

유도전동기에서 다변량 커널 밀도 추정 기반의 이상치 탐지

심영준, 최준규, 김보배, 임성빈*

송실대학교

simyoungjun@gmail.com, cjg@soongsil.ac.kr, kbb@nate.com, *sbi@ssu.ac.kr

Outlier Detection Based on Multivariate Kernel Density Estimation in Induction Motors

Youngjun Sim, Jungyu Choi, Bobae Kim, Sungbin Im*

Soongsil University

요 약

4차 산업 혁명 시대에는 사물 인터넷 기술 보급과 함께 점차 더 다양한 단말들이 사물 인터넷에 연결되고 있고, 그에 따라 단말들이 생성하는 사물 데이터도 급증하는 추세이다. 스마트 기술, 데이터, 자동화, 상호 연결성, 인공지능 등의 기술은 4차 산업 혁명의 핵심 기술이며, 이러한 기술은 공장과 산업 조직의 운영 방식을 혁신적으로 변화시키고 있다. 그 중 예측 유지 보수 기술(Predictive Maintenance)은 현재 가장 눈에 띄는 IIoT(IIoT, Industrial Internet of Things) 적용 사례이자 스마트 팩토리의 확대 및 고도화를 위한 시작점이라 볼 수 있다. 이를 위한 장에 감지 및 프로세스 모니터링은 프로세스 제어의 핵심 기술 중 하나이다. 본 논문에서는 이상 탐지의 기법 중 하나로 유도 전동기의 이상 탐지를 위한 다변량 커널 밀도 추정(MKDE, Multivariate Kernel Density Estimation)을 제안한다. 3상 유도 전동기에서 획득한 전류 데이터를 이용하여 실험한 결과, 특정한 확률분포함수와 상관관계에 대한 정의 없이 오직 데이터만을 사용하여 결합 확률 밀도 함수를 추정하고, 라벨이 없는 데이터에 대해 이상치를 탐지할 수 있다는 점에서 기계 시설물의 이상치 탐지에 적합하다고 판단된다.

I. 서론

4차 산업혁명시대의 스마트 기술, 데이터, 자동화, 상호 연결성, 인공지능 등의 기술은 공장과 산업 조직의 운영 방식을 혁신적으로 변화시키고 있다. 이상 탐지는 스마트 팩토리의 확대 및 고도화를 위한 핵심 기술 중 하나이며 제조업뿐만 기계 센서 모니터링, 휴먼 헬스 모니터링, 금융, 기후 변화, 통신 데이터 등 광범위한 분야에 응용되고 있다. 그 중 유도 전동기는 산업계의 중요 전력설비로서 각종 생산설비 및 유틸리티 설비의 구동을 위한 핵심역할을 수행하고 있으며, 현재 유도 전동기의 고장 진단 방법에 대하여 많은 연구가 이루어지고 있는 추세이다.

II. 본론

푸리에 변환(FFT, Fast Fourier Transfrom)이란 시간에 대한 함수 신호를 주파수 성분으로 분해하는 작업이다. 고속 푸리에 변환(Fast Fourier Transfrom, FFT)는 이산적인 함수 신호에 대한 푸리에 변환과 그 역변환을 빠르게 수행하는 효율적인 알고리즘이다.

함수 신호를 푸리에 변환하여, 주파수 스펙트럼으로 주파수 성분을 분석하였을 때 기본파와 고조파(harmonics) 성분으로 구분할 수 있다. 고조파가 상대적으로 기본파에 비해 클수록 왜곡이 커지며 즉, 이는 순수한 정현파의 파형에서 멀어지게 된다는 의미이다. 전 고조파 왜율(THD, Total Harmonic Distortion)이란 고조파로 인한 파형의 왜곡 정도를 나타내는 지표이다. 고조파 성분이 많을수록 THD 값이 높아지게 되면 정현파로부터의 왜곡이 심하다는 것을 의미한다. THD 계수에 대한 일반 방정식은 다음과 같다.

$$THD(w) = \frac{\sqrt{\sum_{n>1} Y_n^2(w)}}{Y_1} \quad (1)$$

Y_1 은 기본파의 크기이며 Y_n 은 고조파의 크기이다.

