Ostfalia
Hochschule für angewandte Wissenschaften

IARIA

Application of a Maneuver-Based Decision Making Approach for an Autonomous System Using a Learning Approach

Xing, Xin and Ohl, Sebastian

Contact: {xi.xing | s.ohl}@ostfalia.de

Ostfalia Hochschule für angewandte Wissenschaften
– Hochschule Braunschweig/Wolfenbüttel
Salzdahlumer Str. 46/48 · 38302 Wolfenbüttel

Ostfalia
Hochschule für angewandte Wissenschaften

Presenter Resume

Current Role:

- Research Assistant at Ostfalia University of Applied Sciences

Background:

- Bachelor’s in Electrical Engineering from Ostfalia University of Applied Sciences
- Master’s in Electrical Engineering from Technical University of Braunschweig
- 3 years of experience in robotics systems
- Over 1 years of experience in autonomous driving

Research Interests:

- Autonomous vehicles
- Machine learning for decision-making systems

ExerShuttle Project

Collaborative Background: Joint research on autonomous driving technologies by Ostfalia and TU Clausthal.

Project Goal: Providing a practical autonomous driving experience and research platform by showcasing autonomous bus scenario.

Research Significance: Enhances practical teaching by linking theoretical knowledge with practical applications.

Agenda

Ostfalia
Hochschule für angewandte Wissenschaften

Introduction

Generated from Chatgpt AI

Advanced Decision Making

- Safety-critical car-following models
- Adaptive Cruise Control (ACC)
- Automatic Emergency Braking (AEB)

Traditional Decision Making methods

↓

Learning-Based Approaches (RL)

Problem Formulation

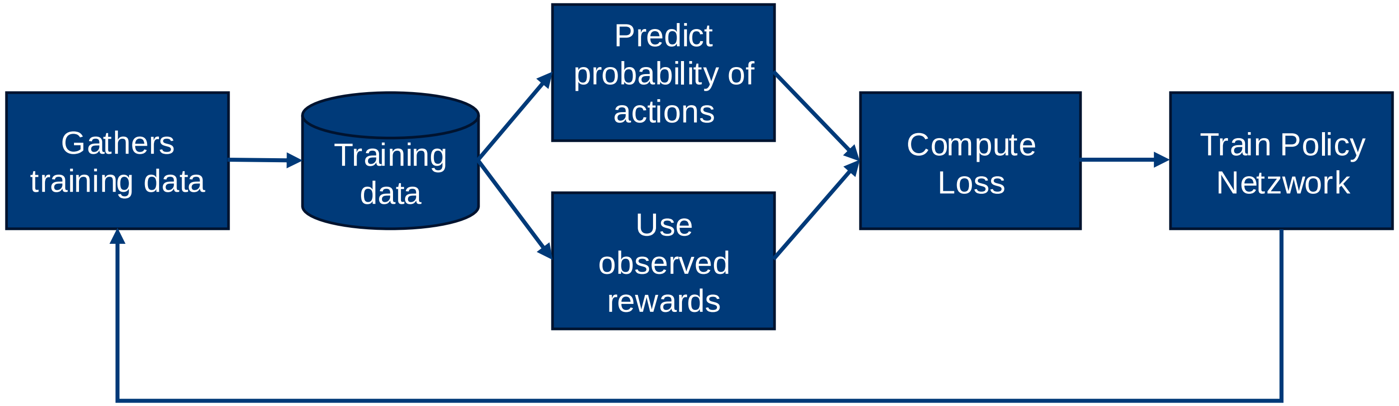
Simulate real-world driving to test ACC and AEB.

