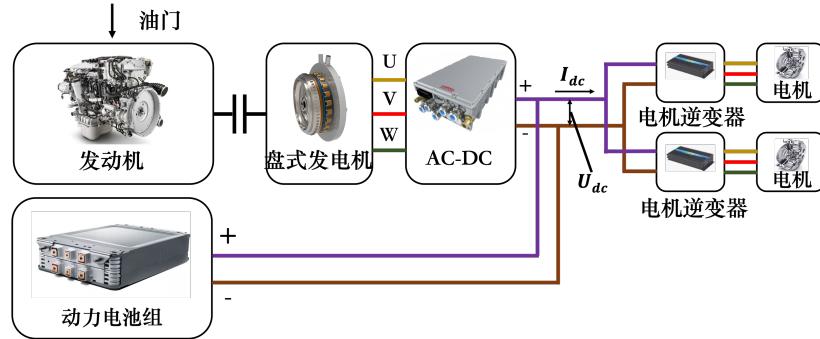


图 1.3 插电式串联混合动力车辆构型

1.2.2 增程式混合动力车辆发展现状

增程式混动汽车动力系统主要包括增程器、动力电池、驱动系统和电动机。与插电式混合动力车不同的是增程式混合动力汽车动力电池组较小，不可进行外界充电，只可由增程器对动力电池进行充电。增程器包括发动机及发电机。增程式混合动力汽车构造如图 1.4 所示，当电池荷电状态较高时，增程器不工作，动力电池驱动电机及驱动系统工作；当电池荷电状态较低时，增程器启动时给动力电池充电，动力电池驱动电机工作，多余的电能将被储存在电池中；当系统需求功率较大时，增程器和动力电池一同工作，驱动电动机。由于增程式混合动力车辆相比于纯电动汽车只多出增程器部分，其余系统架构相对相似，因此增程式混合汽车可以带来与纯电动汽车相似的驾驶体验并且由于增程器的存在，增程式混合动力汽车续驶里程相对纯电动汽车高。



图 1.4 增程式混合动力汽车构型

国内市场有多种增程式混合动力车型，其中理想制造 ONE 具有 1.2L 发动机

和 NEDC 续驶里程 180km 的电池，续驶里程可达到 700km。雪佛兰公司推出的沃蓝达轿车可实现纯电续航里程 80km，总里程达到 500km。增程式混合动力汽车相对于插电式混合动力汽车结构简单，可直接从纯电动汽车平台进行改造，能加续驶里程，保障纯电驾驶感，但由于增程式混动汽车本质上还是燃油驱动车轮，不符合我国节能环保的目标，在我国并不普及。

1.3 混合动力车辆能量管理策略研究现状

能量管理策略是混合动力汽车技术中关键的一环，在目前学界和工业界主流的能量管理技术和方法中，能量管理问题中亟待解决的就是混合动力源的能耗高效性和策略的实时性。随着近年技术的发展，混合动力车辆能量管理方法逐渐被分为两大类和多个分支。能量管理策略的分类如图所示：目前，目标于实现实时在线的能量管理策略是研究的热门。为实现混合动力车辆实时在线能量管理，基于规则的方法最先被提出。

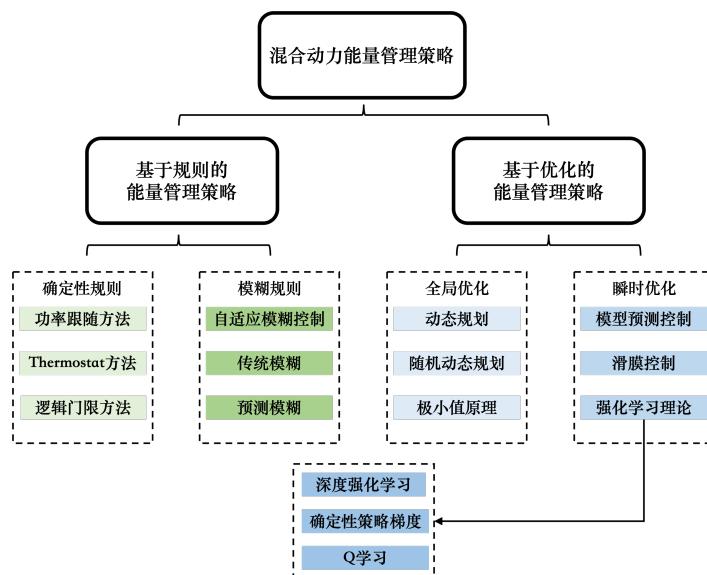


图 1.2 能量管理策略分类

1.3.1 基于规则的能量管理策略

基于规则的能量管理策略开发基于工程师在以往工作中积累的针对发动机、电池工作的经验。将预先制定的逻辑和阈值设计成发动机和动力电池能量分配的规则，在实际应用中直接进行能量分配。从而达到发动机长时间工作在高效工作区间的效果^[1-4]。基于规则的能量管理策略包括确定性规则策略和模糊规则策略两种。

文献[5]提出了基于 thermostat 的策略，基于混合动力车辆对功率的需求、驾驶员的加速需求和电池荷电状态（SOC）的三种状态量进行规则制定。所提出的基于规则的能量管理策略在高速路和城市道路上分别取得了 6% 和 11% 的能耗经济性提升。文献[6]提出基于电车荷电状态 SOC 创建规则阈值的基于规则的能量管理策略，在电池 SOC 低于最低阈值时启动发动机对电器进行充电，当电池 SOC 高于最高阈值时停止发动机，试验结果显示与传统内燃机车相比，该方法在 NEDC 工况及城市道路工况上分别提升了 10.1% 和 28.5% 的燃油经济性。文献[7]提出了一种针对插电式混合动力校车的基于规则的能量管理策略，该策略使用电荷消耗模式/电荷维持模式（CD/CS）控制车辆。为提高基于规则的能量管理策略对现实情况的兼容性，文献[7]提出了使用动态规划考虑驾驶员在 GPS 上预定路线优化基于规则的能量管理策略并将之应用于普锐斯插电式混合动力汽车，通过试验和方针，该方法使发动机工作在更有的工作点，提高了普锐斯在油耗方面的潜力。文献[8]建立了一种考虑车辆驾驶信息和车辆道路能量消耗而非车辆速度曲线的基于规则的能量管理策略，该策略针对不同混合动力汽车进行优化，通过在驾驶过程中实时比较车辆剩余能量与车辆剩余里程能量估计实现对发动机的控制，与传统 CD-CS 策略相比，该方法在燃油经济性上提升了 18.4% 并且显著降低发动机启停次数。文献[10]基于电池能从电网中获得大量能量，提出了一种优先使用电池、减少发动机使用的基于规则的能量管理策略。该方法被应用于普锐斯及另一种并联式混动汽车上，并与传统的基于规则的能量管理策略进行比较，结果表明在普锐斯及并联式混动车上分别取得 16% 和 6% 的燃油经济性提升。

通过预先设置的规则，在实际应用中基于规则的能量管理策略取得了不错的燃油经济性提升。但是，基于规则的能量管理策略对研发人员和工程师的直觉、工作经验要求高，在规则设计中易出现达不到最优燃油经济性效果的情况。因此针对基于规则的能量管理策略难以达到最优的问题，通过预先进行计算找到最优解的基于优化的能量管理策略被研究人员提出。

1.3.2 基于优化的能量管理策略

基于优化的能量管理策略通过建立混合动力源和混动动力车辆目标函数，以系统约束为边界条件，通过优化方法求解目标函数极限值。优化目标函数一般为混合动力车辆燃油消耗和动力电池消耗电量的加权求和。根据应用场景的区别，基于优化的能量管理策略分为全局优化策略和实时优化策略^[11-15]。

遗传算法、动态规划和粒子群算法等都是经过长期研究、日渐成熟的全局优化方法^[16-20]。通过应用全局优化方法，在某一具体工况上，混合动力车辆可以取得理论最优的目标函数求解。文献[21-25]基于贝尔曼方程（Bellman equation），将动态规划应用于能量管理策略开发中。文献[26]提出一种基于动态规划、模糊控制和驾驶模式识别的能量管理策略，利用动态规划算法找出模糊控制滑动特性的规律并将其应用于模糊控制策略中，最终使用驾驶识别将模糊控制应用于城市混合动力汽车。仿真所的结果接近动态规划全局最优，相比于基于规则的能量管理策略提升了 4.64% 的燃油经济性。文献[27]提出了基于遗传算法的能量管理策略，通过车辆驾驶速度将不同驾驶模式分类为不同的交通状况和驾驶方式，并利用遗传算法针对混合动力源组件参数进行优化，从而分别使用不同的能量分配策略，通过与传统优化方法的对比取得了 34% 的燃油经济性提升。文献[28]提出了一种可应用于实时控制的基于增强遗传算法（EGA）和模拟退火的能量管理策略，使用主要成分分析法将历史工况数据分类并构建成典型场景，使用提出的能量管理策略针对典型场景做离线优化并将控制策略在线激活使用，该方法接近动态规划的全局最优策略并优于传统的基于规则的策略。由于动态规划的最优性，其经常被用作于评价新开发算法的优劣。文献[29]提出了一种基于在线校正的预测能量管理策略，该方法对未来工况进行预测并使用动态邻近粒子群算法做离线优化，针对离线预测和在线情况的偏差提出在线校正算法。通过与 CD/CS 方法进行比较，该方法在预测准确情况下可提升燃油经济性 32.39%。文献[30-32]将粒子群算法应用于混合动力能量管理策略的开发中，粒子群算法被应用于在线控制器中用于计算基于规则的能量管理策略的阈值。由于基于全局优化的能量管理策略需要提前得到行程工况，对于道路上行驶的客车和货车的开发并不适用。并且基于全局优化的能量管理策略计算过程中需要多次迭代过程，计算时间长，计算机所需运算能力大，因此并不适用于一般车辆的能量管理策略开发。

目标于实时优化得到最佳能量管理策略，基于实时优化的能量管理策略被提出，包括等效燃油消耗最小策略、模型预测控制等方法^[33-38]。文献[39]通过考虑混合动力车辆燃油经济性和动力电池寿命，基于模型预测控制开发能量管理策略。文献[40]使用现代导航系统做离线路径规划和能量分配优化得到电池 SOC 参考点，并在在线应用中使用模型预测控制滚动预测 SOC 参考点对车辆进行能量管理，通过与 CD/CS 策略进行仿真比较，该方法取得了 15.18% 的燃油经济性提升。文献[41-43]将等效燃油消耗最小策略应用于混合动力车辆能量管理策略研究。文献[44]比较了三种实时能量管

理策略：自适应等效燃油消耗最小策略、最优控制和随机动态规划，并将它们与动态规划和庞德里亚金极小值原理进行比较，结果表明自适应等效燃油消耗最小策略生成的燃油消耗及 SOC 曲线最接近全局最优。

1.3.3 基于学习的能量管理策略

基于学习的能量管理策略是近年来日渐火热的研究方向。基于学习的能量管理策略属于基于实时优化的能量管理策略的一个分支。由于近年来机器学习和人工智能的崛起，深度学习和神经网络被引入至能量管理策略的开发中用于实现在线更新的能量管理策略^[45-48]。

强化学习最先被引入至混合动力车辆能量管理策略开发中^[49-54]。强化学习中使用的状态转移矩阵具有基于马尔可夫理论和免模型的特点。文献[55]提出了一种针对并联式混合动力车辆的基于速度预测和强化学习的预测式能量管理策略。使用模糊编码、最邻近方法和马尔可夫链预测速度，通过使用强化学习求解能量管理策略，该方法相比于基于规则的策略减少 17.54% 燃油消耗。文献[56]提出了一种结合规则和 Q 学习的实时能量管理策略，该方法将 Q 学习和 CD/CS 结合，实现在线状态切换，仿真结果接近 DP。然而，强化学习的运算中经常出现“维度爆炸”的情况，也即当训练过程中状态网格划分过细，强化学习的计算量呈指数增长，带来更多的计算时间和更大的计算机算力需求。深度强化学习通过结合强化学习和深度神经网络解决了这个问题。文献[57]提出了深度强化学习理论和框架，通过训练得到的算法可以在 Atari 游戏中击败专业人类玩家。通过这一理论的启发，深度强化学习被引入至能量管理策略的开发中。在深度强化学习中包含有两个深度神经网络，估值网络和目标网络，深度强化学习具有对于不同工况的自适应性和相对于全局优化算法更少的计算时间。文献[58]提出了一种针对于并联式混合动力公交车的深度强化学习算法，通过与强化学习和动态规划进行对比，在未经训练的全新道路中燃油经济性超过强化学习 5.6%，超过动态规划 90% 并且算法训练时间与强化学习相比减少了 72.3%。

由于深度强化学习具有深度神经网络，网络参数的设置对于深度强化学习的训练至关重要。诸如学习率、经验池大小等超参数对于深度神经网络的训练收敛快慢和训练结果影响巨大。随着基于深度强化学习理论的混合动力车辆能量管理策略发展，深度强化学习参数的增多带来更多的参数调节问题。传统上使用的基于经验的参数调节方法难以应对此问题。如今学术界针对神经网络参数优化问题的解决方法分为网格搜

索和随机搜索两大类。文献[59]提出了基于随机搜索的无模型强化学习参数优化方法。文献[60]提出了结合高斯过程和贝叶斯优化的强化学习超参数优化方法。针对应用于混合动力车辆能量管理策略的深度神经网络超参数优化还有待研究。

1.4 当前存在的问题及论文主要内容

1.4.1 混合动力能量管理存在的问题

混合动力能量管理策略的研究从混合动力汽车问世即得到研究人员广泛关注，诸多成果被应用于实际车辆，关于混合动力车辆能量管理策略相关研究相对充分。通过回顾前人研究成果，混合动力车辆能量管理策略研究有以下不足：

(1) 基于规则的混合动力车辆能量管理策略对复杂多变路况泛化能力不足。其预先制定的发动机、电池工作规则难以适应多种工况，且对于不同型号发动机和动力电池难以进行迁移。该策略缺乏理论支撑，需要由富有经验工程师进行开发且不易达到最优。

(2) 基于全局优化的能量管理策略可在离线状态下达到全局最优解。针对于实时路面情况，不能预先得知全部工况，该策略难以应用且全局优化算法求解时间较长，难以应对真实道路上复杂工况。

(3) 基于学习的能量管理策略可利用深度学习、强化学习等方法针对工况进行学习，在实际路面上针对不同工况进行调整。该方法需对神经网络、Q 表等进行长时间训练，耗时长，难以应用于实时控制。基于深度强化学习的方法需在训练前设定好超参数用于训练神经网络，超参数的选取对训练结果影响较大，影响该方法应用于混合动力车辆。

1.4.2 混合动力车辆调试技术存在的问题

(1) 当今市场上有相关软硬件具有基于 CAN 的混合动力车辆状态监控功能，相关软件价格昂贵，影响混合动力整车开发成本。

(2) 针对于混合动力源开发过程中的参数实时监测、参数实时调整和故障报错等功能尚无相关产品，导致混合动力车辆的开发常常陷入试验难以推进的境地，相关设计和算法也因调试过程漫长而难以推进。

综合上述，混合动力车辆研究中依旧存在相关问题。具有更强实时性、燃油经济性的能量管理策略是当前研究的重点和关键。近年来，机器学习技术被广泛开发和利

用于各行各业，由于机器学习具有学习能力、能在环境中习得知识，开展基于学习的能量管理策略研究是未来混合动力能量管理研究的方向。同时，开发一套能实时监控、调试混合动力源的软件也是研究的当务之急。

1.4.3 论文主要研究内容

本文以串联混合动力履带车为研究对象，针对串联混合动力履带车研究、设计中混合动力能量管理问题和混合动力履带车调试进行研究，提出基于深度强化学习的混合动力车辆能量管理策略和相关参数优化方法，设计一款应用于混合动力履带车调试的实时监控调参软件。通过相关研究以期提高混合动力车辆能量管理策略最优性、适应性和实时性，并助力混合动力履带车调试和能量管理算法的实现。

第二章进行了串联混合动力履带车建模研究，建立了混合动力履带车前向控制模型。根据履带车设计，建立混合动力履带车各组件数学模型，针对发动机、盘式发电机、动力电池和功率电子器件进行数学建模，考虑履带车工作需求针对履带车动力学进行数学建模，该数学模型被分别建立在 python 和 MATLAB/Simulink 环境中。该模型已经过实验室验证。

第三章进行了串联混合动力履带车能量管理算法研究，针对深度强化学习收敛速度慢、易陷入局部最优等问题，提出基于 NAF-DQL+PR 的能量管理策略，其以 NAF-DQL 算法为基础，通过引入经验权值优先回顾功能针对深度神经网络历史训练经验进行带权值回顾，使得神经网络收敛更快，训练变高效，训练效果更好。将 NAF-DQL+PR 应用于混合动力车辆能量管理问题中，构建基于 NAF-DQL+PR 的能量管理策略，利用工况数据对算法网络进行训练并将其与 NAF-DQL 和动态规划算法进行对比，验证其高效性和最优性。

第四章进行了串联混合动力履带车能量管理算法策略超参数优化研究，NAF-DQL+PR 算法中对神经网络训练影响较大的超参数是学习率，针对其进行优化。将学习率编码为二进制遗传因子，利用遗传算法使学习率遗传因子种群自我进化，并将 NAF-DQL+PR 算法收敛所得油耗作为遗传因子适应度进行反馈。通过训练得到最佳表现遗传因子。将该遗传因子应用于 NAF-DQL+PR 训练，与其他学习率进行比较，验证优化算法的最优性。

第五章进行了串联混合动力履带车调试软件开发研究并进行硬件在环仿真实验验证基于 NAF-DQL+PR 的混合动力车辆能量管理策略在实时控制下的效果。针对于

混合动力车辆开发研制中参数监测、调试和故障提示等需求痛点，基于 LabVIEW 软件和高速 CAN 通讯协议搭建混合动力源实时监测调试程序。该程序集合并解析高速 CAN 通讯网络传输的信号，使用 LabVIEW 软件进行实时显示。并建立车辆参数实时修改框架，实现针对于 100 个参数的实时修改。该程序还支持针对混合动力源参数实时绘图显示功能，有利于对动力源状态监测。该程序具有实时性，有利于推进混合动力车辆研究。基于 MotoTron 和 dSPACE 实时仿真平台搭建串联混合动力履带车硬件在环实验平台。基于 MotoTron 控制器和 dSPACE 实时仿真平台软件开发，将 NAF-DQL+PR 和 NAF-DQL 两种算法移植入硬件在环仿真平台，完成两种算法在实时控制下的能耗经济性比较。验证基于 NAF-DQL+PR 的能量管理策略的高效性。利用该实时仿真平台对第五章所研制混合动力源实时调试程序进行硬件在环仿真实验，验证该软件实时性及有效性。

