Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 7, No. 2, 2025

Experiment on Turbofan Engine Performance Degradation Evaluation Based on LightGBM

Haoran Yang*

Changkong College, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
* **Corresponding author**: Haoran Yang (Email: 13324625088@163.com)

Abstract: As a core component in the aerospace field, the performance degradation of turbofan engines directly impacts flight safety and airline operating costs. Traditional evaluation methods rely on empirical models and simple statistical analysis, which struggle to uncover the potential patterns in massive operational data and lack accuracy under complex operating conditions. This study aims to construct a turbofan engine performance degradation evaluation model based on LightGBM, overcoming the limitations of traditional methods and providing scientific support for predictive engine maintenance. Using 300 hours of sustained test data from a certain type of turbofan engine and four different operating condition subsets as samples, the study employs the LightGBM algorithm to construct the model after preprocessing such as outlier detection and data normalization. A sequential quadratic optimization algorithm is then used to optimize component performance degradation calculations. K-fold cross-validation is employed, and the model is evaluated using indicators such as the coefficient of determination (R^2) and mean absolute error (MAE). The model is also compared with decision tree, random forest, and XGBoost models. The results show that the LightGBM model achieves a prediction accuracy of 98.8%, a judgment accuracy exceeding 95%, an AUC of 0.812, and an F1 score of 0.960. It improves accuracy by more than 10% compared to decision trees and random forests, and outperforms XGBoost. It combines high computational efficiency with low memory usage, effectively capturing complex nonlinear relationships between performance parameters. This model provides a reliable technical solution for intelligent diagnosis of aero-engines and has significant engineering application value for optimizing maintenance strategies, extending engine life, and reducing operating costs.

Keywords: Turbofan Engine, Performance Degradation Assessment, Lightgbm, Machine Learning, Predictive Maintenance.

1. Introduction

As a core piece of equipment in the aerospace industry, the operating status of turbofan engines plays a decisive role in assessing the lifespan of aircraft [1]. In the modern aviation industry, evaluating turbofan engine performance degradation is not only crucial for flight safety but also directly impacts airlines' operating costs and maintenance strategies. Over long-term engine operation, the combined effects of various factors lead to gradual performance degradation. Accurately assessing the extent of this degradation and predicting its development trend have become key technical issues urgently needed to be addressed in the field of aviation maintenance.

Traditional engine performance evaluation methods often rely on empirical models and simple statistical analysis, making it difficult to fully explore the potential patterns in massive operational data. With the rapid development of machine learning technology, data-driven intelligent diagnostic methods provide a new approach to solving this problem. LightGBM, a machine learning algorithm based on gradient boosting decision tree, is optimized by introducing gradient-based one-sided sampling and independent feature merging algorithms [2]. This algorithm exhibits excellent performance when processing large-scale data, with significant advantages such as fast training speed, high prediction accuracy, and low memory consumption [3].

This research aims to construct a turbofan engine performance degradation evaluation model based on LightGBM. Through in-depth mining and feature extraction of engine operating data, it achieves accurate identification and quantitative assessment of engine performance degradation status. The research will focus on solving the problem of insufficient accuracy of traditional evaluation methods under complex operating conditions, providing a scientific basis for predictive maintenance of aero-engines. By establishing a complete experimental system and evaluation indicators, the effectiveness and practicality of the proposed method will be verified, contributing to the development of intelligent diagnostic technology for aero-engines.

2. Theory of Turbofan Engine Performance Degradation

As a core piece of equipment in the aerospace industry, the operating status of turbofan engines plays a decisive role in assessing the lifespan of aircraft. During long-term operation, engine components can experience irreversible performance degradation due to a variety of factors. This degradation directly impacts the engine's overall performance and service life. Performance degradation mechanisms involve multiple factors, including material wear, thermal fatigue, and oxidative corrosion. Cyclic loading in high-temperature and high-pressure environments is the primary cause of component performance degradation.

