

Crack Extension Life and Critical Crack Length Prediction Based on XGBoost

Yu Liu ¹, Kaixing Zhao ², Fusheng Hou ¹, Xiaohui Hao ¹

¹COMAC Shanghai Aircraft Design & Research Institute, No.5188 Jinke Road, Shanghai, 200120, China

²Northwestern Polytechnical University, 127 West Youyi Road, Beilin District, Xi'an Shaanxi, 710072, China

Abstract

Damage tolerance design can ensure the structural safety of civil aircraft throughout its life cycle, which requires accurate analysis of crack extension life and crack extension length. This paper proposes an XGBoost-based crack extension life and crack extension length prediction method for civil aircraft structures. The method uses machine learning algorithms to train the structural state prediction model. The advantage of this method is that it can quickly determine the crack life and crack length without relying on the processing technology and engineering diagnosis experience of a large amount of collected data, which provides a more flexible method to determine the crack extension life and crack extension length under various influencing factors. By comparing a variety of machine learning algorithms, XGBoost model obtained the highest test scores, the experimental results show that the method achieves accurate prediction of crack extension life and critical crack length, which can be used for rapid engineering evaluation.

Keywords

Damage tolerance, crack extension, machine learning, XGBoost

1. Introduction

Damage tolerance design is a modern fatigue fracture control method developed and progressively applied since the 1970s to ensure the structural safety of structures throughout their life cycle. The theory assumes that each structural material has internal defects in the course of processing or using, so the rate of expansion of these defects and the remaining strength of the structure needs to be determined using various damage theories (such as fracture mechanics) and a given external load [1]. Damage tolerance is also a major component of the assessment for strength certification of aircraft by NASA, FAA, and other agencies. In the damage tolerance analysis, in addition to determining the critical parts of the damage tolerance and their sensitive parts, the accuracy of the crack extension life calculation and crack extension length calculation for this part is an important part to ensure the correct conclusion of the analysis [2]. However, there are various types of structural details in the airframe structure and their load states are complex, while the material of each part of the structure are different [3]. Therefore, in addition to the traditional fracture mechanics-based analysis methods, a tool that can quickly evaluate the crack extension life and the final crack length of the corresponding structure is also needed.

XGBoost, which is short for eXtreme Gradient Boosting, is an algorithmic toolkit based on the Boosting framework and is very powerful in parallel computation efficiency, missing value handling, and prediction performance [4]. In data science, XGBoost is well suited for performing data mining; in industrial large-scale data, the distributed version of XGBoost has extensive portability and supports running on various distributed environments such as Kubernetes and Hadoop, making it a good solution for industrial large-scale data.

ICBASE2022@3rd International Conference on Big Data & Artificial Intelligence & Software Engineering, October 21-23, 2022, Guangzhou, China

EMAIL: liuyu1@comac.cc (Yu Liu); kaixing.zhao@nwpu.edu.cn (Kaixing Zhao); houfusheng@comac.cc (Fusheng Hou); haoxiaohui@comac.cc (Xiaohui Hao)

ORCID: 0000-0001-7565-519X (Yu Liu); 0000-0002-0000-084X (Kaixing Zhao); 0000-0002-3192-3030 (Fusheng Hou); 0000-0001-5233-7762 (Xiaohui Hao)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

The current theories based on data-driven deep learning and machine learning, as the latest research results in pattern recognition, have achieved fruitful results in big data processing in various fields with powerful modeling and characterization capabilities. Therefore, using XGBoost-based crack extension prediction can get rid of the reliance on the traditional analysis method of mechanism-based fracture mechanics, complete the adaptive extraction of structural states and the construction of complex models, and finally realize end-to-end modeling to complete the prediction of complex indicators [5, 6, 7].

2. Crack extension dataset creation

As a part of the aircraft, the fuselage skin is an important part of the whole aircraft structure. In this paper, crack expansion prediction is performed for the fuselage and wing skin structure, which is usually considered as the central crack of an infinite plate, and the simplified model is shown below.

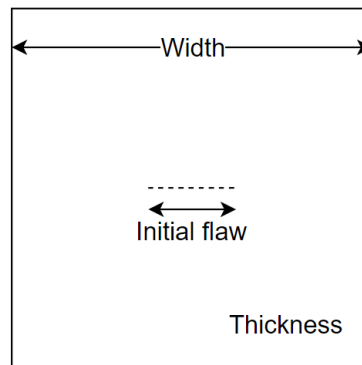


Figure 1: Central crack simplified model

Since the specimen-level test matrix of crack expansion usually cannot meet the data volume demand of machine learning, the data set in this paper is obtained by the crack expansion analysis software. Among them, if only the pressure-filled load of the skin is considered, the loaded state of the structure can be simplified to an equal amplitude spectrum. The specific parameters include structure width (plate width), maximum stress, structure thickness and initial crack size [8, 9].

The specific parameters and range are as following.

- structure width (plate width), 100-1500 mm;
- maximum stress, 100-200 MPa;
- Structure thickness, 0.5-5.5 mm;
- initial flaw(crack) size, 1.27-3.77 mm.