基于以上所述各章节研究内容，本文个章节概要、结构如图 1.6 所示。

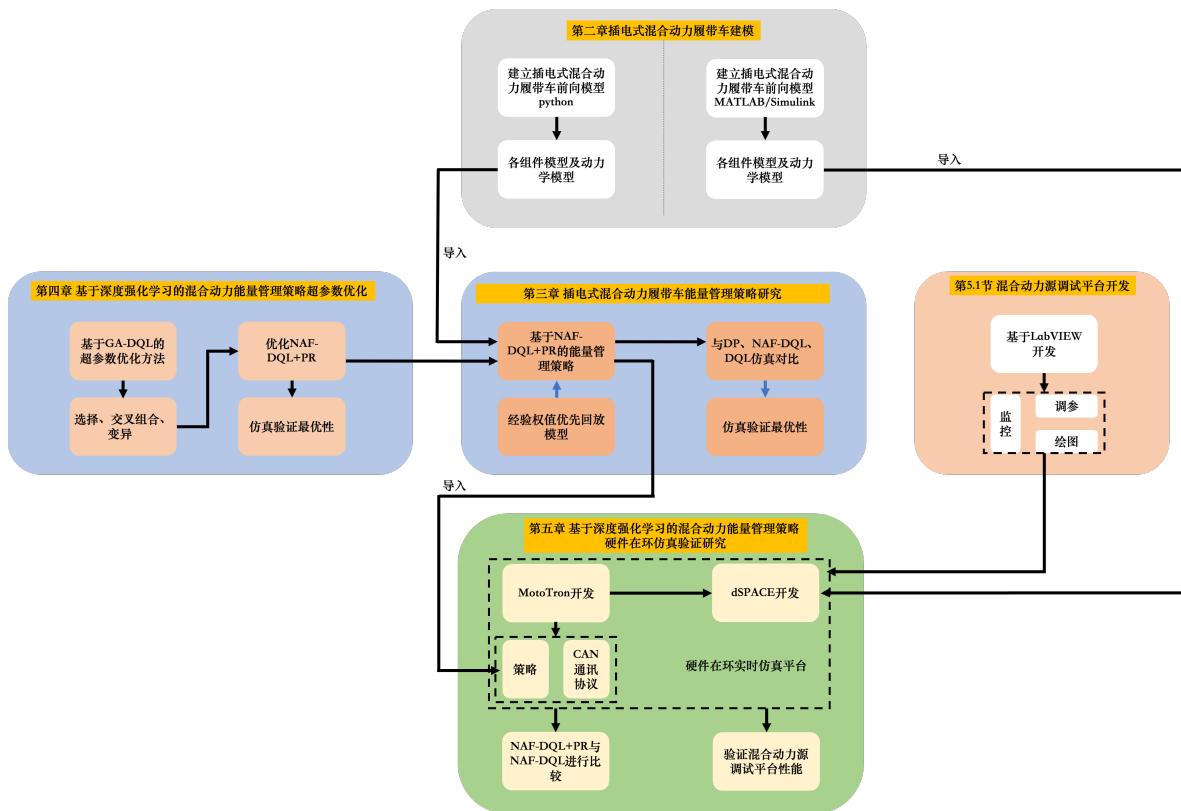


图 1.3 论文各章节内容结构及联系

第2章 串联式混合动力履带车建模研究

本文所研究串联混合动力无人履带车为课题组自研串联式混合动力车，基于使用需要，为方便作战需求，该款混合动力无人履带车被设计为串联形式，为减少安置于底盘的传动装置并充分发挥串联式履带车电气化、轻量化的优势。该款履带车取消了传统履带车汇流分动装置和混合动力车辆上常见的传动变速装置，推进了履带车辆电气化。该款履带车串联式混合动力驱动系统如图 2.1 所示。目标于解决混合动力履带车能量管理问题，针对履带车进行简化，只考虑履带车行驶过程带来的能量需求，忽略散热系统、变压输出系统等能量需求。各组件包括发动机、盘式发电机、动力电池组、高压功率单元、电机驱动系统、高速 CAN 总线、整车控制器、电池管理系统和发动机控制器。其中发动机和盘式发电机组成发动机-发电机组（engine generator set, EGS）。发动机-发电机组和动力电池组所输出电流由高压功率单元整合，从三相交流电变为直流电输入至电机控制器，用电机控制器将直流电转化为三相交流电驱动安装于履带内的永磁无刷直流电机。从工况中计算得来的混合动力履带车辆需求功率被输入至整车控制器中（VCU）。整车控制器通过控制发动机控制器、电池管理系统和电机逆变器实现整车控制并实现针对发动机-发电机组和动力电池组之间的能量分配。整体而言，当履带车需求功率较大时，发动机-发电机组与动力电池组共同驱动电机，当需求功率较小时，发动机-发电机组停止工作，动力电池组驱动电机，当动力电池 SOC 较低时，发动机-发电机组为其充电或停车使用外部充电桩进行补电。本款混合动力无人履带车基本参数如表 2.1 所示。

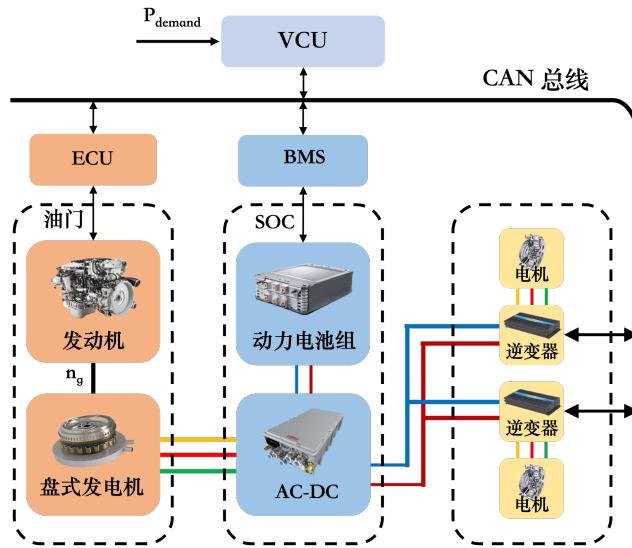


图 2.1 串联混合动力履带车结构

表 2.1 串联混合动力履带车参数

参数名	符号	取值	单位
履带车质量	m	1200	kg
履带中心距	B	1.3	m
驱动轮半径	r	0.25	m
履带接地长度	L	1.6	m
迎风面积	A	2.94	m ²
滚动阻力系数	f	0.0494	/
发电机转动惯量	J_g	2.2	kg m ²
发动机转动惯量	J_e	3.6	kg m ²
发电机电动势系数	K_e	2	Vsrad-2
发电机等效阻抗系数	K_x	0.0098	Nm A-2
电池最低 SOC	SOC_{min}	0.3	/
电池最高 SOC	SOC_{max}	0.8	/
动力电池组容量	C_b	45.5	Ah
空气阻力系数	C_D	0.7	/

2.1 发动机-发电机组及高压功率模块模型构建

本文研究对象串联混合动力无人履带车所使用发动机为电子调速柴油机，发电机采用永磁同步盘式发电机。发动机输出轴和发电机主轴连接，启动发动机带动发电机发电。本文研究能量管理策略目的在于在满足车辆行驶需求前提下节省发动机燃油消耗，通过对发动机特性的了解，发动机动作在高效工作区可实现高能量输出和低燃油消耗从而达到高能效经济性。为方便建模进行研究，将发动机-发电机组和高压功率模块一并进行建模简化，如图 2.2 所示。

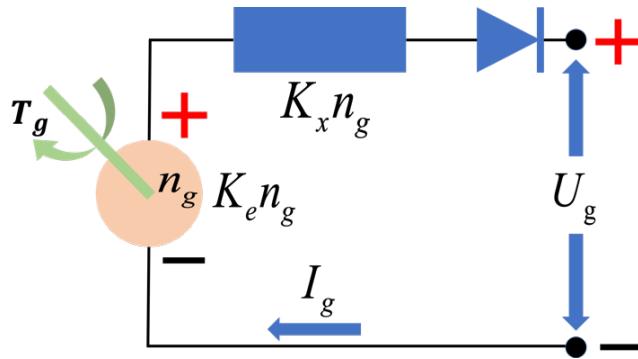


图 2.2 发动机-发电机组和高压功率模块简化结构

在忽略发电机三相绕组内阻值变化和发动机传输至发电机转矩损失的情况下，发电机输出电压 U_g 、输出电流 I_g 、发动机输出扭矩 T_{eng} 及发电机电磁扭矩关系如下：

$$\begin{cases} U_g = K_e n_g - K_x n_g I_g \\ T_g = K_e I_g - K_x I_g^2 \\ \frac{T_{eng} - T_g}{i_{eg}} = \frac{2\pi}{60} \left(\frac{J_e}{i_g^2} + J_g \right) \frac{dn_g}{dt} \\ n_g = i_{eg} n_{eng} \end{cases} \quad (2.1)$$

其中 K_e 表示电动势力系数， n_g 表示盘式发电机转速，单位是 rad/s， $K_x n_g$ 代表发电机等效阻抗，这里 $K_x = 3PL^g/\pi$ ， K_x 是发电机等效阻抗系数， L^g 代表发电机电枢同步电感。 J_e 和 J_g 分别是发动机和发电机转动惯量。 i_{eg} 为发动机与发电机传动比，由于设计上采用直接固连，因此本文中 $i_{eg} = 1$ 。针对发动机-发电机组建模还需考虑其物理特性，也即发动机转速需不能高于其最高转速或低于最低转速，发电机转速不能高于其最高转速或低于最低转速。发动机-发电机组模型被建立于 MATLAB/Simulink 和 python 环境中，MATLAB/Simulink 中仿真模型如图 2.3 所示，python 代码搭建如图 2.4 所示。

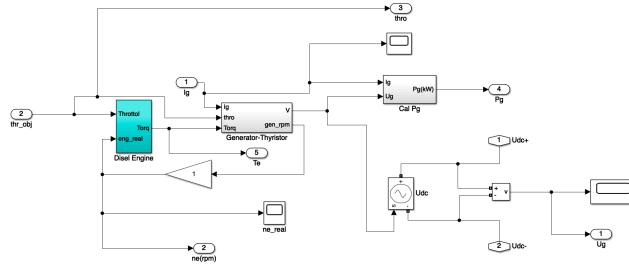


图 2.3 Simulink 发动机-发电机组仿真模型

```

fc_map_spd = np.array([1500, 1600, 1700, 1800, 2000, 2200, 2300, 2450, 2550, 2650, 2800, 3000, 3050]) # Map Speed in r/min
fc_map_trq = np.array([
[0,5,10,15,20,25,30,35,40,45,50,55,60,65,75,80,85,90,95,100,105,110,115,120,125,130]) # Map Torque in Nm
fc_fuel_map = np.array([
[0.000, 0.123, 0.238, 0.295, 0.314, 0.349, 0.389, 0.428, 0.468, 0.511, 0.555, 0.600, 0.646, 0.692,
0.740, 0.781, 0.831, 0.881, 0.931, 0.995, 1.048, 1.194, 1.176, 1.239, 1.304, 1.391, 1.475],
[0.000, 0.131, 0.251, 0.302, 0.330, 0.373, 0.415, 0.457, 0.502, 0.545, 0.592, 0.637, 0.685, 0.733,
0.781, 0.830, 0.881, 0.934, 0.992, 1.048, 1.188, 1.168, 1.239, 1.304, 1.391, 1.475],
[0.000, 0.138, 0.266, 0.331, 0.353, 0.405, 0.443, 0.492, 0.536, 0.585, 0.630, 0.680, 0.728, 0.778,
0.828, 0.879, 0.929, 0.982, 1.036, 1.097, 1.152, 1.215, 1.278, 1.339, 1.406, 1.477, 1.543],
[0.000, 0.146, 0.284, 0.342, 0.402, 0.427, 0.483, 0.529, 0.576, 0.627, 0.674, 0.726, 0.776, 0.827,
0.880, 0.932, 0.986, 1.039, 1.092, 1.146, 1.208, 1.259, 1.317, 1.376, 1.433, 1.545, 1.637],
[0.000, 0.161, 0.311, 0.409, 0.469, 0.498, 0.561, 0.611, 0.665, 0.721, 0.774, 0.833, 0.898, 0.948,
1.006, 1.063, 1.123, 1.183, 1.243, 1.303, 1.363, 1.424, 1.484, 1.550, 1.614, 1.695, 1.822],
[0.000, 0.177, 0.342, 0.450, 0.528, 0.559, 0.630, 0.688, 0.749, 0.813, 0.871, 0.939, 1.002, 1.068,
1.134, 1.199, 1.268, 1.338, 1.405, 1.474, 1.542, 1.614, 1.684, 1.750, 1.836, 1.904, 2.008],
[0.000, 0.185, 0.357, 0.479, 0.545, 0.582, 0.654, 0.712, 0.779, 0.844, 0.908, 0.978, 1.043, 1.115,
1.184, 1.255, 1.326, 1.397, 1.470, 1.541, 1.619, 1.699, 1.778, 1.857, 1.935, 2.024, 2.101],
[0.000, 0.198, 0.381, 0.512, 0.580, 0.621, 0.697, 0.756, 0.828, 0.894, 0.963, 1.036, 1.107, 1.183,
1.257, 1.333, 1.409, 1.486, 1.565, 1.643, 1.727, 1.810, 1.895, 1.980, 2.062, 2.153, 2.241],
[0.000, 0.206, 0.395, 0.535, 0.612, 0.648, 0.736, 0.796, 0.868, 0.946, 1.009, 1.086, 1.159, 1.237,
1.316, 1.395, 1.474, 1.553, 1.637, 1.721, 1.806, 1.882, 1.979, 2.067, 2.156, 2.246, 2.337],
[0.000, 0.214, 0.412, 0.565, 0.646, 0.684, 0.789, 0.837, 0.911, 0.981, 1.059, 1.148, 1.217, 1.299,
1.383, 1.462, 1.546, 1.628, 1.715, 1.802, 1.889, 1.976, 2.064, 2.155, 2.248, 2.343, 2.438],
[0.000, 0.226, 0.436, 0.620, 0.711, 0.746, 0.846, 0.915, 0.991, 1.069, 1.149, 1.236, 1.319, 1.403,
1.493, 1.582, 1.667, 1.759, 1.837, 1.928, 2.017, 2.113, 2.208, 2.305, 2.405, 2.504, 2.601],
[0.000, 0.242, 0.467, 0.694, 0.788, 0.833, 0.933, 1.010, 1.101, 1.191, 1.278, 1.373, 1.462, 1.559,
1.652, 1.746, 1.842, 1.936, 2.026, 2.135, 2.237, 2.339, 2.442, 2.548, 2.659, 2.768, 2.881],
[0.000, 0.247, 0.476, 0.723, 0.802, 0.861, 0.956, 1.032, 1.129, 1.215, 1.310, 1.404, 1.498, 1.595,
1.690, 1.788, 1.883, 1.983, 2.081, 2.183, 2.285, 2.390, 2.498, 2.602, 2.710, 2.821, 2.933])
T, w = np.meshgrid(fc_map_trq, fc_map_spd)
thr = min(1,action[0]) # 节气门开度: thr [0,1]
ng = s[1]* $(3050 - 1500) + 1500$ 
eng_spd = max(ng / ieg, 1500)
eng_spd = min(eng_spd, 3050) # 发动机转速 rpm
ng_rad = eng_spd/60
w_gen = ng_rad * 2*3.14
Tmax = np.interp(eng_spd, [1500, 1600, 1700, 1800, 2000, 2200, 2300, 2450, 2550, 2650, 2800, 3000, 3050], [77.9, 89.7, 98.6, 123.7, 122.8, 122.9, 121.0, 120.3, 122.4, 125.0, 124.5, 120.1, 114.5]) # Engine's max torque
eTq = thr * Tmax
if w_gen == 0:
    Udc = V_bat
else:
    b = -(V_bat / R_in + ke / kx)
    aa = 1 / R_in + 1 / (kx * w_gen)
    Udc = (-b + np.sqrt(b**2 - 4 * aa * Pdem)) / (2 * aa)
    IGS = int(V_bat > ke * w_gen)
    Udc = (1 - IGS) * Udc + IGS * (
        V_bat - (V_bat - np.sqrt(V_bat**2 - 4 * R_in * Pdem)) / 2.)
    Ig = (1 - IGS) * (Pdem / Udc - Ib)
    Pgen = Udc * Ig / 1000.
    V_gen_on = ke * w_gen - kx * w_gen * Ig
    flag2 = int(ng <= 1000 * ieg)
    flag2 = int(ng <= 1000 * ieg)
    ng_ = flag2 * int((thr>0.3)*(spd_lim*ieg) + (1-flag2)*((eTq/ieg*ita2-(ke * Ig - kx * Ig*2) - friction) / c1) + ng

```

图 2.4 python 发动机-发电机组仿真模型代码

2.2 动力电池组模型构建

根据本文研究需求，对动力电池进行性能建模，忽略其电化学、热学和元器件电气变化特性。常用电池性能模型包括 BP 神经网络模型、Rint 模型、Thevenin 模型、PNGV 模型、部分放电模型及其他复合模型^[61]，本文综合考虑多种模型建模效果选取 Rint 模型对混合动力无人履带车动力电池组进行建模。根据 Rint 模型建立动力电池输出电压、输出功率、输出电流和电池荷电状态 SOC 变化率的关系：

$$\begin{cases} U_b = V_{bat} - I_b R_{intdis} \\ P_b = V_{bat} I_b - I_b^2 R_{intdis} \\ \dot{SOC} = \frac{\left(V_{bat} - \sqrt{V_{bat}^2 - 4R_{intdis}P_{bat}(t)} \right)}{2C_b R_{intdis}} \\ I_b = \frac{\left(V_{bat} - \sqrt{V_{bat}^2 - 4R_{intdis}P_b} \right)}{2R_{intdis}} \end{cases} \quad (2.2)$$

其中 V_{bat} 表示动力电池组开路电压，其随电池荷电状态变化而改变， R_{int_dis} 表示动力电池组内阻值， C_b 表示动力电池组容量。该混合动力履带车采用三元锂电池，该电池单体输出电压和电池荷电状态关系如图 2.5 所示，动力电池模型被建立于 MATLAB/Simulink 和 python 环境中，MATLAB/Simulink 中仿真模型如图 2.6 所示，python 仿真模型如图 2.7 所示。

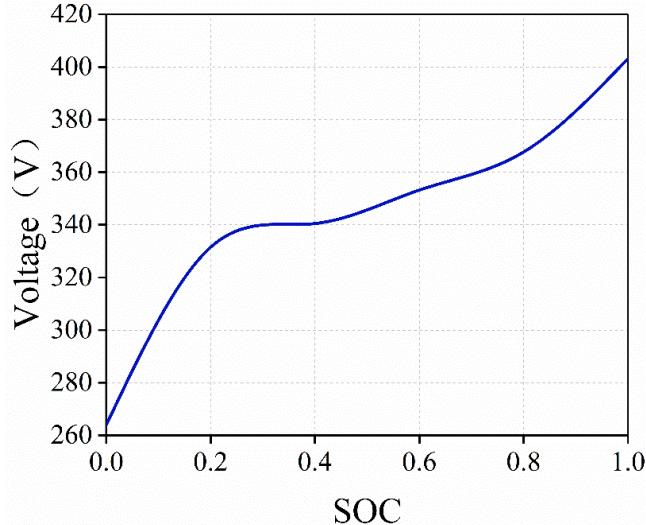


图 2.5 电池单体输出电压和电池荷电状态关系

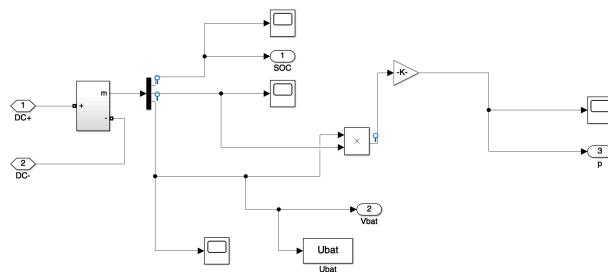


图 2.6 Simulink 动力电池组模型

```

c_batt = 84
xsoc = np.array([0, 0.4, 0.6, 0.8, 1]) # SOC
yvb = np.array([264, 340.5, 353.2, 367.7, 403.2]) # 开路电压
R_in = 0.1
soc = s[0]
soc = max(soc,0)
soc = min(soc,1)
Ib = (1 - IGS) * (V_bat - Udc) / R_in + IGS * (
    V_bat - np.sqrt(V_bat**2 - 4 * R_in * Pdem)) / (2 * R_in)
Pbat = Udc * Ib / 1000.
V_bat_on = V_bat - Ib * R_in
soc_ = soc - Ts * Ib / c_batt
soc_ = min(1., soc_)

