2019-01367 - PhD Position F/M Formal Verification by Abstract Interpretation of Functional Correctness Properties of Neural Networks

Type de contrat : CDD
Niveau de diplôme exigé : Bac + 5 ou équivalent
Fonction : Doctorant

Contexte et atouts du poste

Machine learning techniques, notably deep learning [3], are enjoying tremendous success. They are applied to an ever-growing set of applications (image classification, speech recognition, prediction, to name a few). They are now also being considered for embedded critical applications, including medical devices, autonomous driving, or avionics. Such applications require a high level of insurance that the system behaves reliably and as expected.

In the context of traditional (non-machine learning) software systems, including embedded systems, formal methods have been used in order to provide strong, mathematically-grounded proofs of correctness. They are used at an industrial level in avionics [5], where they have stringent verification requirements mandated by international standards (DO-178C). This success is due to the recognition of formal methods by standards and the availability of effective, efficient verification tools. Among formal verification techniques, static analyzers based on abstract interpretation [1] live in a sweet spot as they are fully automated, efficient, and sound by construction (such as the Atréïre analyzer [2] used to ensure the absence of run-time errors in critical avionics C code). By contrast, the use of formal verification techniques in machine learning is extremely limited.

The goal of the PhD is to explore the use of abstract interpretation to verify neural networks, with a focus on avionic applications. The ANTIQUE team from INRIA & ENS | Université PSL and the APR team from Sorbonne Université are experts in verification by abstract interpretation and application to avionics [2]. The ANTIQUE team also has growing expertise in the analysis of neural networks [4, 5]. We are collaborating with Airbus to explore the use of embedded neural networks in avionics, define their verification requirements, and obtain case studies.

The PhD will be co-supervised by Caterina Urban (INRIA Paris, team ANTIQUE) and Antoine Miné (Sorbonne Université, team APR) and will take place at both École Normale Supérieure | Université PSL and Sorbonne Université. We expect frequent meetings with our industrial contacts at Airbus in Toulouse.

Mission confiée

The state of the art in the formal verification of neural networks is mostly limited to proving local robustness, which only ensures that a network behaves as expected around a small set of isolated points in the domain space. Several techniques have been applied to this problem (SMT, integer programming, etc.), and abstract interpretation has proven particularly effective [1] thanks to its ability to design abstraction with tunable cost vs. precision tradeoffs. However, global robustness (robustness at all points), which would be more suitable for critical software, seems out of reach. Another property, fairness, which is of interest for society-related decision-making applications, has also been tackled using abstract interpretation [3]. These are all non-functional properties, which are implicitly specified.

Our aim with this PhD is to explore functional properties instead. Existing work in this area [4] is even more limited, focusing on linear programming methods, with a single-use case and simple, artificial properties. By contrast, we wish to use abstract interpretation to tackle complex, realistic properties, identified through our collaboration with Airbus.

The PhD will thus address two key challenges currently limiting a broader application of formal verification to data science software [2]: the lack of specifications, and the difficulty to achieve both scalability and precision on a wider class of neural networks. We have already identified two kinds of specifications of interest:

- designer-specified input-output assertions stating high-level properties (such as “when input X is between these bounds, then output Y must be greater than 2”);
- for networks approximating (for performance reasons) a function defined analytically or algorithmically, a bound between the exact and approximate result on all inputs.

We believe that more classes will emerge in the course of the PhD project. Concerning scalability and precision, we are confident that the theory of abstract interpretation provides a large enough design space to create abstractions adapted to each class of networks and specifications (as is the case in software verification).

Informations générales

- Thème/Domaine : Preuves et vérification
- Ville : Paris
- Centre Inria : CRI de Paris
- Date de prise de fonction souhaitée : 2020-10-01
- Durée de contrat : 3 ans
- Date limite pour postuler : 2020-04-03

Contacts

- Equipe Inria : ANTIQUE
- Directeur de thèse : Urban Caterina / caterina.urban@inria.fr

A propos d’Inria

Inria, est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 3900 personnes. Ses 300 équipes-projets agiles, en général communiqués avec des problèmes délégués en ligne par une start-up, de 3500 chercheurs et ingénieurs pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers différents. 900 personnels d’emploi à la recherche contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde.

L’essentiel pour réussir

The PhD will be co-supervised by Caterina Urban (INRIA Paris, team ANTIQUE) and Antoine Miné (Sorbonne Université, team APR) and will take place at both École Normale Supérieure | Université PSL and Sorbonne Université. We expect frequent meetings with our industrial contacts at Airbus in Toulouse.

Consignes pour postuler

Sécurité défense : Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.

Politique de recrutement : Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

Attention : Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d’autres canaux n’est pas garanti.


Principales activités

The expected work will consist of several steps from the choice of case studies to designing, implementing and validating experimentally static analyses:

1. The student will select, with the help of our industrial contacts at Airbus, representative examples of neural networks (starting with smaller ones and growing in size and complexity) and properties of interest.

2. Formal specifications for these classes of properties will need to be defined, possibly in a parametric way to enable configuration by the analysis user.

3. The student will define new abstract domains, which are at the core of all abstract interpreters [1]. Both semantic aspects (expressiveness, abstract algebra) and algorithmic aspects (data-structures, algorithms, complexity) must be considered to allow a theoretical analysis and an effective implementation. Many abstract domains (including numeric domains, which are natural candidates to tackle learning methods) have already been proposed, but mainly for software analysis. They will need at least some adaptation, or a complete redesign, to become effective on neural network semantics and properties. Moreover, each different class of networks and properties may require domain-specific abstractions.

4. The abstractions shall be proved formally sound using the abstract interpretation theory.

5. The student will implement the methods and experiment with them on the case studies. Because abstract domains embed a precision vs. cost tradeoff, experimental proof is paramount to justify the ability of the method to provide meaningful results in reasonable times or realistic networks. Experiments will provide feedback to hone abstract domains and achieve scalability as larger networks are considered.

We expect the student to iterate on steps 3 to 5 in a tight loop.

Compétences

The ideal candidate should have the following qualities and skills:

- strong theoretical and practical knowledge of formal methods, in particular, abstract interpretation;
- basic knowledge of deep neural networks;
- programming skills and the willingness to implement and conduct analysis experiments;
- ability to read, write, and present in English.

Avantages

- Subsidized meals
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training