

**本科毕业设计（论文）**

题目**：**高速路车辆自主跟驰模型及其算法实现

学 院： 计算机科学与工程学院

专业班级： 计算机科学与技术2018级4班

姓 名： 于程

学 号： 201801060423

指导教师： 张福新

完成日期： 2022年月日

教务处制

**BACHELOR'S DEGREE THESIS OF SHANDONG UNIVERSITY**

**OF SCIENCE AND TECHNOLOGY**

**Writing the title of the paper in English here**

College：College of Computer Science and Engineering

Subject： Computer Science and Technology

Name： Yu Cheng

Directed by： Zhang Fuxin Professor

**QINGDAO CHINA**

**郑 重 声 明**

本人呈交的毕业论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本设计（论文）的研究成果不包含他人享有著作权的内容。对本设计（论文）所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本设计（论文）的知识产权归属于培养单位。

本人签名： 日期：

摘 要

由于泥沙与水流的相互作用，使得河流发生演变，因此泥沙特性与水流特性均是河流动力学的重要研究课题。当水流中含有植物时，水流的紊动特性会发生明显的改变，从而引起泥沙的一些特性如沉速发生改变。本文以实验为基础，结合理论分析，研究了在静水条件下刚性植物对泥沙沉速的影响，同时在水槽中通过改变流量来研究在恒定均匀流条件下非淹没植物对泥沙沉降轨迹的影响，得到如下主要结论：

**关键词：**关键词1；关键词2；关键词3

**ABSTRACT**

(**Times New Roman** 小2加粗)

Fluvial river processes evolve over time in response to the constant interaction between sediment and the water column. If vegetation is present within the water column, the change in turbulence characteristics will impact the movement of sediment, in particular the settling velocity. In this paper, the influence of vegetation on the settling velocities of sediment particles is studied experimentally. The non-submerged vegetation friction factor in steady uniform flow is considered by under different flume discharge quantities. The main outcomes can be summarized as follows:

（**Times New Roman** 1.5倍行距）

**KEY WORDS:** sediment; rigid vegetation; settling velocity; turbulence characterize（**Times New Roman** 小4）

**目 录**

**第1章 绪论**

**1.1**  **研究背景与意义**

高速公路是一种中远距离运输的主要方式，具有运量大、速度快、效率高、辐射广等诸多优势。不仅促进了沿线沿线产业结构的优化和区域经济的发展，还带动了运输结构的改善和运输效率的提高。随着高速公路网络的日益完善和高速公路运输量的不断增加，在部分经济发达地区高速公路的设施容量已经无法满足日益增长的交通需求，高速公路的拥堵和安全问题日益突出。不合理的车辆跟驰模型设计容易引发高速路上潜在的冲突。一方面容易造成较高的交通事故发生率,另一方面也降低了周边区域的车辆通行效率。合理的车辆跟驰模型可以辅助驾驶员决策，预防及疏解高速公路拥堵、提高路网运输安全和效率。高速路的交通安全和通行效率问题是世界各国面临的难题。然而目前现在的高速路车辆行驶的主要方式为驾驶员驾驶，考虑到通过驾驶员驾驶来实现车辆跟驰会出现一些问题，比如：驾驶员疲劳，单位时间无法实现大的客流量等。因此高速公路车辆跟驰模型的算法研究具有重要意义。

**1.2** **国内外研究现状**

国内顾海燕深入分析了车联网处于不同发展阶段时交通流中车辆表现出来的不同行驶特性，构建了各阶段对应的车辆跟驰模型，同时将车联网技术与自动控制技术相结合，构建符合车联网不同发展阶段车辆跟驰特性的车辆跟驰模型，并通过仿真探究各模型的稳定性条件[4]。朱冰等针对自适应巡航控制系统在控制主车跟驰行驶中受前车运动状态的不确定性影响问题，在分析车辆运动特点的基础上，提出一种能够考虑前车运动随机性的跟驰控制策略。将车辆跟驰问题构建为一定奖励函数下的马尔可夫决策过程，引入深度强化学习研究主车跟驰控制问题。利用近端策略优化算法建立车辆跟驰控制策略，通过与前车运动随机过程模型进行交互式迭代学习，得到具有运动不确定性跟驰环境下的主车纵向控制策略，实现了对车辆纵向控制的最优决策[1]。李姣在传统交通流元胞自动机模型的基础上,引入了对智能网联车辆和传统人工驾驶车辆在行驶差异性上的考虑,建立了智能网联车辆参与的多车道混合交通元胞自动机模型。基于智能网联车辆集结为车队行驶这一设想,建立了基本图模型、规定了智能网联车队在交通流元胞自动机模型中的形成逻辑和速度控制规则。基于对交通流特性的分析,建立了智能网联车辆参与的混合交通流能耗模型,并对智能网联车辆的两种组织形式分别进行了仿真[3]。

秦严严等针对ACC车辆跟驰模型存在的不足,提出基于非线性动态车头间距的ACC跟驰模型，并分析模型特性。应用随速度非线性动态变化的车头间距策略，建立基于非线性动态车头间距策略的ACC跟驰模型，从交通流均衡态与非均衡态两个层面，分析所提模型在通行能力提升与跟驰模型稳定性两个方面的特性，并与以往模型进行对比分析[5]。丁婉婷根据智能网联汽车的特征和优势,基于传统的换道决策模型,提出适用于智能网联汽车的自主性换道决策模型。对提出的换道决策流程中的单车换道决策进行研究和建模。在单车换道决策中考虑换道行为对原车道和目标车道上多辆后随车的影响,以减少换道行为对原车道和目标车道上上游交通的影响。再次,对提出的换道决策流程中的车辆协同换道决策进行研究,提出一种基于博弈论的车辆协同换道决策模型[6]。李文礼等为提高车辆对行驶环境的自我学习和决策能力，提出了一种基于深度确定性策略梯度（DDPG）的车辆自主避撞决策控制模型。基于 Markov 决策过程的强化学习理论和车辆纵向运动学特性，设计了决策所需目标对象及自车信息的状态空间和自车减速度的动作空间，以安全性、舒适性和效率因素为多目标奖励函数的端到端的车辆自主避撞决策模型。在实现车辆有效避撞的同时，兼顾乘坐舒适性，且性能优于模糊控制[7]。

国外Ngoduy提出了一个宏观模型来描述智能交通流的动态，其中智能车辆比手动车辆更靠近彼此并以车辆排的形式运行，每个车辆排包含几辆汽车。该模型是从跟驰模型发展而来的，并对模型参数集，基于线性稳定性的方法，构建线性稳定性图。分析发现，基于车辆排的驾驶模型受环境影响的扰动较小，增强了交通流的稳定性[8]。Ngoduy还提出了一种适用于手动和自适应巡航控制(ACC)车辆混合交通流的动力学理论。在模型中，ACC车辆的加速/减速是使用微观模型明确指定的。采用线性稳定性方法推导出交通流稳定性的影响条件。同时对ACC车辆的交通流进行建模分析，明确了不同参数对交通流稳定性的影响作用。分析结果表明，ACC车辆有助于提高交通流量的稳定性[9]。Yi J等讨论了ACC车辆宏观交通流传播稳定性的一般方法。提出了一个交通流速度饱和的宏观模型，其中每辆车都由ACC间距策略控制。研究了非线性交通流稳定性准则。针对广义ACC交通流模型，推导出了交通流稳定性与模型参数之间的定量关系。得出的结果与先前使用具有恒定时间间隔(CTH) 策略的微观和宏观模型获得的结果一致[10]。

Tiaprasert等对智能车辆参与的混合交通流中所出现的超车、换道、先进先出等行为进行了考虑，建立了适用于不同类型车辆的CTM模型，并利用数据库验证了模型的科学性[11]。Ma为研究记忆加速度对交通流稳定性和车辆油耗的影响，提出了一种改进的跟驰模型。将控制理论应用于稳定性条件分析，理论分析与数值模拟结果一致。数值结果表明，记忆加速度对交通流的稳定性有积极的影响。同时考虑记忆效应提出了一种改进的微观模型。用于探索降低车辆油耗的有效方法。数值结果表明，在保证记忆步长的情况下，车辆的油耗可以得到较好的控制[12]。Jie Me等为了提高运输系统的安全性和效率，提出了一种有效的车辆编队方法，令其中一组车辆通过最小化每辆车的跟踪误差来保持预先定义的移动模式。在车对车网络中针对车辆的CAM传输和控制参数联合优化无线电资源的分配方案，以在保证可靠性要求的同时最大限度地减少跟踪误差，确保车辆之间的通信和车辆排稳定性。同时提出了一个基于跟踪误差的调度准则，然后通过二分图匹配获得无线资源分配方案[13]。Kiran基于Atari游戏学习的成功演示和Go by谷歌DeepMind，提出了一个使用[深度强化学习](https://so.csdn.net/so/search?q=%E6%B7%B1%E5%BA%A6%E5%BC%BA%E5%8C%96%E5%AD%A6%E4%B9%A0&spm=1001.2101.3001.7020)的自主驾驶框架。它包含了用于信息集成的递归神经网络，使汽车能够处理部分可观察的场景。还集成了最近关于attention模型的工作。该框架在一个名为TORCS的开源3D赛车模拟器中进行了测试。仿真结果表明，自动驾驶汽车能够较好的学习复杂道路的曲率和其他车辆简单的相互作用的场景[14]。

