博弈论期末论文

基于平均场博弈的自动驾驶决策研究

姓　　名：徐国洪

学　　号：124032910146

2024 年 11 月 28 日

Final Thesis on Game Theory

RESEARCH ON AUTONOMOUS DRIVING DECISION MAKING BASED ON MEAN FIELD GAME THEORY

Author: Xu Guohong

StudentID：124032910146

摘 要

本文旨在研究基于平均场博弈理论的自动驾驶车辆决策优化问题。平均场博弈（Mean Field Games, MFG）作为一种数学框架，能够有效模拟多智能体系统中个体行为与群体动态的相互作用，对于自动驾驶领域中的复杂交通环境决策问题具有重要的应用价值。本研究首先建立了车辆动态模型，定义了状态变量和控制变量，并构建了个体优化问题，以最小化包括旅行时间、能耗和安全距离在内的运行成本。进一步地，通过宏观分布函数描述群体行为，并利用 Hamilton-Jacobi-Bellman (HJB) 方程量化个体与群体间的相互作用，为自动驾驶车辆提供了考虑群体效应的最优控制策略。

通过数值方法求解 HJB 方程，并结合 MFGLib 库进行模拟实验，验证了模型的有效性。实验结果表明，所提出的模型能够优化自动驾驶车辆在复杂交通环境中的决策过程，提高了决策的效率和安全性。本研究不仅为自动驾驶车辆在面对复杂交通环境时的决策提供了理论依据和实践指导，也为自动驾驶技术的发展和普及提供了新的视角。未来的研究将进一步探索更复杂的交通场景和环境因素，优化数值求解方法，并结合人工智能和机器学习技术，以提高自动驾驶车辆的决策智能化水平。

关键词：平均场博弈，自动驾驶，决策优化，多智能体系统

Abstract

This article aims to study the decision optimization problem of autonomous driving vehicles based on mean field game theory. Mean Field Games (MFG), as a mathematical framework, can effectively simulate the interaction between individual behavior and group dynamics in multi-agent systems, and has important application value for complex traffic environment decision-making problems in the field of autonomous driving. This study first established a vehicle dynamic model, defined state variables and control variables, and constructed an individual optimization problem to minimize operating costs including travel time, energy consumption, and safety distance. Furthermore, by describing group behavior through macroscopic distribution functions and quantifying the interactions between individuals and groups using the Hamilton Jacobi Bellman (HJB) equation, an optimal control strategy considering group effects is provided for autonomous vehicles.

The HJB equation was solved using numerical methods, and simulation experiments were conducted in conjunction with the MFGLib library to verify the effectiveness of the model. The experimental results show that the proposed model can optimize the decision-making process of autonomous vehicles in complex traffic environments, improving the efficiency and safety of decision-making. This study not only provides theoretical basis and practical guidance for the decision-making of autonomous vehicles in complex traffic environments, but also offers a new perspective for the development and popularization of autonomous driving technology. Future research will further explore more complex traffic scenarios and environmental factors, optimize numerical solution methods, and combine artificial intelligence and machine learning technologies to improve the decision-making intelligence level of autonomous vehicles.

Key words: Mean field game theory, autonomous driving, decision optimization, multi-agent systems

目　录

第 1 章 绪论

1.1 研究意义

近年来，随着科技水平与自动化水平的不断提高，自动驾驶也逐渐成为智能交通系统的研究热点之一，人们对自动驾驶车辆的研究也越来越深入，自动驾驶主要包括三个部分，分别是感知，决策，规划控制，在自动驾驶汽车领域，困难场景如恶劣天气、复杂交通环境、紧急情况等对车辆的感知、决策和控制提出了更高的要求。

本文将博弈论引入自动驾驶决策规划问题，将驾驶决策问题转化一个博弈问题。博弈论可以描绘车辆间的相互作用的决策行为选择，被认为是很有前途的模拟人类决策和建模驾驶决策的方法之一[1]。

其实际意义在于能提高自动驾驶汽车的安全性，在困难场景下，如雨雪天气、大雾、夜间等，自动驾驶汽车的感知系统可能会受到限制，导致决策失误。通过博弈论方法，我们可以模拟和优化车辆在这些情况下的决策过程，从而提高其安全性；增强自动驾驶汽车的适应性，困难场景往往伴随着高度的不确定性和复杂性。博弈论可以帮助自动驾驶汽车在这些场景下做出更加灵活和适应性强的决策，提高其在各种环境下的行驶能力；促进自动驾驶技术的发展和普及，通过在困难场景下提高自动驾驶汽车的性能，可以增强公众对自动驾驶技术的信任，推动其在更广泛的应用场景中的接受和使用。