```

图 2.7 python 动力电池组模型代码

2.3 履带车辆动力学模型

基于混合动力能量管理问题车辆建模的目标在于接收算法输出控制量进行计算并输出下一时刻状态量作为反馈，从而建立一个闭环系统，使得深度强化学习方法得以被训练。因此将混合动力履带车模型视为刚体，简化混合动力履带车辆动力学模型，忽略该款混合动力履带车双侧履带与地面间复杂的动力学关系且只考虑履带车前进、后退和转向运动。通过建立混合动力履带车在 XY 平面上运动模型完成建模。

混合动力履带车推进功率包括两部分：向前行驶功率和转向行驶功率，根据《坦克行驶原理》^[62]建立混合动力履带车动力学模型。

$$P_{\text{demand}} = (F_r + F_i + F_w + F_j)v_{\text{aver}} + M_r\omega \quad (2.3)$$

$$\begin{cases} F_r = mgf\cos\alpha \\ F_i = mgs\sin\alpha \\ F_w = \frac{C_D A v_{\text{aver}}^2}{21.15} \\ F_j = \delta m \frac{dv}{dt} \\ v_{\text{aver}} = \frac{v_1 + v_2}{2} \\ \omega = \frac{v_1 - v_2}{B} \end{cases} \quad (2.4)$$

式中 F_r 、 F_i 、 F_w 和 F_j 分别表示履带车辆所受滚动阻力、坡道阻力、空气阻力和惯性力。 v_1 、 v_2 和 v_{aver} 分别代表左侧履带速度，右侧履带速度和双侧履带平均速度。 $M_r\omega$ 表示履带车转向功率，其中 M_r 代表履带车辆转向阻力矩， ω 是履带车转向角， B 代表双侧履带中心距离， C_D 是车辆空气阻力系数， A 表示履带车迎风面积， m 表示履带车质量， dv/dt 是履带车辆加速度， g 是重力加速度， f 是履带车滚动阻力系数。履带车辆转向阻力矩 M_r 计算过程如下：

$$\begin{cases} M_r = \frac{1}{4} \mu mgL \\ \mu = \frac{\mu_{max}}{0.925 + 0.075\rho} \\ \rho = \frac{R}{B} \end{cases} \quad (2.5)$$

其中 L 代表单侧履带接地长度, ρ 代表履带车相对转向半径, μ 和 μ_{max} 分别代表履带车转向阻力系数和履带车做 $0 \sim B/2$ 转向时转向阻力系数。履带车辆动力学模型被建立于 MATLAB/Simulink 和 python 环境中, MATLAB/Simulink 中仿真模型如图 2.8 所示, python 仿真模型如图 2.9 所示。

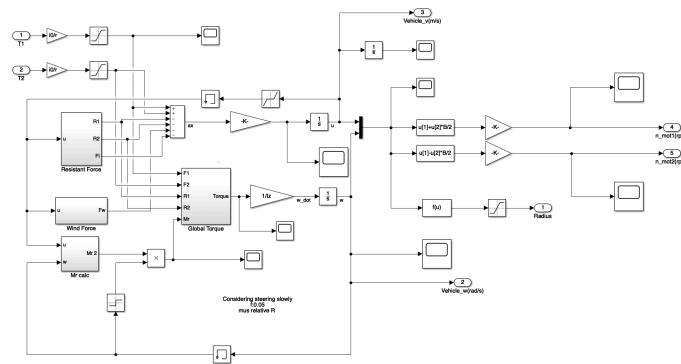


图 2.8 Simulink 履带车辆动力学模型

```

def initCycle(self):
    m = 1200
    g = 9.8
    A = 0.91 # area
    Cd = 0.7
    f = 0.0494
    Ts = 1 # for training
    with open('./data/CYC_US06.csv') as f_obj:
        cycinfo = pd.read_csv(f_obj, usecols=['Time', 'cyc_kmh'])
    time = cycinfo['Time']
    Velocity = cycinfo['cyc_kmh']
    Pdem = np.zeros((len(time) - 1, 1))
    Acc = np.zeros((len(time) - 1, 1))
    # Generate power from cycle
    for i in range(1, len(time)):
        Acc[i-1] = (Velocity[i] - Velocity[i-1])/3.6/Ts # 加速度 m/s^2
        Faero = (Cd*A * (Velocity[i - 1])**2)/21.15 # 空气阻力
        Ff = m*g*f # 滚动阻力
        Pdem[i-1] = (Faero + Ff + m*Acc[i-1])*Velocity[i-1]/3.6/0.78/1000 # 需求功率 Kw
    Pdem = np.append(Pdem, 0)
    Acc = np.append(Acc, 0)
    Velocity = Velocity.values.reshape(len(time), 1)
    Pdem = Pdem.reshape(len(time), 1)
    Acc = Acc.reshape(len(time), 1)
    CycInfo = np.hstack((Velocity, Pdem, Acc))
    return CycInfo

```

图 2.9 Simulink 履带车辆动力学模型

2.4 仿真模型最优问题构建

针对混合动力履带车能量管理问题, 基于深度强化学习建立能量管理策略, 其中代价函数、状态变量和控制量的选取和构建非常关键。本文选取混合动力履带车发动机电子油门开度为控制量, 其由控制策略计算生成。选取混合动力履带车发动机-发

电机组转速、动力电池组荷电状态及混合动力履带车需求功率为系统状态量。在系统优化时间范围 $[t_0, t]$ 内，系统状态变量 $S_t = (n_g(t), SOC(t), P_{demand})$ 满足如下关系：

$$\dot{S}_t = f(S_t, a_t) \quad (2.6)$$

式中 f 为状态方程，其构建基于式 (2.1) 和 (2.2)， a_t 是控制变量序列，在离散的时间内，一系列的控制序列使得在此时间范围内运行的混合动力履带车性能指标达到最优，本文考虑履带车在选取工况上燃油消耗量和电池荷电状态 SOC 波动两方面因素构建串联混合动力无人履带车能量管理代价函数：

$$J = - \left[\int_{t_0}^t \alpha \dot{f}_{rate}(t) dt + \beta [SOC(t) - SOC(t_0)]^2 \right] \quad (2.7)$$

其中 \dot{f}_{rate} 表示发动机燃油消耗率， $[t_0, t]$ 表示履带车工作时间。 α 及 β 是两 0、1 之间的正参数且 $\alpha + \beta = 1$ ； $SOC(t_0)$ 表示初始电池荷电状态，代价函数的创建为减少履带车燃油消耗并减少动力电池 SOC 剧烈波动，由于使用串联混合动力履带车，不需要动力电池组在行驶完具有初始电量，仅需考量电池寿命，减少剧烈充放电次数。

此外，混合动力源的运行还需满足以下约束：

$$\begin{aligned} P_{b,min} &\leq P_b \leq P_{b,max} \\ n_{en,min} &\leq n_{en}(t) \leq n_{en,max} \\ T_{en,min} &\leq T_{en} \leq T_{en,max} \\ \dot{S}_t &= f(S_t, a_t) \\ SOC_{min} &\leq SOC(t) \leq SOC_{max} \end{aligned} \quad (2.8)$$

式中， $P_{b,max}$ 、 $P_{b,min}$ 分别是动力电池组最大和最小功率， $n_{en,max}$ 、 $n_{en,min}$ 分别是发动机最大转速和最小转速， $T_{en,max}$ 、 $T_{en,min}$ 分别是发动机最大和最小输出转矩， SOC_{max} 、 SOC_{min} 分别是动力电池组最大和最小 SOC 值。

表 2.2 第二章所使用参数

参数名称	符号	参数名称	符号
发电机输出电压	U_g	履带车惯性力	F_j
发电机输出电流	I_g	履带车左侧履带速度	v_1
发动机输出扭矩	T_{eng}	履带车右侧履带速度	v_2
发电机电磁转矩	T_g	履带车双侧履带平均速度	v_{aver}
发电机电动势系数	K_e	履带车转动阻力矩	M_r
发电机转速	n_g	履带车转向角速度	ω
发电机等效阻抗系数	K_x	双侧履带中心距	B
发动机转动惯量	J_e	履带车迎风面积	A
发电机转动惯量	J_g	履带车质量	m

参数名称	符号	参数名称	符号
主减速比	i_{eg}	重力加速度	g
动力电池组开路电压	V_{bat}	滚动阻力系数	f
动力电池组输出电流	I_b	单侧履带接地长度	L
动力电池组内阻	R_{int_dis}	相对转向半径	ρ
动力电池组容量	C_b	履带车转向阻力系数	μ
履带车滚动阻力系数	F_r	履带车 $0 \sim B/2$ 转向阻力系数	μ_{max}
履带车坡度阻力系数	F_i	燃油消耗率	f_{rate}

2.5 本章小结

本章建立了串联混合动力履带前向仿真模型，完成针对于混合动力履带车发动机、发电机、动力电池组、高压功率模块及履带车辆动力学建模，该数学模型被建立于 python 及 MATLAB/Simulink 环境中，便于之后进行能量管理策略开发和硬件在环仿真实验。该模型具有相当仿真精度，可被用于混合动力能量管理策略训练及硬件在环仿真实验验证中，该模型有助于完成能量管理算法的训练及验证。

第3章 串联混合动力履带车能量管理策略研究

基于强化学习的混合动力能量管理策略能实现针对目标工况自我更新和学习，其具有基于规则的能量管理策略所不具备的对不同工况的适应性和比基于全局优化的能量管理策略训练更快的训练速度。本章针对基于强化学习和因其发展而来的深度强化学习理论的能量管理策略之不足，开展基于深度强化学习能量管理策略开发。基于 NAF-DQL 算法开展优化工作，提出结合经验权值优先回放的混合动力车辆能量管理策略，实现深度强化学习算法快速收敛并确保其最优性，实现混合动力能量管理策略燃油经济性并减少运算时间，推进基于深度强化学习的能量管理策略实时性研究。本章首先阐述强化学习理论和深度强化学习理论，之后开展基于深度强化学习的能量管理算法研究，详尽阐述 NAF-DQL+PR 算法在混合动力车辆能量管理问题中的训练框架构建、网络更新方法、经验权值优先回放方法及训练过程。最后通过实验工况数据训练优化算法，验证其最优性。

3.1 强化学习及深度强化学习理论概述

强化学习是一种无监督学习，无需对数据进行标定，通过在环境中训练智能体即可得到最佳策略。强化学习中包含智能体和环境两主要部分。智能体中存放强化学习算法，其产生动作，动作被馈入之环境中得到相应奖励或惩罚，智能体通过该奖励评价该动作优劣从而对动作进行修改，不断的试错使得智能体最终找到最优策略。

强化学习中， Q 值被定义为动作状态值函数，也即在状态 s 下进行动作 a 所得回报 $Q(s, a)$ 。强化学习中针对状态和动作及 Q 值建立的映射关系被称为 Q 表，通过不断训练更新 Q 表最终找到对应 Q 值最大的动作集输出为策略。 Q 学习中为建立 Q 表将状态 s 和动作 a 进行离散，当系统为追求高控制精度将离散精度提高时， Q 表维度呈平方增大， Q 表更新所需算力及时间呈指数增长，针对此问题，深度强化学习理论被提出。作为深度强化学习的一种，深度 Q 学习（DQL）通过深度神经网络，建立状态量和 Q 值的映射关系，实现强化学习表格到深度强化学习中神经网络的转变，有效提升计算效率，减少对计算机算力需求。建立双神经网络结构，通过计算目标网络与估值网络差值更新估值网络参数，实现深度神经网络训练。通过引入经验回放策略，实现深度神经网络对历史经验再次学习，降低训练数据间相关性，有效避免陷入局部最优。深度 Q 学习通过训练神经网络，求得最大 Q 值及其对应已离散动作，将一系

列在对应状态下具有最大 Q 值的动作按时序组合成为策略输出。

深度 Q 学习实现连续状态量输入，对控制量仍进行离散化处理，所得控制量精度较低，针对连续和高精度控制问题，难以达到相应控制精度。此外，通过 Q 值回溯动作较为耗时，影响该策略应用于混合动力车辆能量管理策略。基于 NAF-DQL 的混合动力能量管理策略因而被提出，该策略沿用深度 Q 学习框架，针对深度神经网络输出端添加归一化优势函数 (normalized advantage function, NAF)，该函数将深度神经网络输出 Q 值与动作直接建立映射，也即建立了系统状态量和动作之间的直接联系，此种改变带来的优势有两点，一是减少通过 Q 值寻找动作的耗时，通过深度神经网络训练直接确定动作，二是避免对于动作控制量离散化处理，经由 NAF-DQL 训练得到的动作策略是连续的，可以实现针对连续、高精度系统的控制。在混合动力车辆上，可以实现连续的油门输出，考虑到发动机高效工作区的存在，连续且高精度的油门控制量有助于混合动力车辆能量管理策略提升混合动力车辆燃油经济性。

3. 2 基于 NAF-DQL+PR 的能量管理策略研究

针对深度 Q 学习的不足，本文根据 NAF-DQL 算法进行改进，通过加入经验权值优先回放模块 (PR) 与 NAF 和 DQL 有机结合，实现针对深度 Q 学习的能量管理策略在控制最优性和实现高效性上的大幅提升。本文以串联混合动力履带车为研究对象，详尽阐述基于 NAF-DQL+PR 的混合动力能量管理算法深度神经网络基本原理、经验权值优先回放过程及深度神经网络更新方法。

3. 2. 1 基于深度强化学习的能量管理方法构建

深度强化学习由于嵌入深度神经网络结构，需考虑神经网络训练收敛问题，为此引入双神经网络结构，即估值网络和目标网络。利用目标网络作为估值网络的监督。估值网络的结构与目标网络完全一致，在网络更新上目标网络延迟几个周期，通过延迟更新避免网络训练陷入局部最优。此外深度强化学习算法中还包含对历史经验的再次训练，也即 replay-buffer 结构。

本文目标于串联混合动力履带车能量管理策略开发，选取发电机转速 n_g ，动力电池组荷电状态 SOC 和第二章混合动力履带车建模中求得的当前需求功率 P_{demand} 作为状态量输入至深度神经网络：

$$\text{state} = \left\{ n_g, SOC, P_{\text{demand}} \right\}^T \quad (3.1)$$

考虑到以上三个状态量变化范围不同会对深度神经网络训练速度起到较大影响，式中三个状态量都被进行归一化处理以期加快网络收敛速度。鉴于发电机转速在大多数时间大于或等于零、混合动力车辆会出现刹车情况瞬间功率需求可能低于零，将 n_g 和 P_{demand} 的变化范围分别规定为[0,1]和[-1,1]，电池荷电状态 SOC 处于[0.3,0.8]之间故不做处理。状态量归一化处理如下：

$$\begin{cases} n'_g = \frac{n_g - n_{g,\min}}{n_{g,\max} - n_{g,\min}} \\ P'_{\text{demand}} = \frac{P_{\text{demand}}}{P_{\text{demand},\max}} \end{cases} \quad (3.2)$$

选取发动机节气门开度作为控制量 a ，不同于传统的基于深度强化学习方法，本文开发的方法结合了归一化优势方程，控制变量和状态量都无需进行离散化处理，经过归一化优势函数处理的深度强化学习策略输出控制量在[0,1]范围内。

鉴于 NAF-DQL+PR 属于一种无监督学习，为反映动作 a 在时间 t 时的动作效果并且帮助算法更新网络，有必要建立奖励函数。考虑到对燃油经济性、电池电荷状态 SOC 稳定性的需求和插电池混合动力履带车无需电量保持策略，奖励函数以如下方式建立：

$$J = - \left[\int_{t_0}^t \alpha \dot{f}_{\text{rate}}(t) dt + \beta [SOC(t) - SOC(t_0)]^2 \right] \quad (3.3)$$

式中 $R(s, a)$ 代表智能体在 s 状态下经由动作 a 所获得的瞬时奖励，其中根据强化学习理论，累计回报和瞬时回报的关系如下：

$$G = \sum_{k=0}^{\infty} \gamma^k R_k(s, a) \quad (3.4)$$

式中 γ 代表折扣函数，意味着当前时刻所获得奖励在时间上的衰减。

由于系统状态量数量为三个，因此深度神经网络输入层节点数量为三，输入层节点获取状态量后将其传输至网络中第一个隐藏层 H_1 ，隐藏层 H_1 将数据通过对数据进行线性变换后输入至 Sigmoid 激活方程进行激活和处理，之后将数据输入至网络第二个隐藏层 H_2 。第一隐藏层 H_1 的计算过程如下所示：

$$\begin{cases} g_j^{(1)}(t) = \sum_{i=1}^n \theta_{ij}^{(1)}(t) x_i(t) + b_j^{(1)}, j = 1, 2, \dots, N_{h1} \\ h_j^{(1)}(t) = \text{Sigmoid}(g_j^{(1)}(t)) = \frac{1}{1 + e^{-g_j^{(1)}(t)}} \end{cases} \quad (3.5)$$

式中 N_{h1} 是隐藏层 H_1 中神经元数量，在本算法中， $N_{h1} = 15$ 。

网络中第二隐藏层 H_2 对接收到的数据进行进一步处理，同样经过线性变换之后输入进 Sigmoid 激活函数进行激活，第二隐藏层 H_2 处理过程如下所示：

$$\begin{cases} g_k^{(2)}(t) = \sum_{j=1}^{N_{h1}} \theta_{kj}^{(2)}(t) h_j^{(2)}(t) + b_k^{(2)}, \quad k = 1, 2, \dots, N_{h2} \\ h_k^{(2)}(t) = Sigmoid(g_k^{(2)}(t)) = \frac{1}{1 + e^{-g_k^{(2)}(t)}} \end{cases} \quad (3.6)$$

经过第二隐藏层 H_2 处理的数据被输送至输出层，数据在输出层经过线性变换后得到三个输出量，分别是状态价值 $V(s)$ 、控制量 μ 和用来构建归一化优势函数的下三角矩阵 L ，以上三个输出量计算过程如下：

$$\begin{aligned} V &= \sum_{m=1}^{N_{h2}} \theta_m^{(3)}(t) h_m^{(3)}(t) + b_m^{(3)}, \quad m = 1, 2, \dots, N_{h2} \\ \mu &= \sum_{m=1}^{N_{h2}} \theta_m^{(3)}(t) h_m^{(3)}(t) + b_m^{(3)}, \quad m = 1, 2, \dots, N_{h2} \\ L_{0,n} &= \sum_{m=1}^{N_{h2}} \theta_{nm}^{(3)}(t) h_m^{(3)}(t) + b_m^{(3)}, \quad m = 1, 2, \dots, N_{L_0} \\ N_{L_0} &= \frac{\mu_{dim}(\mu_{dim} + 1)}{2} \end{aligned} \quad (3.7)$$

