

LSTM-AutoEncoder를 활용한 선박 메인엔진의 이상 탐지 및 라벨링

Outlier Detection and Labeling of Ship Main Engine using LSTM-AutoEncoder

김도희¹ · 한영재¹ · 김혜미¹ · 강성필² · 김기훈¹ · 배혜림[†]

부산대학교 산업공학과 산업데이터공학융합전공¹, 랩오투원²

요약

운송 산업은 우리나라의 3면이 바다로 둘러싸여 있는 지리적 요건과 자원 소비량의 대부분을 수입에 의존하는 자원 빈곤 문제로 인해 중요한 산업 중 하나이다. 그 중에서도 해운업의 비중은 운송 산업의 대부분을 차지할 정도로 크며, 해운업에서의 유지보수는 선박의 운영 효율성 개선 및 비용 감소에 있어서도 중요하다. 그러나 현재 선박이 유지보수를 위해 일정기간 주기로 검사가 시행되고, 이에 따라 시간과 비용이 발생하며 원인규명도 제대로 되지 않고 있다. 따라서 본 연구에서는 제안 방법론인 LSTM-AutoEncoder를 활용해 실제 선박 운항 데이터에 대해 시점을 고려하여 선박 고장의 원인이 될 수 있는 이상 탐지를 수행한다. 또한 클러스터링을 통해 근집화를 수행하고 이상치에 대해 요인별로 그룹화를 통해 선박 메인엔진 고장의 잠재 원인을 규명한다. 이는 선박의 다양한 정보에 대해 보다 빠르게 모니터링이 가능하고, 이상 정도를 식별할 수 있다. 또한 현재 선박의 고장 감시시스템에 있어서도 구체화된 경보 점 설정과 고장 진단 체계를 갖추고, 유지보수시점을 찾는 데 도움을 줄 수 있을 것이다.

■ 중심어 : 선박, 이상 탐지, LSTM-AutoEncoder, 클러스터링, 빅데이터

Abstract

The transportation industry is one of the important industries due to the geographical requirements surrounded by the sea on three sides of Korea and the problem of resource poverty, which relies on imports for most of its resource consumption. Among them, the proportion of the shipping industry is large enough to account for most of the transportation industry, and maintenance in the shipping industry is also important in improving the operational efficiency and reducing costs of ships. However, currently, inspections are conducted every certain period of time for maintenance of ships, resulting in time and cost, and the cause is not properly identified. Therefore, in this study, the proposed methodology, LSTM-AutoEncoder, is used to detect abnormalities that may cause ship failure by considering the time of actual ship operation data. In addition, clustering is performed through clustering, and the potential causes of ship main engine failure are identified by grouping outlier by factor. This enables faster monitoring of various information on the ship and identifies the degree of abnormality. In addition, the current ship's fault monitoring system will be equipped with a concrete alarm point setting and a fault diagnosis system, and it will be able to help find the maintenance time.

■ Keyword : Ship, Outlier detection, LSTM-AutoEncoder, Clustering, Big-data

2022년 05월 09일 접수; 2022년 05월 27일 수정본 접수; 2022년 06월 08일 게재 확정.

* 이 과제는 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음

† 교신저자 (hrbae@pusan.ac.kr)

I. 서론

우리나라는 3면이 바다로 둘러싸여 있는 지리적 요건과 자원 소비량의 93%인 대부분을 수입에 의존하는 자원 빈곤 문제로 인해 운송 산업이 발달하였다. 그 중에서도 해운업의 비중은 운송 산업의 99.7%를 차지할 정도로 매우 크다. 해운업에서 유지보수는 비용적인 측면과 원인규명에 있어서 매우 중요하다. 유지보수 비용은 선박 운용비용 중 약 10%를 차지하고, 메인 엔진 하나의 가격 역시 선박 가격의 10~15%를 차지할 정도로 유지보수 시점은 선박의 운영 효율성 개선 및 비용 감소에 있어서 중요하다(김동현 외(2020)). 현재 선박의 유지보수는 일정 기간 주기로 정비가 시행되며, 운행을 하지 않음에 따라 시간과 비용이 발생하며 원인 규명 또한 제대로 이루어지고 있지 않다(Deris, S.(1999)). 선박의 효율성을 위해서는 선박 고장 원인이 되는 이상 패턴을 찾고, 이를 통해 고장 진단 및 유지보수 시점을 예측하여 사전에 예방해야 한다.

본 연구는 선박 메인엔진에서 수집된 데이터를 바탕으로 고장의 잠재원인을 규명하는 데 그 목적이 있다. 선박 메인엔진 시계열 데이터를 활용하여 비정상 특징을 추출하고, 이를 이상탐지 알고리즘에 적용하여 고장의 원인 및 시점에 대한 고려가 가능한 LSTM-AutoEncoder(LSTM-AE) 방법론을 제안하였다. 이 후 이상치들에 대해 k-means clustering을 통하여 군집화를 수행하고, 서로 영향을 미치는 요인들을 파악하여 그룹화를 함으로써 고장의 잠재원인에 대해 분석하였다.

본 연구의 결과물을 활용하여 메인 엔진에 이상이 발생할 경우 이상 정도를 식별할 수 있으며 이는 기존 유지보수의 문제점이었던 사고 후 처리에 따른 원인규명이 되지 않았던 점도 해결할 수 있을 것이라 기대된다.

