

배기가스 온도 데이터를 활용한 선박 메인엔진 결함 진단 및 예지

Main engine fault diagnostics and prognostics using exhaust gas temperature data

김태경 석사¹, 김세하 석사¹, 전창재 교수²

¹세종대학교 인공지능학과, ²세종대학교 인공지능데이터사이언스학과

¹rlaxorud970@gmail.com, ¹1155622@naver.com, ²cchun@sejong.ac.kr

Abstract

선박 메인 엔진의 결함을 조기에 발견하고 대처하는 것은 효율적이고 안전한 운항을 위해 매우 중요하다. 결함이 발생할 경우 경제적 손실과 안전의 위험이 증가할 수 있다. 그러나 현재 시점에서의 결함 진단은 이미 발생한 결함에 대해서만 대처할 수 있다는 한계가 있다. 따라서 본 논문에서는 실제 운항 환경에서 수집된 배기가스 온도 데이터를 활용하여 현재 시점에서의 결함 진단에서 더 나아가 미래 시점에서의 결함 진단으로 확장하여 선박 메인 엔진의 결함 진단을 위한 모의실험을 수행한다. 연구는 세 단계로 진행된다. 첫째, 현재 시점에서의 결함 진단에 여러 머신러닝 모델을 적용하여 가장 적합한 모델을 선정한다. 둘째, 미래의 시계열 데이터를 예측하는 Transformer 기반 모델을 설계하여 성능을 평가한다. 마지막으로 예측된 시계열 데이터를 기반으로 미래 시점에서의 결함 진단을 진행한다. 실험 결과, 예측된 시계열 데이터를 활용하여 결함 진단을 진행하여 78.160%의 정확도를 달성하였다.

Keywords: Fault diagnosis, Time-series prediction, Marine systems, Exhaust gas temperature, Machine Learning, Transformer

1. 서론

선박은 수많은 하위 시스템과 구성 요소로 이루어진 복잡한 시스템이다. 그중 메인 엔진은 가장 중요한 하위 시스템이며, 추진력을 공급하는 역할을 한다(Ji *et al.*, 2023). 이는 안정적이고 효율적인 운항에 매우 필수적이다(Park and Oh, 2022). 그러나 선박은 통상적으로 건조 후 25년 이상 운항하며, 장기적인 운항에 따른 고

장이나 결함으로 인해 냉각수의 과열, 윤활유의 과열 및 배기가스의 과열 등의 문제가 발생할 수 있다(Cheliotis, 2020). 이러한 결함은 조기에 대처하지 않을 경우 막대한 금전적 손실과 심각한 인명 피해로 이어질 수 있다(Kim *et al.*, 2020). 따라서 메인 엔진의 결함을 사전에 발견하고 대처하는 것은 매우 중요하다.

최근 인공지능의 발전으로 다양한 산업에서 머신러닝을 활용한 결함 탐지 및 결함 진단 연구가 이루어졌다. 선박의 메인 엔진에서도 머신러닝을 적용한 결함 진단 연구들이 등장하였다. Kocak *et al.*(2023)에서는 Linear Regression, Support Vector Machine 및 Tree 기반 모델 등 머신러닝 기반 모델들을 활용하여 선박 메인 엔진 이상 탐지 연구를 진행하였다. 또한 Tsaganos *et al.*(2018)에서는 Naïve Bayes 및 Support Vector Machine 등 머신러닝 기반 모델들을 사용하여 해양 엔진 결함 진단 연구를 진행하였다.

이처럼 결함 탐지 및 결함 진단 연구는 많이 이루어졌지만, 결함의 정도나 상태를 식별하는 연구는 실제로 수집된 결함 데이터가 정상 데이터에 비해 현저히 부족하기 때문에 많이 진행되지 못하였다(Velasco-Gallego and Lazakis, 2022). Kocak *et al.*(2023)와 Tsaganos *et al.*(2018)에서도 이상 상태에 대한 탐지만 수행하였으며, 결함에 대한 식별은 고려하지 않았기에 결함의 정도나 상태를 알 수 없었다. 이러한 데이터의 한계를 극복하고자 이상 상황을 모의하여 결함을 식별하는 연구가 진행되었다(Campora *et al.*, 2018).

