Reinforcement learning-based motion planning methods

Overview of PhD thesis by:

Árpád Fehér

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

Supervisor:
Szilárd Aradi, PhD
Department of Control for
Transportation and Vehicle Systems
Budapest University of Technology and
Economics

March 31, 2023

1 Introduction

The development of self-driving vehicles can bring significant benefits to transportation. These vehicles can improve safety by reducing accidents caused by human error, improve efficiency by communicating with each other and adapting to changing road conditions, and improve accessibility for the elderly and disabled [1]. Furthermore, self-driving cars can reduce environmental effects by optimizing routes and driving habits. Overall, the development of self-driving vehicles can positively impact society [2].

Developing highly automated vehicle functions relies heavily on artificial intelligence and machine learning [3]. These tools enable the vehicle to improve its performance by learning from data. Artificial intelligence enables vehicles to adapt to different traffic situations and cope with environmental uncertainty, and it is also necessary for systems to provide a safe and comfortable journey [4]. However, it is essential to note that several challenges must be overcome before the broad deployment of full self-driving.

Vehicle motion planning is the process of determining the safe and efficient route and speed profile to follow. The motion planning process typically consists of several steps [5]. First, the vehicle's environment, including the location of obstacles and other vehicles, is described and represented as some kind of static or dynamic map. This can be generated using sensor data, such as lidar or radar, classical maps, or various V2X communications. The vehicle then needs to determine a global route through its environment that avoids static obstacles while considering constraints such as the vehicle's kinematic limitations and the road's geometry. For this task, there are several different route planning algorithms, such as Rapidly Exploring Random Trees (RRT), Probabilistic Roadmap (PRM), or variants of the A* [6] algorithm. In addition, the vehicle must generate and periodically update a local route that follows the global reference route, considering instantaneous constraints, vehicle dynamics, and dynamically moving objects. The car must follow the local path using control algorithms, thus determining the steering,

acceleration, and braking commands.

These steps need to be done in real-time, considering uncertainties and the constantly changing environment of the vehicle. Several motion planning algorithms have been developed for vehicles. In one classification, there are geometric, sampling, and optimization-based algorithms. Each algorithm has its strengths and weaknesses, and the choice of method depends on the application's specific requirements [5, 7].

Another approach is broadly classifying motion planning algorithms into model-based and model-free categories. Model-based algorithms use a mathematical model of the vehicle dynamics and the environment to plan the trajectory. Some model-based kinodynamic designers consider the vehicle's physical constraints, such as maximum acceleration and turning radius, to design a physically feasible trajectory. Other algorithms use model-based optimization techniques [8] that minimize a cost function to find the best trajectory, which may include factors such as time, energy consumption, or risk.

In contrast, model-free algorithms do not use a mathematical model of the vehicle or the environment. Instead, these algorithms learn from experience and generate trajectories based on previous examples. One model-free approach is Imitation Learning (IL), which can learn from the behavior of human drivers and adapt to traffic situations [9]. Another popular model-free approach is Reinforcement Learning (RL), which designs a trajectory by maximizing the reward value. The agent learns from experience and improves its performance over time by adapting to the environment.

Reinforcement learning is part of the machine learning toolkit, which is particularly suited to solving problems related to decision-making and control. An agent learns to act in an environment to maximize a reward value. This makes it suitable, for example, for solving tasks related to self-driving cars, where the goal is to control the car's movements for safe and efficient navigation. Reinforcement learning has been applied in self-driving cars in various ways, in motion planning, decision making, or even low-level control [10].

One of the biggest challenges in reinforcement learning is the so-called sim-to-real problem [11], i.e., translating the agent trained from the simulation into reality, which does not behave correctly due to differences between the modeled environment and reality. Agents are often trained in a simulation environment because it is safe, fast, and cheaper than reality training. Overall, the sim-to-real problem is a challenging and active research area requiring a combination of machine learning, robotics, and control theory. The results presented in this thesis also seek an answer to this problem. Therefore, testing the algorithms developed on a real vehicle was important to demonstrate their effectiveness.

This thesis deals with motion planning techniques for highly automated vehicles using reinforcement learning as a machine learning approach.

Reinforcement learning is an exciting and dynamically evolving area of machine learning. It has proven effective in achieving better than ever and even better than human performance in a wide range of scientific and application areas[12, 13]. It is based on a Markov decision process, which assumes an entirely discrete space of states, events, and actions in its original form. Accordingly, the initial algorithms of RL are defined in this form, which makes them difficult to apply to many vehicle control problems. The renaissance of this approach, like many other machine learning algorithms, occurred in the early 2010s, along with the revival of artificial neural networks. This change also opened the way for creating algorithms that already provided the user with a continuous state space and decision options.

The four theses presented in this dissertation illustrate the path that led to the creation of a hierarchical agent for complex dynamic obstacle avoidance, starting from an introductory feasibility study.

