

令和5年度 卒業論文

Traffic Intersection Negotiation using Multi-Agent Policy Optimization

主指導教員

山崎達也 教授

新潟大学 工学部 知能情報システムプログラム

KHEANG Kim Ang

Abstract

This study investigates negotiation tasks at intersections for autonomous vehicles and proposes a multi-agent reinforcement learning approach to tackle this challenge. Negotiating interactions at intersections is crucial for safe and efficient autonomous driving, requiring intelligent decision-making. However, traditional approaches, such as Proximal Policy Optimization (PPO) with Gaussian distribution, face limitations in capturing the complex dynamics of negotiation scenarios. We introduce the utilization of the Beta distribution as an alternative to the Gaussian distribution. The Beta distribution offers several advantages, including greater flexibility in modeling continuous actions and a more accurate representation of the uncertainty associated with negotiation tasks. By leveraging the properties of the Beta distribution, we can enhance the agents' decision-making capabilities, leading to improved negotiation outcomes. Our approach employs a centralized learning phase and decentralized execution. In the centralized learning phase, agents learn from a global perspective, considering the interactions and dynamics of all vehicles at the intersection. This enables the agents to learn effective negotiation strategies and understand the impact of their actions on the overall traffic flow. During decentralized execution, agents make decisions independently in real time, allowing for adaptive negotiation behaviors based on local observations. Moreover, we focus on a dual-channel architecture that predicts both the alpha and beta values of the Beta distribution. This architecture provides fine-grained control over the agents' behavior, allowing them to dynamically adjust their negotiation strategies based on the specific context of the intersection and the behavior of other vehicles. By accurately predicting the alpha and beta values, our agents can make informed decisions and exhibit more nuanced negotiation behaviors. To further improve negotiation performance, we introduce a shared value advantage function that incorporates the joined actions and state of the agents. This shared value function captures the collective impact of all agents, promoting coordination and cooperation in negotiation tasks. By considering the holistic effects of the agents' actions, our approach leads to better negotiation outcomes and improved traffic flow efficiency. The experiment is done through

simulation which is a low-cost option to train autonomous vehicles. Through extensive experimentation and comparisons, we demonstrate that our proposed approach outperforms PPO with Gaussian distribution in terms of learning rate and overall performance. The utilization of the Beta distribution, combined with the dual-channel architecture and the shared value advantage function, enables our agents to exhibit interesting emergent behaviors at intersections. These behaviors include slowing down to let other vehicles pass or strategically cutting through intersections, showcasing the agents' adaptability and effective negotiation skills. Overall, this thesis presents a comprehensive multi-agent reinforcement learning approach for negotiation tasks at intersections for autonomous vehicles. By utilizing the Beta distribution, employing a centralized learning decentralized execution framework, adopting a dual-channel architecture, and incorporating a shared value advantage function, we achieve superior negotiation performance. The emergent behaviors exhibited by the agents highlight the efficacy and adaptability of our approach in complex negotiation scenarios. Our findings contribute to the advancement of negotiation tasks in intersection scenarios for autonomous vehicles, facilitating safer and more efficient traffic flow. By leveraging multi-agent reinforcement learning techniques and addressing the limitations of traditional approaches, we pave the way for the development of intelligent and socially aware autonomous vehicles capable of effectively navigating intersections and engaging in successful negotiation protocols.

目次

| | |
|---|----------|
| 第 1 章 Introduction | 1 |
| 1.1 Background | 1 |
| 1.2 Problem Statement | 2 |
| 1.3 Objectives and Scope of the Study | 2 |
| 1.4 Motivation | 4 |
| 1.5 Thesis Structure | 5 |
| 第 2 章 Literature Review | 6 |
| 第 3 章 Methodology | 7 |
| 3.1 Research Design | 7 |
| 3.2 Data and Simulator Choices | 7 |
| 3.3 Preliminary | 8 |
| 3.3.1 Markov Decision Process | 8 |

第1章 Introduction

1.1 Background

Reinforcement learning (RL) has emerged as a transformative paradigm in the field of control engineering, enabling agents to learn how to act in complex environments through observation and interaction. This learning approach has seen significant advancements in recent years, largely driven by the exponential growth in computing capability. The integration of deep learning with RL has further accelerated progress, allowing agents to be trained end-to-end by directly processing raw sensor data through neural networks to predict appropriate actions [1, 2, 3, 4, 5].

