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 :

— le comparer avec des approches de pointe dont les sorties sont des distributions ;
— utiliser leurs sorties dans des simulations stochastiques pour estimer un indicateur particulier, comme par exemple le coût total d'une politique de maintenance.

Pour résoudre ce problème, notre proposition appelée double-ML-Weibull (double machine learning pour le calcul de la loi de Weibull) repose sur l'utilisation de l'écart entre ce qui est prédit par le régresseur et la réalité, et à l'utiliser pour recréer une distribution centrée sur la sortie de ce premier régresseur.

2 Etat de l'art

2.1 Remaining Useful Life (RUL)

Comme expliqué précédemment, les méthodes stochastiques traditionelles permettant d'estimer le temps restant avant défaillance d'un actif (RUL) utilisent les données des défaillances passées pour ajuster leurs paramètres. Soit $S(t) = P(T > t)$ la fonction de survie avec T l'instant de la défaillance. A titre d'exemple, nous proposons les fonctions de survie suivantes :

— la loi exponentielle :
$$S(t; \lambda) = 1 - e^{-\lambda.x}$$
avec λ le taux de casse ;
— la loi de Weibull :
$$S(t; k, \lambda) = 1 - e^{-(x/\lambda)^k}$$
avec k le paramètre de forme et λ le paramètre d'échelle.

Ces méthodes, qui ne nécessitent pas beaucoup de données pour être réglées puis utilisées, créent une distribution de risque, permettant de donner, pour chaque instant, une probabilité d'échec avant un temps t / respectivement une probabilité de survie après un temps t.

D'autres approches cherchent à estimer la RUL d'un composant en utilisant des modèles plus complexes avec des données historiques à l'instar de Si et al. (2011), mais ne parviennent pas à donner un intervalle de confiance pertinent qui serait utile pour les essais de scénarios (Singleton et al. (2013); Zheng et al. (2017)) nécessaires en simulation stochastique, lorsque l'on cherche à optimiser la politique de maintenance d'un système.

2.2 Regression probabiliste

Lorsque la valeur cible sur laquelle l'apprentisssage est réalisé est un scalaire, le passage à une distribution de probabilités n'est basée que sur une seule information : l'incertitude estimée du prédicteur.

La regression probabiliste la plus utilisée pour l'estimation de la RUL est la régression quantile Tian et Wang (2021). Cette dernière, proposée initialement par Koenker et Bassett Jr (1978), consiste à orienter la prédiction dès l'apprentissage, pour que la valeur cible ne soit plus la moyenne (Mean Squared Error). Si le quantile est de 0.5, alors ce cas particulier revient à prédire la médiane. En proposant plusieurs regressions quantiles, il est possible de recréer une distribution de probabilités.

Néanmoins, intégralement reposer la structure de la distribution sur les quantiles du regresseur ne nous parait pas pertinent, dans la mesure où le manque de données est un problème récurrent en maintenance prédictive, et mène à des distribution irréalistes. D'où la nécessité de proposer une nouvelle approche, dans laquelle la forme de la distribution est cadrée, oritntée selon les comportements habituels observés en maintenance, et dont nous justifions la pertinence lors des expérimentations.

3 Du scalaire à la distribution de probabilité : une nouvelle proposition

3.1 Idée maitresse

Elle consiste à utiliser conjointement deux modèles d'apprentissage automatique :
- le premier est un régresseur classique qui prédit la durée de vie utile restante (RUL) de l'actif étudié ;
- le second utilise, à partir du régresseur précédent, ses entrées (données réelles) et sa sortie (RUL estimée) pour prédire l'erreur qu'il commet.

Suite à cela, nous disposons de valeurs qui peuvent être utilisées pour la création de la densité de probabilité, comme l'illustre la Fig. 1.