본 논문은 실제 환경에서 Instantaneous angular speed를 통해 수집된 2행정 디젤 엔진 데이터셋을 활용하여 결함 진단 모의실험을

진행한다. 결함 데이터의 부족 문제를 해결하기 위해서 데이터를 정상 및 위험 구간으로 이분화하여서 이상 상황을 모의하였으며, 각 구간을 6개의 세부 구간으로 나누어서 결함 진단을 진행한다. 이 과정을 통해 결함의 다양한 상태 및 정도를 보다 세밀하게 분석할 수 있다. 더 나아가 현재 시점에서의 결함 진단을 확장하여 미래 시점에서의 결함 진단을 진행한다. 연구는 세 단계로 구성된다. 첫째, 현재 시점에서의 결함 진단에 적합한 모델을 선정하기 위해 머신러닝 기반의 모델을 활용하여 실험을 진행한다. 둘째, Transformer 기반 모델을 설계하여 시계열 데이터를 예측한다. 마지막으로 예측된 데이터를 기반으로 미래 시점에서의 결함 진단을 진행한다. 실험 결과, 미래 시점에서의 결함 진단에서 78.160%의 정확도를 달성하였다.

2. 본론

본 장에서는 데이터를 전처리하는 방법을 설명하고, 결함 진단을 위해 활용된 머신러닝 기반 모델들, 그리고 시계열 예측을 위해 설계한 Transformer 기반 모델에 대해 기술한다. 본 연구에서는 결함을 탐지할 뿐만 아니라 결함의 상태 및 정도를 식별하기 위해서 데이터를 총 12개의 구간으로 분할하였으며, 각 구간은 상태 및 정도를 나타내는 클래스로 정의한다. 데이터 전처리에 대한 자세한 내용은 섹션 2.1에서 설명한다. 또한 본 연구는 현재 시점에서의 결함 진단을 넘어 미래 시점에서의 결함 진단으로 확장한다. 이를 위해 현재 및 미래 시점에서의 결함 진단에 적용한 머신러닝 기반 모델들을 다루고, 미래의 시계열 데이터를 예측하기 위한 Transformer 기반 모델을 설계한다. 각각의 세부 사항은 섹션 2.2와 2.3에서 설명한다.

2.1 데이터셋

본 연구에서는 Instantaneous angular speed를 통해 실제로 수집된 2행정 해양 디젤 엔진 데이터셋을 사용하여 결함 진단을 진행한다. 1년 동안 수집된 배기가스 평균 온도 데이터를 사용하였으며, 운항하지 않을 때의 데이터를 고려하여 전체 평균 온도보다 낮은 값은 제거하였다. 실험에 사용된 데이터에는 이상 데이터가 정상 데이터에 비해 매우 적다. 따라서 데이터를 전처리하여 이상 상황을 모의하였다. 구체적으로 전체 데이터를 정상 및 위험 구간으로 이분화하였고, 결함의 정도를 식별하기

위해 각 구간을 6개의 세부 구간으로 나누었다. 각 세부 구간의 평균 온도는 정상 구간의 첫 번째 세부 구간의 평균 온도와 일치하도록 조정하였다. 이후 위험 구간의 첫 번째 세부 구간의 온도를 증가시켜서 그 구간의 평균 온도가 위험 온도에 도달하도록 처리하였다. 이때, 위험 온도를 600으로 설정하였다. 정상 구간의 첫 번째 세부 구간의 평균 온도와 위험 구간의 첫 번째 세부 구간의 평균 온도를 기준으로 각 세부 구간의 온도를 증가하여 각 구간의 평균 온도가 점진적으로 상승하도록 계산하였다. 각 세부 구간의 평균 온도는 표 1에서 확인할 수 있다.