The specific dataset was created as follows.

1. Carrying out NASGRO input file generation based on a script that can automatically generate input files in Python, based on the planned test matrix;
2. Performing batch calculations;
3. The script obtains the cycle number and crack length information in each output file and summarizes them as the output result data table.

Over 1200 sets of data were eventually collected, including four input parameters as well as crack extension life and critical crack length [10].

3. Crack extension cycle and critical crack prediction based on XGBoost

This section covers the modeling and prediction of crack extension life/critical crack length based on XGBoost, and also includes the modeling process, data preprocessing, parameter optimization, model training, model testing results and comparison of other algorithms.

3.1. Modeling Process

The overall XGBoost prediction model can be divided into several parts: data reading and preprocessing, parameter optimization for cross-validation, model training, model testing, and result output. The flow is shown in Figure 2.

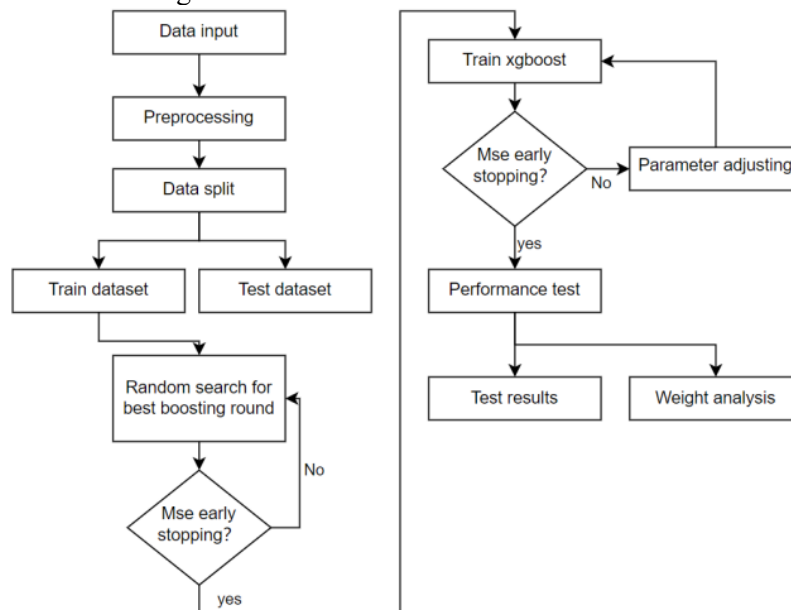


Figure 2: Modeling Process

3.2. Data preprocessing and parameter optimization

The original data was saved in csv format, and after removing the samples containing missing values and samples with too high lifetimes, the final 900 data sets were left. The total data set was partitioned 8:2, in which 80% of the training set (720 samples) was used for training the xgboost model and 20% of the test set (180 samples) was used for performance testing.

In modeling and training with xgboost, where the boosting round as an important parameter can be searched using the cross-validation function that comes with the model. In this paper, we use Random Search to find the optimal boosting round, with the upper limit of 500 cycles, Early stopping and automatic termination after 5 times of no change in the optimization index, tolerance of 0.01, cross-validation of 10fold, and optimization index of MSE. The final best value of boosting round is 143.

3.3. Model training and model testing

The XGBoost model is trained using the training dataset, where the boosting round is 143. the rest of the model parameters are listed in the following table [11].

Table 1

XGBoost model parameters

Parameters	Value	Parameters	Value
colsample_bytree	1.0	lambda	1.0
scale_pos_weight	1.0	eta	0.3
base_score	0.5	grow_policy	depthwise
max_depth	6	alpha	1.0
colsample_bynode	1.0	objective	reg:squared error

The model performance was tested using the full test set data, and the test set prediction scores are shown in the figure. After 143 boosting rounds of training, the performance metrics of the model are shown in the following table.

Table 2
XGBoost model test score

Categories	RMSE	MAE	R square
Cycles	119.528	70.703	0.9999
Crack size	0.526	0.269	0.9997

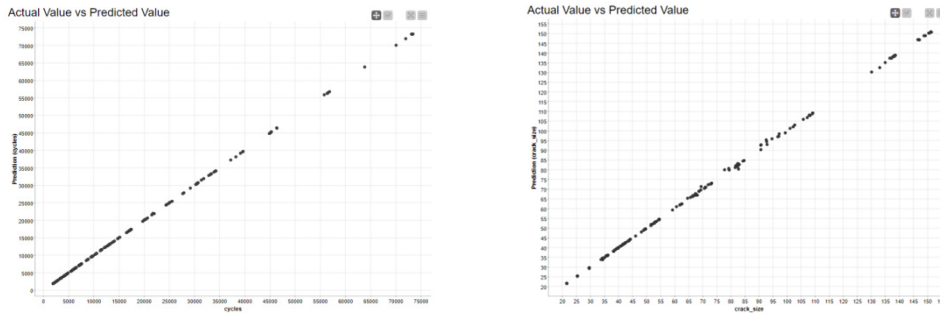


Figure 3&4: Actual value vs predicted value of cycles(left) and crack size(right)

The predicted model weights are analyzed. As shown in figure 5, the weights vary widely among the different attributes, which is of greater significance for analyzing the key influencing factors and optimizing the model inputs, and the relatively high-weighted attribute such as width may provide more information about the overall structural state of the aircraft.

