

Stage de recherche en informatique : Reconnaissance de programmes par réseaux de neurones de graphes

Sid TOUATI*
Enrico FORMENTI†
Christophe ALIAS‡

2024-2025

Lieu

Ce stage peut se dérouler en partie à l'INRIA-Lyon et en partie à Sophia Antipolis.

Motivations et contexte

Un résultat fondamental en informatique (calculabilité) nous affirme que des problèmes indécidables existent, c'est-à-dire pour lesquels il n'existe aucun algorithme qui soit capable d'en fournir la solution. Pour un informaticien, cela n'est pas une bonne nouvelle. La situation est encore pire. En effet, le théorème de Rice nous dit que tous les problèmes sont indécidables sauf ceux qui sont triviaux (au sens de la calculabilité).

Un exemple classique de problème indécidable – et que nous allons étudier dans le cadre de ce stage – consiste à prendre deux programmes et se demander s'ils calculent la même chose.

Que faire donc lorsqu'on veut s'attaquer à un problème indécidable ? Il faut être pragmatique : Il faut accepter de faire des erreurs de temps en temps. Le programme qui essaye de résoudre un problème indécidable fonctionnerait bien pour certaines instances et calculerait un résultat incorrect pour d'autres. Il faudrait aussi veiller à que les cas des calculs incorrects soient aussi rares que possible.

*Contact principal. Professeur à l'université Côte d'Azur. Sid.Touati@inria.fr

†Professeur à l'université Côte d'Azur.

‡Chargé de recherche INRIA.

Ce qu'on vient d'écrire est donc un cadre d'application idéal des réseaux de neurones qui, par nature, calculent leurs résultats à une marge d'erreur près.

Concrètement, dans ce stage nous voudrions construire un outil (grâce aux réseaux de neurones) qui devine ce que calcule un autre programme.

Ce problème de reconnaissance automatique des programmes, bien qu'indécidable, reste un enjeu fondamental en informatique, dont les applications sont nombreuses : *reverse engineering*, compilation, analyse de sécurité de code, etc. Essayer de deviner si un programme calcule telle ou telle fonction précise est une question qui se pose en pratique. Des travaux de recherche existent et ont tenté de reconnaître les programmes de manière algorithmique. Ce projet fait partie d'une nouvelle action exploratoire INRIA appelée ProgRecco, susceptible de proposer une bourse de doctorat en informatique à l'issue du stage.

Présentation du sujet de stage de recherche

Ce stage a pour objectif de réfléchir à une nouvelle démarche de reconnaissance de programmes, qui n'est pas intuitive. L'intelligence artificielle est un domaine classique issu des mathématiques et des sciences informatiques, qui a montré son efficacité en pratique dans des situations de reconnaissance de formes graphiques, de parole, de texte, d'interaction homme-machine, etc. Nous souhaitons développer un outil pour reconnaître des programmes basé sur l'intelligence artificielle. Nous appliquerons des méthodes de réseaux de neurones pour les classification des graphes (Graph Neural Network). Les étapes de travail sont :

1. Modifier le code source d'un compilateur existant (LLVM) pour extraire le graphe de flot de données d'un programme sous un format texte approprié.
2. Générer via compilation plusieurs variantes d'un programme, et entraîner un réseau de neurones avec la méthode GNN (Graph Neural Network).
3. Tester si et à quel point un tel réseau de neurones peut reconnaître d'autres variantes du même programme.

Prérequis souhaités

Le candidat idéal pour ce stage est un étudiant qui a suivi une formation en informatique fondamentale (compilation, programmation avancée, réseaux de neurones, graphes). Concernant la partie technique de compilation avec LLVM, la maîtrise du C++ est nécessaire.