

기계 학습을 활용한 모터 진동의 고장 징후 분석

변준호*, 변시우**

*국민대학교 전자공학부, **안양대학교 소프트웨어학과
e-mail:cdcd1115@gmail.com,swbyun@anyang.ac.kr

Fault Sign Analysis for Motor Vibration Using Machine Learning

Joonho Byun*, Siwoo Byun**

*Dept. of Electronic Engineering, Kookmin University

**Dept. of Software, Anyang University

요약

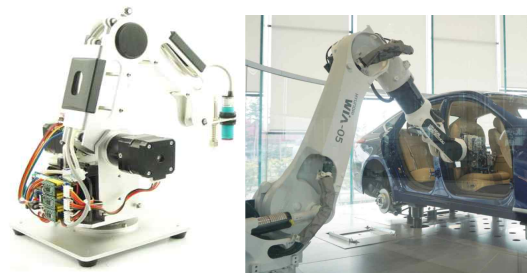
최근 IoT 센서 응용 기술이 여러 분야에서 활성화되고 있다. 그러나 IoT 시스템 복잡도가 많이 증가하여 고장 분석이 더욱 어려워지고 있다. 이를 극복하고자 최근 기계 학습 기법이 고장진단에 시도되고 있다. 본 논문에서는 고장 진단에 효과적으로 적용할 수 있는 기계 학습용 알고리즘을 살펴보고, 그중에 kNN과 SVM 알고리즘을 사용하여 고장진단을 모델을 구축하였다. 또한 모터 고장 예측용 공공 빅데이터를 대상으로 구글 CoLab으로 구현한 후, 그 성능과 정확도를 측정하고 출력 결과를 분석하였다. 진단 예측 실험의 목적은 고장에 영향을 미치는 여러 개의 센서 데이터 중에서 관련성이 큰 센서와 적은 센서를 자동으로 찾고 재학습에 연결시키기 위함이다. 실험 분석 결과, 센서 종류의 추가 및 제거 작업의 최적 조합을 SVM과 kNN 모델로 지도 학습 후 선별 예측하여, 검출률 91.7%, 오인식 오류율 0%, 정확도 95%를 얻었다. 반복 실험을 통하여 연관성이 부족한 센서 데이터를 기계 학습으로 제거하는 지능적인 필터링이 효과가 있고, 블록 단위의 최적의 연관도 학습법이 매우 필요함을 알 수 있었다. 또한 부수적으로 센서 데이터 저장 및 후속 수치 연산량의 절감 효과를 볼 수 있다.

1. 서론

최근 다양한 센서 노드로 구성된 IoT 센서 네트워크 기술이 여러 분야에서 빠르게 적용되어 크게 활성화되고 있다. 특히, IoT 빅데이터가 많이 발생하는 위험한 산업 현장에서 다양한 고장 징후를 실시간으로 예측하여 큰 사고에 대처하게 하는 등 유용하게 활용되며, 그 과정에서 매우 다양한 센서 데이터가 지속해서 대량으로 생성되고 있다[1,2,3].

산업 IoT 현장에서 고장 진단의 목적은 상존하는 고장 기기의 상세 위치와 그 고장의 원인을 밝혀내는 것이다. 그러나 최근 IoT의 시스템 복잡도가 대폭 증가하고, 측정 기기가 점차 소형화, 다양화됨에 따라서 고장 징후 분석(fault sign analysis)도 더욱 어려워지고 있다[4,5].

고 있다. IoT 분야에서도 마찬가지로 산업 장비의 상태를 실시간 감시 및 보고하고, 고장을 예측하는 기술인 장비 건진성 예측 및 예방 관리 기술도 기계 학습을 기반으로 한 연구들이 진행되고 있다[6,7].



[그림 1] 산업계 고장진단의 예 (로봇팔, 감속기, 모터)

2. 관련 연구

2.1 IoT 산업계의 고장진단 분야

최근 4차 산업 혁명과 관련하여 지능적인 기계 학습이 차지하는 비중은 날이 증가하고 있다. 성공적인 기계 학습을 위해서는 방대한 데이터양을 나타내는 빅데이터가 필수적인데, 이러한 빅데이터는 공학적 분야뿐만 아니라 여러 상업적 분야에 적극적인 활용과 심도 있는 연구개발이 꾸준히 진행되

2.2 기계 학습을 활용한 고장진단 기법

산업현장에서도 아래와 같이 다양한 기계 학습 기법이 고장 진단에 활용되고 있으며, 최근 들어서 점차 그 활용 범위가 크게 증가하고 있다[8,9].