또한 유지보수 및 고장 진단 시점을 예측하기 위한 비정상 데이터에 대한 라벨링이 가능하며,

이를 선박관리자가 활용한다면 선박의 다양한 정보에 대해 단순한 데이터 확인이 아닌 모니터링이 가능하고, 유지보수 시점을 예측하는 데도 도움을 줄 수 있을 것이다. 또한 선박관리자가 이상 탐지 기준을 어떻게 정하느냐에 따라 선박 관리 수준을 정하여 이를 정보 점으로 활용할 수 있을 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 선박 성능 모니터링에 관련된 선행 연구를 살펴본다. 3장에서는 본 연구에서 제안된 방법론에 대해 설명한다. 4장에서는 본 연구에 사용된 데이터에 대해 살펴본 후 데이터 정제와 제안 방법론을 통한 이상 탐지 결과를 제시한다. 이후 이상치에 대해 클러스터링을 통한 군집 및 라벨링 결과를 제시하고, 5장에서는 결론 및 시사점을 정리하고, 향후 연구에 대해 제시한다.

II. 선행 연구

선박 모니터링 관련 선행 연구는 선박 성능에 대한 연구가 주를 이뤘으며, 이상을 탐지하고 고장을 진단하는 연구는 거의 없었다. Mak, L. et al.(2014)는 연비를 개선하기 위해 선박 성능 모니터링 및 분석 연구를 수행하였으며, Key Performance Indicators(KPI)를 활용하여 자동으로 연료 측정이 가능하게끔 하였다. Aldous, L. et al.(2015)는 Monte Carlo 방법론을 활용하여 불확실성 요인을 분석하여 선박 성능 모니터링의 불확실성에 대해 분석하였다. Perera, L. P. & Mo, B.(2017)는 선박 성능 모니터링을 위해 선박 엔진 중심으로 데이터를 분석하였는데 가우스 혼합 모델을 활용하여 엔진 작동영역을 분류하여 보다 정교하게 성능을 모니터링 하였다.

Boullosa, D. et al.(2017)은 선박용 디젤엔진 실린더 윤활공정에 대해 T^2 hotelling 기법을 통해 통계적으로 이탈된 관측치에 대한 공정 변수 개별 모니터링을 수행하였다. 그러나 이 연구는 선박 메인엔진에 영향을 끼치는 변수 중 7개의 변

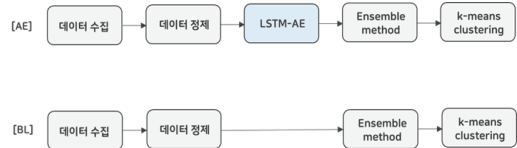
수만을 사용하여 연구하였으며, 시퀀스를 가지는 연속적인 데이터가 아닌 데이터 수집시기가 각각 이루어짐으로써 시점에 대한 고려를 하지 못했다는 한계가 존재한다. 배영목 외(2018)는 선박 엔진 데이터를 붓스트랩 기반의 T^2 다변량 관리도 작성을 통해 선박 메인 엔진 고장의 잠재원인을 규명할 수 있는 제어 상한 설정 사례 연구를 수행하였다. 그러나 이 연구는 고장의 잠재원인을 분석하였지만, 온도에 관해 서만 규명을 함으로써 여러 잠재원인에 대해 분석하지 못했다는 한계점이 존재한다. 김동현 외(2020)는 이상치 탐지 알고리즘을 앙상블한 방법론을 제안하고, 이를 통해 선박 메인엔진의 이상치를 탐지하였다.

본 연구에서는 김동현 외(2020)의 이상 탐지 알고리즘을 Baseline 모델로 정하였다. 이 연구에서는 메인엔진의 이상치를 탐지하였으나 고장의 잠재원인에 대한 분석이 이루어지지 않았고, 시퀀스를 고려한 이상 탐지가 이루어지지 않았다는 한계점이 존재한다. 그러나 이상 탐지에 대해 여러 가지 방법론을 앙상블하여 특정 알고리즘이 지나치게 과적합하는 문제를 해결하였고, 이해관계자들의 도메인 지식을 바탕으로 데이터 분석을 수행하였기 때문에 Baseline 모델로 정하였으며, 이를 본 연구를 통해 구현하고 비교를 수행하여 제안방법론의 성능을 검증하였다.

III. 연구 방법론

선박 메인엔진의 이상 탐지를 위해 본 논문에서는 LSTM-AE를 활용하여 특징을 추출하고, Baseline의 이상 탐지 앙상블 방법론(Baseline)을 통해 이상 탐지를 수행한다. 이후 k-means clustering을 통해 이를 군집화하고, 요인 별 그룹화를 통해 고장 원인에 대해 규명한다. 그림 1은 제안 방법론과 비교방법론인 Baseline 모델의 연구절차 흐름도를 보여주며, 제안방법론은 AE, Baseline은 BL이라 명명하겠다. AE와 BL의 가장

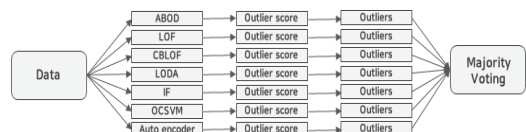
큰 차이점은 본 연구의 방법론은 시계열성을 고려한다는 점이다. 본 연구의 최종 목적은 선박 운항데이터를 활용하여 이상을 탐지 하고, 이를 통해 고장의 잠재원인을 규명하는 것이다. 따라서 BL 방법론을 통해 이상 탐지를 수행할 경우 시계열성은 고려하지 않고 진행되며, 이상 탐지 후 잠재원인을 규명하는 데 있어 설명력이 부족하다. 또한 향후 연구에서 선박 메인엔진에 대한 고장 시점 예측 시 추가적으로 시계열성을 고려해줘야 한다는 한계가 있는 점에서 본 연구와의 차이가 있다.