Touch on the specific RL methods used, such as Policy Gradient (PG) and Proximal Policy Optimization (PPO)

Partially Observable Markov Decision Processes (POMDP) for representing interactions in the driving environment

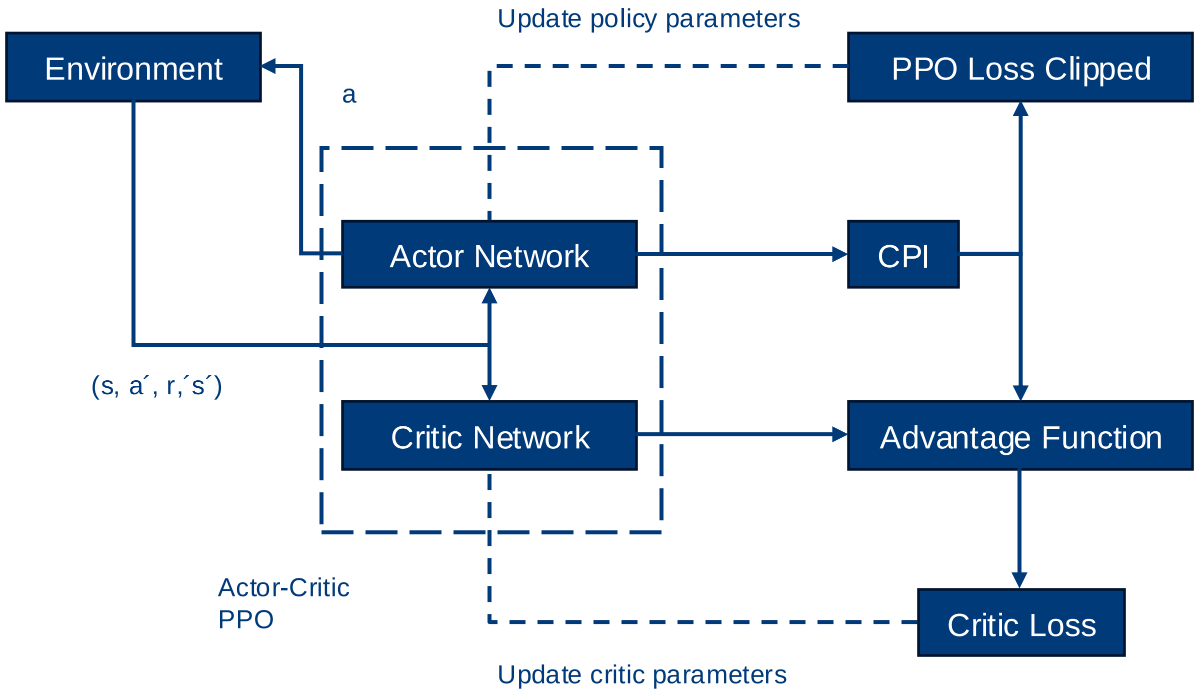
Policy-based Reinforcement Learning

- **Policy Gradient (PG)**:
  - **Overview**: A reinforcement learning technique that adjusts policy directly based on the gradient of the expected reward.
  - **Key Feature**: Uses gradient ascent to incrementally improve policy decisions based on rewards.
  - **Application**: Ideal for environments where the policy needs continuous refinement.



Policy-based Reinforcement Learning

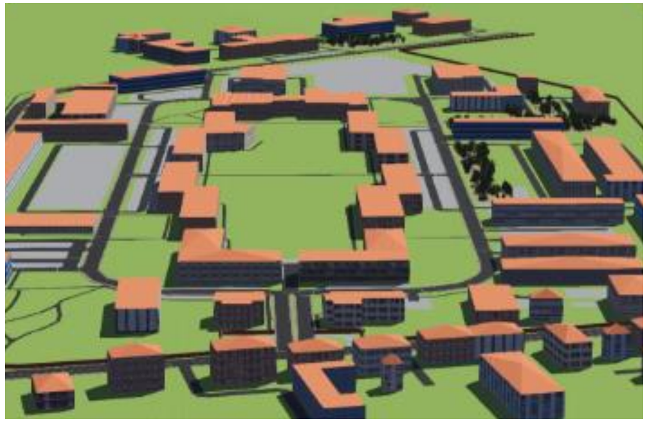
- **Proximal Policy Optimization (PPO)**:
  - **Overview**: An advanced policy gradient method that improves upon earlier techniques by limiting changes in policy updates.
  - **Key Feature**: Utilizes a clipping mechanism to prevent too drastic policy updates, ensuring more stable learning.
  - **Advantages**: Provides better sample efficiency and more consistent learning performance compared to standard PG.

