|  |
| --- |
|  |
|  |
| 머신러닝을 기반으로 한 HiMSEN엔진의 고장 진단 및 원인 시스템 분석 |
| Feb. 2021 |
| 김태균, 여승협, 서재원  3Miroa |

머신러닝을 기반으로 한 HiMSEN엔진의 고장 진단 및 원인 시스템 분석

Abstract

대형선박의 내연기관으로 사용되는 ‘HiMSEN엔진’의 상태를 나타내는 25가지의 센서데이터를 기반으로 하여 HiMSEN엔진의 고장 여부 및 고장 위치를 파악하는 알고리즘을 Python library를 이용해 개발하였다. 데이터는 turbo charger(T/C), Air Cooler, Combustion, Cooling Water의 네 가지 system으로부터의 압력과 온도를 1Hz로 측정된 25가지의 측정값으로 되어있었으며, 연료모드와 정상/고장 상태를 포함하였다. 데이터를 단순조건(ㄱ), 확률분포(ㄴ), 일반데이터(ㄷ) 그리고 일반데이터의 상호작용으로 유도된 변수(ㄹ)로 나누었고, 각각 단순조건식(ㄱ), 확률분포에 대한 z점수(ㄴ), 신경망과 앙상블 기법(ㄷ,ㄹ)로 개별 변수 및 상호작용하는 변수에 대해 점수를 책정하였다. 그 결과 ANN과 앙상블 알고리즘만으로 각 네 가지 system의 건강지수에 대한 기준을 제안, 전체 및 부분 시스템의 건강지수를 측정하는 방식을 제안하였다. 기존의 선행연구들에서 사용한 Wavelet Transform 및 Fast Fourier Transform과 같은 주파수 영역의 데이터를 사용하지 않고 고장을 실시간으로 진단할 뿐만 아니라, 세부 시스템의 고장여부에 대해 라벨링 되지 않은 고장에 대해서도 고장의 위치와 원인을 밝힌다는 점에서 기존 연구들과의 차별성이 있을 것으로 보인다.

Keywords: 엔진, Turbo Charger, 머신러닝, 앙상블, 고장 진단, 건강지수

## Introduction

## 빅데이터 및 인공지능의 발달로 기체에 센서를 부착하여 산업용 기계의 고장을 진단하는 알고리즘들이 개발되고 있다. 이는 현장에서 근로자들의 숙련도에 따른 고장진단 능력 차이를 보완할 뿐만 아니라, 대형 시스템에서 부분적으로 인해 연쇄적으로 발생할 수 있는 문제를 예방하는 매력적인 방식으로 부상하고 있다.

## 이러한 연구는 CNC나 용접기계 등 산업용 기계로부터 선박과 자동차와 같은 엔진 시스템에 이르기까지 다양하게 진행되고 있다. 특히 선박과 같은 대형 엔진을 갖춘 시스템의 경우, 고장에 선별적으로 대응하지 못할 시 전체 system에 shutdown 현상이 일어날 수 있어, 향후 실시간 고장 진단에 대한 수요가 증가할 것으로 예상할 수 있다. 또한 반대로 실시간 고장 알고리즘을 개발하고 개선함으로써 고장을 예방하고 전체 시스템을 보존하는 큰 기대효과를 얻을 수 있다.

## 본 연구에서 분석하고자 하는 엔진은 현대의 HiMSEN엔진으로 디젤과 천연가스를 선택적으로 사용할 수 있는 내연기관이다. HiMSEN엔진은 Turbo Charger(T/C), Air Cooler, Combustion, Cooling Water의 네 가지 시스템으로 구성되어 있으며 각 부분의 온도와 압력, 부하, 속도 등의 시계열 데이터를 1Hz로 측정하는 시스템을 갖추고 있다. 본 연구에서는 통계적 방법을 통해 각 시스템에서의 이상 현상을 정의하는 방법을 제안하고, Python의 머신러닝과 앙상블 기법들을 활용, 전체 시스템의 고장뿐만 아니라 전체데이터로부터 고장이 생긴 시스템을 식별할 수 있는 알고리즘을 제안하려 한다.

## 본 분석 보고서는 선행연구의 한계점을 보완하기 위해 크게 두 가지의 방향을 갖는다. 첫 번째로, 각 시스템별로 이상을 진단하는 기법을 제안해 특정 시스템에 문제가 생겼을 때 그 시스템에 대해 선별적, 선제적 대응이 가능하도록 한다. 둘째, 앞서 언급한 네 가지 시스템(엔진, T/C, Air Cooler, Cooling Water)에 대해 개별적 건강지수와 통합 건강지수를 제안해 엔진의 종합적 상태를 진단하는 알고리즘을 제안하고자 한다.

# Method

팀에서는 이 문제를 비지도학습(각 시스템에서의 고장/정상 클러스터링)과 지도학습(정상/고장 이진분류)이 결합되고, 분류(정상/고장 분류)와 회귀(건강 지수)가 결합된 복합 문제로 보았다. 우리 팀에서는 다음과 같이 문제를 정의하였다.

## 전체 시스템의 고장여부와 시스템에 속한 센서 데이터들을 사용해 “개별 시스템의 고장 여부”를 판단하고 각각을 점수화하는 기준을 마련한다.

1. A에서 판단한 개별시스템의 결과를 종합하여 낸 결과가 실제 전체 시스템의 고장 여부와 얼마나 일치하는지 확인하고, 이를 점수화하는 기준을 마련한다.

## 2.1 데이터의 전처리

데이터셋은 train과 test 두 엔진의 데이터로 지급되었다. 우선 데이터 연산을 용이하게 하기 위해 ‘normal’ 과 ‘abnormal’의 state를 1과 0으로 처리하였다. 고장의 원인(고장이 난 시스템)은 따로 주어지지 않았다. 분석의 주 목적은 정상에서 고장이 일어나는 순간을 포착하기 위한 것이므로, 정상에서 고장으로 바뀌는 순간을 찾기 위해 datetime에 대해 오름차순으로 정렬하였다. 그러나 datetime상에서 정상에서 고장으로 바뀌는 시기는 관찰되지 않았다. 또한 결측값이 있는 데이터와 datetime이 같은 데이터에 대하여 전처리를 진행하였다. datetime이 같은 데이터를 삭제할 수 있는 이유는 데이터가 1Hz 측정된 데이터이고 datetime이 같은 데이터들을 비교한 결과 차이가 전체 데이터의 표준편차보다 약 1000배정도 작으므로 무시할 수 있어 제거해도 무방하다고 판단하였기 때문이다. 결측값이 있는 데이터는 없는 것으로 파악되었고 test 데이터에서는 하나의 데이터가, train 데이터에서는 1085개의 데이터가 datetime이 같은 것임을 찾아 해당 데이터를 삭제하였다. 이후 데이터를 정상과 고장데이터로 나누었으며, 각각을 엔진모드에 따라 분류하였다.

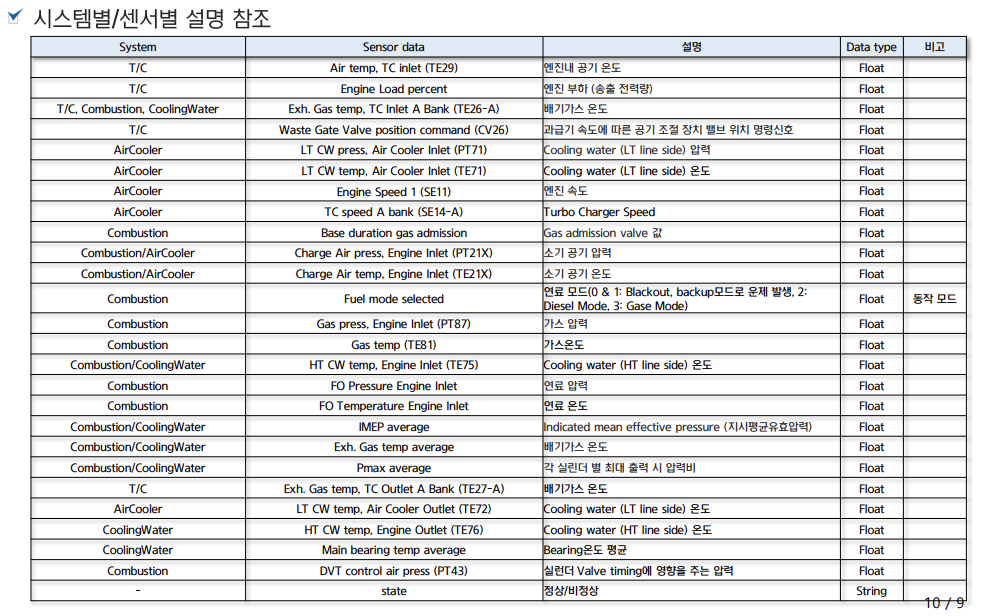


그림 1 시스템별/센서별 설명 참조 (대회측 제공)