1) SVC (Support Vector Classification) 알고리즘은 SVM (Support Vector Machine)에 속해있는 알고리즘으로 범주형 변수일 경우에 사용된다. 2개 이상의 데이터를 분류하는 분류 알고리즘이며, 높은 일반화 성능을 기대할 수 있다.

2) LR(Logistic Regression) 알고리즘은 특정 사고가 발생할 확률을 추정하는 분석 방법의 일종으로, 반응변수가 범주형일 때 사용되는 대표적인 회귀 분석 모델이고, 설명변수와 종속변수가 서로 연속형 변수일 때 회귀계수를 이용하여 관계를 분석한다.

3) DT(Decision Tree) 알고리즘은 Data Mining에 일반적으로 사용되는 방법론적 분류 기법으로 입력 변수를 바탕으로 목표변수의 값을 예측하는 모델을 생성하는 방식이다.

4) Gaussian NB(Gaussian Naive Bayes) 알고리즘은 설명변수가 연속형일 때 사용하는 기법으로, 조건부 확률에 베이즈 정리를 적용하고, 데이터를 구성하는 각각의 요소들이 등장할 확률에 대한 독립성을 가정하여 입력 벡터를 확률적으로 분류하는 방식이다.

5) kNN(k-Nearest Neighbor) 알고리즘은 학습 또는 예측하고자 하는 값으로부터 가장 가까운 거리에 있는 k개만큼의 레이블을 확인 후, 가장 많이 도출된 레이블을 값으로 취하는 방식이다. 즉, 새로운 사례 데이터와 사용 가능한 사례 사이의 유사성을 가정하고 사용 가능한 범주와 가장 유사한 범주에 새로운 사례를 넣는 기법이다.

kNN 알고리즘은 사용 가능한 모든 데이터를 저장하고 유사성을 기준으로 새로운 데이터 포인트를 분류한다. kNN 알고리즘은 분류뿐만 아니라 회귀 분석에도 사용할 수 있지만 대부분 분류 문제에도 사용된다. 학습 세트에서 즉시 학습하지 않고 데이터 세트를 저장하며 분류할 때 데이터 세트에 대한 작업을 수행하기 때문에 게으른 학습자 알고리즘이라고도 한다. kNN 알고리즘은 데이터 세트를 저장하고 새로운 데이터를 받으면 해당 데이터를 새로운 데이터와 매우 유사한 범주로 분류한다.

2.2 고장진단 모델의 성능 평가

고장진단 모델의 성능은 그림2의 분류 성능 평가지표를 이용하여 평가할 수 있다. 여기서 TP는 True Positive로 모델이 True라고 예측한 것이 실제 정답은 True임을 의미하고, TN은 True Negative로 모델이 False라고 예측한 것이 실제 정답은 False임을 표시한다. 또한, FP는 False Positive로 모델이 True라고 예측한 것이 실제 False임을 의미하고, FN은 False Negative로 모델이 False라고 예측한 것이 실제 정답은 True임을 표시한다.

		True Label	
		True	False
Predicted Label	True	True Positive	False Positive
	False	False Negative	True Negative

[그림 2] 성능 평가 지표상의 분류

Accuracy는 올바르게 예측된 데이터의 수를 전체 데이터의 수로 나눈 값이고, Recall은 실제의 수를 의미하며, Precision은 모델이 True로 예측한 데이터 중 실제 True인 데이터의 수이다. F1-score는 Precision과 Recall의 조화 평균이다.

3. 기계 학습을 활용한 고장 징후 분석

3.1 실험 환경

본 연구에 사용된 알고리즘은 여러 후보 중에 kNN과 SVM 알고리즘을 사용하여 고장 예측을 진행하였다. kNN 알고리즘은 다른 알고리즘과 비교 시 간단하면서 훈련 속도가 빨라 사용이 쉽다는 장점이 있다. SVM은 일반적으로 다양한 작업에 사용할 수 있는 보편성과 성능상의 장점이 있다. 본 실험은 아래의 실험 환경을 구축하고, 공공데이터 일부를 활용하여, 인공지능 학습 기반의 SVM 및 kNN 기법의 성능과 정확도를 구글 CoLab으로 측정 후, 그 결과를 분석하였다.