구간 범주	구간 번호	구간별 평균 온도
정상	1	287.849
	2	339.874
	3	391.899
	4	443.924
	5	495.950
	6	547.975
위험	7	600
	8	652.025
	9	704.050
	10	756.076
	11	808.101
	12	860.126

표 1. 구간별 평균 온도

2.2 현재 및 미래 시점에서의 결함 진단을 위한 머신러닝 기반 모델

머신러닝 기반 모델은 과학 및 공학 문제에 널리 사용되었다(Tsaganos *et al.*, 2018). 본 연구에서는 결함 진단에 가장 적합한 모델을 선정하기 위해 대표적인 머신러닝 기반 분류 모델인 Support Vector Machine (SVM), Artificial Neural Network (ANN), K-Nearest Neighbor (K-NN), Decision Tree, Light Gradient Boosting Machine (LGBM) 및 Random Forest를 사용하여 실험하였다. 해당 모델들은 입력된 데이터로부터 클래스를 분류하여 결함의 상태 및 정도를 식별한다.

2.3 미래의 시계열 데이터를 예측하기 위한 Transformer 기반 모델

현재 시점에서의 결함 진단은 결함이 이미 발생한 후에 대처가 가능하다는 한계가 있다. 이는 조기에 결함을 발견한다고 하더라도 이미 경제적 손실과 안전의 위험에 노출될 수 있다. 사전에 결함을 예측하여 대응할 수 있다면 더욱 안정적인 운항이 가능할 것이다. 따라

서 본 연구에서는 미래의 시계열 데이터를 예측하는 Transformer 기반 모델을 제안한다. 이 모델은 과거의 시계열 데이터를 기반으로 미래의 시계열 데이터를 예측한다. 예측된 시계열 데이터는 미래 시점에서의 결함 진단을 위한 입력 데이터로 사용한다.

실험에 사용한 Transformer 기반 모델은 다음과 같이 설계하였다. 우선, positional encoding을 통해 시계열 데이터의 시간적 순서를 학습할 수 있도록 하였다. 위치 정보가 인코딩된 입력 벡터는 Vaswani(2017)에서 제안된 multi-head attention과 feed forward로 구성된 Transformer encoder를 통과한다. Multi-head attention에서는 총 10개의 head를 사용하였고, 입력 벡터의 크기는 250으로 설정하였으며, 오직 1개의 Transformer encoder를 사용하였다. 마지막으로 하나의 feed forward layer를 사용하여 Transformer encoder로부터 추출한 특징을 결함 진단을 위한 입력 데이터의 크기와 동일하게 변환하였다. 추가로 Dropout을 사용하여 과적합을 방지하였으며, 학습의 안정성을 높였다.

3. 실험 결과

3.1 실험 환경

본 논문의 실험은 Linux 20.08 환경에서 진행되었다. 하드웨어는 NVIDIA A6000 GPU와 intel RCore TM i9-13900K CPU를 사용하였다. CUDA 버전은 12.0이며, Python과 Pytorch 버전은 각각 2.2.2와 3.10.0이다.

3.2 현재 시점에서의 결함 진단 결과

본 실험에서는 현재 시점에서의 가장 높은 결함 진단 정확도를 가진 모델을 선정하기 위해 섹션 2.2에서 언급한 머신러닝 기반 모델들을 사용하여 실험하였다. 결함 진단 실험을 진행하기 위해서 데이터의 window size와 stride를 120과 40으로 설정하였다. 이때, window size는 입력으로 사용되는 시계열 데이터의 타임 스텝을 의미하고, stride는 데이터 샘플 간의 간격을 나타낸다. 이후 데이터에서 9개의 통계적 특징(최댓값, 최솟값, 평균, 표준편차, 중앙값, 왜도, 첨도, 25 백분위수 및 75 백분위수)을 추출하여 모델의 입력 벡터로 사용하였다. 시물레이션은 총 30번 실험하였으며, 평균 정확도를 표 2에 기록하였다. 표 2에서 확인할 수 있듯이, Random Forest가 92.383%로 가장 높은 average accuracy를 기록하였고, 평균적으로 다른 모델들에 비해 5.647% 더 높은 정확도를

가지는 것을 보여준다.