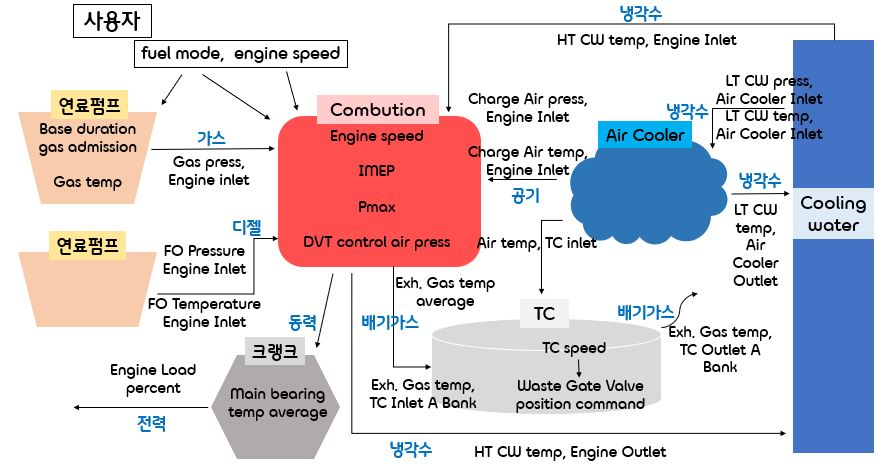


그림 2 전체 시스템 개념도

결측값의 보간과 이상치의 제거는 데이터의 종류를 감안하여 진행되었다. 데이터는 크게 이산적인 분포를 띄는 값과 연속적인 분포를 띄는 값으로 나눌 수 있다. 이산적인 분포를 띄는 엔진 모드, DVT control, mode2의 Waste Gate Valve, Base Duration등은 결측값이 생겼을 때 전 행과 다음 행이 같은 값이면 그 값으로, 그렇지 않으면 그 행의 데이터를 삭제하였다. 연속적인 분포를 띄는 데이터들은 데이터들의 분포를 고려 5초 이하의 결측값을 linear interpolation을 사용하여 보간하였다. Cubic spline과 같은 곡선을 이용한 방식을 사용하지 않은 이유는 구간의 끝값에서 급격한 증감이 있는 상태로 결측값이 발생한 경우 극단적으로 주변 값과 이질적인 값이 나오는 경향이 발생할 수 있었기 때문이다. 선형 보간을 5초 이하로 설정한 것은, 지나치게 긴 구간을 선형으로 보간하면 실제 경향을 모사하지 못할 우려가 있었기 때문이다. 이상치의 경우 IQR을 이용하여 제거하려 하였고, 고장데이터가 의도적으로 많은 이상치를 포함하고 있는 것으로 보여 고장데이터에는 IQR을 사용하는 것이 적절치 않다고 판단, 상식적인 범주를 벗어나는 이상치(절대 영도 이하의 온도, 압력이 0보다 작은 경우)만을 제거하였다. 위 그림1을 참조하면 Air temp, TC inlet (TE29), Charge Air temp, Engine Inlet (TE21X) 등 12개의 온도 센서 데이터가 존재하고 Charge Air press, Engine Inlet (PT21X), Gas press, Engine Inlet (PT87) 등 6개의 압력 센서 데이터가 존재한다. 또한 데이터를 분석했을 때 온도 센서 중 Air temp, TC inlet (TE29)의 평균이 약 40도인 것으로 보아 온도센서의 단위가 이므로 모든 센서 데이터가 -273.15보다 커야 하고, 압력 센서 데이터는 0보다 크거나 같아야 한다. 그러나 DVT control air press (PT43) 센서 데이터는 주변 구동기기에 붙은 센서들로 gauge센서 이므로 0보다 작아도 된다고 판단하였다. 이를 기반으로 해서 분석한 결과 온도 센서 데이터에서는 이상치가 발견되지 않았다. 그러나 Gas press, Engine Inlet (PT87) 센서 데이터에서 66행이 0보다 작아 해당 행들을 모두 제거하였다. DVT control air press (PT43) 데이터는 많은 데이터가 0보다 작아 이상치 처리를 하는 것이 오히려 문제가 발생할 수 있다고 판단하여 이상치 처리를 진행하지 않았다. 이후 다른 데이터에 대하여 분석을 하였다. Engine Load percent, Engine Speed 1 (SE11), MEP average, Pmax average는 각각 엔진 부하, 엔진 속도, 압력비 이므로 0보다 크거나 같아야 한다고 판단하였다. 그러나 실제 데이터에서는 Engine Load percent 데이터에서는 4개의 데이터가 0보다 작고, IMEP average에서는 277개의 데이터가 0보다 작아 해당 행들을 모두 제거하였다.

마지막으로 데이터의 단위, 크기 차이로 인한 영향을 없애기 위해 전체 데이터를 스케일링하였다. 스케일링 방식은 Robust scaler를 사용하여, 중앙값을 0, 1~3분위의 간격을 1로 적용하였다. Robust Scaler를 사용한 이유는 이상치의 영향을 최소화하는 기법이며, 데이터가 정규분포를 따르지 않아도 유연한 적용이 가능한 방식이기 때문이다.

## EDA(탐색적 데이터 분석)

전반적인 고장의 유형과 변수간 상호작용 방식은 D.T.Hountalas의 연구에서 Air Cooler, Combustion, Turbo Charger의 고장 유형과 상호작용 방식을 참고하여 결정하였다.[[1]](#footnote-1) 예를 들어 T/C compressor의 성능저하는 30%만 이루어 지더라도 T/C speed, maximum pressure, air temperature에 모두 영향을 주지만, Air Cooler의 Pressure는 0~100%까지 변하더라도(pressure drop) air temperature, gas temperature, T/C speed에 10%의 영향도 주지 못하였다. 이러한 결과로부터 T/C의 고장여부를 T/C speed, maximum pressure, air temperature의 변동을 통해 역추적 할 수 있는 요소로 생각할 수 있고, air cooler의 pressure는 엔진의 다른 부분 및 고장 여부에 크게 영향을 주지 않는 독립적인 요인으로 간주할 수 있다.

## 2.2.1 데이터의 분류 및 새로운 변수의 생성

## 내연기관의 특징에서부터 데이터의 특징을 추출하여 다음과 같이 세 가지로 데이터를 분류하였다.

표 1 변수 분류

|  |  |  |
| --- | --- | --- |
| Mode2 | Mode3 | 변수명 |
| ㄱ | ㄴ | Base duration gas admission |
| Waste Gate Valve position command (CV26) |
| ㄴ | ㄴ | Main bearing temp average |
| LT CW press, Air Cooler Inlet |
| LT CW temp, Air Cooler Inlet (TE71) |
| LT CW temp, Air Cooler Outlet (TE72) |
| HT CW temp, Engine Inlet (TE75) |
| HT CW temp, Engine Outlet (TE76) |
| Main bearing temp average |
| Charge Air press, Engine Inlet (PT21X) |
| Charge Air temp, Engine Inlet (TE21X) |
| ㄷ | ㄷ | TC speed A bank (SE14-A) |
| Air temp, TC inlet (TE29) |
| Exh. Gas temp, TC Outlet A Bank (TE27-A) |
| Engine Load percent |
| Engine Speed 1 (SE11) |
| Exh. Gas temp average |
| IMEP average |
| Pmax average |
| DVT control air press (PT43) |
| Exh. Gas temp, TC Inlet A Bank (TE26-A) |
| ㄴ | FO Pressure Engine Inlet |
| ㄴ | FO Temperature Engine Inlet |
| ㄴ | ㄷ | Gas press, Engine Inlet (PT87) |
| Gas temp (TE81) |
| ㄹ | ㄹ | (Load-b)/(IMEP+1) |
| (Load)/((IMEP+1)x(Engine speed +1)) |
| (Pmax average-b)/(IMEP+1) |

## ㄱ의 특징은 고정된 값을 갖는다는 점이다. Mode 2(디젤모드)의 경우 디젤을 주 원료로 사용하고 가스를 사용하지 않는다. 따라서 가스를 엔진으로 공급하는 신호에 해당하는 Base duration gas admission센서의 값은 항상 0이여야 한다. 마찬가지 이유로 Waste Gate Valve position command(CV26)또한 항상 0이다. 이 경우 학습시키는 것이 아무런 의미가 없을 뿐 더러, 완전 이산적인 값이기 때문에 따로 분리하여 특정 값이 나오지 않으면 고장으로 분류하였다.

