Social DriveNet: Integrating DDQN with Social Attention for Autonomous Traffic Navigation

Leone Lage Perdigão
Department of Computer Science
University of Bath
Bath, BA2 7AY
llp31@bath.ac.uk

1 Problem Definition

This research project is inspired by and builds upon the paper "Social Attention for Autonomous Decision-Making in Dense Traffic" Leurent and Mercat (2019) with significant modifications aimed at exploring the efficacy of Double Deep Q-Network (DDQN) in complex traffic scenarios. This study aims to validate a possible decision-making enhancement in dense traffic by integrating social attention mechanisms, a novel approach designed to model and predict the behaviors of surrounding drivers in a dynamic and densely populated environment.

As such, the project enfolds both a baseline Deep Q-Network (DQN) model and a DDQN model with Social Attention. These models are trained and evaluated leveraging the 'highway-env' simulation environment, a specialised tool for autonomous driving research that provides realistic traffic scenarios and the ability to simulate intricate vehicular interactions Leurent (2018).

Environment Description:

The 'highway-env' simulation environment, developed by Edouard Leurent, offers a versatile setting for testing and developing autonomous driving algorithms. It simulates a multi-lane highway where the autonomous agent (the ego vehicle) must navigate through traffic, avoid collisions, and make strategic driving decisions Leurent (2018). The environment supports various traffic densities and complex driving behaviors, making it an ideal platform for studying the effects of social attention mechanisms in reinforcement learning models.

States:

- **Observations**: The state of the environment is captured through observations that include features such as the presence, position (x, y coordinates), velocity (vx, vy components), and orientation (cosine and sine of the heading angle) of nearby vehicles. The observation space is standardized to include a fixed number (e.g., 15) of nearby vehicles to maintain consistent dimensionality across different traffic situations. These observations can be formatted as vectors in kinematics or as matrices in occupancy grids, detailing the spatial distribution of traffic around the ego vehicle.

Actions:

- **Discrete Actions**: The action space is discrete, encompassing a set of quantized decision options that the learning agent can execute. These actions typically include:

  $$a_t \in \{\mathbf{LANE\_LEFT}, \mathbf{IDLE}, \mathbf{LANE\_RIGHT}, \mathbf{FASTER}, \mathbf{SLOWER}\}$$

  Here, 'LANE_LEFT' and 'LANE_RIGHT' correspond to lane change maneuvers, 'FASTER' and 'SLOWER' adjust the vehicle's speed, and 'IDLE' maintains the current state without changes. The discrete nature of these actions simplifies the decision-making process and aligns with the common control strategies used in vehicular navigation systems.

Transition Dynamics:

- **Vehicle Dynamics and Road Interactions**: The transition dynamics describe how the state of the environment evolves in response to the actions taken by the autonomous agent. This includes the kinematics of the vehicle, such as changes in speed and direction, as well as interactions with the road infrastructure and other vehicles, which are all influenced by the physical properties of the vehicle and traffic regulations.

Reward Function:

- **General Formulation**: The reward function is designed to encourage optimal driving behavior by combining a velocity component and a collision penalty:

$$R(s_t, a_t) = \alpha \left(\frac{v_t - v_{\mathbf{min}}}{v_{\mathbf{max}} - v_{\mathbf{min}}} \right) - \beta \cdot \mathbf{collision}$$

  where v_t is the current velocity, $v_{\mathbf{min}}$ and $v_{\mathbf{max}}$ are the minimum and maximum allowable speeds, respectively, and α and β are coefficients weighting the importance of speed maintenance and collision avoidance.
- **In Goal-Oriented Scenarios**: For tasks such as parking, the reward function might also include a goal proximity term, which measures the agent's effectiveness in navigating towards a specific target:

$$R(s_t, a_t) = -\gamma \|\mathbf{s}_t - \mathbf{s}_{\mathbf{goal}}\|_p - \beta \cdot \mathbf{collision}$$

  where $\| \cdot \|_p$ denotes the p-norm distance to the goal state $\mathbf{s}_{\mathbf{goal}}$, and γ is a weighting factor. Nonetheless, this reward function variation is not directly in the scope of this work.