其理论意义在于可以丰富博弈论在自动驾驶领域的应用，博弈论在经济学、社会学等领域已有广泛应用，但在自动驾驶领域的应用还相对较少；推动自动驾驶决策理论的发展，自动驾驶汽车的决策过程涉及多车辆、多目标的复杂交互。博弈论提供了一种有效的框架来分析和优化这些交互，有助于推动自动驾驶决策理论的发展。

1.2 国内外研究现状

在人机共享驾驶博弈控制的研究中，Na X 等[1] [1]提出了基于博弈论的算法，比较了分散控制、不合作纳什算法、不合作 Stackelberg 算法与合作 Pareto 算法。这些算法通过预测驾驶员反应来优化自动驾驶汽车的避障控制和轨迹跟踪。Flad M 等[1] [1]进一步提出了“合作共享控制”概念，将驾驶员和自动控制系统的

交互视为纳什平衡问题，以寻求满足设计者需求的最优解。Wand JX 等[1]在网联车辆协同路径规划中，将路径规划问题转化为博弈问题求解。Zhang KR 等[1]分析了人−车系统的动力学特性，并将路径跟踪问题转化为博弈问题求解。

而在换道控制研究中，利用博弈模型对车辆变道决策进行机理性分析，可对特定运行状态下，车辆变道成功与否的概率进巧预测。薛春铭等[1] [1] 针对车辆换道行为，提出了包含车辆协同机制的换道策略。他们将换道视为非合作博弈行为，并根据换道行为特点建立了支付函数，该函数依赖于安全前提下可能获得的加速空间以及后车对前车的影响。此外，策略使用收益再调整方案，通过量化分析车辆的行驶风格，赋予可能影响换道的车辆不同权重参与收益计算。在 SUMO 仿真软件上的测试结果显示，使用博弈换道模型的车辆在不同车流密度环境下平均通行数量提高 5.6%，平均通行时间减少 8.4%，未发生事故，表明模型具有感知、判断人类驾驶风格的能力，可在保证安全的前提下提高通行效率。Kita[1]等提出基于博弈的并道-让路交互模型，认为待换道车辆与目标车道车辆之间是相互影响的关系。他们使用两人非合作博弈建模，以避免碰撞时间为基础建立博弈收益，通过寻找博弈模型的 Nash 均衡得到车辆的最优策略计划采用的博弈模型。Talebpour[1]等基于非零和博弈对强连通环境下的换道行为进行建模，重点关注了车联网强通信环境下信息流对车辆换道决策的作用。他们提出了一个简单版本的服务框架及一种基于模拟矩量法标定的校正方法，使用真实车流数据校正，所提出的框架具备预测换道行为的能力。

第 2 章 平均场博弈建模

2.1 车辆动态模型建立

2.1.1 变量定义

1.在自动驾驶车辆的决策优化模型中，状态变量是描述车辆在任意时刻物理状态的关键参数。这些变量包括：

- **位置** ($x(t)$)：车辆在道路上的位置，通常表示为一维或二维坐标。
- **速度** ($v(t)$)：车辆的速度，即位置随时间的变化率。
- **加速度** ($a(t)$)：车辆的加速度，即速度随时间的变化率。
- **方向** ($\theta(t)$)：车辆的行驶方向，通常以弧度表示。
- **安全距离** ($d_s(t)$)：车辆与其前方车辆之间的距离，以确保安全行驶。
- **相对速度** ($v_r(t)$)：车辆与其前方车辆之间的速度差。

