Formal Verification of Parameterised Neural-symbolic Multi-agent Systems (Extended Abstract)

Panagiotis Kouvaros[1,*], Elena Botoeva[2] and Cosmo De Bonis-Campbell[2]

[1]Department of Information Technologies, University of Limassol, Cyprus
[2]School of Computing, University of Kent, UK

Abstract

We study the problem of verifying multi-agent systems composed of arbitrarily many neural-symbolic agents. We introduce a novel parameterised model, where the parameter denotes the number of agents in the system, each homogeneously constructed from an agent template equipped with a neural network-based perception unit and a traditionally programmed action selection mechanism. We define the verification and emergence identification problems for these models against a bounded fragment of CTL. We put forward an abstraction methodology that enables us to recast both problems to the problem of checking Neural Interpreted Systems with a bounded number of agents. We present an implementation and discuss experimental results obtained using a social dilemma game based on guarding.

Keywords

Neural-symbolic, Multi-agent Systems, Parameterised Verification

1. Introduction

Safety concerns stemming from the increasing development of Multi-agent Systems (MAS) have been put under mathematical scrutiny by automated methods that ascertain their correct behaviour. Verification methods based on SAT and BDDs [1, 2] have resulted in push-button engines such as Verics, MCK and MCMAS [3, 4, 5]. In conjunction with increasingly sophisticated state-space reduction techniques, such as predicate abstraction [6] and partial order reductions [7], the verifiers have been able to scale to the analysis of systems with very large state spaces.

While the different methods target the provision of effective solutions to different types of analyses, e.g., fast search for counterexamples [8] as opposed to fast correctness proofs [9], and for different classes of MAS, e.g., MAS defined over infinite-state as opposed to finite-state variables [6], all methods make two fundamental assumptions. The first is that the MAS under analysis is composed of a known number of agents specified at design time. The second is that the agents composing the MAS are specified using traditional programming languages.

The approaches cannot therefore be used to verify important classes of MAS, where either the systems have arbitrarily many participants or the agents are endowed with machine learning components. The former class of systems includes open systems, where agents can join and leave the system at runtime, and applications designed irrespective of the number of participants, such as robot swarms. The latter class comprises forthcoming neural-symbolic applications such as autonomous vehicles.

More recent methods have addressed the verification of systems with an unbounded number of constituents [10, 11]. While the various methods in the area address different communication primitives for the agents, most of them rely on abstractions whereby the unbounded verification problem is reduced to analysing a finite state-space. The resulting techniques formed the backbone of various formal reasoners such as those targeting fault-tolerance [12] and data-aware systems [13].

In a different line of work, verification methods for MASs comprising agents with neural components were put forward [14, 15]. To deal with the real-valued operational domain of the neural network models, the methods recast verification queries for bounded properties into Mixed-Integer-Linear-Programming.

LNSAI 2024: First International Workshop on Logical Foundations of Neuro-Symbolic AI, August 05, 2024, Jeju, South Korea
*Corresponding author.
pkouvaros@uol.ac.cy (P. Kouvaros); e.botoeva@kent.ac.uk (E. Botoeva); cd586@kent.ac.uk (C. D. Bonis-Campbell)

While these lines of work independently tackle unbounded and neural-symbolic MAS, none of the underlying methods can be used for analysis of systems that are both unbounded and neural-symbolic. In this paper we overcome this limitation. Specifically, we introduce *Parameterised Neural Interpreted Systems (PNIS)*, a formal model for modelling unbounded neural-symbolic MAS. We develop an abstraction methodology for PNIS whereby we derive sound and complete procedures for the verification and emergence identification problems with respect to bounded universal and existential CTL formulae. We utilise an implementation of these procedures to analyse a social dilemma scenario.

