

생산 설비의 이상탐지를 위한 불규칙 샘플링 시계열 데이터 보정 기법

신강현 · 진교홍*

창원대학교

Irregularly-Sampled Time Series Correction Method for Anomaly Detection in Manufacturing Facility

Kang-hyeon Shin · Kyo-hong Jin*

Changwon National University

E-mail : sheen98@changwon.ac.kr / khjin@changwon.ac.kr

요 약

제조 설비에서 짧은 주기로 수집된 제조 데이터는 시간 간격이 일정하지 않은 불규칙 샘플링 시계열이고 값이 불안정하여 큰 분산을 가지는 경우가 많다. 본 논문에서는 단순이동평균법을 이용하여 불규칙 시계열의 시간 간격을 일정하게 보정함과 동시에 값의 분산을 줄이는 보정 기법을 제안하고, 제안된 보정 기법이 생산 설비의 이상탐지의 성능 향상에 효과가 있음을 확인하였다.

ABSTRACT

There are many irregularly-sampled time series in the manufacturing data which are collected from manufacturing facilities by short intervals. Those time series often have large variance. In this paper, we propose irregularly-sampled time series correction method based on simple moving average. This method corrects time intervals between neighboring values in time series regularly and reduces the variance of the values at the same time. We examine that this method improves performance of anomaly detection in manufacturing facility.

키워드

Manufacturing Data, Anomaly Detection, Irregularly-Sampled Time Series, Simple Moving Average, Data Preprocessing

1. 서 론

스마트 팩토리를 위한 어플리케이션 중 하나인 이상탐지 시스템은 인공지능을 활용하여 설비 이상 초기에 이상 상태를 실시간으로 탐지하고 정비를 실시하여 기회 손실을 방지하고 정비 비용을 절감하는 것이 목적이다[1, 2].

제조 설비의 실시간 이상탐지를 위해서는 설비의 변화를 바로 인지할 수 있을 정도로 짧은 주기의 데이터 수집이 요구된다. 이런 환경에서 하나의

설비에서 수집되는 모든 센서 값들은 정확한 간격으로 동일한 순간에 측정되기가 어렵기 때문에 데이터 시간 간격이 일정하지 않은 불규칙 샘플링 시계열 데이터인 경우가 많다. 서로 다른 센서로부터 수집된 여러 불규칙 샘플링 시계열 데이터를 머신러닝에 적용하기 위해서는 시간 간격을 일정하게 보정해주는 작업이 필요하다. 이렇게 시간 간격이 일정하지 않은 시계열 데이터를 보정하는 방법으로는 균일한 시간 간격으로 값을 선형 보간법으로 리샘플링하는 기법이 널리 사용된다[3, 4].

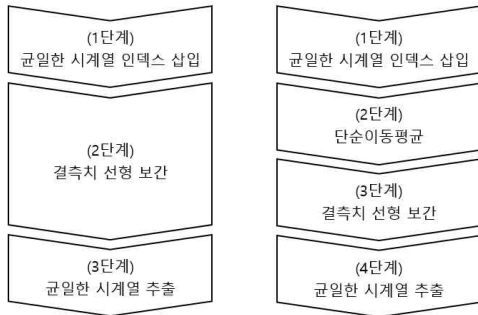
한편, 제조 데이터에는 값이 불안정하여 큰 분산을 가지는 경우도 있다. 이렇게 분산이 크면 데이

* corresponding author

터가 일반화되어있지 않아서 답러닝 모델이 큰 값에 민감해지고 일부 데이터에 과적합할 수 있다[5]. 본 논문에서는 생산 설비에서 수집되는 불규칙 시계열의 시간 간격을 보정하는 과정에 단순이동평균법을 적용하여 동시에 분산을 줄일 수 있도록 하고, 제안된 보정 기법이 생산 설비의 이상 탐지 모델 성능에 보이는 효과를 확인하였다.

II. 단순이동평균을 이용한 불규칙 샘플링 시계열 데이터 보정 기법

기존 보정 기법은 선형 보간법을 이용하여 균일한 시간 간격마다의 값으로 앞뒤 값과 선형적인 위치에 있는 값을 구한다[4]. 제안된 보정 기법은 단순이동평균법을 이용하여 균일한 시간 간격마다의 값으로 앞선 일정한 구간 동안의 평균을 구한다.