2.控制变量是车辆可以调整以优化其行为的输入，这些通常包括：

- **加速度 ($u_a(t)$)**：车辆的加速度控制输入。
- **转向角度 ($u_\theta(t)$)**：车辆的转向控制输入，影响车辆的方向。

3.为了建立车辆动态模型，需要定义状态变量如何根据控制变量随时间演化。通过以下微分方程组来描述：

位置方程：$\dfrac{dx(t)}{t} = v(t)$

速度方程：$\dfrac{dv(t)}{dt} = u_a(t)$

方向方程：$\dfrac{d\theta(t)}{dt} = w(t)$

2.1.2 个体优化问题的构建

在自动驾驶车辆的决策过程中，个体优化问题的核心在于最小化车辆的运行成本。这一成本可能涵盖多个方面，包括但不限于旅行时间、能耗以及与安全相关的距离考

量。为了实现这一目标，我们首先定义了成本函数，它综合了状态变量和控制变量的影响。

成本函数J可以定义为状态变量 $x(t),v(t),\theta(t),d_s(t)$和控制变量 $u_a(t)$，$u_\theta(t)$ 的函数：

$$J(x,v,\theta,d_s,u_a,u_\theta)=\int_{t_0}^{t_f}L(x(t),v(t),\theta(t),d_s(t),u_a(t),u_\theta(t))dt+\Phi(x(t_f)) \quad (1)$$

其中L是拉格朗日函数，表示在每个瞬间的成本，Φ是终端成本，表示到达目的地时的成本，t_0和t_f分别是初始和最终时间。

拉格朗日函数L进一步分解为多个组成部分，以反映不同的成本因素：

$$L(x,v,\theta,d_s,u_a,u_\theta)=L_T(v,u_a)+L_E(u_a)+L_s(d_s) \quad (2)$$

L_T是**旅行时间成本**，与速度 vv 成反比，鼓励车辆以更快的速度行驶以缩短旅行时间：

$$L_T(v,u_a)=\alpha_T\left(\frac{1}{v}\right) \quad (3)$$

L_E是**能耗成本**，与加速度 u_a的平方成正比，以减少能耗：

$$L_E(u_a)=\alpha_E u_a^2 \quad (4)$$

L_s是**安全距离成本**，与安全距离d_s成反比，确保车辆保持安全行驶距离：

$$L_s(d_s)=\alpha_s\left(\frac{1}{d_s}\right) \quad (5)$$

其中，α_T，α_E，α_s分别是时间成本、能耗成本和安全距离成本的权重系数。

终端成本$\Phi(x(t_f))$表示车辆到达目的地时的额外成本或奖励，可以表示为：

$$\Phi(x(t_f))=\beta(x(t_f)-x_d)^2 \quad (6)$$

这里，x_d代表目的地的位置，而β是终端成本的权重系数。

通过构建这样的个体优化问题，我们能够为自动驾驶车辆提供一个明确的决策目标，即在确保安全的前提下，通过调整加速度和转向角度等控制变量，来最小化整体的运行成本。

2.1.3 优化问题

自动驾驶车辆的决策优化中，我们的目标是找到一组控制变量 ua(t)$ua(t)$ 和 uθ(t)$u\theta(t)$，使得成本函数 J 最小化。成本函数 J 综合考虑了旅行时间、能耗、安全距离等因素，具体表达式如下：

$$\min_{u_a(t),u_\theta(t)} J(x,v,\theta,d_s,u_a,u_\theta) \quad (7)$$

受到如下约束：

位置方程：$\dfrac{dx(t)}{t}=v(t)$

速度方程：$\dfrac{dv(t)}{dt}=u_a(t)$

方向方程：$\dfrac{d\theta(t)}{dt}=w(t)$

2.2 群体动态建模

在自动驾驶车辆的决策优化中，考虑群体动态是至关重要的。平均场博弈（Mean Field Games, MFG）为我们提供了一个强有力的数学框架，以模拟和优化自动驾驶车辆（AVs）在复杂交通环境中的决策问题。

2.2.1 群体行为的宏观描述

在 MFG 框架下，每辆自动驾驶车辆的行为不仅受到个体状态的影响，还受到群体平均行为的显著影响。群体行为可以通过宏观分布函数$m(x,t)$来描述，该函数随时间演化，反映了交通流的变化。具体来说$m(x,t)$描述了在时间 t 时，位置为 x 的车辆密度。这一连续函数的动态演化，通过以下偏微分方程（PDE）来刻画：