밀도 추정이란 데이터로부터 변수가 가질 수 있는 모든 값의 확률, 즉 확률 밀도 함수(PDF, Probability Density Function)를 추정하는 행위이다. 확률 밀도 함수는 확률 변수의 분포를 나타낸 함수인데, 구간 $[a, b]$ 에서 확률 변수 X 가 해당 구간에 있을 확률은 다음과 같이 표현이 가능하다..

$$P(a < X < b) = \int_a^b f(x)dx \quad (2)$$

이러한 밀도추정 방법은 다양한 종류가 존재하는데, 크게 Parametric estimation 과, Non-parametric estimation으로 나뉜다. Non-parametric estimation 은 순수 관측 데이터만으로 확률 밀도 함수를 추정하는 기법인데, 가장 간단한 형태는 히스토그램(Histogram)형태이다. 하지만 히스토그램을 이용한 밀도 추정은 불연속성이 나타나고, 고차원 데이터에는 메모리 등의 문제가 있다. 이를 보완한 대표적인 추정법에는 본 논문의 주제인 Kernel Density Estimation이 있다. KDE는 커널(Kernel) 함수를 이용한 밀도 추정방법이다. 커널 함수는 원점을 중심으로 대칭이며 적분 값이 1인 양의 함수로 정의할 수 있다.

KDE는 이들을 사용하여 주어진 데이터의 분포를 예측하는데 다음과

같은 수식을 사용한다 [1].

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (3)$$

K는 커널 함수를 뜻하고, h는 커널함수의 Bandwidth 값으로 커널의 분산을 조절하는 파라미터이다. Bandwidth 값이 작을수록 커널이 뾰족한 형태를 띄고, 클수록 커널이 완만한 형태가 된다. 데이터의 확률 밀도 함수를 히스토그램으로 추정했을 때와 KDE를 이용했을 때의 그림 1과 같다. 초록색은 값이 작은 Bandwidth를 사용한 것이고 붉은색은 값이 큰 Bandwidth를 사용한 것이다.

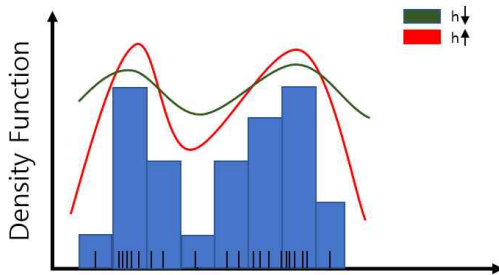


그림 1 히스토그램, 커널 함수를 이용한 확률 밀도 함수

다변량 커널 밀도 추정(MKDE)은 특정한 확률 분포 함수(CDF, Cumulative Distribution Function)와 상관 관계에 대한 정의 없이 오직 데이터만을 사용하여 결합 확률 밀도 함수 (JPDF, Joint Probability Density Function)를 추정하는 비 모수적 통계 모델링 방법이다. MKDE는 분포함수의 형태가 고정되지 않고 데이터에 따라서 유연하게 표현 할 수 있기 때문에 분포 함수의 표현의 자유도가 매우 높은 기법이다. MKDE는 각 데이터에 동일한 커널 함수 (kernel function)를 정의하고 모든 커널 함수를 최적의 대역폭에 따라서 합하면 최종적으로 결합 확률 밀도 함수가 추정되게 된다 [2]. Silverman's rule of thumb에 의하면 최적의 대역폭을 구하는 식은 다음과 같다 [3].

$$h_i = \sigma_i \left\{ \frac{4}{n(d+2)} \right\}^{\frac{1}{(d+4)}}, \quad i = 1, 2, \dots, d, \quad (4)$$

σ_i 는 i 번째 feature에 대한 표준편차이며 d 는 차원의 수이다.