**1.3** **论文主要工作**

通过认真查阅与车辆网技术发展相关的专业论文和相应的学术研究成果，对国内外相关学者已经成功实现的车辆跟驰模型进行了深入的研究和分析后。本文通过采取了一种已经实现了的有效的车辆编队方法，并结合深度强化学习算法在自动驾驶领域取得大量成就的研究经验，对车辆跟驰算法进行了进一步的研究，实现了一种基于深度强化学习的车辆编队方法。

本文主要完成了以下几个方面的工作：

（1）深入理解学习了车辆的跟驰模型，通过采取将车辆编队的方式来控制车辆。将多个车辆按照既定的数量划分为一个车辆排，以车辆排中的相邻两辆车为控制单位。车辆排中的所有车辆都采取恒定间距策略，通过最小化每个控制单位中车辆的跟踪误差来保持预先定义的移动模式，从而使整个车辆排保持稳定移动模式。

（2）将车辆跟驰问题构建为一定奖励函数下的马尔可夫决策过程，引入深度强化学习算法来优化车辆跟驰控制问题，确定了以车辆编队的稳定控制为目标，建立了车辆追踪误差目标函数，通过使车辆追踪误差最小化，获取最优的车辆控制动作，实现车辆编队的稳定控制。仿真结果表明，与现有方案相比，所提方案在减小跟踪误差的同时也保证了车辆排的稳定。

（3）利用Python实现相应的算法，并使用PyCharm进行仿真实验测试。

**1.4** **论文章节结构**

第1章 绪论。介绍了本文的研究背景，国内外研究的现状，研究完成的目的与意义以及本文完成的主要工作。

第2章 基本知识介绍。介绍了本文涉及的深度强化学习算法“DQN”,“DDPG”以及“车辆跟驰模型”的概念。

第3章 车辆跟驰模型的构建。介绍了本文的跟驰模型的构建过程，详细描述了采用车辆编队的方式构建模型的步骤。

第4章 基于深度强化学习的车辆跟驰算法设计。介绍了本文采用的两种深度强化学习算法，并介绍了如何将深度强化学习应用到建立的模型中，如何在模型中定义强化学习的状态空间、动作和奖励。

第5章 实验验证。进行仿真实验并对实验的结果进行展示，同时对实验结果进行详细的对比分析。

第6章 总结与展望。对各章内容要点进行简要分析总括，并详细分析各种算法的具体利弊。

**第2章 基本知识介绍**

**2.1** **车辆跟驰模型**

**2.1.1 车辆跟驰模型的概述**

车辆跟驰的概念最早可以追溯到20世纪50年代初期，学者们运用运筹学的知识成功解释了跟驰模型以来，跟驰理论的发展已经持续了超过半个世纪之久。跟驰模型是利用动力学的方法，研究在无法超车的单一车道上，车辆列队行驶时后车跟随前车行驶状态的一种理论。跟驰模型在微观仿真、通行能力分析、自适应巡航控制、交通安全评价等领域有着广泛的应用，因此，近60年来，吸引了交通工程学、系统工程、自动控制、车辆工程等诸多领域的专家及学者的深入研究，并取得了众多的富有意义的研究成果。车辆跟驰理论已经成为道路交通流理论的核心内容之一。

车辆跟驰模型可以用于研究人、车单元的运动并在相互作用的层次上分析车道交通流的特性，描述交通流的宏观特性，应用于交通规划、交通管理与控制，充分发挥交通设施的功效，解决交通问题。

**2.1.2 车辆跟驰模型分类**

车辆跟驰模型方面的重要研究成果可以从交通工程、统计物理等不哦那个角度分成按图1.1所示的7类；刺激-反应模型、安全距离模型、生理-心理模型、人工智能模型、优化速度模型、智能驾驶模型以及元胞自动机模型。

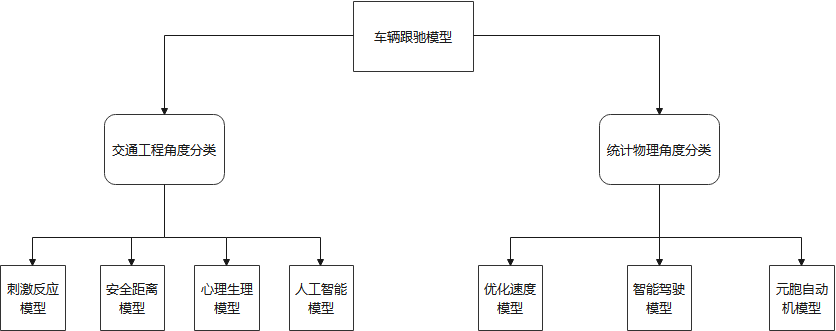


图1.1 车辆跟驰模型

一、按交通工程角度分类

（1）刺激反应模型是最基本的跟驰行为建模框架，这类模型是最早的也是对后续研究和发展较大的模型之一。刺激反应车辆跟弛模型的基本思想是在车辆跟驰过程中，驾驶员刺激反应来源于当前车辆与其前车之间的速度差，经过反应延迟后，驾驶员以改变加速度作为对该刺激的反馈，反馈对刺激的敏感性程度用敏感系数来表示。

（2）安全距离模型也叫做防撞模型，安全距离模型的主要思想是在考虑车辆性能限制、驾驶人特性的条件下，车辆驾驶人结合自身的经验判断并试图以一定的距离跟随前车行驶，这个距离可以满足在前车紧急刹车的情况下，后车的驾驶员依旧有足够的反应时间作出反应，进行紧急制动并防止发生碰撞。

（3）人工智能模型是20世纪90年代伴随着人工智能的发展出现的一种新型车辆跟驰模型。和传统的车辆跟驰模型相比，人工智能模型提高了模型的灵活性，也更适合于描述人的复杂心里与生理行为，是今后车辆仿真模型发展的重要方向之一。不过人工智能模型也存在一些问题，比如无法预先确定稳定性，神经网络模型需要学习的计算量很大，实时性难以符合等。

二、按统计物理角度分类

（1）智能驾驶模型是交通流仿真中一个经典的跟驰模型，包括自由状态下的加速度趋势和考虑与前导车碰撞的减速趋势，人工智能模型中车辆下一刻加速度由当前速度和最大速度的偏差决定，通过将此时刻实际值与期望值做商来衡量两者之间偏差的相对值，并用幂计算来模拟驾驶员对这种偏差的敏感程度，最终实现跟驰加速度的确定，进而确定下一时刻车辆的速度和位置，所有的车辆遵循此规则即完成交通流的跟驰建模。

（2）最优速度模型解决了刺激反应模型在停车启动过程中加速度过大的问题，本质上是对刺激反应模型的拓展延伸。在诸多学者的大量的理论分析和数值模拟下，最优速度模型拓展了很多种类，主要包括考虑驾驶人的反应时间，复杂的交通信号，多个前导车跟驰等环境因素的影响。最优速度模型经常被应用到模拟宏观交通流的各种现象。

（3）元胞自动机模型是在一个离散的、有限状态的元胞组成的元胞空间里，以一定的局部规则在离散的时间维度上演化的动力学模型。因为交通元素从本质上来说是离散的，所以元胞自动机理论可以很好的模拟车辆的跟驰行为，具有独特的优越性。

**2.1.3 车辆跟驰模型的发展历程**

（1）20世纪50年代，通用汽车实验室在汽车跟驰模型理论上做了大量的实验并取得了众多的研究成果，车辆跟驰模型慢慢开始走进科研界的历史舞台。Chandler基于早期的刺激反应模型RV没有考虑安全间距容易发生碰撞的缺陷做了改良，引入了安全间距等因素，建立了跟驰车辆加速度与前后车相对速度之间关系的模型，这便是GM模型的原型。此后，Gazis对GM模型进行了积分，得到了宏观流量-速度的关系模型，使得跟驰模型在交通流微观模型和宏观理论之间建立起了联系。这一系列的研究成果使得GM模型成为了最重要、最经典的跟驰模型。在跟驰模型的微观交通流理论中占据着主导地位。

（2）20世纪60年代，随着认知心理学和视觉知觉理论的发展，许多的学者逐渐认识到跟驰行为并不能被看作纯粹的机械性精确过程，而应该更加关注驾驶人在驾驶过程中的感知、决策与处理过程。所以心理学的相关理论被引入到了跟驰模型中，驾驶人通过分析前导车寸大小的变化来感知车辆之间相对速度的变化，一旦感知超过了阀域，驾驶人就会做出相应的行为。这就是心理-生理类模型的雏形。生理-心理模型促进了交通工程领域与心理学之间的相互交叉融合。

（3）20世纪末，跟驰模型日益受到许多著名的统计物理学家的关注，他们尝试通过对微观跟驰行为的建模来展示宏观交通流中的诸多复杂交通情形的非线性特性，并用此来解释交通阻塞形成和消散的机理。在此基础上出现了广义力模型、智能驾驶模型、全速度差模型等。将跟驰理论的研究引入到了新领域。1999年，一篇有关跟驰模型的经典综述性文章的发表，引发了新一轮对跟驰行为建模的广泛关注和激烈讨论，这让许多的持有不同观点的交通心理和交通工程方面的学者对跟驰行为建模进行深刻的反思。