Within the realm of RL, various algorithms have been developed [4, 6, 7, 8], each dictating how an agent learns from its environment. These algorithms can broadly be categorized as model-based and model-free methods. In the context of autonomous vehicles navigating intersections, agents must acquire the skills to negotiate interactions effectively. They should be capable of adapting their behavior to be either collaborative, cooperative, or competitive, depending on the situation, in order to safely and efficiently cross intersections.

However, a notable limitation of conventional RL algorithms is that they were primarily designed with the assumption of single-agent environments, where one agent interacts with the environment in isolation. In contrast, intersection scenarios involve multiple agents (vehicles) interacting with one another, necessitating the development of multi-agent reinforcement learning (MARL) techniques.

Efforts have been made to extend model-free methods for MARL purposes, aiming to enable agents to learn in collaborative and competitive multi-agent environments. Nev-

ertheless, many of these MARL methods have tended to simplify policy architectures, overlooking essential aspects such as the choice of probability distribution density function or the design of agents' policy architectures.

1.2 Problem Statement

The problem of negotiating interactions at intersections for autonomous vehicles poses significant challenges in the field of autonomous driving. Traditional reinforcement learning (RL) algorithms, primarily designed for single-agent environments, struggle to address the complexities of multi-agent negotiation scenarios. While some efforts have been made to extend model-free methods for multi-agent reinforcement learning (MARL) purposes, these approaches often overlook crucial aspects, such as the choice of probability distribution functions and the design of policy architectures. Additionally, conventional RL algorithms typically rely on Gaussian distributions, which may not adequately capture the complexities of negotiation tasks. As a result, there is a need for a novel and sophisticated approach that leverages the advantages of multi-agent reinforcement learning, centralized learning with decentralized execution, and the Beta distribution to effectively train autonomous agents to negotiate intersections collaboratively, competitively, or in a mixed setting. Addressing these challenges will significantly advance the capabilities of autonomous vehicles, leading to safer and more efficient traffic flow at intersections.

1.3 Objectives and Scope of the Study

The primary objective of this study is to develop and evaluate a novel multi-agent reinforcement learning approach for negotiating interactions at intersections for autonomous vehicles. The specific objectives are as follows:

- To design a robust and efficient multi-agent reinforcement learning framework that enables autonomous vehicles to navigate intersections safely and efficiently by learning collaborative, competitive, and mixed negotiation strategies.

- To explore the advantages of the Beta distribution over traditional Gaussian distributions in modeling continuous actions and uncertainty during negotiation tasks, thereby enhancing the adaptability and decision-making capabilities of the autonomous agents.
- To propose a dual-channel architecture for predicting both the alpha and beta values of the Beta distribution, empowering the agents to fine-tune their negotiation strategies based on the specific context of the intersection and the behavior of other vehicles.
- To develop a shared value advantage function that incorporates joint actions and states of the agents, promoting coordination and cooperation among the autonomous vehicles during negotiation scenarios, leading to improved negotiation outcomes and overall traffic flow efficiency.
- To conduct extensive experiments and comparisons with conventional PPO algorithms using Gaussian distributions to demonstrate the superior performance and learning rate achieved by our proposed methodology.

