

# 매트릭스 프로파일을 이용한 제조 시계열 데이터 패턴 추출

김태현 · 진교홍\*

창원대학교

## Pattern Extraction of Manufacturing Time Series Data Using Matrix Profile

Tae-hyun Kim · Kyo-hong Jin\*

Changwon National University

E-mail : 20227259@gs.cwnu.ac.kr / khjin@changwon.ac.kr

### 요 약

제조업에서 생산 설비의 상태를 모니터링하기 위해 각종 센서를 부착하고 있으며, 이를 통해 획득된 데이터의 경우 시계열 데이터인 경우가 많다. 생산 설비의 이상 여부를 판단하기 위해서는 시계열 데이터로부터 패턴을 추출하는 과정이 선행되어야 하며 다양한 방법이 연구되고 있다. 본 논문에서는 수집된 다변량 시계열 데이터로부터 패턴을 추출하기 위해 매트릭스 프로파일 알고리즘을 적용하였으며, 이를 통해 현재 CNC 머신으로부터 수집 중인 다중 센서 데이터의 패턴을 추출하였다.

### ABSTRACT

In the manufacturing industry, various sensors are attached to monitor the status of production facility. In many cases, the data obtained through these sensors is time series data. In order to determine whether the status of the production facility is abnormal, the process of extracting patterns from time series data must be preceded. Also various methods for extracting patterns from time series data are studied. In this paper, we use matrix profile algorithm to extract patterns from the collected multivariate time series data. Through this, the pattern of multi sensor data currently being collected from the CNC machine is extracted.

### 키워드

Pattern Extraction, Manufacturing Time Series Data, Matrix Profile, Anomaly Detection

### I. 서 론

스마트 공장의 보급에 따라 제조업체는 설비의 이상 탐지, 예지보전을 목적으로 설비에 각종 센서를 부착하여 데이터를 수집하고 있다[1]. 제조 공정은 설비를 가동하는 동안 끊임없이 동작하기 때문에 센서 데이터는 시계열 데이터 형태로 수집되는 경우가 많다. 한편, 제조 현장에서 생산 설비의 가동이 중단되거나 이상이 발생하게 되면 제품의 생산에 영향을 주어 기업에 막대한 손해가 발생할 수 있으므로 획득된 시계열 데이터로부터 패턴을 추출하고 정상 패턴과 이상 패턴을 구분하여 생산 설비의 이상 여부를 판단하는 방법이 연구되고 있다[2].

기존의 패턴 추출 방법으로는 매트릭스 프로파일(Matrix Profile), DTW(Dynamic time warping)[3], RPM(Representative Pattern Mining)[4] 등이 있다.

본 논문에서는 임의 길이의 시계열 데이터에서 모든 모티프(Motif)를 발견할 수 있는 매트릭스 프로파일 알고리즘[5]을 사용하여 CNC 머신으로부터 수집된 다변량 시계열 데이터에서 반복되는 패턴인 모티프를 찾아보고자 한다. 패턴추출시 사용되는 거리함수로는 민코프스키 거리(Minkowski distance)[6]를 활용하였다.

본 논문의 구성은 다음과 같다. 2장에서 매트릭스 프로파일 알고리즘과 민코프스키 거리를 통한 패턴 추출 과정을 설명하고, 3장에서는 패턴 추출 결과를 확인하였다. 마지막으로 4장에서 결론을 기술하였다.

\* corresponding author

## II. 매트릭스 프로파일을 이용한 데이터 패턴 추출

### (1) CNC 제조 데이터

CNC 머신에 부착된 센서로부터 0.1초 단위로 톨의 사용량, 스피들 부하, 스피들 속도, 서보모터 부하 등의 데이터를 수집하였다. 이 중 일정한 범위 안에서 값이 반복되는 스피들 부하와 서보모터 부하 데이터를 사용하였다.

