View Reviews

Paper ID
6775

Paper Title
Formal Verification of Parameterised Neural-symbolic Multi-agent Systems

Track Name
IJCAI2024 Main Track

Reviewer #1

Questions

1. Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
The authors investigate formal verification of neural-symbolic multi-agent systems. In particular, they propose an abstraction-based method to verify systems with unboundedly many neuro-symbolic agents against next-step branching-time specifications.

2. What are the main strengths of the paper? Please focus on novelty (how novel are the concepts, problems addressed, or methods introduced in the paper), soundness (is the paper technically sound), clarity (is the paper well-organized and clearly written), and significance (comment on the likely impact the paper may have on the AI research community as a whole or on its own sub-field).
The subject is very relevant and timely, and the idea interesting. The new contribution is solid (though with important limitations that should be mentioned more explicitly), and the experimental evaluation very promising. In particular, the abstraction by zero-one agent is very interesting, and seems a powerful tool.

3. What are the weaknesses of the paper? Please focus on novelty, soundness, clarity, and significance.
On the down side:

- The property specification language is very restricted - essentially, it is the next-time fragment of CTL (the authors allow formulas of type A(X^k)\phi and E(X^k)\phi, but those can be equivalently expressed by (AX)^k\phi and (EX)^k\phi, respectively).

This does not mean that the work is uninteresting, but I believe the authors should emphasize the limitation more clearly.

- If I understand correctly, the neural part is given as a black box (e.g., a genuine neural net, external to the symbolic formal model of the system). If that is the case, nothing in the proposal is specific to neuro-symbolic systems. Instead, it is a verification algorithm for models partially given by an external black box, and with applications to neuro-symbolic systems. This is not a bad thing (if anything, the scope of the method is broader than the authors claim), but the way the authors present it somewhat obscures its real nature.

- There is no comparison to literature on parameterized verification of CTL (see questions for rebuttal).

4. Please carefully list the questions that you would like the authors to answer during the author-feedback period. List only questions about specific issues that 1) could directly influence your evaluation of the paper and 2) do not require new results.

1. How is obs given in practice? As a concrete ML model (e.g., neural net)? Is it assumed to be deterministic?

2. bICTL allows no negation and no existentially quantified variables. Please explain why.

3. Can you comment on the relationship of your proposal to the literature on parameterized temporal verification of MAS, e.g.,
Benjamin Aminof, Aniello Murano, Sasha Rubin, Florian Zuleger: Automatic Verification of Multi-Agent Systems in Parameterised Grid-Environments. AAMAS 2016: 1190-1199
and later works by the same authors?

4. Do I understand correctly that Table 1 reports non-parameterized verification, and Table 2 parameterized verification?

5. You show no experimental evaluation of emergence identification, right? Or am I missing something?

AFTER THE REBUTTAL

I have seen the authors' responses, and I find them convincing.

5. Are the results of this paper easily reproducible? (Please refer to our reproducibility guidelines https://ijcai24.org/reproducibility/).
CONVINCING : I am convinced that the obtained results can be reproduced, possibly with some effort. Key resources (e.g., proofs, code, data) are already available, will be made available upon acceptance, or good reasons as to why they are not (e.g., proprietary data or code) are reported in the paper. Key details (e.g., proofs, experimental setup) are sufficiently well described but their exact recovery may require some work.

7. Overall Assessment
Clear Accept. Interesting. A very good submission. I learned a lot from this paper. I vote and argue for acceptance. (Top 50% of accepted IJCAI papers.)

8. Justify your score in a few lines. Please focus on novelty, soundness, clarity, significance, and credibility with respect to reproducibility.
see above

Minor comments:

- p.2, right column, l.-4: mention what t and e are

- What you assume is a synchronous model of a homogeneous multi-agent system (i.e., all the agents have the same structure). There has been much work on the logical properties of such MAS, e.g.,
Pedersen, T., Dyrkolbotn, S.K. (2013). Agents Homogeneous: A Procedurally Anonymous Semantics Characterizing the Homogeneous Fragment of ATL. In: PRIMA 2013: Principles and Practice of Multi-Agent Systems. PRIMA 2013. Lecture Notes in Computer Science, vol 8291.
Please discuss the connection briefly.