（4）21世纪以来，智能交通系统和现代检测技术发展迅速，所以研究智能交通系统下的跟驰模型就显得十分重要。这类模型更加注重研究车辆与车队之间的互动与协同建模，即“车路协同”。分析在环境的作用下跟驰行为对交通流特性的影响，有力的促进了微观交通仿真和辅助驾驶控制理论的研究。

**2.2 深度强化学习**

深度强化学习是将强化学习和深度学习结合在一起，用强化学习来定义问题和优化目标，用深度学习来解决状态表示、策略表示等问题。分别用深度神经网络来建模强化学习中的值函数、策略，然后用误差反向传播算法来优化目标函数。

**2.2.1强化学习简介**

强化学习模型如图2.1所示主要由智能体（Agent）、环境（Environment）、状态（State）、动作（Action）、奖励（Reward）组成。智能体执行了某个动作后，环境将会转换到一个新的状态，对于该新的状态环境会给出奖励信号（正奖励或者负奖励）。随后，智能体根据新的状态和环境反馈的奖励，按照一定的策略执行新的动作。上述过程为智能体和环境通过状态、动作、奖励进行交互的方式。智能体通过强化学习，可以知道自己在什么状态下，应该采取什么样的动作使得自身获得最大奖励。由于智能体与环境的交互方式与人类与环境的交互方式类似，可以认为强化学习是一套通用的学习框架，可用来解决通用人工智能的问题。因此强化学习也被称为通用人工智能的机器学习方法。

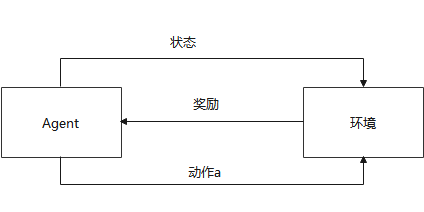


图2.1 强化学习模型

**2.2.1 马尔可夫决策过程**

马尔科夫决策过程由一个五元组（S, A, P, R, γ）来描述的，它是在马尔科夫奖励过程的基础上添加了动作集（A）而改进来的。在强化学习中，智能体与环境是通过执行动作来进行交互的。因此，需要添加动作集（A）进来，这样从一个状态转移到另一个状态就是因为执行了动作才会产生的。

其中：S为环境状态集，A为有限的动作集，P为状态转移概率，R为回报函数，γ为折扣因子，用来计算累计回报，随着状态的转移而逐渐衰减。

价值函数是折扣奖励总和的期望值，用公式（2.1）表示

（2.1）

强化学习的最终目标是找到最大化累计奖励期望的策略，用公式（2.2）表示

（2.2）

这样的话强化学习的关键就在于找到一个策略，使整个过程的累计奖励值最大。

**2.2.2 Q-learning**

Q-learning是解决马尔可夫决策过程的常用算法之一，Q-learning可以看作一种增量式动态规划，它通过直接优化一个可迭代计算的动作值函数Q(s, a) , 来找到一个策略使得期望折扣报酬总和最大的E(s)，其中E(s)=maxQ(s,a)。这样，智能体在每一次的迭代中都需要考察每一个行为，可确保学习过程收敛。其中Q(s, a)指状态s执行完动作a后希望获得的累积回报，它取决于当前的立即回报和期望的延时回报。

Q-learning算法用公式（2.3）来表示。

（2.3）

**2.2.3 Sarsa**

Sarsa算法也是解决马尔可夫决策过程的常用算法之一，Sarsa算法被称为改进的Q-learning算法。Sarsa与Q-learning的差别在于Q-learning采用的是值函数的最大值进行迭代, 而Sarsa则采用的是实际的Q值进行迭代。Sarsa 在每次学习过程中，智能体依据当前Q值确定下一状态时的动作；而Q-learning则依据修改后的Q值确定动作，因此称Sarsa是一种on-policy学习方式。Sarsa算法用公式（2.4）来表示。

（2.4）

**2.3 本章小结**

在车辆跟驰模型方面，对模型的概述、分类、发展历程三个方面进行了详细的介绍。在深度强化学习方面，对深度强化学习的概念进行了简单的介绍并对强化学习的相关算法和理论进行了深入的阐述。

**第3章 车辆跟驰模型的构建**

本文提出的车辆跟驰模型是通过将车辆编队来实现的，车辆编队的构建方法主要从车队的控制策略、车队的运动学模型、车队的稳定性三个方面进行阐述。

具体控制流程如图3.1所示。

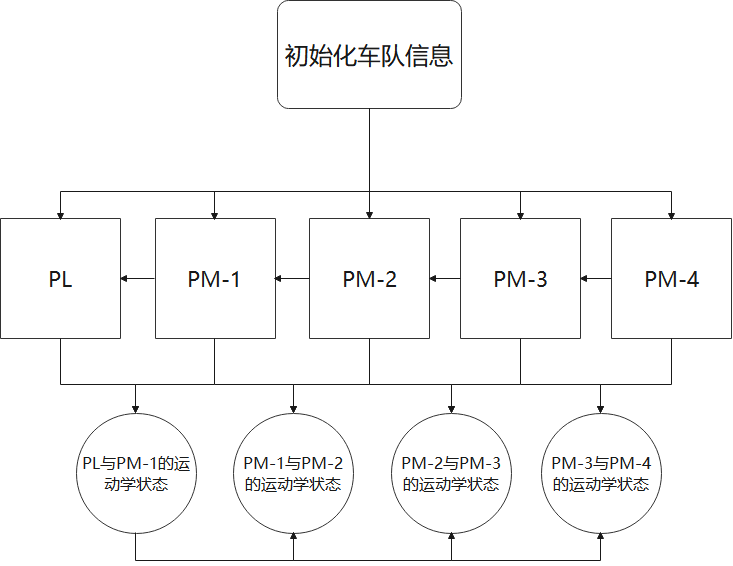


图3.1 车队的控制流程

本文的模型以5辆车建立车队，车队由头车PL和跟随车PM-i（0<i<5）组成。PL是车队的队头车辆，PM-i（0<i<5）是车队的其他车辆。通过控制PL的状态信息来调整PM-i（0<i<5）的运动状态并实时采集每辆车的状态信息，使PM-i（0<i<5）保持和PL相近似的运动状态以达到跟驰的目的。下面将详细阐述模型的构建过程。

**3.1 车队的控制策略**

**3.1.1 间距的控制策略**

车辆编队是一组共享共同移动模式的车辆，其中每个队成员车辆PM-i跟随队头车辆PL并保持与前一个PM-i-1的目标车辆间距。车辆间距控制策略一般有两种：定时距策略和定间距政略。如图3.2所示其目的都是为了得到两辆车之间的一个合理的期望间距。

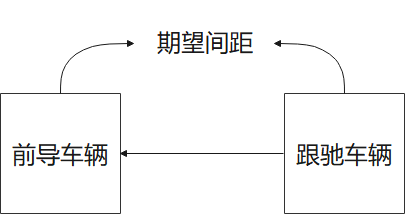


图3.2 间距的控制策略

（1）定时距策略是通过事先规定好的长度相等的时间块，综合考虑前导车辆和跟驰车辆的速度，一般采取两辆车速度差的绝对值和时间块相乘来计算期望间距。

（2）定间距策略是通过直接规定一个期望间距，无需考虑其他外在因素，让两辆车之间一直保持该间距。

定时距策略的期望车距与车辆速度有关。在定间距政策中，期望间距是一个恒定值并且与车辆速度无关，这会导致较高的通行能力。从提高通行能力的角度出发，我们在车队中实施定间距策略。

**3.1.2 队列的控制策略**

车辆的队列控制是一个非常重要的问题，也是一个具有挑战性的问题。在车队中，每个PM-i（0<i<5）应该通过控制其速度和加速度来保持预定义的驾驶模式。队列控制的另一个关键要求是实现车队的稳定性，即确保PL的状态改变不会导致队列中PM的间距沿着队列下游逐渐扩大。实现队列控制主要有三种控制策略：前驱跟随控制策略、对称双向控制策略、先导-前驱跟随控制策略。

