

이종 모델간 앙상블을 이용한 수소충전소 다이어프램 압축기 고장 진단에 관한 연구

홍영우¹, 김성은², 신덕식³, 유동영⁴

홍익대학교 소프트웨어융합학과 석사과정¹, 교수⁴

(주)에니트 차세대인공지능연구소²

한국전자기술연구원 스마트전장제어연구팀³

hyw1021@g.hongik.ac.kr¹, kse55@enitt.co.kr², dssshin@keti.re.kr², ydy@hongik.ac.kr⁴

A study on diagnosis of failure of hydrogen refueling station diaphragm compressor using heterogeneous model ensemble

Young-Woo Hong¹, Seong-Eun Kim², Duck-Shick Shin³, Dong-Young Yoo⁴

^{1,4}Dept. of Software and Communications Engineering, Hongik University

²Next Generation Artificial Intelligence Research Institute, Enitt Co., Ltd.

³Smart Electrics Research Center, Korea Electronics Technology Institute

요 약

우리나라의 수소연료전지 차량의 점유율이 매년 증가하고 있으나, 수소충전소 설비의 잦은 중단으로 수소연료전지 차량 운전자들이 제때 차량을 충전하지 못하는 불편이 발생하고 있다. 본 논문에서는 수소충전소 설비 중 Diaphragm을 사용하는 압축기의 이상 패턴을 탐지하는 Ensemble 모델을 통해 수소충전소에서 2023년 1월 1일부터 2023년 6월 28일 동안 수집된 데이터를 분석하였으며, 해당 기간 동안 발생했던 고장에 대해 2일전부터 이상 패턴이 10,000 이상 탐지되는 결과를 얻었다.

1. 서론

SNE리서치의 2023년 2월 Global FCEV Monthly Tracker[1]에 따르면, 우리나라의 수소연료전지 차량의 점유율이 전년 대비 4.4% 성장하며 가장 높은 점유율을 기록하며, 수소연료전지차 운전자 수는 증가하고 있다. 하지만, 수소충전소 설비의 잦은 고장으로 인한 운영 중단과 유지보수 지연으로 수소연료전지 차량 운전자들이 제때 수소를 충전하지 못해 불편을 겪고 있으며, 무리한 충전 작업으로 안전사고 우려 역시 증가하며 이는 수소경제에 대한 불신을 가져올 수 있다고 지적하였다.

본 논문에서는 수소충전소에서 사용되는 설비 중 Diaphragm 압축기에서 수집할 수 있는 전력 PQM (Power Quality Monitoring)와 흡입 압력, 토출 압력, 토출 온도로 구성된 운영 데이터에서 발생하는 고장 징후를 탐지하기 위해 서로 다른 유형의 Machine Learning 모델 간 앙상블을 통해 고장 징후를 조기에 탐지 방안을 연구하였다.

2. 관련 연구

Diaphragm 압축기는 3장의 금속 재질 판으로 구

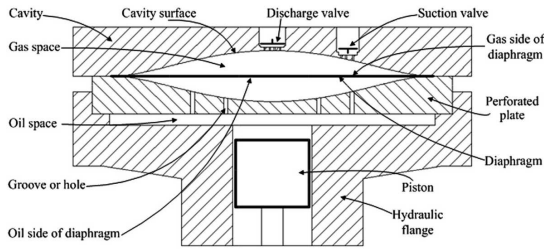
성된 Diaphragm을 통해 유압을 조절하여 온도 높은 압축 수소를 생산하는 방식으로, 유압이 가해지는 Diaphragm에 대해 고장이 자주 발생하고 있다.

2.1 Diaphragm 압축기에 관한 연구

2.1.1 Study on the diaphragm fracture in a diaphragm compressor for a hydrogen refueling station [5]

수소충전소에 사용되는 Diaphragm 압축기의 Diaphragm 피로 수명에 영향을 주는 주요 요인은 방사형 응력인데, 이 연구에서는 Diaphragm의 외측 금속판의 구멍이 방사형 응력에 끼치는 영향을 분석하였으며, 금속판 구멍의 윤곽과 폭, 금속판의 두께가 최대 방사형 응력에 영향을 준다고 밝혔다. 따라서, Diaphragm 금속판의 두께를 두껍게 만들고나, 구멍의 폭을 줄이면 Diaphragm이 받는 피로도를 줄일 수 있다는 사실을 증명하였다.