〈그림 1〉 제안방법론(AE)과 Baseline(BL) 연구 절차

3.1 Ensemble method(BL)

본 연구의 Baseline모델인 김동현 외(2020)는 이상탐지 알고리즘의 과적합 문제를 해결하기 위해 다양한 이상 탐지 알고리즘을 앙상블 하는 방법론을 제안하였다. 각 이상 탐지 알고리즘을 통해 이상치 점수가 상위 1%에 속하는 데이터를 추려내어 이상치로 가정해 각각의 이상치를 구한 후 전체 7개의 모형 중 과반인 4개 이상의 모형에서 이상치로 검출된 데이터 개체를 이상치로 판단하여 이상을 탐지한다. 7개의 모형은 Angle-based Outlier Detection(ABOD), Local Outlier Factor(LOF), Cluster-based Local Outlier Factor(CBLOF), Lightweight Online Detector of Anomalies(LODA), Isolation Forest(IF), One-class

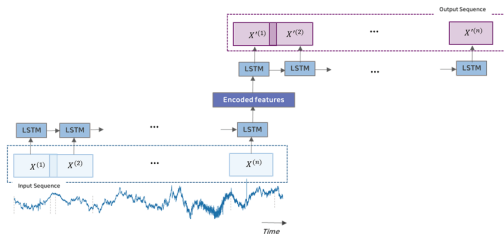


〈그림 2〉 앙상블 기법을 활용한 이상탐지 모델

SVM(OCSSVM), Autoencoder-based Outlier detection(Auto-encoder)를 사용했다. 그림 2는 앙상블 기법을 활용한 이상탐지 모델을 보여준다.

3.2 LSTM-AutoEncoder(AE)

AutoEncoder는 라벨이 없는 훈련 데이터를 사용하여 입력 데이터의 표현을 효율적으로 학습할 수 있는 인공신경망이다(김현재 외(2021)). 일반적으로 입력보다 낮은 차원을 가지므로 차원축소에 효과적이며, 특성 추출기처럼 작동하기 때문에 심층 신경망의 비지도 사전학습에 사용될 수 있다. 또한 대부분의 도메인에서 발생하는 비정상데이터가 충분하지 않은 데이터 불균형 문제에 있어서도 적용이 가능하기 때문에 이상 탐지 방법론으로 많이 활용된다.



〈그림 3〉 LSTM-AutoEncoder(LSTM-AE)의 일반적 구조

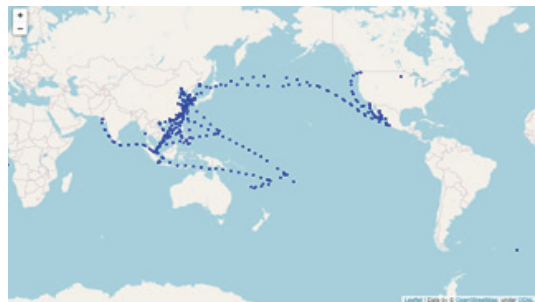
본 연구에서 사용된 LSTM-AE는 인코더에서 입력받은 시퀀스 데이터로부터 특징을 추출하고 저차원의 벡터로 변환하여 학습한다(Nguyen, H.D.(2021)). 디코더에서 학습된 패턴을 기반으로 원 데이터의 시퀀스를 재구성한다. 그림 3은 LSTM-AE 구조를 보여준다.

IV. 실 험

4.1 데이터 수집

본 연구에서는 해양수산빅데이터 거래소의 선

박 운항 데이터를 활용하였다. 데이터 기간은 2019년 1월 1일 2021년 11월 1일까지 10초 주기로 총 5,733,171개이며, 해당 선박에 부착되어 있는 센서 및 운항 데이터로 실시간 모니터링이 가능하다. 본 연구의 데이터 기간 내 선박 항해 경로를 시각화한 그림은 다음과 같다.



〈그림 4〉 선박 항해 경로 시각화(2019.01.01. ~2021.11.01)

4.2 데이터 정제

본 연구에서는 수집된 총 데이터를 선박은 10초 주기로 많은 변동 일어나지 않기 때문에 다운 샘플링하여 평균값 처리를 통해 95,674개로 축소하였다. 이후 데이터의 결측치가 많아 총 데이터 수의 반 이상인(> 47,837) 요인만 선정하여 총 31개의 요인에서 15개의 요인으로 차원 축소하였으며, 표 1의 표시된 부분과 같다.

또한 항해 상태일 때의 고장 진단의 원인이 되는 이상을 탐지하기 위해 기준 RPM이 80이상 120이하인 경우 안정적인 항해 상태로 판단할 수 있다는 전문가의 의견에 따라 데이터를 추출하였고, 센서 이상으로 인해 잘못 수집된 값을 제거하였다. 결측치가 많은 요인에 대해 제거를 통해 1차 선정한 후에도 각 요인 별 데이터양이 다른 부분에 대해서는 시계열성 고려를 위해 보간법(Interpolation)을 활용하여 결측치를 대체하였다.