其中 N_{L_0} 代表下三角矩阵的行数。

目标网络结构与估值网络一致但没有归一化优势函数，其输出值为该状态量 Q 值。目标网络参数复制于估值网络但延迟几个周期。

输出层输出了三个用于构建归一化优势函数的输出量 $V(s)$ 、 μ 及下三角矩阵 L 。基于归一化优势函数思想，即为确保深度神经网络可直接输出控制量且此输出量所得 Q 值是所有可能输出量中最大的，将 Q 值进行重新表达， Q 值不再只是动作回报，而是被划分为两部分—当前状态的价值函数 $V(s)$ 及当前神经网络输出动作优势函数 $A(s, a | \theta^A)$ ，新表征的 Q 值计算方式如下：

$$Q(s, a | \theta^Q) = V(s | \theta^V) + A(s, a | \theta^A) \quad (3.8)$$

式中， θ^Q 、 θ^V 、 θ^A 分别是关于 Q 值函数的参数、状态价值函数参数和动作优势函数参数。在此时中，使用 $A(s, a | \theta^A)$ 而非传统上的 Q 值衡量动作 a_t 在状态 s_t 处的优势。在每一个确定的状态 s_t 下， $V(s_t)$ 的价值是确定的，因此在一确定状态下获得最大 Q 值与在此状态下获得具有最大动作优势值的动作是等价的。于是深度强化学习中找到具有最大 Q 值的动作 a 的问题被转化为在此状态下获取拥有最大动作优势值 $A(s, a | \theta^A)$ 的动作 a 。归一化优势函数的构建如下所示：

$$A(s, a | \theta^A) = -\frac{1}{2} (a - \mu(s | \theta^\mu))^T P(s | \theta^P) (a - \mu(s | \theta^\mu)) \quad (3.9)$$

式中, a 代表智能体输出至环境的控制量, $\mu(s|\theta^{\mu})$ 代表在状态 s 下智能体所能输出的最优控制量。 $P(s|\theta^P)$ 是一对称正定矩阵, 其由下三角矩阵正定矩阵 L 构造得来。由于 $P(s|\theta^P)$ 是一对称正定矩阵, 因此动作优势函数 $A(s, a|\theta^A)$ 的值恒小于等于零, 为取得 $A(s, a|\theta^A)$ 最大值 0, 当且仅当智能体输出控制量 $a = \mu(s|\theta^{\mu})$ 。对称正定矩阵构造方法如下:

$$P(s|\theta^P) = L(s|\theta^P)L(s|\theta^P)^T \quad (3.10)$$

式中 $L(s|\theta^P)$ 由前文介绍神经网络输出, 且其是 $P(s|\theta^P)$ 的唯一 Cholesky 分解。

3.2.2 经验权值优先回放模型构建

由于混合动力车辆控制过程符合马尔可夫理论, 也即车辆运行状态 $(n_g, SOC, P_{\text{demand}})$ 只与上一时刻状态有关而与之前的任何历史状态无关, 因此在应用深度强化学习理论并构建经验权值优先回放模型时也基于此。每一次训练都可被视为从状态 S_{t-1} 通过动作 a_t 到达状态 S_t 并且获得一个及时奖励 R_t 。因此将每一次训练的转移(transition)视作历史经验储存在经验池中(replay-buffer), 其形式如下:

$$\text{transition} = (S_{t-1}, a_t, S_t, R_t) \quad (3.11)$$

在一般的深度强化学习和深度学习框架中, 为打破数据集中的相关性, 每隔一段时间被储存于经验池中的历史训练数据被随机取出一个进行再次训练。训练数据中较稀有的历史训练经验可被重新训练从而使智能体学习。经验回放模型可以减少神经网络训练时间并释放出更多的计算机算力和内存。然而随机抽取并不考虑训练历史数据之间的差异性, 也即对于训练效果较好和较差的数据一视同仁随机取出, 导致智能体难以从经验回放中有效习得知识, 从而训练获得的策略所获奖励较少, 陷入一种被称为“Blind Cliffwalk”的困境, 在此情况下, 智能体需要进行更多经验回放训练、花费更多时间才能摆脱这一困境, 得到更多奖励。

估值网络和目标网络在 Q 值计算上得到的差距 TD-error 被用于评价历史经验数据的“惊喜度”。TD-error 表示着该历史经验数据在估值网络中训练所得 Q 值与目标网络的偏差, 偏差越大则该历史经验数据越具有重新被训练的价值。TD-error 的计算如下:

$$\delta_j = R_j + \gamma_j \hat{Q}(S_j, A_j) - Q(S_j, A_j), \quad j = 1, \dots, N_{mb} \quad (3.12)$$

式中, $\hat{Q}(S_j, A_j)$ 是目标网络计算所得最大 Q 值, $Q(S_j, A_j)$ 是估值网络计算所得最大 Q

值，前文已介绍计算方法。 N_{mb} 是每次取出小片段（minibatch）大小，在本文中 $N_{mb} = 128$ 。

为使具有更大 TD-error 的历史训练数据被更容易取出进行再次训练，每一储存于经验池中的历史训练数据都被赋予一优先级，优先级的计算如下所示：

$$P(j) = \frac{p_j^\alpha}{\sum_k p_k^\alpha} \quad (3.13)$$

其中 $P(j)$ 是第 j 个历史经验数据的相对优先级， p_j 是该数据优先级， $p_j = \delta_j$ ，指数 α 意味着优先级在随机抽取中的重要性。

为方便对于历史经验数据的带优先级抽取，使用 Sumtree 结构将历史经验数据以列表形式储存。每一历史经验数据及其优先级 $P(j)$ 被储存于最底层叶节点上，每一树枝都只有两个分支，每一节点的数值都是其两分枝数值之和。当随机抽取开始时，叶节点被分为 n_{reg} 个部分便于进行抽取，其计算方法为：

$$n_{reg} = \sum_k p_k / n_{batchsize} \quad (3.14)$$

其中 $n_{batchsize}$ 是经验池大小，在 Sumtree 结构最底端的根节点数值被划分为 n_{reg} 个数值区域，在每一数值区域进行抽取。对于第 n_i ($i = 1, \dots, n_{reg}$) 个数值区域，搜索从根节点开始并遵循如下规律：如果左侧叶节点数值大于 n_i ，则搜索向左侧叶节点进行，将 n_i 移至左侧叶节点，否则用 n_i 减去左侧节点数值并将搜索方向转向右侧叶节点，并将该处计算所得移至右侧叶节点直到到达最上层叶节点。经验权值回顾模型如图 3.1 所示。Sumtree 图中根节点位于最上方，叶节点位于最下方。

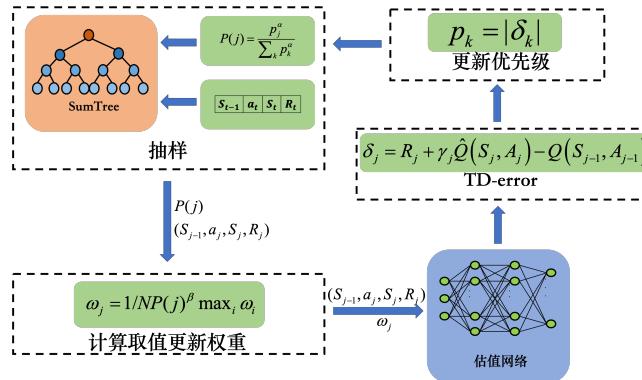


图 3.1 经验权值优先回顾

3.2.3 基于经验权值优先回放的深度神经网络更新方法构建

经验权值优先回放模型将具有更大训练偏差、训练效果较差的历史经验数据从经

验池中提取进行深度神经网络重复训练，为使得神经网络更有效习得这些历史经验数据的“知识”，对传统上的神经网络更新方法进行改良，使得神经网络在学习具有更大学习意义、TD-error 更大的历史经验数据之后进行更大胆的更新。首先损失函数被定义为带有权重的估值网络和目标网络 Q 值之差的平方，其数学表达如下：

$$Loss = \omega_j \left(\left[r_t + \gamma \max_{a_{t+1}} \hat{Q}(s_{t+1}, a_{t+1}; w') \right] - Q(s_t, a_t; w) \right)^2, \quad t = 1, \dots, T \quad (3.15)$$

式中， ω_t 是取值更新权重，其左右在于神经网络进行更新时，针对不同历史经验数据的 TD-error 进行带权值更新， ω_t 的计算方法如下：

$$\omega_j = \frac{1}{NP(j)^\beta \max_i \omega_i} \quad (3.16)$$

式中， β 的值本文取 1。

由于目标网络参数值从估值网络复制而来，于是网络更新计算仅针对于估值网络。估值网络更新基于梯度下降法。神经网络参数在损失函数减少的方向进行更新，估值神经网络更新方法如下：

$$\begin{cases} w_{mn}^{(k)}(t+1) = w_{mn}^{(k)}(t) + l_r(t) \Delta w_{mn}^{(k)}(t) \\ \Delta w_{mn}^{(k)}(t) = -\frac{\partial L(w)}{\partial w_{mn}^{(k)}(t)} \end{cases} \quad (3.17)$$

式中， $l_r(t)$ 为估值神经网络学习率，其取值范围为(0,1)。 $-\partial L(w)/\partial w_{ij}^{(k)}(t)$ 是损失函数对神经网络参数矩阵的偏导数的负值，其计算如下：

$$\begin{aligned} \Delta w_{mn}^{(k)}(t) &= -\frac{\partial L(w)}{\partial w_{mn}^{(k)}(t)} \\ &= E \left[\left(r_j + \gamma \max_{a_{j+1}} \hat{Q}(s_{j+1}, a_{j+1}; w') - Q(s_j, a_j; w) \right) \right] \frac{\partial Q_{s;a;w}(t)}{\partial w_{mn}^{(k)}(t)} \\ &= E_0 \frac{\partial Q_{s;a;w}(t)}{\partial w_{mn}^{(k)}(t)} \end{aligned} \quad (3.18)$$

本文所研究的基于 NAF-DQL+PR 的混合动力车辆能量管理策略实现流程如表 3.1 所示：

表 3.1 NAF-DQL+PR 算法计算流程

Algorithm 1 NAF-DQL+PR

- 1: **输入:**串联混合动力履带车需求功率 P_{demand} ，动力电池组荷电状态 SOC ，发电机转速 n_g
-

2: 设置经验池大小 h , 每次取出经验片段数量 N_{mb} , 经验权值优先回顾周期
 T_r , 参数 α 、 β , 神经网络最大训练次数 M_{max} , 随机初始化双神经网络参
数及学习率 l_r , 目标网络参数更新周期 T_c 。

3: **for** $episode = 1$ to M_{max} **do**

4: 初始化动作随机探索参数 N

5: 从训练数据集中选择训练数据

6: **for** $t = 1$ to T **do**

7: 从混合动力车辆模型观测状态量 (n_g, SOC, P_{demand})

8: 将状态量储存于经验池 SumTree 结构

9: **if** $t \equiv 0 \bmod T_r$ **then**

10: **for** $j = 1$ to h **do**

11: 基于 $j \sim P(j) = p_j^\alpha / \sum p_i^\alpha$ 从 Sumtree 随机抽取经验片段

12: 计算该经验片段取值更新权重 $\omega_j = (N \cdot P(j))^{-\beta} / \max_i \omega_i$

13: 将该数据片段输入深度神经网络进行计算

14: 计算该片段 TD-error $\delta_j = R_j + \gamma_j \cdot \hat{Q}(S_j, A_j) - Q(S_j, A_j)$

15: 使用 TD-error 更新该数据优先级 $p(j) \leftarrow |\delta_j|$

16: 神经网络参数更新值 $\Delta \leftarrow \Delta + \omega_j \cdot \delta_j \cdot \nabla_\theta Q(S_{j-1}, A_{j-1})$

17: **end for**

18: 更新神经网络参数 $\theta \leftarrow \theta + l_r \cdot \Delta$ 并将 Δ 重置为 0

19: **end if**

20: **if** $t \equiv 0 \bmod T_c$ **then**

21: 将估值网络参数复制至目标网络 $Q_{target} \leftarrow \theta_{eval}$

22: **end if**

23: 利用归一化优势函数计算控制量 A_t

24: 施加探索于控制量 $A_t \leftarrow A_t + N$

25: A_t 输出至混合动力汽车模型观测瞬间所得奖励 R_t

26: 观测下一时刻状态 $state_ = (n_{g_}, SOC_ , P_{demand_})$

27: **end for**

3.3 算法仿真及结果分析

在本章节中，使用仿真实验方法对基于 NAF-DQL+PR 的串联混合动力能量管理方法进行评估。为验证本章所提出算法，采用第二章建模的 1.2 吨串联混合动力无人履带车作为仿真对象对 NAF-DQL+PR 进行离线仿真及测试，并将训练结果与基于 NAF-DQL、基于 DQL 的能量管理策略进行对比验证其改进，并与全局最优 DP 进行对比验证本算法的次优性。

3.3.1 离线仿真工况

为更好验证基于 NAF-DQL+PR 的串联混合动力能量管理策略的高效性、节油性，使用 US06 工况对算法进行离线验证，该工况速度及需求功率如图 3.2 所示。图中，US06 工况持续 600 秒，全程长 12.89 公里，其平均车速为 48.4km/h，最大车速为 80.3km/h 满足本款串联混合动力履带车运行需求。工况中速度为需求功率为负值部分为混合动力车辆制动导致。

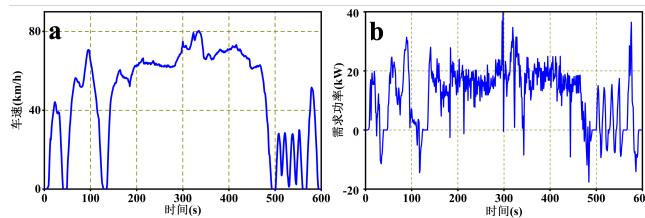


图 3.2 离线仿真工况 (a) 车速 (b) 需求功率

3.3.2 深度强化学习网络参数设定及训练环境

基于深度强化学习的混合动力车辆能量管理策略需提前对网络及各模型参数进行初始化设置，初始参数的选取对深度强化学习的训练影响较大，良好的参数选取可使得神经网络训练收敛快速且收敛至良好训练效果，关于深度强化学习参数的优化在第五章进行讨论。基于 NAF-DQL+PR 的串联混合动力能量管理方法参数选取如表 3.2 所示：

表 3.2 NAF-DQL+PR 神经网络参数表

网络参数	值
折扣因子	0.99
学习率	0.01

网络参数	值
每次取出经验片段数量	128
每周期最大训练次数	200
目标网络更新延迟	5
最长训练周期	100

该深度强化学习基于 Keras 及 Tensorflow 人工神经网络训练库被实现于 python 环境中。训练使用本课题组搭载双路 NVIDIA GTX 1080 GPU 针对于深度神经网络训练的工作站。

3.3.3 深度强化学习训练结果及对比

基于深度强化学习的能量管理策略训练效果如图 3.3 所示。图 (a) 为智能体在环境中所得奖励，训练效果初期不理想，从初始值 -900 向下降，此时智能体正在对环境进行探索，由于网络初始参数为随机和人为给定，其效果较差，算法尚不能输出较好结果。经过 5 个周期的训练，智能体逐渐从环境中学习得到“知识”，网络参数的优化使得智能体从环境中得到的奖励逐渐增加并在第十个周期后稳定下来并逐渐收敛。图 (b) 为深度神经网络训练损失曲线图，如图所示，训练伊始网络损失值较大，此时目标网络和估值网络之间偏差较大，也即估值网络未收敛需多进行学习，之后随着训练步数增多，网络训练损失逐渐减小接近于零，此时目标网络与估值网络偏差较小，意味着深度神经网络训练已收敛。横向对比图 (a) 可见此时深度强化学习模型所得奖励已稳定，代表网络训练达到最优值附近。

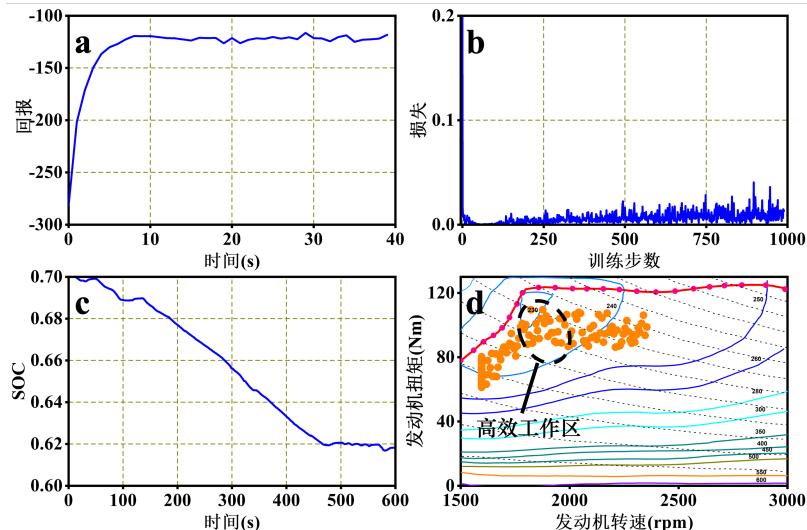


图 3.3 基于 NAF-DQL+PR 的混合动力车辆能量管理策略训练结果

NAF-DQL+PR 策略产生的电池荷电状态 SOC 曲线图如图 (c) 所示, 由于未设置电池荷电状态维持策略, 总体上动力电池组在进行放电, 中间有少数充电状况出现, 动力电池组 SOC 从 0.7 降到 0.62 附近且并无明显充放电情况出现, 整体动力电池运行过程平稳, 表示基于 NAF-DQL+PR 的混合动力车辆能量管理策略有助于维护动力电池寿命。NAF-DQL+PR 策略产生的发动机工作点如图 (d) 所示, 发动机高效工作区已在图中标出。图中粉色曲线为本款发动机外特性曲线, 彩色实现为等燃油消耗线, 虚线为等功率曲线。由图 (d) 可见, 大部分工作点位于中等燃油经济区, 部分工作点进入高燃油经济区, 没有工作点位于低燃油经济区, 因此基于 NAF-DQL+PR 的能量管理策略有助于发动机维持工作点在中高燃油经济区, 在输出功率的同时减少燃油消耗。深度强化学习训练效果表明, 基于 NAF-DQL+PR 的能量管理策略能快速收敛至最优值, 其输出电池 SOC 无剧烈充放电, 有利于保护动力电池并延长其使用周期。该策略有助于发动机工作在高效经济区, 有助于减少燃油消耗。

