

산업인공지능 개론

MINI PROJECT No. 3

산업인공지능학과 | 정원용

2021. 05. 06 (목)

목 차

1. 분석 대상 및 목적
2. 데이터 정의
3. 분석 수행
4. 분석 결과



I. 분석 대상 및 목적

1) 분석 대상

- 수소충전소에서 사용되는 설비로는 압축기, 압축가스설비, 냉각설비, 충전설비 등의 다양한 설비 중 안전성과 효율성이 중요시 되는 장비로 수소충전용 냉각기(Chiller)를 과제의 대상으로 선정함.

2) 분석 목적

- 칠러의 센서에서 수집된 데이터를 활용하여 설비의 위험도를 판별할 수 있다면, 현업에서 관리하는 모니터링 시스템에 이를 적용해 시스템을 개선하고자 함.



II. 데이터 정의

1) 데이터 정의

- 보유 데이터는 수치 값을 갖는 연속형 데이터로 1차 함수 형태로 나타내기 적합
- 칠러의 센서에서 수집된 데이터로 센서의 정상 범위가 정의되어 있음.
- 보유 데이터는 총 5,753건 임.
- 분석에 사용할 데이터는 “응축기 냉각수 입구온도”, “응축기 냉각수 출구온도” 임.

| 번호 | 항목 | 설명 | 단위 | 관리 범위 |
|----|------|---------------|-----|--------|
| 1 | EWf | 오일 차압 | kpa | |
| 2 | EIWP | 압축기 AMPS | A | |
| 3 | EOWP | 퍼지흡입 온도 | ℃ | |
| 4 | EEWT | 냉수 입구온도 | ℃ | 10 |
| 5 | ELWT | 냉수 출구온도 | ℃ | 5 |
| 6 | CAT | 응축기 머프로치 온도 | ℃ | 2.5이하 |
| 7 | EAT | 증발기 머프로치 온도 | ℃ | 2.5이하 |
| 8 | ERP | 증발기 냉매압력 | kpa | -65이상 |
| 9 | SAET | 증발기 냉매포화 온도 | ℃ | 0이상 |
| 10 | CRP | 응축기 냉매압력 | kpa | 600이하 |
| 11 | SCT | 응축기 냉매포화 온도 | ℃ | 400이하 |
| 12 | CWF | 압축기 모터권선 온도 | ℃ | |
| 13 | COWP | 압축기 냉매토출 온도 | ℃ | |
| 14 | CEWT | 응축기 냉각수 입구온도 | ℃ | 25~32 |
| 15 | CLWT | 응축기 냉각수 출구온도 | ℃ | 25~32 |
| 16 | IGVP | IGV1 Position | % | 0~100 |
| 17 | OTP | 오일탱크 압력 | kpa | -750이상 |
| 18 | DOP | 오일펌프 토출압력 | kpa | 40~120 |
| 19 | OTT | 공급 오일온도 | ℃ | 40~60 |

<표1. 칠러(chiller) 수집 센서 항목>

| TIME | EWf | EIWP | EOWP | EEWT | ELWT | CAT | EAT | ERP | SAET | CRP | SCT | CWF | COWP | CEWT | CLWT | IGVP | OTP | DOP | OTT |
|------------|-----|------|------|------|------|-----|-----|-----|------|-----|-----|-----|------|------|------|------|-----|-----|-----|
| 2020-05-12 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 60 |
| 2020-05-12 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 |
| 2020-05-12 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 60 |
| 2020-05-12 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 |
| 2020-05-12 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 62 |
| 2020-05-12 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 60 |
| 2020-05-12 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 |
| 2020-05-12 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 61 |
| 2020-05-12 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 |
| 2020-05-12 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 23 | 24 | 26 | 26 | 25 | 0 | -50 | -50 | 62 |
| 2020-05-12 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 26 | 25 | 0 | -50 | -50 | 60 |
| 2020-05-12 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -17 | 23 | 24 | 26 | 26 | 25 | 0 | -50 | -50 | 63 |
| 2020-05-12 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -17 | 23 | 24 | 26 | 26 | 25 | 0 | -50 | -50 | 60 |
| 2020-05-12 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -17 | 23 | 24 | 26 | 26 | 25 | 0 | -50 | -50 | 63 |
| 2020-05-12 | 0 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -17 | 23 | 24 | 26 | 26 | 25 | 0 | -50 | -50 | 61 |
| 2020-05-12 | 0 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -17 | 23 | 24 | 26 | 26 | 25 | 0 | -50 | -50 | 60 |
| 2020-05-12 | 0 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -17 | 23 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 |
| 2020-05-12 | 0 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -17 | 23 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 60 |
| 2020-05-12 | 0 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -17 | 23 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 |
| 2020-05-12 | 0 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 61 |
| 2020-05-12 | 0 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 23 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 62 |
| 2020-05-12 | 0 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 62 |
| 2020-05-12 | 0 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 26 | 25 | 0 | -50 | -50 | 60 |