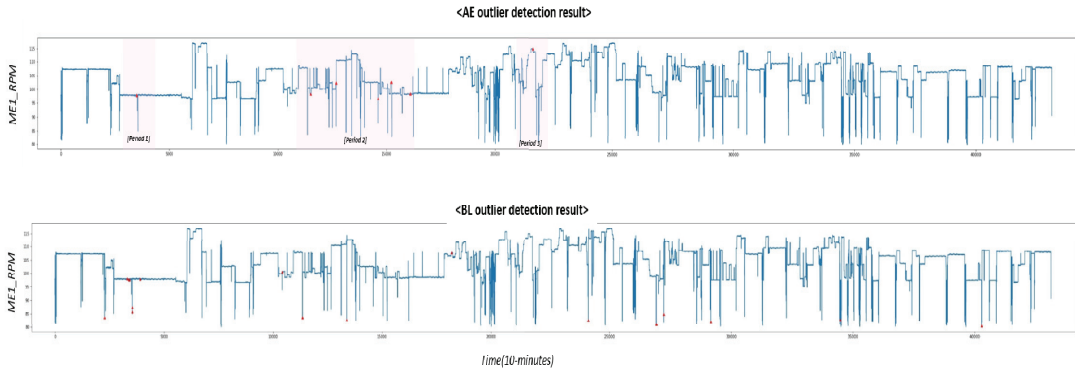
4.3 선박 메인엔진 이상 탐지(Outlier detection)

LSTM-AE를 통해 추출된 요인에 대해 서로 다른 이상 탐지 알고리즘을 사용하여 탐지하고, 다수결을 통해 이상치를 검출하였다. 이상 탐지 기

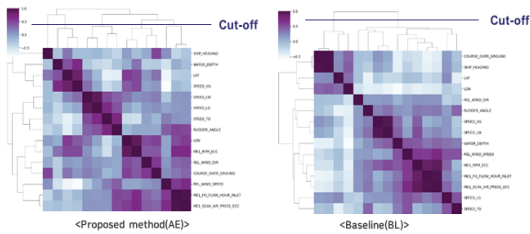
준은 1%에 속하는 데이터를 이상치로 가정하였다. 그림 4는 이상 탐지 결과를 나타내며 제안방법론인 AE의 경우 총 21개, Baseline 모델인 BL의 경우 총 19개의 이상치가 검출되었다. Appendix A를 살펴보면, AE의 경우 시퀀스를 고려하였기

〈표 1〉 선박 메인 엔진 관련 요인 설명 및 데이터 수

No.	Variable	Description	Amount of data (94,775)
1	LAT	위도	94,775
2	LON	경도	94,775
3	SPEED_VG	대지속도	95,049
4	SPEED_LG	경도기준 속도	93,069
5	SPEED_TG	위도기준 속도	93,344
6	SPEED_LW	대수속도	93,069
7	REL_WIND_DIR	바람의 방향	93,895
8	REL_WIND_SPEED	바람의 속도	95,674
9	COURSE_OVER_GROUND	실제 선박의 진행 방향(기준: 지면)	95,674
10	SHIP_HEADING	선수의 방향	95,049
11	WATER_DEPTH	수심	65,696
12	RUDDER_ANGLE	선체 중심선과 키가 이루는 각 (타각)	93,175
13	ME1_FO_FLOW_HOUR_INLET	주기관 연료유 입구 흐름양	95,674
14	ME1_SCAV_AIR_PRESS_ECC	주기관 소기 외부 압력	95,674
15	ME1_RPM_ECC	주기관 RPM	95,674
16	ME1_SCAV_AIR_PRESS	주기관 소기 압력	45,342
17	ME1_RPM_SHAFT	주기관 RPM 축	427
18	ME1_SHAFT_POWER	주기관 축동력	427
19	ME1_SHAFT_TORQUE	주기관 축토크	427
20	ME1_TC1_RPM	주기관 과급기 RPM	45,342
21	ME1_CYL1_EXH_GAS_OUTLET_TEMP	주기관 실린더1 출구 온도	44,642
22	ME1_CYL2_EXH_GAS_OUTLET_TEMP	주기관 실린더2 출구온도	44,642
23	ME1_CYL3_EXH_GAS_OUTLET_TEMP	주기관 실린더3 출구온도	44,642
24	ME1_CYL4_EXH_GAS_OUTLET_TEMP	주기관 실린더4 출구온도	44,642
25	ME1_CYL5_EXH_GAS_OUTLET_TEMP	주기관 실린더5 출구온도	44,642
26	ME1_CYL6_EXH_GAS_OUTLET_TEMP	주기관 실린더6 출구온도	44,642
27	ME1_TC1_EXH_OUTLET_TEMP	주기관 과급기 윤활유 출구온도	44,642
28	ME1_TC1_EXH_INLET_TEMP	주기관 과급기 윤활유 입구온도	44,642
29	ME1_LO_INLET_PRESS	주기관 윤활유 입구 압력	45,342
30	ME1_FO_INLET_PRESS	주기관 연료유 입구 압력	45,342
31	ME1_FO_INLET_TEMP	주기관 연료유 입구 온도	44,640



<그림 5> 선박 메인엔진 이상 탐지 결과(上: AE, 下: BL)



<그림 6> Correlation matrix 결과(左:AE, 右:BL)

<표 2> 피어슨 상관계수를 활용한 AE Sub-grouping 결과

Group No.	Label name	Variable
Sub-group 1	대지 속도	LAT
		WATER_DEPTH
		SPEED_VG
Sub-group 2	대수 속도	SPEED_LW
		SPEED_LG
		SPEED_TG
		RUDDER_ANGLE
Sub-group 3	경도 관련	REL_WIND_DIR
		ME1_RPM_ECC
		COURSE_OVER_GROUND
		LON
Sub-group 4	선박의 방향	SHIP_HEADING
Sub-group 5	메인 엔진 요소	ME1_FO_FLOW_HOUR_INLET
		ME1_SCAV_AIR_PRESS_ECC
		REL_WIND_SPEED

때문에 총 3구간으로 시점 별로 나타나는 것을 볼 수 있고, BL의 경우 메인엔진의 RPM 기준으로 이상치가 나타나는 것을 볼 수 있다.