2.1.2 A non-destructive fault diagnosis method for a diaphragm compressor in the hydrogen refueling station [6]

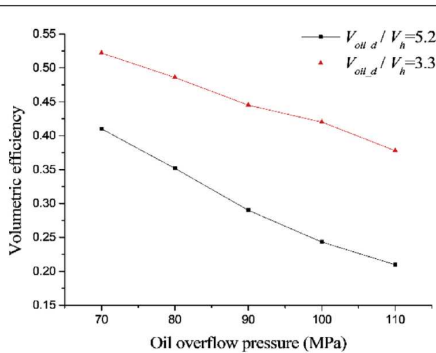


(그림 1) Diaphragm 압축기의 구조

수소충전소 Diaphragm 압축기에서 발생하는 비과외 유형 고장 중 Flywheel과 Oil head를 탐지하기 위해 Acoustic Emission(AE) 신호를 활용하였으며, Short-term Fourier transform (STFT)를 통해 각도별 스펙트로그램 이미지를 생성하고, 이를 분석하여 압축기의 동작과 Oil head 발생 여부를 분석하였다.

2.1.3 Effect of hydraulic oil compressibility on the volumetric efficiency of a diaphragm compressor for hydrogen refueling stations [7]

수소충전소 Diaphragm 압축기에서 발생하는 고장 유형 중 하나인 Oil overflow가 발생하는 조건을 연구하기 위해 Test rig를 구성 후, Oil overflow가 발생하는 흡입 압력, 흡입 온도, 토출 압력 수치를 실험하였다. 이를 통해, 흡입 압력, 흡입 온도, 토출 압력의 증가는 Oil Volumetric 효율성을 감소시켜 Oil overflow를 발생시킬 수 있다는 결론을 내렸다.



(그림 2) Oil Volumetric 효율성과 Oil overflow pressure 간 상관관계

2.2 Ensemble 모델에 대한 연구

2.2.1 Classifier Ensemble Design for Imbalanced Data Classification: A Hybrid Approach [8]

불균형한 데이터는 학습 모델을 단순화시켜 성능을 저하시키며, 분류기는 잘못된 결과를 산출할 수 있는데, 이러한 문제를 해결하기 위해 Uma R. Salunkhe 외 1인은 Multiple Classifier

System(MCS)를 제안한다. 이때 제안했던 방식 중 서로 다른 학습 알고리즘을 가진 분류기 Ensemble 시키는 Stacking 기법을 통해 불균형한 데이터로 발생할 수 있는 Over Fit의 위험을 줄일 수 있다고 밝혔다.

2.2.2 Breast Cancer Classification using Deep Learning-based Ensemble [9]

Uma R. Salunkhe 외 1인[8]의 MCS 방식에 따라, WOBC(Wisconsin original breast cancer)와 WDBC(Wisconsin diagnostic breast cancer)의 유방 진단 데이터를 대상으로 서로 다른 모델을 앙상블 시키는 Stacking 기법을 적용하여 정상과 유방암을 분류하는 Ensemble 모델을 생성하였다. 이 Ensemble 모델을 WOBC, WDBC 데이터에 적용하였을 때, 단일 모델에 비해 Accuracy와 c-statistics 측면에서 더 우수하다.

3. 데이터 분석 및 문제 정의

본 연구에서는 앞선 관련연구[5], [7]에서 Diaphragm 금속판의 피로 수명에 영향을 줄 수 있는 흡입 압력, 흡입 온도, 토출 압력 데이터로 구성된 운영 데이터와 전류(current), 전류 불평형율(current_unbalance), R,S,T상 역률의 평균(pf_average)로 구성된 PQM 데이터를 Diaphragm 압축기를 사용하는 실제 수소충전소의 영업시간 중 수집된 데이터를 활용하였다.