The first thesis performs a feasibility analysis of an end-to-end tracking task. A simulation environment is implemented for efficient and fast learning, including a kinematic vehicle model, ideal sensor models detecting angular and distance errors, and a random trajectory generator of different difficulties. A reward function was designed to calculate the reward obtained by passing through equally

spaced gates. I also developed a method for the sim-to-real problem, which allows testing an agent programmed in Python using industrial simulation software and an actual vehicle. A virtual CAN network implements an interface between the simulation software, the physical vehicle, and the Python environment.

Although the end-to-end approach seems correct, it does not provide information on the long-term performance of maneuvers and is, therefore, not very useful in vehicle control tasks. Accordingly, in the second thesis, I aimed to examine how to implement longer maneuver planning methods than single-step control using reinforcement learning in a way that provides some feasibility metrics. To do this, I defined the continuous reinforcement learning task as given initial and final states of a vehicle's trajectory, between which the agent must determine the parameters of a continuous curve that will be the path to follow. To learn this, the agent using the self-learning method, a controller must drive the vehicle along this planned path in simulation, during which the reward function can be evaluated and also teaches the agent to estimate the feasibility.

Building on the previous results, the third thesis attempts to implement a more complex motion planning task. In this case, the aim is to optimize the parameters of a prototype maneuver, which I have demonstrated using the dual lane-changing problem. In order to implement it, the parameters of not one but several continuously connected curves must be defined. As in the previous task, the sim-to-real problem - i.e., the behavior of the agent trained in simulation in real life - is even more critical in this case. To solve this problem, I used the domain-randomization technique, i.e., I randomly influenced the simulation conditions so that the agent would eventually be able to adapt to real-world environmental situations. As in the previous thesis, I validated the results with a field test, which showed that the trained agent outperforms the capabilities of an average driver.

The fourth thesis expands the static object avoidance problem to encompass dynamic objects. A pedestrian model that randomly crosses the path of the vehicle was employed to demonstrate the

development of the algorithm. Avoiding the dynamic obstacle is no longer a one-step path-planning task. In this scenario, the method formulation resembles a model predictive control scheme, where a complete path is planned based on the vehicle's state and the pedestrian's movement, but only a small portion of the path is executed. Then, based on the new situation, the entire planning task is restarted. The trained agent was tested in a simulation environment and compared to the performance of human drivers.

2 Contributions of the Thesis

Thesis I

I have developed a method suitable for "end-to-end" path tracking of automated road vehicles using reinforcement learning. The method uses tracking error and yaw angle error, with the discrete set of steering angle values as actions. A Deep Q-network based agent determines the intervention of steering angle based on these high-level sensor signals.

Related publications: [Feher1], [Feher2]

Remarks

- For training, I developed a simulation environment that includes the vehicle model, taking into account its kinematic limitations, ideal sensor models of heading angle error and tracking error, and a random track generator with parameterizable difficulty.

- I developed the reward function, which determines the reward based on the result of passing through gates placed equidistantly.

- I developed a method for the sim-to-real problem; with which the agent implemented in the python environment can be effectively tested in industrial simulation softwares and real vehicles.

- I developed an interface using virtual CAN network between python and the simulation software in the evaluation phase, which can be replaced with real CAN network during the field tests.

Thesis II

I have developed a method for a trajectory planning task with initial conditions and a predefined target state, as well as for estimating the feasibility of the trajectory. The planner uses a Deep Deterministic Policy Gradient agent, whose training is based on a novel one-step training method. During the training of the agent, the designed path is evaluated with an internal closed-loop controller. The method includes a new reward function that takes into account the vehicle's side slip in addition to the lateral error and the angular error.

Related publications: [Feher3] [Feher4]

Remarks

- In addition to determining the waypoints of the polynomial trajectory, the trained agent provides an estimation of the dynamic feasibility.
- The algorithm using the previously developed "sim2real" method was tested in a simulation environment and on a real vehicle.
- I illustrated that a combination of machine learning and classical methods could be suitable for the design of the motion control of autonomous vehicles.

Thesis III

I have developed an avoidance maneuver method for automated road vehicles. The method is demonstrated through a double lane change maneuver as a critical traffic situation. The developed route planner strives to design a dynamically feasible path, which is able to avoid obstacles appearing in front of the vehicle. The solution uses the dynamic parameters of the vehicle and the geometric parameters of the problem as an input, and defines a set of curves as a path to be follow as an output, by using reinforcement learning.

Related publication: [Feher5]

Remarks

- The planning was carried out by the Deep Deterministic Policy Gradient agent, whose inputs are the speed of the vehicle and the geometric parameters of the occupied areas provided by the environmental sensors.
- The outputs of the agent are the parameters of the curve sections describing the trajectory.
- The reward values are calculated when driving a non-linear dynamic vehicle model along the track.
- I evaluated the design by performing the moose test defined in the ISO 3888-2 standard.
- I compared the results of the different approaches and the performance of human drivers.

Thesis IV

I have developed a local trajectory planning and control method for road vehicles, which designs the route and speed profile to avoid a moving objects crossing the road taking into account the vehicle dynamics and the geometric limitations of the road. The trained Twin Delayed Deep Deterministic Policy Gradient agent continuously redesigns the trajectory at a given horizon, which is performed by a model predictive controller.