（1）前驱跟随控制策略是以相邻的两辆车为控制单位，如图3.3所示跟驰车通过收集前导车的状态信息来调整自身的状态。

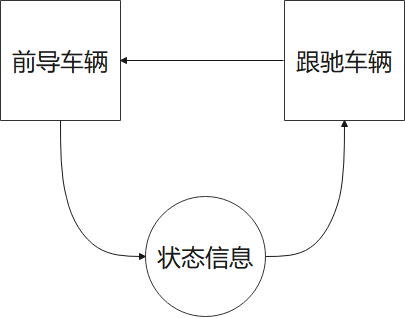


图3.3 前驱跟随控制策略

（2）对称双向控制策略是以跟驰车辆和其前后相邻车辆为控制单位，如图3.4所示跟驰车通过收集前后相邻车辆的状态信息来调整自身的状态。

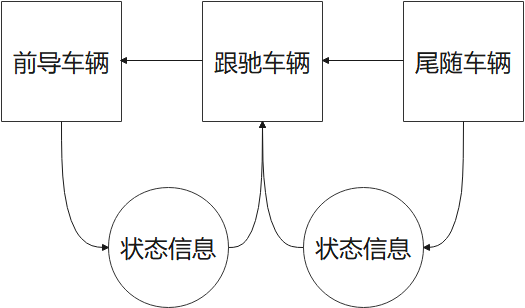


图3.4 对称双向控制策略

（2）先导-前驱跟随控制策略是以队头车辆PL和前导车辆为控制单位，如图3.5所示跟驰车通过收集队头车辆PL的状态信息和前导车辆的状态信息来调整自身的状态。

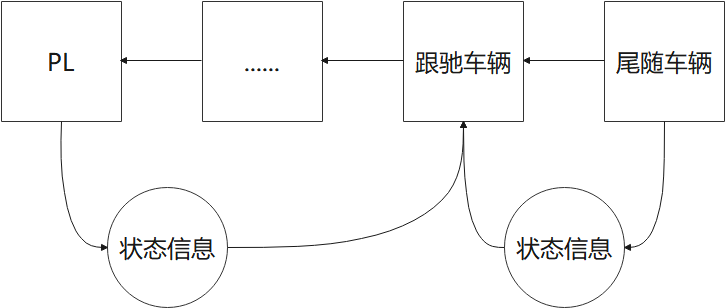


图3.5 先导-前驱跟随控制策略

在前驱跟随控制策略中，每个PM只收集前导车辆的信息来做出控制决策。然而，这种控制策略可能会由于沿着车队下游车辆的间距误差放大而导致车队的不稳定性。在对称双向控制策略中，PM的控制动作取决于前后相邻PM的运动学状态。随着车队规模的增加，这种控制架构会导致更大的车辆间距误差。先导-前驱跟随控制策略的控制动作由队头车辆PL和前导车辆共同决定，可以在避免车队的不稳定性的同时不至于导致较大的车辆间距误差。所以本文采取先导-前驱跟随控制策略来进行队列控制。

**3.2 车队的运动学模型**

车辆的动力学通常是复杂和非线性的，包括发动机、传动系统、制动系统、空气阻力、轮胎摩擦等多种因素的影响。所以应该使用三个合理的假设来获得一个简洁的模型分析：

（1）由于车队是由相同类型的车辆组成，我们假设车辆动力学彼此相同。

（2）车队在直线车道内移动，在高速公路场景中不采取变道行动。

（3）队内各车辆的加减速可控。

第i个(0<=i<5)车辆的位置、速度分别用zi、vi来表示，其中第0辆车代表PL其他车辆是PM。用向量xi = [zi, vi]表示第i个车辆的运动学状态。每辆车的长度相同，记为L。因此，可以用公式（3.1）来描述车辆的运动学状态。

（3.1）

其中表示PM的指令加速度，即加油门和踩刹车两种动作。由于车辆的推力和制动能力通常是有限的，对控制输入有（3.2）的约束条件。

（3.2）

在定间距策略下，每个PM需要跟踪其前一个PM的轨迹。因此，为了评估每台PM的队列控制性能，应该引入几个重要的指标。 首先，PMi（0<i<5）相对于前一个PMi-1的间距误差用公式（3.3）来定义。

（3.3）

其中常数为目标车辆间距，此外PMi相对于前一个车辆PMi-1的速度误差用公式（3.4）来定义。

（3.4）

为了保持期望车距，队列控制的目标可以表示为使每个PM的跟踪误差尽可能小。PM的跟踪误差用公式（3.5）表示。

（3.5）

在车队系统中，通过定期收集相关车辆的运动学状态来控制每个PM的纵向位置。此外，连续时间的车队系统是用采样时间Δt来离散化的，Δt为相邻两次采样的时间间隔。

**3.3 车队的稳定性**

**3.3.1 稳定性的定义**

在车队系统中，非常有必要考虑当前控制输入对车队后续车辆的运动状态的影响。针对这个问题，我们提出了稳定性来规范车队的长期行为。稳定性是指PL的状态改变不会导致队列中PM的间距沿着队列下游逐渐扩大，即PMi的间距误差|δi|不应大于间距其前一个PMi-1的间距误差|δi-1|。车队的稳定性用公式（3.6）来定义。

（3.6）

如果,则认为车队处于稳定状态。对于任何一个时间间隙Δt，由于实际原因，稳定性对于存在许多不确定因素的车队系统至关重要。

**3.3.2 公式化问题**

如前文所述，采用将车辆编队的方式会影响跟驰模型的性能。因此，我们需要定义性能指标来衡量模型的性能。本文旨在通过最小化每个PM的跟踪误差（即间距、速度误差）来保持恒定的车辆间距。

首先，PMi的目标函数考虑了它的跟踪误差，用公式（3.7）来表示。

（3.7）

其中，正对角矩阵Q为跟踪误差的权重。矩阵Q定义了间距误差和速度误差的权重。

根据各个PM的目标函数，车队的目标函数可以定义为公式（3.8）。

（3.8）

**3.4 本章小结**

首先对如何进行车辆编队进行了分析，介绍了车队的组成部分，以及如何通过队头车辆PL来控制其他车辆PM，并给出了模型的控制流程。然后，介绍了车队的控制策略包括车距控制和队列控制。本文在车距控制上选择定间距策略，在队列控制上选择先导-前驱跟随控制策略。之后，介绍了车队的运动学模型并给出了模型的运动学公式。最后，添加了车队的稳定性因素，来规范车队的长期行为。

**第4章 基于深度强化学习的车辆跟驰算法设计**

强化学习是一种基于马尔可夫决策过程的决策算法，它通过与环境进行交互式学习，得到设定奖励函数下的最优决策。深度强化学习算法是通过结合深度神经网络的特征提取能力和强化学习的决策能力。本章我们将使用两种深度强化学习算法DQN和DDPG对车辆跟驰问题进行描述。在引入深度强化学习后车队的控制流程如图4.1所示。

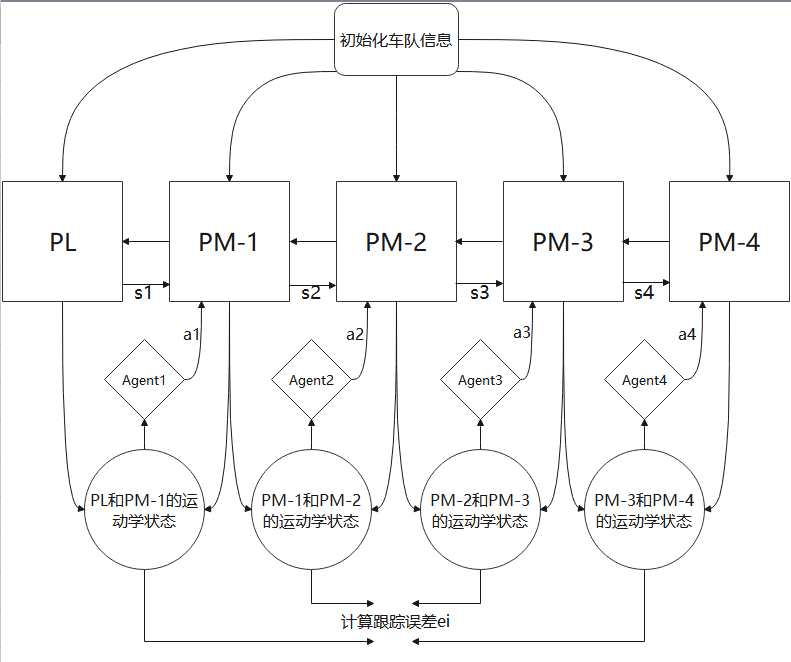


图4.1 车队控制流程

**4.1 Deep Q Network**

DQN算法全称为Deep Q Network，是强化学习有力的工具之一，本小节将针对Q-learning的局限性和使用DQN算法进行跟驰控制两个方面对DQN算法进行详细的阐述。

**4.1.1 Q-learning的局限性**

如表4.1所示Q-learning算法是使用Q Table来存储记忆库每个记忆由当前状态、下一个状态、Q值和动作组成。Q-learning的算法流程图如图4.2所示。

车辆跟驰是一个复杂的问题，状态空间十分庞大。如果全部使用Q Table来存储，对计算机的内存性能是一个很大的挑战，而且每次在这么大的表格中搜索对应的状态也是一件很耗时的事情。不过，在机器学习中有一种方法对这种事情很在行，那就是神经网络。我们可以将状态和动作当成神经网络的输入，然后经过神经网络分析后得到动作的Q 值, 这样我们就没必要在表格中记录Q值，而是直接使用神经网络生成Q值。然后按照Q-learning的方法, 直接选择拥有最大值的动作当做下一步要做的动作。DQN的算法流程图如图4.3所示。

表 4.1 Q Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Q Table | | 动作集 | | | |
| 动作1 | 动作2 | 动作3 | 动作4 |
| 状态id | 状态特征 | 左 | 右 | 上 | 下 |
| 1 | S1 |  |  |  |  |
| 2 | S2 |  |  |  |  |
| 3 | S3 |  |  |  |  |
| 4 | S4 |  |  |  |  |
| 5 | S5 |  |  |  |  |
| 6 | S6 |  |  |  |  |
| 7 | S7 |  |  |  |  |
| 8 | S8 |  |  |  |  |
| 9 | S9 |  |  |  |  |
| 10 | S10 |  |  |  |  |
| 11 | S11 |  |  |  |  |
| 12 | S12 |  |  |  |  |
| 13 | S13 |  |  |  |  |
| 14 | S14 |  |  |  |  |
| 15 | S15 |  |  |  |  |
|  | | | | | |

