2024-IEEE-Intelligent Vehicles Symposium - Article 1

A Safe Reinforcement Learning driven Weights-varying Model Predictive Control for Autonomous Vehicle Motion Control

Baha Zarrouki, Marios Spanakakis, Johannes Betz

1. Abstract

Determining the optimal cost function parameters of Model Predictive Control (MPC) to optimize multiple control objectives is a challenging and time-consuming task. Multiobjective Bayesian Optimization (BO) techniques solve this problem by determining a Pareto optimal parameter set for an MPC with static weights. However, a single parameter set may not deliver the most optimal closed-loop control performance when the context of the MPC operating conditions changes during its operation, urging the need to adapt the cost function weights at runtime. Deep Reinforcement Learning (RL) algorithms can automatically learn context-dependent optimal parameter sets and dynamically adapt for a Weightsvarying MPC (WMPC). However, learning cost function weights from scratch in a continuous action space may lead to unsafe operating states. To solve this, we propose a novel approach limiting the RL actions within a safe learning space representing a catalog of pre-optimized BO Pareto-optimal weight sets. We conceive a RL agent not to learn in a continuous space but to proactively anticipate upcoming control tasks and to choose the most optimal discrete actions, each corresponding to a single set of Pareto optimal weights, context-dependent. Hence, even an untrained RL agent guarantees a safe and optimal performance. Experimental results demonstrate that an untrained RL-WMPC shows Pareto-optimal closed-loop behavior and training the RL-WMPC helps exhibit a performance beyond the Pareto-front.

2. Contextual Challenges Introduction

In autonomous driving and complex system control, nonlinear Model Predictive Control (MPC) is a commonly used method. However, designing an effective MPC system poses several challenges, especially when setting the loss function. Here are some difficulties related to this:

2.1. Manual Tuning and Expert Knowledge Requirement

Setting the loss function for MPC requires extensive manual tuning and expert knowledge. This process is time-consuming and complex, often leading to suboptimal solutions, which can affect the overall performance of the system.

2.2. Impact of Weights on Safety Performance

The choice of weights in the loss function is crucial for safety performance. Inappropriate weights can lead to catastrophic accidents, especially in high-risk applications like autonomous driving.

2.3. Specificity to Operating Points

Even if suitable loss function weights are determined, they are typically only effective for specific operating points. Once operating conditions change, the reliability and effectiveness of these weights may significantly decrease. In summary, designing an MPC system requires significant effort in setting the loss function to ensure its safety and effectiveness under different operating conditions.

3. Related Works

In the field of Model Predictive Control (MPC), previous research has primarily focused on weight tuning to enhance system performance.

3.1. Traditional Approaches

In traditional approaches, Genetic Algorithm (GA) methods enhance closed-loop control performance by handling constraints, while Cooperative Particle Swarm Optimization (PSO) methods focus on improving trajectory tracking performance, widely applied in Unmanned Aerial Vehicle (UAV) domains. These methods optimize weights to enhance system stability and accuracy. However, they often require significant computational resources and are sensitive to initial parameter settings, which may lead to slow convergence or being trapped in local optima.

3.2. Bayesian Optimization Approaches

Bayesian Optimization (BO) aims to address the mismatch between real systems and models. Due to model simplifications, environmental changes, and parameter preset values not aligning with actual conditions, there is often a significant gap between the model and real-world conditions. BO captures uncertainties by introducing Gaussian processes and dynamically updates in real-time, helping to overcome these adverse factors and ensuring model reliability in different environments. However, BO methods are highly data-dependent and may face high computational complexity in high-dimensional spaces.

3.3. Reinforcement Learning Approaches

Additionally, the introduction of Reinforcement Learning (RL) has further advanced MPC development. By determining the prediction horizon of MPC through RL, system complexity can be reduced. Furthermore, RL is used to learn MPC's meta-parameters, further reducing complexity. Some studies have proposed variable-weight nonlinear MPC, which can adjust the weight matrix of the cost function online, but safety is not guaranteed. There are also approaches using RL to learn control law parameters to determine controller outputs in real-time, but safety assurance remains insufficient.