4.4 요인 간 상관관계를 고려한 Sub-grouping

이상치 탐지 후 요인 간에 미치는 영향을 파악하기 위해 피어슨 상관계수를 활용하여 요인의 Sub-grouping을 실시하였다. 그림 5는 요인 간 상관관계에 대한 매트릭스 시각화를 나타내며 덴드로그램을 통해 Cut-off 설정 후 요인 별 라벨링을 위해 Sub-group으로 그룹화를 수행했다.

표 2는 제안 방법론인 AE의 Sub-grouping 결과를 나타내며 상관계수 0.5 기준으로 총 5개의 그룹으로 구성되었다. 표 3은 Baseline 모델인 BL의 Sub-grouping 결과를 나타내며 상관계수 0.5 기준으로 총 4개의 그룹으로 구성되며 각 방법론의 Sub-group은 서로 맵핑되는 것을 볼 수 있다.

4.5 k-means clustering을 활용한 군집화

이상치 탐지 후 고장 원인에 대한 라벨링을 위해 k-means clustering을 수행하였다. Elbow method를 통해 k를 설정하였으며 k-means clustering을 수행한 후의 결과를 제안방법론 AE와 Baseline BL을 비교한다. 그림 6은 Elbow method 결과 두 방법론 모두 k가 3으로 설정되었음을 보

여준다(Marutho, D. et al(2018)).

표 4는 제안방법론 AE의 군집화 결과를 보여 준다. AE-Cluster 1은 16개, AE-Cluster 2는 1개, AE-Cluster 3은 4개로 총 21개의 이상치가 군집

화 되었다. 그림 4의 이상치 탐지 결과의 시점과 같이 군집화가 수행되는 것으로 보아 시퀀스 요인이 크게 작용한 것을 알 수 있다. 군집화가 잘 되었는지 확인하기 위해 군집 간 거리를 식(1)의 유클리디안 거리로 구하였다.

<표 3> 피어슨 상관계수를 활용한 BL Sub-grouping 결과

Group No.	Label name	Variable
Sub-group 1	선박의 방향 및 위치	LAT
		LON
		COURSE_OVER_GROUND
		SHIP_HEADING
Sub-group 2	대수 속도	SPEED_LG
		SPEED_TG
Sub-group 3	외부 영향	REL_WIND_DIR
		RUDDER_ANGLE
		SPEED_VG
		SPEED_LW
Sub-group 4	메인 엔진 요소	WATER_DEPTH
		ME1_FO_FLOW_HOUR_INLET
		ME1_SCAV_AIR_PRESS_ECC
		REL_WIND_SPEED
		ME1_RPM_ECC

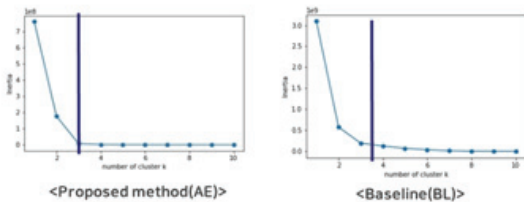
$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

유클리디안 거리는 두 점 (x_1, y_1) , (x_2, y_2) 이 있을 때 두 점 사이의 거리가 식(1)의 d와 같다. AE 군집 간거리가 매우 크기 때문에 군집화가 잘 된 것을 알 수 있으며 거리는 표 5에 나타난다.

<표 5> 제안방법론(AE) 클러스터 간 유클리디안 거리

Euclidean Distance	AE-Cluster1	AE-Cluster2	AE-Cluster3
AE-Cluster1		23529.52	7285.98
AE-Cluster2	23529.52		29941.70
AE-Cluster3	7285.98	29941.70	

각 클러스터에 대해서 sub-grouping 결과와 함께 분석하면, AE-Cluster 1은 총 16개의 이상치 중 각각 특정한 날에 1시간 동안 6개, 5개, 5개 이상치가 나타나고 있으며, 이들은 한 달 정도의 주기를 가지고 있다. 또한 선박의 위치가 멕시코 Cortez 해에 위치하고 있으며, 이 지역은 조수간만의 차가 큰 지역이다. 조수간만의 차로 인해 연료소비량에 비해 속도와 RPM이 높으며, 외부 요인에 대해서도 민감하게 반응한다. AE-Cluster 2는 1개의 이상치만 존재하며, 어느 군집으로도 묶이지 않는 이상치이다. 외부 요인인 바람이 아주 크게 작용하는 것으로 보이며, 이로 인해 다른 요인의 값들도 영향을 받은 것으로 보아 고장의 잠재적 원인이 될 수 있다. AE-Cluster 3은 4개의 이상치가 존재하며 모두 같은 날 오전 8시 반부터 한 시간 사이에 측정된 값이다. RUDDER_ANGLE 요인 값이 굉장히 크므로 급히 방향을



<그림 7> Elbow method 결과 그래프(k=3)

<표 4> 제안방법론(AE) k-means clustering 결과

Cluster No.	Count
AE-Cluster1(=Period 1)	16
AE-Cluster2(=Period 2)	1
AE-Cluster3(=Period 3)	4

조정해야하는 상황이므로 이 또한 고장의 잠재적 원인이 될 수 있다.

표 6은 Baseline BL의 군집화 결과를 보여준다. BL-Cluster 1은 7개, BL-Cluster 2는 8개, BL-Cluster 3은 4개로 총 19개의 이상치가 군집화 되었다.

〈표 6〉 Baseline(BL) k-means clustering 결과

Cluster No.	Count
BL-Cluster1	7
BL-Cluster2	8
BL-Cluster3	4

각 클러스터 간의 거리는 유클리디안 거리로 구했으며, 표 7과 같다.

