Surrogate Reloaded: LSTM-Based Failure Prediction for Testing DRL Agents

Background

Deep Reinforcement Learning (DRL) agents learn by a trial-and-error approach. A common approach of training agents is by exposing them to many different scenarios (environments). After training, the agents need to be tested if they will be used in non-simulated scenario's. However, testing is hard because it's unknown when the agent will fail. This results in running the agent on many randomized environments, which is costly if the agent takes seconds to minutes each run.

Indago is a state-of-the-art testing framework that finds new failures faster [1]. It reduces failure finding time by training a surrogate model on the DRL training data. This surrogate acts as a proxy by predicting whether an agent would fail in an environment.

The explored surrogate is a Multi-Layer Perceptron (MLP), training on the static initial environment configuration of the agent training data. We explore a Long Short-Term Memory (LSTM) model with the aim of exploiting the temporal aspect within the agent training data.

RQ: How effectively can LSTM surrogate models predict failure configurations for testing DRL agents compared to MLP surrogates?

Methodology

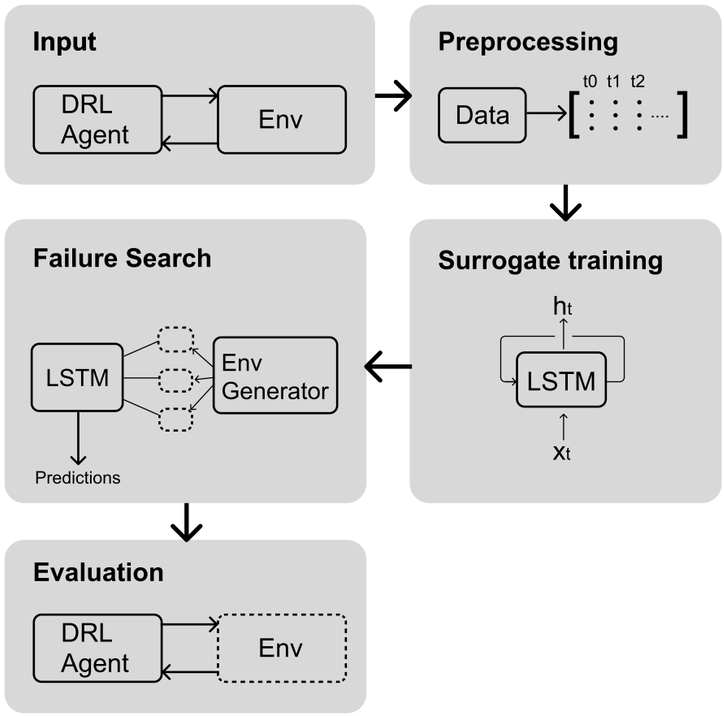

Fig 1. The approach for finding new failures using a surrogate model. The figure is adjusted from the approach in the paper by Biagiola et al. [1].

Results

As can be seen in Table 1, the best LSTM surpasses the MLP in both mean failure count and input diversity. The latter metric quantifies the variety of failure configurations, suggesting the LSTM's capacity to generate a broader range of environmental conditions.

However, the LSTM demonstrates no significant improvement in output diversity, which measures the variation in agent behavior in response to these new environments.

An analysis of the top LSTM models reveals that 11 out of 15 exclusively utilize the initial two timesteps of the agent's training data, implying that longer sequences reduce predictive accuracy.

Future research should prioritize methods for more efficiently capturing the initial state transition or examine the underlying causes for the observed loss of predictive power with longer sequences.

Metric	MLP (baseline)	LSTM (ours)
Mean failures (%)	30	**34.6**
Input coverage (%)	58.0	**99.0**
Input entropy (%)	8.28	**90.39**
Output coverage (%)	**73.55**	68.97
Output entropy (%)	**75.37**	57.59

Table 1. A comparison between the best LSTM and the MLP baseline

[1] Biagiola, M., & Tonella, P. (2023). Testing of Deep Reinforcement Learning Agents with Surrogate Models. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2305.12751

Author: Steven van den Wildenberg (s.j.vandenwildenberg@student.tudelft.nl)
Responsible Professor: Dr. A. Panichella
Supervisors: A.J. Bartlett