Summary: Although these methods have made progress in performance optimization, the issue of determining optimal cost function weights based on safety has not yet been resolved.

4. Proposed Work

This paper conceptualizes a weight-variable Model Predictive Control (MPC) driven by a safety reinforcement learning agent. The reinforcement learning agent autonomously adjusts the MPC cost function weights under different control conditions through foresight design, selecting the most suitable weight set from the Pareto front optimized by Multi-Objective Bayesian Optimization (MOBO). This approach allows for flexible weight adjustments in dynamic environments, ensuring optimal performance of the control system.

In experiments, this method demonstrated its adaptability in dynamically adjusting nonlinear MPC weights, particularly in simulating the control of a full-scale autonomous vehicle to follow an optimal track. Safety evidence provided by the experiments shows that even with untrained agents, this method can exhibit Pareto optimality, proving its effectiveness and safety in real-world applications.

Additionally, the paper conducts a context-related analysis of the decision-making process of the reinforcement learning agent. By testing in unseen environments, the method's generalization and robustness capabilities were evaluated, exploring the impact of continuous learning in unfamiliar settings. This research provides new insights into the application of reinforcement learning in complex control systems and demonstrates its potential to maintain performance and safety across different environments.

5. My Perspective on This Paper

This article cleverly combines the strengths of Bayesian optimization and reinforcement learning to propose an innovative approach. In summarizing previous work, the article accurately points out the widespread application and distinct shortcomings of Bayesian optimization and reinforcement learning in this field. By deploying reinforcement learning agents on the results of multi-objective Bayesian optimization, the authors successfully leverage the strengths and construct a more effective model. This combination strategy not only demonstrates unique innovation but also helps readers naturally understand the advantages of the method, showcasing a writing technique worth learning.

However, the article falls short in terms of visual presentation. Compared to other articles of the same level, some charts lack detailed data support, providing only intuitive and rough data displays. This is an area for improvement, as adding more detailed data would enhance the article's persuasiveness and scientific rigor. Overall, despite areas for improvement, the article has significant highlights in methodological innovation and writing technique.

2025-IEEE Access - Article 2

Safe Reinforcement Learning via Episodic Control

ZHUO LI, DERUI ZHU, JENS GROSSKLAGS

1. Abstract

Safe reinforcement learning (Safe RL) aims to learn policies capable of learning and adapting within complex environments while ensuring actions remain free from catastrophic consequences. This is a critical consideration in domains such as robotics, autonomous vehicles, and healthcare. Unlike traditional RL, which focuses mainly on maximizing episodic rewards, Safe RL integrates safety constraints to balance rewards and safety. Current Safe RL algorithms, while promising, lack sample efficiency as they require extensive environmental interactions to perform multi-objective optimization. This paper introduces an episodic-control-based method to enhance sample efficiency in safe policy optimization. In RL, methods based on episodic control involve the storing and replaying of past experiences or episodes, aiding in more efficient and precise policy optimization. Our proposed method involves clustering, measuring, and storing previous states based on a joint metric of returns and safety. Subsequently, we retrieve these state measurements and incorporate them into the policy optimization process through reward shaping. This approach effectively guides the policy towards high-return and safe decisions. We evaluate the performance of our method on established Safe RL benchmarks, including six safety-critical agent control tasks. The results demonstrate that our method can concurrently achieve higher episodic returns and fewer violations of safety constraints compared to the baseline methods, suggesting an effective balance between earning rewards and safety.

2. Contextual Challenges Introduction

Reinforcement Learning (RL) is effective in sequential decision-making but often risks unsafe actions. Safe RL addresses this by incorporating safety constraints. However, it faces **sample inefficiency** due to the need for extensive interactions to satisfy multiple objectives.