### (2) 매트릭스 프로파일

매트릭스 프로파일은 시계열 내의 모든 서브시퀀스 간 Z-정규화된 유클리드 거리(Z-Normalized Euclidean Distance)를 사용하여 유사성을 탐색하는 알고리즘이다. 전체 길이가 n인 시계열 T에서 길이가 m인 서브시퀀스를 선택하여 윈도우를 하나씩 이동하며 n-m+1개의 서브시퀀스를 얻는다. 모든 서브시퀀스에 대해서 유클리드 거리를 구하여 각 위치에서 계산된 거리의 최솟값만을 저장한 결과인 매트릭스 프로파일을 만든다. 이 중 가장 작은 값을 가지는 서브시퀀스를 시계열의 모티프라고 한다.

### (3) 패턴 추출 과정

CNC 머신의 공정시간이 일정하지 않아 공정시간의 평균값 120을 서브시퀀스의 길이 m으로 설정하고 매트릭스 프로파일 알고리즘을 수행하여 다변량 시계열 데이터에서 모티프를 발견하였다. 그 후 모티프와 데이터 간 민코프스키 거리를 계산하여 거리 행렬을 만들었다. 두 데이터 모두 거리 행렬의 값이 11을 초과하는 경우 패턴의 형태가 크게 달라졌다. 따라서 유효한 패턴이 아닌 것으로 간주하고 임계값을 11로 설정하여 패턴을 추출하였다. 그림 1은 스피들 부하 데이터에 위 과정을 적용한 결과이고, 그림 2는 서보모터 부하 데이터에 적용한 결과이다.

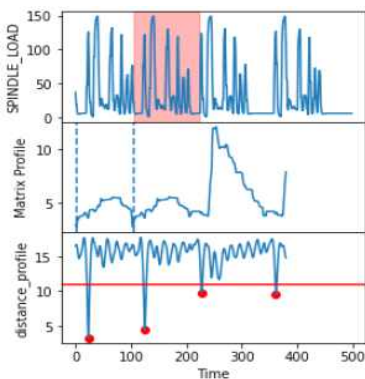


그림 1. 스피들 부하 데이터의 거리 행렬 생성 결과

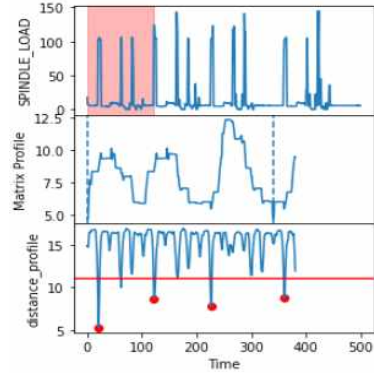


그림 2. 서보모터 부하 데이터의 거리 행렬 생성 결과

## III. 패턴 추출 결과

거리 행렬 각 값의 인덱스로 패턴을 추출한 결과 총 3가지 종류의 패턴이 추출되는 것을 확인하였으며 추출한 패턴의 수를 표 1에 나타내었다.

표 1. 데이터별 추출 패턴 수

데이터	패턴 1	패턴 2	패턴 3
스피들 부하	172	6	2
서보모터 부하	172	5	12

패턴 1은 반복되는 유사한 형태의 패턴이고, 패턴 2는 유사도가 낮은 형태의 패턴이다. 패턴 3은 임계값이 초과된 거리 행렬의 패턴(주기성이 없는 패턴)을 의미한다.

그림 3에서 스피들 부하 데이터에서 추출한 패턴 1의 대표 패턴과 패턴 2의 일부를 나타내었다. 그림 4는 동일 데이터의 패턴 3 결과이고, 패턴 3의 인덱스를 그림5에 표시하였다.