$$\frac{\partial m}{\partial t}+\frac{\partial(mv)}{\partial x}=0 \tag{8}$$

此方程体现了车辆密度的局部变化率等于车辆流动的散度，这是从流体力学中的连续性方程引申而来的，适用于描述车流中的质量守恒原理

2.2.2 个体与群体的相互作用

在 MFG 中，个体车辆的最优策略不仅取决于其即时状态，还与群体的平均行为紧密相关。这意味着车辆的控制变量$u_a(t)$和$u_\theta(t)$应基于宏观分布函数$m(x,t)$来确定。为了找到个体车辆的最优控制策略，我们为每辆车建立 Hamilton-Jacobi-Bellman (HJB) 方程，该方程综合考虑了个体成本和群体平均行为：

$$\frac{\partial V}{\partial t}+\min_{u_a,u_\theta}\left[L(x,v,\theta,d_s,u_a,u_\theta)+\frac{\partial V}{\partial x}v+\frac{\partial V}{\partial v}u_a\right]=0 \tag{9}$$

其中$V(x,v,\theta,t)$是每个车辆的价值函数，它表示从时间 t 开始，从状态(x,v,θ)出发的最小化成本。

求解 MFG 模型涉及建立并求解 HJB 方程和宏观动态方程。这通常需要采用数值

方法，如有限差分法、有限元法或谱方法。通过这些方法，我们能够确定最优控制策略$u_a^*(t)$和$u_\theta^*(t)$：

$$u_a^*(t) = arg\min_{u_a}[L(x,v,\theta,d_s,u_a,u_\theta) + V_a u_a] \tag{10}$$

$$u_\theta^*(t) = arg\min_{u_\theta}[L(x,v,\theta,d_s,u_a,u_\theta) + V_a u_\theta] \tag{11}$$

这些策略是价值函数梯度的函数，指导车辆在遵守交通规则和确保安全的前提下，以最经济的方式行驶

2.2.3 困难场景下的优化

在困难场景下，自动驾驶车辆需要做出更加复杂和精细的决策。平均场博弈（MFG）可以用来模拟和优化这些场景下的个体车辆行为，以优化整体交通流。在困难场景下，成本函数 L 需要包含额外的项来反映特定的挑战。例如，增加与车辆密度相关的成本项以应对交通拥堵，增加与避障动作相关的成本项以鼓励安全的避障策略，以及增加与天气条件相关的成本项以鼓励在恶劣天气下减速并增加安全距离。除了调整成本函数，还可以通过引入额外的约束条件来模拟困难场景，比如**速度限制**，在交通拥堵或恶劣天气条件下，实施更低的速度限制，以降低事故风险；在紧急避障和恶劣天气条件下，强制执行更大的安全距离，确保车辆有足够的反应时间；在交通拥堵或施工区域，限制车辆进入特定路段，或引导车辆选择替代路线，以减少交通压力。

以交通拥堵为例，我们可以增加一个与车辆密度相关的成本项。假设车辆密度由宏观分布函数 $m(x,t)$ 给出，我们可以这样调整成本函数：

$$L(x,v,\theta,d_s,u_a,u_\theta) = L_T(v,u_a) + L_E(u_a) + L_s(d_s) + a_C m(x,t) \tag{12}$$

其中a_C是与车辆密度相关的成本的权重系数，然后据调整后的成本函数对应修改 Hamiltonian HH 和 HJB 方程。

2.3 本章小结

本章深入探讨了平均场博弈（MFG）在自动驾驶决策建模中的应用。通过定义关键的状态变量和控制变量，本研究构建了一个多智能体交互的动态模型。状态变量包括位置、速度、加速度、方向、安全距离和相对速度，而控制变量则涵盖加速度和转向角度。进一步地，本章通过构建成本函数，将个体优化问题形式化为一个最小化问题，该函数综合了旅行时间、能耗和安全距离的成本。此外，引入了拉格朗日函数和终端成本，以确保车辆在到达目的地时的决策效率和安全性。

在群体动态的考量上，本章通过宏观分布函数 $m(x,t)$ 描述了车辆密度的演化，

并通过偏微分方程捕捉了群体行为的动态变化。个体与群体之间的相互作用通过 Hamilton-Jacobi-Bellman (HJB) 方程得以量化，从而为自动驾驶车辆提供了一个考虑群体效应的最优控制策略。针对复杂交通场景，本章提出了调整成本函数和引入额外约束条件的方法，以模拟交通拥堵、紧急避障和恶劣天气条件下的决策优化。这些调整确保了模型在面对实际交通挑战时的适应性和有效性。

