

Résistance au bruit et à la rareté de la détection d'anomalies par arbre de décision de systèmes physiques simulés

Nesrine Bannour*, Anne Jeannin-Girardon*
Nicolas Lachiche*, Etienne Schneider*

* ICube, Université de Strasbourg
300 Bd Sébastien Brant
67400 Illkirch-Graffenstaden
bannour.nesrine@gmail.com,
{anne.jeannin, nicolas.lachiche, etienneschneider}@unistra.fr

Résumé. La détection d'anomalie est une tâche d'apprentissage dans laquelle les anomalies sont beaucoup plus rares que les comportements normaux. Notre objectif est de détecter une anomalie, en l'occurrence une fuite de fluide, le plus tôt possible, avant l'arrêt préventif de la machine. Dans cet article, nous étudions la résistance au bruit et à la rareté des anomalies d'une technique d'apprentissage supervisée, les arbres de décision. Nous considérons des données artificielles représentatives d'anomalies de systèmes physiques comme la crevaisson d'un pneumatique ou la fuite de fluide réfrigérant d'une pompe à chaleur. Nos tests montrent qu'un arbre de décision est capable d'apprendre un seuil sur la pression observée, en présence de bruit, qui s'adapte à des fréquences très faibles d'anomalies, jusqu'à 1 pour 100 000.

1 Introduction

La détection d'anomalies est définie comme la recherche de structures dans un jeu de données qui ne correspondent pas au comportement attendu (Chandola et al., 2009).

Dans cet article, nous étudions la résistance au bruit et à la rareté de la détection d'anomalies par arbre de décision dans le cas de systèmes physiques simulés. En fait, la question est de savoir si l'apprentissage supervisé est adapté à des problèmes de détection d'anomalies, où la classe positive (anomalie) est extrêmement moins fréquente que la classe négative (normale), en présence de bruit bien sûr. La fonction cachée étant simple (un hyperplan, sur une seule variable), toutes les techniques d'apprentissage supervisées pourraient être utilisées (perceptron, SVM, plus proche voisin, etc.). Nous avons choisi les arbres de décision car ils fournissent un modèle explicite et peuvent gérer un grand nombre de données.

La suite de notre article s'organise comme suit. La Section 2 expose le contexte et la problématique de la détection d'anomalies des systèmes physiques que nous considérons : le pneumatique et la pompe à chaleur. Ensuite, nous détaillons dans la Section 3 notre générateur de données artificielles. La Section 4 présente notre modèle de détection d'anomalies et les différents résultats obtenus suite à son évaluation. Enfin, une conclusion est établie dans la Section 5 et propose quelques perspectives.