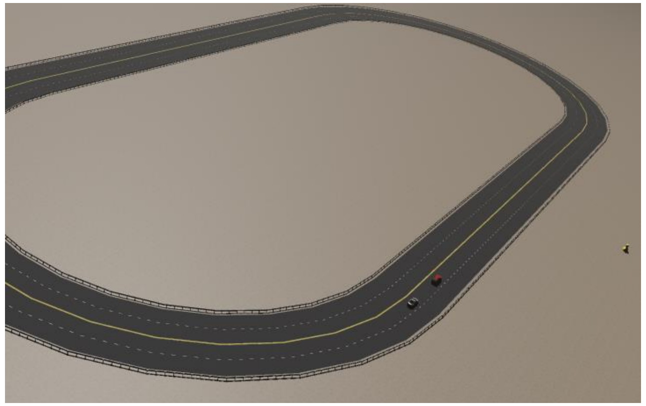


Methodology – Simulation Environment

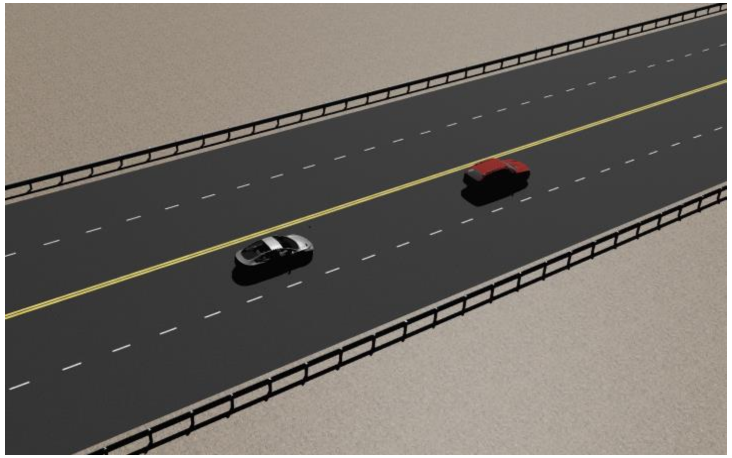
Intelligent Driver Model (IDM) for ACC:

- Desired velocity: $30\ km/h$
- Safe time headway: $1.5\ s$
- Minimum distance: $7\ m$
- Acceleration: $\pm 1.5\ m/s^2$

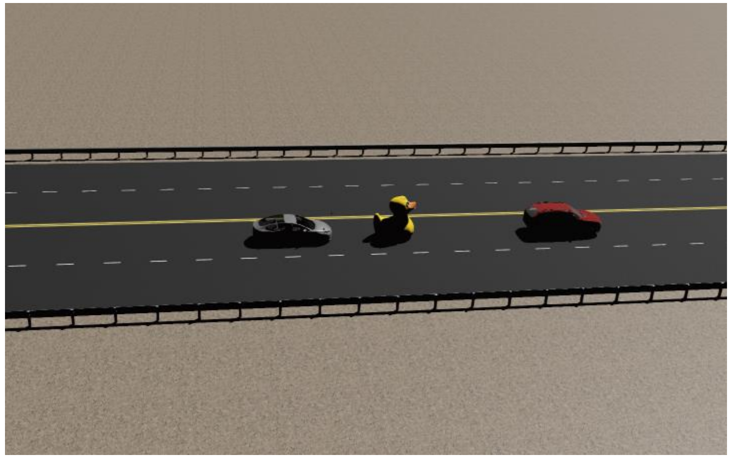




Methodology – Simulation Environment



Leading Vehicle with
speed of 20 km/h

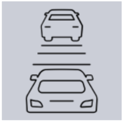


A yellow duck appears randomly
after 4s

If the car does not brake in time,
the car will collide with
the duck.