3. Related Works

In Safe Reinforcement Learning (Safe RL), previous research has focused on the following three approaches to improve sample efficiency:

3.1. Safety-Constrained Optimization

These methods explicitly integrate safety constraints into the optimization objectives, ensuring that safety boundaries are respected during learning.

3.2. Primal-Dual Methods

By transforming the original constraint problem into a dual problem, these methods address challenges where the original problem is difficult to optimize directly.

3.3. Model-Based Methods

By learning models of the environment, these approaches simulate potential safety issues to reduce sample requirements. While effective in reducing sample complexity, they often fail to fully utilize past interaction experiences.

Summary: The three aforementioned approaches have contributed to a reduction in sample demand to some extent, but they have not fully utilized past interaction experiences to enhance learning efficiency.

4. Proposed Work

This paper propose an episodic-control method to improve sample efficiency in Safe RL. This involves:

1.State abstraction to handle infinite state spaces.

2.Using reward and cost scores to evaluate states.

3.Joint optimization of these metrics to guide policy development.

5. My Perspective on This Paper

This paper addresses the issue of sample inefficiency in Safe Reinforcement Learning (Safe RL) by proposing an episodic-control-based Safe RL scheme. The approach focuses on leveraging past interaction experiences to enhance learning efficiency and innovatively constructs a closed-loop architecture of state abstraction $\rightarrow$ dual-dimensional measurement $\rightarrow$ memory-guided optimization.

This data-driven architectural design is highly instructive. The paper effectively improves sample utilization and reinforcement learning training speed by using an abstractor to abstract safety issues in the autonomous driving domain.

An integrated framework for motion planning and trajectory optimization of AGVs using spatio-temporal safety corridors

Xi Zhang, Yaomin Lu,Zhiyang Ju

1. Abstract

Efficiently generating safe and smooth trajectories for autonomous ground vehicles (AGVs) is a crucial and challenging task, particularly in dynamic environments with moving obstacles. This paper proposes an integrated motion planning and trajectory optimization (MPTO) framework that employs an optimization-based spatio-temporal safety corridors (STSC) to ensure trajectory smoothness and safety from a three-dimensional spatio-temporal perspective. The proposed MPTO framework comprises two layers. In the first layer, a multi-objective quadratic programming (MOQP) method was developed with the objective of rapidly generating smoothly varying STSC. The multi-objective cost function provides a comprehensive evaluation of the corridors in terms of their size, direction, and smoothness. Additionally, a convex polygonal feasible area (CPFA) was proposed to provide a linear obstacle-avoidance constraint for the MOQP. The smooth STSC provides within-corridor constraints for trajectory optimization, thereby ensuring collision avoidance of obstacles and reducing the dependence of trajectory optimization on the reference trajectory. In the second layer, an optimal trajectory generation method using polynomials is proposed to generate smooth and efficient trajectories. With smooth STSC constraints, the trajectory optimization model primarily focuses on smoothness, ensuring that the trajectory remains safe and smooth even with sudden changes in the feasible area. Finally, the proposed MPTO framework is validated through simulations and real vehicle experiments.

2. Contextual Challenges Introduction

The complexity of generating safe and smooth trajectories for Automated Ground Vehicles (AGVs) in dynamic environments arises from the presence of moving obstacles. Path planning must consider not only spatial dimensions but also the temporal aspect, making trajectory planning significantly more complex than path planning. Additionally, the dynamic changes in feasible areas have a significant impact on safety, requiring a balance between safety and smoothness. Although high-quality reference trajectories or larger safety thresholds can address this issue, they lead to increased time consumption or reduced free space. Therefore, the authors emphasize the necessity of developing theoretically sound frameworks to ensure the smoothness and safety of AGV paths in three-dimensional space-time without adding extra complexity.

3. Related Works

In this article, previous work is summarized into three approaches: search-based methods, optimization-based methods, and two-stage methods.

