Proposition de stage :
De la synthèse de programmes aux tests, et vice-versa!

Nathanaël Fijalkow Antoine Rollet

Encadrants : Nathanaël Fijalkow et Antoine Rollet (CNRS, LaBRI, Bordeaux)

Localisation : LaBRI (Bordeaux)

Sujet du stage : Ce stage propose d'étudier le problème de la synthèse de programme en s'inspirant de techniques de tests de programmes, et réciproquement, d'utiliser la synthèse pour faire du test.

Thèmes : synthèse de programme, apprentissage, test

La motivation du problème étudié dans ce stage est la synthèse de programme : à partir d'une spécification, on cherche à construire automatiquement un programme satisfaisant cette spécification. Ce problème a de nombreuses applications en pratique, par exemple le récent TF-coder pour générer du code Tensorflow [SBS20]. Un exemple d'application concrète est donné en Figure 1 : à partir d'une paire d'input / output, trouver un programme transformant l'input en l'output. Les techniques récentes de synthèse de programme s'appuient sur l'apprentissage automatique, et nécessitent en particulier la construction de bases de données dites synthétiques [CGP+20]. Les techniques de tests peuvent être utilisées pour générer ces bases de données.

```
a ← [int]                 An input-output example:
b ← SORT a                Input:
c ← FILTER (>0) b         [-17, -3, 3, 11, 0, -5, -9, 13, 6, 6, -8, 11]
d ← HEAD c                Output:
e ← DROP d b              [-5, -3, 0, 3, 6, 6, 11, 11, 13]
```

Figure 1: Un exemple du type de programme synthétisé

Objectifs du stage

L'objectif de ce stage est d'explorer les liens riches entre synthèse de programmes et test, et de les mettre en pratique. Il peut s'agir par exemple d'utiliser des techniques de couverture de code pour générer des paires d'input / output ayant un pouvoir de distinction entre plusieurs programmes, ou encore d'étudier des techniques d'apprentissage pour automatiser la génération de cas de tests.

References

[CGP+20] Judith Clymo, Adrià Gascón, Brooks Paige, Nathanaël Fijalkow, and Haik Manukian. Data generation for neural programming by example. In Silvia Chiappa and Roberto Calandra, editors, *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy]*, volume 108 of *Proceedings of Machine Learning Research*, pages 3450–3459. PMLR, 2020.

[SBS20] Kensen Shi, David Bieber, and Rishabh Singh. Tf-coder: Program synthesis for tensor manipulations. *CoRR*, abs/2003.09040, 2020.