- p.4, left column, l.17 & ff, "is in the scope of either a universal or an existential quantifier": you do not allow for existential variable quantification here

- p.4, left column, l.-13: I think it should be a instead of i

- p.4, Def.11: You might emphasize that parameterized verification is a decision problem, but emergence identification is a function problem.

Reviewer #2

Questions

1. Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).

Parameterized verification is an important topic in model checking and formal verification. It concerns about systems consisting of templated-objects, and such systems may have (potentially) infinitely many states, such as (running of) network protocols, and/or multi-agent systems.

In this paper, against neural-symbolic multi-agent systems, the authors present a new verification framework, in such framework, a parameterized neural-symbolic system is described as a PNIS, which consists of the description of the agent-template, the description of the environment, together with a proposition-labeling function. In this definition, agents can be instantiated into arbitrarily many agents, each is equipped with local states, an observation function, as well as a local state transition function.

Subsequently, a variable of CTL, called bICTL is presented to be the specification language, this logic employs a tuple being of the form (p, i) as atomic formulas, and it is closed under Boolean operations like conjunction and disjunction, it also involves temporal connectives like EX and AX coordinating with some concrete step, more importantly, it allows quantifiers over the indices of agents.

Then, a verification procedure of PNIS against bICTL is put forward. Main results of the model checking algorithm are yielded by a series of theorems (i.e., Thm. 1 - Thm. 4). By proposing the notion of zero-one abstraction and bounded simulation, the authors show how to detect the emergence threshold and how to throughly solve the parameterized model checking problem.

Finally, the proposed approach is experimentally evaluated.

2. What are the main strengths of the paper? Please focus on novelty (how novel are the concepts, problems addressed, or methods introduced in the paper), soundness (is the paper technically sound), clarity (is the paper well-organized and clearly written), and significance (comment on the likely impact the paper may have on the AI research community as a whole or on its own sub-field).

First and foremost, the goal of this paper is to solve an interesting and important problem. The authors present a systematic model checking framework against parameterized neural-symbolic systems (modeled by PNIS), the importance of such

systems is daily increasing. In this paper, the authors present a novel approach to detecting the emergence threshold of parameterized model checking, the approach is interesting, and I believe that might be heuristic to model checking for other unbounded systems.

3. What are the weaknesses of the paper? Please focus on novelty, soundness, clarity, and significance.

First, I find that the expressiveness of the newly presented specification language (namely, bICTL) might be rather limited: From the temporal perspective, it has only “Next”, but no “Until” operator, and fixpoint- or recursion-operator is not employed. Therefore, is it essentially first-order modal-logic, I wonder if it can express any “complex-enough” temporal properties?

Second, no complexity analysis of the algorithm is given. According to the experimental results, it seems that the scalability of the model checking algorithm is not satisfactory: it only works w.r.t. small cases.

Lastly, though the newly presented system is very flexible in describing, yet it is quite involved and complicated in syntactic definition.

4. Please carefully list the questions that you would like the authors to answer during the author-feedback period. List only questions about specific issues that 1) could directly influence your evaluation of the paper and 2) do not require new results.

Q1: in P4, C1, Def.8, when defining satisfaction of the atoms (q, a), ‘a’ does not appear in the definition, should ‘i’ be ‘a’ here?

Q2: What is the complexity of PNIS model checking against bICTL? e.g. PSPACE-hard, EXP-hard? Of course, this question is optional to answer in the presence, I just intend to remind the importance of this problem. You might want to work on it in the future.