## ㄴ의 특징은 특정 범위를 벗어나지 않으면 안전하다는 점이다. 베어링의 경우 마찰열에 의해서 상승하는 온도와 전도와 복사에 의해 소산되는 열이 양적을 같게 되는 평형온도가 존재하게 된다. 소산되는 열의 양이 적으면 열이 점차 축적되어 원활한 운전을 방해하거나 베어링을 태워 고착시킬 수 있다[[2]](#footnote-2). 이 데이터의 경우 주변 조건이 구해지지 않아 정확하게 평형온도를 추산할 수는 없지만 정상상태의 온도를 바탕으로 짐작해 볼 수는 있다. 그리고 정상상태의 범위를 벗어난다면(베어링의 경우는 단측으로 온도가 높은 경우만을 고려) 고장으로 진단할 수 있다. 그 방식은 표준정규분포를 사용하였다. 변수를 X, 정상데이터에서 X의 표준편차를 라고 하였을 때 변수는 X+1.96 와 X-1.96 사이에 95%의 데이터가 분포하는 정규분포의 형태를 띄게 된다. 이를 바탕으로 전체 정상 데이터의 범위 밖에 해당하는 값은 고장으로 판단할 수 있다.

## ㄷ은 대부분 TC와 Combustion에 관련된 센서로 부분적인 연관성을 갖는 데이터가 포함되었다는 특징이 있다. 예를 들어 3 가지의 배기가스 간에는 어느정도 경향이 일치하며 특정 구간에서는 비례하는 특징이 있다. 하지만 세 변수 사이에는 뚜렷한 선형적 관계나 상관관계는 나타나지 않는다. 셋 모두 엔진 스피드와 연관이 있다고는 볼 수 있으나 엔진스피드와도 복잡한 상관관계를 가져 수학적 모델링이 쉽지 않다. 이와 같이 서로 독립적이지 않으며, 관계를 간단한 수식으로 나타내기 어려운 변수의 경우 신경망 알고리즘을 통해 학습시키는 것이 해법이 될 수 있다. 일례로 Joseba Castresana et al은 BMEP와 Engine speed inlet을 신경망으로 학습시켜 BSFC, exhaust temperature, CO 발생량과 NOx발생량을 높은 정확도로 예측하였다.[[3]](#footnote-3)

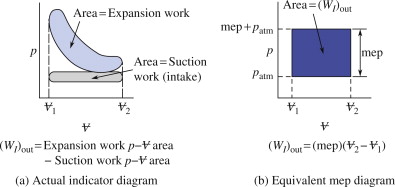
ㄹ은 내연기관의 특성을 사용하여 유도된 값으로 다음의 과정을 거쳐 유도되었다. 내연기관이 한 boundary work은 PV diagram에서 다음 그림에서와 같은 면적을 나타내며, 평균 유효 압력을 통해 다음과 같이 나타낼 수 있다[[4]](#footnote-4)

그림 3 PV diagram(참고문헌 4)

## V2-V1은 실린더 내부에서 피스톤이 왕복운동을 하는 부피로 이 값이 일정치 않으면 엔진에 결함이 있는 것으로 판단할 수 있다. W는 엔진이 한 일로, 이 값에 에너지 효율을 곱하면 엔진이 송출하는 에너지가 된다. 엔진 로드(전력)를 시간에 대해 적분하면 엔진이 송출하는 에너지를 얻을 수 있으므로 이 경우 1초 단위로 Engine Load Percentage를 시간에 대해 적분하였다. 이에 해당하는 변수들을 대입하여 정리하면 다음과 같다. (단 k는 상수)

## 그러나 두 변수의 비율을 그대로 적용하게 되면 분모(IMEP)가 0이 되는 순간이 생겨 모델링하기에 곤란해진다. 따라서 정상 데이터에 대해 IMEP에 대기압인 1을 더해 gauge pressure로 변환한 후, 선형회귀분석을 진행한 뒤 선형회귀식의 기울기를 새로운 변수로 사용하기로 하였다. 또한 (1)식을 시간에 대해 미분하면 다음과 같은 식을 얻을 수 있다.[[5]](#footnote-5)

## C는 상수로 4행정기관에선 2이고, N은 기관의 분당 회전수(RPM), n은 실린더의 수를 의미한다. 위의 수식을 정리하고 해당되는 센서값들을 대입하면 다음과 같은 관계식을 얻을 수 있다. (단 k2는 상수)

## 분모가 0이 되는 경우를 방지하기 위해 Engine Speed에 1rpm을 더하였다. (엔진스피드는 모드3에서 약 720, 모드 2에서 약 25000rpm으로 1을 더한 오차는 1%의 영향도 주지 않을 것으로 예상하였다.) 이와 같은 관계식으로부터, 정상데이터에 대한 세 변수의 비율로 새로운 변수를 정의하였다.

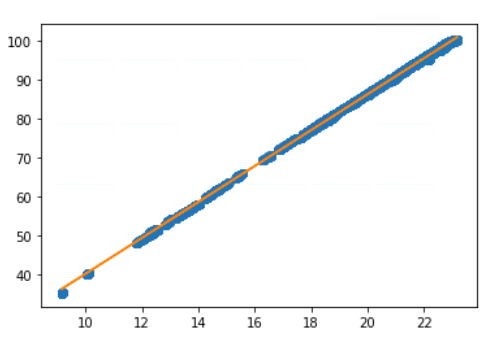
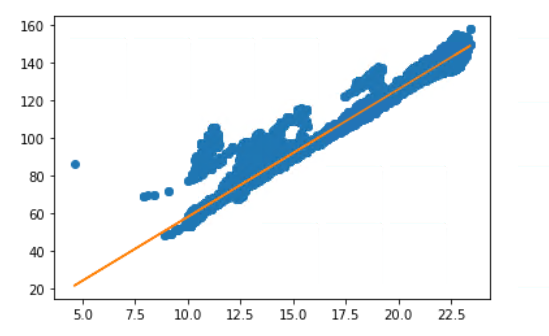
 

그림 4 IMEP-Engine load percentage 그림 5 Pmax average-IMEP

## 또한 최대 압력의 평균인 Pmax와 IMEP간에도 유의미한 상관관계를 발견하여 두 변수의 비율도 새로운 변수로 정의하였다.

## 선행 연구들에서 가장 많이 사용한 방식은 Fourier Transform, Wavelet Transform과 같은 주파수 계열 분석으로, 일정 시간간격동안 입력된 데이터를 주파수영역으로 변환하거나 raw data 안에서 impulse factor, shape factor와 같은 feature를 추출(extract)해 학습시키는 방식을 사용했다.[[6]](#footnote-6)[[7]](#footnote-7) 하지만 이 방식은 매우 높은 sampling rate의 센서를 사용할 때 사용하는 방식으로, 현재의 시스템에는 적용하기 어렵다고 판단하였다. Nyquist Theorem에 의하면 특정 신호를 관측하기 위해서는 적어도 그 신호의 2배 이상 되는 sampling rate로 측정하여야 하는데, 1Hz의 sampling rate로는 수백~수천 rpm에 달하는 엔진의 주기적 신호를 정확하게 측정할 수 없다고 판단하였다.

## 또한 특정 시간동안 들어온 정보에 대해 추출하는 방식은 샘플의 정규성을 보장하기 위해 최소 30개 이상의 연속된 샘플에 대한, 최소 30개 이상의 데이터 묶음(30개 sample \* 30개 묶음)이 필요하다. 하지만 지금과 같은 시스템의 경우 30개 이상의 샘플을 확보하기 위해 30초의 시간이 필요한데, 엔진 입장에서 한 묶음이 다 들어오기까지 걸리는 30초는 엔진의 속도나 모드만 변하더라도 크게 요동할 수 있는 긴 시간이기 때문에 정규성을 가진 집합이라고 보기 어렵다. Train의 일부 mode와 test 데이터의 경우는 심지어 연속된 데이터가 많지 않아 충분한 묶음 확보조차 어려웠다. 이러한 한계로 인해 시간 구간에 대해서는 시계열 분석은 포기하였고, 개별 시점에 대한 측정값만을 학습하였다.

