

Écrêter la valeur cible ou filtrer les données en maintenance prévisionnelle : exemple de C-MAPSS

Nassime Mountasir^{*,**}, Baptiste Lafabregue^{*},
Bruno Albert^{**}, Nicolas Lachiche^{*}

^{*}Laboratoire ICube, 300 Boulevard Sebastien Brant, 67400 Illkirch-Graffenstaden

^{**}ENGLAB, 3 rue Denis Papin, 67300 Schiltigheim

1 Contexte

La maintenance prévisionnelle consiste à prédire lorsqu'une panne risque d'arriver sur une machine en prédisant une *Remaining Useful Life* (RUL), correspondant au temps avant la défaillance. Pour prédire le RUL, il n'est souvent possible de détecter l'apparition d'une défaillance que tardivement, car la grande majorité des données acquises correspondent à des données sans dégradation.

Notre contribution porte ainsi sur l'influence de la préparation des données sur l'apprentissage d'un modèle de prédiction de RUL. Pour comparer les différentes approches avec la nôtre, nous utilisons le jeu de données C-MAPSS. Il s'agit d'un jeu de données d'essai (*benchmark*) typique du domaine de la maintenance prévisionnelle (Ramasso et Saxena, 2014). Il s'intéresse à des simulations de dégradation de turboréacteurs et est développé par la NASA, décrit dans Saxena et al. (2008). Afin de visualiser des prédictions sur ce jeu de données, on utilise une forêt aléatoire avec les paramètres par défaut. Nous choisissons également une fenêtre de taille 15 en prenant donc en considération, pour un individu, les valeurs des paramètres des 15 pas de temps précédents. La phase de dégradation, elle, est souvent fixée à 125 pas de temps avant la défaillance, car il s'agit de la valeur utilisée par Heimes (2008) dans son approche, largement citée dans la littérature. Nous considérerons donc ce seuil dans la suite de l'article.

La Figure 1, correspondant à des exemples de prédictions réalisées sur des individus du jeu de données C-MAPSS, permet de visualiser le phénomène d'impossibilité à prédire correctement dans la phase normale. On peut constater sur la colonne de gauche que si l'on entraîne un modèle sans prendre en compte cette spécificité, le modèle se trompe lourdement, atteint un plateau et commence à diminuer son erreur sur les prédictions. Le modèle ayant appris à prédire un RUL sur une quantité importante de données ne comportant aucune dégradation, beaucoup d'erreurs sont réalisées. Il est donc nécessaire de se focaliser uniquement sur la phase de dégradation et non sur la totalité des données si l'on cherche à avoir un modèle performant. Pour réduire l'impact des données collectées en fonctionnement normal, l'approche traditionnelle consiste à écrêter le RUL (Heimes, 2008) afin qu'il ne dépasse pas un seuil. Écrêter permet ainsi de réduire la plage de valeurs du RUL en transformant les valeurs supérieures au seuil.

La deuxième colonne de la Figure 1 montre un exemple de prédiction avec écrêtage. Nous observons que modifier la valeur de la variable cible diminue l'erreur sur la phase normale,

Écrêter la valeur cible ou filtrer les données en maintenance prévisionnelle

mais déforme les prédictions sur la phase dégradée également. Le modèle, pour minimiser son erreur, va prédire très souvent des valeurs autour du seuil d'écrêtage. Plus la taille des séquences augmente, et plus la dégradation des prédictions s'accroît.

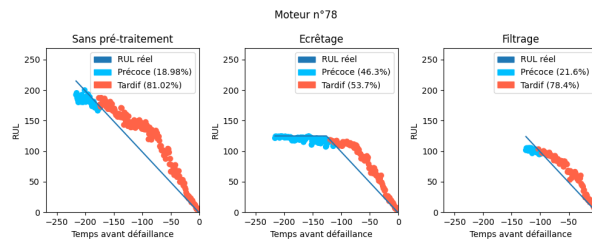


FIG. 1 – Visualisation des différentes prédictions sur un moteur du jeu de données C-MAPSS

2 Notre proposition

Nous avons vu précédemment qu'utiliser les données de la phase normale conduit à plus d'erreurs, car il n'est pas possible de prédire le RUL dans cette phase, et que l'écrêtage réduit les erreurs sur la phase normale, mais conduit à déformer le modèle y compris sur la phase dégradée. Nous proposons ainsi une approche, appelée filtrage. Elle consiste à considérer uniquement les données de la phase dégradée et ce sans changer les valeurs de la variable cible. Le fait de filtrer les données permet de s'intéresser uniquement aux informations correspondant à une dégradation, sans prendre en compte les données de fonctionnement normal qui empêcheraient le modèle d'extraire les caractéristiques présentes dans les données.

La troisième colonne de la Figure 1 permet de visualiser un exemple de prédiction réalisée avec les données filtrées. On constate que les prédictions sont plus resserrées et plus proches de la vérité terrain. Le modèle ayant été entraîné seulement sur des données comportant des dégradations, il n'a pas été brouillé par les données du fonctionnement normal. Cela permet au modèle d'être plus performant sur la phase de dégradation car il n'a été entraîné que sur ces données qui comportent des variations perceptibles, mais également d'éviter d'ajouter de l'erreur en prédisant un RUL approximatif pour la phase stationnaire.

Références

- Heimes, F. O. (2008). Recurrent neural networks for remaining useful life estimation. In *2008 International Conference on Prognostics and Health Management*, pp. 1–6.
- Ramasso, E. et A. Saxena (2014). Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets. *International Journal of Prognostics and Health Management* 5(2). Number : 2.
- Saxena, A., K. Goebel, D. Simon, et N. Eklund (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 International Conference on Prognostics and Health Management*, pp. 1–9.