- 1) 실험 데이터: (블록당 100개*20)*2,340파일
- 2) 실험 횟수: 1차 세부 실험 반복 수:20회, 이후 1,000회.
- 3) 측정지표: 정확도(accuracy), 검출 성공률(hit-ratio), 검출오류율(false-alarm)

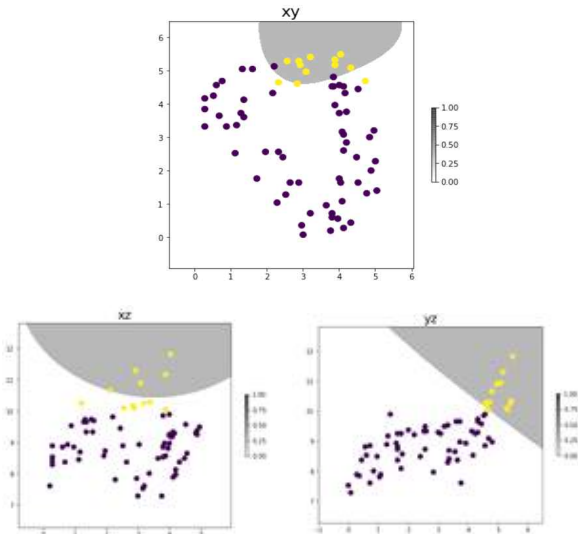
실제로는 10개 이상의 센서 데이터를 사용하나, 실험의 편의상 이들 중 연관성이 상대적으로 높아 보이는 3종(x,y,z)과 연관성이 낮은 나머지 중 대표로 1개(h)를 포함하여, 네 종류의 센서 데이터를 사용하였다.

본 실험에서 두 가지 고장진단 알고리즘을 적용한 결과를 그림 3과 그림 4에 그래프로 표시하였다.

3.2 SVM 기반 모델의 실험 분석

이 실험은 여러 종류의 센서 데이터를 입력 값으로 한 지도 학습에 대한 성능 평가이다. 실험의 목적은 고장에 영향을 미치는 여러 개의 센서 데이터 중에서 관련성이 큰 센서와 적은 센서를 자동으로 찾기 위함이다. 즉, 가능한 한 높은 예측 성향(징후)을 보이는 고장과의 관련성이 큰 센서 후보들을 함께 묶고, 반대로 관련성이 적은 센서 후보는 미리 제거하는 지능적인 데이터 필터링 작업을 수행하고자 한다. 결과적으로 유용한 센서 데이터만 선택되고, 불필요한 센서 데이터가 사전에 제거되므로, 추후 데이터 저장 공간 및 분석 작업 부하가 줄고 전체적인 시스템 성능이 향상된다.

그림3은 센서 조합별 지도 학습된 SVM 기법의 실험 결과이다. 실제 현장은 전체 10여 가지의 센서가 존재하여 그 조합이 너무 많으므로, 그중에서 가장 중요한 3개(x,y,z)를 사전에 민감도 테스트를 통하여 1차 선별하였고, 그중에서 최적의 두 가지 조합(xy, xz, yz)을 찾는 것이다.



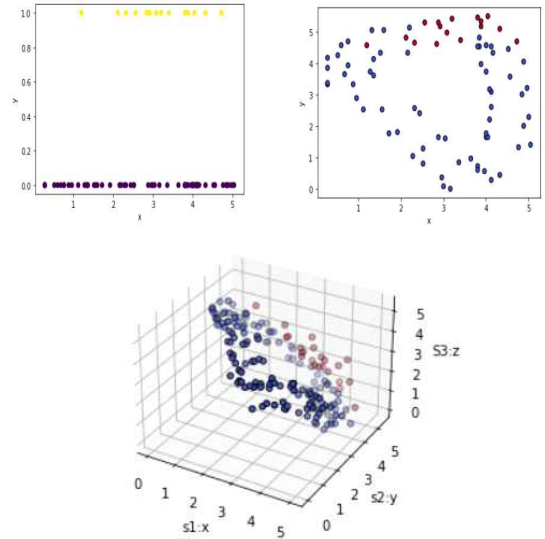
[그림 3] SVM 기법을 활용한 블록 단위 실험 결과의 예

예를 들어, 그림 3의 상단 그래프는 센서 x와 y 조합의 결과이며, 한 파일당 20개의 블록으로 구성하고, 약 100개 단위로 블록화 하여 반복 실험한 결과이다. 노란 점은 고장으로 인식된 데이터이고, 보라색은 정상으로 인식된 데이터이다. 회색으로 분리된 영역에 노란 점 12개 중에서 10개의 고장을 정확히 인식(hit)하였고, 보라색 점 2개는 고장으로 오인식(false alarm) 되었다.

