Adversarial retraining pour les systèmes configurables

Paul Temple*[1]

[1]PReCISE research center, University of Namur – Belgique

Résumé

Les systèmes deviennent de plus en plus complexes et peuvent être configurés afin de répondre au mieux aux attentes des utilisateurs tout en capitalisant sur la réutilisabilité du code. Le problème étant que les espaces de configuration deviennent si grand qu'il est impossible d'explorer, générer, tester toutes les configurations possibles de ces systèmes. L'utilisation de modèles de Machine Learning (ML) permet alors d'élager cet espace suivant un but donné (e.g., garder les configurations qui n'utilisent pas plus d'un certain seuil en espace mémoire). Le but étant de réduire drastiquement mais également de manière précise le nombre de configurations à évaluer. En parallèle, depuis le début des années 2010, l'évolution extrêmement rapide du ML à poser de nombreuses questions y compris concernant le fonctionnement de ces modèles. Le domaine de l'adversarial machine learning a alors émergé et cherche à comprendre comment des personnes malveillantes peuvent exploiter les faiblesses des modèles de ML et comment s'en défendre. Dans sa réalisation, l'adversarial machine learning forge de nouvelles données ayant un impact maximal sur les modèles de ML (e.g., maximiser les chances d'être mal classée). Nous proposons d'utiliser ces techniques sur des systèmes configurables (et les modèles de ML associées) afin de générer automatiquement de nouvelles configurations du système. Cette génération permet entre autre : de rendre le modèle de ML, utilisé pour filtrer les configurations, plus précis; d'explorer davantage l'espace de configuration du système. Ce travail a été publié à EMSE 2021 (DOI: https://doi.org/10.1007/s10664-020-09915-7).

*Intervenant