总体而言，本章为自动驾驶车辆在复杂交通环境中的决策提供了一个坚实的理论基础，并为后续的模拟和实验验证章节铺平了道路。

第 3 章 方程求解和模拟实验

3.1 HJB 方程求解

利用 python 代码模拟 HJB 方程的数值求解，采用了有限差分法来近似求解 HJB 方程。有限差分法是一种常见的数值方法，用于求解偏微分方程，它通过将连续的状态空间和时间空间离散化为离散的网格点来实现。在每个时间步和状态变量上，我们都使用了均匀的网格划分，以确保模型的计算精度和稳定性。得到如图 3- 1 价值函数随着位置 x 和速度 v 的变化图，该图是一个等高线图（contourf 图），它展示了在初始时间 t=0 时，价值函数 $V(x,v,t=0)$ 随着位置 x 和速度 v 的变化情况。

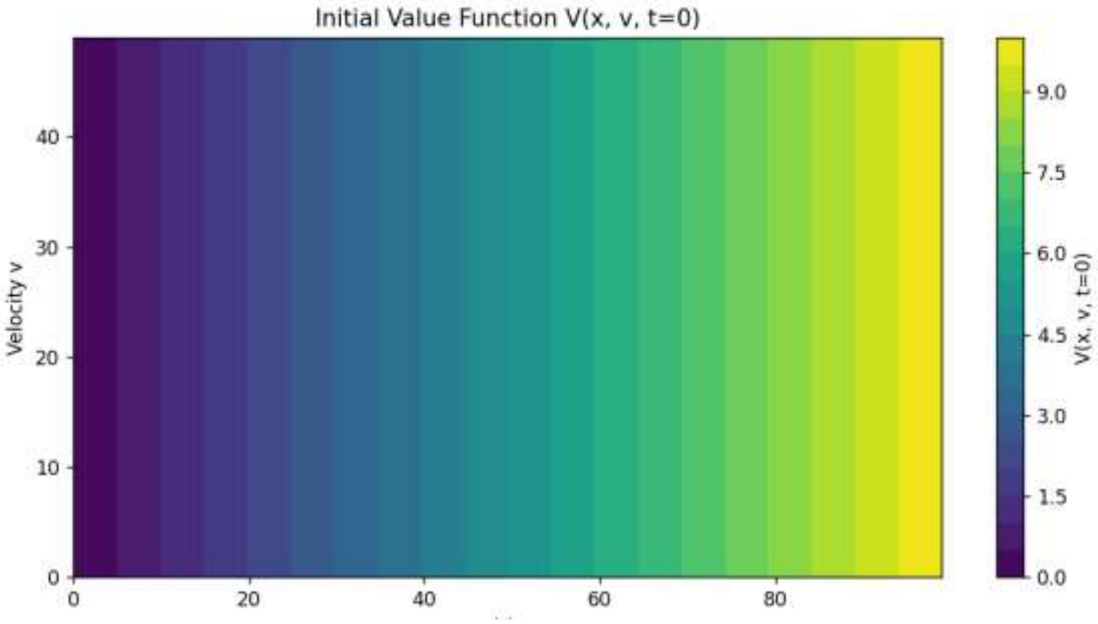

图 3- 1 价值函数随着位置 x 和速度 v 的变化图

等高线图是一种常用来表示三维数据在二维平面上的投影的方法，其中颜色的变化代表了函数值的变化。

在这张图中：

- 横轴（x 轴）表示车辆的位置 x，范围从 0 到大约 90。
- 纵轴（y 轴）表示车辆的速度 v，范围从 0 到 50。
- 颜色条（colorbar）表示价值函数 V 的值，范围从 0 到 9.5。

颜色的变化从紫色到黄色，代表了价值函数 V 的值从低到高。紫色区域表示较

低的价值函数值，而黄色区域表示较高的价值函数值。这意味着在图中，随着位置 x 和速度 v 的增加，价值函数 V 的值也在增加。

3.2 平均场模拟

在通过仿真实验验证自动驾驶车辆在复杂交通环境中的决策模型。我们采用了平均场博弈（Mean Field Games, MFG）理论，并通过 mfglib 库实现仿真，以探索车辆如何在保持安全的同时优化其行驶策略。

