

Proposition de stage recherche M2 en laboratoire 2019-2020

Titre : Analyse discriminante à noyau incrémentale et compressible pour la détection de nouveautés. Application à l'industrie 4.0

Description du sujet : Le sujet du stage concerne l'étude de la détection d'anomalies en mode supervisée. De manière plus précise, on considère que la base d'apprentissage est composée de plusieurs classes connues et à partir de cette connaissance il est possible de classer une nouvelle donnée parmi les classes déjà apprises ou non. Dans le deuxième cas, la donnée est considérée comme « nouveauté ». L'analyse discriminante à noyau (ADN) est connue comme un outil adapté et performant pour la détection d'anomalies en mode supervisée. Cependant, lorsque la base d'apprentissage atteint des tailles et des dimensions importantes ou que les données arrivent en flux, l'ADN devient inadaptée. La principale raison est que cette technique est basée sur une décomposition en valeurs singulières de la matrice à noyau qui présente une complexité en temps de $O(n^3)$ (n étant le nombre d'objets à apprendre) et donc devient trop coûteux lorsque la taille de l'échantillon augmente fortement. Associé à ce problème numérique, on peut ajouter aussi des coûts de stockage en mémoire vive pour la matrice à noyau qui deviennent vite prohibitifs. Dans ce stage de recherche nous proposons de mettre en place une version incrémentale de l'ADN adaptée aux données arrivant de manière séquentielle. Dans un second temps une stratégie de compression sera développée de manière à réduire l'information redondante et ainsi faire face à une matrice à noyau de taille toujours croissante.

Nous proposons d'étudier les performances de ce détecteur dans une application de cybersécurité dans l'industrie 4.0 où les données arrivent en flux. L'objectif est de modéliser les différents types de trafic de données au cours d'une phase intitulée « phase d'apprentissage », ce qui permettrait de détecter des anomalies pouvant indiquer une intrusion ou un vol de données par des cybercriminels.

Mots-clés : analyse discriminante à noyau, espace nul, implémentation incrémentale, compression, industrie 4.0

Encadrant(s) : F. Dufrenois, D. Hamad

email(s) : franck.dufrenois@univ-littoral.fr