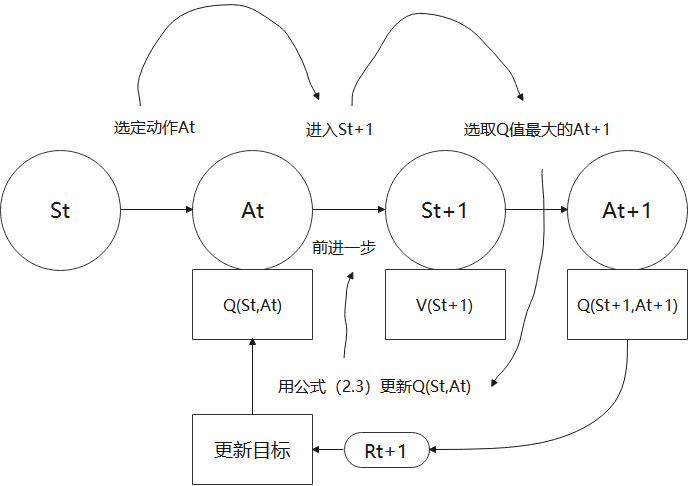


图4.2 Q-learning流程图

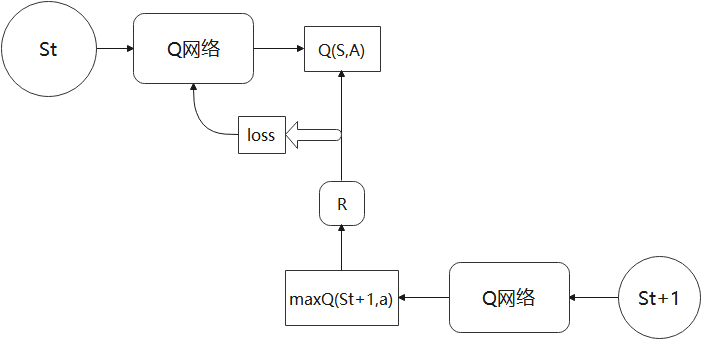


图4.3 DQN流程图

**4.1.2 Double DQN**

DQN有一个显著的问题，就是DQN估计的Q值往往会偏大。这是由于我们的Q值是以一下一个状态的Q值得最大值来估算的，但是下一个状态的Q值也是一个估算值，也依赖于他的一下一个状态的Q值。这就导致了Q值往往会有偏大的情况出现。

Double DQN采用两个神经网络来解决上述问题。一个网络叫eveluate network，用来估算当前状态的Q值，另一个网络叫target network，用来估计下一个状态的Q值。算法的流程图如图4.4所示。

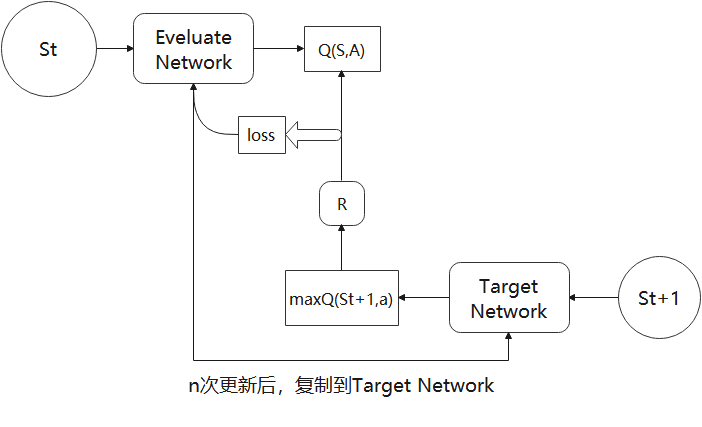


图4.3 Double DQN流程图

**4.1.3 基于DQN算法的跟驰控制建模**

由于本文采取将车辆编队来进行跟驰建模，所以车辆跟驰控制是一个设计跟踪性、稳定性等目标在内的多目标优化问题。基于车辆的跟踪误差来设计目标函数。我们根据车辆真实运动的实际情况设定了公式（3.7）中Q矩阵的权重系数。

将间距误差的权重Qz定义为100，速度误差的权重Qv定义为5。所以公式（3.7）可以用公式（3.8）来表示。

（4.1）

强化学习最重要的是定义合理的状态空间、动作和奖励，接下来我们将定义跟驰场景的状态空间、动作和奖励。

（1）状态空间：由于车辆跟驰是在一条车道上进行的，而且我们假定在高速路的环境下车辆是做直线行驶的。所以状态空间包括车辆自身的速度、位置两个因素。

（2）动作空间：在现实生活中，驾驶人都是通过刹车和油门来改变车辆的运动状态。这两个动作最直接影响的就是车辆的加速度，所以我们以刹车、油门以及不采取任何行动这三个动作作为动作空间。由于，在车辆的运动学模型中我们定义了加速度的约束（3.2），所以刹车和油门两个动作所对应的加速度值不能超过其最大值和最小值。

（3）奖励函数：为了保证车队的稳定性、跟踪误差，本文设定了稳定性指标、跟踪性指标、惩罚项三个性能指标来表示车辆跟驰的奖励函数。

稳定性是衡量车辆通行效率的一个重要指标，在定间距策略下，良好的稳定性可以实现较高的通行效率。我们用公式（4.2）来表示。

（4.2）

跟踪误差是衡量车辆通行效率的另一个重要指标，最小化跟踪误差是跟驰控制的最终目标。我们用公式（4.3）来表示。

（4.3）

为了保证车队中的车辆不会发生碰撞，设置了惩罚项来进行约束。我们用公式（4.4）来表示。

（4.4）

在建立跟驰模型的时候，将前后相邻的两辆车作为一个控制单位，所以这里我们将其作为强化学习的单个智能体。单个智能体的总体奖励函数用公式（4.5）来表示。

（4.5）

由于车队中的所有车辆都需要进行控制，所以车辆跟驰模型是一个多智能体的强化学习。我们用公式（4.6）来表示车队的奖励函数。

（4.6）

**4.2 Deep Deterministic Policy Gradient**

我们知道Q-learning只能处理离散的状态和动作。DQN算法是基于Q-learning进行改进的，解决了Q-learning不能处理连续状态空间的问题。但是，DQN算法继续沿用了Q-learning的算法公式（2.3）,公式中的只能处理离散的动作，所以DQN算法也无法处理连续的动作空间。而在现实生活中车辆行驶中的加速度变化显然是一个连续的动作。所以本节提出一种可以处理连续动作空间的强化学习方法DDPG。

**4.2.1 Policy Gradient**

强化学习是一个通过奖惩机制来学习正确行为的方法。家族中有很多不同的成员，其中由学习奖惩值根据价值选择行为的。比如Q-learning，DQN，也有不通过分析奖惩值直接输出行为的方法。比如Policy Gradient。我们可以给Policy Gradient加上一个神经网络来输出预测的动作，对比以值为基础的方法，Policy Gradient直接输出动作的最大好处就是可以在一个连续的区间内挑选动作。

**4.2.2 Actor Critic**

Policy Gradient方法可以处理连续的动作空间，但是Policy Gradient是回合更新的，相比于Q-learning的单步更新，Policy Gradient降低了学习效率。所以我们使用了一种将两者整合起来的方法Actor Critic。Actor Critic对应的是两个神经网络。Actor的作用是在连续的动作空间中选取合适的动作，所以Actor使用的是Policy Gradients，Critic的作用是对Actor选出来的动作进行评价，是以值为基础的学习法，所以使用的是Q-learning。

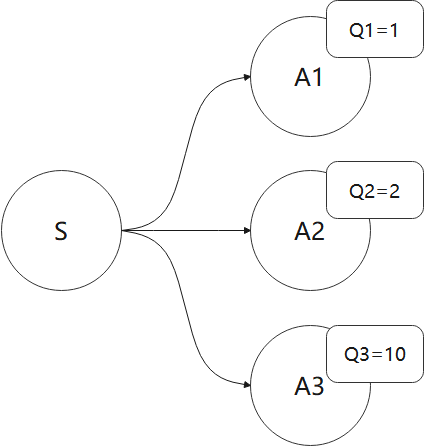


图4.5 基于Q值的TD算法

TD-error是Actor Critir的重点。这里的TD-error和DQN中的TD-error是不同的。在DQN中预估的是Q值，而在Actor Critic中估算的是V值。如图4.5所示，如果我们使用神经网络来估算Q值作为更新值，在开始的时候，我们采用平均策略，假定随机到了A1。于是我们用策略梯度的带权重方法来更新策略，这里的权重就是Q值。策略会更倾向于选择A1，意味着更大概率选择A1。这就掉进了正数陷阱。我们明明希望A3能够获得更多的机会，最后却是A1获得最多的机会。为了解决这个问题，我们可以通过减去一个平均值，使得到的权重有正的也有负的。如图4.6所示。