## 2.3 알고리즘의 적용

## 전체 알고리즘을 다음과 같이 구성하였다.

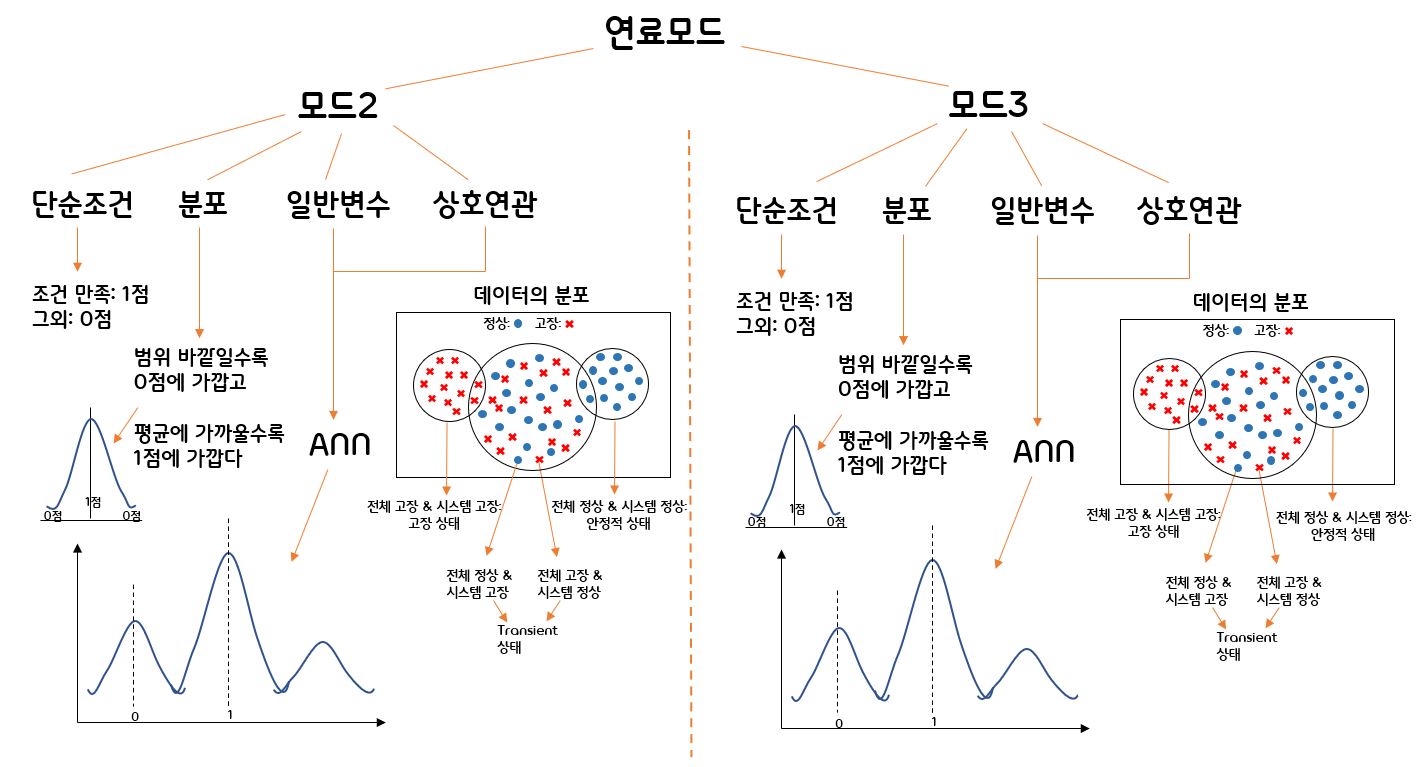


그림 6 알고리즘 개형도

## 2.3.1 단순 조건

## 2.2.1의 ㄱ에 해당하는 데이터의 경우 특정 엔진모드에서는 특정 값만을 가져야한다. 따라서 단순 조건식을 통해 특정 값에 해당되면 정상(1) 그렇지 않으면 고장(0)으로 분류하였다.

## 2.3.2 분포/통계적 기법 사용

2.2.1의 ㄴ에 해당하는 데이터의 경우 먼저, 해당 변수의 데이터가 정규분포를 따르지 않을 수 있으므로 데이터를 표준화하는 작업을 거친다. 표준화에는 우리가 흔히 수능의 표준점수 산출에 사용되는 Z점수를 이용하며, Z점수 = (데이터 변량 - 데이터 평균) / 데이터 표준편차 이다. 이렇게 하면 각 변수마다 데이터의 표준편차와 평균이 달라도 하나의 수치화된 점수로 표현이 가능하다. 그러면 데이터 변량이 평균과 같을 때 점수가 0이 된다. 이 때, 평균은 추산된 평형온도(압력)이라고 한다. 따라서 우리는 평형온도(압력), 즉 가장 정상이라고 볼 수 있는 값을 1이라고 정하고, 고장(비정상)일 때의 값을 0이라고 한다. Z점수의 범위는 0~1이 아니기 때문에 우리는 점수를 변환해야한다. 먼저, Z점수가 양수일 때, 해당 데이터의 최대값일 때의 Z점수를 분모로 하여 Z' = Z / max(Z) 라고 하여 최대값을 1로 만든 후, Z점수가 음수일 때, 해당 데이터의 최소값일 때의 Z점수를 분모로 하여 Z' = Z / min(Z)라고 하여 최대값을 1로 만든다. Z점수가 양수, 음수일 때 분포가 달라 최대값과 최소값의 절대값이 다를 수 있기에 양수일 때의 경우, 음수일 때의 경우로 나눠서 한다. 이 때, 고장에 가까울수록 점수가 1이 되는데 우린 고장일 때 값을 0으로 설정했으므로 한 번 더 점수를 변환한다. 변환된 점수 Z'에 대해 Z'' = 1 - Z'라고 하여 최종 점수를 설정한다.[[8]](#footnote-8)

## 2.3.3 일반변수/상호연관: T/C와 Combustion system의 각각의 고장진단: ANN 알고리즘의 사용

그림 7 문제상황: 예상되는 데이터의 두 가지 분포

## 과제의 목적 중 하나는 고장이 일어난 원인이 되는 시스템을 찾는 것이기에 데이터는 다양한 고장의 경우를 포함하고 있을 것이라 예측할 수 있으나, 주어진 데이터는 state를 고장과 정상으로 나누고 있을 뿐, 고장의 원인과 고장이 일어난 시스템을 명시하고 있지 않다. 이러한 상황에서 전체 데이터로 머신러닝 혹은 딥러닝 모델을 적용하게 된다면 고장과 정상을 구분할 수는 있으나 고장의 원인이 되는 시스템을 찾을 수 없다. 고장데이터라 하여도 한 가지의 시스템만 고장이 나고 나머지는 정상인 것이 대부분이기 때문에, 전체 고장/정상으로 라벨링 된 데이터를 시스템별로 나누어 모델을 학습시키면 한 시스템 안에서 정상과 고장 데이터에 유의미한 차이가 없어 분류율이 떨어지거나 과적합 문제가 발생할 가능성이 매우 커진다. Clustering을 통해 시스템 단위로 재분류를 하는 방식도 고안해 보았지만, 고장데이터가 다양하게 분포할 수 있어 clustering 된 결과를 완전히 신뢰할 수 없다는 점, 고장이 시작되는 위치를 찾아 점수화하기 어렵다는 점 때문에 포기하게 되었다. 이 보고서에서는 가장 보편적인 알고리즘 중 하나인 ANN기법을 사용하여 세 가지 문제를 해결하려 하였다

## Python3 Tensorflow2.0의 ANN 알고리즘을 정상1 고장 0으로 하는 회귀방식을 사용했다. 학습에 사용된 데이터의 input은 EDA에서 선택된 ‘ㄷ’, ’ㄹ’ 데이터의 feature이며, output은 전체 시스템의 정상(1)과 고장(0)이었다. 학습과 검증, 테스트에 쓰이는 데이터는 train : validation : test를 6:2:2 로 분할하여 과적합을 방지하였다.