3.2.1 实验环境定义

利用 MFGLib 构建了一个自定义的自动驾驶环境 AutonomousDrivingEnvironment，该环境能够模拟车辆的状态和动作，以及它们之间的相互作用。MFGLib 是一个开源的 Python 库，致力于解决通用平均场博弈（MFG）的纳什均衡（NE），具有用户友好和可定制的界面，旨在促进 MFG 的应用和研究。一方面，它有助于创建和分析任意用户定义的 MFG 环境，只需对 MFG 有极少的先验知识。另一方面，它充当模块化和可扩展的代码库，供社区轻松构建原型并实施 MFG 的新算法和环境及其变体和泛化。

在实验中，本文定义了一个包含位置和速度维度的状态空间，以及与之相关的控制变量，如加速度和转向角度。环境的动态由奖励函数和状态转移函数描述，这些函数综合考虑了旅行时间、能耗和车辆密度等因素。利用 MF-OMO 算法，求解了定义的环境。该算法通过迭代过程寻找最优控制策略，以最小化车辆的旅行时间和能耗，同时考虑到车辆密度对决策的影响。算法的迭代次数设置为 300，以确保收敛。

3.2.2 结果分析

图 3- 2 实验结果图展示了仿真过程中可利用性分数随运行时间的变化。可利用性分数是衡量系统对策略变化敏感性的指标，分数越低，表示系统越稳定，对外部扰动的抵抗力越强。