이 실험에서 지도 학습용 SVM 커널은 여러 개 중에서 3종류(linear, poly, RBF)를 선별하여 사용되었으며, 이중 최적의 커널이 산출한 결과를 종합하여 분석하였다. 실험 분석 결과, 센서 데이터 yz의 조합이 검출률 91.7%, 오류율 0%, 정확도 95%로서 가장 이상적인 조합이었다. 참고로 수십 가지의 전체 센서 종류를 모두 포함하면 최고의 고장 인식이 될 것 같지만, 실상은 그렇지 않다. 연관성이 부족한 센서 데이터가 포함되면 기계 학습을 방해하여, 실질적인 고장 인식물을 저하시키기 때문이다. 또한, 전체 xyz 3종류에서 2종류로 센서 데이터를 줄임으로써, 단순히 계산해도 1/3의 저장 및 수치 연산의 절감 효과를 볼 수 있다.

3.3 kNN 기반 모델의 실험 분석

kNN 알고리즘 기반의 실험도 위의 모델과 같은 방식으로 진행하였다. 추가로 데이터를 100개 단위에서 600개 단위로 학습에 참여시켜주면 편차가 5% 정도까지 내려갔다. 별도로, kNN에서는 고장 연관 센서를 1개, 2개, 3개로 점차 증가시키면서 측정해 보았다. 센서를 1개 사용 시 0.81, 2개 사용 시 0.94, 3개 사용 시 0.943으로서, 2개만 사용해도 충분히 우수한 결과가 나왔다. 그림4는 1개, 2개, 3개를 적용한 결과 그래프이다.



[그림 4] kNN 기법을 활용한 블록 단위 실험 결과의 예

추가로 다른 센서 하나를 더 투입하여 세 가지(s1,s2,s4) 조합에 대하여 실험해 보았다. 결과는 0.9로서 두 가지만 포함한 (s1,s4) 조합의 결과치인 0.85 보다 높은 0.9가 나왔다. 하지만, 연관성인 높은 두 가지 조합인 (s1,s2)의 결과치인 0.94 보다 낮은 결과치이다.

위의 실험 결과를 종합해 보면, 많은 센서 종류를 포함하여 판단하는 것은 오히려 정확도를 떨어뜨리며, 불필요한 데이터 저장과 연산량 부하를 촉발함을 알 수 있었다. 또한, 연관성이 높은 센서 종류가 여러 개 있더라도 그중 일부만 잘 선택하여 사용하면 충분히 우수한 정확도를 얻을 수 있으며, 데이터와 후속 연산 부하를 줄일 수 있음을 알 수 있었다.

즉, 관련성이 높은 센서 종류를 잘 선별하고, 불필요한 센서 종류를 제거하는 지능적인 학습법이 효과가 있고, 매우 중요함을 알 수 있었다. 특히, 세부적으로 센서 데이터 스트림의 추이를 잘 분석하고, 최적의 센서 조합(복합 센서 컨텍스트)과 이 조합에 맞는 최적의 학습 모델(SVM 커널과 kNN)과 적절한 연관도 튜닝 인자를 찾는 것이 중요함을 알 수 있다.

[코드 1] 기계 학습과 테스트를 위한 파이선 모듈

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import classification_report,
confusion_matrix,accuracy_score
from mlxtend.plotting import plot_decision_regions

#Function to load data from an excel file and select
columns specified by col
def load_data(col): # Select sensor data block
    data=pd.read_excel('data.xlsx',usecols=col,header=8)
    SensorDataBlock=data.loc[0:BlockSize]
    return SensorDataBlock

# Function to train an SVM model using X train and
v train.
# predict labels for X test using the trained model.
# and return v train and v pred
def SVM(X train,X test,v train,v test):
    svm=SVC()
```

```

svm.fit(X_train,y_train)
# Train the model using X_train and y_train
y_pred=svm.predict(X_test) #Predict labels for X_test
return y_train,y_pred

# Function to split data into training and test sets,
# train an SVM model using the training set,
# predict labels for the test set using the trained
model, and return the accuracy score

def Train_and_Test(data):
    data=np.array(data) #Convert data to a numpy array
    X=data[:, :-1]
    y=data[:, -1].astype(int)

    #Split the data into training and test sets
    X_train,X_test,y_train,y_test=train_test_split(X,y,
test_size=0.2)
    #Train the SVM model and predict labels for the test
set
    y_train,y_pred=SVM(X_train,X_test,y_train,y_test)
    accuracy = accuracy_score(y_test,y_pred)
    return accuracy