为验证基于 NAF-DQL+PR 的能量管理策略的优越性, 对基于 NAF-DQL 的能量管理策略、基于 DQL 的能量管理策略和基于动态规划的能量管理策略分别进行训练, 通过横向对比进行验证。如图 3.4 所示, 图 (a)、图 (b) 分别展示了应用四种方法后发动机工作点及电池荷电状态 SOC 变化曲线。如图 (a) 所示, 基于 NAF-DQL+PR、NAF-DQL 及 DQL 的混合动力能量管理策略都采用相同初始网络参数, 这三种能量管理策略的分布从上到下分别是 NAF-DQL+PR、NAF-DQL 及 DQL, 与其余两种能量管理策略相比, 基于 NAF-DQL+PR 的能量管理策略使得更多的发动机工作点集中于中高效工作区, 其输出发动机工作点呈分层和分散两种状态, 在低转速区集中的工作点表示此时对动力电池使用较多, 进入高效工作区呈分散状态的工作点表示此时策略学习到发动机高效工作区“知识”从而减少动力电池使用, 更多利用发动机进行功率输出。基于 NAF-DQL 的能量管理策略工作点分布呈分层式, 部分工作点集中在低转速区间, 部分分散于中转速区, 这是由于基于 NAF-DQL 的能量管理策略倾向于使用动力电池进行供电。基于 DQL 的能量管理策略工作点分布于低效经济区, 由于没有归一化优势函数的高精度控制和经验权值优先回顾的历史经验挖掘, 其难以学习到最优控制策略。相比于以上三种能量管理策略, 基于 DP 的能量管理策略分布点较为靠上且更多处于高效工作区, 因 DP 代表全局最优策略, 其工作点表现也是此工况下发动机最好表现。

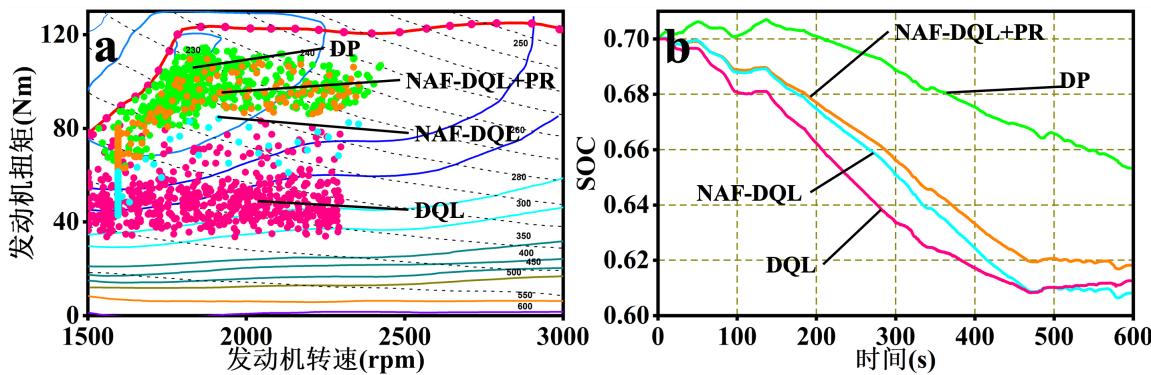


图 3.4 四种能量管理策略训练结果

从图（b）可看出，由于未规定电池电荷保持策略，四种能量管理策略都表现出 SOC 总体下降趋势，动力电池组工作四种能量管理策略所得电池 SOC 曲线中，基于 DP 的能量管理策略 SOC 最为平缓，因 DP 具有全局最优性。相比于其余两种基于学习的能量管理策略，基于 NAF-DQL+PR 的能量管理策略输出 SOC 最为平缓且无明显充放电现象。NAF-DQL 表现与 NAF-DQL+PR 类似，放电趋势类似，因都适用归一化优势函数，所以输出控制精度较高，但其 SOC 曲线相对斜率更高。而相比于其他三种能量管理策略，基于 DQL 的能量管理策略输出 SOC 曲线较陡且有明显剧烈充放电现象，因 DQL 没有学习到最佳控制策略。因此基于 NAF-DQL+PR 的能量管理策略能输出逊于 DP 但优于其余两种能量管理策略的 SOC 曲线，从而验证了基于 NAF-DQL+PR 的能量管理策略可以有效保护电池、延长电池工作寿命。

四种能量管理策略的燃油消耗量如表 3.3 所示。相比于其余三种能量管理策略，基于 NAF-DQL+PR 的能量管理策略表现仅略逊于基于 DP 的能量管理策略而优于其余基于学习的能量管理策略，从表中可见，NAF-DQL+PR 消耗燃油 397.64 克，相比于 DP 燃油消耗的 389.08 克仅多消耗 2.15%，基于 DP 的全局最优性，其燃油经济性被定义为 100%，相对于 DP, NAF-DQL+PR、NF-DQL 及 DQL 分别具有 DP 的 97.85%、95.24% 及 92.59% 的燃油经济性，而相对于 NAF-DQL 及 DQL，基于 NAF-DQL+PR 的能量管理策略分别提升了 2.61% 及 5.26%。

表 3.3 四种能量管理策略燃油消耗及燃油经济性

策略	燃油消耗(g)	燃油经济性(%)
NAF-DQL +PR	397.64	97.85
DP	389.08	100

策略	燃油消耗(g)	燃油经济性(%)
NAF-DQL	408.54	95.24
DQL	420.20	92.59

此外,为推进对混合动力能量管理策略实时性的研究,充分发挥基于深度强化学习的能量管理策略自学习的优势,本文统计了四种能量管理策略运算至收敛的时间,如表 3.4 所示。基于 NAF-DQL+PR 的能量管理策略训练所需收敛时间为 77 秒, DP 所需训练时间为 1756 秒, NAF-DQL 训练至收敛所需时间为 982 秒, DQL 训练至收敛所需时间为 1019 秒, 相对于 DP、NAF-DQL 及 DQL, 基于 NAF-DQL+PR 的训练收敛时间大幅缩短,由于引入经验权值优先回顾结构,深度神经网络对训练工况进行了更有效的学习,习得具有更高回报的策略速度更快,网络收敛更快。因此相比于其他算法,基于 NAF-DQL+PR 的能量管理策略具有收敛更快的优势。

表 3.4 四种能量管理策略训练至收敛耗时

策略	训练至收敛耗时(s)
NAF-DQL + PR	77
DP	1756
NAF-DQL	982
DQL	1019

由以上针对 NAF-DQL+PR、DP、NAF-DQL 及 DQL 在发动机工作点、燃油消耗、电池 SOC 曲线和计算至收敛所需时间四方面的比较可见,基于 NAF-DQL+PR 的能量管理策略在发动机工作点分布、电池 SOC 曲线变化及燃油消耗方面都稍逊于基于全局优化的 DP 但优于其余两种算法,在计算至收敛所需时间方面远优于其余三种方法,从而验证了 NAF-DQL+PR 具有全局次优性的能量管理效果但具有全局最优的训练速度,因此验证了基于 NAF-DQL+PR 的混合动力履带车辆能量管理算法的最优性和有效性。

3.4 本章小结

本章以基于 NAF-DQL 的能量管理算法为基础,提出经验权值优先回顾模型并将其与 NAF-DQL 算法有机结合,提高基于深度强化学习的能量管理算法的寻优能力及计算速度,推进了基于深度强化学习的能量管理策略实时性研究。主要工作总结如下:

- (1) 为提升 NAF-DQL 算法运算效率及最优性,对经验回放模型进行改进,构

建经验权值优先回放模型，促进深度神经网络挖掘历史经验数据，学习更优策略。通过定义历史经验数据相对优先级赋予其权值，通过构建 Sumtree 结构实现对数据带优先级提取，通过构建取值更新权重改进深度神经网络更新方法，最终完成 NAF-DQL+PR 算法的构建并完成对算法的模型训练

(2) 为验证基于 NAF-DQL+PR 的能量管理策略的高效性控制，将其训练结果与动态规划、NAF-DQL 及 DQL 在发动机工作点、电池组 SOC、燃油消耗及训练时间四个维度进行对比，结果表明相对于其他两种基于学习的能量管理策略，NAF-DQL+PR 可实现更好的控制效果，其燃油经济性可达 DP 的 97.85%，相对于 NAF-DQL 及 DQL 分别提升 2.61% 和 5.26%，验证了其燃油经济性。在训练所需时间方面，NAF-DQL+PR 大幅优于其余三种算法，训练收敛所需时间短，具有更快的训练速度，证明了经验权值优先回顾模块的引入对于算法实时性的加强，推进了基于深度强化学习的能量管理策略实时性研究。

第4章 基于深度强化学习的混合动力能量管理策略超参数优化

第三章讨论了基于深度强化学习的串联混合动力车辆能量管理方法，通过引入经验权值优先回顾的方法加速神经网络训练并提升训练效果。基于深度强化学习的能量管理策略训练需提前设置初始化参数，包括学习率、经验池大小、训练周期等多个超参数都需提前进行设置。调试过程表明，不同超参数及超参数组合的设定对深度强化学习算法训练影响较大，其中学习率由于决定了深度神经网络更新学习快慢对深度强化学习训练效果影响突出，不同的学习率选取影响到深度强化学习训练收敛效果及收敛快慢，并进一步影响到基于深度强化学习的混合动力车辆能量管理策略的效果。本章为推进基于 NAF-DQL+PR 的串联混合动力履带车能量管理算法研究，针对 NAF-DQL+PR 算法中学习率参数进行优化，使用遗传算法 (genetic algorithm, GA) 通过进化理论寻找最优 NAF-DQL+PR 学习率。本章介绍了优化结构 GA-DQL，并进行不同学习率的训练效果对比，以调查学习率对网络训练效果的影响并验证 GA-DQL 优化效果，以此推进基于深度强化学习的能量管理算法开发。

4.1 基于遗传算法的深度强化学能量管理算法超参数优化

遗传算法是一种启发型算法，基于达尔文进化论中自然选择和物种变异机理，遗传算法模拟自然进化的过程，将优化目标表征为某一种染色体并创建种群，该种群中每一个体都携带有该染色体但是基因型各异，该种群位于被预先定义的“自然环境”中进行进化。根据适者生存理论，难以适应自然的个体将被淘汰，能够在自然中存活的个体将会进行繁殖和变异从而产生新个体，在遗传算法中能在训练中存活的个体将进行组合交叉及变异。在历代进化之后产生新的种群在“自然选择”中趋向同一，此时个体的基因型全部相同从而完成优化。

根据遗传算法理论，针对基于深度强化学习的串联混合动力履带车的超参数问题进行优化求解。通过建立优化目标基因、种群及构造环境，建立基于遗传算法的 NAF-DQL+PR 模型 (GA-DQL)，对深度强化学习算法学习率进行优化。

首先随机初始化种群及基因，基因编码使用二进制编码规则便于进行组合交叉及变异。其次建立每个个体基因型与其表现型对应关系，也即将二进制编码基因转化为十进制数字并在学习率变化范围内一一对应。基于历史经验，学习率的选取范围选择在 0.01~0.0001 之间，完成定义种群、基因及表现型后计算每一个体的适应度，鉴于

优化针对混合动力能量管理问题，将策略所获总回报定义为个体适应度，也即在训练工况下得到最多奖励、消耗最少燃油的学习率被选择出来进行下一轮进化。基于个体适应度进行挑选，个体适应度挑选基于优先级，具有更高适应度的个体具有更大几率被选择出进入下一轮进化。不同于直接删除低适应度个体，具有低适应度有更低几率被选择进入一下轮进化，为低适应度个体保留变异进化出高适应度个体的机会。所有被选择个体都有一定几率进行组合交叉，随机从种群中挑选两个个体进行随机点位遗传因子交换生成子代，子代的产生伴随随机出现的基因变异，也即子代个体中随机位置出现遗传因子变化，在本优化中是 $0 \rightarrow 1$ 或 $1 \rightarrow 0$ ，产生的新子代将进入下一轮进化，直到种群中所有个体基因型一致，进化过程停止。GA-DQL 优化算法结构如图 4.1 所示。GA-DQL 优化算法实现流程如表 4.1 所示。

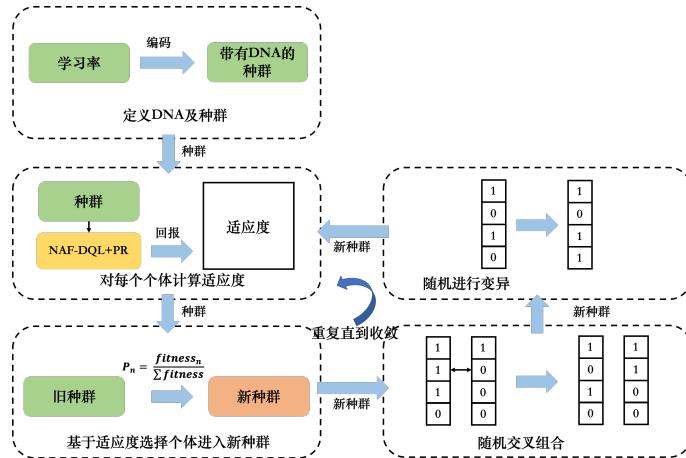


图 4.1 遗传算法计算流程

表 4.1 基于 GA -DQL 的深度强化学能量管理策略实现过程

Algorithm 2 GA -DQL

- 1: **Input:** 混合动力车辆需求功率 P_{demand} , SOC , 发电机转速 n_g
- 2: 设置种群及 DNA 大小, 设置最多进化次数 N_g , 设置最大训练周期 M_{\max} , 随机初始化网络参数
- 3: 随机初始化种群及 DNA
- 4: **for** $generation = 1$ to N_g **do**
- 5: 将 DNA 转化为对应的学习率输入进 NAF-DQL+PR
- 6: **for** $episode = 1$ to M_{\max} **do**
- 7: 初始化动作探索过程 N

```

8:      训责训练工况
9:      for  $t = 1$  to  $T$  do
10:     接收初始训练观测状态量( $n_g, SOC, P_{\text{demand}}$ )
11:     将状态量输入进估值网络
12:     计算  $L, V, \mu$ 
13:     计算归一化优势函数  $A = -\frac{1}{2}(a - \mu)^T LL^T(a - \mu)$ 
14:     计算动作  $a$ 
15:     对动作施加随机探索过程  $a \leftarrow a + N$ 
16:     计算  $Q$  值  $Q = V + A$ 
17:     实施动作  $a$  并观测回报  $R_t$ 
18:     得到下一时刻状态  $state\_ = (n_{g\_}, SOC\_, P_{\text{demand}\_})$ 
19:     if  $t \equiv 0 \bmod 5$  then
20:       基于经验权值优先回顾抽取历史训练数据
          ( $n_g, SOC, P_{\text{demand}}$ )
21:       计算损失函数、TD-error 并带权值更新神经网络参数
22:       将估值网络参数克隆至目标网络
          end if
23:     end for
24:   end for
25:   得到每个 DNA 的适应度  $fitness = reward$ 
26:   得到表现最佳 DNA
27:   基于适应度随机抽取个体进入新种群  $P_n = \frac{fitness_n}{\sum fitness}$ 
28:   基于随机交叉组合和变异生成新种群
29: end for
30: return 最佳 DNA
31: 将最佳 DNA 转化为学习率

```

4. 2 超参数优化算法验证

为验证基于 GA-DQL 的混合动力车辆能量管理超参数优化方法的有效性和最优性，本章节中，将 GA-DQL 算法优化所得学习率在基于 NAF-DQL+PR 的串联混合动力车辆能量管理策略中进行离线仿真测试，使用第二章介绍 1.2 吨串联混合动力无人履带车作为仿真对象，并与其他学习率进行横向比较，验证 GA-DQL 算法在基于深度强化学习的能量管理策略中的优化效果。

4. 2. 1 离线仿真工况

与第三章验证 NAF-DQL+PR 算法选取相同工况，进行验证，该工况特性已在第四章进行介绍，因其符合本款串联混合动力履带车使用工况需求，故在本章节继续使用 US06 进行算法验证。

4. 2. 2 超参数算法验证参数设定及环境

为控制变量保证比较有效性，对照组和实验组除学习率外参数都设置为一样。基于 GA-DQL 的超参数优化算法验证试验参数设定如表所示：

表 4.2 GA-DQL 算法验证试验参数设定

网络参数	值
折扣因子	0.99
每次取出经验片段数量	128
每周期最大训练次数	200
目标网络更新延迟	5
最长训练周期	100

4. 2. 3 GA-DQL 算法验证及对比

基于 GA-DQL 优化所得最优学习率值为 0.00020645，将此数值输入进 NAF-DQL+PR 算法进行训练得到结果如图所示。由于第三章使用结果为已优化好学习率，因此此处结果与第三章结果一致，如图 4.2 所示：

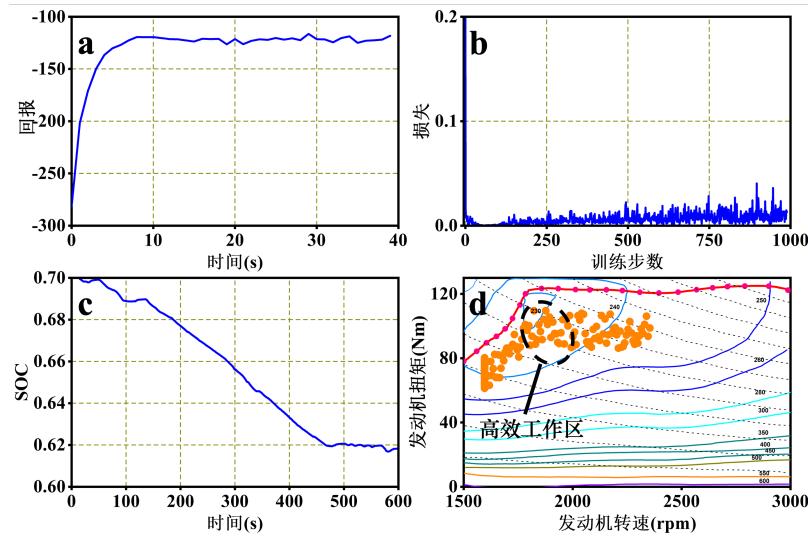


图 4.2 经优化的 NAF-DQL+PR 训练效果

如图可见，5 个训练周期之后，网络训练所得回报趋于稳定，估值网络和目标网络之间损失函数趋近于零，训练收敛快。动力电池组 SOC 下降平稳，无剧烈充放电。大部分发动机工作点位于中高效经济区，由此可见该学习率设置所得结果具有高燃油经济性，能延长动力电池寿命并加速神经网络训练。

为验证 GA-DQL 优化算法所得学习率由于其他学习率，建立其他三组学习率仿真试验。三组学习率选择分别为 0.01, 0.001 和 0.0001。仿真实验结果如图 4.3 所示，从图（a）可见，相对于其他三组学习率训练结果，学习率为 0.00020645 的能量管理策略所产生工作点更多位于中高效工作区，学习率为 0.01 的策略工作点大部分位于低效工作区，其原因是学习率选择较大导致错过最佳策略，学习率 0.001 和 0.0001 同理，因为学习率选择不合适而导致只有少部分工作点进入中等效率区，未能使得发动机工作点进入高效工作区。因此与其他三组学习率相比，优化过的学习率具有更优发动机工作点分布。从图（b）可见，由于都采用 NAF-DQL+PR 算法，四条 SOC 曲线变化趋势相近且无明显剧烈充放电现象。相比于其余三组学习率，学习率为 0.00020645 的 NAF-DQL+PR 算法输出动力电池 SOC 曲线更为平缓。因此验证 GA-DQL 优化算法有助于提升 SOC 变化平稳性。