Methodology – Simulation Environment

- Action Space:
  - ACC
  - AEB
- State Space
  - V_{AV}
  - V_{LV}
  - G
  - A



Reward Function for ACC

- Reward for distance between vehicles
- Reward for velocity difference between vehicles

Reward Function for AEB

- Reward for the occurrence of a collision
- Reward for distance between vehicles

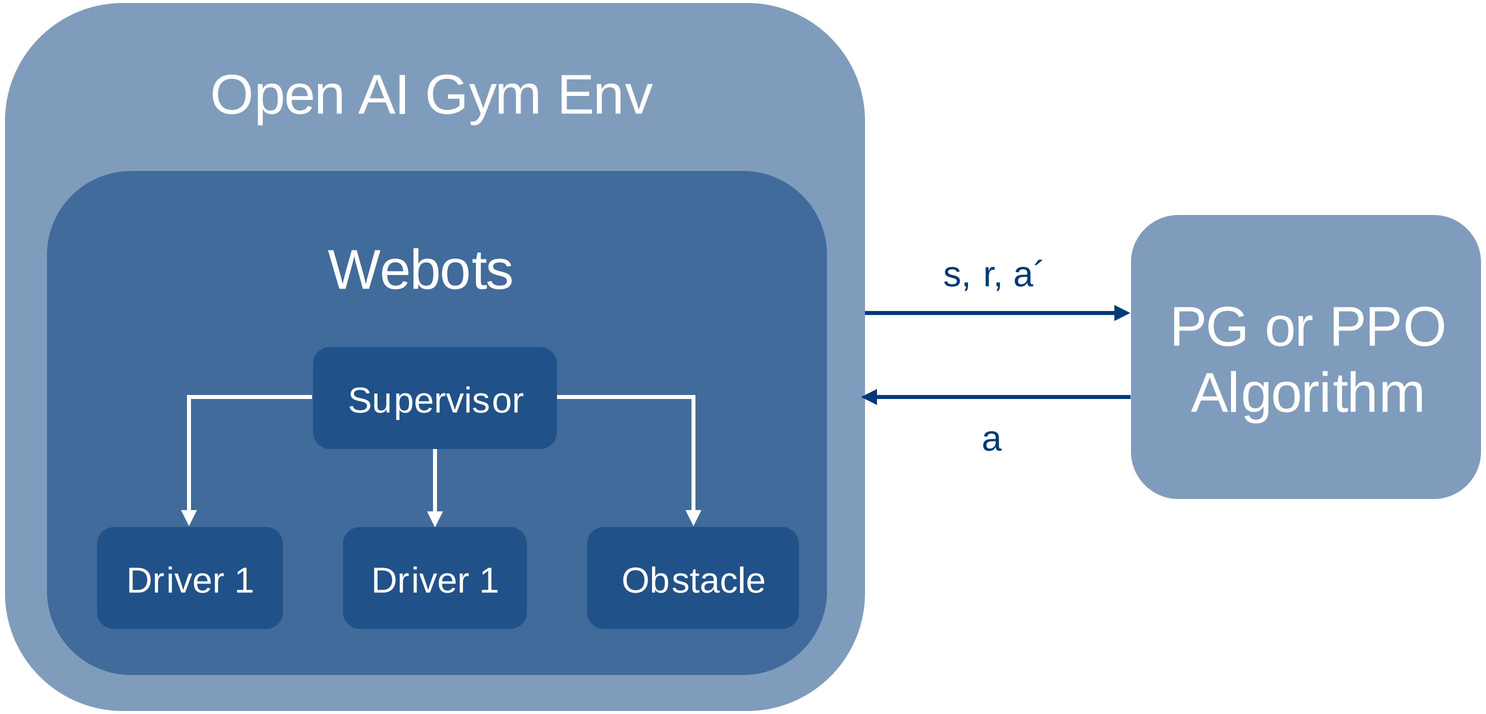