〈표 7〉 Baseline(BL) 클러스터 간 유클리디안 거리

Euclidean Distance	BL-Cluster1	EL-Cluster2	EL-Cluster3
EL-Cluster1		27812.92	13965.49
EL-Cluster2	27812.92		14148.40
EL-Cluster3	13965.49	14148.40	

각 클러스터의 특징을 파악하기 위해 sub-grouping 결과와 함께 분석해보면, BL-Cluster 1은 총 7개의 이상치가 존재하며 해안이 아닌 태평양 한가운데 위치해 있고, RPM이 80대이다. BL-Cluster 2는 총 8개의 이상치가 존재하며 아시아 연안에 위치하여 빠른 속도 변화가 일어났으며 전후 데이터를 통해 메인엔진에 영향을 미친 것으로 보인다. 또한 RPM이 90대로 군집화 되었다. 마지막으로 BL-Cluster 3는 BL-Cluster 1과 같이 한가운데에 위치했지만 가속되고 있는 상태로 보이며 RPM 100대로 나타났다. Appendix B를 통해 각 방법론에 대한 군집을 선으로 표현하고 라벨링 된 결과를 박스로 표현하였다. 같은 그룹으로 묶인 요인에 대해서는 군집의 패턴 역시 비

슷함을 확인할 수 있다.

세 군집으로 묶이는 점은 두 방법론 모두 같지만 Baseline 모델의 경우 고장의 잠재원인을 규명하는 데 있어 종합적인 해석이 불가능하고, 메인 엔진 RPM에 따라 나뉘는 것을 볼 수 있다. 이는 선박의 가속 또는 감속 상태를 알 수 있지만 선박 고장의 원인 규명을 위해서는 불분명한 요인이다. 반면에 제안방법론의 경우, 시계열성을 고려하여 군집화가 진행되었으므로 고장의 잠재원인이 나타나는 주기를 파악할 수 있으며, 일정한 주에 따라 군집되는 것을 볼 수 있다. 이는 그 시기에 선박이 어디를 운항하고 있었는지에 따라 지리적 요건 및 외부환경 요인을 고려할 수 있으며, 이에 따라 영향을 미치는 요인에 대해 관계를 파악함으로써 고장 원인에 대해 조금 더 구체화된 원인 규명이 가능하다는 것을 보여준다.

V. 결론

본 연구는 우리나라의 대표적인 운송 산업인 해운 산업의 선박 운영 효율성 및 비용 감소를 위해 선박 메인엔진의 이상을 탐지하고, 이를 군집화하고 요인별 영향관계를 파악함으로써 고장의 잠재원인을 규명하였다. 본 연구의 제안 방법론인 AE는 LSTM-AutoEncoder를 통해 실제 선박 운항 데이터를 활용하여 시계열성을 고려하여 특징을 추출하고, 이를 이상 탐지 알고리즘을 통해 이상치를 검출한다. 이후 이상탐지 결과를 k-means clustering을 통해 군집화하고 요인간의 상관관계를 분석하여 Sub-grouping을 수행한다. 또한 Baseline 모델인 BL 결과와 이를 비교한다. 비교를 통해 Baseline의 RPM 속도에 따라 이상치가 묶여 다른 영향 요인들을 고려한 종합적인 분석이 불가능하고, 원인 규명이 되지 않는 한계점을 보완하였다.

또한 실제 데이터에 대한 적용을 통해 질 낮은 데이터에 대한 데이터 품질을 개선하였으며, 고

장 원인 규명에 있어 보다 구체화된 결과를 제시하였다. 하지만 이 연구가 산업적인 측면에서 유의미하기 위해서는 선박 관리자의 의사결정을 대체하는 것이 아니라 실시간으로 전달되는 데이터에 대해 이상을 식별하는 도구로써 활용될 수 있으며, 고장 원인으로 진단하기 위해서는 전문가의 의견이 추가되어야 할 것이다.

향후 연구에서는 센서 이상 등으로 인해 품질이 낮은 데이터의 품질을 높일 수 있는 데이터 복원 연구와 이상 탐지 기준에 대한 최적화 연구가 필요할 것이다. 마지막으로 이상 검출 결과를 바탕으로 고장 진단 및 유지보수 시점에 대한 시계열 예측 연구를 수행함으로써 현재 임의로 시행되는 선박 유지보수 시점에 대해 비용을 최소화하는 구체적인 시점 예측이 필요할 것이다.

참 고 문 헌

- [1] 김동현, 이지환, 이상봉, & 정봉규. (2020). 앙상블 기법을 이용한 선박 메인엔진 빅데이터의 이상치 탐지. *수산해양기술연구*, 56(4), 384-394.
- [2] Deris, S., Omatu, S., Ohta, H., Kutar, L. C. S., & Abd Samat, P. (1999). Ship maintenance scheduling by genetic algorithm and constraint-based reasoning. *European Journal of Operational Research*, 112(3), 489-502.
- [3] Mak, L., Sullivan, M., Kuczora, A., & Millan, J. (2014, September). Ship performance monitoring and analysis to improve fuel efficiency. In 2014 Oceans-St. John's (pp. 1-10). IEEE.
- [4] Aldous, L., Smith, T., Bucknall, R., & Thompson, P. (2015). Uncertainty analysis in ship performance monitoring. *Ocean Engineering*, 110, 29-38.
- [5] Perera, L. P., & Mo, B. (2017). Marine engine-centered data analytics for ship performance monitoring. *Journal of Offshore Mechanics and Arctic Engineering*, 139(2).
- [6] Boullosa, D., Larrabe, J. L., Lopez, A., & Gomez, M. A. (2017). Monitoring through T2 Hotelling of cylinder lubrication process of marine diesel engine. *Applied Thermal Engineering*, 110, 32-38.
- [7] Bae, Y. M., Kim, M. J., Kim, K. J., Jun, C. H., Byeon, S. S., & Park, K. M. (2018). A case study on the establishment of upper control limit to detect vessel's main engine failures using multivariate control chart. *Journal of the Society of Naval Architects of Korea*, 55(6), 505-513.
- [8] 김현재, 김동훈, 임채욱, 신용탁, 이상철, 최영진, & 우승범. (2021). 해양 이상 자료 탐지를 위한 오토인코더 활용 기법 최적화 연구. *한국해양·해양공학학회논문집*, 33(6), 265-274.
- [9] Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *International Journal of Information Management*, 57, 102282.
- [10] Marutho, D., Handaka, S. H., & Wijaya, E. (2018, September). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In 2018 international seminar on application for technology of information and communication (pp. 533-538). IEEE.

