EVALUATING THE ROBUSTNESS OF SAC UNDER DISTRIBUTIONAL SHIFT IN DRIVING DOMAIN

AUTHOR

Lazar Polovina
email: L.Polovina@student.tudelft.nl

SUPERVISORS

Dr. Mustafa Celikok
Dr. Frans Oliehoek

EXAMINER

Dr. Annibale Panichella

INTRODUCTION

Reinforcement Learning (RL) is used for decision-making in domains like driving and finance.
Models for RL are usually trained in stable conditions, which makes algorithm harder to predict when they experience changes in environment.
This study is here to show how SAC behaves under distributional shifts and it focuses on driving domain

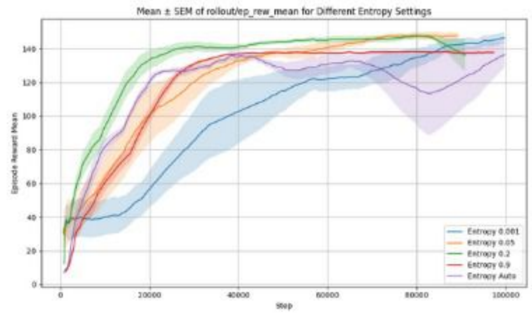
RESEARCH QUESTION

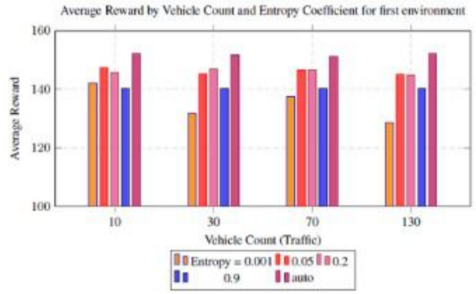
HOW DOES THE PERFORMANCE OF SAC-TRAINED AGENTS DEGRADE UNDER INCREASING DISTRIBUTIONAL SHIFT, AND HOW DOES THIS RELATE TO THE ENTROPY REGULARIZATION COEFFICIENT?

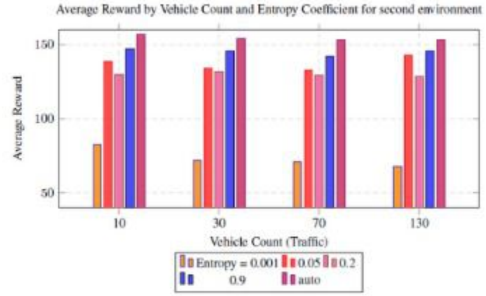
ANALYSIS

SAC Policy Objective

$$J_\pi(\theta) = \mathbb{E}_{s_t \sim D, a_t \sim \pi_\theta} \left[Q(s_t, a_t) - \alpha \log \pi_\theta(a_t | s_t) \right]$$







METHODOLOGY

HighwayEnv driving simulator was used to evaluate SAC under various entropy coefficients—0.001, 0.05, 0.2, 0.9—and automatic tuning. SAC agents were trained under consistent traffic conditions and tested across two environments:

Environment 1: only traffic density varied.

Environment 2: additional coefficients were modified, driver behavior was changed to aggressive and number of lanes was equal to 2.

Each agent was evaluated across 4 traffic densities (10, 30, 70, 130 vehicles) over multiple seeds. Metrics included average reward and crash rate.

RESULTS/FINDINGS

Training Phase: Entropy 0.2 yielded the fastest convergence and highest training rewards. Auto-tuned entropy performed comparably with slightly more variance.

Testing - Environment 1: Auto entropy agent outperformed all others in reward and crash rate. Moderate fixed settings (0.05, 0.2) performed well, while high entropy (0.9) maintained perfect crash safety with moderate rewards

.
Testing - Environment 2: Auto entropy again delivered top results. High entropy (0.9) became more advantageous under complex conditions, while low entropy (0.001) led to unsafe, brittle behavior.

CONCLUSION

Adaptive entropy tuning in SAC enables robust generalization under distributional shift, maintaining high performance and safety across scenarios.

Moderate fixed entropy values can also be effective, but performance is environment-dependent.

Low entropy harms adaptability, while high entropy supports safe exploration in complex domains.

Future Work: Combine entropy tuning with risk-sensitive objectives and apply to more realistic or adversarial driving simulations.