初始阶段（0-1 秒）：在仿真开始时，可利用性分数较高，表明系统对策略变化较为敏感。

中间阶段（1-2 秒）：可利用性分数急剧下降，显示出系统策略的快速收敛。

后期阶段（2-4 秒）：可利用性分数继续下降，但速度减缓，最终趋于稳定，表明系统达到平衡状态。

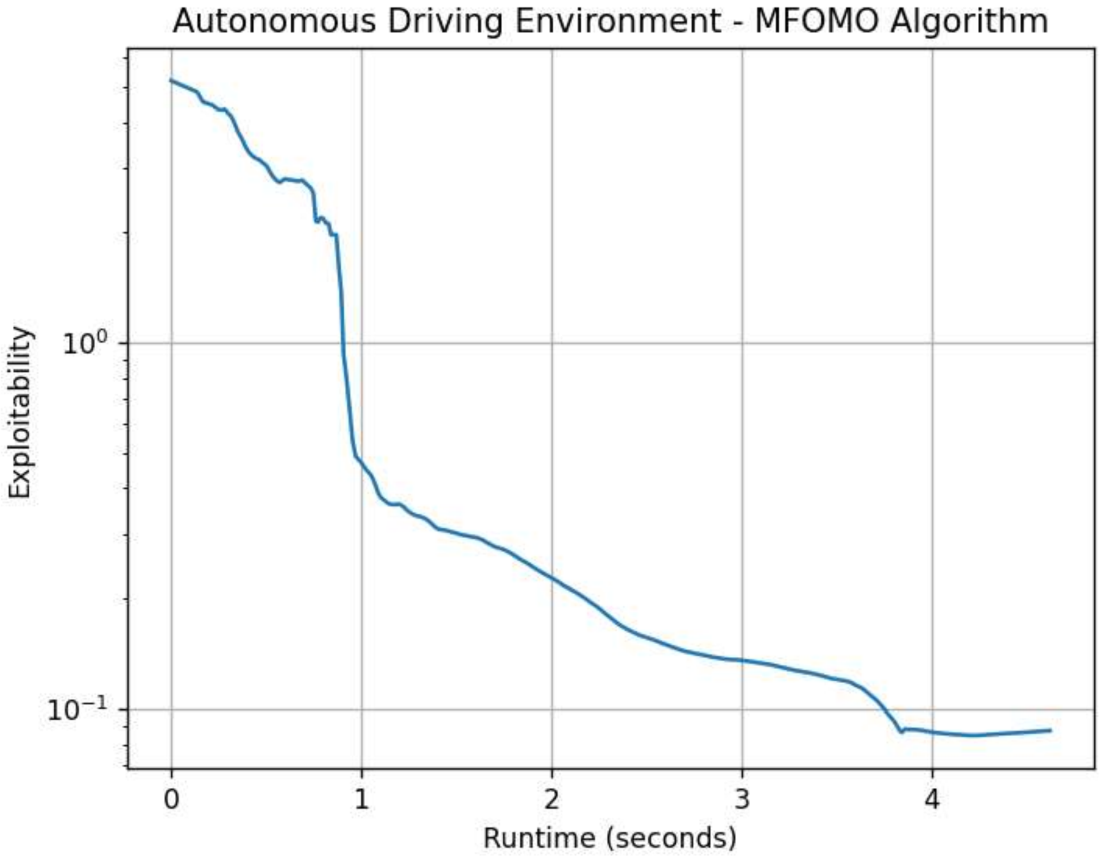

图 3- 2 实验结果图

3.3　本章小结

在本章中，通过数值求解 Hamilton-Jacobi-Bellman 方程和构建平均场博弈模拟环境，验证了基于平均场博弈理论的自动驾驶决策模型的有效性。利用有限差分法对 HJB 方程进行近似求解，并结合 MFGLib 库对自动驾驶车辆的动态行为进行了模拟。实验结果表明，所提出的模型能够优化自动驾驶车辆在复杂交通环境中的决策过程，提高了决策的效率和安全性。这些发现为自动驾驶技术的发展提供了理论支持，并为未来在更广泛的交通场景中的应用奠定了基础。

第 4 章 全文总结

4.1 主要结论

本文通过引入平均场博弈理论，对自动驾驶车辆在复杂交通环境中的决策问题进行了深入研究。研究结果表明，该理论框架能有效模拟多智能体系统中个体行为与群体动态的相互作用，为自动驾驶车辆提供了一种新的决策优化方法。通过构建个体优化问题和群体动态模型，本文不仅定义了关键的状态变量和控制变量，还构建了成本函数，将旅行时间、能耗和安全距离等因素纳入考量。此外，通过数值求解 Hamilton-Jacobi-Bellman 方程和进行模拟实验，验证了模型的有效性和实用性。本研究的结论为自动驾驶车辆在面对复杂交通环境时的决策提供了理论依据和实践指导。

4.2 研究展望

未来的研究可以探索更复杂的交通场景和更多的环境因素，如不同天气条件、道路状况以及车辆间的通信能力等。此外，可以进一步优化数值求解方法，提高模型的计算效率和精度。研究还可以扩展到实际的车辆测试，以验证模型在现实世界中的适用性和鲁棒性。最后，结合人工智能和机器学习技术，可以进一步提高自动驾驶车辆的决策智能化水平，使其更好地适应动态变化的交通环境。这些研究方向将为自动驾驶技术的发展提供新的视角和动力。

参考文献

[1] 陈华.基于博弈论的自动驾驶车辆协同换道分析[J].武汉理工大学学报(交通科学与工程版),

[2] Na X, Cole D J. Game-theoretic modeling of the steering interaction between a human driver and a vehicle collision avoidance controller [J]. IEEE Transactions on Human-Machine Systems, 2015, 45(1): 25-38.

[3] Na X, Cole D J. Linear quadratic game and non-cooperative predictive methods for potential application to modelling driver–AFS interactive steering control [J]. Vehicle System Dynamics, 2013, 51(2): 165-198.

[4] Flad M, Otten J, Schwab S, et al. Necessary and sufficient conditions for the design of cooperative shared control[C]. 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2014: 1253-1259.

[5] Flad M, Otten J, Schwab S, et al. Steering driver assistance system: A systematic cooperative shared control design approach[C]. 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2014: 3585-3592.

[6] Wang J, Dai M, Yin G, et al. Output-feedback robust control for vehicle path tracking considering different human drivers' characteristics[J]. Mechatronics, 2018, 50: 402-412.

[7] Zhang K, Wang J, Chen N, et al. A non-cooperative vehicle-to-vehicle trajectory-planning algorithm with consideration of driver's characteristics[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2018: 1-16.

[8] 薛春铭. 基于博弈的车辆协作换道策略研究[D].大连:大连理工大学, 2017

[9] 薛春铭，谭国真，丁男，等. 基于博弈论的人类驾驶与无人驾驶协作换道模型[J]. 计算机工程, 2017(12):267-272.

[10] Kita H. A merging giveway interaction model of cars in a merging section: a game theoretic analysis[J]. Transportation Research Part A Policy & Practice, 1999, 33(3–4):305-312

[11] Talebpour A, Mahmassani H S, Hamdar S H. Modeling Lane-Changing Behavior in a Connected Environment: A Game Theory Approach[J]. Transportation Research Part C Emerging Technologies, 2015, 59:216--232.