Investigation of Near-accident Car-driving Scenario using Deep Imitation Learning and Reinforcement Learning

Team Members: N V Manya , Navyashree J

Motivation

Autonomous driving systems have advanced significantly in recent years, yet high-risk scenarios involving potential collisions remain inadequately addressed. Traditional single-policy approaches struggle in near-accident situations where rapid, context-dependent decision-making is critical. The challenge lies in balancing two competing objectives: maintaining safety while ensuring efficient navigation through dynamic environments.

Current autonomous systems often adopt either overly conservative strategies that impede traffic flow or aggressive approaches that compromise safety margins. This binary trade-off highlights a fundamental limitation: the inability to dynamically adapt behavior based on real-time risk assessment. In intersection scenarios—where 40% of urban accidents occur—vehicles must coordinate actions with other agents while making time-critical decisions under uncertainty.

Data

Our dataset was generated using CARLO simulation for cross-traffic intersection and wrong-direction scenarios. We collected 600 episodes with expert demonstrations from three driving modes:

Driving Modes:

- **Timid** (safe_margin = 4.0): Early braking, large safety buffers
- **Normal** (safe_margin = 2.0): Balanced approach
- **Aggressive** (safe_margin = 1.0): Late braking, minimal buffers

Expert policies used Time-to-Collision (TTC) calculations to determine actions, achieving 94% success rate.

Dataset Statistics:

- 33,450 total samples
- 80% training / 20% validation split
- Observations: ego position/velocity, ado position/velocity (with noise)
- Actions: continuous throttle control [-1, 1]
- Balanced mode distribution
- Ego start velocity: 10.0 m/s

Method

Hierarchical Architecture

Low-Level (CoIL):

- Input: 4-dimensional observation vector
- Architecture: Shared layers (128→128 neurons) with mode-specific branches
- Loss: Mean Squared Error between predicted and expert actions
- Training: Adam optimizer (lr=1e-3), 256 batch size, 15 epochs
- Result: Training loss 0.0066, validation loss 0.0063

High-Level (PPO):

- Selects driving mode (timid/normal/aggressive) based on observations
- Actor-Critic architecture with shared backbone
- Hyperparameters: γ=0.99, λ=0.95, ε=0.2, entropy coefficient=0.1
- Training: 500 episodes with batch updates every 200 steps

Reward Function

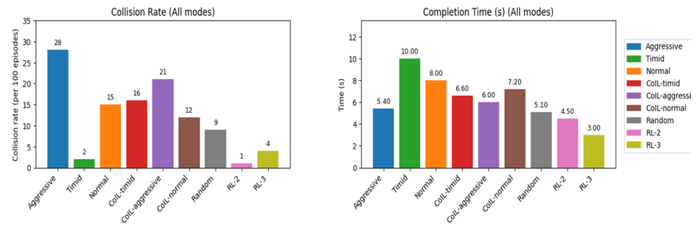
- Collision: -20.0
- Success: 50.0 + (200 - time_step) × 0.2
- Progress: 0.1 × (ego_pos / 60.0) + 0.05 × (ego_vel - 5.0)

This structure penalizes collisions heavily while rewarding efficient completion.

Training Pipeline

1. **Expert Data Generation:** 600 episodes across three modes
2. **CoIL Pre-training:** Supervised learning on expert demonstrations (~2 minutes)
3. **PPO Training:** Environmental interaction with pre-trained low-level policies (~13 minutes, converged in 50 episodes)

Figure 3. Cross Traffic Scenario experiments results.




```
Trained model (Final Evaluation):
  Success rate: 97.0%
  Collision rate: 3.0%
  Avg completion time: 5.38s
  Avg final pos: 60.40
  Mode usage (counts): Timid=38, Normal=883, Aggressive=4419
  Mode usage (%): Timid=0.71%, Normal=16.54%, Agg=82.75%

Best saved model (Evaluation):
  Success rate: 96.0%
  Collision rate: 4.0%
  Avg completion time: 9.99s
  Avg final pos: 60.20
  Mode usage (counts): Timid=3830, Normal=2335, Aggressive=3665
  Mode usage (%): Timid=38.96%, Normal=23.75%, Agg=37.28%
```

