

고속 푸리에 변환(FFT) 기반 주기 추출 및 윈도우 구성을 활용한 GELU CNN-GRU AE 모델의 산업 전력 시계열 이상치 탐지

박범도¹, 권우현¹, 이충호², 허태욱², *이상금¹

*국립한밭대학교¹, 한국전자통신연구원²

{pbeomdo, mfireon0520}@gmail.com, {leech, htw398}@etri.re.kr, sangkeum@hanbat.ac.kr

FFT-Based Periodicity Extraction and Window Configuration with a GELU
CNN-GRU AE Model for Anomaly Detection in Industrial Power Time-Series

Beomdo Park¹, Woohyeon Kwon¹, ChungHo Lee², TaeWook Heo², and *Sangkeum Lee¹

*Hanbat National University¹, Electronics and Telecommunications Research Institute²

Abstract

본 연구는 산업 현장에서 수집된 전력 데이터의 이상 패턴을 탐지하기 위한 접근 방식을 제안한다. 원본 데이터의 비정상성을 완화하며 시계열 변화를 강조하기 위해 차분을 적용한다. 이후, 데이터에 FFT(Fast Fourier Transform)를 적용하여 차분을 적용하지 않은 데이터보다 이상 상황에 대한 분해능을 높여 주요 주파수 영역의 상위 5개의 주기 성분을 식별한다. GELU(Gaussian Error Linear Unit) 활성화 함수를 활용하고 CNN(Convolutional Neural Network)과 GRU(Gated Recurrent Unit) 기반 AE(Autoencoder)를 결합 모델을 제안한다. 입력과 재구성된 윈도우 간 오차는 시간적 왜곡에 강인한 DTW(Dynamic Time Warping)로 계산한 재구성 오차를 기반으로 Isolation Forest 알고리즘으로 최종 이상치를 판단한다. 방법론은 산업 전력 데이터의 이상 상황을 정밀하게 포착함으로써, 스마트 팩토리 구축을 위한 기반 기술로 활용될 수 있으며, 실시간 모니터링 및 자동화된 이상 대응 체계 개발에 기여한다.

I. 서론

산업 설비의 예지 보전(Predictive Maintenance)과, 에너지 효율화 및 시스템 안정성 확보를 위해 대규모 시계열 전력 데이터에서 이상 패턴을 효과적으로 탐지하는 것이 중요한 과제이다. 전력 사용 패턴의 이상 징후를 조기에 감지하면 예기치 못한 설비 고장 등으로 인한 생산 손실의 예방이 가능하다[1]. 산업 전력 데이터의 이상 탐지는 방대한 규모, 복잡한 패턴, 비정상성 등으로 분석에 어려움이 따른다. 전통적 통계 기반 방식은 한계가 있으며, 딥러닝 기반 AE(Autoencoder)가 효과적인 대안으로 부상하고 있다[2]. 특히 RNN(Recurrent Neural Network) 계열인 LSTM(Long Short-Term Memory)[3] 또는 GRU(Gated Recurrent Unit)[4] 모델이 활용되고 있다.

본 논문에서는 산업 전력 시계열의 특성을 고려하여, 데이터의 추세 제거와 변화 강화를 위해 전체 시계열에 1차 차분을 적용하고 FFT(Fast Fourier Transform)를 적용하여 급격한 변화를 나타내는 주요 주파수 영역의 주기 성분을 추출한다. 각 주기 성분을 윈도우 크기로 설정하고 GELU(Gaussian Error Linear Unit) CNN(Convolutional Neural Network)-GRU AE 모델에 입력되어 주기별 패턴을 학습하며, DTW(Dynamic Time Warping) 기반 재구성 오차와

Isolation Forest를 결합한 이상치 탐지 방법론을 제안한다.

II. 본론

2.1. FFT 기반 주기 추출 및 윈도우 구성

전체 시계열 데이터의 주기적 특성을 분석하고 다양한 윈도우 크기 선정에 활용하기 위해 FFT를 사용한다. 전력 데이터의 이상치는 갑작스러운 사용량 변화와 같은 단기적이고 급격한 패턴으로 나타나는데, 이는 주파수 영역에서 진폭이 증가하는 피크(peak)와 같은 형태로 표현된다. 원본 데이터의 FFT 분석 결과(그림 1)와 비교하여, 1차 차분을 적용한 데이터의 FFT 분석 결과(그림 2)는 급격한 변화 패턴을 나타내는 주파수 영역에 더욱 집중되어 더 작은 주기 영역에서의 분해능이 향상된다.