This framework sets up the autonomous system to learn effective driving policies through simulation, optimizing for both efficiency and safety in complex, densely populated traffic scenarios. The use of social attention mechanisms aims to dynamically incorporate the intentions and relative positions of other traffic participants into the decision-making process, enhancing the overall intelligence and responsiveness of the driving system.

2 Background

2.1 Introduction to Reinforcement Learning

Reinforcement Learning (RL) provides a framework for learning optimal policies in sequential decision-making problems, where an agent learns to achieve goals by interacting with a dynamic environment. This learning process involves observing the state of the environment, selecting actions, and receiving rewards based on the outcomes of these actions González et al. (2016).

2.2 Reinforcement Learning in Autonomous Driving

In the context of autonomous driving, RL can adaptively refine driving strategies by continually processing vehicular dynamics and environmental feedback. The primary goal is to enhance safety, efficiency, and adaptability in complex and unpredictable traffic scenarios González et al. (2016).

2.3 Relevant RL Methods and Their Applications

2.3.1 Deep Q-Networks (DQN)

DQN integrates deep neural networks with the Q-learning algorithm, allowing agents to approximate the optimal action-value function in high-dimensional state spaces typical of driving environments. The algorithm updates the Q-values based on the Bellman equation as follows:

$$Q^{new}(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

Despite its success, DQN is prone to overestimations of Q-values due to the max operation in its update rule, which can lead to suboptimal policy learning, particularly in complex driving scenarios where precise action evaluation is critical van Hasselt, Guez and Silver (2015).

2.3.2 Double Deep Q-Networks (DDQN)

To address the overestimation issue in DQN, DDQN modifies the Q-value update mechanism by decoupling the action selection from the target Q-value generation:

$$Q^{new}(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma Q(s_{t+1}, \arg\max_a Q(s_{t+1}, a)) - Q(s_t, a_t) \right]$$

This approach reduces bias in the learning process, enhancing the stability and reliability of the learning outcomes, which is crucial for safety-critical applications like autonomous driving Zhang et al. (2018).

2.3.3 Policy Gradient Methods

Policy gradient methods such as Proximal Policy Optimization (PPO) and Trust Region Policy Optimization (TRPO) optimize the policy directly by estimating the gradient of the expected return. These methods are particularly useful in continuous action spaces, such as steering angle and acceleration control in autonomous vehicles:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}\left[\sum_{t=0}^{T} \nabla_\theta \log \pi_\theta(a_t|s_t)(R_t - b(s_t)) \right]$$

PPO and TRPO provide mechanisms to maintain exploration while avoiding large updates that could lead to unstable learning, making them well-suited for the adaptive and interactive nature of driving behaviors González et al. (2016).

2.4 Integration of Social Attention Mechanisms

The integration of social attention mechanisms, as explored in the work by Leurent et al. (2019), represents a significant advancement in modeling interactions among multiple agents, such as vehicles in dense traffic. These mechanisms weight the influence of surrounding agents, enhancing the capability of RL models to make context-aware decisions:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

By focusing on relevant vehicles and their behaviors, social attention enables more nuanced and predictive driving strategies, potentially reducing accidents and improving traffic flow efficiency Leurent and Mercat (2019).

2.5 Assessment of Techniques

While traditional and advanced RL methods offer substantial benefits in autonomous driving, their effectiveness heavily relies on the quality of the simulation environment, the accuracy of the state representation, and the scalability of the algorithms. Continuous research and development efforts are necessary to address these challenges, ensuring that the RL-based systems can operate reliably in real-world traffic conditions.

3 Method

This study employs Reinforcement Learning (RL) to enhance decision-making in dense traffic for autonomous vehicles. We build on the baseline model of a standard Deep Q-Network (DQN), incorporating a Double Deep Q-Network (DDQN) with a novel social attention mechanism. This hybrid approach aims to address overestimation biases and improve interaction awareness with surrounding traffic.