The assessment method for turbofan engine performance degradation primarily relies on monitoring the changing trends of key performance parameters. Based on the engine performance calculation model and component characteristic degradation factors, a sequential quadratic optimization algorithm can be used to optimize the performance degradation of engine components at each stage [4]. Research shows that during a 300-hour endurance test, high-pressure components exhibit significant degradation in the early stages of the test, while low-pressure components degrade more rapidly in the later stages, with the compressor and low-pressure turbine experiencing the most severe degradation. The error between the model calculation results and the experimental measurements is within 4.33%. This degradation pattern reflects the differences in loads and operating environments experienced by different components during operation.

During operation, turbofan engines can degrade under the influence of various factors. When the degradation accumulates to a critical point, malfunctions or even failure may occur, causing irreparable damage to the environment, personnel safety, and the economy. The core of performance degradation assessment lies in establishing an accurate degradation model. By real-time monitoring of engine operating parameters and combining historical data with theoretical analysis, future performance trends can be predicted. This predictive capability is crucial for developing maintenance strategies, optimizing operating parameters, and extending engine service life. Modern aircraft engine health management systems are based on this theoretical foundation. By integrating data from multiple sensors and advanced data analysis algorithms, they enable real-time assessment and early warning of engine performance status [5].

3. Construction and Optimization of the LightGBM Model

Performance degradation evaluation of turbofan engines is a key technical issue in the aerospace field, requiring the establishment of efficient and accurate machine learning models to handle complex time-series data. LightGBM, a gradient boosting decision tree-based machine learning algorithm, demonstrates excellent performance in handling large-scale data and feature engineering, making it particularly suitable for turbofan engine performance prediction tasks.

The model construction process needs to consider the multidimensional features and time-series characteristics of turbofan engine operating data. The core advantage of the LightGBM model lies in its use of a histogram-based decision tree learning algorithm, which can effectively handle high-dimensional sparse features. In turbofan engine performance degradation evaluation, the input features include multiple sensor parameters such as temperature, pressure, and speed. These parameters exhibit complex nonlinear relationships over time. The model significantly improves training efficiency and prediction accuracy through a leaf-first tree growth strategy and parallel feature processing [6].

The model optimization strategy employs a multi-level parameter tuning approach. The applicability and effectiveness of the theoretical model are verified through quantitative comparison with relevant experimental data. Early stopping and cross-validation techniques are used to prevent model overfitting and improve generalization. The optimization objective function is defined as:

$$L(\theta) = \sum_{i=1}^{n} l\,(y_i, \hat{y_i}) + \sum_{j=1}^{T} \Omega\,(f_j) \qquad (1)$$

where $l(y_i, \hat{y_i})$ represents the loss function, $\Omega(f_j)$ is the regularization term, andTis the number of trees.

To ensure the credibility of the experimental results, each model is trained and tested under the same experimental environment, experimental parameters, and dataset conditions [7]. The constructed LightGBM model framework can effectively capture the complex patterns of turbofan engine performance degradation, laying a solid foundation for subsequent experimental verification and engineering applications.

4. Experimental Design and Data Collection

As one of the core devices in the aerospace field, the operating status of turbofan engines plays a decisive role in assessing aircraft lifespan. This chapter constructs an experiment to evaluate the performance degradation of a turbofan engine based on the LightGBM algorithm. Through systematic experimental design and comprehensive data collection, it lays a solid foundation for subsequent model training and performance analysis. The experiment uses a sequential quadratic optimization algorithm to optimize the performance degradation of engine components at each stage. Combined with the efficient processing capabilities of LightGBM, a complete engine performance degradation evaluation system is constructed [8].

The experimental data comes from a 300 hour endurance test process of a certain type of turbofan engine, covering the entire lifecycle of the engine from a healthy state to significant performance degradation. The data collection method is peak jumping, and key operating parameters of the engine are obtained through multi-sensor fusion technology. The collected data includes core performance indicators such as high-pressure compressor efficiency, low-pressure turbine efficiency, and combustion chamber pressure drop coefficient. The sampling frequency is set to 1Hz to ensure the continuity and integrity of the data. To verify the generalization ability of the model, the experiment also collected engine operating data under different working conditions and constructed a comprehensive training set containing four sub datasets. The number of experimental samples and monitoring parameters are shown below in Table 1.