Reward Function for comfort

- Reward for harsh braking

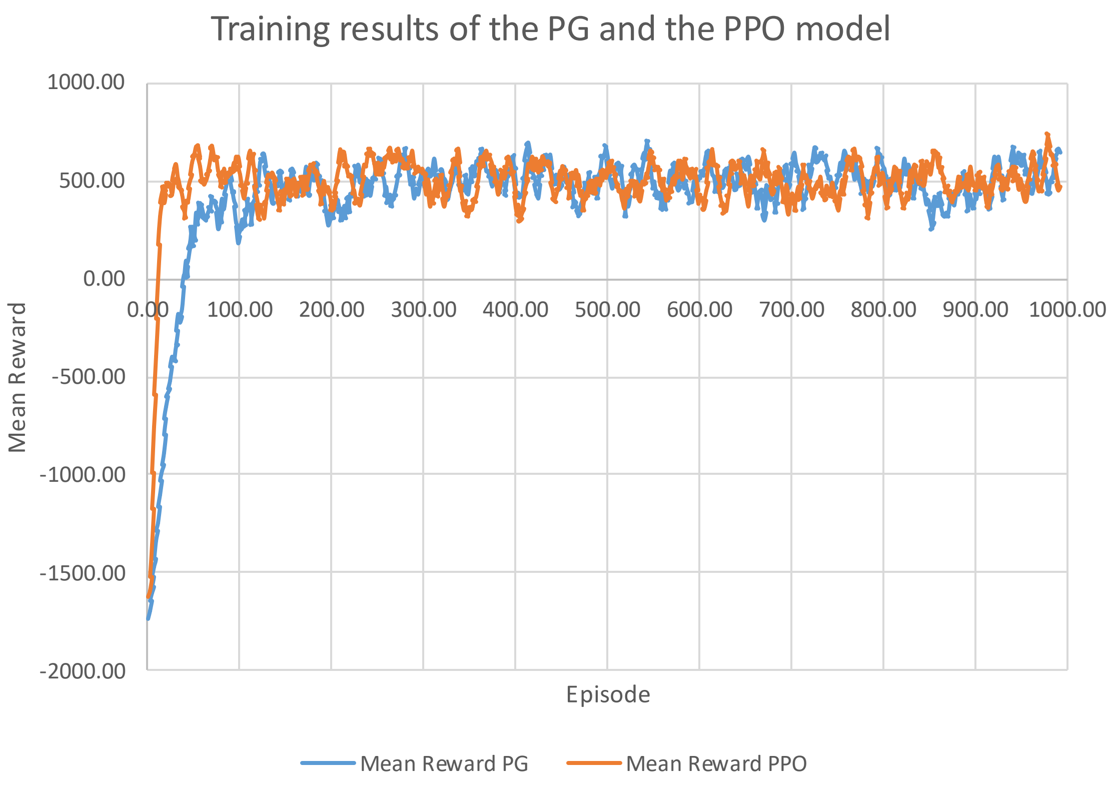
Ostfalia
Hochschule für angewandte Wissenschaften

Training Architecture

- Training Environment:
    - Creating Gym Environment
- Components from Webots [3]:
    - Driver and Supervisor Modules
    - Sensors and Actuators
- RL Framework:
    - Policy-based Algorithms
    - Simulation Feedback Loop