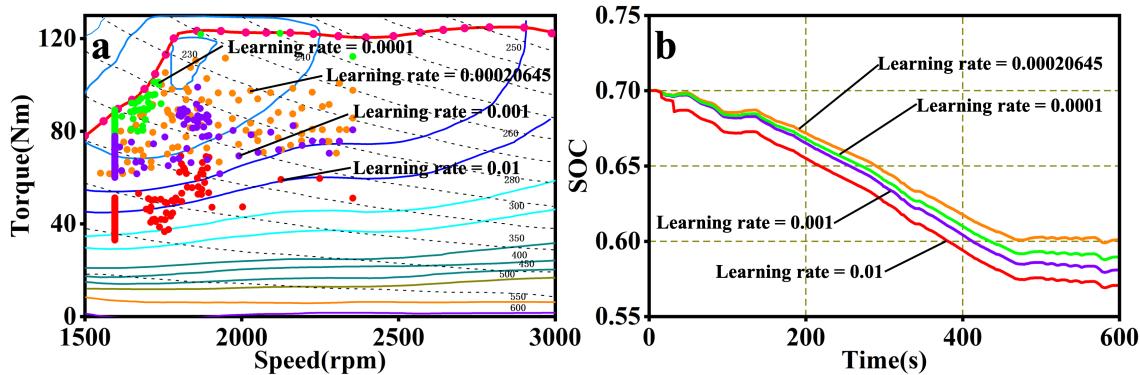


图 4.3 四种不同学习率训练效果

四组学习率训练后所得燃油消耗量如表 4.3 所示，相比于其他三组学习率，经过优化的学习率所产生燃油消耗最低，为 397.64 克，相比于学习率 0.01 的 457.21 克、学习率 0.001 的 426.62 克和学习率 0.0001 的 404.21 克分别提升了 15.1%、7.3% 和 2.1%。由此可验证基于 GA-DQL 的优化方法可有效提升串联混合动力履带车燃油消耗表现。

表 4.3 四种学习率燃油经济性比较

学习率	油耗(g)	燃油经济性(%)
0.00020645	397.64	100
0.0001	404.21	97.9
0.001	426.62	92.7
0.01	457.31	84.9

此外，由于基于 NAF-DQL+PR 的能量管理策略具有训练收敛快的特点，有必要针对不同学习率训练收敛时间进行比较和算法最优化验证。各组学习率训练至收敛耗时如表 4.4 所示。经过比较，经优化的学习率训练至收敛耗时最短，为 77 秒，少于其余三组学习率训练耗时，相对于学习率为 0.01、0.001 及 0.0001 的训练时间分别提升 55.8%、23.4% 和 6.49%。由此可见，适当的学习率选取能对经验权值回顾模型效果起到增益作用，实现“1+1>2”的效果。

表 4.4 四种学习率训练高效性比较

学习率	训练耗时(s)	训练高效性 (%)
0.00020645	77	100

学习率	训练耗时(s)	训练高效性 (%)
0.0001	82	93.51
0.001	95	76.6
0.01	120	44.2

从本章节针对不同学习率设定进行仿真试验对各组学习率进行发动机工作点、SOC 曲线、燃油消耗及训练时间进行横向对比的结果可见，基于 GA-DQL 的深度强化学习能量管理策略超参数优化方法可实现针对学习率的优化，其优化所得学习率可在以上四个方面达到最优效果，由此验证了该方法实施效果。

4. 3 本章小结

本章针对深度强化学习能量管理策略中学习率的选取进行研究。提出了一种基于遗传算法的深度强化学习学习率优化方法，该方法优化所得学习率被输入进 NAF-DQL+PR 在离线仿真中与学习率未经优化的 NAF-DQL+PR 进行发动机工作点、电池 SOC、油耗和训练收敛速度四个维度的比较。结果表明经优化的 NAF-DQL+PR 可输出最佳燃油经济性且训练耗时最短，因此验证了 GA-DQL 优化算法的有效性。

第5章 基于深度强化学习的混合动力能量管理策略硬件在环仿真验证研究

为验证本文第三章所提出基于 NAF-DQL+PR 能量管理策略、第四章所提出基于 GA-DQL 的深度强化学习能量管理策略超参数优化方法，本章设计混合动力源实时调试软件程序，使用 dSPACE 实时仿真平台及 MotoTron 控制器建立串联混合动力履带车硬件在环仿真平台，对经过优化的基于 NAF-DQL+PR 混合动力车辆能量管理策略进行硬件在环仿真验证，验证本文研究的混合动力能量管理算法在实时控制环境下对串联混合动力履带车节能性的效果。

5.1 混合动力源调试程序开发

传统上进行混合动力源调试依靠 dSPACE 等相关软硬件，通过采集卡收集高速 CAN 总线数据进行几个参数的读取，或是使用电流计、电压表、功率分析仪和示波器等设备进行实时测量和读取，对于混合动力源样机调试和制造中出现少至几个，多至十几个参数的情况，很难做到全局兼顾、面面俱到。由于混合动力源调试过程中需要使用发动机、发电机、动力电池、高压功率模块和电池柜等组件，其中高温、高压模块的存在使得人工采集和测量十分危险。混合动力源设计试制常常面临新型结构、新算法等尝试，在混合动力源的调试中需时刻监视各组件包括电压、电流、温度和故障等信息，因此传统上所使用的测量数据方法危险且不适用。采用机器代替人工读取各组件信息、调试人员只需关注设备反馈信息的方法无疑有助于提升混合动力源调试效率并保护工程师人身安全。面对调试中出现的不理想结果，传统上工程师会将混合动力源做停机处理之后进行数据的输入和刷写，在每次调试时都需重新检查各组件确保其安全，这种方法往往需要耗费一定时间并且干扰调试工作节奏和减缓调试进度，往往一天的工作时候只能针对一个参数进行改进优化。因此针对于混合动力源实际调试中出现的问题、隐患和需求，研发一个具有实时监控、故障显示并且支持多个参数修改功能的混合动力源实时调试平台亟待开发。本章为解决在混合动力源能量管理策略调试中出现的需求及隐患，介绍一混合动力源实时调试程序。本文开发的混合动力源实时调试程序基于 LabVIEW 设计，通信基于高速 CAN 通信协议。高速的通信通路和及时的软件响应是实时调试软件的基础。

5.1.1 系统总体设计

本文开发的混合动力源实时开发需求和目标明确，也即，开发一款基于 LabVIEW 和 CAN 协议的能实现混合动力源各组件状态实时监控、多个参数实时修改功能的轻型软件且无需安装过多应用程序。

针对监控需求，需进行界面开发及实时 CAN 协议解析并展示在面板上。针对实时调试功能，需能从面板及时读取被修改参数信息，通过高速 CAN 协议发送至对应地址并及时返回该参数状态确认修改成功。后端程序中，需针对每秒从 CAN 总线上传来的 1000~5000 帧信息进行解析。在后端程序中，存在两种读取 CAN 协议数据帧方式，一种为基于用户事件触发生成事件，对事件进行操作并最终结束事件的用户事件方法，该方法应用于解析 CAN 数据时，会对每一传进电脑的 CAN 帧建立用户事件并最终销毁用户事件，该方法应用于数据密度较低情况，较为有效，面对庞大的数据量难以在短时间内处理相当数量的数据，造成 CAN 数据在系统缓存中大量堆积，前面板难以实时显示相应数据，因而基于用户事件开发的方法不能应用于本软件开发。另一种方法为定时触发读取，也即每隔一定时间从 CAN 总线信息流中提取一帧数据进行解析，将解析所得数据反映在前面板后等待下一个定时触发。考虑人眼视觉极限频率为 60hz，也即当前面板数据刷新在 10ms 以下时，人眼不能看到前面板中所有刷新出的数字，考虑对于计算机算力，解析一帧 CAN 数据不需 10ms，定时触发模型被应用于本软件开发。在总体设计框架如图 5.1 所示。

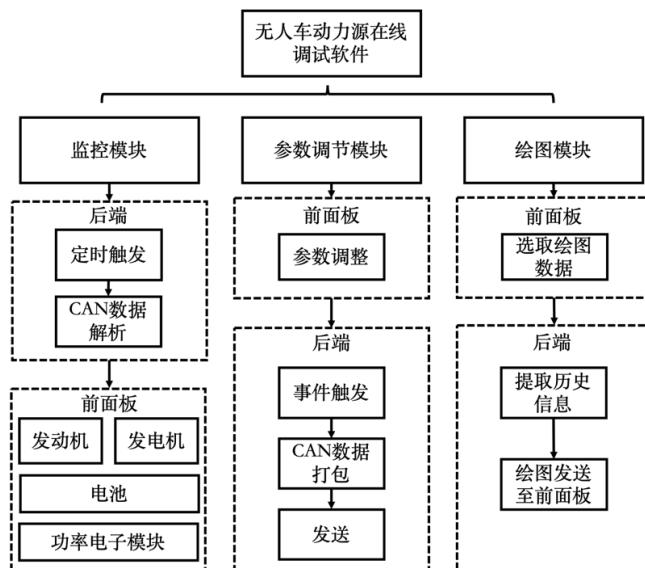


图 5.1 混合动力源实时调试程序结构

5.1.2 软件设计

如图 5.2 所示，考虑到调试过程中需同时兼顾混合动力源各组件、监控混合动力源实时调试程序及进行参数修改，图中将监控界面和参数调节界面集成于同一选项卡中，有利于调试工作顺利进行，其中监测模块被分为 4 部分，分别监测发动机、发电机、高压功率模块、动力电池组及驱动电机的运行状态。其中针对发动机，监测其冷却水温确保发动机冷却回路运转正常。针对发电机，通过给定目标转速监测其实际转速跟随情况、工作模式及故障码。针对高压功率模块监测其变压器磁钢温度、变压器线圈温度及整流管温度确保局部过热不会发生。针对动力电池组，由于动力电池组对温度敏感且易出现故障，该软件设计监测其整体输出电压、电流、运行状态、充电状态和电池组中电池差异性。为更直观展示各组件运行状态及反映相关 CAN 数据帧及时发送到该软件，在前面板上针对各数据帧接收和故障码报错设置指示灯便于调试工程师直观了解到混合动力源运行状态。

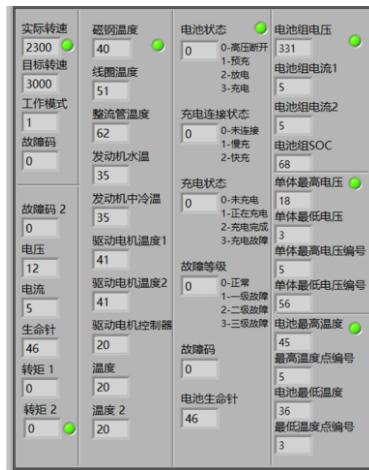


图 5.2 混合动力源实时调试软件监控界面

后端程序运行如图 5.3 所示。图中 VCI_CAN_OBJ_N_2 从底层接收 CAN 报文，每过十毫秒，后端程序对 VCI_CAN_OBJ_N_2 进行一次读取，解析出其中 8 位数据帧，根据对应地址分别发送至对应的显示框并将训练状态显示在监控界面指示灯中。

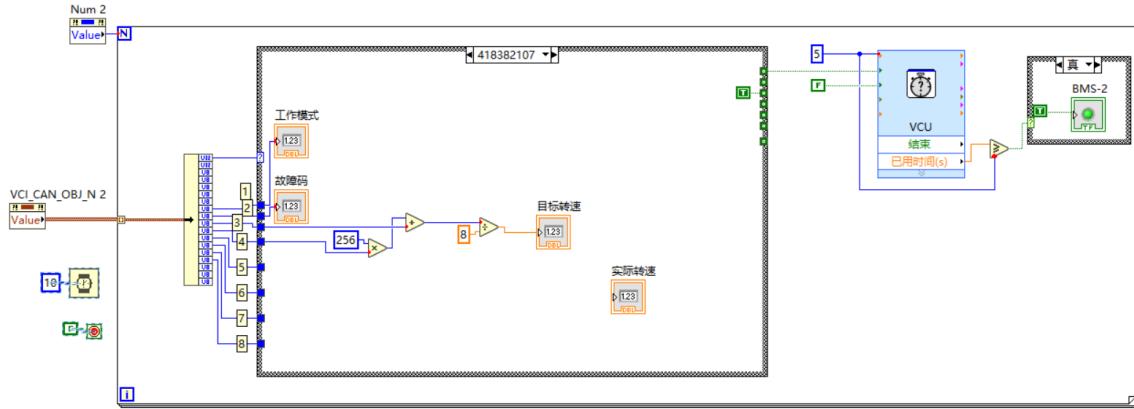


图 5.3 混合动力源实时调试软件监控后端程序图

由图 5.4 可见，在软件仿真部分可输入不同坡度、车速等参数调整动力源运行环境，有利于进行各种情况下的完整测试。在参数运行状态监测栏中，可实时读取混合动力源 100 个可调参数值，根据各参数原始值可进行在线修改。通过在数据框中输入相应参数序号和修改值可直接进行在线实时参数调整，有利于加速混合动力源调试进度。

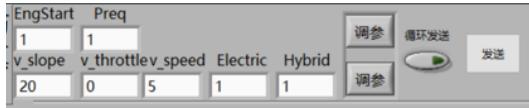


图 5.4 混合动力源实时调试软件环境设置界面

在参数运行状态监测栏中，可实时读取混合动力源 100 个可调参数值，根据各参数原始值可进行在线修改。通过在数据框中输入相应参数序号和修改值可直接进行在线实时参数调整，有利于加速混合动力源调试进度。如图 5.5 所示。鉴于参数修改功能的使用频率不及监控功能，使用用户事件触发对参数在线修改功能进行实现。参数调节功能实现如图 5.6 所示。每次调试工程师输入需修改参数序号及修改值时触发一个用户事件，该事件对准备修改的参数序号和参数值进行处理，将其变为 CAN 地址及相应的十六进制字符串，并通过 CAN 发送模块发送至远端特定地址处。待发送完成，该用户事件被销毁，该模块进入静默等待直到下一次参数修改。同时，参数在线监控模块实时刷新 100 个参数值，确认该参数值已被成功写入。如图 5.7 所示。该模块可实现 100 个参数实时在线修改，并及时反馈修改状况确保通讯正常，有利于确保混合动力源调试工作安全高效进行。

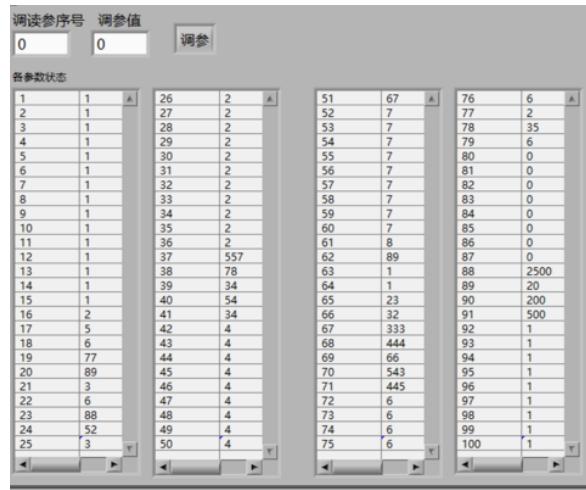


图 5.5 混合动力源调试软件调参界面

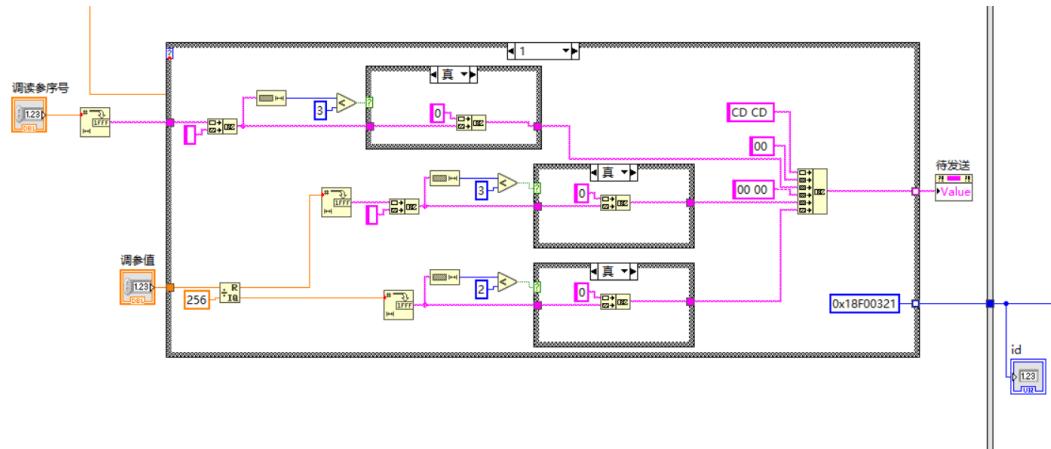


图 5.6 混合动力源调试软件调参程序

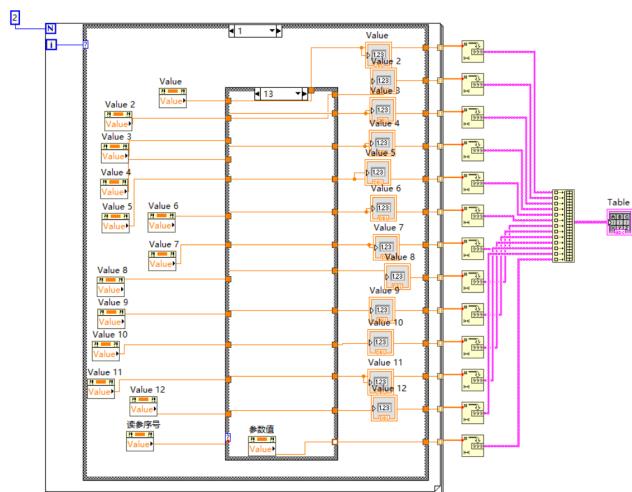


图 5.7 混合动力源调试软件调参显示程序

绘图模块设计目标于辅助混合动力源调试过程中及调试工作完成后需对所监测数据进行趋势分析等后续工作。绘图模块中含有四块可实时显示的画图面板，可实现四个参数在线实时显示功能。通过在前面板输入相应数据参数，可将其实时绘制在画板上。该程序设计如图 5.8 所示。XY 图程序接收数组数据，将历史数据及新增数据组成数组后显示在绘图界面上。

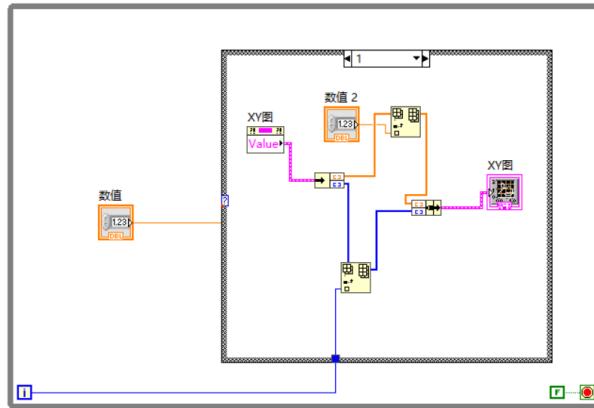


图 5.8 混合动力源调试软件绘图程序

针对于前面板触发绘图事件，该程序使用事件触发，对输入参数序号进行提取，之后对该参数所在地址数据信息进行解析，该解析过程使用定时触发，也即类似上文前面板各组件监控部分实现。通过这种事件触发和定时触发相结合的方法可分别接受前面板输入和进行 CAN 通讯解析。将解析数据接入画图模块即可实现在线绘图功能，该功能有利于调试工程师实时了解参数变化趋势并进行试验后分析。