III. 실험

본 논문에서는 가우시안 함수를 커널 함수로 사용한 다변량 커널 밀도 추정을 이용해 유도 전동기의 이상을 탐지하였다. 이 연구에 사용된 데이터 세트는 AI Hub에서 제공한 “기계 진동 및 전류 데이터”로 기계 시설물의 상태 모니터링 및 고장 진단을 위해 활용하는 진동 및 전류 데이터이다 [4]. 총 11948개의 전류 신호 데이터 중 정상 데이터와 비 정상 데이터의 수는 각각 10000개, 1948개 이다. Train set과 Test set을 각각 5974개로 나누어 진행 하였다. FFT 변환 후, 신호를 분석하여 기본 주파수와 전 고조파 왜율 (THD) 을 특징 추출에 이용하였다. R, T, S의 3상의 신호에 대한 3개의 120Hz 기본 주파수와, THD를 계산하여 총 6개의 feature를 생성하였다. Bandwidth는 Silverman's rule of thumb를 이용하여 계산하였다. 그림 2는 전류 신호 데이터를 FFT하여 나타낸 주파수 스펙트럼, 그림 3은 다변량 커널 밀도 추정 기반의 이상치 탐지 흐름도이다.

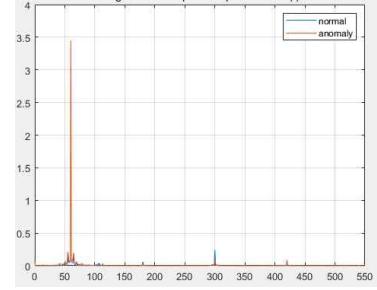


그림 2 전류 신호 FFT 분석

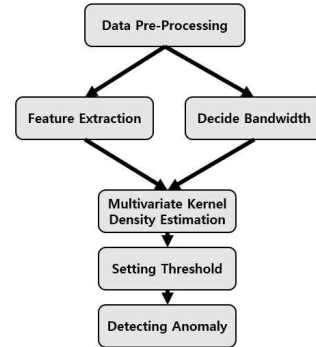


그림 3 다변량 커널 밀도 추정에 기반한 이상치 탐지 흐름도

다변량 커널 밀도 추정을 이용한 이상치 확률과, Silverman's rule of thumb를 이용한 대역폭 값을 조정하면서 최적의 Threshold를 계산하였다. 그 결과 98.93% 정확도를 얻을 수 있다.

IV. 결론

본 논문에서는 유도 전동기의 이상치 탐지를 위해 다변량 커널 밀도 추정 기법을 사용했다. 총 5974개의 test 데이터에 대해 정확도 평가를 수행하였고, 98.9%의 결과를 얻었다. 다변량 커널 밀도 추정은 라벨이 없는 데이터로 모델의 훈련이 가능하다. 또한 상대적으로 적은 수의 이상 데이터를 사용하여 데이터 불균형이 있는 상황에서도 좋은 성능을 내는 것으로 보인다. 따라서 기계시설물에 대한 이상치 탐지에 적합하다고 판단된다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021R1F1A1061403, 반도체 식각 공정에서 TadGAN을 이용한 Predictive Maintenance 기반의 EPD 추정)

참고 문헌

- [1] Parzen, Emanuel. "On estimation of a probability density function and mode." The annals of mathematical statistics 33.3 (1962): 1065-1076.
- [2] Y.-J. Kang, Y. Noh, and O.-K. Lim, "Development of MKDE-ebd for Estimation of Multivariate Probabilistic Distribution Functions," Journal of the Computational Structural Engineering Institute of Korea, vol. 32, no. 1, pp. 55 - 63, Feb. 2019.
- [3] Scott, David W. Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons, 2015.
- [4] "AI 허브", 기계시설물 고장 예지 센서. 2021년 6월 18일 수정, 2021년 8월 30일 접속, <https://aihub.or.kr>.