

Double-ML-Weibull : du Machine Learning à la RUL, vers une distribution de probabilité

François Meunier*

*1 Chem. de la Porte des Loges, 78350 Les Loges-en-Josas
Air Liquide Innovation Campus Paris
francois.meunier@airliquide.com

Résumé. Les méthodes classiques d'estimation de la durée de vie utile restante (Remaining Useful Life) proposent une distribution de probabilité du risque de défaillance, qui permet en conséquence de donner une probabilité de défaillance avant chaque instant. Cependant, les méthodes récentes d'apprentissage automatique, qui utilisent des modèles plus complexes pour mieux comprendre les liens de causalité éventuelle entre les données disponibles et l'indicateur ciblé, proposent uniquement une régression. Dans cet article, nous introduisons une transformation de la valeur de sortie d'un régresseur basée sur l'apprentissage automatique en le complétant par un autre qui, en parallèle, calcule l'erreur estimée de ce modèle et l'utilise pour créer une distribution grâce à une loi de Weibull. Cette approche, appelée double-ML-Weibull, est un bien meilleur outil pour proposer une simulation dans un contexte stochastique, au lieu d'utiliser telles quelles des valeurs scalaires comme la "Mean Time To Failure" ou la "Mean Time Between Failure".

1 Introduction

L'apprentissage automatique est un ensemble de techniques qui permettent une meilleure compréhension et valorisation des variables latentes. Une puissance de calcul croissante combinée à des inférences statistiques utilisant des données toujours plus nombreuses permettent de grandes améliorations lors de la tentative de prédire l'avenir. La durée de vie utile restante (ou *RUL*) est un concept clé, en particulier lorsque l'on souhaite proposer un plan de maintenance optimisé, impliquant ainsi des économies de coûts et une diminution des risques. Les approches classiques pour estimer la *RUL* d'un actif particulier reposent généralement sur un modèle assez simple qui nécessite quelques variables à régler. Les lois exponentielles et de Weibull résumées par Bhattacharya (2011) sont de bons exemples de cet ensemble d'approches. Leurs résultats sont une distribution du risque de défaillance (qui peut facilement être traduite en distribution de fiabilité). Cependant, lorsque l'on examine ce que peuvent faire les algorithmes de régression (une branche des techniques d'apprentissage automatique supervisé), il apparaît qu'il existe une incompatibilité avec l'état de l'art actuel. En effet, la sortie d'une régression est un scalaire, car il est demandé au modèle de prédire une seule valeur (Li et al. (2018)). Lorsque l'on essaie d'utiliser ces outils en maintenance prédictive, et en raison de la différence de type de données (distribution vs scalaire), il devient impossible de :