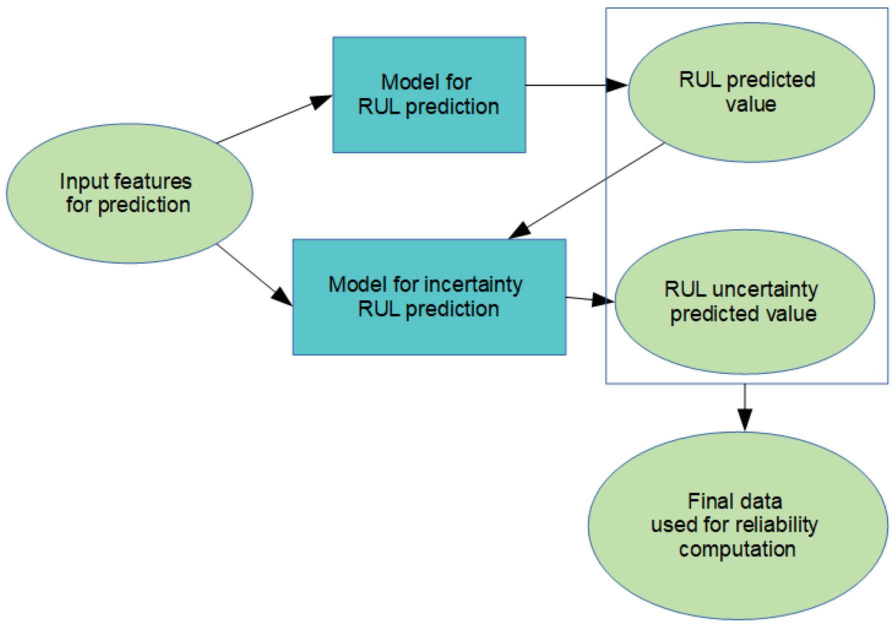

FIG. 1: Notre proposition : calcul de l'incertitude de la prédiction de la RUL pour une meilleure estimation de la fiabilité

3.2 Algorithme

Pour la partie de création de chaque modèle, il est proposé de procéder comme suit :

Algorithm 1 Approche globale de création de paramètres de densité de probabilité

$RUL_{scalar} \leftarrow Regressor_{scalar}(features)$

$Regressor_{scale}(features, RUL_{scalar})$
$\leftarrow Error_{Regressor_{scalar}(features)}$

$RUL_{distribution}$
$\leftarrow DistributionFunction(RUL_{scalar},$
$Error_{Regressor_{scalar}})$

return $RUL_{distribution}$

— Utilisation d'un ensemble de données pour entraîner $Regressor_{scalar}$ et obtenir la valeur d'erreur de chaque prédiction ;
— Utilisation de cet ensemble de données pour entraîner $RegressorRUL_{scalaire}$, l'autre modèle, en utilisant les features utilisées par $RegressorRUL_{distribution}$ ainsi que sa prédiction, et les valeurs de ses erreurs de prédiction comme cibles.

Bien sûr, l'apprentissage et le test sur le même jeu de données ne sont pas pertinents pour l'évaluation, d'où la nécessité de réaliser une validation croisée.

Avec :
— $CrossVal(items)$: fonction qui calcule une validation croisée sur $item$;
— $.fit(X, Y)$ méthode d'ajustement pour entrainer un modèle d'apprentissage automatique ;
— $.predict(X, Y)$ méthode de prédiction pour utiliser un modèle d'apprentissage automatique entrainé.

Une fois entraînés, les deux prédicteurs sont utilisés pour le calcul de la distribution de la RUL. En effet, à ce stade, nous avons
— un prédicteur pour une valeur scalaire de la RUL,
— un prédicteur de l'erreur du prédicteur précédent.

Une distribution normale classique utilisant la moyenne et l'écart type serait inappropriée pour plusieurs raisons (la symétrie n'est pas réaliste dans ce cas), mais la plus évidente est qu'il ne peut y avoir un risque de défaillance non nul avant le moment actuel $t = 0$. La fonction existante qui peut prendre en considération l'échelle λ et la forme k est la loi de Weibull.

Notre proposition est donc d'utiliser les résultats précédemment calculés pour retrouver les paramètres optimaux de cette fonction, puis de trouver un moyen de l'évaluer en comparant la sortie, une distribution, avec la valeur réelle connue, qui est un scalaire.