(A) 기존 보정 기법 (B) 제안된 보정 기법
 그림 1. 불규칙 시계열 보정 기법

기존 보정 기법은 그림 1-(A)에 나타난 방식으로 진행된다. 먼저, 원하는 일정한 시간 간격의 시계열 인덱스를 불규칙 샘플링 시계열 인덱스에 병합하여 균일한 시계열 인덱스를 삽입한다. 그러면 일정한 시간 간격을 따라 값이 없는 시점마다 결측치가 생긴다. 다음 단계로, 생겨난 결측치를 선형 보간법으로 보간한다. 마지막으로 처음에 삽입했던 균일한 시계열 인덱스에 있는 값만 추리면 균일한 시계열 데이터를 얻을 수 있다.

표 1과 같은 불규칙 샘플링 시계열 데이터를 예로 들어 기존 보정 기법을 적용하여 시간 간격이 1초로 균일한 시계열 데이터로 보정한 예시를 표 2에 나타내었다. 1단계에서는 시계열 인덱스에 1초 단위의 새로운 인덱스를 삽입한다. 원래 값이 있던 4.0초를 제외한 대부분의 새로운 인덱스에는 결측치가 생긴다. 2단계에서는 생겨난 결측치를 선형 보간법으로 보간한다. 3단계에서 1초 단위 인덱스에 있는 값만 가져오면 원본 데이터를 1초 단위로

표 1. 불규칙 시계열 예시

| 시간(초) | 원본 |
|-------|------|
| 0.7 | 2.00 |
| 1.6 | 3.00 |
| 2.2 | 1.00 |
| 2.8 | 2.00 |
| 4.0 | 1.00 |
| 6.3 | 4.00 |
| 7.0 | 3.00 |

표 2. 기존 보정 기법 적용 예시

| 시간(초) | (1단계) | (2단계) | (3단계) |
|-------|-------|-------|-------|
| 0.7 | 2.00 | 2.00 | |
| 1.0 | NaN | 2.33 | 2.33 |
| 1.6 | 3.00 | 3.00 | |
| 2.0 | NaN | 1.67 | 1.67 |
| 2.2 | 1.00 | 1.00 | |
| 2.8 | 2.00 | 2.00 | |
| 3.0 | NaN | 1.83 | 1.83 |
| 4.0 | 1.00 | 1.00 | 1.00 |
| 5.0 | NaN | 2.30 | 2.30 |
| 6.0 | NaN | 3.61 | 3.61 |
| 6.3 | 4.00 | 4.00 | |
| 7.0 | 3.00 | 3.00 | 3.00 |

표 3. 제안된 보정 기법 적용 예시 (단순이동평균 윈도우 크기 2초)

| 시간(초) | (1단계) | (2단계) | (3단계) | (4단계) |
|-------|-------|-------|-------|-------|
| 0.7 | 2.00 | 2.00 | 2.00 | |
| 1.0 | NaN | 2.00 | 2.00 | 2.00 |
| 1.6 | 3.00 | 2.50 | 2.50 | |
| 2.0 | NaN | 2.50 | 2.50 | 2.50 |
| 2.2 | 1.00 | 2.00 | 2.00 | |
| 2.8 | 2.00 | 2.00 | 2.00 | |
| 3.0 | NaN | 2.00 | 2.00 | 2.00 |
| 4.0 | 1.00 | 1.33 | 1.33 | 1.33 |
| 5.0 | NaN | 1.00 | 1.00 | 1.00 |
| 6.0 | NaN | NaN | 3.31 | 3.31 |
| 6.3 | 4.00 | 4.00 | 4.00 | |
| 7.0 | 3.00 | 3.50 | 3.50 | 3.50 |

보정한 결과를 얻을 수 있다.

반면, 제안된 단순이동평균을 이용한 보정 기법은 그림 1-(B)에 나타난 방식으로 진행된다. 제안된 보정 기법은 기존 보정 기법과 달리 2단계에서 시계열 데이터 전체에 단순이동평균법을 적용하여 시계열 데이터의 분산을 줄이고 추세로 일반화하고자 한다. 그러면 1단계에서 생겨난 결측치를 포함한 모든 값이 해당 시점에서부터 윈도우 크기만큼 떨어진 과거까지 구간의 평균값으로 대체된다. 다만, 윈도우 크기 이상의 구간 동안 기록된 값이 하나도 없다면 해당 시점에는 결측치가 남게 된다.

따라서 남아있는 결측치를 보간하는 단계가 추가로 필요하다. 2단계에서 결측치가 남는 경우는 많지 않아서 3단계에서 보간에 많은 연산이 필요하지는 않으므로 제안된 보정 기법이 기존 보정 기법보다 많은 단계로 이뤄진다고 해서 연산에 소요되는 시간이 더 긴 것은 아니다.