3.1 Baseline Model

Initially, our approach involves using a baseline DQN model as described by Mnih et al. Mnih et al. (2015). The DQN serves as a foundational model from which we assess enhancements brought about by DDQN and social attention mechanisms.

3.2 Simulation Environment

The ‘highway-env‘ simulation platform Leurent (2018), offers a dynamic environment to test autonomous driving algorithms. It simulates various traffic conditions, enabling the testing of our models under different traffic densities and behaviors.

3.3 Advanced Model Implementation

- **DDQN Model:** We extend the DQN framework by incorporating a dual estimator strategy to mitigate the overestimation bias, enhancing the accuracy of action-value predictions.
- **Social Attention Mechanism:** This mechanism is integrated into the DDQN model. It prioritizes computational resources on vehicles that most significantly influence the ego vehicle's decision-making process, improving interaction modeling with other traffic participants.

3.4 Evaluation Metrics

Models are evaluated based on safety (collision rates), compliance and efficiency (traffic rule adherence and optimization of travel time), and operational effectiveness (through metrics like average speeds and maneuver correctness).

3.5 Hyperparameter Optimization

Using the Optuna framework Akiba et al. (2019), we tune learning rates, batch sizes, and network architectures to optimize model performance under complex simulation conditions.

3.6 Training and Validation

Models undergo rigorous episodic training within the ‘highway-env‘, focusing on optimizing policies for safety and efficiency. Validation assesses model robustness against predefined safety and efficiency benchmarks.

3.7 Rationale for Method Choices

Our choice of integrating DDQN with social attention is driven by their demonstrated potential in accurately predicting and managing interactions in dynamic traffic scenarios, significantly enhancing policy reliability over traditional DQN approaches.

4 Results

A presentation of your results, showing how quickly and how well your agent(s) learn (i.e., improve their policies). Include informative baselines for comparison (e.g. the best possible performance, the performance of an average human, or the performance of an agent that selects actions randomly).

5 Discussion

An evaluation of how well you solved your chosen problem.

6 Future Work

A discussion of potential future work you would complete if you had more time.

7 Personal Experience

A discussion of your personal experience with the project, such as difficulties or pleasant surprises you encountered while completing it.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining*.

González, D., Pérez, J., Milanés, V. and Nashashibi, F., 2016. A review of motion planning techniques for automated vehicles. *Ieee transactions on intelligent transportation systems*, 17(4), pp.1135–1145. Available from: https://doi.org/10.1109/TITS.2015.2498841.

Hasselt, H. van, Guez, A. and Silver, D., 2015. Deep reinforcement learning with double q-learning. *Corr*, abs/1509.06461. 1509.06461, Available from: http://arxiv.org/abs/1509.06461.

Leurent, E., 2018. An environment for autonomous driving decision-making. https://github.com/eleurent/highway-env.

Leurent, E. and Mercat, J., 2019. Social attention for autonomous decision-making in dense traffic. 1911.12250.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. and Hassabis, D., 2015. Human-level control through deep reinforcement learning. *Nature*, 518(7540), February, pp.529–533. Available from: http://dx.doi.org/10.1038/nature14236.

Zhang, Y., Sun, P., Yin, Y., Lin, L. and Wang, X., 2018. Human-like autonomous vehicle speed control by deep reinforcement learning with double q-learning. *2018 ieee intelligent vehicles symposium (iv)*. pp.1251–1256. Available from: https://doi.org/10.1109/IVS.2018.8500630.

Appendices

If you have additional content that you would like to include in the appendices, please do so here. There is no limit to the length of your appendices, but we are not obliged to read them in their entirety while marking. The main body of your report should contain all essential information, and content in the appendices should be clearly referenced where it's needed elsewhere.

Appendix A: Example Appendix 1

Appendix B: Example Appendix 2