On utilise donc la valeur d'écart-type calculée par le second regresseur pour trouver la valeur du paramètre de forme k optimale. L'idée est de partir de valeur qui s'avère être centrale pour la loi de Weibull, soit 4 (équivalent à une loi équilibrée, celle qui se rapprocherait d'une distribution pseudo-gaussienne), et d'y appliquer la distribution des valeurs obtenues lors de la prédiction de l'écart-type :

$$k_{ajuste} = abs(ErreurCalculee) * 4 * \frac{nombre_{ErreurCalculee} > 0}{nombre_{ErreurCalculee} < 0} \tag{1}$$

Algorithm 2 Processus complet pour trouver les fonctions qui calculent les valeurs utiles pour les paramètres de la loi de distribution

Input $features$: features décrivant le contexte et la santé de l'actif ; RUL : Remaining Useful Life
Output $RegressorRUL_{scalar}$: régresseur de la RUL, sortie scalaire ; $RegressorRUL_{distribution}$: régresseur de l'erreur du premier régresseur

Begin

for $train, test \in CrossVal(features)$ **do**
$RegressorRUL_{scalar}.fit(features_{train}, RUL_{train})$
$Liste_{Errors} \leftarrow \oslash$

for $feature, RemLife \in features_{test}, RUL_{test}$ **do**
$Liste_{Errors} \leftarrow Liste_{Errors} \cup$
$|RegressorRUL_{distribution}(feature) - RemLife|$
end for

end for
$RegressorRUL_{distribution}.fit([features, RUL],$
$Liste_{Errors})$

return $Regressor_{scalar}$, $Regressor_{distribution}$

End

Une fois la valeur k retrouvée, il est possible, disposant également de ce que l'on considère comme la médiane de la distribution trouvée par le premier régresseur, de retrouver la valeur λ manquante pour la loi de Weibull. La relation est obtenue par la formule suivante :

$$mediane = \lambda * (ln2)^{\frac{1}{k}} \tag{2}$$

Ce qui donne, si l'on pose $mediane = resultat_{regression}$:

$$\lambda = resultat_{regression} * ln(2)^{\overline{\frac{1}{k_{ajuste}}}} \tag{3}$$

Par ce biais, on obtient finalement les deux paramètres de la loi de Weibull souhaitée.

3.3 Métrique proposée

Chercher à comparer une valeur scalaire à une distribution de probabilités n'est pas chose évidente. En effet, une distance classique n'est pas applicable, et les métriques existantes qui traitent de distributions tendent à comparer deux probabilités comme la divergence de Kullback Leibler (Kullback et Leibler (1951)) ou la distance de Bhattacharyya (Bhattacharyya (1946)), et non pas une probabilité et un scalaire comme l'impose notre cas d'étude. Par ailleurs, les

TAB. 1: Comparaison des résultats avec comme score la hauteur de la distribution à l'instant de la RUL réelle

Approche traditionnelle avec statistiques descriptives	0.22 ± 0.037
Méthode directe avec utilisation d'un k égal à 4	0.31 ± 0.029
Méthode par regression quantile	0.27 ± 0.022
Méthode avec calcul spécifique de k par un régresseur secondaire (notre proposition)	0.42 ± 0.057

méthodes d'évaluation d'intervalles de confiance comme l'Average Coverage Error et le Prediction Interval Normalized Average Width ne sont pas pertinentes, car nous souhaitons évaluer une distribution complète et non un simple intervalle.

Intuitivement, on cherche à faire en sorte que notre courbe de probabilité de survie "encadre" la valeur réelle de la RUL, c'est-à-dire que la valeur réelle corresponde à la valeur de densité de probabilité la plus élevée (cette convention pourrait tout à fait être remplacée par une autre, mais elle est la plus agnostique).

On cherche donc à maximiser la hauteur de densité de probabilité de notre courbe à l'instant cible, qui n'est autre que la valeur scalaire réelle de la RUL. Cela donne :

$$hauteur_{proba} = \frac{k}{\lambda} * \frac{scalaire}{\lambda}^{k-1} * \exp^{-\frac{scalaire}{\lambda}^{k-1}} \tag{4}$$

4 Expérimentations

Les expérimentations sont réalisées sur le datachallenge PHM 2008, dont les données disponibles s'avèrent être un bon cas d'utilisation pour évaluer intuitivement la pertinence de notre approche, n'ayant pour l'heure pas de référentiel pour cette tâche. Le jeu de données en question est de dimension 24 (soit 24 capteurs différents), composé d'une partie apprentissage de 100 multi-séries temporelles (x24) ayant 150 à 500 mesures chacune, et une partie test de 100 multi-séries (x24) de 80 à 500 mesures chacune. Les performances du modèle de regression utilisé pour calculer la durée de vie utile restante ne sont pas le point central, mais servent de base essentielle pour évaluer les paramètres de Weibull (conjointement avec la seconde régression supervisée). Pour la régression, xgboost, proposé par Chen et Guestrin (2016), est choisi, étant l'un des meilleurs algorithmes de l'état de l'art. L'essentiel est donc de comparer la distribution de la RUL et la valeur réelle.