5.2 串联混合动力履带车硬件在环仿真平台

串联混合动力履带车硬件在环仿真平台由实时仿真环境和整车控制器组成，其中实时仿真环境中集成串联混合动力履带车前向仿真模型，其建模在第二章中已进行介绍。整车控制器中集成基于 NAF-DQL+PR 开发的混合动力车辆能量管理策略，负责向实时仿真平台发送控制量并接收从实时仿真平台反馈的状态量。

5.2.1 实时仿真被控系统搭建

实时仿真被控系统搭建基于 dSPACE 公司产品 dSPACE AutoBox 实时仿真系统。dSPACE 配套开发软件包括 (Real-Time Interface, RTI) 及 ControlDesk，由于 dSPACE 基于 MATLAB/Simulink 开发，在 MATLAB/Simulink 上的代码可通过 RTI 实现 C 代

码生成并进行编译下载到硬件中。ControlDesk 是安装于电脑的上位机监视软件，在前文关于混合动力车辆调试平台测试中使用过，该软件可实现针对采集数据实时显示并绘图功能。由于 dSPACE 具有强大的计算能力和扩展能力，其常被用做混合动力车辆能量管理策略硬件在环仿真平台。本文所使用 dSPACE 实物如图 5.9 所示。



图 5.9 dSPACE AutoBox

在硬件在环仿真试验中，串联混合动力履带车前向模型通过 RTI 软件进行 C 代码生成及编译并下载至 dSPACE。该前向控制模型包含有发动机-发电机组、动力电池组及高压功率模块模型。该模型通过接收整车控制器所发送能量分配指令进行计算，得到相关混合动力车辆状态变量反馈至整车控制器形成闭环。

串联混合动力履带车模型在 dSPACE 中由两部分构成，程序系统配置模块及主程序。系统配置模块中包含有 CAN 通讯模块参数设置，主程序中包含串联混合动力履带车前向模型及 CAN 收发模块，CAN 收发模块可实现与整车控制器通讯。dSPACE 上位机软件基于 ControlDesk 进行开发，该软件用于加载配置文件和实时监测模型运行状况，发动机-发电机组转速、动力电池电压及温度等状态可实时反映在上位机软件中。硬件在环仿真测试基于 CAN 通讯协议运行，从实时仿真环境向整车发送数据为状态量发电机转速 n_g 、动力电池组 SOC 及整车需求功率 P_{demand} ，硬件在环平台同时也向混合动力源实时调试平台发送动力源各组件运行状态，并在运行过程中实时接收实时调试平台的控制信号。

5.2.2 整车控制器开发及搭建

产品 MotoTron ECM565-128 硬件控制器，该控制器所采用处理器型号为 32 位 Freescale MPC565，该微处理器具有强大计算能力，可支持整车控制器开发。基于该

设备 CAN 2.0B 高速通信接口，该控制器可与 dSPACE 下位机进行通信，整车控制器发出控制指令为发动机油门开度。

MotoTron 软件开发基于 MATLAB/Simulink 平台，基于 stateflow 所创建逻辑可通过配套软件 MotoHawk 实现代码生成并与底层控制器进行交互。基于配套上位机软件 MotoTune 和 Canbootloader 将所生成代码下载至整车控制器中。将第五章优化训练所得估值神经网络参数从 python 平台提取并将其复现于整车控制器中，该控制模型输入为车辆状态量输出为车辆控制量即油门开度。整车控制器开发流程如图 5.10 所示：

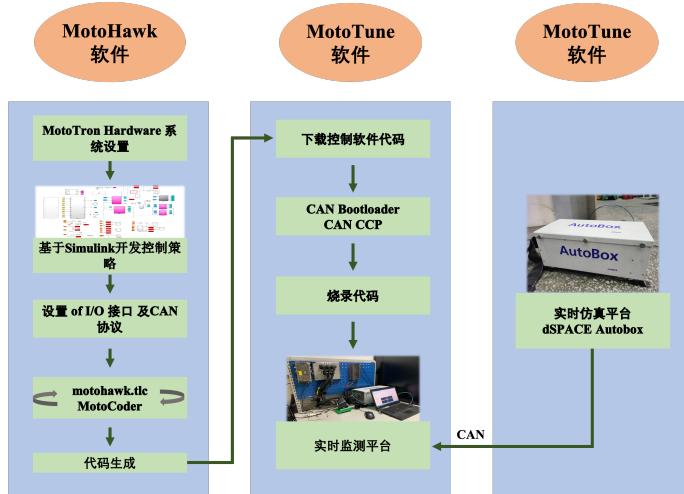


图 5.10 整车控制器开发流程

5.2.3 硬件在环仿真平台架构

将章节 6.1.1 及 6.1.2 对硬件在环被控系统及整车控制器进行连接可形成串联混合动力履带车辆硬件在环仿真系统。该硬件在环仿真系统包含有整车控制器 MotoTron、dSPACE 实时仿真平台及搭载于笔记本电脑上的监视软件 ControlDesk。通过 CAN 通讯方式实现整车控制器与被控系统的连接，通过以太网连接上位机与 dSPACE。

5.3 基于 NAF-DQL+PR 的混合动力能量管理策略硬件在环验证

为考察基于 NAF-DQL+PR 的混合动力车辆能量管理策略高效性和迁移性，硬件在环仿真实验所选取工况不同于训练工况。测试工况如图 5.11 (a) 所示。该工况持续时间 400s，最高车速 78 千米/小时，平均车速 41.7 千米/小时，该训练工况不同于训练所使用 US06 工况但符合本款履带车设计需求。

通过将经过第四章优化的第三章所介绍介于 NAF-DQL+PR 混合动力能量管理策略进行训练所得估值网络参数复制至整车控制器 MotoTron 实现能量管理策略在平台

间迁移。为验证本文所提出基于 NAF-DQL+PR 的能量管理算法改进，将基于 NAF-DQL 的能量管理策略加入进硬件在环仿真实验，通过对比两算法在发动机工作点、动力电池组 SOC 及燃油消耗量验证本文研究方法的最优性。

如图 5.11 所示，图 (b) 为基于 NAF-DQL+PR 的能量管理策略油门开度及基于 NAF-DQL 的能量管理策略油门开度，由图可见，相比于 NAF-DQL，本文研究算法在硬件在环仿真实验中输出油门开度在 50% 以上，而 NAF-DQL 算法输出油门开度普遍低于 50%，因此相对于 NAF-DQL 算法，NAF-DQL+PR 算法可使发动机工作更为充分而非 NAF-DQL 算法所展示的发动机在启停状态摇摆。从图 (c) NAF-DQL+PR 及 NAF-DQL 算法发动机工作点图也可对图 (b) 结果进行佐证，即 NAF-DQL+PR 算法发动机工作点位于中高效工作区，部分接近于最高效工作区，发动机输出功率较高，燃油消耗较低。NAF-DQL 算法发动机工作点多集中于低效工作区，发动机输出功率较低，燃油消耗较高。从图 (d) NAF-DQL+PR 及 NAF-DQL 动力电池电荷状态 SOC 曲线可看出，两算法初始 SOC 都为 0.7，对比 NAF-DQL+PR，由于 NAF-DQL 算法发动机工作在低效区，输出功率较低，动力电池输出功率辅助发动机工作，导致 NAF-DQL 算法的 SOC 曲线波动较 NAF-DQL+PR 明显，伴随有相对剧烈的充放电现象。从油门开度、发动机工作点及 SOC 变化曲线结果图可见，基于 NAF-DQL+PR 的能量管理策略能使发动机更多工作中相对高效工作区，并可确保动力电池 SOC 无剧烈充放电现象存在。

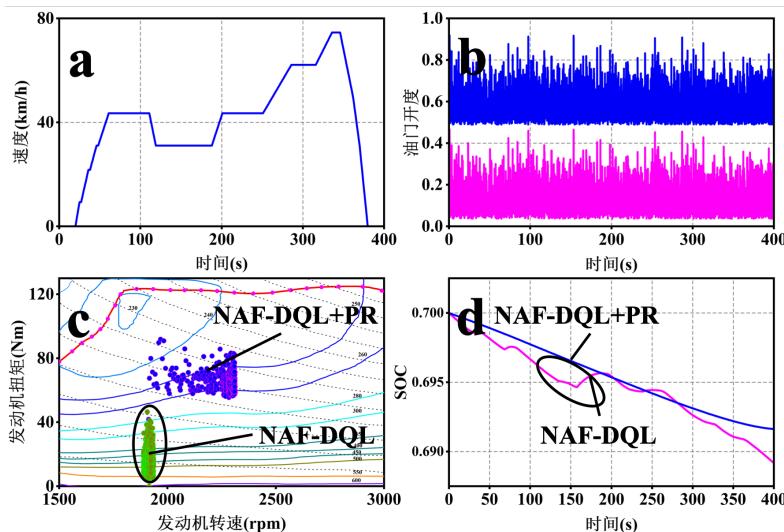


图 5.11 NAF-DQL+PR 与 NAF-DQL 硬件在环仿真结果

为进一步验证基于 NAF-DQL+PR 的能量管理策略优于基于 NAF-DQL 的能量管

理策略，在硬件在环仿真实验中对两算法油耗及电量消耗做比较，如表 5.1 所示。

表 5.1 NAF-DQL+PR 与 NAF-DQL 燃油及电量消耗对比

策略	NAF-DQL	NAF-DQL+PR
百公里燃油消耗(L/100km)	9.94	7.56
百公里电量消耗(kWh/100km)	15.45	11.97
初始 SOC	0.7	0.7
终止 SOC	0.68919	0.69163

从表中可见，NAF-DQL 算法每百公里耗油 9.94 升，消耗电量 15.45kWh，NAF-DQL+PR 每百公里耗油 7.56 升，消耗电量 11.97kWh。与 NAF-DQL 算法相比，NAF-DQL+PR 百公里燃油消耗量及电量消耗分别减少了 35.7% 和 36.8%。由于基于 NAF-DQL 的能量管理策略未能学习到迁移性更佳的策略，因此其发动机工作点集中于低效工作区，燃油消耗大且功率输出低，导致对动力电池的功率需求较大。基于 NAF-DQL+PR 的能量管理策略具有更佳的迁移性，所训练得到的深度神经网络针对未训练过的陌生工况同样表现不俗，使大部分发动机工作点集中于中高效工作区并使得动力电池输出平稳。因此，此硬件在环仿真实验结果验证了基于 NAF-DQL+PR 的能量管理策略优于基于 NAF-DQL 的能量管理策略，本文所提出的经验权值优先回顾模型及基于 GA-DQL 的深度强化学习超参数优化方法有效帮助深度强化学习方法找到最优能量管理策略，经验权值优先回顾模型及基于 GA-DQL 的深度强化学习超参数优化方法的高效性得以证明。

5.4 混合动力源实时调试程序试验调试

为测试混合动力源实时调试程序实时接收混合动力源数据的能力，使用本章 5.1 节所搭建硬件在环实时仿真系统对该软件进行硬件在环试验，基于 dSPACE 和 MotoTron 的实时仿真系统模拟混合动力源向混合动力源实时调试软件发送搭载于 CAN 通信的数据，以测试该软件能否实时接收、解析和显示数据以及能否对参数进行实时更改。

试验调试结果如图 5.12 所示，各路 CAN 数据正常发送至上位机并被混合动力源调试程序正确解析。针对于监控模块，没有错误数据出现且该模块响应及时并无肉眼可见延时或卡顿现象出现。针对于调试模块，混合动力源调试软件可将 100 个参数实时展示在表格中，参数调节修改及时，上位机和下位机之间并无明显延迟现象。



图 5.12 混合动力源调试软件显示界面调试结果

绘图模块调试结果如图 5.13 所示，绘图模块可针对要求读取参数进行实时折线图绘制，该图准确显示了四个状态参数的实时变化曲线，从而显示混合动力源状态量变化趋势，便于调试工程师实时掌握数据变化规律。

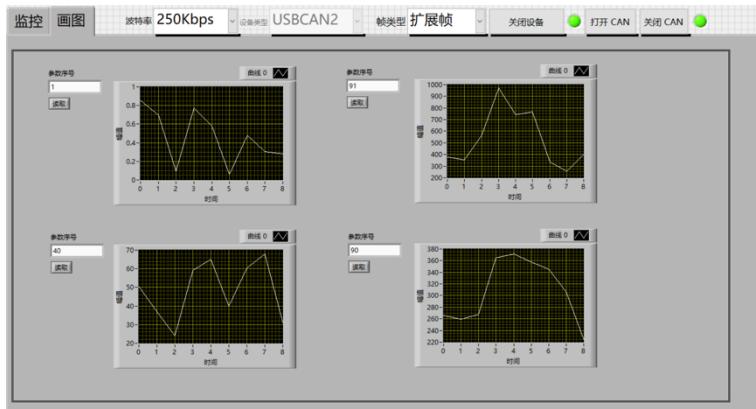


图 5.13 混合动力源调试软件绘图界面调试结果

5.5 本章小结

本章针对第四、五章提出并优化的基于 NAF-DQL+PR 的混合动力能量管理算法进行硬件在环仿真实验。通过使用 MotoTron 做为整车控制器，使用 dSPACE 搭载被控系统完成硬件在环系统搭建。将第三、四章完成优化的基于 NAF-DQL+PR 的能量管理策略神经网络参数移植入基于 MATLAB/Simulink 的 MotoTron 开发平台，通过设置 CAN 通信协议完成硬件在环实验准备。在硬件在环仿真实验中，对 NAF-DQL 及 NAF-DQL+PR 两算法在燃油消耗、电量消耗、发动机工作点分布和 SOC 曲线特点进行比较，结果表明基于 NAF-DQL+PR 的能量管理策略具有更好的燃油经济性，进一步验证基于 GA-DQL 的深度强化学习能量管理策略优化方法的有效性。由于硬件在环仿真实验环境使用陌生工况，因此可验证相对于 NAF-DQL 算法，NAF-

DQL+PR 算法具有更好的迁移性和适应性，又因此算法训练收敛快，因此该方法具有应用于实时混合动力车辆能量管理的潜力。通过使用基于 dSPACE 和 MotoTron 的硬件在环实时仿真平台对混合动力源实时调试软件进行硬件在环试验，实验结果表明此软件接收、解析和显示混合动力源数据及时，正确无误对 100 个参数进行修改，绘图功能准确及时，验证了此软件的有效性。

结论

全文总结

二十一世纪以来，我国经济腾飞，国内汽车保有量逐年升高，伴随着环境污染和全球能源短缺、我国石油不能自给自足等问题日益严重。此外，我国汽车工业随体量庞大但因历史原因自主创新能力较差，急需发展新能源汽车技术进行弯道超车。因此，环境、能源和产业三方面因素使得发展新能源汽车技术成为中国工业发展的当务之急。混合动力汽车结合了纯电动汽车和传统燃油汽车的优点，保证续驶里程同时实现节能减排。混合动力汽车通过能量管理技术对车辆内不同动力源实现能量分配，保证动力输出同时实现节能减排效果，满足中国市场需求。本文以课题组自研的串联混合动力履带车为研究对象，提出基于 NAF-DQL+PR 的混合动力车辆能量管理算法及基于 GA-DQL 的深度强化学习能量管理策略超参数优化方法，设计了一款面向于混合动力车辆动力源在线调试的软件程序。通过软件仿真和进行硬件在环仿真实验，完成能量管理策略、超参数优化方法及在线调试软件程序的验证与测试。全文主要研究成果总结如下：

(1) 完成了串联混合动力履带车系统建模。针对课题组自研串联混合动力履带车在 python 环境及 MATLAB/Simulink 中进行建模，完成混合动力履带车发动机-发电机组、动力电池组及高压功率模块等主要组件数学建模。在 dSPACE 实时仿真平台中搭建串联混合动力履带车前向模型用于硬件在环试验。

(2) 研发了面向混合动力车辆动力源在线调试的程序。基于 LabVIEW 平台及 CAN 通信协议设计具有实时监控、故障报警、在线参数调节和实时绘图的混合动力源在线调试系统。该软件设计目标于解决混合动力源调试中痛点问题，基于事件触发和定时触发，设计了针对发动机、发电机、动力电池及高压功率模块的监控显示界面及其后端逻辑，设计了针对 100 个参数的在线参数调节界面及其后端逻辑，设计了参数实时绘图模块。基于 dSPACE 实时仿真平台，完成了对混合动力源在线调试软件测试的测试，验证了其系统实时性和稳定性。

(3) 提出了基于 NAF-DQL+PR 的串联混合动力履带车能量管理策略。目标于推进基于深度强化学习的能量管理策略实时性研究，加速深度强化学习能量管理策略训练并实现基于深度强化学习的能量管理策略节能性，通过引入经验权值优先回顾模型，基于 TD-error 赋予训练历史经验数据优先级，引入 Sumtree 数据存储及提取模型，

引入基于取值更新权重的带权重神经网络更新模型，对深度强化学习模型历史训练数据进行带权重优先回顾。结合归一化优势函数，加速了基于深度强化学习能量管理策略训练，提升了基于深度强化学习能量管理策略训练效果。通过对 NAF-DQL+PR 算法训练并将训练结果与动态规划、NAF-DQL 及 DQL 进行对比，完成对算法最优性的验证。

(4) 提出了基于 GA-DQL 的深度强化学习能量管理策略超参数优化方法。由于学习率对深度强化学习算法影响较大，目标于解决 NAF-DQL+PR 能量管理算法超参数学习率设定问题，基于遗传算法提出基于 GA-DQL 的超参数优化方法。引入遗传算法，将学习率编码为二进制“基因”并初始化种群，将 NAF-DQL+PR 算法所得奖励作为个体适应度，利用个体适应度对个体进行筛选并通过交叉组合和变异完成子代生成进入下一轮进化。选择三组未经优化学习率的 NAF-DQL+PR 算法结果与优化过 NAF-DQL+PR 进行发动机工作点、动力电池 SOC 曲线及燃油消耗量的对比，验证了基于 GA-DQL 的超参数优化方法的有效性。结果表明经优化的基于 NAF-DQL+PR 的混合动力履带车能量管理策略在以上三个方面均有显著提升。

(5) 基于硬件在环仿真平台进行了基于 NAF-DQL+PR 能量管理算法的实时控制验证。为验证基于 NAF-DQL+PR 能量管理算法在实时控制环境下的最优性、迁移性及进一步验证基于 GA-DQL 超参数优化方法的有效性，将 MotoTron 快速原型控制器作为整车控制器，完成整车控制器及被控系统的 CAN 总线通讯协议开发。将已优化的基于 NAF-DQL+PR 能量管理算法网络参数复现于整车控制器。将整车控制器、上位机及已搭建好串联混合动力履带车前向仿真模型的 dSPACE 连接进行硬件在环仿真实验。通过在硬件在环仿真实验中将 NAF-DQL+PR 与 NAF-DQL 算法进行发动机工作点、油门开度、动力电池 SOC、燃油和电量消耗五个维度的对比，验证 NAF-DQL+PR 算法的节能效果。硬件在环实验结果表明，基于 NAF-DQL+PR 的能量管理策略具有好燃油经济性和迁移性。基于 GA-DQL 的超参数优化方法有效提升了基于 NAF-DQL+PR 的能量管理策略表现。

