Guiding Safe Reinforcement Learning Policies Using Structured Language Constraints

Bharat Prakash, Nicholas Waytowich, Ashwinkumar Ganesan, Tim Oates and Tinoosh Mohsenin.
UMBC, Baltimore
Motivation

- It is important that agents acting in the real world have the capability to understand instructions and warning signs in the form of natural language.

- We present an architecture which can learn mappings between simple structured language constraints and agent behavior.

- This framework is then used to shape rewards and train agents to interpret simple language constraints and act safely.
Architecture

The proposed architecture has two main components.

- **The constraint checker module**: accepts a trajectory segment and a language constraint and outputs a binary label which denotes violations.
- **RL policy module** ($\pi$): accepts constraint, state and outputs action
Methods

• Data Collection:
  • The data to train the constraint checker needs to be collected using human annotators (although automated in our experiment).
  • Each sample consists of a sequence of states, $S_1$ to $S_n$ which corresponds to the agent’s behavior, text constraint and a binary label denoting violation.
• The constraint checker is trained in a supervised fashion from the data.
• During RL training, the constraint checker receives the state sequence and the structured language constraint and outputs a reward $R_c$. 
Experiments and Results
Thank you!