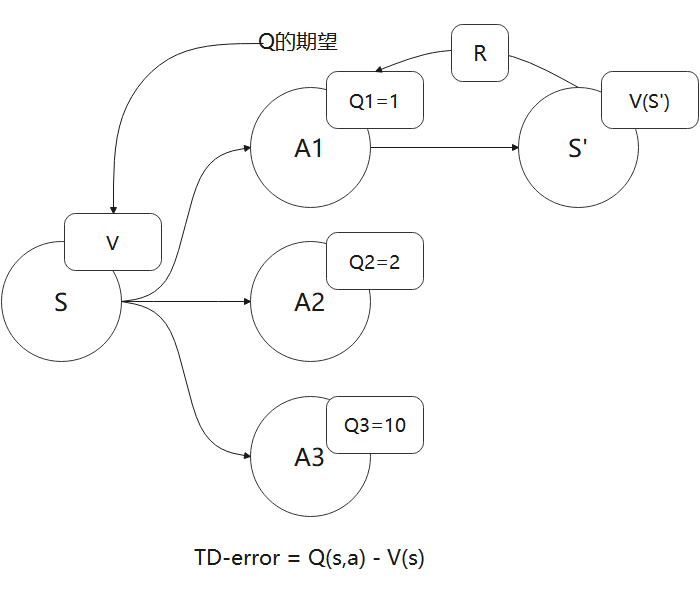


图4.6 基于V值的TD算法

Actor Critic的算法流程如图4.6所示。

（1）定义两个network：Actor 和 Critic。

（2）从状态s开始，执行动作a，得到奖励r，进入状态s'。

（3）把记录的数据输入到Critic网络，求TD-error，并缩小TD-error。

（4）把Critic网络得到的结果输入到Actor，并计算策略分布。

（5）更新状态s，从（2）开始重新迭代直到回合结束。

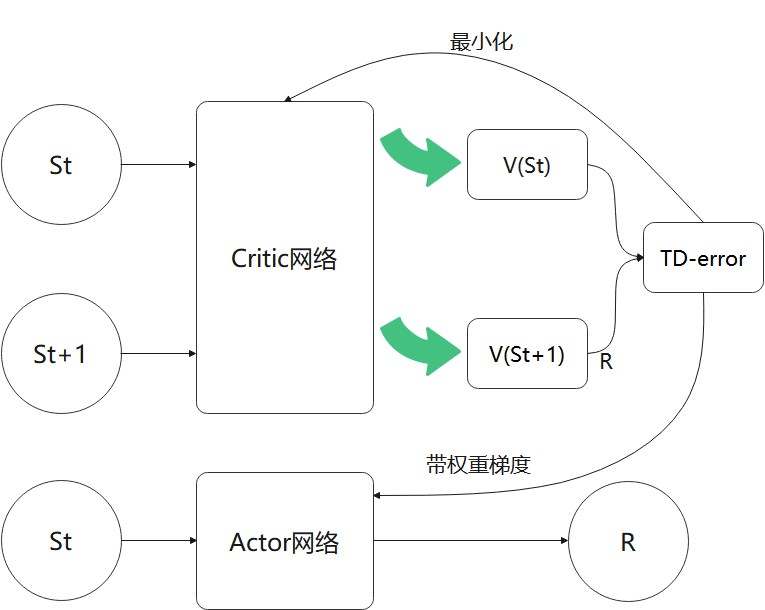


图4.7 Actor Critic算法流程图

**4.2.3 基于DDPG算法的跟驰控制建模**

DDPG算法是一种基于Actor Critic方法，也可以处理连续的动作空间。所以需要对用DQN算法进行跟驰控制的动作空间进行修改。由于车辆的刹车和油门两个动作是有约束限制的，所以我们的动作空间定义为。关于状态空间和奖励函数本节沿用DQN算法。关于DDPG算法的神经网络的更新本节沿用Double DQN算法，采用双神经网络进行更新。DDPG算法的流程图如图4.8所示。

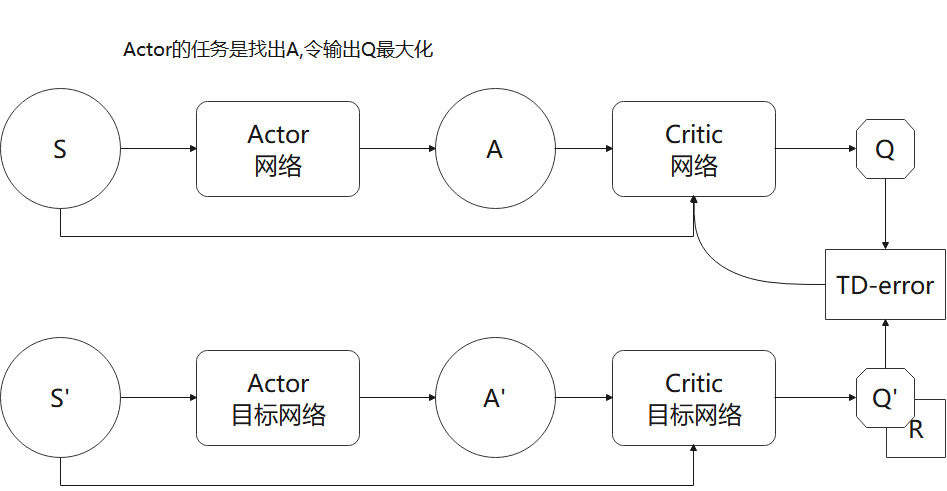


图4.7 DDPG算法流程图

**4.3 本章小结**

首先对DQN算法进行了详细的介绍，然后基于DQN算法定义了车辆跟驰的状态空间、动作空间、奖励函数。然后针对DQN算法的局限性引出了DDPG算法，并使用DDPG算法重新定义了车辆跟驰的三要素。最后基于Double DQN的思想对DDPG算法进行了改进。

**第5章 实验验证**

本章将针对本文提出的模型进行仿真实验，设置了车队的累计回报、车辆的跟驰间距和跟驰速度、车辆的期望间距和实际间距三个性能指标。对DQN算法和DDPG算法的控制效果进行了对比分析。

**5.1 驾驶数据采集**

本文中的车辆行驶数据来自计算机的模拟场景，设置仿真步长Δt，即相邻两次采样时间的间隔。基于公式（3.1）可以得到每辆车辆速度位置随时间的变化关系，用公式（5.1）表示。

（5.1）

**5.2 DQN算法和DDPG算法的回报曲线对比**

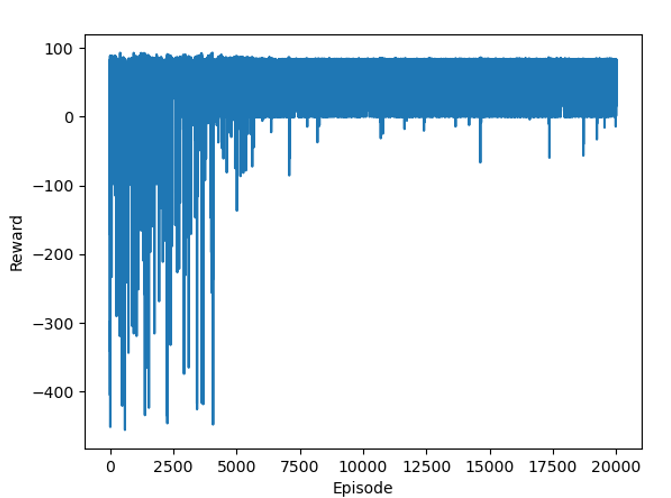


图5.1 20000次训练的DQN回报曲线

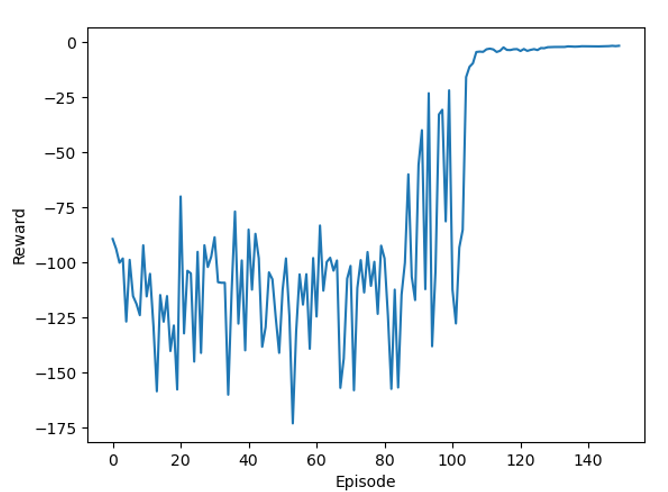


图5.2 150次训练的DDPG回报曲线

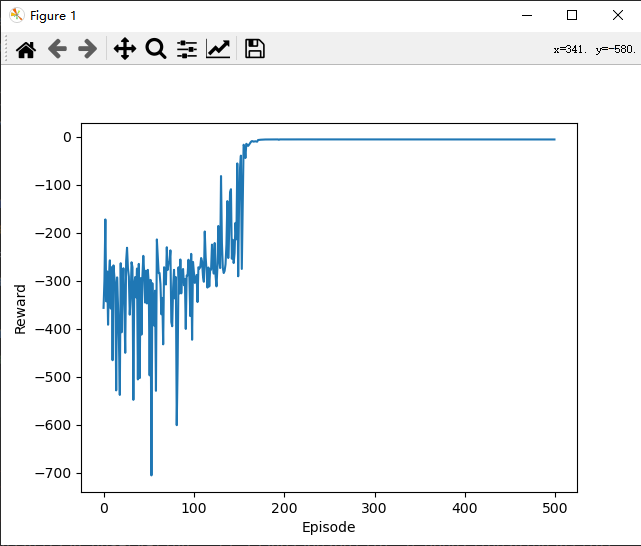


图5.3 500次训练的DDPG回报曲线

从图5.1可以看出基于DQN算法的跟驰模型的回报曲线最终是一种振荡的收敛，训练效果并不理想，而从图5.2和图5.3中可以看出基于DDPG算法的跟驰模型的回报曲线收敛于一个稳定值。说明DDPG算法更适合于车辆的跟驰控制。