## ANN을 사용한 이유는 다음과 같다. 우선 ‘전체 시스템은 고장이지만, 특정 system에선 정상인’ 고장데이터의 영향을 최소화할 수 있다. SVM(Support Vector Machine), QDA(Quadratic Discriminant Analysis)와 같이 두 클래스 사이에 경계선을 긋는 기법의 경우 ‘해당 시스템은 정상’인 고장데이터와 정상데이터까지도 분류를 시도하게 된다. 따라서 위와 같이 경계선을 그어 이산적으로 분류를 하는 방법은 해당 시스템의 고장을 진단하는 데에는 적절치 않다. 그에 비해 ANN의 회귀방식을 사용한다면 0~1사이의 값을 output으로 반환하게 되고 1에 가까울수록 정상, 0에 가까울수록 고장으로 판단할 수 있다. 일반적인 분류와는 달리 특정 시스템 내에서는 정상과 차이가 없는 고장데이터도 많은 비율을 차지하기에 ANN은 정상 값에 대해서는 대부분 1에 가까울 만큼 높은 수치가 아닌 0.6~ 0.5정도의 낮은 값을 내게 된다. (그림8 참고) 시스템 별로 최종 건강지수 도출할 때에는 이 점을 보완하여 앙상블 과정에서 0~1로 나오도록 스케일을 조정하였다.(그림9 참고)

그림 9 ANN 회귀 분포로부터 점수화

그림 8 ANN 회귀를 사용한 군집 분류

## ANN은 다음과 같이 구성되었다. 우선 T/C와 Combustion의 두 시스템에 대해 각각 ANN을적용하였다. 데이터는 정상과 고장을 1:1로 샘플링하여 사용하였다. Epoch의 수는 100으로 지정하였다. Hidden layer의 수가 4인 sequential model로 정하였고, 노드의 수는 Kolmogorov rule에 의해 정하였다. Kolmogorov rule에 따라 n개 차원 입력데이터에 대해 2n+1개 이상의 노드를 사용해야 하므로[[9]](#footnote-9), 각 시스템별로 입력데이터 차원에 맞도록 노드의 수를 조정하였다. Mode 2와 mode 3에서 4개 차원의 input을 가진 T/C는 9개, mode 2에서 12개, mode3에서 9개 차원을 갖는 Combustion은 각각 25개와 19개 노드를 갖도록 조정했다. 또한 이 알고리즘의 핵심은 과적합을 방지하며 (보편적인 경우라면 매우 낮은) 적절한 정확도에서 학습을 멈추도록 하여야 라벨이 고장이지만 특정 시스템에서는 정상으로 동작하는 데이터를 오판하지 않는다. 따라서 validation과 10%이상 차이가 날 시 과적합방지를 위해 조기에 학습을 종료하도록 지정하였다.

## 2.3.4 앙상블을 통한 개별 시스템 알고리즘의 성능 보완

## 시스템의 건강지수를 나타내기 위해 앙상블 기법을 사용하였다. 데이터는 전체 시스템에 대한 고장유무만이 라벨링 되어있기에 특정 시스템에서의 고장여부를 확인할 수 없기 때문에 다수의 알고리즘을 앙상블 하는 것이 과적합을 예방할 수 있는 방법이라 판단하였다. 일반적인 방법으로 앙상블 기법의 후보로 가장 간단한 형태의 다수결에 해당하는 hard voting과 0~1사이의 비율로 반환된 값을 평균내는 soft voting[[10]](#footnote-10)이 있으며, 경우에 따라 로지스틱 회귀[[11]](#footnote-11), ANN[[12]](#footnote-12)을 사용하는 경우도 있었다. 일반적으로 머신러닝 기법들을 사용한 기법들이 정확도가 높을 것으로 예상할 수 있으나, 지금처럼 특수한 경우에 머신러닝 방식을 사용하여 앙상블을 하게 된다면 ‘특정 시스템에서는 정상범주에 있는 고장데이터’를 특정시스템에서도 모두 고장으로 처리해버릴 가능성이 높아진다. 이런 상황을 방지하기 위해 soft-voting 기법을 참고하여, 특정 시스템에 대해서만 학습한 다수의 ANN의 output들을 평균 낸 후 다시 0~1로 스케일링하여 특정 시스템의 최종 output으로 결정하였다.

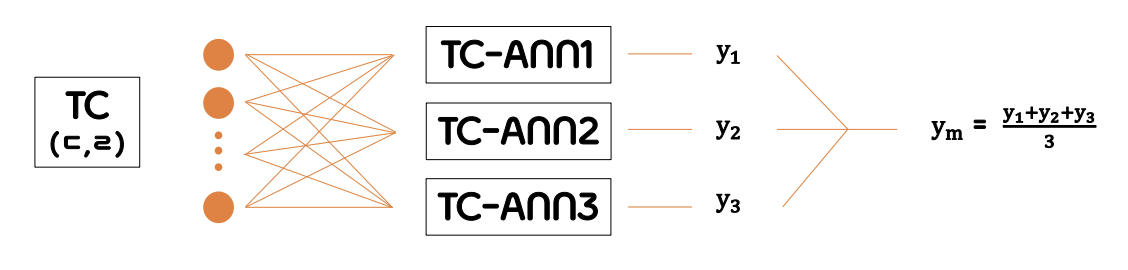


그림 10 앙상블 과정(TC 시스템 예시)

## 2.3.5 최종모델: 종합 건강지수의 도출

각 시스템에 별로 100\*(개별 요소들의 output 평균)을 시스템의 건강지수로 정의하였다. 엔진의 최종 건강지수를 나타내기 위해 전체 4개 시스템의 건강지수에 각각 변수의 개수만큼 가중치를 두어 평균내는 방식을 취하였다.

.

# 3. Result

# 3.1 각 항목별 고장 진단

# 3.1.1 ‘ㄱ’과 ‘ㄴ’ 데이터

표 2 train mode2

|  |  |  |
| --- | --- | --- |
|  | **밀도그래프**  **(정상 – 파랑 / 비정상 – 빨강)** | **표준정규그래프**  **(좌 - 치환 전 / 우 - 치환 후)** |
| ㄱ  LT CW press, Air Cooler Inlet (PT71) |  | 없음 |
| ㄱ  Base duration gas admission |  | 없음 |
| ㄱ  Waste Gate Valve position command (CV26) |  | 없음 |
| ㄴ  Main bearing temp average |  |  |
| ㄴ  LT CW temp, Air Cooler Inlet (TE71) |  |  |
| ㄴ  HT CW temp, Engine Inlet (TE75) |  |  |
| ㄴ  Charge Air press, Engine Inlet (PT21X) |  |  |
| ㄴ  Charge Air temp, Engine Inlet (TE21X) |  |  |
| ㄴ  Gas press, Engine Inlet (PT87) |  |  |
| ㄴ  Gas temp (TE81) |  |  |
| ㄴ  HT CW temp, Engine Outlet (TE76) |  |  |
| ㄴ  LT CW temp, Air Cooler Outlet (TE72) |  |  |
| ㄴ  DVT control air press (PT43) |  |  |

3.1.2 ‘ㄷ’, ’ㄹ’ 데이터

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mode2-TC-ANN1 | Mode2-TC-ANN2 | Mode2-TC-ANN3 |
| 히스토  그램 |  |  |  |
| 산점도 |  |  |  |

표 3 train mode2 - TC

Table train-mode2-TC

표 4 train mode2 - Combustion

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mode2-Combustion-ANN1 | Mode2-Combustion-ANN2 | Mode2-Combustion-ANN3 |
| 히스토  그램 |  |  |  |
| 산점도 |  |  |  |

Table train-mode2-Combustion

개별 그래프의 분포는 다소 차이가 있으나 3가지 데이터의 정규분포가 모두 공통적으로 나타나는 양상을 보인다.

표 5 train data 전체

|  |  |  |
| --- | --- | --- |
| Mode2 | TC | Combustion |
| 히스토그램 |  |  |
| 산점도 |  |  |

전체의 경우도 개별 그래프에서와 같이 세 가지의 정규분포가 나타나는 양상을 보인다. 이때 히스토그램 중앙부 가장 높은 도수를 가진 히스토그램의 계급값(0.5)을 만점(1, 정상), 히스토그램의 좌측 군집에서 가장 높은 도수를 가진 히스토그램의 계급값(0.1)을 0점(0, 고장)으로 설정한 뒤 그 범위에 속하지 않은 값들은 좌측이면 0, 우측이면 1로 설정하였다. 그리고 0과 1사이를 linear scaling 하여 0과 1사이 값을 갖도록 하고 100을 곱해 최종 건강지수를 산출하였다.

# 3.2 전체 엔진에 대한 고장 진단

# [truth table]

# 4. Discussion

4.1 의의

우선 수많은 센서 데이터와 내연기관의 운동에 대한 지식을 바탕으로 다양한 고장지표를 제안하였다. 이전까지의 많은 고장진단 알고리즘이 무겁고 복잡한 알고리즘을 사용하여 고장여부를 판단하였다. 하지만 본 알고리즘의 경우 탐색적 데이터 분석을 통해 특정 온도, 압력조건만 유지된다면 고장의 원인이 되지 않는 간단한 지표로부터, mode에 따라 이산적인 값들을 갖는 데이터, 상호연관성 때문에 직접적인 분석이 어려웠던 데이터까지 신경망 분석을 통해 결과를 냄으로써 복잡한 내연기관 시스템을 다소 단순화할 수 있었다. 또한 열역학적 지식을 기반으로 하여 의미 있는 변수들을 유도하였다.