(그림 3) PQM 데이터의 Pearson 상관관계수 Heat Map

PQM 데이터에 대해 Pearson 상관관계수 분석, 데이터 분포 분석, 수기 분석을 통해 Diaphragm 압축기의 첫 기동 또는 이상 기동시 전류 불평형율과 평균 역률이 상승하는 패턴이 확인되었으며, 운영 데이터의 경우, 영업일 기준 Diaphragm 고장 하루 전 관찰되는 패턴을 중심으로 데이터 라벨링을 수행하

였으며, 데이터 일부가 불균형하게 수집되는 경우가 있어, 모델의 Over Fitting을 줄이기 위해[8] Stacking 기법을 적용한 Ensemble 모델을 사용하였다.

4. 모델 적정성

4.1 PQM 데이터 평가 및 Base Estimator 모델별 성능 평가

Base Estimator 모델로 적절한 모델을 선정하기 위해 PQM 데이터를 통해 LR (Logistic Regression) 모델, MLR(Multiple Linear Regression) 모델, DT(Decision Tree) 모델, RF 모델, RF(Random Forest) 모델, KNN(K-Nearest Neighbor) 모델, DNN (Deep Neural Network, Binary Classifier) 모델을 대상으로 성능 평가를 수행하였다.

각 모델별 성능 평가 결과는 다음 < 표 1 > ~ < 표 6 > 과 같다.

< 표 1 > LR 모델

Columns	A	B	C
R, S, T 상 평균 전류 불평형율	O	O	O
전류	O	O	X
역률	O	O	O
R, S, T 상 개별 전류 불평형율	O	X	X
Accuracy	100 %	100 %	100 %

< 표 2 > MLR 모델

Columns	A	B	C
R, S, T 상 평균 전류 불평형율	O	O	O
전류	O	O	X
역률	O	O	O
R, S, T 상 개별 전류 불평형율	O	X	X
Accuracy	98.64 %	97.30 %	91.04 %

< 표 3 > DT 모델

Columns	A	B	C
R, S, T 상 평균 전류 불평형율	O	O	O
전류	O	O	X
역률	O	O	O
R, S, T 상 개별 전류 불평형율	O	X	X
Result			
Accuracy	1 0 0 %	1 0 0 %	100 %

< 표 4 > DNN 모델

Columns	A	B	C
R, S, T 상 평균 전류 불평형율	O	O	O
전류	O	O	X
역률	O	O	O
R, S, T 상 개별 전류 불평형율	O	X	X
Accuracy	100 %	86.14 %	81.04 %

< 표 5 > RF 모델

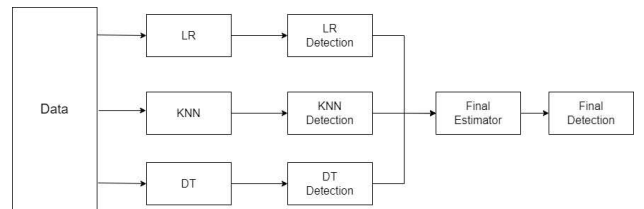
Columns	A	B	C
R, S, T 상 평균 전류 불평형율	O	O	O
전류	O	O	X
역률	O	O	O
R, S, T 상 개별 전류 불평형율	O	X	X
Accuracy	100 %	100 %	100 %

< 표 6 > KNN 모델

Columns	A	B	C
R, S, T 상 평균 전류 불평형율	O	O	O
전류	O	O	X
역률	O	O	O
R, S, T 상 개별 전류 불평형율	O	X	X
Accuracy	100 %	100 %	100 %

4.2 Ensemble 모델 설계

PQM 데이터를 통한 모델별 성능 평가에 따라, LR 모델, KNN 모델, DT 모델을 대상으로 각 Base Estimator로 사용하였고, Base Estimator의 각 결과를 Decision Tree 기반 Final Estimator 모델의 입력 값으로 사용하는 Stacking Classifier 기법을 다음 그림과 같이 적용하였다.



(그림 4) Ensemble 모델 구조도

본 실험에서는 동일한 구조의 모델을 PQM 데이터만 학습한 모델, 운영 데이터만 학습한 모델로 구분하여 독립적으로 학습과 테스트를 수행하였다. 1월 5일 고장 사례를 중심으로 1월 1일에서 1월 6일까지의 데이터에 라벨링 작업을 수행하여 모델 학습

데이터로 활용하였다.