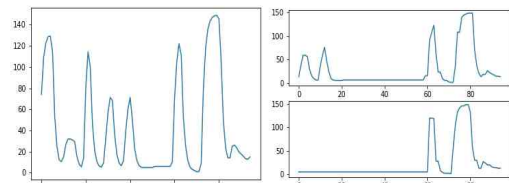


그림 3. 스피들 부하 데이터 패턴 1과 패턴 2

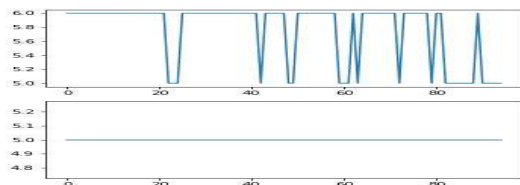


그림 4. 스피들 부하 데이터 패턴 3

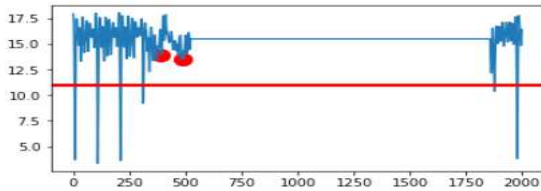


그림 5. 스핀들 부하 데이터 패턴 3의 인덱스

그림 6, 7, 8은 위와 동일한 과정을 서보모터 부하 데이터에 적용한 결과이다.

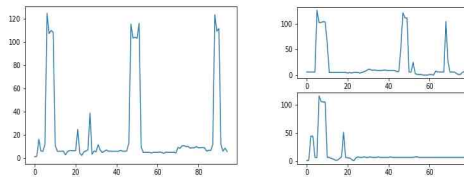


그림 6. 서보모터 부하 데이터 패턴 1과 패턴 2

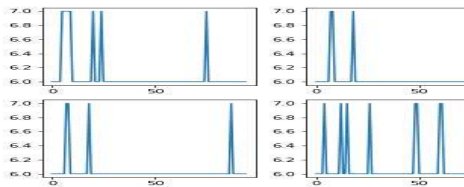


그림 7. 서보모터 부하 데이터 패턴 3

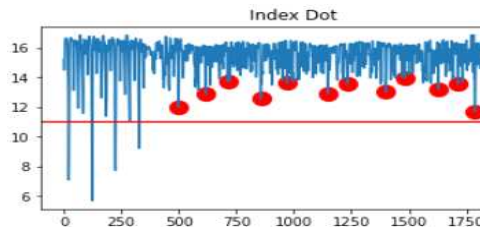


그림 8. 서보모터 부하 데이터 패턴 3의 인덱스

#### IV. 결론

본 논문에서는 매트릭스 프로파일 알고리즘과 민코프스키 거리를 사용하여 CNC 머신으로부터 수집한 다변량 시계열 데이터에서 패턴을 추출하였고 총 세 종류의 패턴이 있는 것을 확인하였다.

향후에는 추출된 패턴을 딥러닝 모델에 적용하여 이상 탐지 모델을 구축할 계획이다.

#### Acknowledgement

본 연구는 2022년도 중소벤처기업부의 기술개발 사업 지원에 의한 연구임 [S3266312]

#### References

- [1] D. H. Hu, W. H. Jeong, and K. C. Im, "The Study on CNC Facility Predictive Maintenance System Based on Manufacturing Data," *The Society of Convergence Knowledge Transactions* Vol. 9, No. 4, pp. 179-187, 2021.
- [2] M. Bai, J. Liu, J. Chai, X. Zhao and D. Yu, "Anomaly detection of gas turbines based on normal pattern extraction," *Applied Thermal Engineering*, Vol. 166, Feb. 2016.
- [3] SAKOE, Hiroaki; CHIBA, Seibi. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26.1: 43-49, 1978.
- [4] X. Wang, J. Lin, P. Senin, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, and S. Frankenstein, "RPM: Representative Pattern Mining for Efficient Time Series Classification.," in *EDBT*, pp. 185-196, Mar. 2016.
- [5] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen and E. Keogh, "Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets," in *2016 IEEE 16th international conference on data mining (ICDM)*, pp. 1317-1322, 2016.
- [6] A. Singh, A. Yadav, and A. Rana, "K-means with Three different Distance Metrics," *International Journal of Computer Applications*, Vol. 67, No. 10, Apr. 2013.