Engineering Safe AI Models with Dependently Typed Languages

Contacts
Humbert Fiorino (Humbert.Fiorino@imag.fr) LIG-Marvin
Damien Pellier (Damien.Pellier@imag.fr) LIG-Marvin

Keywords
AI Planning, PDDL language, dependently typed languages, safety, provable AI.

Context
Cobotics is an emerging evolution of industrial robotics where robots are devised to interact with human operators in order to perform collaborative tasks in shared spaces by assisting professional gestures and combining the physical capabilities of the robots with the cognitive abilities of the human operators. In cobotics, the industrial robots are no longer physically isolated from humans (cages, protection barriers, sensors etc.) and Human-Robot Interactions (HRI) are at the core of cobotics.

HRI are mostly based on AI models of (inter)actions hand-coded by robot experts. One possible modelling approach is AI planning [1]. In this case, the (inter)actions models are written in Planning Domain Description Language (PDDL). Given this PDDL model, the world state perceived by its sensors and the task to achieve (assigned by the human operator), the robot is able to compute which action to trigger. The details of this computation based on planning systems are out of the scope of this research project (see [1]). The main point is that the robot decisions to perform this or that action are based on high level symbolic representations encoded in PDDL by error-prone human experts.

Clearly, as HRI involve humans sharing spaces with machines making autonomous decisions, safety has become a major concern for cobot engineers.

Objective
It is well known that debugging and maintaining high level symbolic code like PDDL models is a difficult task. Unlike other modelling languages, PDDL lacks suitable tools for model validation and proofs. The purpose of this research project is to study how to encode PDDL models in provable languages. Due to their logical proximity with PDDL, the languages targeted are languages with dependent types.

Expected results
The expect results are the following:

- A state-of-the-art bibliography on model validation based on type checking,
- A method to encode (a fraction of) PDDL into (a fraction of) a dependently typed language,
- A parser from this dependently typed language to PDDL,
- An evaluation of this parser based on PDDL benchmarks.

Reference

[1] M. Ghallab, D. Nau and P. Traverso, “Automated Planning”, Morgan-Kaufman, 2004.