

딥러닝 기반 상관계수 예측 및 Modified Z-Score를 활용한 이상 징후 감지

남다운, 신요안*

승실대학교 전자정보공학부

dwnam@soongsil.ac.kr; *yashin@ssu.ac.kr

(*교신저자)

Anomaly Detection Using Deep Learning-Based Correlation Prediction and Modified Z-Score

Dawoon Nam, Yoan Shin*

School of Electronic Engineering, Soongsil University

(*Corresponding author)

요약

본 논문은 시계열 센서 데이터들의 상관계수를 예측하기 위해 LSTM (Long and short-Term Memory) 딥러닝 모델을 활용하고, Modified Z-Score를 활용하여 임계값을 지정함으로써 이상 징후를 감지하는 방안을 제시한다. 레이블링 된 사출 성형기의 이상치 데이터를 통해 이상 징후 감지 여부를 검증하였다.

I. 서론

최근 산업 현장에서는 다양한 센서 데이터를 실시간으로 모니터링하여 공정의 이상 여부를 감지하는 것이 중요해지고 있다. 특히 사출 성형 공정과 같은 복잡한 제조 공정에서는 딥러닝과 같은 인공지능 알고리즘을 도입하여 다양한 센서 데이터를 통합적으로 분석하여 이상 징후를 조기에 탐지하는 것이 필요하다[1]. 본 연구에서는 시계열 데이터로 이루어진 센서 데이터의 상관계수를 예측하기 위해 LSTM (Long and short-Term Memory) 딥러닝 모델을 활용하고, Modified Z-Score 기법을 통해 이상 징후를 감지하는 방안을 제안한다.

II. 제안 방식

실제 작동하는 공장의 사출 성형기에서 수집한 시계열 데이터의 교차되는 입력 특성을 제거하고, 데이터 간의 상관관계를 분석함으로써 일반화 성능을 향상시킬 수 있다[2]. 본 논문에서는 91개의 특성들 중 주요 특성으로 선정한 “최대사출압력”, “바렐 온도” 두 특성을 슬라이딩 윈도우 기법을 이용하여 상관계수로 이루어진 데이터 세트를 도출하고, LSTM 딥러닝 모델을 이용하여 예측을 진행하였다. LSTM 모델은 시계열 데이터를 처리하는 데 특화된 인공 신경망 구조로, 장기 의존성을 효과적으로 학습할 수 있어 상관계수 예측에 적합하다[3].

그림 1은 LSTM 예측 모델의 상관계수 예측 결과를 도시한다. 평균제곱오차를 활용하여 예측 결과를 확인한 결과 0.1650으로 적절한 예측 성능을 확인하였다.

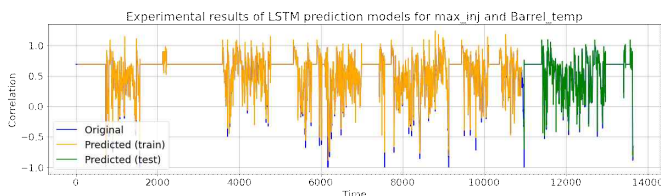


그림 1. LSTM 예측 모델 실험 결과

예측된 상관계수와 라벨링된 이상 데이터를 바탕으로 Modified Z-Score 기법[4]을 적용하여 이상 데이터를 탐지한다. Modified Z-Score는 데이터의 중앙값과 MAD (Median Absolute Deviation)를 활용하여 이상치를 탐지하는 방법으로, 표준 편차 기반의 Z-Score보다 이상치에 민감하게 반응한다. 본 연구에서는 Modified Z-Score의 임계값을 설정하고, 이를 통해 이상 데이터를 탐지하였다.

III. 실험 결과 및 결론

그림 2는 실제 이상 징후가 일어난 데이터를 붉은 선으로 시각화한 결과이다. 그림 3은 Modified Z-Score를 이용하여 이상 탐지 결과로 그림 2인 레이블링 된 이상치 데이터와 잘 일치하였으며, 임계값 조정을 통해 탐지 성능을 최적화 할 수 있었다. 제안된 방법은 실제 공정 데이터에 적용하여 높은 정확도로 이상 징후를 감지할 수 있을 것으로 기대된다.

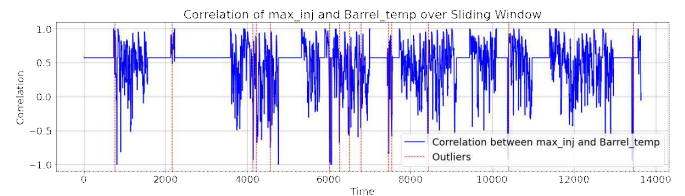


그림 2. 레이블링된 이상치 데이터 시각화

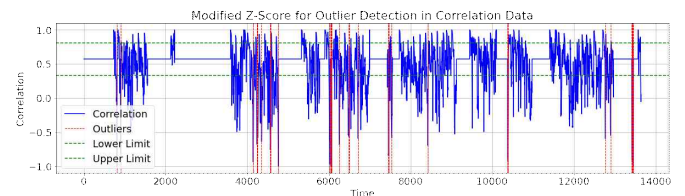


그림 3. Modified Z-Score의 이상 탐지 결과

ACKNOWLEDGMENT

본 논문은 정부 (과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2023-00251595)

참 고 문 헌

- [1] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," *Jour. Manufacturing Syst.*, vol. 48, pp. 157-169, July 2018.
- [2] 남다운, 이재승, 김병주, 신요안, "시계열 빅데이터 예측을 위한 상관분석 기반 딥러닝 학습 데이터 선정," *2023년도 한국통신학회 동계종합학술발표회 논문집*, pp. 803-804, 2023년 2월.
- [3] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, Mar. 2020.
- [4] M. Wu, M. Sun, F. Zhang, L. Wang, N. Zhao, J. Wang, and W. Huang, "A fault detection method of electric vehicle battery through Hausdorff distance and modified Z-score for real-world data," *Jour. Energy Storage*, vol. 60, Apr. 2023.