Provable Safe Reinforcement Learning

Advisor: **Bettina Könighofer**

Motivation

Are you interested in Reinforcement Learning?
Are you interested in Logics?
Let's combine them to get the best of both!

As project, almost any learning problem is interesting where safety is an issue and/or the problem involves a lot of planning/thinking ahead. A few examples:

- **FM+RL for PAC-MAN**
  RL: Learns to play PAC-MAN.
  FM: Protects PAC-MAN from getting eaten by ghosts.
  (Similar: 2048, 2 player snake, atari games like flappy brid, seaquest . . .)
- **FM+RL for autonomous driving**
  RL: An autonomous taxi has to learn to drive smoothly and has to visit several points to pick up passengers.
  FM: Finds out the optimal route and avoids accidents.
- **FM+RL for multi-agent systems**
  RL: Learns to perform complex missions.
  FM: Prevents congestions and collisions.
- **FM+RL for smart cities**
  RL: Learns a complex traffic light controller for the entire city. FM: Ensures that emergency vehicles will have a clear road, etc.
- **FM+RL for robocup logistics league** . . .

Goals and Tasks

- Pick a project that excites you.
- Let's figure out together, how FM could assist a learning agent (SAT, SMT, Model Checking, Synthesis).
- Implement everything and play with it.

Literature

- M. Alshiekh et al.
  Safe Reinforcement Learning via Shielding
  Conference on Artificial Intelligence (AAAI-18)
  https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17211
- S. Pranger et al.
  Adaptive Shielding under Uncertainty
  CoRR 2020
  https://arxiv.org/abs/2010.03842

Courses & Deliverables

- ☑ **Introduction to Scientific Working**
  Short report on background
  Short presentation
- ☑ **Bachelor Project**
  Project code and documentation
- ☑ **Bachelor's Thesis**
  Project code
  Thesis
  Final presentation

Recommended if you're studying

☑ CS ☑ ICE ☑ SEM

Prerequisites

- Interest in Logics

Advisor / Contact

bettina.koenighofer@iaik.tugraz.at