Related Work. The contribution is related to the two lines of work discussed above, namely parameterised verification and verification for neural-symbolic MAS. Previous models in parameterised verification targeted either arbitrarily many agents operating in fixed environments [10, 11, 16] or a fixed number of agents living in environments of arbitrary size [17]. None of these models include neural components. Systems comprising homogeneous agents were also analysed within the framework of Alternating-time Temporal Logic but in a fixed, non-parameterised and purely symbolic setting [18].

Existing verification methods for MAS with neural components [14, 15] take as input systems with a known number of agents. The main theoretical finding of this work is that verification for an unbounded number of agents can be reduced to the verification of (abstract) systems with a bounded number of agents.

2. Parameterised Neural-symbolic Interpreted Systems

Interpreted systems are a standard semantics for describing multi-agent systems [19]. They provide a natural setup to interpret specifications in a variety of languages such as temporal-epistemic logic. Parameterised interpreted systems is a parametric extension of interpreted systems put forward to reason about unbounded multi-agent systems [20]. The parameter in a system of this kind denotes the number of agents composing the system, each homogeneously constructed from an agent template. We extend parameterised interpreted systems to PArameterised Neural-symbOlic interpreted Systems (PANoS), where the template for the agents is not purely symbolic but (i) comprises a perception mechanism that is implemented via neural networks, (ii) it is coupled with a symbolic action mechanism. This neural-symbolic treatment of the agents follows the Neural Interpreted Systems (NIS) model from [8]. Differently from PANoS however, NIS are limited to standard non-parametric systems with a pre-defined number of agents.

A PANoS $\mathcal{S}$ consists of the descriptions of an agent template, from which an unbounded number of concrete agents may be constructed, and of an environment in which the agents operate. A PANoS gives a parametric description of an unbounded collection of concrete NIS. In particular, for any value $n \geq 1$ of the parameter, the concrete system $\mathcal{S}(n)$ composes n copies of the agent template with the environment following the agents' composition presented in NIS [21]. Each concrete system $\mathcal{S}(n)$ is associated with a temporal model $\mathcal{M}_{\mathcal{S}(n)}$, as standard in interpreted systems [19]. The model can be used to interpret properties in an indexed and bounded variant of Computation Tree Logic (CTL), henceforth bICTL. The logic (i) introduces indexed atomic propositions that are quantified over the agents of the concrete system that a formula in question is evaluated on; and (ii) permits only the construction of formulae whose evaluation can be realised on paths of bounded lengths. The former extends CTL by allowing the formulation of properties irrespective of the concrete system on which they are evaluated. The latter restricts CTL to bounded formulae [21]. In this work we answer the parameterised verification and emergence identification problems for PANoS and bICTL.

Definition 1 (Parameterised verification problem). *Given a PANoS $\mathcal{S}$ and a* bICTL formula φ, determine whether φ holds on every concrete system $\mathcal{S}(n)$ instantiated from $\mathcal{S}$.

Definition 2 (Emergence identification problem). *Given a PANoS $\mathcal{S}$ and a* bICTL formula φ, compute whether there is an emergence threshold $th \in \mathbb{N}$ such that φ holds on every concrete system $\mathcal{S}(n)$ instantiated from $\mathcal{S}$ with $n \geq th$ agents.

3. Parameterised Verification Procedure

We put forward procedures for solving the parameterised verification and emergence identification problems. The procedures recast the parameterised verification and emergence identification problems for a PANoS $\mathcal{S}$ and a bICTL formula φ to a number of (standard) verification problems for abstract and concrete NIS. We show that the satisfaction status of φ on these systems determines the satisfaction and existence of emergence thresholds for the formula on $\mathcal{S}$. This enables us to use previously established methodologies for the verification of NIS against bCTL [14] to analyse PANoS.

Towards this we construct the zero-one abstraction of the systems generated from $\mathcal{S}$. The zero-one abstraction is a NIS comprising a zero-one agent, which is an abstraction for arbitrarily many concrete agents, m concrete agents, where m is the number of index variables present in the specification to check, and the environment. In other words, the zero-one agent in this abstract NIS encodes how an arbitrary number of agents may interfere with the temporal evolution of m concrete agents. A correspondence can then be established between the abstract model and the concrete models. We show in particular that (i) the abstract model simulates every concrete model with at least $m + 1$ agents; (ii) there is always a concrete model with a sufficient number of agents that simulates the abstract model up to a bounded number of time-steps; (iii) a concrete model always simulates a smaller concrete model.