This study focuses on addressing negotiation tasks in unsignalized intersections for autonomous vehicles. The research scope encompasses the development of a multi-agent reinforcement learning approach, specifically centered on Proximal Policy Optimization (PPO) with continuous actions and the utilization of the Beta distribution. The investigation will primarily involve autonomous vehicles acting as the agents, aiming to negotiate interactions with other vehicles in a dynamic traffic environment. The scope includes the exploration of various policy architectures, with particular emphasis on the dual-channel architecture for predicting the alpha and beta values of the Beta distribution. The shared value advantage function will also be integrated to enable agents to consider the collective impact of their actions and enhance coordination during negotiation. The performance of the proposed methodology will be evaluated through extensive simulations and comparisons with traditional PPO algorithms utilizing Gaussian distributions. The evaluation

metrics will include learning rate, negotiation outcomes, and the emergence of interesting behaviors such as slowing down to let other vehicles pass or cutting through intersections. It is important to note that this study focuses specifically on the intersection negotiation tasks for autonomous vehicles and does not cover other aspects of autonomous driving, such as motion planning or perception. Additionally, while the study aims to achieve superior performance in negotiation scenarios, it may not address all possible scenarios that autonomous vehicles may encounter in real-world urban environments.

1.4 Motivation

The motivation behind this research stems from the urgent need to address the challenges associated with negotiation tasks at intersections for autonomous vehicles. As autonomous driving technology continues to advance rapidly, intersections remain one of the most complex and critical environments for autonomous vehicles to navigate safely and efficiently. Traditional approaches, built on single-agent assumptions and Gaussian distributions, fall short in capturing the intricacies of multi-agent negotiation scenarios.

The potential benefits of effective intersection negotiation are significant, including enhanced traffic flow, reduced congestion, and increased road safety. By enabling autonomous vehicles to negotiate intersections collaboratively, competitively, or in mixed settings, we can create a more harmonious and efficient traffic ecosystem. Furthermore, successful negotiation protocols at intersections are crucial for the widespread adoption and acceptance of autonomous driving technology.

The adoption of multi-agent reinforcement learning and the utilization of the Beta distribution present promising avenues to revolutionize intersection negotiation. The ability to fine-tune negotiation strategies, consider the joint impact of all agents, and accurately model uncertainty using the Beta distribution are all factors that can greatly improve the negotiation capabilities of autonomous vehicles.

The research seeks to push the boundaries of existing techniques and bridge the gap between single-agent and multi-agent reinforcement learning. By designing a robust and

adaptable negotiation framework, we aim to empower autonomous vehicles to make well-informed and context-aware decisions at intersections, leading to safer, smoother, and more efficient traffic flow.

Ultimately, this study’s motivation lies in contributing to the development of socially-aware and intelligent autonomous vehicles that can effectively interact with other agents and adapt to various negotiation scenarios at intersections. By overcoming the limitations of conventional approaches, we aspire to pave the way for the widespread integration of autonomous vehicles in urban environments, revolutionizing transportation and advancing the future of mobility.

1.5 Thesis Structure

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry’s standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

第2章 Literature Review

Lorem Ipsum is simply dummy text of the printing and typesetting industry [5]. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of [9] Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

第3章 Methodology

3.1 Research Design

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

3.2 Data and Simulator Choices

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

3.3 Preliminary

3.3.1 Markov Decision Process

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

References

- [1] J. Chen, S.E. Li, and M. Tomizuka, “Interpretable end-to-end urban autonomous driving with latent deep reinforcement learning,” 2020.
- [2] M. Toromanoff, E. Wirbel, and F. Moutarde, “End-to-end model-free reinforcement learning for urban driving using implicit affordances,” 2019.
- [3] H. Xu, Y. Gao, F. Yu, and T. Darrell, “End-to-end learning of driving models from large-scale video datasets,” 2016.
- [4] B. Peng and etal., “End-to-end autonomous driving through dueling double deep q-network,” *Automotive Innovation*, vol.4, no.3, pp.328–337, Aug. 2021.
- [5] A.P. Capasso, P. Maramotti, A.Dell’Eva, A. Broggi, “End-to-end intersection handling using multi-agent deep reinforcement learning,” 2021.
- [6] V. Mnih and etal., “Asynchronous methods for deep reinforcement learning,” 2016.
- [7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” 2017.
- [8] J. Schulman, S. Levine, P. Moritz, M.I. Jordan, and P. Abbeel, “Trust region policy optimization,” 2015.
- [9] J.K. Terry and etal., “Pettingzoo: Gym for multi-agent reinforcement learning,” 2020.