Q3: To me, the specification language is a bit too simple. My third question is: what is the consequence if we add some additional operator into the logic (for example, if AU and EU (until-operators) are employed)? Does it lead to undecidability to the verification? The authors needn’t provide a rigorous answer presently, but it is appropriate to mention something about this question in the feedback.

5. Are the results of this paper easily reproducible? (Please refer to our reproducibility guidelines https://ijcai24.org/reproducibility/).

CONVINCING : I am convinced that the obtained results can be reproduced, possibly with some effort. Key resources (e.g., proofs, code, data) are already available, will be made available upon acceptance, or good reasons as to why they are not (e.g., proprietary data or code) are reported in the paper. Key details (e.g., proofs, experimental setup) are sufficiently well described but their exact recovery may require some work.

7. Overall Assessment
Weak Accept. Useful. A good paper. The results and insights will benefit the field. I believe it should be accepted.

8. Justify your score in a few lines. Please focus on novelty, soundness, clarity, significance, and credibility with respect to reproducibility.
I now justify the reason why I suggest the weak acceptance of the paper.

First of all, this paper intends to tackle an interesting problem, namely, parameterized model checking of neural-symbolic system, and the problem itself is important in verification.
The proposed approach is novel and sound, and it might be referred to other parameter model checking problems.

Nevertheless, it is not satisfactory for the absence of complexity analysis. Meanwhile, as I have mentioned, the modeling approach is too complicated, whereas the specification language is not so expressive. In addition, the experimental results also witness the high cost of the presented approach: the scalability of such verification is not satisfactory, and it can hardly deal with large-scale models.

Reviewer #5

Questions

1. Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
The paper considers a model of MAS in which each agent has local states and transitions, and an observation function is implemented as a neural network. A parameterized neural interpreted system (PNIS) is a composition of such agents, where the number of agents is parameterized. For a specification, the paper considers a variant of CTL in which atomic propositions care indexed by agents and

permits the construction of formula that can be realized on paths of bounded lengths (biCTL). The paper studies two problems: verification of PNIS against a biCTL formula, and the problem of finding a cutoff number of agents after which the formula is valid (called “emergence threshold”). The paper develops a sound and complete framework for universal and existential biCTL formula. Experiments are conducted on a small example.

2. What are the main strengths of the paper? Please focus on novelty (how novel are the concepts, problems addressed, or methods introduced in the paper), soundness (is the paper technically sound), clarity (is the paper well-organized and clearly written), and significance (comment on the likely impact the paper may have on the AI research community as a whole or on its own sub-field).

I find it natural and nice to parameterize the number of agents.
The emergence question is interesting, in my opinion.

3. What are the weaknesses of the paper? Please focus on novelty, soundness, clarity, and significance.

I found the paper notation heavy.
I am not sure how significant or novel the results are.
It's not clear to me whether NN are inherent in the model. They seem to me to be simply a source of non-determinism.

4. Please carefully list the questions that you would like the authors to answer during the author-feedback period. List only questions about specific issues that 1) could directly influence your evaluation of the paper and 2) do not require new results.

n/a

5. Are the results of this paper easily reproducible? (Please refer to our reproducibility guidelines https://ijcai24.org/reproducibility/).

CREDIBLE: I believe that the obtained results can, in principle, be reproduced. Even though key resources (e.g., proofs, code, data) are unavailable at this point, the key details (e.g., proof sketches, experimental setup) are sufficiently well described for an expert to confidently reproduce the main results, if given access to the missing resources.

7. Overall Assessment

Borderline Accept. Marginally above the acceptance threshold. Technically correct, but not particularly exciting or inspiring. Could be accepted more or less in its current form. Not a big loss if it is not included in the program. Please use sparingly.

8. Justify your score in a few lines. Please focus on novelty, soundness, clarity, significance, and credibility with respect to reproducibility.

The results do not seem very significant or novel to me. The presentation is good, but I found the paper notation heavy.