Comme il n'existe pas de référence, nous proposons par la suite de choisir comme baseline l'utlisation d'un k valant 4 dans un premier temps, ce qui correspond à une loi de Weibull relativement symétrique (donc se rapprochant le plus possible d'une loi Gaussienne). Notre seconde baseline est la courbe de Weibull correspondant à un calcul par statistiques descriptives traditionnelle. Enfin, la troisième tire partie de la régression quantile mentionnée dans l'état de l'art.

Les résultats se retrouvent dans la Table 1.

Obtenus sur un tel jeu de données, ils permettent de conclure que les performances de notre approche surpassent très largement :

1. celles d'une approche basée sur des statistiques descriptives, méthode traditionnelle de "fitting" d'une loi de Weibull ;
2. celles d'une approche de calcul de la RUL avec une regression et un paramètre de forme fixé arbitrairement ;
3. enfin, celles issues de l'utilisation de la regression par quantile, parce que plus fidèle aux données, est bien plus sensible à ces dernières, donc à leurs aléas et carences.

5 Conclusion

Afin de permettre l'utilisation de RUL probabilistes dans des systèmes de simulation stochastiques, nous avons développé Double-ML-Weibull, une nouvelle méthodologie permettant de proposer un meilleur calibrage de la loi de Weibull pour l'estimation d'une durée de vie résiduelle. Ceci est réalisé par l'utilisation de deux modèles d'apprentissage automatique, qui permettent de mieux estimer les paramètres de la loi de Weibull en utilisant la prédiction (apprentissage automatique) au lieu de la description (statistiques traditionnelles). Les expérimentations réalisées sur des données ouvertes montrent l'intérêt de cette approche, en mettant en évidence la meilleure estimation des bornes de ces paramètres.

Références

Bhattacharya, P. (2011). Weibull distribution for estimating the parameters. In P. Bhattacharya (Ed.), *Wind Energy Management*, Chapter 1, pp. 1–10. Rijeka : IntechOpen.

Bhattacharyya, A. (1946). On a measure of divergence between two multinomial populations. *Sankhyā : The Indian Journal of Statistics (1933-1960) 7*(4), 401–406.

Chen, T. et C. Guestrin (2016). Xgboost : A scalable tree boosting system. *CoRR abs/1603.02754*, 785–794.

Koenker, R. et G. Bassett Jr (1978). Regression quantiles. *Econometrica : journal of the Econometric Society 46*(1), 33–50.

Kullback, S. et R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics 22*(1), 79–86.

Li, R., W. J. C. Verhagen, et R. Curran (2018). A comparative study of data-driven prognostic approaches : Stochastic and statistical models. In *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–8.

Si, X.-s., W. Wang, C. hu, et D.-H. Zhou (2011). Remaining ueful life estimation—a review on the statistical data driven approaches. *European Journal of Operational Research 213*, 1–14.

Singleton, R. K., E. G. Strangas, et S. Aviyente (2013). Time-frequency complexity based remaining useful life (rul) estimation for bearing faults. In *2013 9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED)*, pp. 600–606.

Tian, Q. et H. Wang (2021). Predicting remaining useful life of rolling bearings based on reliable degradation indicator and temporal convolution network with the quantile regression. *Applied Sciences 11*(11), 4773.

Zheng, S., K. Ristovski, A. Farahat, et C. Gupta (2017). Long short-term memory network for remaining useful life estimation. In *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 88–95.

Summary

Classical methods to estimate the Remaining Useful Life propose a distribution of probabilities of failure risk, which consequently makes it possible to give a probability of failure before each moment. However, recent machine learning methods, which use more complex models to better understand the possible causal links between the available data and the target indicator, only propose a regression. In this article, we introduce a transformation of the output value of a regressor based on machine learning by completing it with another one which calculates the estimated error of this model and uses it to create a distribution thanks to a Weibull law. This approach, called double-ML-Weibull, is a much better tool to provide simulation in a stochastic context, instead of using scalar values like "Mean Time To Failure" or "Mean Time Between Failure".