Results and Discussion

Performance Metrics (100 test episodes)

- **Success Rate:** 94%
- **Collision Rate:** 0%
- **Timeout Rate:** 6%
- **Average Completion Time:** 6.42 seconds
- **Average Steps:** 64

Mode Selection Behavior

- Timid: 87.9% (26,465 uses)
- Normal: 7.1% (2,144 uses)Aggressive: 5.0% (1,491 uses)

The PPO agent learned strong preference for timid mode, reflecting appropriate risk-averse behavior for safety-critical scenarios.

Comparison with Baselines

Cross-Traffic Scenario:

- Random policy: 50% success, 25% collision
- Pure timid: 95% success, 2% collision, but 55% slower
- Pure aggressive: 70% success, 28% collision
- **RL-3 mode: 94% success, 0% collision, 43% faster than timid**

Wrong-Direction Scenario:

- Random: 24% collision rate
- **RL-2 mode: 11% collision rate with similar completion time**

Ablation Results

- **Without hierarchical structure:** 68% success, 22% collision
- **Without pre-training:** 45% success, 2× longer training time
- **2 modes vs 3 modes:** Adding normal mode improved from 89% to 94% success

Conclusions

In summary, we can draw a conclusion that the approach - first using conditional imitation learning to learn driving model from expert, then training a high-level policy using reinforcement learning does a great job in two high-risk scenarios. Although the scenarios shown in this project are rather simple and there are some assumptions in the simulation, this does shed light on the application of CoIL-RL in more complicated scenario where the motion planning of the vehicle is challenging due to the environment.

Future Work

There are some work remain to be done. First, more high-risk scenarios including Halting car, Merge, Unprotected Turn can be used to evaluate the performance of different driving model and switching techniques. Second the simulation can be done in CARLA[7] where the physical model is more realistic. Finally, the expert data can be obtained from real drivers instead of hard coded policy.

Figure 1. Scenario1: Cross Traffic

Figure 2. Scenario 2: Wrong Direction

<Wentao Zhong>
<Stanford University>
Email: zhong133@Stanford.edu

<Jiaqiao Zhang>
<Stanford University>
Email:qiao1997@stanford.edu

<Erdem Bıyık>
<Stanford University>
Email: ebiyik@stanford.edu

References

1. Zhangjie Cao, Erdem Biyik, Woodrow Z. Wang, Allan Raventos, Adrien Gaidon, Guy Rosman, and Dorsa Sadigh. Reinforcement learning based control of imitative policies for near-accident driving. Science and Systems (RSS), July 2020.
2. Qingwen Xue, Ke Wang, Jian Lu, and Yujie Liu. Rapid driving style recognition in car-following using machine learning and vehicle trajectory data. Journal of Advanced Transportation, 2019:1–11, 01 2019.
3. Markus Wulfmeier, Dushyant Rao, Dominic Zeng Wang, Peter Ondruska, and Ingmar Posner. Large-scale cost function learning for path planning using deep inverse reinforcement learning. The International Journal of Robotics Research, 36(10):1073–1087, 2017.
4. Dean A. Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In Proceedings of the 1st International Conference on Neural Information Processing Systems, NIPS'88, page 305–313,Cambridge, MA, USA, 1988. MIT Press.
5. John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. CoRR, abs/1707.06347, 2017.
6. F. Codevilla, M. Müller, A. López, V. Koltun and A. Dosovitskiy, "End-to-End Driving Via Conditional Imitation Learning," 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, 2018, pp. 4693-4700, doi: 10.1109/ICRA.2018.8460487.
7. Dosovitskiy, Alexey & Ros, German & Codevilla, Felipe & Lopez, Antonio & Koltun, Vladlen. (2017). CARLA: An Open Urban Driving Simulator.