두 번째로, 통계적인 지표를 활용해 점수를 부여하였다. 베어링의 온도, 냉각수의 온도와 같이 특점 범위 내에 분포하면 정상으로 간주되는 지표는 수능점수를 내는 원리인 정규화를 통해 계산하였다. 이를 통해 보다 수학적으로 유의미한 점수 지표를 제안하려 노력했다.

세 번째로, 비지도적 학습법이 혼재된 문제를 신경망의 회귀방식을 사용해 단순화하고, 분포를 분석했다는 점이다. 이는 신경망 모델이 비슷한 값은 비슷한 결과를 내고, 정규분포를 따른다는 가정하에, 부분 시스템의 고장여부가 라벨링되지 않은 데이터의 고장여부를 진단하기 위한 방식이었다. 이 방식이 유효한지 증명하기까지는 현재 우리의 수학적, 공학적 지식으로는 무리가 있겠지만, 적절한 가정이 뒷받침된다면 주효한 분류 및 회귀 방식이 될 것이라 생각한다.

4.2 한계점

경험부족으로 인해 예상시간을 정확하게 파악하지 못했다. Train data의 Mode2는 다른 데이터와는 다르게 양이 매우 적었다. 따라서 다양한 parameter를 사용해 볼 수 있었고, 시간 또한 많이 소모하지 않았다. 하지만 모드3과 테스트 데이터는 train mode2의 수백배에 달하는 양이 있었고, 이를 전처리하고 신경망에 학습시키는 데에는 매우 많은 시간을 소모하였다. 심지어는 100번의 에포크를 반복하기 위해 train data의 mode3을 학습시켰으나 validation accuracy에 근접하지 못했고, 1000번의 에포크를 실행하여도 validation accuracy에 근접하지 못하고 10시간 가까운 시간을 소모하였다. 주어진 시간 내에 끝내야 함을 감안했다면 과정의 엄밀함을 포기하더라도 보다 가볍고 빠른 알고리즘을 사용하는 것이 타당했을 것이다.

주어진 데이터의 낮은 sampling rate(1Hz)로 인해 이상연소에 대한 정확한 판단이 힘들다. 이상연소는 폭발단계보다 먼저 폭발이 일어나거나 배기단계에서 폭발이 늦게 일어나는 현상인데 이를 알기 위해서는 데이터가 4단계 과정 전부 보여주어야 되고 최소한 4단계 과정에 대한 데이터가 필요하다. 이 HiMSEN엔진은 엔진속도가 수백rpm에 달해 1분에 약 500번 이상 반복된다는 것이다. 앞서 탐색적 데이터 분석에서 언급하였듯, 주어진 데이터가 1Hz이므로 1분에 60개의 데이터가 존재하므로 Nyquist Theorem에 의해 이것은 데이터가 4단계 과정을 정확히 반영한다고 보기 어렵다.

해당 데이터는 state, 즉 전체 시스템이 고장인지 정상인지만 알 수 있도록 구성 되어있다. 과정에서 나름의 논리적 접근을 통해 시스템의 고장 여부를 판단하고 점수화를 하려하였지만, 각 시스템의 정상/고장여부는 알려지지 않아 답을 확신할 수는 없다. 즉, 주어진 데이터로 학습을 시키면 정확히 전체 시스템이 고장인지 정상인지에 관해서만 판단할 수 있을 뿐, 어느 시스템이 문제가 발생하였는지는 확실하게 판단할 수 없을 것이다. 설령 이 데이터 중 abnormal한 데이터가 모든 시스템에 문제가 있었던 데이터라고 하였더라도 문제는 있을 것이다. 왜냐하면 만약에 Combustion에 문제가 있으면 이 문제에 의해서 다른 시스템도 연달아서 생길 가능성이 있기 때문이다. 또한 시스템이 완전 독립적이지 않다는 부분도 분석을 매우 어렵게 만들었다. 일부 센서 데이터 여러 시스템에 걸쳐 있는 데이터가 존재하는데 이 센서에서 이상이 발견되면 어떻게 처리하는지 알 수 없다. 예를 들어 Exh. Gas temp, TC Inlet A Bank (TE26-A)라는 센서 데이터는 T/C, Combusiton, CoolingWater에 걸쳐 있는 데이터이다. 만약 이 센서 데이터가 문제가 발생한다면 세 시스템 중 어느 것이 문제가 있었는지, 이 데이터가 무슨 시스템에 영향을 많이 받는지 알 수 없어 정확하게 판단할 수 없다. 데이터 분석시에 이러한 데이터는 센서데이터가 속한 모든 시스템에 포함하였으나, 완벽하게 이를 반영하였다고는 보기 어렵다.

Feature Selection method를 사용하지 못하는 것이 아쉬웠다. Feature selection은 주어진 구간에서 데이터의 통계적인 특징들을 뽑아서 t-test으로 특징들을 선별한 후 특징들로 머신러닝 모델을 학습시키는 방법이다. 데이터를 효과적으로 압축하여 사용하고, 특징들로 하는 방법이기에 학습시간이 짧다는 것이 장점이어서 사용하고 싶었지만 앞서 method의 초반부에서 언급한 복합적 상황에 의하여 사용할 수 없었다. 이 방법은 데이터가 각 시스템이 정상인지 고장인지 정확히 알아야 사용할 수 있는 방법인데 앞에 언급한 것처럼 전체 시스템이 고장인지 정상인지만 알 수 있어서 개별 시스템에는 사용할 수 없다. 또한 실제 t-test를 한 결과 p-value가 0인경우가 많았다. 이는 과적합(전체 상태는 고장이지만 특정시스템에서는 정상인 데이터를 모두 고장으로 처리해버리는 상황)이라는 문제를 야기할 수 있어 이 방법으로 하는 것이 부적절하다고 판단하였다.

4.3 보다 나은 데이터 수집을 위한 제안

4.3.1 가스 센서

많은 질소 산화물과 일산화탄소는 불완전 연소에 의해 생겨나게 된다. 불완전 연소는 연료가 정상적으로 낼 수 있는 에너지 보다 적은 에너지를 발산하도록 하여, 엔진 내부 압력과 피스톤의 상하운동에 영향을 주고, 엔진의 성능을 저하시키는 요인이 된다. method에서 언급했듯 ANN 알고리즘을 이용, Joseba Castresana et al은 BMEP와 Engine speed inlet을 신경망으로 학습시켜 BSFC, exhaust temperature, CO 발생량과 NOx발생량을 높은 정확도로 예측하였다.[[13]](#footnote-13) 이러한 점을 활용하여 engine speed inlet과 BMEP로부터 추정한 CO, NOx의 양과 측정되는 양이 다르다면 엔진의 고장을 예측해볼 수 있다.

4.3.2 진동, 관성 센서

앞서 많은 부분에서 언급했듯 대부분의 센서가 온도와 압력을 측정하는 센서이기 때문에 낮은 sampling rate을 갖게 되는데, 이로 인해 선행연구에서 쓰이는 Fourier Transform 및 Wavelet Transform과 같은 주파수 영역 분석을 전혀 시도할 수 없었다. 많은 선행연구에서 엔진의 고장을 진단하기 위해 사용한 대표적인 방법인 만큼, 진동센서를 사용한다면 보다 다양한 고장원인을 분석할 수 있을 것으로 생각된다. 특히나 낮은 sampling rate로 높은 RPM의 엔진속도에서 나오는 신호들을 측정하기 어려운 한계가 있었는데, 이러한 점을 극복하여 엔진의 이상연소를 감지할 수 있는 매력적인 방안으로 생각된다.

또한 크랭크 축의 시간, 엔진속도에 따른 기울기변화도 보편적으로 엔진성능을 좌우하는 요소로 알려져 있다. 적절한 sampling rate의 관성센서가 존재하는지는 알기 어렵지만, 적절히 crank angle을 추적한다면 보다 정밀한 이상연소추정이 될 것이라 생각한다.