표 1의 시계열 데이터에 제안된 보정 기법을 적용하여 시간 간격이 1초로 균일한 시계열 데이터로 보정한 예시는 표 3에 나타내었으며, 1단계는 기존 보정 기법과 동일하다. 2단계에서는 1단계에서 만들어진 시계열 데이터에 단순이동평균법을 적용한다. 예시에서 단순이동평균의 윈도우 크기는 2초로 하였다. 그러면 대부분의 결측치가 사라지지만 6.0초에서는 과거 2초 전부터 기록된 값이 하나도 없기 때문에 결측치가 남게 된다. 따라서 3단계에서 해당 결측치를 선형 보간법으로 보간하였다. 4단계에서 1초 단위 인덱스에 있는 값만 가져오면 원본 데이터를 1초 단위로 보정한 결과를 얻을 수 있다. 그림 2는 앞선 예시에서 두 보정 기법 적용한 결과로 도출된 시계열 데이터를 시각화한 것이다.

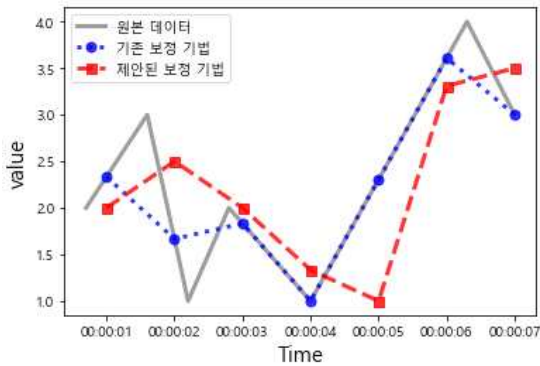


그림 2. 불규칙 샘플링 시계열 데이터 보정 결과 비교

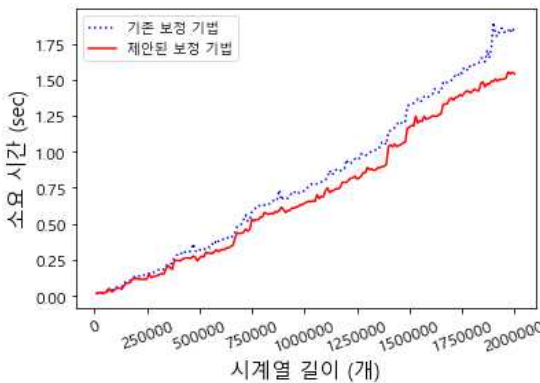


그림 3. 시계열 길이에 따른 불규칙 시계열 보정 기법 연산 소요 시간

기존 보정 기법과 제안된 보정 기법을 적용할 때 시계열의 길이에 따른 열(Column) 당 연산 소요 시간을 그림 3에 나타내었다. 가로축은 시계열의 길이이고, 세로축은 열 당 소요 시간이다.

단순이동평균법이 선형 보간법보다 연산이 간단하기 때문에 제안된 보정 기법이 기존 보정 기법에 비해 한 단계가 많음에도 불구하고 전처리 속도가 더 빠름을 알 수 있다.

III. 제안된 보정 기법 성능 평가

생산 설비의 이상 탐지를 위한 제안된 보정 기법의 효과를 평가하기 위해 기존 보정 기법과 제안된 보정 기법을 생산 설비로부터 수집한 실제 제조 데이터에 적용하였다. 성능 평가를 위해 CNC 설비로부터 취득된 전력 데이터와 스피들 진동 데이터, 사용된 톨 종류, 톨 불량 여부를 사용하였고, 단순이동평균 윈도우 크기는 10초로 설정하였다.

그림 4는 생산 설비로부터 수집한 전력 데이터와 스피들 진동 데이터의 원본 일부를 시각화한 것이다. 이 원본 데이터에 기존 보정 기법과 제안된 보정 기법을 적용하면 각각 그림 5, 6과 같은 형태로 보정된다.