These results enable the derivation of procedures for solving the parameterised verification and emergence identification problems. In the case of universal properties, verification can be conducted by constructing and checking the abstract model and the concrete model with m agents. The specification is satisfied by the abstract and concrete models if and only if the specification is satisfied in general for any number of agents. The satisfaction of the specification by the abstract model is also connected by biconditional implication with the existence of an emergent threshold for the specification. In particular, $m + 1$ is an emergence threshold if the abstract model satisfies the specification; otherwise, there is no emergence threshold.

For the case of existential properties, verification can be performed by enumerating all concrete models, identifying the smallest one that simulates the abstract model up to the temporal depth of the specification in question, and checking all concrete models up to the simulating one. The specification is satisfied by all these concrete models if and only if the specification is satisfied in general for any number of agents. The satisfaction of the specification by the concrete model that simulates the abstract model is also connected by biconditional implication with the existence of an emergent threshold for the specification. In particular, if n is the size of the simulating model, then n is an emergence threshold if the simulating model satisfies the specification; otherwise, there is no emergence threshold.

4. Evaluation

We present an evaluation of the parameterised verification procedures on a guarding game, an instance of a social dilemma game characterised by tension between individual and collective rationality [22]. We used the parameterised verification procedures to verify existential and universal specifications pertaining to a colony of agents surviving after a number of time steps. We used the VENMAS toolkit [14] for checking the concrete and abstract systems prescribed by the procedures. The results conclude that (i) for $k \in \{3, 4, 5\}$ time steps, there need to be at least 3 agents present in the colony to ensure temporal evolution paths along which the colony is alive (i.e., 3 is an emergence threshold for the existential viability property); (ii) there is always a temporal evolution path where the colony is no longer alive (i.e., the universal viability property is not satisfied by all concrete systems). The parameterised verification procedures established these results by checking only the abstract system and the concrete systems with up to 3 agents. In contrast, traditional verification approaches would need to verify the properties in question for any number of agents in the system. This is computationally intractable as verification times grow exponentially with the number of agents in the system.

5. Conclusions

Advances in interconnectivity of autonomous services and machine learning fuel the development of MAS with arbitrarily many neural-symbolic components, thereby creating a pressing need for their verification. Towards addressing this need, in this paper we put forward a number of automated procedures for the formal analysis of parameterised, neural-symbolic MAS. The procedures enable conclusions to be drawn on the satisfaction of temporal properties irrespective of the number of agents composing the MAS. They can additionally identify emergence thresholds expressing sufficient conditions on the of number agents for a property to be realised. The theoretical results have driven the implementation of a parameterised, neural-symbolic verifier, which we used to reason about a simple social dilemma game. More generally, the techniques here developed can be used to analyse properties of policies learned to deal with real-life challenges that come in the form of collective risk dilemmas, as well as properties in swarm scenarios and open systems in general.

In future work we target the development of parameterised methods for interleaved semantics for neural-symbolic MAS and strategic properties.

References

[1] M. Kacprzak, A. Lomuscio, W. Penczek, Verification of multiagent systems via unbounded model checking, in: Proceedings of the 3rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS04), ACM, 2004, pp. 638–645.

[2] F. Raimondi, A. Lomuscio, Automatic verification of multi-agent systems by model checking via OBDDs, Journal of Applied Logic 5 (2005) 235–251.

[3] P. Gammie, R. van der Meyden, MCK: Model checking the logic of knowledge, in: Proceedings of 16th International Conference on Computer Aided Verification (CAV04), volume 3114 of *Lecture Notes in Computer Science*, Springer, 2004, pp. 479–483.