4.3.3 고장 원인의 라벨링

이 역시도 많은 부분에서 언급했듯, 현재 이 보고서의 알고리즘으로는 통계적 추정만이 가능할 뿐, 실제 고장 여부는 알 수 없다. 이진웅 외 4인은 논문에서 spot-welding이라는 아주 단순한 반복작업을 하는 산업 기계의 고장을 진단하는 데에도 그 기계의 여러 고장 변수 중 하나인 용접 각도 misalignment를 알아내는 인공지능 알고리즘에 모든 지면을 할애했다. 또한 논문에서 고장의 형태인 misalignment 각도도 정상을 포함한 4가지를 포함하여 심각한 수준의 고장에서부터 경미한 고장을 모두 포함하였다. [[14]](#footnote-14) 아무리 단순한 기계의, 한 가지의 고장이라도 원인을 분석하기 위해서 그만한 어려움이 있음을 시사한다. 현재 대회에서는 대형 선박엔진이라는 거대하고 복잡한 시스템을 다룬다. 이런 상황에서 개별 시스템의 고장 여부에 대한 라벨링 없이 개별 시스템을 분석하는 것은 매우 어려운 일로 보이며, 실제로 정답을 알 수 없다는 측면에서 매우 리스크가 큰 방식이라고 생각한다. 물론 여러 변수가 상호작용한다는 점에서 spot-welding과 차이점이 있지만, 적어도 몇 가지 뚜렷한 고장 모드를 나누어 함께 라벨링한다면 보다 효과적인 방식의 분석이 가능할 것으로 생각한다.

4.3.4 데이터의 균형과 system 분류

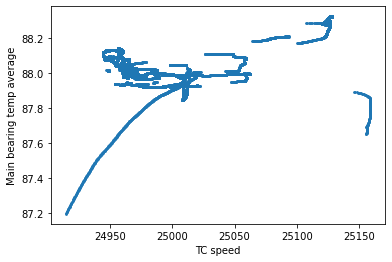
일반적으로 베어링의 온도는 베어링이 지지하고 있는 축의 회전속도에 비례한다. 따라서 약간의 차이가 있을 지라도 베어링의 온도는 엔진속도나 TC속도 혹은 엔진 로드에 비례하는 결과가 나오는 것이 일반적일 것이다. 하지만 현재의 상황은 전혀 상관관계가 없는 것처럼 분포한다. 여러가지 상황들을 추측해 볼 수 있으나, 애초에 엔진 스피드가 특정 구간(평형온도 인근)에 절대 다수가 분포해 엔진속도에 따라 상승하는 베어링의 온도를 적절히 반영하지 못하는 것으로 보인다. 또한 적절한 관계가 나오더라도 구간이 너무 좁아(TC speed가 1% 정도, 온도가 1도 정도 변한다) 이를 상관관계라고 보기 어렵다. 이런 상황에서 특정 상황에만 맞추어 상관관계를 정의하면 베어링온도나 속도가 다를 때 이를 설명하지 못하는 현상이 발생한다. 선박의 엔진이 대부분 일정한 속도로 운행하게 되겠지만, 낮은 속도에서의 적절한 상관관계가 유도되지 않는다면 엔진속도가 변하는 상황에서 고장을 감지하지 못할 가능성이 생긴다.

그림 11 엔진스피드와 베어링 온도 그림 12 TC스피드와 베어링 온도

4.4 그 밖에 다른 방법들

현재 분석한 기법보다 다양한 방법들이 후보로 존재하였으나, 한달이라는 제한된 시간동안 모든 방법을 시도하지는 못하여 다소 아쉬움이 남는다. 다음은 후보로 제시되었던 여러 방식들이다.

1. GMM 클러스터링: ANN을 통해 얻은 output이 정규분포를 따르는 것으로 보이는 세 개의 집합으로 이루어져 있으므로, GMM을 통해 세 개의 그룹으로 나눈다면 세 가지 데이터를 효과적으로 분류할 수 있을 것이다. 하지만 이 경우 ANN의 output이 정규분포를 따르는 3개의 군집으로 이루어진 것을 알아내기까지의 시간이 매우 오래 걸려 현실적으로 이를 구현할 만한 시간적 여유가 없었다. (직접 분석하기 전까지는 안정상태를 제외한 2개 군집만을 예상했다.) 만약 가능하다면 첫번째 군집의 평균값과 두 번째 군집의 평균값을 각각 0과 1로 놓고 점수화하는 방법을 시도할 수 있었을 것이다.
2. Settling time과 감쇠계수: test데이터를 분석하던 중, 연속된 데이트타임에서 연료모드, 혹은 엔진속도가 변할 때 마치 진동 시스템의 overdamping과 같은 현상으로 주변의 변수들이 변하는 것을 발견하였다. 연료모드, 혹은 엔진속도가 step function과 같이 동작한다면 그와 연관된 변수들은 변화된 연료모드, 엔진속도에 영향을 받아 그 상태에 적합한 상태로 변해간다. 이 경우, 연료모드와 엔진속도는 급격히 변하고 다른 변수들은 상대적으로 완만히 변하기 때문에 기존 알고리즘에선 이를 고장으로 인식할 수 있는 여지도 생긴다. 이를 보완하기 위해 진동의 settling time(임팩트나 변화 이후 안정되기까지의 시간)과 감쇠계수(안정되는 경향)개념을 도입하면 그러한 fuel mode(engine speed)-transient 상태를 보다 효과적으로 분류할 수 있을 것이다. 예를 들어 fuel mode가 변했을 때 exh-gas temp가 정상상태의 settling time이 지나서도 안정화되거나 목표값으로 오지 않는다면 이를 고장으로 진단할 수 있고, 목표값으로 이동하는 기울기를 감쇠계수를 통해 구해 적절한 기울기가 나오지 않는다면 이를 고장으로 진단할 수도 있을 것이다. 이 방식은 transient phase데이터를 얻고 처리하는 과정이 지나치게 오래 걸릴 것으로 예상되어 실행에 옮기지 못했다.
3. Massive-fusion: 각 시스템에 대해 매우 많은 이진분류모델들(Logistic Regression, k-NN, Decision Tree, Random Forest, Support Vector Machine, Quadratic Discriminant Analysis 등)을 다양한 parameter로 학습시켜 그 평균값을 건강지수로 평가하는 방법이다. 라벨링은 전체 시스템의 고장여부만을 확인하기 때문에, 특정 시스템에서 고장여부를 확인하고 점수화하기는 어렵다. ‘전체 시스템은 고장이나 특정 시스템은 정상인’ 데이터는 머신러닝 알고리즘에서 정상으로 판별할 수 있는 가능성도 높다는 점에서 착안한 것으로, 많은 알고리즘이 정상으로 판별한다면 건강지수를 높게 주는 방식이다. 원래 처음으로 떠올려 시도하려 했던 방법이나, 이 경우 데이터 해석에 있어서 매우 난해하고 모델이 거대해지며, 모델 설명에 있어서 수학, 과학적 엄밀성이 떨어지는 문제가 생길 수 있어 배제하였다.

# 5. 맺음말

김태균: 팀원들과 함께 이 대회에 참가할 수 있어서 영광이었고, 평소에 해보고 싶은 주제 중 하나였던 고장 진단을 실제 데이터로 해볼 수 있는 기회를 얻어서 매우 의미 있는 시간이었습니다. 대회를 진행하면서 스스로의 부족함도 느끼고, 팀원들과 많은 지식을 공유하고, 또 많은 학술 자료들을 탐독하며 대회 기간동안 스스로 성장해감을 느꼈습니다. 팀장으로서, 데이터 관계와 신경망 분석을 맡았던 팀원으로서 실수도 많았지만 이해해주고 함께 노력해준 또한 팀에게 고맙고 미안합니다. 팀원들간 소통하고 함께 작업하며 많은 추억을 갖고 가게 되었습니다. 대회 전반부에 전반적인 내연기관에 대한 공부를 하느라 전체적으로 인공지능 학습과 다양한 알고리즘을 사용에 시간을 많이 할애해보지 못해, 원했던 것만큼 효과적이고 엄밀한 알고리즘을 만들어 내지는 못한 것은 다소 아쉽지만, 이 과정을 통해 많은 지식과 깨달음을 얻어 유익한 시간이었습니다. 이런 대회를 열어 기회를 주신 많은 분들께 한 번 더 감사를 표합니다. 감사합니다.

여승협: 경진대회를 주최하여 주신 현대중공업그룹 한국조선해양 관계자 분들께 감사의 말씀을 드립니다. 경진대회를 주최하셨기 때문에 이전에 배웠던 지식들을 토대로 새로운 Challenge를 할 수 있었습니다. 다른 곳에서 얻을 수 없는 데이터를 가지고 하는 분석은 매우 값진 경험이 되었습니다. 또한 이번 대회는 뛰어난 동료와의 협력을 통해 스스로를 발전시킬 수 있었던 하나의 큰 계기가 되었습니다. 부족한 저를 이끌고 끝까지 함께 해준 3Miroa팀에 감사합니다. 마지막으로 대회 중 질문에 성실히 답하여 주신 경진대회 사무국 관계자 분들께도 감사의 말씀을 전합니다.

