

Verification of unbounded multi-agent systems via parameterised model checking

Panagiotis Kouvaros, Alessio Lomuscio, Edoardo Pirovano
(Talk given by Alessio Lomuscio)

Department of Computing, Imperial College London, UK

Over the past ten years, there has been growing interest in trying to verify formally the correctness of AI systems. This has been compounded by recent public calls for the development of “responsible” and “verifiable” AI. Indeed, since the development of ever more complex and pervasive AI systems including autonomous vehicles, the need for higher guarantees of correctness for the systems has intensified. Formal verification is one of the techniques used in Computer Science to debug systems and certify their correctness. It is therefore expected that formal methods will contribute to provide guarantees that AI systems behave as intended.

One of the challenges of present AI systems is that often their number of components is unbounded or cannot be predicted at runtime. This is the case, for example, in swarm robotics, intelligent environments, and beyond. In this talk I will share with the audience the highlights of the work conducted by Panagiotis Kouvaros and myself on this subject.

Specifically I will like to start by summarising the key results from [KL16] from 2016. In that paper we study the problem of verifying multi-agent systems where the number of components cannot be determined at design time. We give a semantics that captures parameterised, generic multi-agent systems and identify three notable classes that represent different ways in which the agents may interact among themselves and with the environment. While the verification problem is undecidable in general, we put forward cutoff procedures for the classes identified. The method is based on the existence of a simulation between the templates for the agents and the template for the environment in the system. We show that the cutoff identification procedures as well as the general algorithms that we propose are sound; for one class we show the decidability of the verification problem and present a complete cutoff procedure. We report experimental results obtained on MCMAS-P, a novel model checker implementing the parameterised model checking methodologies here devised.

I will carry on by summarising some recent work from [KL17] where we develop a technique to evaluate the fault-tolerance of a multi-agent system whose number of agents is unknown at design time. In this work we inject a variety of non-ideal behaviours, or faults, studied in the safety-analysis literature, into the abstract agent templates. We define the parameterised fault-tolerance problem as the decision problem of establishing whether any concrete system, in which the ratio of faulty versus non-faulty agents meets a given ratio, satisfies a given temporal-epistemic specification. We also put forward a sound and complete technique for solving the problem for the semantical set-up considered. As above,

I will report on an implementation and a case study identifying the threshold under which the alpha swarm aggregation algorithm is robust to faults against its temporal-epistemic specifications. This work has recently been extended in [KLP18] where a synthesis procedure for determining the tolerance ratio for a swarm is given.

References

- [KL16] P. Kouvaros and A. Lomuscio. Parameterised verification for multi-agent systems. *Artificial Intelligence*, 234:152–189, 2016.
- [KL17] P. Kouvaros and A. Lomuscio. Verifying fault-tolerance in parameterised multi-agent systems. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI17)*, pages 288–294. AAAI Press, 2017.
- [KLP18] P. Kouvaros, A. Lomuscio, and E. Pirovano. Symbolic synthesis of fault-tolerance ratios in parameterised multi-agent systems. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence and 23rd European Conference on Artificial Intelligence (IJCAI-ECAI18)*, pages 324–330. IJCAI, 2018.