이에 따라, 차분이 적용된 시계열의 FFT를 적용해 큰 진폭을 갖는 상위 주요 피크들을 식별한다. 이를 기반으로 주기를 계산하여 복수의 윈도우 크기를 결정하는 방식으로 다양한 시간 스케일에서의 이상 패턴 탐지 가능성을 높인다. 각 윈도우 크기에 따라, 모델 학습 및 추론을 위한 실제 입력 윈도우를 생성한다. 개별 윈도우를 평균 0, 표준편차 1로 표준화하고 모델의 최종 입력으로 사용하여, 시계열 데이터의 서로 다른 시간 주기에 걸쳐 나타나는 특징적인 패턴들을 학습한다.

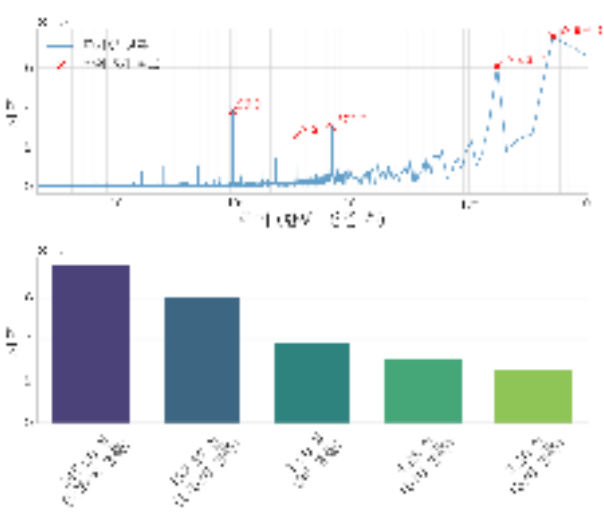


그림 1. 원본 데이터의 FFT 분석 및 피크 식별 결과

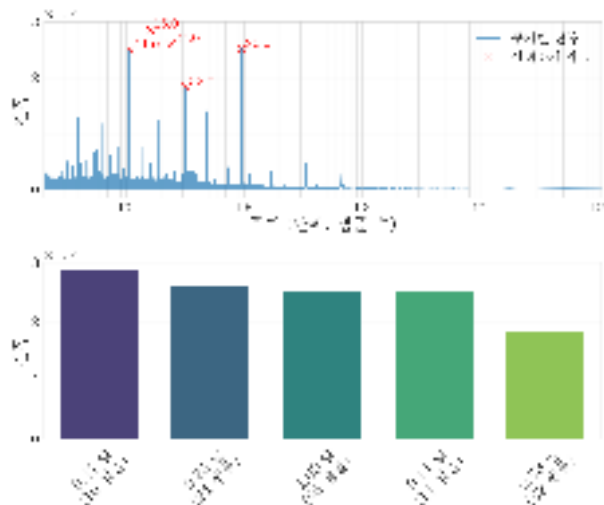


그림 2. 차분 데이터의 FFT 분석 및 피크 식별 결과

2.3 GELU CNN-GRU AE모델

구분	항목명	값 또는 설명
입력	입력 차원	1 (단변량 시계열)
	출력 채널 수	32
CNN 구조	커널 크기	7
	Stride	윈도우 크기의 25%
	활성화 함수	GELU
	Dropout 비율	0.1
GRU 구조	은닉층 차원	64
	GRU 계층 수	1
잠재 공간	차원	32
학습 설정	Epochs	50
	Batch Size	128
	Learning Rate	0.0001
	Optimizer	Adam
손실 함수	손실 함수	평균 제곱근 오차 (RMSE Loss)

표 1. GELU CNN-GRU AE 모델 주요 파라미터

입력 윈도우는 먼저 CNN 계층을 통과하여 지역적 패턴을 추출한다. CNN은 시계열 내 짧은 구간의 형태적 특징을 잡아내는 데 유용하다[5]. CNN 계층 후에는 GELU 활성화 함수와 Dropout이 적용된다. GELU는 ReLU(Rectified Linear Unit)에 비해 부드러운 비선형성을 제공하고 모든 지점에서 미분 가능하여 안정적인 학습과 입력 크기에 따라 활성화 정도를 확률적으로 조절하여 일반화 성능을 향상시킬 수 있다[6].

CNN을 통과한 특징 시퀀스는 GRU 계층의 입력으로 전달된다. GRU는 시계열 데이터의 시간적 순서 정보를 학습하여 저차원 잠재 벡터로 압축한다. GRU는 LSTM 대비 구조가 단순하고 파라미터 수가

적어 계산 효율성이 높아, 짧은 윈도우 입력에 대한 맞춤형 모델 최적화에 유리하다[4].