서재원: 먼저 제가 배운 것들을 적용해보며 도전할 수 있도록 기회를 주신 현대중공업그룹 한국조선해양 관계자 분들께 감사의 말씀드립니다. 이번 대회에서 수업시간에 배운 PHM 관련 이론들을 실제 데이터로 적용할 수 있어서 너무 좋았습니다. 대회를 진행하면서 내연기관에 대하여 다시 공부할 수 있었고 데이터를 분석, 전처리하는 방법에 대하여 다시 복습할 수 있었습니다. 또한 팀원들과 협력하면서 함께 지식을 공유하며 좋은 추억을 만들 수 있었습니다. 제가 생각한 알고리즘과 여러 기법들이 여러 사정 때문에 잘 사용되지 못하여 아쉬웠지만 그 과정 속에서 많은 것들을 배울 수 있었습니다. 이 대회 속에서 정말 많은 것들을 배웠고 스스로 많이 발전되었습니다. 마지막으로 부족한 저를 많이 도와주신 3Miroa팀원들과 부족한 질문에 성실하게 답해주신 대회 관계자 분들 감사합니다. 마지막으로 다시한번 이런 좋은 대회를 주최해주시고 저희에게 기회를 주신 현대중공업그룹 한국조선해양 관계자 분들께 감사의 말씀드립니다.

# References

1. Dimitrios T Hountalas, Prediction of marine diesel engine performance under fault conditions, Applied Thermal Engineering, Volume 20, Issue 18, 2000, Pages 1753-1783, ISSN 1359-4311,https://doi.org/10.1016/S1359-4311(00)00006-5.
2. 정남용, 기계설계, 학진북스, 2020, p37
3. Joseba Castresana, Gorka Gabiña, Leopoldo Martin, Zigor Uriondo, Comparative performance and emissions assessments of a single-cylinder diesel engine using artificial neural network and thermodynamic simulation, Applied Thermal Engineering, Volume 185, 2021, 116343, ISSN 1359-4311, <https://doi.org/10.1016/j.applthermaleng.2020.116343>.
4. Robert T. Balmer, Modern Engineering Thermodynamics, 2011
5. Robert T. Balmer, Modern Engineering Thermodynamics, 2011
6. Lee, Yang Min, Lee, Kwang-Young, Bae, Seung-Hyun, Jang, Hwi, and 이재기, “다채널 진동 센서를 이용한 선박 엔진의 진동 감지 및 고장 분류 시스템,” *정보처리학회논문지A*, vol. 17A, no. 2, pp. 81–92, Apr. 2010.
7. Xu Wang, Yanping Cai, Aihua Li, Wei Zhang, Yingjuan Yue, Anbo Ming, Intelligent fault diagnosis of diesel engine via adaptive VMD-Rihaczek distribution and graph regularized bi-directional NMF, Measurement, Volume 172, 2021, 108823, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2020.108823.
8. Educational and Psychological Measurement and Evaluation, 8th Edition. - Kenneth Hopkins, University of Colorado
9. L.H. Feng, J. Lu, The practical research on flood forecasting based on artificial neural networks, Expert Syst. Appl. 37 (2010) 2974–2977.
10. 김민수1906, context\_KKP, 제주감귤, DB분석가, TNT, 데이콘 경진대회 1등 솔루션, 위키북스, 2021
11. Yifeng Yang, Long Wei, Ying Hu, Yan Wu, Liangyun Hu, Shengdong Nie, Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning, Journal of Neuroscience Methods, Volume 350, 2021, 109019, ISSN 0165-0270, https://doi.org/10.1016/j.jneumeth.2020.109019.
12. Su Hyun Youn, Taeyong Sim, Ahnryul Choi, Jinsung Song, Ki Young Shin, Il Kwon Lee, Hyun Mu Heo, Daeweon Lee, Joung Hwan Mun, Multi-class biological tissue classification based on a multi-classifier: Preliminary study of an automatic output power control for ultrasonic surgical units, Computers in Biology and Medicine, Volume 61, 2015, Pages 92-100, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2015.03.021.
13. Joseba Castresana, Gorka Gabiña, Leopoldo Martin, Zigor Uriondo, Comparative performance and emissions assessments of a single-cylinder diesel engine using artificial neural network and thermodynamic simulation, Applied Thermal Engineering, Volume 185, 2021, 116343, ISSN 1359-4311, https://doi.org/10.1016/j.applthermaleng.2020.116343.J
14. iwoong Lee, Inwoong Noh, Sang Ik Jeong, Yongho Lee, Sang Won Lee, Development of Real-time Diagnosis Framework for Angular Misalignment of Robot Spot-welding System Based on Machine Learning, Procedia Manufacturing, Volume 48, 2020, Pages 1009-1019, ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2020.05.140.

1. Dimitrios T Hountalas, Prediction of marine diesel engine performance under fault conditions, Applied Thermal Engineering, Volume 20, Issue 18, 2000, Pages 1753-1783, ISSN 1359-4311,

   https://doi.org/10.1016/S1359-4311(00)00006-5. [↑](#footnote-ref-1)
2. 정남용, 기계설계, 학진북스, 2020, p37 [↑](#footnote-ref-2)
3. Joseba Castresana, Gorka Gabiña, Leopoldo Martin, Zigor Uriondo, Comparative performance and emissions assessments of a single-cylinder diesel engine using artificial neural network and thermodynamic simulation,

   Applied Thermal Engineering, Volume 185, 2021, 116343, ISSN 1359-4311, https://doi.org/10.1016/j.applthermaleng.2020.116343. [↑](#footnote-ref-3)
4. Robert T. Balmer, Modern Engineering Thermodynamics, 2011 [↑](#footnote-ref-4)
5. Robert T. Balmer, Modern Engineering Thermodynamics, 2011 [↑](#footnote-ref-5)
6. Lee, Yang Min, Lee, Kwang-Young, Bae, Seung-Hyun, Jang, Hwi, and 이재기, “다채널 진동 센서를 이용한 선박 엔진의 진동 감지 및 고장 분류 시스템,” *정보처리학회논문지A*, vol. 17A, no. 2, pp. 81–92, Apr. 2010. [↑](#footnote-ref-6)
7. Xu Wang, Yanping Cai, Aihua Li, Wei Zhang, Yingjuan Yue, Anbo Ming, Intelligent fault diagnosis of diesel engine via adaptive VMD-Rihaczek distribution and graph regularized bi-directional NMF, Measurement,

   Volume 172, 2021, 108823, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2020.108823. [↑](#footnote-ref-7)
8. Educational and Psychological Measurement and Evaluation, 8th Edition. - Kenneth Hopkins, University of Colorado [↑](#footnote-ref-8)
9. L.H. Feng, J. Lu, The practical research on flood forecasting based on artificial neural networks, Expert Syst. Appl. 37 (2010) 2974–2977. [↑](#footnote-ref-9)
10. 김민수1906, context\_KKP, 제주감귤, DB분석가, TNT, 데이콘 경진대회 1등 솔루션, 위키북스, 2021 [↑](#footnote-ref-10)
11. Yifeng Yang, Long Wei, Ying Hu, Yan Wu, Liangyun Hu, Shengdong Nie,

    Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning,

    Journal of Neuroscience Methods, Volume 350, 2021, 109019, ISSN 0165-0270,

    https://doi.org/10.1016/j.jneumeth.2020.109019. [↑](#footnote-ref-11)
12. Su Hyun Youn, Taeyong Sim, Ahnryul Choi, Jinsung Song, Ki Young Shin, Il Kwon Lee, Hyun Mu Heo, Daeweon Lee, Joung Hwan Mun, Multi-class biological tissue classification based on a multi-classifier: Preliminary study of an automatic output power control for ultrasonic surgical units,

    Computers in Biology and Medicine, Volume 61, 2015, Pages 92-100, ISSN 0010-4825,

    https://doi.org/10.1016/j.compbiomed.2015.03.021. [↑](#footnote-ref-12)
13. Joseba Castresana, Gorka Gabiña, Leopoldo Martin, Zigor Uriondo, Comparative performance and emissions assessments of a single-cylinder diesel engine using artificial neural network and thermodynamic simulation,

    Applied Thermal Engineering, Volume 185, 2021, 116343, ISSN 1359-4311, https://doi.org/10.1016/j.applthermaleng.2020.116343. [↑](#footnote-ref-13)
14. Jiwoong Lee, Inwoong Noh, Sang Ik Jeong, Yongho Lee, Sang Won Lee,

    Development of Real-time Diagnosis Framework for Angular Misalignment of Robot Spot-welding System Based on Machine Learning, Procedia Manufacturing, Volume 48, 2020, Pages 1009-1019, ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2020.05.140. [↑](#footnote-ref-14)