Sparse and robust training for neural network verification

Keywords: PyRAT, Neural Network, Abstract Interpretation

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 a large number of 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 lab extended its activities on the topic of AI trustworthiness and gave birth to a new research group on the topic: AISER (Artificial Intelligence Safety, Explainability and Robustness). Developping, tools such as PyRAT, AIMOS or CAISAR to improve the safety of AI systems as a whole.

Objectives

Numerous tools such as PyRAT can be used to verify the safety or security of a neural network. While based on different methods and implementations, each of them may find themselves limited when facing larger or more complex networks. The increasing number of parameters and layers introduce important imprecision in the analysis performed and thus does not allow to fully verify the network. Some first result on training a robust network through formal methods showed that it was possible for the training not only to influence the verification but to drastically reduce the time taken by it while also increasing the precision. The influence of other training techniques such as quantified network, sparse, Lipschitz on the verification process may be a way to surmount the complexity of larger networks.

The aim of this internship is to help in the study of the link between these training techniques and the verification process. This can be divided into two parts: the training of different neural networks on different dataset with the various techniques in mind and corresponding verification of these network. For the verification, depending on the architecture of the network, additional implementation might be undertaken on the verification tool. After testing the existing training techniques, new techniques might be extrapolated and tested. This work will be executed in collaboration with a PhD student.

Qualifications

- **Minimal**
  - Master student or 2nd or 3rd year of engineering school
  - knowledge of Python
  - Neural network training/AI and one AI framework (PyTorch, Keras, TF,...)
  - ability to work in a team
- **Preferred**
  - some knowledge of abstract interpretation or formal method

Characteristics

- **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

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:

- Augustin Lemesle (augustin.lemesle@cea.fr)
- Zakaria Chihani (zakaria.chihani@cea.fr)