추가로 서로 다른 window size와 stride에 따른 성능의 차이를 비교하기 위해서 총 7가지 조합의 경우를 가정하여 실험하였다. Random Forest를 활용하여 총 30번 실험하였으며, 실험 결과를 표 3에 기록하였다. 표 3에 기록된 바와 같이, window size가 120이고, stride가 40일 때 가장 높은 average accuracy를 가지는 것을 확인하였다. 이를 통해 미래 시점에서의 결함 진단 실험을 진행할 때 window size가 120이고, stride가 40인 데이터에서 학습한 Random Forest를 활용하는 것이 적합하다는 것을 확인하였다.

Model	Average Accuracy
SVM	81.720%
ANN	82.796%
K-Neighbors	87.869%
Decision Tree	90.312%
LGBM	90.984%
Random Forest	92.383%

표 2 머신러닝 기반 모델들의 결함 진단 실험 결과

Window size	Stride	Average Accuracy
240	80	89.260%
120	40	92.383%
90	30	91.982%
60	20	92.307%
30	10	91.734%
20	6	91.254%
10	3	90.395%

표 3. 총 7가지 경우의 window size와 stride의 조합에 따른 Random Forest를 활용한 결함 진단 비교 실험 결과

3.3 미래 시점에서의 결함 진단 결과

본 실험에서는 미래 시점에서의 결함 진단을 진행하기 위해서 Transformer 기반 모델을 활용하여 시계열 데이터 예측 결과를 평가하고 예측된 데이터를 활용하여 결함 진단을 진행한다

3.3.1 Transformer 기반 모델을 활용한 시계열 데이터 예측 결과

시계열 예측을 진행하기 위해서 가장 적절한 입력 데이터의 타임 스텝을 선정해야 한다. 이를 위해 총 5가지 경우로 가정하여 비교 실험을 진행하였다. 모델의 출력 크기는 예측된 데이터를 통해 결함 진단을 진행해야 하므로

120으로 고정하였다. 평가 지표는 Root Mean Squared Error (RMSE) 및 Mean Absolute Error (MAE)를 사용하였다. 표 4에 기록된 바와 같이 window size가 240일 때 가장 낮은 RMSE와 MAE를 가지는 것을 확인하였다. 따라서 window size가 240인 데이터에서 학습한 모델이 예측한 데이터를 활용하여 미래 시점에서의 결함 진단을 진행한다.

Window size	RMSE (↓)	MAE (↓)
120	12.563	12.377
150	14.832	14.712
180	11.877	11.745
210	13.931	13.810
240	10.233	10.085

표 4. 총 5가지 경우의 window size에 따른 Transformer 기반 모델을 활용한 시계열 데이터 예측 결과

3.3.2 예측 데이터를 활용한 미래 시점에서의 결함 진단 결과

미래 시점에서의 결함 진단을 진행하기에 앞서 섹션 3.2와 섹션 3.3.1에서 가장 높은 성능을 보인 설정을 채택하였다. 즉, window size의 크기가 240인 데이터로 학습한 Transformer 기반 모델이 예측한 미래의 시계열 데이터에 window size가 120이고, stride가 40인 데이터로 학습한 Random Forest를 적용하여 결함 진단을 진행한다. 총 30번의 시뮬레이션을 실험하였으며, 평균 정확도를 계산하였다. 실험 결과, 미래의 시계열 데이터를 활용하여 결함 진단을 진행했음에도 78.160%의 평균 정확도를 달성하는 것을 확인하였다. 이는 우리가 제안한 미래 시점에서의 결함 진단 방법이 거의 80% 정확하게 진단할 수 있다는 것을 보여준다.