**5.2 DDPG算法的车辆的跟驰间距和跟驰速度**

本小节进行了两组实验，分别针对主车PL匀速运动和变速运动进行实验，观察跟驰车的速度和位置变化。通过这两个指标测试模型的性能。

**5.2.1 PL匀速运动**

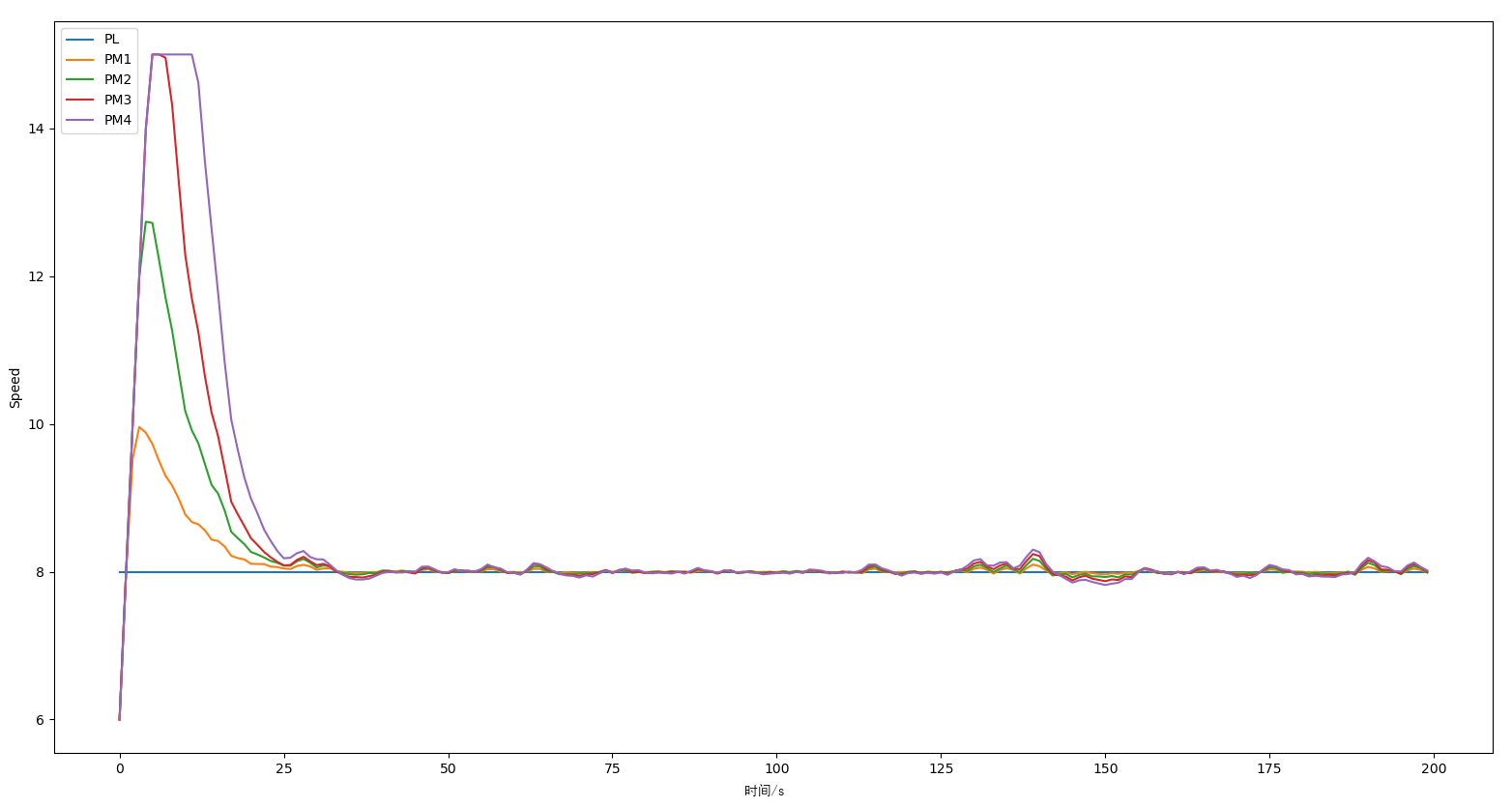


图5.4 PL匀速时车队的速度变化曲线

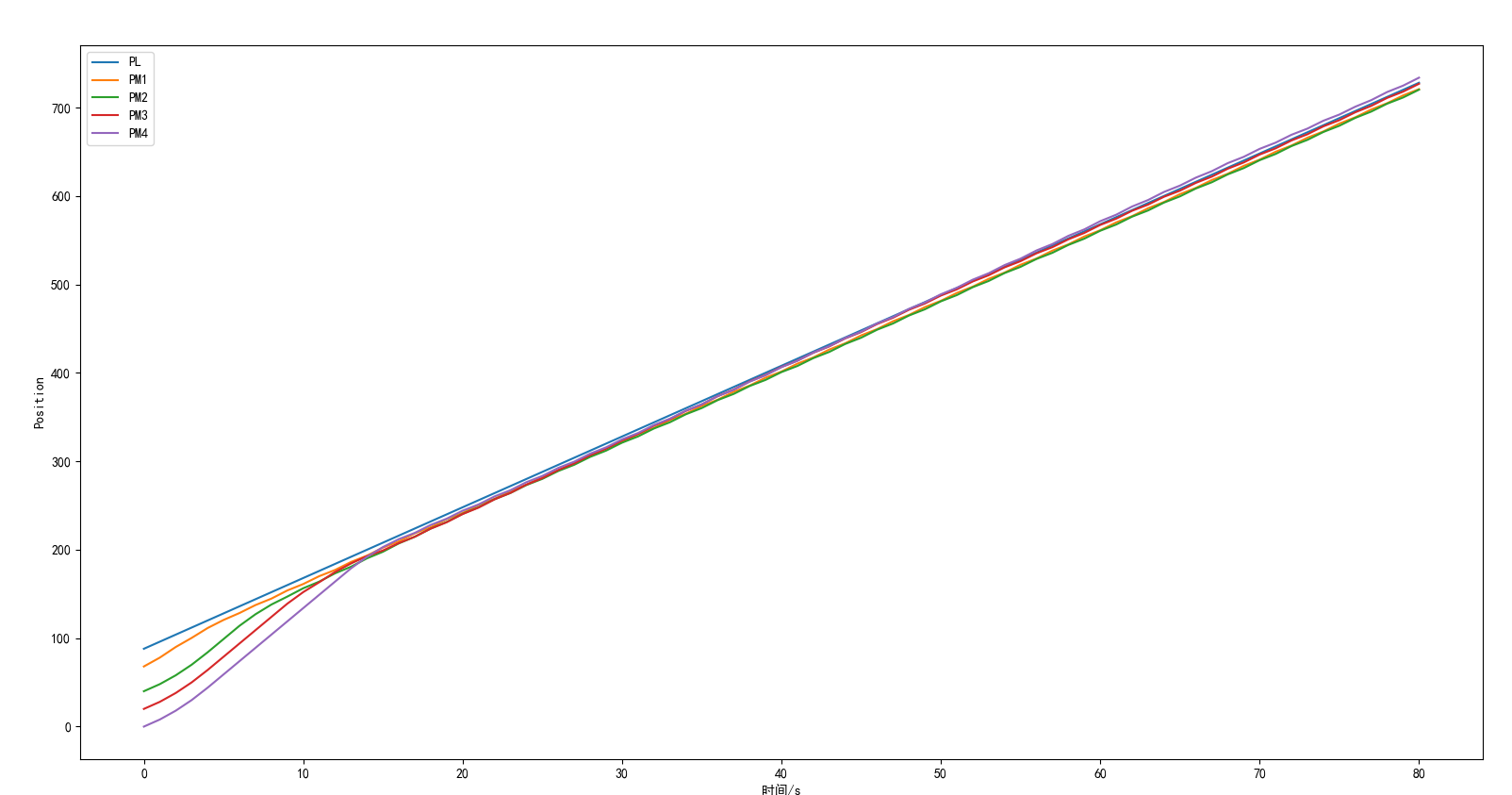


图5.5 PL匀速时车队的位置变化曲线

**5.2.1 PL变速运动**

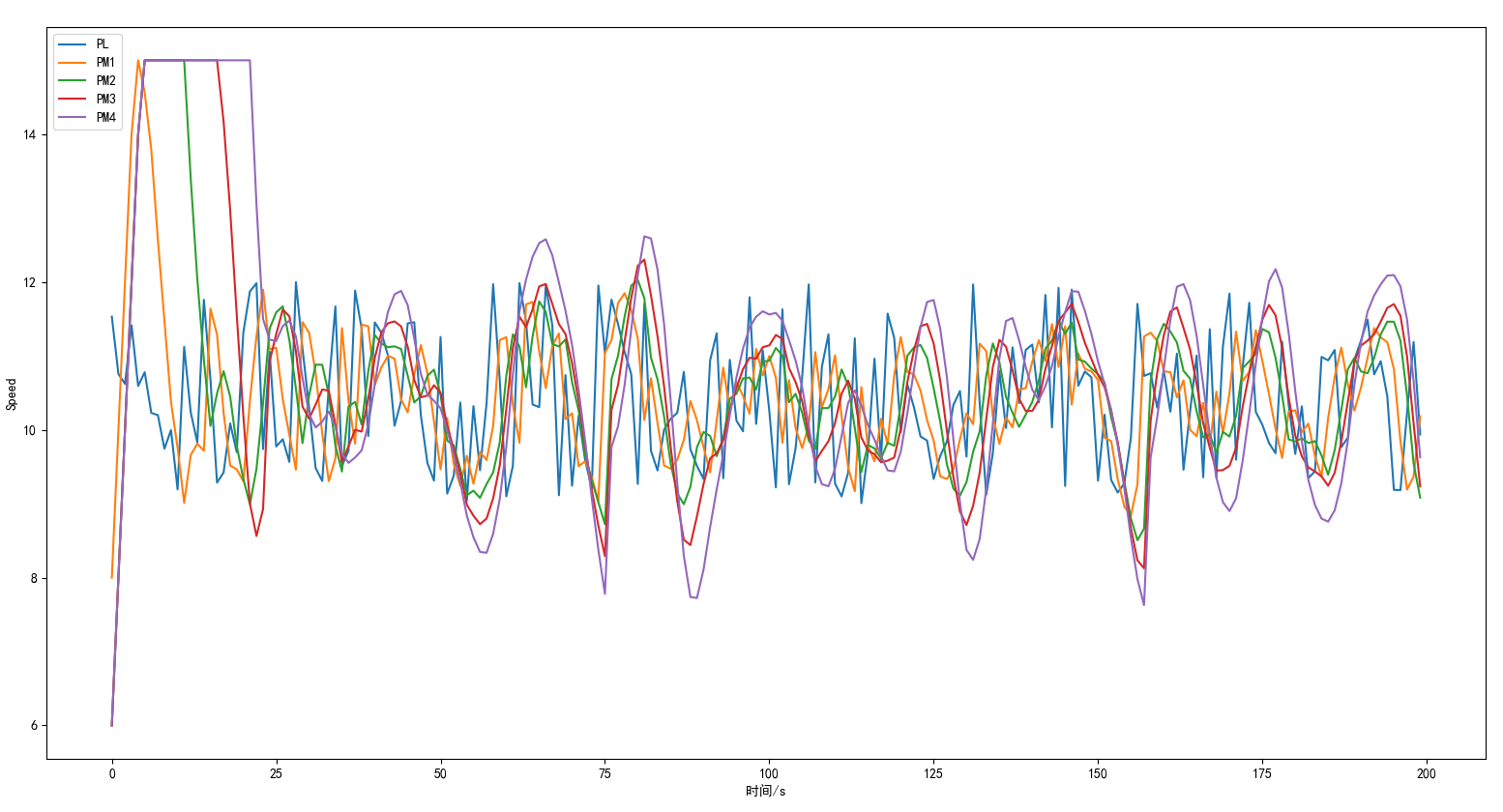


图5.5 PL变速时车队的速度变化曲线

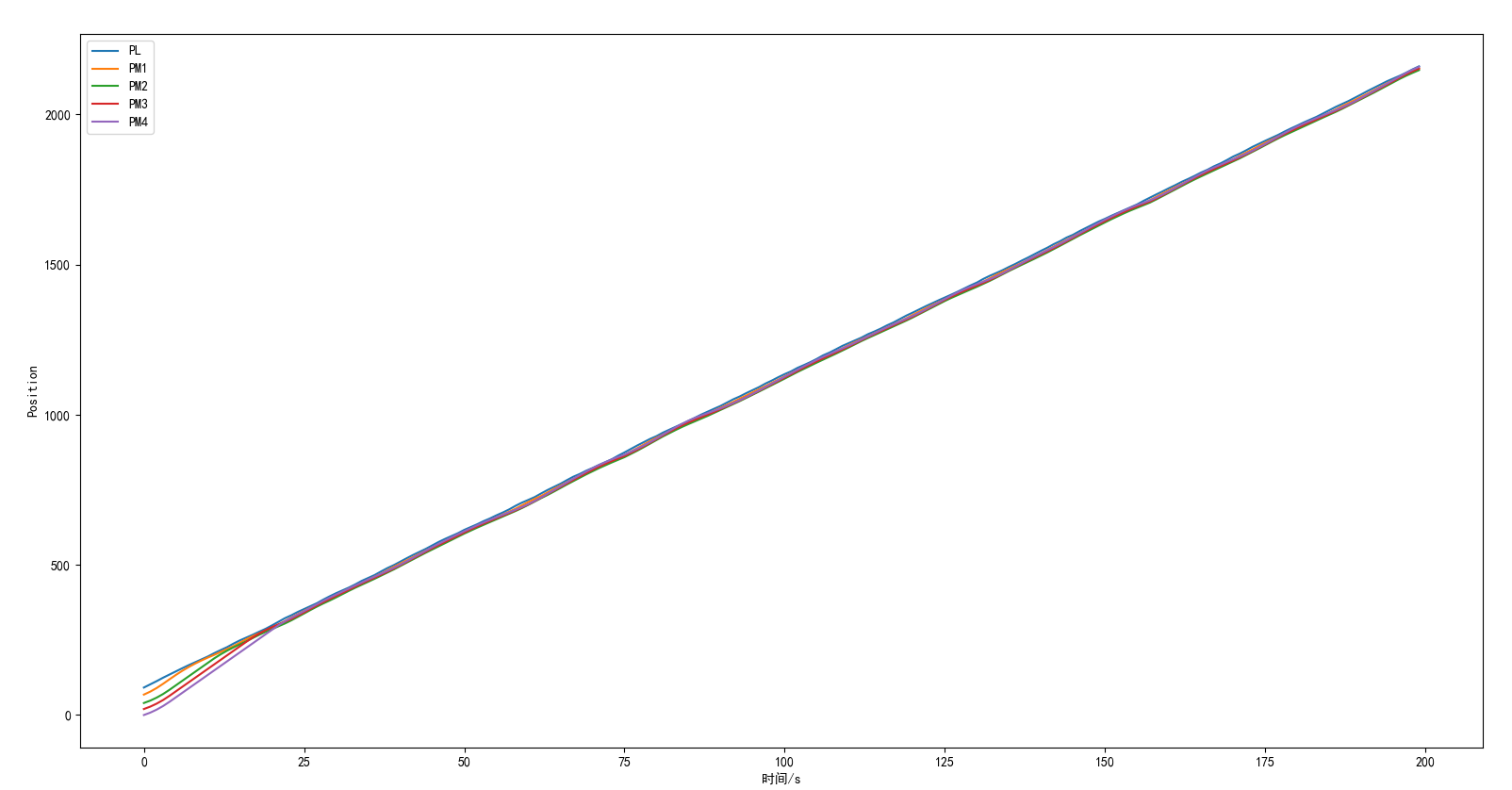


图5.6 PL变速时车队的位置变化曲线

参考文献 (黑体小二)

[1] 钱宁, 万兆惠. 泥沙运动力学[M]. 北京：科学出版社, 2005.

[2] 唐洪武, 闫静,吕升奇. 河流管理中含植物水流问题研究进展[J]. 水科学进展，2007,18(5): 785-792.

[3] Kouwen N, Unny T E, Hill H M. Flow Retardance on vegetated channels [J]. Journal of the Irrigation and Drainage Division,1969,95(IR2):329-342.

[4] Gourlay M R. Discussion of Flow Resistance in vegetated channels by Kouwen etal. [J]. Journal of the Irrigation and Drainage Division,1970,96(IR3):351-357.

(中文用宋体小四，数字及字母用Time New Roman小四，1.5倍行距)

参考致谢示例：

致谢 (黑体小二)

本论文是在\*\*\*老师的悉心指导下完成的，感谢\*\*\*老师在论文的撰写中给予的细心的指导与帮助。老师那严谨的治学态度、精益求精的科研作风给我们树立了很好的榜样，这将时刻激励我在工作与学习中继续进步。

感谢朝夕相处的同学，、、、

感谢、、、、

感谢同小组的成员，感谢他们在毕业设计过程中给予的鼓励和帮助，他们对于学术上严谨的态度和不懈的追求令我难以忘怀，在此向他们表示衷心的感谢！