[4] M. Kacprzak, W. Nabialek, A. Niewiadomski, W. Penczek, A. Pólrola, M. Szreter, B. Woźna, A. Zbrzezny, VerICS 2007 - a model checker for knowledge and real-time, Fundamenta Informaticae 85 (2008) 313–328.

[5] A. Lomuscio, H. Qu, F. Raimondi, MCMAS: A model checker for the verification of multi-agent systems, Software Tools for Technology Transfer 19 (2017) 9–30.

[6] A. Lomuscio, J. Michaliszyn, Verifying multi-agent systems by model checking three-valued abstractions, in: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, 2015, pp. 189–198.

[7] W. Jamroga, W. Penczek, T. Sidoruk, P. Dembiński, A. Mazurkiewicz, Towards partial order reductions for strategic ability, Journal of Artificial Intelligence Research 68 (2020) 817–850.

[8] W. Penczek, A. Lomuscio, Verifying epistemic properties of multi-agent systems via bounded model checking, in: Proceedings of the second international joint conference on Autonomous agents and multiagent systems, 2003, pp. 209–216.

[9] T. Ball, O. Kupferman, An abstraction-refinement framework for multi-agent systems, in: 21st Annual IEEE Symposium on Logic in Computer Science (LICS'06), IEEE, 2006, pp. 379–388.

[10] P. Kouvaros, A. Lomuscio, Parameterised verification for multi-agent systems, Artificial Intelligence 234 (2016) 152–189.

[11] P. Kouvaros, A. Lomuscio, Parameterised model checking for alternating-time temporal logic, in: Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI16), IOS Press, 2016, pp. 1230–1238.

[12] P. Kouvaros, A. Lomuscio, E. Pirovano, Symbolic synthesis of fault-tolerance ratios in parameterised multi-agent systems, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI18), AAAI Press, 2018, pp. 324–330.

[13] F. Belardinelli, P. Kouvaros, A. Lomuscio, Parameterised verification of data-aware multi-agent

systems, in: Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI17), AAAI Press, 2017, pp. 98–104.
[14] M. Akintunde, E. Botoeva, P. Kouvaros, A. Lomuscio, Verifying strategic abilities of neural-symbolic multi-agent systems, in: Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR20), AAAI Press, 2020, pp. 22–32.
[15] M. Akintunde, E. Botoeva, P. Kouvaros, A. Lomuscio, Formal verification of neural agents in non-deterministic environments, Journal of Autonomous Agents and Multi-Agent Systems 36 (2022).
[16] P. Felli, A. Gianola, M. Montali, Smt-based safety verification of parameterised multi-agent systems, arXiv preprint arXiv:2008.04774 (2020).
[17] B. Aminof, A. Murano, S. Rubin, F. Zuleger, Automatic verification of multi-agent systems in parameterised grid-environments, in: Proceedings of the 2016 international conference on autonomous agents & multiagent systems, 2016, pp. 1190–1199.
[18] T. Pedersen, S. Dyrkolbotn, Agents homogeneous: A procedurally anonymous semantics characterizing the homogeneous fragment of atl, in: International Conference on Principles and Practice of Multi-Agent Systems, Springer, 2013, pp. 245–259.
[19] R. Fagin, J. Y. Halpern, M. Vardi, A nonstandard approach to the logical omniscience problem, Artificial Intelligence 79 (1995).
[20] P. Kouvaros, A. Lomuscio, Verifying emergent properties of swarms, in: Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI15), AAAI Press, 2015, pp. 1083–1089.
[21] M. Akintunde, E. Botoeva, P. Kouvaros, A. Lomuscio, Formal verification of neural agents in non-deterministic environments, in: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS20), ACM, 2020.
[22] P. A. Van Lange, J. Joireman, C. D. Parks, E. Van Dijk, The psychology of social dilemmas: A review, Organizational Behavior and Human Decision Processes 120 (2013) 125–141. doi:https://doi.org/10.1016/j.obhdp.2012.11.003, social Dilemmas.