저자 소개



김도희(Dohee Kim)

- 2019년 : 부산대학교 산업공학과 (공학사)
- 2019년~현재 : 부산대학교 산업공학과 (석박사통합과정)
- 관심분야 : 빅데이터, 인공지능, 최적화 등



한영재(Yeongjae Han)

- 2019년 : 경북대학교 통계학과 (이학사)
- 2019년~현재 : 부산대학교 산업공학과 (석사과정)
- 관심분야 : 머신러닝, 딥러닝 등



김혜미(Hyemee Kim)

- 2019년 : 부산대학교 산업공학과 (공학사)
- 2019년~현재 : 부산대학교 산업공학과 (석박사통합과정)
- 관심분야 : 하이퍼파라미터 최적화, 딥러닝 등



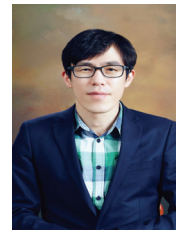
강성필(Seong-Phil Kang)

- 2021년~현재 : 부산대학교 산업공학과 (박사과정)
- 관심분야 : AI, 빅데이터, 운송



김기훈(Ki-Hun Kim)

- 2012년 : 포항공과대학교 산업경영공학과(공학사)
- 2019년 : 포항공과대학교 산업경영공학과(박사)
- 2021년~현재: 부산대학교 산업공학과(조교수)
- 관심분야: 머신러닝, 데이터마이닝, 스마트시스템



배혜림(Hyerim Bae)

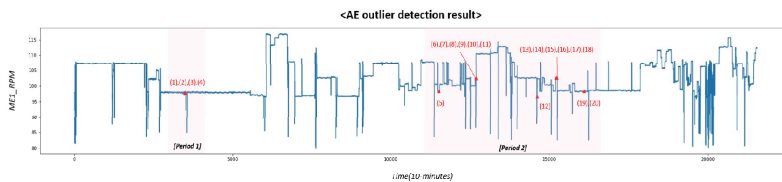
- 1996년 : 서울대학교 산업공학과 (공학사)
- 1998년 : 서울대학교 산업공학과 (석사)
- 2002년 : 서울대학교 산업공학과 (박사)
- 2004년~현재 : 부산대학교 산업공학과 (교수)
- 관심분야 : 프로세스 마이닝, 빅데이터, 인공지능 등

Appendix A. 선박 메인엔진 이상 탐지 결과

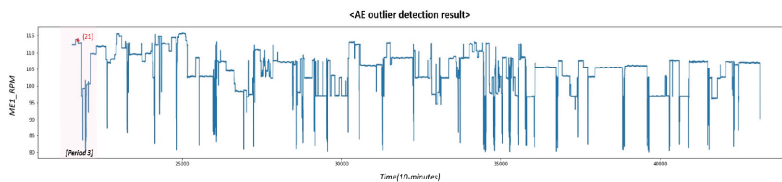
A-1. 제안방법론(AE) 선박 메인엔진 이상 탐지 결과

LSTM-AutoEncoder(AE)를 활용하여 특징을 추출한 후 BL의 앙상블 방법론을 활용하여 이상 탐지한 결과를 나타낸다. 데이터는 2019년 1월 1일부터 2020년 11월 1일까지의 선박운항데이터를 활용하였다. 이상 탐지 기준은 데이터 기준 1%로 하였으며, 가독성을 위해 그래프를 이등분하여 시각화하였다. 이상치는 총 21개로 탐지되었다.

① 데이터 기간: 2019.01.01. 01:20:00~2019.11.10. 07:00:00



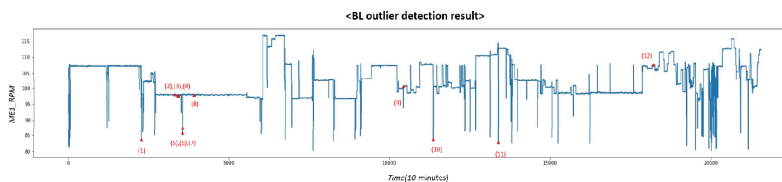
② 데이터 기간: 2019.11.10. 07:10:00 .~ 2020.11.01. 00:30:00



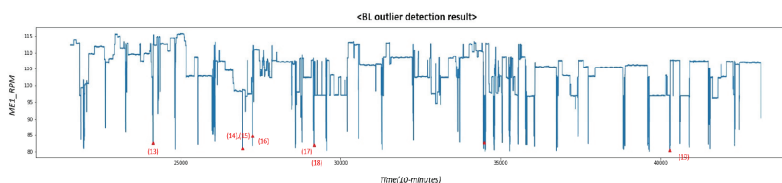
A-2. Baseline(BL) 선박 메인엔진 이상 탐지 결과

Baseline(BL)의 앙상블 방법론을 활용하여 이상 탐지한 결과를 나타낸다. 데이터는 위와 동일하게 2019년 1월 1일부터 2020년 11월 1일까지의 선박운항데이터를 활용하였다. 이상 탐지 기준은 데이터 기준 1%로 하였으며, 가독성을 위해 그래프를 이등분하여 시각화하였다. 이상치는 총 19개로 탐지되었다.