5. 실험 결과

데이터 측정 기간 중 실제 Diaphragm 고장이 있었던 날짜별 데이터 모델들의 탐지 결과는 다음과 같다. 고장 당일은 영업 중지로 데이터 측정이 없었던 경우는 결과에서 배제하였다.

< 표 7 > 2023년 1월 5일, 압축기 Diaphragm 고장

Date	PQM 데이터 Ensemble 모델	운영 데이터 Ensemble 모델
2023-01-01	56,980	0
2023-01-02	24,944	0
2023-01-03	25,897	12,748
2023-01-04	31,349	10,706
2023-01-05	N/A	
2023-01-06	31,188	0
2023-01-07	36,884	0

< 표 8 > 2023년 2월 16일, 압축기 Diaphragm 고장

Date	PQM 데이터 Ensemble 모델	운영 데이터 Ensemble 모델
2023-02-13	38,748	0
2023-02-14	42,431	11,343
2023-02-15	49,380	12,854
2023-02-16	N/A	
2023-02-17	49,423	0
2023-02-18	55,180	0
2023-03-19	46,521	2

< 표 9 > 2023년 1월 23일, 압축기 Diaphragm 고장

Date	PQM 데이터 Ensemble 모델	운영 데이터 Ensemble 모델
2023-01-20	48,828	0
2023-01-21	46,421	10,380
2023-01-22	83,907	43,015
2023-01-23	N/A	
2023-01-24	72,836	0
2023-01-25	44,726	0
2023-01-26	46,526	0

< 표 10 > 2023년 3월 22일, 정상 운영

Date	PQM 데이터 Ensemble 모델	운영 데이터 Ensemble 모델
2023-03-18	52,442	0
2023-03-19	50,451	0
2023-03-20	44,586	0
2023-03-21	38,240	0
2023-03-22	40,770	0
2023-03-23	40,469	0
2023-03-24	49,380	0

실험 결과, 운영 데이터 Ensemble 모델에서 학습 데이터로 활용하지 않았던 1월 23일과 2월 16일 고장 사례에 대해 2일 전부터 10,000건 이상의 이상 패턴이 탐지되었으며, 정상 운영 기간 중 10건 이하로 탐지되는 사례와 비교하면 대조적이다.

6. 결론

운영 데이터 Ensemble 모델에서 고장 하루 전 Diaphragm 고장에 대해 탐지 건수가 최초로 10,000 이상 측정되었을 때, 2일 후 실제 고장이 발생하였다. 반면, PQM 데이터 Ensemble 모델은 1월 23일 고장 사례에서 고장일 전후로 평소보다 20,000건 이상 탐지되었다.

참고문헌

[1] SNE Research, "Global FCEV Monthly Tracker", "https://www.sneresearch.com/en/insight/release_view/105/page/12?s_cat=%7C&s_keyword=", Accessed: Sep 2023

[5] Xiaohan Jia, Jiahao Chen, Han Wu, Xueyuan Peng, "Study on the diaphragm fracture in a diaphragm compressor for a hydrogen refueling station", In International Journal of Hydrogen Energy 27 April 2016 41(15):6412-6421

[6] Xueying Li a, Jiahao Chen a, Zhizhong Wang a, Xiaohan Jia a, Xueyuan Peng, "A non-destructive fault diagnosis method for a diaphragm compressor in the hydrogen refueling station", In International Journal of Hydrogen Energy 13 September 2019 44(44):24301-24311

[7] Shengdong Ren, Xiaohan Jia, Jiacheng Jiang , Shengtao Zhang ,Bin Zhao, Xueyuan Peng, "Effect of hydraulic oil compressibility on the volumetric efficiency of a diaphragm compressor for hydrogen refueling stations", In International Journal of Hydrogen Energy 22 April 2022 47(34):15224-15235

[8] Uma R. Salunkhea, Suresh N. Malib, "Classifier Ensemble Design for Imbalanced Data Classification: A Hybrid Approach", In Procedia Computer Science 2016 85:725-732

[9] Do-yeon Choi, Kwang-mo Jeong, Dong Hoon Lim, "Breast Cancer Classification using Deep Learning-based Ensemble", Journal of Health Informatics and Statistics. May 30, 2018 43(2):140