Formalization and Implementation of Multiple Solvers Protocols for Artificial Intelligence verification and validation

Keywords: Formal Verification, Neural Network, Why3, CAISAR

Institution

The French Alternative Energies and Atomic Energy Commission (CEA) is a key player in research, development, and innovation. Drawing on the widely acknowledged expertise gained by its 16,000 staff spanned over 9 research centers with a budget of 4.1 billion Euros, CEA actively participates in more than 400 European collaborative projects with numerous academic (notably as a member of Paris-Saclay University) and industrial partners. Within the CEA Technological Research Division, the CEA List institute addresses the challenges coming from smart digital systems.

Among other activities, CEA List's Software Safety and Security Laboratory (LSL) research teams design and implement automated analysis in order to make software systems more trustworthy, to exhaustively detect their vulnerabilities, to guarantee conformity to their specifications, and to accelerate their certification. Recently the field of activity of the laboratory has been extended to artificial intelligence safety and security verification.

Context

Critical systems are the next frontier for data-based Artificial Intelligence (AI) systems. Although their remarkable performance in a profusion of fields, deep neural networks lack the necessary guarantees to be embedded in critical systems. Defining the conformity of an AI system regarding a specification is an open research question (Mattioli et al. 2021). Specifically, formal verification is still struggling to be applied to realistic AI programs. The main issues are:

- the NP-completeness of formal neural network analysis (Katz et al. 2017)
- the lack of scalability of existing tools (SAT and SMT solvers, abstract interpretation. . .)
- the difficulty to formally specify the input distribution and specifications to assert the safety of an AI system (Girard-Satabin et al. 2020)

Over the last five years, those problems were investigated, leading to the birth (and death) of a profusion of tools and techniques (see for instance ERAN (Singh, Gehr, Mirman, et al. 2018), CROWN (Wang et al. 2021), Reluplex (Katz et al. 2017) and its successor Marabou (Katz et al. 2019)). Those tools use vastly different techniques (Satisfiability Modulo Theory, Constraint Programming, Abstract interpretation) and live in isolation. There is currently little work existing on using the result of one tool to inform the other, despite this approach being at the hearth of various verification tools such as the award-winning Why3 verification environment, or the C static code analyzer Frama-C (Baudin et al. 2021), also developped at the lab. CAISAR is an AI verification platform under active development at the lab that aims to tackle those issues. Its goal is to provide a prover-agnostic way to certify the AI components of a system.

Objectives

CAISAR currently supports sending proof tasks to each solver separately and provide a synthesis of their answers. Successful approaches to artificial intelligence verification in the literature involve combining multiple solver technologies to further refine the overall result (Singh, Gehr, Püschel, et al. 2018). More generally, it is common in formal verification to combine results of several methods. An illustrative example is abstract interpretation, which combines and refine the answer of several domains reduced product. CAISAR aims to bridge several approaches under an unified way to provide an AI-system level certification: it thus needs ways to communicate between provers and combine their answers. This internship can be described by the following goals:

- familiarization with the CAISAR platform
- state-of-the art study of current cross-solvers best practices
- prototyping examples of communications between solvers

- testing the proposed approaches against state-of-the-art benchmarks, for instance the VNN-COMP (Bak, Liu, and Johnson 2021)
- optionnaly, contributing to visualisation features

Qualifications

The candidate will work at the crossroads of formal verification and artificial intelligence. As it is not realistic to be expert in both fields, we encourage candidates that do not meet the full qualification requirements to apply nonetheless.

- **Minimal**
  - Master student or equivalent (2nd/3rd engineering school year) in computer science
  - notions of AI and neural networks
  - ability to work in a team, some knowledge of version control
- **Preferred**
  - knowledge of OCaml
  - knowledge of formal verification in general, of SMT solving in particular
  - knowledge of Why3

Characteristics

The candidate will be monitored by two research engineers of the team.

- **Duration:** 5 to 6 months from early 2023
- **Location:** CEA Nano-INNOV, Paris-Saclay Campus, France
- **Compensation:**
  - €700 to €1300 monthly stipend (determined by CEA compensation grids)
  - maximum €229 housing and travel expense monthly allowance (in case a relocation is needed)
  - CEA buses in Paris region and 75% refund of transit pass
  - subsidized lunches
  - 2 days of remote work

Application

If you are interested in this internship, please send to the contact persons an application containing:

- your resume;
- a cover letter indicating how your curriculum and experience match the qualifications expected and how you would plan to contribute to the project;
- your bachelor and master 1 transcripts;
- the contact details of two persons (at least one academic) who can be contacted to provide references.

Applications are welcomed until the position is filled. Please note that the administrative processing may take up to 3 months.

Contact persons

For further information or details about the internship before applying, please contact:

- Julien Girard-Satabin (julien.girard2@cea.fr) (also available on LinkedIn)
- Zakaria Chihani (zakaria.chihani@cea.fr)

References

Bak, Stanley, Changliu Liu, and Taylor Johnson. 2021. "The Second International Verification of Neural Networks Competition (VNN-COMP 2021): Summary and Results." *arXiv:2109.00498 [Cs]*, August. https://arxiv.org/abs/2109.00498.

Baudin, Patrick, François Bobot, David Bühler, Loïc Correnson, Florent Kirchner, Nikolai Kosmatov, André Maroneze, et al. 2021. "The Dogged Pursuit of Bug-Free C Programs: The Frama-C Software Analysis Platform." *Communications of the ACM* 64 (8): 56–68. https://doi.org/10.1145/3470569.

Girard-Satabin, Julien, Guillaume Charpiat, Zakaria Chihani, and Marc Schoenauer. 2020. "CAMUS: A Framework to Build Formal Specifications for Deep Perception Systems Using Simulators." In *ECAI 2020 - 24th European Conference on Artificial Intelligence*. Santiago de Compostela, Spain. https://hal.inria.fr/hal-02440520.

Katz, Guy, Clark Barrett, David Dill, Kyle Julian, and Mykel Kochenderfer. 2017. "Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks." *arXiv Preprint arXiv:1702.01135*.

Katz, Guy, Derek A. Huang, Duligur Ibeling, Kyle Julian, Christopher Lazarus, Rachel Lim, Parth Shah, et al. 2019. "The Marabou Framework for Verification and Analysis of Deep Neural Networks." In *Computer Aided Verification*, edited by Isil Dillig and Serdar Tasiran, 11561:443–52. Cham: Springer International Publishing.

Mattioli, Juliette, François Terrier, Loic Cantat, Julien Chiaroni, M Barreteau, Y Bonhomme, C Guettier, and C Alix. 2021. "IA de confiance: condition nécessaire pour le déploiement de l'IA dans les systèmes de défense." *APIA (Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle)*.

Singh, Gagandeep, Timon Gehr, Matthew Mirman, Markus Püschel, and Martin Vechev. 2018. "Fast and Effective Robustness Certification." In *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, 10802–13. Curran Associates, Inc. http://papers.nips.cc/paper/8278-fast-and-effective-robustness-certification.pdf.

Singh, Gagandeep, Timon Gehr, Markus Püschel, and Martin Vechev. 2018. "Boosting Robustness Certification of Neural Networks." In *International Conference on Learning Representations*.

Wang, Shiqi, Huan Zhang, Kaidi Xu, Xue Lin, Suman Jana, Cho-Jui Hsieh, and J. Zico Kolter. 2021. "Beta-CROWN: Efficient Bound Propagation with Per-neuron Split Constraints for Complete and Incomplete Neural Network Robustness Verification." October 31, 2021. http://arxiv.org/abs/2103.06624.