<표2. 칠러(chiller) 수집 센서 항목의 수치 데이터>

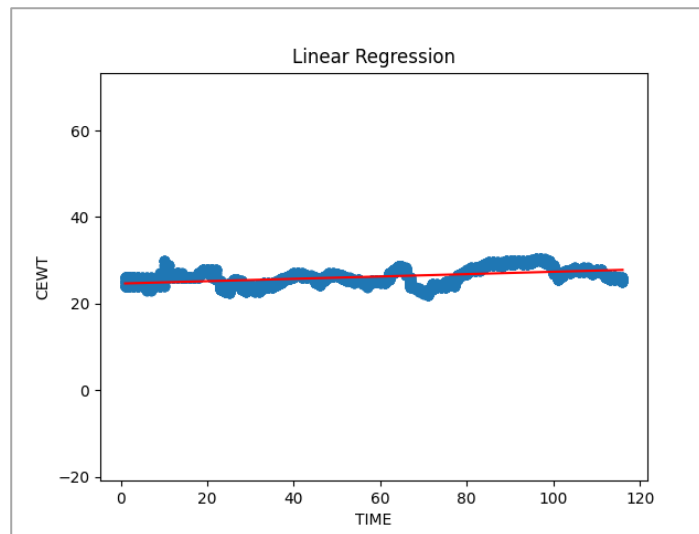
III. 분석 수행

1) 선형 회귀 분석

- 보유 데이터를 날짜를 기준으로 분석할 경우, X축 라벨이 중첩되는 문제가 발생.
- 라벨의 날짜 중첩 문제를 해결하기 위해 "TIME_NO" 컬럼으로 날짜에 대한 인덱스 컬럼 추가 함.

```
1 # 필요 Library]
2 import pandas as pd
3 import numpy as np
4 from sklearn.linear_model import LinearRegression
5 from matplotlib import pyplot as plt
6
7 # 분석용 데이터 불러오기
8 dataset = pd.read_csv('source_data.csv')
9 val = dataset.iloc[:, [1, 15]].values
10
11 data = np.array(val)
12 plt.scatter(data[:, 0], data[:, 1])
13 plt.title("Linear Regression")
14 plt.xlabel("TIME")
15 plt.ylabel("CEWT")
16 plt.axis("equal")
17
18 x1 = data[:, 0].reshape(-1, 1)
19 y1 = data[:, 1].reshape(-1, 1)
20
21 # Linear Regression 모델 학습
22 model = LinearRegression()
23 model.fit(x1, y1)
24
25 # Linear Regression 예측값 계산
26 y_pred = model.predict(x1)
27 plt.plot(x1, y_pred, color='r')
28 plt.show()
```

| 1 | TIME | TIME_NO | WF | EIWP | EOWP | EEWT | ELWT | CAT | EAT | ERP | SAET | CRP | SCT | CWF | COWP | CEWT | CLWT | IGVP | OTP |
|----|------------|---------|----|------|------|------|------|-----|-----|-----|------|-----|-----|-----|------|------|------|------|-----|
| 2 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 3 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 4 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 5 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 6 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 7 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 8 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 9 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 10 | 2020-05-12 | 1 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 |
| 11 | 2020-05-12 | 1 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 23 | 24 | 26 | 26 | 25 | 0 | -50 |
| 12 | 2020-05-12 | 1 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 26 | 25 | 0 | -50 |



- 분석 데이터가 정상 관리 범위(25℃ ~ 32℃)에 있으며, 이상 데이터는 확인되지 않음.