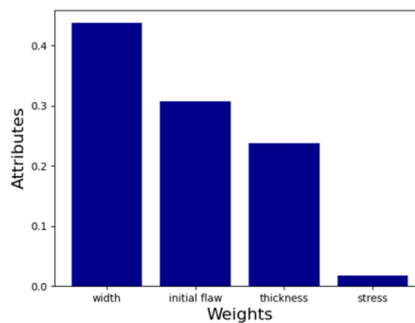


Figure 5: Attributes weights

3.4. Comparison with other algorithms

This study also applies other machine learning methods for modeling, including generalized linear regression models, decision tree, and support vector machine, to model the prediction of crack expansion number and critical crack length as well. The prediction models were all trained using the same 80% dataset (720 samples) and tested using 20% dataset (180 samples). One of the decision tree models, max_depth, was kept consistent with the XGBoost parameters above. The corresponding evaluation indexes are given for the performance of XGBoost and the other three prediction models in the test set. The prediction result scores for the number of crack expansion cycles are shown in Table 3, and the prediction result scores for the critical crack length are shown in Table 4 [12, 13, 14].

Table 3

Cycles prediction test score

Categories	RMSE	MAE	R square
XGBoost	119.528	70.703	0.999
Generalized linear	15682.212	12117.993	0.764
Decision tree	862.681	702.061	0.997
SVM	16436.034	12781.517	0.754

Table 4

Crack length prediction test score

Categories	RMSE	MAE	R square
XGBoost	0.526	0.269	0.997
Generalized linear	15.612	12.453	0.742
Decision tree	1.665	1.214	0.997
SVM	15.772	11.050	0.740

4. Conclusion

The XGBoost model has the best test score regardless of the number of crack expansion cycles or the length of critical cracks, and the model accuracy can meet engineering applications. Among the other tested models, the decision tree model has the closest score to Xgboost, which is also consistent with the performance consistency of the same tree model on the same data set.

In this paper, the Xgboost-based model achieves the prediction of the extended life and critical crack length of the center penetration crack of a flat plate. The model can be applied to the rapid assessment of damage tolerance in engineering. However, the model data is obtained based on fracture mechanics software simulations, which will have some deviation when applied to real aircraft structures, and can be combined with crack extension test data for further migration learning.

5. References

- [1] Reifsnider K L, Case S W. Damage tolerance and durability of material systems[M]. 2002.
- [2] Jones R. Fatigue crack growth and damage tolerance[J]. *Fatigue & Fracture of Engineering Materials & Structures*, 2014, 37(5): 463-483.
- [3] Toor P M. A review of some damage tolerance design approaches for aircraft structures[J]. *Engineering fracture mechanics*, 1973, 5(4): 837-880.
- [4] Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting[J]. *R package version 0.4-2*, 2015, 1(4): 1-4.
- [5] Konda N, Verma R, Jayaganthan R. Machine Learning Based Predictions of Fatigue Crack Growth Rate of Additively Manufactured Ti6Al4V[J]. *Metals*, 2021, 12(1): 50.
- [6] Pierson K, Rahman A, Spear A D. Predicting microstructure-sensitive fatigue-crack path in 3D using a machine learning framework[J]. *Jom*, 2019, 71(8): 2680-2694.
- [7] Raja A, Chukka S T, Jayaganthan R. Prediction of fatigue crack growth behaviour in ultrafine grained al 2014 alloy using machine learning[J]. *Metals*, 2020, 10(10): 1349.
- [8] Wang B, Zhao W, Du Y, et al. Prediction of fatigue stress concentration factor using extreme learning machine[J]. *Computational Materials Science*, 2016, 125: 136-145.
- [9] Mettu S R, Shivakumar V, Beek J M, et al. NASGRO 3.0: A software for analyzing aging aircraft[C]//The Second Joint NASA/FAA/DoD Conference on Aging Aircraft. 1999 (Pt. 2).
- [10] NASGRO F M. Fatigue Crack Growth Analysis Software[J]. Reference manual, 2002.
- [11] Brownlee J. XGBoost With python: Gradient boosted trees with XGBoost and scikit-learn[M]. *Machine Learning Mastery*, 2016.
- [12] Noble W S. What is a support vector machine?[J]. *Nature biotechnology*, 2006, 24(12): 1565-1567.

- [13] Myles A J, Feudale R N, Liu Y, et al. An introduction to decision tree modeling[J]. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 2004, 18(6): 275-285.
- [14] Nelder J A, Wedderburn R W M. Generalized linear models[J]. *Journal of the Royal Statistical Society: Series A (General)*, 1972, 135(3): 370-384.