创新点

综合全文内容，本文在针对串联混合动力履带车控制器在环仿真研究方面取得部分成果，创新点总结如下：

(1) 建立串联混合动力履带车前向模型。在 python 和 MATLAB/Simulink 环境

中搭建串联混合动力履带车前向模型用于能量管理策略开发与验证。

(2) 将经验权值优先回顾模块引入基于深度强化学习的能量管理策略。根据历史经验数据训练效果进行带权重抽取重复训练并对神经网络进行带权重更新。该模型的引入提升了深度强化学力量管理策略燃油经济性，提高了其训练速度并推进深度强化学力量管理策略实时性研究。

(3) 基于遗传算法对深度强化学力量管理策略超参数进行优化。引入遗传算法寻找最优深度强化学力量算法学习率，进一步优化了基于 NAF-DQL+PR 的能量管理策略，提升其燃油经济性及训练效率。

(4) 设计研发目标于串联混合动力履带车调试软件程序。基于 LabVIEW 设计在线调试程序，基于事件触发和定时触发，实现针对混合动力源各组件状态实时监控，实现对 100 个参数实时在线调整功能，实现组件状态实时绘图功能。

工作展望

本文对串联混合动力履带车能量管理问题进行深入研究并完成相关算法验证，推进了相关算法研究进展。总结本文工作内容及所的研究成果，以下几个方面还值得进一步研究：

(1) 本文所使用基于 NAF-DQL+PR 的能量管理策略，计算时间短但距离实时应用尚有距离，未来可基于本算法开展联合物联网、智能车等相关研究实现在线更新的基于学习的能量管理策略。

(2) 受限于实验条件，本文提出的能量管理策略未能进行实车验证，未来可基于(1)所开发的在线能量管理策略进行实车验证，验证本文所提出算法的有效性。

(3) 本文所提出深度强化学力量参数优化方法受限于运算能力只对学习率进行重点探究，未来通过使用随机搜索等方法实现多个超参数优化，提升深度强化学力量管理策略训练效果。

参考文献

- [1] Y Wang, Z Sun and Z Chen. Development of energy management system based on a rule-based power distribution strategy for hybrid power sources[J]. Energy 175(2019) 1055-1066,2019
- [2] T Pippia, J Sijs and B Schutter. A Single-Level Rule-Based Model Predictive Control Approach for Energy Management of Grid-Connected Microgrids[J]. IEEE Transactions on Control Systems Technology.
- [3] A R Bhattiab and Z Salam. A rule-based energy management scheme for uninterrupted electric vehicles charging at constant price using photovoltaic-grid system[J]. Renewable Energy, Volume 125,2018,384-400
- [4] Y Liu, J Gao, D Qin,Y Zhang and Z Lei.Rule-corrected energy management strategy for hybrid electric vehicles based on operation-mode prediction[J]. Journal of Cleaner Production. Volume 188, 1 July 2018, Pages 796-806
- [5] S G Wirasingha, A Emadi. Classification and review of control strategies for plug-in hybrid electric vehicles[J]. IEEE Transactions on Vehicular Technology, 2011, 60(1): 111–122.
- [6] N Jalil, N A Kheir, M Salman. Rule-based energy management strategy for a series hybrid vehicle[C]. American Control Conference. 1997.
- [7] J Peng, H Fan, H He and D Pan. A Rule-Based Energy Management Strategy for a Plug-in Hybrid School Bus Based on a Controller Area Network Bus[J]. Energies 2015, 8, 5122-5142
- [8] C J Mansour. Trip-based optimization methodology for a rule-based energy management strategy using a global optimization routine: The case of the Prius plug-in hybrid electric vehicle[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2016, 230(11): 1529-1545.
- [9] B Padmarajan, A McGordon, P Jennings. Blended rule based energy management for PHEV: System Structure and Strategy[J]. IEEE Transactions on Vehicular Technology, 2016, 65(10): 8757-8762.
- [10] H Banvait, S Anwar, Y Chen. A rule-based energy management strategy for plugin hybrid electric vehicle (PHEV)[J]. Proceedings of the American Control Conference, 2009: 3938-3939

- [11] L Xi, X Zhang, C Geng. Energy management strategy optimization of extended-range electric vehicle based on dynamic programming[J]. Journal of Traffic & Transportation Engineering, 2018.
- [12] Ossart F, Qi J, Marchand C. Comparative Study of Real-Time HEV Energy Management Strategies[J]. IEEE Transactions on Vehicular Technology, 2018, PP(99):1-1.
- [13] H He, J Guo, J Peng. Real-time global driving cycle construction and the application to economy driving pro system in plug-in hybrid electric vehicles[J]. Energy, 2018, 152.
- [14] W Zhen, J Xu, D Halim. HEV Energy Management Fuzzy Logic Control Based on Dynamic Programming[C] IEEE Vehicle Power & Propulsion Conference. IEEE, 2015.
- [15] B Jreddi, Y Mishra, G Ledwich. Dynamic programming based home energy management unit incorporating PVs and batteries[C] IEEE Power & Energy Society General Meeting. IEEE, 2018.
- [16] L Li , Y Zhang , C Yang. Hybrid genetic algorithm-based optimization of powertrain and control parameters of plug-in hybrid electric bus[J]. Journal of the Franklin Institute, 2015, 352(3):776-801.
- [17] Favre L, Robyr J L, Gonon F, et al. Improvement of the environmental impact of the global energy management of buildings by genetic algorithm optimization[C]//2018 5th International Conference on Electric Power and Energy Conversion Systems (EPECS). IEEE, 2018: 1 -6.
- [18] L Kang, Y Zhang, H Dou and R Zou. Design of an integrated energy management strategy for a plug-in hybrid electric bus[J]. Journal of Power Sources. Volume 448, 1 February 2020, 227391
- [19] A Panday, H O Bansal. Optimal fuel control of series-parallel input split hybrid electric vehicle using genetic algorithm based control strategy[C]. International Conference on Energy Economics and Environment - 1st IEEE Uttar Pradesh Section Conference, UPCON-ICEEE 2015, 2015.
- [20] T Zhu, B Li, C Zong, Y Wu. Series hybrid electric vehicle power system optimization based on genetic algorithm[J]. IOP Conference Series: Materials Science and Engineering, 2017, 235:012013.

- [21] X Jiao, T Shen. SDP Policy Iteration-based energy management strategy using traffic information for commuter hybrid electric vehicles[J]. *Energies*, 2014, 7(7): 4648-4675.
- [22] J Liu, H Peng. Control optimization for a power-split hybrid vehicle[J]. *IEEE, Proceedings of the American Control Conference*, , 2006: 466-471.
- [23] C C Lin, H Peng, J W Grizzle. A stochastic control strategy for hybrid electric vehicles [C]. *IEEE, American Control Conference*. 2004: 4710-4715 vol.5.
- [24] C C Lin, H Peng, J W Grizzle, J Kang. Power management strategy for a parallel hybrid electric truck[J]. *IEEE Transactions on Control Systems Technology*, 2004, 11(6): 839-849.
- [25] L Li, C Yang, Y Zhang, L Zhang, J Song. Correctional DP-based energy management strategy of plug-in hybrid electric bus for city-bus route[J]. *IEEE Transactions on Vehicular Technology*, 2015, 64(7): 2792-2803.
- [26] Z Wei, J Xu, D Halim. HEV power management control strategy for urban driving[J]. *Applied Energy*, 2017, 194: 705-714
- [27] H K Roy, A McGordon, P A Jennings. A generalized powertrain design optimization methodology to reduce fuel economy variability in hybrid electric vehicles[J]. *IEEE Transactions on Vehicular Technology*, 2014, 63(3): 1055-1070.
- [28] T Liu, H Yu, H Guo, T Qin, Y Zou. Online energy management for multimode plug-in hybrid electric vehicles[J]. *IEEE Transactions on Industrial Informatics*, 2019, 15(7): 4352-4361.
- [29] Z Chen, R Xiong, C Wang, J Cao. An on-line predictive energy management strategy for plug-in hybrid electric vehicles to counter the uncertain prediction of the driving cycle[J]. *Applied Energy*, 2017, 185: 1663-1672.
- [30] Z Chen, R Xiong, J Cao. Particle swarm optimization-based optimal power management of plug-in hybrid electric vehicles considering uncertain driving conditions[J]. *Energy*, 2016, 96: 197-208.
- [31] Z Lei, D Cheng, Y Liu, D Qin, Y Zhang, Q Xie. A dynamic control strategy for hybrid electric vehicles based on parameter optimization for multiple driving cycles and driving pattern recognition[J]. *Energies* 2017, 10, 54.
- [32] Z Chen, R Xiong, K Wang, B Jiao. Optimal energy management strategy of a plug-in hybrid electric vehicle based on a particle swarm optimization algorithm[J]. *Energies*, 2015, 8

- (5): 3661-3678.
- [33] C Gong, M Hu, S Li, et al. Equivalent consumption minimization strategy of hybrid electric vehicle considering the impact of driving style[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2019, 233(10): 2610-2623.
- [34] H Li, Ravey A, A N'Diaye. A novel equivalent consumption minimization strategy for hybrid electric vehicle powered by fuel cell, battery and supercapacitor[J]. Journal of Power Sources, 2018, 395: 262-270.
- [35] F Zhang, H Liu, Y Hu. A supervisory control algorithm of hybrid electric vehicle based on adaptive equivalent consumption minimization strategy with fuzzy PI[J]. Energies, 2016, 9 (11): 919.
- [36] C Xiang, F Ding, W Wang. MPC-based energy management with adaptive Markov-chain prediction for a dual-mode hybrid electric vehicle[J]. Science China Technological Sciences, 2017, 60(5): 737-748.
- [37] M Shahverdi, M Mazzola, S Abdelwahed, et al. MPC-based power management system for a plug-in hybrid electric vehicle for relaxing battery cycling[C]//2016 IEEE Transportation Electrification Conference and Expo (ITEC). IEEE, 2016: 1-6.
- [38] K Han, J Tao, L Xie. Rule and MPC based Hybrid Energy Allocation System for Hybrid Electric Vehicle[C]//2019 Chinese Automation Congress (CAC). IEEE, 2019: 57-62.
- [39] K Williams. Real-Time Stochastic Predictive Control for Hybrid Vehicle Energy Management[J]. 2018(May).
- [40] D Shen, V Bensch, S Müller. Model predictive energy management for a range extender hybrid vehicle using map information[J]. IFAC-PapersOnLine, 2015, 28(15): 263–270.
- [41] S Zhang, R Xiong, C Zhang. Pontryagin's Minimum Principle-based power management of a dual-motor-driven electric bus[J]. Applied Energy, 2015, 159: 370-380.
- [42] Y Zou, T Liu, F Sun, H Peng. Comparative study of dynamic programming and pontryagin's minimum principle on energy management for a parallel hybrid electric vehicle[J]. Energies, 2013, 6(4): 2305-2318.
- [43] T Feng, L Yang, Q Gu, Y Hu, T Yan, B Yan. A supervisory control strategy for plug-in hybrid electric vehicles based on energy demand prediction and route preview[J]. IEEE Transactions on Vehicular Technology, 2015, 64(5): 1691-1700.

- [44] Q Jiang, F Ossart, C Marchand. Comparative study of real-time HEV energy management strategies[J]. IEEE Transactions on Vehicular Technology, 2017, 66(12): 10875-10888.
- [45] Y Li, H He, J Peng. Energy Management Strategy for a Series Hybrid Electric Vehicle Using Improved Deep Q-network Learning Algorithm with Prioritized Replay[J]. DEStech Transactions on Environment, Energy and Earth Sciences, 2018 (iceee).
- [46] Y Li, H He, J Peng. Deep Reinforcement Learning-Based Energy Management for a Series Hybrid Electric Vehicle Enabled by History Cumulative Trip Information[J]. IEEE Transactions on Vehicular Technology, 2019, 68(8): 7416-7430.
- [47] H He, Cao J, X Cui. Energy optimization of electric vehicle's acceleration process based on reinforcement learning[J]. Journal of Cleaner Production, 2020, 248: 119302.
- [48] T Liu, X Hu, S Li. Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle[J]. IEEE/ASME Transactions on Mechatronics, 2017, 22(4): 1497-1507.
- [49] T Liu, X Hu, W Hu. A heuristic planning reinforcement learning-based energy management for power-split plug-in hybrid electric vehicles[J]. IEEE Transactions on Industrial Informatics, 2019, 15(12): 6436-6
- [50] J Cao, R Xiong. Reinforcement learning-based real-time energy management for plug-in hybrid electric vehicle with hybrid energy storage system[J]. Energy Procedia, 2017, 142: 1896-1901.
- [51] X Hu, T Liu, X Qi. Reinforcement Learning for Hybrid and Plug-In Hybrid Electric Vehicle Energy Management: Recent Advances and Prospects[J]. IEEE Industrial Electronics Magazine, 2019, 13(3): 16-25.
- [52] Z Chen, H Hu, Y Wu. Energy management for a power-split plug-in hybrid electric vehicle based on reinforcement learning[J]. Applied Sciences, 2018, 8(12): 2494.
- [53] Y Li, H He, J Peng. Power management for a plug-in hybrid electric vehicle based on reinforcement learning with continuous state and action spaces[J]. Energy Procedia, 2017, 142: 2270-2275.
- [54] J Wu, Y Zou, X Zhang. An Online Correction Predictive EMS for a Hybrid Electric Track ed Vehicle Based on Dynamic Programming and Reinforcement Learning[J]. IEEE Access, 2019, 7: 98252-98266.

- [55] T Liu, X Hu, S Li, D Cao. Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle[J]. IEEE/ASME Transactions on Mechatronics, 2017, 22(4): 1497-1507.
- [56] X Qi, G Wu, K Boriboonsomsin, M J Barth. A novel blended real-time energy management strategy for plug-in hybrid electric vehicle commute trips[J]. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, IEEE, 2015, 2015-Octob (September): 1002-1007.
- [57] V Mnih, K Kavukcuoglu, D Silver, A A. Rusu, J Veness, M G Bellemare, A Graves, M Riedmiller, A K Fidjeland, G Ostrovski, S Petersen, C Beattie, A Sadik, I Antonoglou, H King, D Kumaran, D Wierstra, S Legg, D Hassabis. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529.
- [58] J Wu, H He, J Peng, Y Li, Z Li. Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus[J]. Applied Energy, 2018, 222: 799-811.
- [59] Mania H, Guy A, Recht B. Simple random search provides a competitive approach to reinforcement learning[J]. arXiv preprint arXiv:1803.07055, 2018.
- [60] Barsce J C, Palombarini J A and Martínez E C 2018 Towards autonomous reinforcement learning: automatic setting of hyper-parameters using Bayesian optimization CoRR abs/1805.04748
- [61] 郑旭, 郭汾. 动力电池模型综述[J]. 电源技术, 2019, 43(03):167-170.
- [62] 汪明德, 赵毓芹, 祝嘉光. 坦克行驶原理 [M]. 国防工业出版社, 1983.

攻读学位期间发表论文与研究成果清单

学术论文

- [1] Runnan Zou, Yuan Zou, Yanrui Dong, Likang Fan. A Self-Adaptive Energy Management Strategy for Plug-in Hybrid Electric Vehicle Based on Deep Q Learning[C]. 4th International Conference on Artificial Intelligence, Automation and Control Technologies, 2020. (EI 会议, 已录用)

参与科研项目

1、陆军装备部混合动力源重点研发计划 国防基础科研项目

主要负责混合动力源实时调试程序开发和能量管理策略开发及动力源性能测试。

2、陆军装备部“跨越险阻 2018”地面无人系统挑战赛

主要负责混合动力履带车结构设计及各组件选型匹配。

3、教学用实物平台开发 企业横向课题

主要负责双质量振动试验台电气组件参数匹配和选型、操作界面开发及平台控制代码编写。

致谢

本文主体部分完稿于 2020 年 4 月疫情中的北京家中。万语千言，在写完严谨的学术论文后竟也一时噎住。在与北理工离别之际，回首硕士三年和在北理工求学的七年，感慨实多。从大四进实验室到现在研究生即将毕业，我的成长和蜕变很大，这里都离不开学校、老师、同学和家人对我的帮助。

首先要感谢邹渊教授接受我为他的学生，在四年间我接受他无微不至的学术指导，从一个门外汉进入到学术的殿堂。邹老师严谨的治学精神和一丝不苟的研究态度一直深深地影响着我。在邹老师的带领和指导下我进入到各种科研项目中，在为国家做贡献的同时学习知识和技术。每一周的工作汇报、每一次的学术讨论和小组讨论，邹老师切中肯綮的观点总能带给我新的学术思考。是邹老师一次又一次为我指明研究方向和技术路线，带领我进行科研。本文的工作都在邹老师的带领下完成，因此本文也凝结着邹老师在这三年来对我付出的心血。在本文即将完成之际，学生由衷地感谢邹渊老师对我多年来的培养与信任。

我还要感谢实验室的张旭东老师。作为课题组年轻的老师，总能在工作中给予我技术上的帮助。在混合动力源装备研制的项目中，张老师的亲力亲为激励我在艰苦的实验条件下继续拼搏。此外，张老师新颖的学术思考也帮助我在混合动力能量管理策略开发中找到突破口。在此向张旭东老师致以最真诚的谢意。

学术上离不开两位导师的指导，生活上离不开我的母亲杜秋艳女士的支持。感谢我的母亲几十年如一日地信任、支持和爱护我，您是我努力工作、认真生活的后盾。感谢我的母亲为我进行学术研究所付出的一切。我身为人的存在离不开母亲，我成长为研究生的过程离不开母亲的养育。我愿将此篇拙作作为礼物献给我的母亲。在人生第二十五个夏天，我向我的母亲杜秋艳女士表达最殷切的爱意，希望您永远健康快乐。

我还要感谢实验室的同学们，感谢武金龙、张涛、焦飞翔和何定波四位师兄对我学术上无微不至的帮助，感谢杜广泽、郑壮壮和翟瑞卿三位同学在三年研究生生活中的陪伴，感谢吴喆、郭凌雄和刘仕杰对本文相关工作的贡献，感谢魏守洋、刘超、陈昶、刘献韬、樊杰、李晓良、肖慕、孔泽慧、白一帆、张彬、邹震宇、郭宁远、肖鸿磊、赵世奇、杜国栋、孙文景、董国顺、徐福康、李圆圆在实验室工作中对我的帮助。实验室的同学们陪我走过研究生三年岁月，感谢你们的帮助。

在生活中，我要感谢室友张海鸣、王琛和程明。三位善良的室友让我在充实的科

研之余有一个港湾可以放松身心，感谢你们的陪伴。感谢学校 Au-time Fitness 的杨思源、邬家骥、武方虎和姚远，感谢你们提供的优越的训练场所让我能保持身心健康，在科研之路上越战越勇。

特别感谢胡晨星、樊利康和郑思宇三位老师和学长对我科研之路的巨大帮助，感谢你们在过去一年里的指导和陪伴，是你们带我走出迷茫找到人生前进的方向。

此外，我想在这里特别感谢我的女朋友董妍睿小姐。感谢你在近一年来的陪伴、信任和包容，感谢你相信我的豪言壮语，感谢你对我研究的帮助。能够与你作伴是我的幸运，希望我们能一直这样平淡地相互守望。未来不远，吻你万千。

感谢对本文提出建议的专家，你们的建议是我进步的基石，你们的肯定是我科研之路的动力。

感谢北理工给予我七年学习生活之地，母校的养育学生终身难忘。再见北理工，我会继续努力做一个实事求是的研究者。