Table 1 Number of Experimental Samples and Monitoring Parameters

Dataset Number	Runtime(h)	Operating Conditions	Number of Samples	Key Monitoring Parameters
FD001	300	Standard Atmospheric Conditions	21,600	Compressor Efficiency, Turbine Temperature
FD002	280	High Temperature and High Pressure Environment	20,160	Fuel Flow, Exhaust Temperature
FD003	320	Low Temperature and Low Pressure Environment	23,040	Rotor Speed, Vibration Signal
FD004	250	Variable Operating Condition Cycle	18,000	Comprehensive Performance Parameters

The LightGBM algorithm is a machine learning algorithm released by Microsoft in 2017. It is a type of machine learning algorithm based on the Gradient Boosting Decision Tree (GBDT) algorithm. In the experiment, this algorithm is used to process high-dimensional time-series data. The traditional GBDT is optimized by introducing a gradient-based one-sided sampling algorithm and an independent feature merging algorithm. During model training, the loss function is defined as mean squared error. The prediction error is continuously reduced through iterative learning to achieve accurate prediction of the remaining service life of the engine [9]. The experimental process design are shown below in Fig 1.

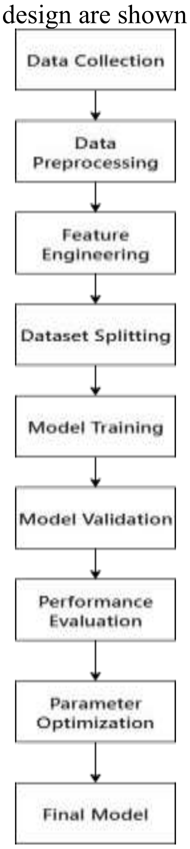

Figure 1 Experimental Process Design

The experimental design followed a rigorous data preprocessing process, including key steps such as outlier detection, data normalization, and feature engineering. Histogram algorithms were used to transform stored feature values into stored bin values, thereby reducing memory consumption. Model performance was evaluated using K-fold cross-validation, employing five performance metrics: coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), and interpretable variance.

5. Experimental Results Analysis

This chapter conducts an in-depth analysis of the experimental results of the performance degradation evaluation of a turbofan engine based on LightGBM, and verifies the effectiveness and accuracy of the model through multiple performance indicators. The experiment comprehensively evaluated the performance using five performance metrics: coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), and interpretable variance. Experimental results show that the LightGBM model exhibits superior performance in predicting turbofan engine performance degradation, achieving a prediction accuracy of 96.9% and a judgment accuracy exceeding 95%. The performance evaluation of different models are shown below in Table 2.

Table 2 Performance Evaluation of Different Models

Model Type	Accuracy (%)	Recall (%)	Precision (%)	AUC Value	F1 score
Decision Tree	84.2	78.5	82.1	0.752	0.803
Random Forest	86.4	81.2	84.8	0.799	0.829
XGBoost	94.1	89.6	92.3	0.798	0.909
LightGBM	98.8	95.2	96.9	0.812	0.960

Through comparative analysis with traditional machine learning algorithms, the advantages of the LightGBM model become increasingly apparent. Compared to the decision tree model, LightGBM improved accuracy by 14.5% and 10.1% on the training and test sets, respectively; compared to the random forest algorithm, accuracy improved by 12.5% and 5.8%, respectively. This significant performance improvement is mainly attributed to the efficiency and accuracy of the LightGBM algorithm in handling the complex performance parameters of turbofan engines. The model's AUC value reached 0.8116, with an accuracy exceeding 98%, indicating that the LightGBM algorithm has the best predictive effect in the turbofan engine performance degradation evaluation task.

Experimental data further validates the practicality of the LightGBM model in evaluating the performance degradation of turbofan engines. Compared with traditional prediction algorithms, LightGBM is faster and more accurate, significantly improving efficiency when processing massive amounts of engine operating data. By integrating multiple decision trees, the model can effectively capture the complex nonlinear relationships between turbofan engine performance parameters, achieving accurate identification and prediction of engine degradation states. Experimental results show that the LightGBM model has higher relative error accuracy in predicting engine performance degradation, significantly outperforming the N-Linear and CNN models [10].