III. 분석 수행

2) 로지스틱 회귀 분석

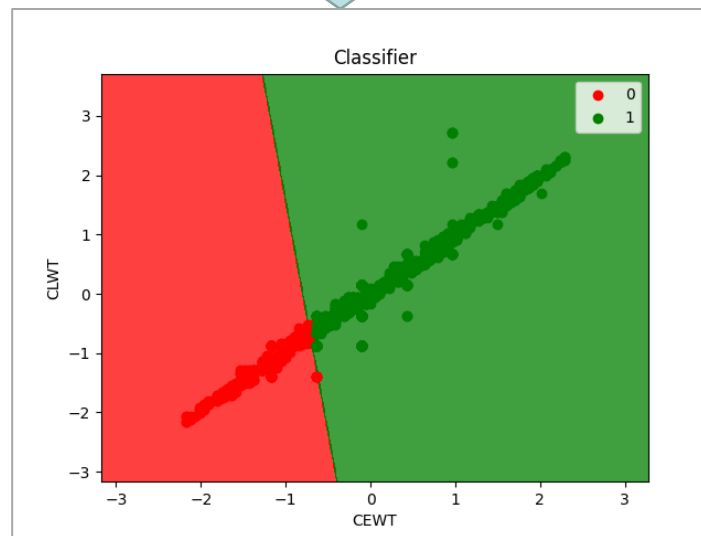
- 보유 데이터에는 **범주형 데이터가 없으므로 해당 분석을 위한 가정이 필요.**
- 가정) 센서 수집 데이터의 정상 범위가 정의되어 있으므로, 이에 해당하지 않는 데이터와 구분하여 '정상(1)', '비정상(0)'으로 별도의 컬럼을 추가함.

```

1 # 필요 Library
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5
6 # 분석용 데이터 불러오기
7 dataset = pd.read_csv('source_data.csv')
8
9 x = dataset.iloc[:, [15,16]].values
10 y = dataset.iloc[:, 23].values
11
12 from sklearn.model_selection import train_test_split
13 xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.25, random_state=0)
14
15 from sklearn.preprocessing import StandardScaler
16 sc_x = StandardScaler()
17 xtrain = sc_x.fit_transform(xtrain)
18 xtest = sc_x.transform(xtest)
19 print(xtrain[0:10, :])
20
21 from sklearn.linear_model import LogisticRegression
22 classifier = LogisticRegression(random_state=0)
23 classifier.fit(xtrain, ytrain)
24 y_pred = classifier.predict(xtest)
25
26 from sklearn.metrics import confusion_matrix
27 cm = confusion_matrix(ytest, y_pred)
28 print("혼동행렬 : \n", cm)
29
30 from sklearn.metrics import accuracy_score
31 print("정확도 : ", accuracy_score(ytest, y_pred))
32
33 from matplotlib.colors import ListedColormap
34 x_set, y_set = xtest, ytest
35 x1, x2 = np.meshgrid(np.arange(start=x_set[:, 0].min() - 1, stop=x_set[:, 0].max() + 1, step=0.01),
36 np.arange(start=x_set[:, 1].min() - 1, stop=x_set[:, 1].max() + 1, step=0.01))
37
38 plt.contourf(x1, x2, classifier.predict(
39 np.array([x1.ravel(), x2.ravel()]).T).reshape(x1.shape), alpha=0.75, cmap=ListedColormap(('red', 'green')))

```

| 1 | TIME | TIME | EWf | EWp | EOWPEWT | ELWT | CAT | EAT | ERP | SAET | CRP | SCT | CWF | COWfCEWT | CLWT | IGVP | OTP | DOP | OTT | PTPT | RUNTIME | RESULT | | |
|----|------------|------|-----|-----|---------|------|-----|-----|-----|------|-----|-----|-----|----------|------|------|-----|-----|-----|------|---------|--------|---|---|
| 2 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 60 | 414 | 0 | 1 |
| 3 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 | 414 | 0 | 1 |
| 4 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 60 | 414 | 0 | 1 |
| 5 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 | 414 | 0 | 1 |
| 6 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 62 | 414 | 0 | 1 |
| 7 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 60 | 414 | 0 | 1 |
| 8 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 | 414 | 0 | 1 |
| 9 | 2020-05-12 | 1 | 1 | 0 | 26 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 61 | 414 | 0 | 1 |
| 10 | 2020-05-12 | 1 | 1 | 0 | 25 | 9 | 10 | 0 | 0 | -48 | 10 | -16 | 24 | 24 | 26 | 25 | 25 | 0 | -50 | -50 | 63 | 414 | 0 | 1 |



```

[[-0.10334494  0.14784216]
 [ 1.1704553  1.17703544]
 [ 2.28503051  2.25768839]
 [ 0.37433015  0.35368082]
 [-0.05026993 -0.00653683]
 [-0.63409504 -0.67551247]
 [ 0.95815526  0.6624388 ]
 [ 2.28503051  2.25768839]
 [ 0.21510512  0.19930182]
 [-0.05026993 -0.00653683]]

```

혼동행렬 :
 [[384 11]
 [0 1043]]
 정확도 : 0.9923504867872045

- 분석 대상 항목 특성
 기준이 명확하고, 정상범
 위 조건에 따른 결과가
 명확하여 정확도가 높게
 나왔다고 생각됨.

IV. 분석 결과

1) 분석 결과

- 보유 데이터에 대한 특성을 명확히 알아야 그에 걸맞는 분석 기법을 통해 의미 있는 결과를 도출할 수 있다는 것을 알게 되었음.
- 보유 데이터를 통한 위험성 판별을 위해 로지스틱 회귀 분석이 현업에 더 적합하다고 판단되며, 로지스틱 회귀 분석을 위한 결과 데이터(범주형 데이터)의 확보하지 못한 부분에 대해 아쉬움이 남는 과제였음.