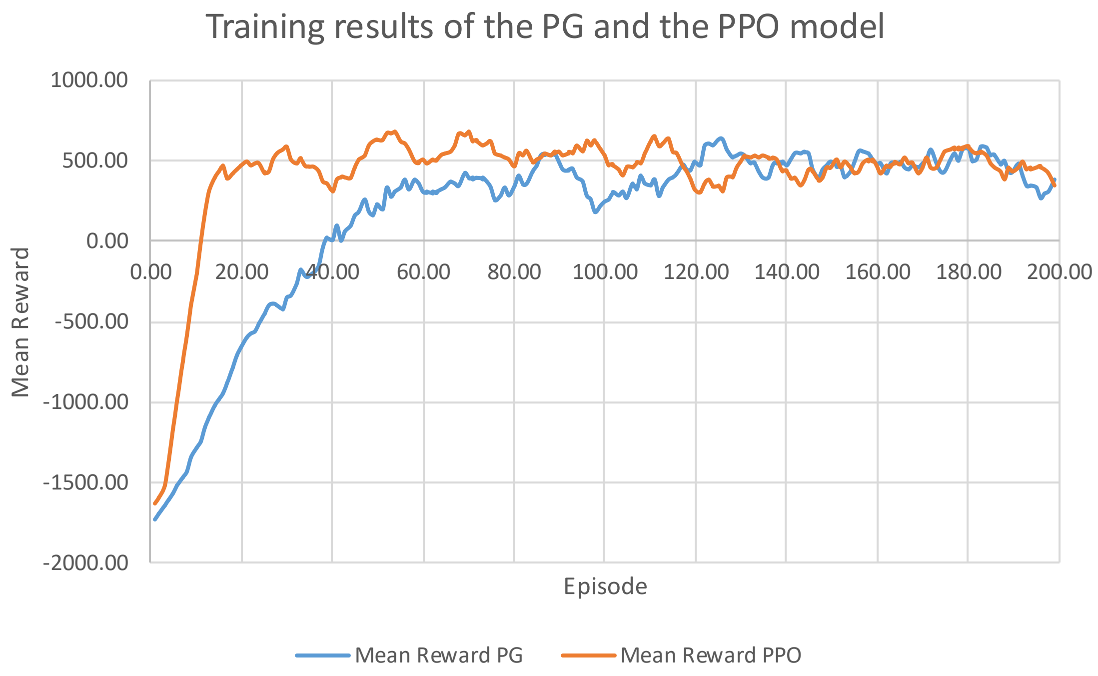


Evaluation of Results



The reward values of both algorithms stabilize at approximately 500

Evaluation of Results



PPO shows faster convergence compared to PG

Evaluation of Results

Algorithm	Wrong behavior or collision / %	AEB Selection / %
PG	1.5	24.85
PPO	0.3	1.0

PPO leads to better overall system reliability and response accuracy compared to PG

Conclusions and Future Directions

- **Key Results**:
  - **Effective Selection**: Both PG and PPO successfully manage ACC and AEB system selections for routine and emergency maneuvers.
  - **Algorithm Performance**:
    - PPO shows faster convergence, achieving stable reward values significantly quicker than PG.
    - PPO maintains a lower error rate (0.3%) in follow-up tests compared to PG (1.5%).
  - **Insights**: Validate the feasibility of RL in automating maneuver-based decision-making for driving.

Conclusions and Future Directions

- **Current Limitations and Future Work**:
  - **Simulation Complexity**: Current simulations are relatively simple and may not fully represent complex real-world driving scenarios.
  - **Sensor Technology**: Emphasize the need for more advanced sensor integration to enhance simulation accuracy and applicability.
  - **Further Developments:**
    - Suggest expanding training environments to include more diverse traffic conditions and overtaking scenarios.
    - Plan to validate and optimize models within the ExerShuttle project in real-world conditions.
  - **Broader Integration**: Advocate for incorporating a wider range of driving behaviors into training models, ensuring comprehensive testing and validation.

Ostfalia Hochschule für angewandte Wissenschaften

References

[1] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, “Policy gradient methods for reinforcement learning with function approximation”, in Advances in Neural Information Processing Systems, S. Solla, T. Leen, and K. Muller, Eds., vol. 12, MIT Press, 1999, pp. 1057–1063.

[2] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms”, arXiv, 2017.

[3] O. Michel, “Webots: Professional mobile robot simulation”, Journal of Advanced Robotics Systems, vol. 1, no. 1, pp. 39–42, 2004.