# Main function to iterate over different combinations
of columns to be used for training, load data,
# calculate accuracy, and store the mean accuracy for
each combination of columns in a dictionary

def main():
    columns = {
        1: 's1', # x Sensor data
        2: 's2', # y Sensor data
        3: 's3', # z Sensor data
        # ...
        10: 's4', # h Sensor data
        12: 'fault sign', # fault sign ...
    }
    acc_dict = {}

    #loop through different combinations
    for cols in ['all combination of sensor data']:
        col_names = [columns[col] for col in cols]
        data = load_data(cols) # load data
        acc = []
        for i in range(100): # repeat N times
            accuracy=Train_and_Test(data)#accuracy
            acc.append(accuracy)
        mean_acc = sum(acc) / len(acc) # mean accuracy
        acc_dict[tuple(col_names)] = mean_acc

    sorted_acc = sorted(acc_dict.items(), key=lambda x:
x[1], reverse=True) #sort by accuracy
    for cols, acc in sorted_acc:
        print(f"Variables: {cols}, Mean Accuracy:
{acc:.4f}")
#...

```

4. 결론

본 논문에서는 산업 IoT 환경에서 고속 결함 진단과 효과적인 센서 데이터의 필터링을 위하여 kNN과 SVM 알고리즘을 활용한 고장 진단 모델을 구현 및 분석하였다.

실험을 위하여 모터 고장 관련 공공 빅데이터와 구현 도구로서 구글 CoLab 사용하였다. 또한 고장에 영향을 미치는 여러 개의 센서 중에서 영향력이 큰 센서 조합과 작은 센서를 자동으로 선별하기 위하여 고장 연관성에 대한 기계 학습을 반복하였다. 실험 결과 고장 연관성을 기계 학습으로 블록단위로 추적하면서 반복적으로 조합 및 제거하는 지능적인 지도 학습 기법이 효과가 있었고, 예측 정확도도 그대로 유지함을 알 수 있었다. 센서 종류의 최적 조합을 지도 학습하여, 검출률 91.7%, 오류율 0%, 정확도 95%를 얻었다.

향후 연구과제로서 보다 다양한 센서 필터링 지능을 탑재한 하이브리드 딥 러닝 구현과 실제 산업 현장의 노후된 장치

를 고려한 현장 실험과 분석을 추진하고자 한다.

참고문헌

- [1] C. Sekhar, N.Suneetha, A Study on Data Categorization for Data Analytics, International Journal of Internet of Things and Big Data. 3(1), 1-6, 2018.
- [2] L. Ashwini, H. Lee, M. Hwang., Implementation of IoT Application using Geofencing Technology for Protecting Crops from Wild Animals. Asia-pacific Journal of Convergent Research Interchange. 6(6),13-23, 2019.
- [3] J. Lee, H. Kim, Design and Implementation of IoT Devices to Foster Emotions and Prevent Dementia for Senior Citizens. Journal of Next-generation Convergence Technology Association. 5(3), 347-353, 2021.
- [4] S. Mansfield, E. Vin & K. Obraczka, An IoT-Based system for autonomous, continuous, real-time patient monitoring and its application to pressure injury management. International Conference on Distributed Computing in Sensor System, 66-68. DOI : 10.1109/DCOSS52077.2021.00024, 2021.
- [5], H. Lee, S. Kim, S. Kang, Y. Lim, A Fault Dropping Technique with Fault Candidate Ordering and Test Pattern Ordering for Fast Fault Diagnosis, Semiconductor and Devices, 46(3). 32-40, 2009.
- [6] Choi, J. H., "Introduction of Failure Prognosis and PHM", Journal of The Korean Society of Mechanical Engineers, Vol. 53, No. 7, pp. 24-34, 2013.
- [7] Joon-Hyuk Lee, Min-Gyeong Kwon, Young-Beom Kim, Jang-Wook Hur, Failure Diagnostics of Camera Image Sensor For Vehicle Using CNN, Journal of the Korea Academia-Industrial cooperation Society Vol. 23, No. 12 pp. 877-884, 2022.
- [8] Myagila, Kasian, Hassan Kilavo. "A comparative study on performance of SVM and CNN in Tanzania sign language translation using image recognition." Applied Artificial Intelligence, Vol. 36, Issue. 1, 2022.
- [9] Seon-Mo Kim, Ha-Ye-Rim Lee, In-Soo Hwang, Jang-Wook Hur, Analysis of Fault Diagnosis Algorithm for Thermal Imaging Camera Circuit Board Using Machine Learning, Journal of the Korea Academia-Industrial, cooperation Society, Vol. 23, No. 10 pp. 118-124, 2022