인코더에서 생성된 잠재 벡터는 완전 연결(FC) 계층과 GELU를 거쳐 GRU가 처리할 차원으로 확장된 후, 디코더 GRU를 통해 특징 시퀀스를 복원한다. 최종적으로 출력 FC 계층이 원본 윈도우와 동일 차원의 시퀀스로 재구성하여 출력한다.

2.4 이상치 탐지

시계열 데이터의 재구성 오차 측정 시, 일반적인 MSE 등은 시간 축의 미세 변동에 민감하다. 이를 극복하고자 DTW를 활용한다. DTW는 두 시계열 간 비선형적 정렬을 통해 시간 축 뒤틀림이나 속도 차이에 강인하게 형태적 유사성을 측정하므로, 단순 시간 변동과 실제 이상 패턴을 정밀하게 구분할 수 있다[7]. 각 윈도우에 대해 원본 시퀀스(차분 및 스케일링 적용)와 재구성된 시퀀스 간 DTW 거리를 재구성 오차로 정의한다. 재구성 오차 값들의 분포로부터 실제 이상치를 판별하기 위해 Isolation Forest 알고리즘을 적용한다. Isolation Forest는 이상치가 적은 분할로 쉽게 고립되는 특성을 이용하며, 비지도 학습 방식으로 데이터 분포 가정이 불필요하다[8]. 학습 시 오염 비율(contamination)을 0.005로 예상 이상치 비율의 민감도를 조절한다. Isolation Forest로 특정 윈도우가 이상으로 판별되면, 해당 윈도우 내부에서 보다 정밀한 이상 시점 식별을 수행한다. 구체적으로, 각 시점의 입력 값과 재구성된 값 사이의 절대 오차를 계산한다. 이 계산된 절대 오차가, 해당 윈도우의 기간에 대한 평균 제곱근 오차(RMSE)의 2.5배를 초과하는 경우, 해당 시점(인덱스)을 최종적인 이상 발생 지점으로 확정한다. 이렇게 식별된 모든 이상 시점의 인덱스는 원본 시계열의 개별 시간 인덱스로 변환하여 기록한다.

III. 결과

3.1 데이터셋

본 연구에서 사용된 데이터는 3년간 15분 단위의 특정 산업 현장(1개 기업)에서 수집된 전력 사용량 시계열 데이터(2020~2022년)이다. 데이터는 다양한 정상 운영 패턴과 간헐적인 이상 상황을 포함하고 있어 원시 시계열에 대한 전처리가 필요하다. 해당 데이터 중 0을 결측치(NaN)로 변환하고 KNN(K-Nearest Neighbors) Imputer를 사용하여 주변 데이터 패턴을 고려한 결측치 보간(Interpolation) 기법으로 데이터의 연속성을 확보한다.

FFT를 적용해 주요 주파수 성분을 계산하고, scipy.signal.find_peaks를 이용해 큰 진폭을 나타내는

피크를 추출하여 이를 통해 추출된 상위 5개의 주요 주기는 [11, 16, 24, 32, 96](15분 단위의 샘플 수 기준)이며, 각각 모델의 입력 윈도우 크기로 설정한다.

3.2 모델 학습 및 이상치 탐지

윈도우 크기별로 독립적인 GELU CNN-GRU AE모델을 개별적으로 학습시키고 이상치 탐지를 수행한다. 학습 이후, 각 윈도우 크기에 대응하는 모델을 사용하여 전체 데이터 기간에 걸친 DTW를 계산하여 재구성 오차로 사용한다. 재구성 오차 값들은 해당 시간 윈도우의 이상 정도를 나타내는 점수가 된다. 이 오차 점수 분포에 Isolation Forest를 적용하여 이상 가능성이 높은 윈도우들을 식별한다.