Related publication: [Feher6]

Remarks

- During the training the longitudinal Stanley and the lateral MPC controllers are used to perform the trajectory tracking.
- The inputs of the agent are the relative distance and speed of the moving object provided by the environmental sensors, as well as the ego vehicle's motion state.
- The outputs of the agent are the parameters of a polynomial function describing the route and the parameters of the speed profile.
- The method was compared with the performance of human drivers in a simulation environment.
- Although the developed algorithm was defined for road vehicles with pedestrians crossing, its principles can be applied to many other dynamic object avoidance problems.

3 Future work

Research into reinforcement learning, including its application to traffic and vehicle control tasks, is a fascinating field, and accordingly, its international results are developing at an extraordinary pace. Both the basic learning algorithms and the complementary meta-heuristics and artificial neural network structures are being innovated year by year. The results presented in this dissertation can be used to implement real-time vehicle control tasks, but their robustness is not yet guaranteed, which is not surprising as this is one of the most pressing challenges in the whole field at present. Accordingly, testing and validating algorithms is a particularly attractive area of research. In addition, more accurate implementation of different sensing models in the system, multi-agent learning, and many other vehicle control tasks could be important development aspects. In my further research, I plan to focus on integration tasks, the design of different feasible control schemes, and the development of systems that can be tested in real life.

Publications related to the thesis

[Feher1] Árpád Fehér, Szilárd Aradi, and Tamás Bécsi. Q-learning based reinforcement learning approach for lane keeping. In *2018 IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI)*. IEEE, nov 2018.

[Feher2] Árpád Fehér, Szilárd Aradi, and Tamás Bécsi. Fast Prototype Framework for Deep Reinforcement Learning-based Trajectory Planner. *PERIODICA POLYTECHNICA-TRANSPORTATION ENGINEERING*, ONLINEFIRS:1–6, 2020.

[Feher3] Arpad Feher, Szilard Aradi, Ferenc Hegedus, Tamas Becsi, and Peter Gaspar. Hybrid DDPG Approach for Vehicle Motion Planning. In *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics, ICINCO 2019, Vol. 1 Setúbal, Portugália : SciTePress, (2019)*, pages 422–429. Scitepress, 8 2019.

[Feher4] Árpád Fehér, Szilárd Aradi, Tamás Bécsi, Péter Gáspár, and Zsolt Szalay. Proving Ground Test of a DDPG-based Vehicle Trajectory Planner. In *2020 European Control Conference (ECC)*, pages 332–337, 2020.

[Feher5] Árpád Fehér, Szilárd Aradi, and Tamás Bécsi. Hierarchical evasive path planning using reinforcement learning and model predictive control. *IEEE Access*, 8:187470–187482, 2020.

[Feher6] Árpád Fehér, Szilárd Aradi, and Tamás Bécsi. Online trajectory planning with reinforcement learning for pedestrian avoidance. *Electronics*, 11(15):2346, 2022.

References

[1] Daniel Watzenig and Martin Horn. *Automated driving: safer and more efficient future driving*. Springer, 2016.

[2] Tamas Tettamanti, István Varga, and Zsolt Szalay. Impacts of autonomous cars from a traffic engineering perspective. *Periodica Polytechnica Transportation Engineering*, 44:244–250, 10 2016.

[3] Yifang Ma, Zhenyu Wang, Hong Yang, and Lin Yang. Artificial intelligence applications in the development of autonomous vehicles: a survey. *IEEE/CAA Journal of Automatica Sinica*, 7(2):315–329, 2020.

[4] Biagio Ciuffo, Konstantinos Mattas, Michail Makridis, Giovanni Albano, Aikaterini Anesiadou, Yinglong He, Szilárd Josvai, Dimitris Komnos, Marton Pataki, Sandor Vass, and Zsolt Szalay. Requiem on the positive effects of commercial adaptive cruise control on motorway traffic and recommendations for future automated driving systems. *Transportation Research Part C: Emerging Technologies*, 130:103305, 2021.

[5] Brian Paden, Michal Cap, Sze Zheng Yong, Dmitry Yershov, and Emilio Frazzoli. A Survey of Motion Planning and Control Techniques for Self-driving Urban Vehicles. *CoRR*, pages 1–27, 2016.

[6] Michael Montemerlo, Jan Becker, Suhrid Bhat, et al. Junior: The stanford entry in the urban challenge. *Journal of field Robotics*, 25(9):569–597, 2008.

[7] Laurène Claussmann, Marc Revilloud, Dominique Gruyer, and Sébastien Glaser. A review of motion planning for highway autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 21(5):1826–1848, 2019.

[8] Ferenc Hegedüs, Tamás Bécsi, and Szilárd Aradi. Dynamically feasible trajectory planning for road vehicles in terms of sensitivity and robustness. *Transportation Research Procedia*, 27:799 – 807, 2017. 20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017, Budapest, Hungary.

[9] Abdoulaye Oumar Ly and Moulay A Akhloufi. Learning to drive by imitation: an overview of deep behavior cloning methods. *IEEE Transactions on Intelligent Vehicles*, 2020.

[10] Szilárd Aradi. Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(2):740–759, 2022.

[11] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 737–744, 2020.

[12] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

[13] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.