3.1. Search-Based Methods

Advantages: These methods can easily handle constraints in high-dimensional configuration spaces. The basic idea is to discretize the continuous space and then apply algorithms on the discretized data to find the optimal solution. This approach is naturally suited for high-dimensional constraint problems and is insensitive to the complexity of constraints. However, it faces challenges related to time consumption and trajectory quality.

3.2. Optimization-Based Methods

Advantages: These methods solve problems in continuous space and are advantageous for generating high-precision trajectories. The basic idea is to describe the trajectory planning problem as an optimization problem, defining objective functions and constraints to achieve high-precision trajectories. However, the computational complexity is high, especially when dealing with non-convex constraints, which can easily lead to local optima.

3.3. Two-Stage Methods

Stage 1: Use search-based methods to identify the most promising homotopy class paths globally. Stage 2: Use optimization-based methods to find a local optimal solution. This approach combines global path planning with local path adjustment but may lead to drastic changes in the global path, affecting vehicle safety and comfort.

Summary: By introducing additional safety constraints, path changes can be somewhat suppressed; however, the aforementioned methods increase computational costs and restrict the vehicle's driving space.

4. Proposed Work

1.Proposed MPTO Framework (Motion Planning and Trajectory Optimization)

This framework includes two lightweight optimization models: one for motion planning and one for trajectory optimization. It ensures that the process of generating spatio-temporal trajectories is efficient, resulting in paths that are both safe and smooth.

2.Proposed STSC Framework (Spatio-Temporal Safety Corridor)

This framework effectively addresses the collision-prone issues of traditional corridors. By dynamically adjusting corridors, optimizing space, and designing reference trajectory deviation penalties, STSC flexibly generates freer and safer corridors. Note: A corridor refers to the potential area where a driving path may be laid out, acting as a safety channel. It is designed to ensure that paths within this range are considered safe. The workflow of this paper follows the approach of "corridor generation $\rightarrow$ path generation $\rightarrow$ corridor adjustment $\rightarrow$ path constraint."

3.Conducted Simulations and Real-World Tests

The effectiveness and practicality of the proposed MPTO method were validated through simulations and real vehicle experiments.

5. My Perspective on This Paper

The article addresses the issue of increased computational costs and reduced path freedom due to safety constraints, proposing a framework that maintains low overall costs, high reliability, and generates smooth paths. The transition from previous research achievements to the author's own findings is seamless, similar to what is seen in article 1, making it a valuable point of study. Additionally, the article features excellent illustrations, particularly the following two figures. Figure 1 describes the process of corridor generation and optimization, while Figure 2 showcases the safety performance of different RL methods. Both are aesthetically pleasing and intuitive.

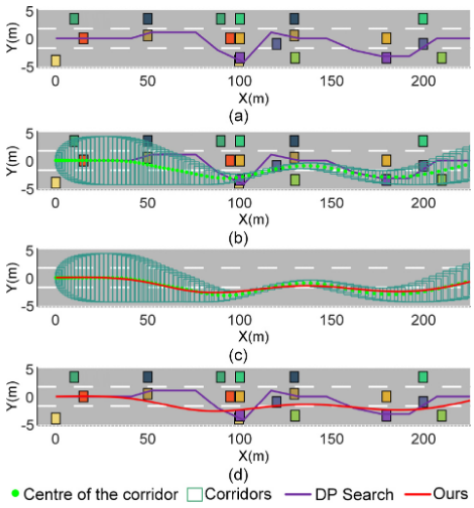

Figure 1. The process of corridor generation and optimization.

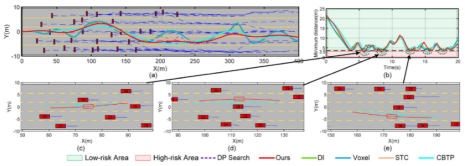

Figure 2. The safety performance of different RL methods.