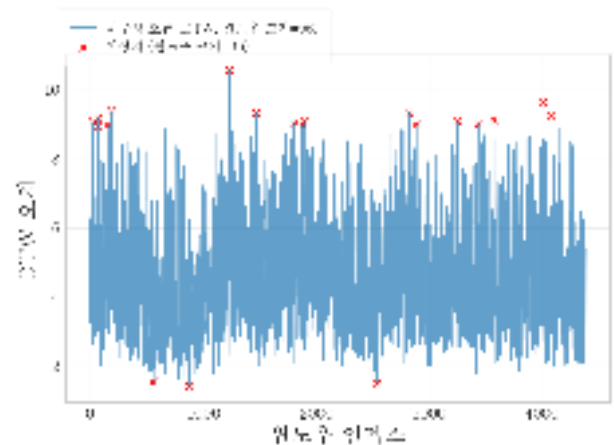


그림 3. (윈도우 크기 96) DTW 재구성 오차 분포

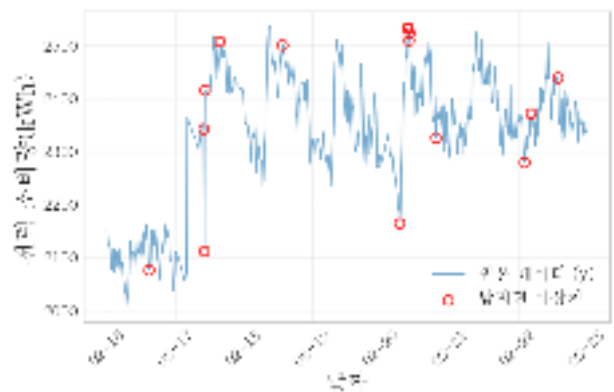


그림 4. (2020년 2월 16일 ~ 2020년 2월 22일)의 이상치 검출 결과

그림 3은 윈도우 크기 96(=1일)에서의 DTW 오차 분포와 Isolation Forest에 의해 탐지된 이상치 윈도우의 인덱스 예시이다. 최종적인 이상치 탐지 결과는 모든 개별 윈도우 크기에서 탐지된 이상 정보를 구성한다. 각 윈도우 크기에 대응하는 모델의 Isolation Forest에 의해 이상으로 판별된 윈도우 내에서, 최종

이상 발생 시점을 특정하고 인덱스를 통한다.

그림 4는 전체 입력 데이터 중에서 이를 통해 검출된 이상치가 가장 많은 1주일 간의 데이터 구간 시퀀스(2020년 2월 16일 ~ 2020년 2월 22일)의 개별 이상치 인덱스에 따른 시각화이다. 분석 결과, 제안된 방법론은 전력 사용량이 평소의 주기적 패턴에서 급격하게 벗어나는 순간적인 피크 또는 강하, 비정상적인 변동성을 보이는 구간, 또는 예상치 못한 수준의 전력 소비가 지속되는 구간 등 다양한 형태의 이상치를 탐지한다. 이는 실제 산업 현장에서 발생할 수 있는 특정 설비의 갑작스러운 가동 중단, 예기치 않은 피크 부하 발생, 공정 라인의 비정상 가동, 또는 장기간에 걸친 미세한 에너지 누수나 설비 효율 저하 등 다양한 이상 상황에 해당될 수 있다. 특히, FFT 분석을 통해 데이터 특성에 기반하여 선정된 여러 윈도우 크기를 종합적으로 활용함으로써, 짧은 시간 스케일의 급격한 이상치 뿐만 아니라 상대적으로 긴 시간 스케일에서 나타나는 미묘한 이상 패턴까지 탐지할 수 있는 가능성을 보인다. 이러한 결과는 산업 현장의 잠재적 위험 요소를 조기에 감지하고 대응하는데 중요한 정보를 제공할 수 있다.

IV. 결론

본 연구는 산업 전력 시계열 이상치 탐지를 위해 1차 차분, 차분 데이터 기반 FFT, GELU CNN-GRU AE 모델, DTW 기반 재구성 오차, Isolation Forest를 결합한 다단계 접근 방식을 제안한다. 1차 차분은 비정상성 완화 및 변화 강조, FFT는 급격한 변화에 해당하는 주기 성분 추출 및 분석 윈도우 크기를 설정하고 모델의 윈도우 입력에 기여한다. DTW는 시간 왜곡에 강인한 재구성 오차 측정을, Isolation Forest는 효율적인 이상치 판별을 가능하게 한다. 이러한 요소들의 결합은 산업 현장의 기업별, 용도별 이상 상황을 정교하고 강건하게 탐지할 가능성을 제시한다. 향후 다양한 산업 데이터셋 검증 및 탐지된 이상 원인 분석, 예측 모델 연동을 통해 예지 보전 시스템의 실효성을 높이는 연구가 필요하다.

참고문헌

[1] Henning, S., Hasselbring, W., Burmester, H., Möbius, A., & Wojcieszak, M. (2021). Goals and measures for analyzing power consumption data in manufacturing enterprises. *Journal of Data, Information and Management*, 3(1), 65-82.

[2] Guo, Y., Liao, W., Wang, Q., Yu, L., Ji, T., & Li, P.

(2018, November). Multidimensional time series anomaly detection: A gru-based gaussian mixture variational autoencoder approach. In *Asian conference on machine learning* (pp. 97-112). PMLR.

[3] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.

[4] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

[5] Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., ... & Qiao, H. (2018, April). Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In *Proceedings of the 2018 world wide web conference* (pp. 187-196).

[6] Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.

[7] Sakoe, H., & Chiba, S. (2003). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1), 43-49.

[8] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In *2008 eighth IEEE international conference on data mining* (pp. 413-422). IEEE.