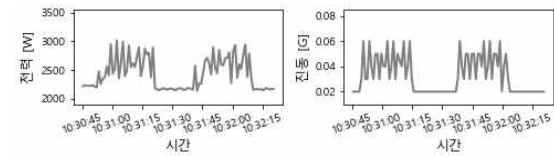


그림 4. 원본 제조 데이터

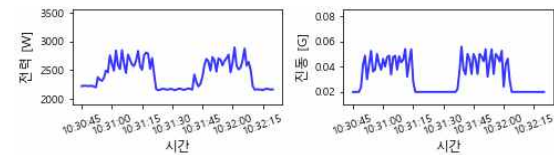


그림 5. 기존 보정 기법을 적용한 제조 데이터

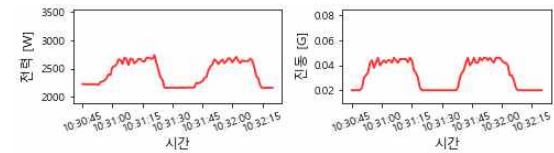


그림 6. 제안된 보정 기법을 적용한 제조 데이터

두 보정 기법으로 보정한 제조 데이터를 각각 그림 7과 같은 간단한 딥러닝 모델로 학습하여 모델의 성능을 평가하였다. 전력 데이터와 스피들 진동 데이터를 60초 단위로 분할하여 입력 데이터로 사용하였으며, 사용된 톨의 불량 여부를 탐지하도록 학습하였다.

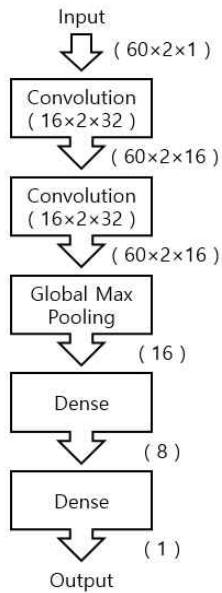


그림 7. 이상 탐지 딥러닝 모델

랜덤 변수에 따른 이상 탐지 모델의 성능 차이를 줄이기 위해 각 보정 기법 별로 실험을 100회씩 반복하였다. 데이터를 전처리한 후에 각 실험을 시행했으며, 새로운 랜덤 시드를 설정하는 것부터 딥러닝 모델을 생성하여 훈련 및 평가하는 것까지를 1회의 실험으로 진행하였다.

각 보정 기법 별로 100회씩 실험하여 도출한 이상 탐지 성능의 평균을 표 4에 나타내었다. 제안된 보정 기법으로 전처리하였을 때가 기존 보정 기법에 비해 F1-score가 약 2.7% 향상되었음을 확인할 수 있다.

표 4. 불규칙 시계열 보정 기법에 따른 이상 탐지 모델 성능 평가

| | Precision | Recall | F1-score |
|-----------|-----------|--------|--------------|
| 기존 보정 기법 | 0.864 | 0.919 | 0.889 |
| 제안된 보정 기법 | 0.897 | 0.939 | 0.916 |

IV. 결 론

본 논문에서는 생산 설비에서 수집한 불규칙 시계열의 시간 간격을 보정하는 방법으로 단순이동 평균법을 이용한 기법을 제안하고, 기존에 널리 사용되는 선형 보간법을 이용한 기법과 비교하였다.

불규칙 시계열의 특징을 가지는 제조 데이터를 전처리할 때, 제안된 보정 기법이 기존 보정 기법

에 비해 전처리 속도가 빠르고 이상 탐지 모델의 성능 향상에 효과가 있음을 확인하였다.

본 연구에서는 CNC 제조 데이터에 제안된 보정 기법을 적용해 간단한 딥러닝 모델을 학습시켜 효과를 확인하였다. 향후에는 제안된 기법을 다양한 제조 시계열과 학습 모델에 적용하여 성능을 확인할 것이다.

Acknowledgement

본 과제(결과물)는 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 사회맞춤형 산학협력 선도대학(LINC+) 육성사업의 연구결과입니다.

References

- [1] K. S. You, and Y. M. Moon, "Trends in the Development of Smart Factory Facility Diagnosis and Predictive Maintenance Technology," *The Journal of The Korean Institute of Communication Sciences*, Vol. 37, No. 7, pp. 36-42, Jun. 2020.
- [2] B. G. Goh, and J. G. Baek, "Anomaly Detection with Variational Autoencoder to Prevent System Malfunctions," *Journal of the Korean Institute of Industrial Engineers*, Vol. 45, No. 2, pp. 138-145, Apr. 2019.
- [3] K. Rehfeld, N. Marwan, J. Heitzig, and J. Kurths, "Comparison of correlation analysis techniques for irregularly sampled time series," *Nonlinear Processes in Geophysics*, Vol. 18, pp. 389-404, Jun. 2011.
- [4] Python regularise irregular time series with linear interpolation [Internet]. Available : <https://stackoverflow.com/questions/25234941/python-regularise-irregular-time-series-with-linear-interpolation/>.
- [5] W. R. Daasch and R. Madge, "Variance reduction and outliers: statistical analysis of semiconductor test data," in *IEEE International Conference on Test*, Austin: TX, pp. 1-9, Nov. 2005.