4. 결론 및 향후 연구

본 논문에서는 실제 운항 환경에서 IAS를 통해 수집된 2행정 해양 디젤 엔진의 배기가스 온도 데이터를 활용하여 메인 엔진의 결함 진단을 위한 모의실험을 진행하였다. 연구는 세 가지 실험으로 구성하였다. 첫째, 현재 시점에서의 결함 진단에 다양한 머신러닝 기반 모델들을 적용하여 실험 결과를 비교하였고, 가장 높은 평균 정확도를 가진 모델을 미래 시점에서의 결함 진단에 가장 적합한 모델로 선정하였다. 둘째, 미래의 시계열 데이터를 예측하는 Transformer 기반 모델을 설계하여 성능을 평가하였으며, 마지막으로, 예측된 시계열 데

터를 활용하여 미래 시점에서의 결함 진단을 진행하였다. 미래 시점에서의 결함 진단 실험의 결과, 예측된 시계열 데이터를 활용하여 결함 진단을 진행했음에도 불구하고 78.160%의 평균 정확도를 가지는 것을 확인하였다. 이는 우리가 제안한 미래 시점에서의 결함 진단 방법이 거의 80% 정확하게 진단할 수 있다는 것을 보여주며, 이로 인해 선박의 안정성이 크게 향상할 수 있을 거라 기대한다. 향후 우리는 더욱 정확한 결함 진단 위해서 복잡한 데이터에서도 고수준의 특징을 학습할 수 있는 딥러닝 기반 모델을 설계하여 연구를 확장하고, 보다 정확한 시계열 데이터 예측이 가능한 알고리즘으로 발전시키는 것을 목표로 연구를 계획한다.

Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송혁신인재양성(메타버스융합대학원) 사업 연구 결과로 수행되었음 (IITP-2024-RS-2023-00254529)

참고 문헌

- Campora, U., Cravero, C., and Zacccone, R. (2018). Marine gas turbine monitoring and diagnostics by simulation and pattern recognition. *International journal of naval architecture and ocean engineering*, 10(5), 617-628.
- Cheliotis, M., Lazakis, I., and Theotokatos, G. (2020). Machine learning and data-driven fault detection for ship systems operations. *Ocean Engineering*, 216, 107968.
- Cheliotis, M., Lazakis, I., and Cheliotis, A. (2022). Bayesian and machine learning-based fault detection and diagnostics for marine applications. *Ships and Offshore Structures*, 17(12), 2686-2698.
- Ji, Z., Gan, H., and Liu, B. (2023). A deep learning-based fault warning model for exhaust temperature prediction and fault warning of marine diesel engine. *Journal of Marine Science and Engineering*, 11(8), 1509.
- Kim, D., Lee, S., and Lee, J. (2020). An ensemble-based approach to anomaly detection in marine engine sensor streams for efficient condition monitoring and analysis. *Sensors*, 20(24), 7285.
- Kocak, G., Gokcek, V., and Genc, Y. (2023). Condition monitoring and fault diagnosis of a marine diesel engine with machine learning techniques.

Pomorstvo, 37(1), 32-46.

Park, J., and Oh, J. (2022). Analysis of collected data and establishment of an abnormal data detection algorithm using principal component analysis and K-nearest neighbors for predictive maintenance of ship propulsion engine. *Processes*, 10(11), 2392.

Tsaganos, G., Papachristos, D., Nikitakos, N., Dalaklis, D., and Ölcer, A. I. (2018, April). Fault detection and diagnosis of two-stroke low-speed marine engine with machine learning algorithms. In *Proceedings of 3rd international symposium on naval architecture and maritime*.

Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

Velasco-Gallego, C., and Lazakis, I. (2022). Development of a time series imaging approach for fault classification of marine systems. *Ocean Engineering*, 263, 112297.