① 데이터 기간: 2019.01.01. 01:20:00~2019.11.10. 07:00:00



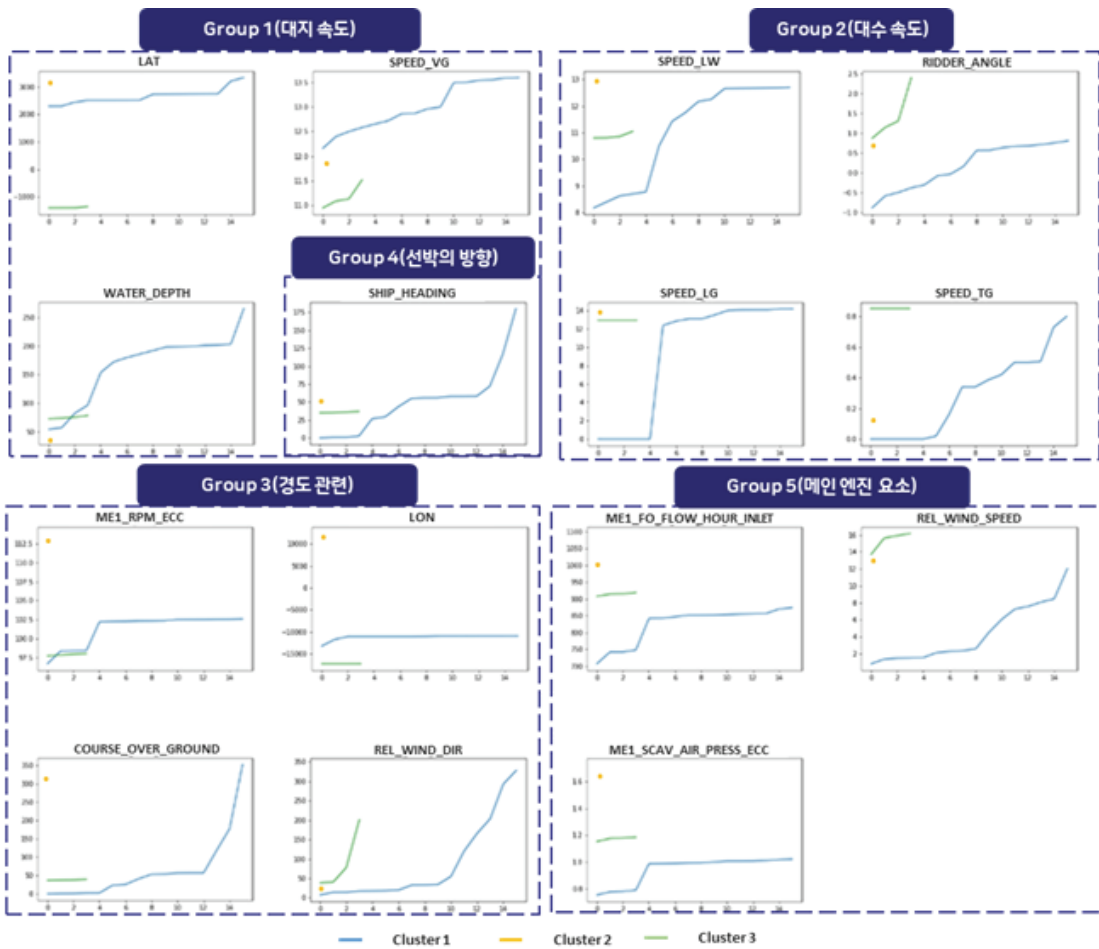
② 데이터 기간: 2019.11.10. 07:10:00 .~ 2020.11.01. 00:30:00



Appendix B. 선박 메인엔진 이상 군집화 및 라벨링 결과 시각화

B-1. 제안방법론(AE) 선박 메인엔진 이상 군집화 및 라벨링 결과 시각화

제안방법론(AE)를 활용하여 선박 메인엔진 요인에 대한 상관관계를 고려한 그룹화 결과 위 표 2와 같이 5개로 분류되었다. 또한 k-means clustering을 활용하여 군집화한 결과 세 군집으로 나뉘어졌으며, 위 표 4와 같다. 아래 그림의 박스는 그룹화 결과를 나타내며, 이상치에 대해 각 요인에 대해 시각화한 그래프이다. 파란색 선은 16개의 이상치가 포함되어있는 군집 1, 초록색 선은 4개의 이상치가 포함되어있는 군집 3을 각각 나타내며, 군집 2의 경우 한 개의 이상치만 존재하기 때문에 점으로 표시되어 보인다.



B-2. Baseline(BL) 선박 메인엔진 이상 군집화 및 라벨링 결과 시각화

Baseline(BL)을 활용하여 선박 메인엔진 요인에 대한 상관관계를 고려한 그룹화 결과 위 표 3와 같이 4개로 분류되었다. 또한 k-means clustering을 활용하여 군집화한 결과 세 군집으로 나뉘어졌으며, 위 표 5와 같다. 아래 그림의 박스는 그룹화 결과를 나타내며, 이상치에 대해 각 요인에 대해 시각화한 그래프이다. 파란색 선은 7개의 이상치가 포함되어있는 군집 1, 주황색 선은 8개의 이상치가 포함되어있는 군집 2, 초록색 선은 4개의 이상치가 포함되어있는 군집 3의 각각 결과를 나타낸다.