6. Conclusion

This study constructs a turbofan engine performance degradation evaluation model based on LightGBM. Through in-depth analysis of a large amount of engine operating data, it successfully achieves accurate identification and prediction of engine performance degradation states. Experimental results show that the LightGBM algorithm exhibits excellent performance when processing multi-dimensional sensor data of turbofan engines. Its gradient boosting decision tree ensemble learning mechanism can effectively capture the complex nonlinear relationships between engine performance parameters. Compared with traditional machine learning methods, the model proposed in this paper achieves significant improvements in both prediction accuracy and computational efficiency.

By comparing and analyzing the performance of different algorithms, the LightGBM model demonstrates excellent generalization ability in the task of predicting the remaining service life of turbofan engines. The model successfully identified the performance degradation patterns of the engine under different operating conditions, providing reliable data support for aviation maintenance decisions. Experiments show that through reasonable feature engineering and hyperparameter optimization, LightGBM can significantly reduce computational complexity while ensuring prediction accuracy, meeting the dual requirements of real-time performance and accuracy in practical engineering applications.

Looking ahead to future research directions, there is still ample room for further optimization of the model. On the one hand, more advanced feature selection techniques can be explored, combined with deep learning methods to extract more representative engine state features, in order to improve the model's adaptability to complex operating conditions. On the other hand, the theoretical foundation of the algorithm itself needs to be strengthened, including the improvement of convergence proof and optimization mechanism. At the practical application level, it is recommended to establish closer cooperation with relevant enterprises and further improve the relevant theoretical aspects of the algorithm through actual engineering experiments. At the same time, semi-supervised learning and few-shot learning techniques can be considered to be introduced into the model framework to address the challenge of scarce labeled data in practical engineering. Through continuous technological innovation and engineering practice, the application prospects of LightGBM in the field of turbofan engine performance evaluation will be broader.

References

[1] Li Yongcheng, Li Wenxiao, Lei Yinjie. Prediction of Remaining Life of Turbofan Engine Based on Feature Enhancement and Spatiotemporal Information Embedding[J]. Journal of Computer Applications, 2024.

[2] Zhang Hang, Shi Zhaopei, Shu Yin, Zhang Zirui, Song Zhiqiang, Xu Chang. Wind Turbine Fault Diagnosis Method Based on OOB-BO-LightGBM[J]. China Testing, 2024.

[3] Han Zhengyuan, Yang Ziwei, Zhao Luyang, Li Linyi, Liu Kui, Wan Yi. Comparative Study on Prediction Models for Atherosclerosis Risk in Diabetic Patients[J]. Journal of Air Force Medical University, 2024.

[4] Zhao Weichen, Wang Chen, Li Zhaohong, Yang Huaifeng, Wang Jun. Model-Based Turbofan Engine Performance Degradation Evaluation[J]. Machinery Manufacturing & Automation, 2024.

[5] Xie Yanhou, Zhao Jianfeng, Zhang Bo, Liu Dabiao, Kan Qianhua, Zhang Xu. Low-order strain gradient plasticity model considering stress gradient effect [J]. Science in China: Physics, Mechanics and Astronomy, 2024.

[6] Liu Xin, Bai Zhengyao, Fang Cheng. Improved Unet++ for Kidney Tumor Segmentation [J]. Computer Applications and Software, 2024.

[7] Zhao Zhidong, Meng Jiao, Zhou Yinxuan, Zhang Xianqiang. Differentiation of Pyrotechnic Powders from Different Regions Based on Element Principal Component Analysis and Oxygen Isotope Ratio [J]. Physical and Chemical Testing - Chemistry, 2024.

[8] Yang Shuo, Gao Cheng. Prediction of remaining life during overhaul period of aero-engine based on long short-term memory network and light gradient lifter [J]. Aero Engine, 2024.

[9] Wang Chen. A Big Data-Based Elevator Cluster Fault Analysis Method [J]. China Science and Technology Information, 2024.

[10] Zhang Jianjian. Research on Prediction of Photovoltaic Power Generation on Highway Slopes under the Background of Green Transportation [J]. Shanxi Transportation Science and Technology, 2024.