

2025학년도 소중한 JUMP-UP Labs 결과보고서(공개용)

과제 현황					
과제명	AI(인공지능)를 활용한 산업체 전력 사용량 이상치 탐지				
팀명	ECO AI LAB				
과제 수행기간	2025년 4월 ~ 2025년 11월				
Github주소 (해당시)	https://github.com/HBNU-SWUNIV/jum-up-labs25-ecoai-2025				
지도교수					
소속	국립한밭대학교 컴퓨터공학과				
성명	이상금				
전화					
E-mail					
참가인원(팀원)					
No.	성명	소속학과	학번	전화번호	E-mail
1	이지상	컴퓨터공학과			
2	김준연	컴퓨터공학과			
3	박범도	컴퓨터공학과			
4	장현석	컴퓨터공학과			
5	박준성	컴퓨터공학과			
6	권우현	컴퓨터공학과			
7	정민성	컴퓨터공학과			
8	유동건	컴퓨터공학과			
9	이정섭	컴퓨터공학과			

과제 세부 설명

1. 개발 동기 및 목적, 필요성

전 세계적으로 탄소 중립이 주요 과제로 대두됨에 따라, 대한민국 역시 2050년까지 탄소 중립 실현을 목표로 하고 있다. 이산화탄소 배출의 상당 부분은 화석 연료와 산업 활동에서 비롯되며, 이에 대한 해결책으로 산업체의 전력 데이터를 기반으로 최적의 솔루션을 제공하는 '그린버튼(Green Button)' 플랫폼이 주목받고 있다. 하지만 산업 공정 데이터는 다양한 환경 요인과 비정상적 운영으로 인해 신뢰성을 저해하는 이상치(anomaly)를 포함할 가능성이 높다. 본 연구는 이 산업체 전력 소비 데이터를 분석하고 이상치를 효율적으로 탐지하여 그린버튼 플랫폼의 신뢰성과 정확성을 높이는 방안을 제시하는 것을 목적으로 한다.



2. 과제 수행 내용

본 과제는 산업 데이터의 비정상성(Non-stationarity)과 클래스 불균형 문제를 해결하기 위해 다음과 같은 다각도의 접근 방식을 수행했다.

- GELU CNN-GRU AE 기반 이상치 탐지 모델 개발: 시계열의 지역적 특징(CNN)과 시간적 흐름(GRU)을 동시에 학습하는 오토인코더(Autoencoder) 모델을 설계했다. 활성화 함수로는 GELU를 사용하여 비선형성을 강화했으며, 재구성 오차 측정 시 시간 왜곡에 강인한 DTW(Dynamic Time Warping)를 적용하고 Isolation Forest로 최종 이상치를 판별하는 하이브리드 방법론을 제안했다.
- 데이터 해상도에 따른 모델 최적화 연구: 데이터의 수집 주기(시간 해상도)가 모델 성능에 미치는 영향을 분석했다. 고해상도 데이터에서는 LSTM Autoencoder가, 중-저해상도 데이터에서는 적대적 학습(Adversarial Training) 기반의 USAD(Unsupervised Anomaly Detection) 모델이 더 우수한 성능을 보임을 규명했다.
- MI-FGSM 기반 데이터 증강 (Data Augmentation): 이상치 데이터가 극히 부족한 화학 산업 데이터의 클래스 불균형 문제를 해결하기 위해, 정상 데이터에 미세한 노이즈를 주입하여 적대적 샘플을 생성하는 MI-FGSM(Momentum Iterative-Fast Gradient Sign Method) 기법을 적용했다. 이를 통해 탐지 모델의 F1 Score를 비약적으로 향상시키는 임계값을 도출했다.

<p>3. 과제 수행 결과</p>	<p>본 연구를 통해 도출된 정량적 및 정성적 성과는 다음과 같다.</p> <ul style="list-style-type: none"> ○ 산업 특화 이상치 탐지 성능 확보: 화학 산업: MI-FGSM 증강 기법을 LSTM-AE에 적용하여, 최적 노이즈 임계값에서 F1 Score 0.9778을 달성하며 클래스 불균형 문제를 효과적으로 해결했다. 산업 전반: GELU CNN-GRU AE 모델을 통해 급격한 피크 부하와 미세한 에너지 누수 패턴을 모두 탐지하는 데 성공했으며, DTW 적용으로 단순 시간 밀림에 의한 오탐지를 방지했다. ○ 해상도별 모델 선정 가이드라인 제시: 실제 산업 현장의 중·저해상도 데이터 환경에서는 USAD 모델이 미세한 이상 변동을 더 민감하게 포착함을 입증하여, 현장 데이터 특성에 맞는 모델 선택 기준을 수립했다. ○ 학술적 성과: 본 과제 수행 결과를 바탕으로 총 3편의 학술대회 논문을 발표하였으며, 이를 통해 제안 방법론의 독창성과 우수성을 검증받았다. FFT 기반 주기 추출 및 GELU CNN-GRU AE 모델 (박범도 등) LSTM AE와 USAD의 해상도별 성능 분석 (박준성 등) MI-FGSM 데이터 증강 기법 (장현석 등)
<p>4. 팀원 역할 분담</p>	<ul style="list-style-type: none"> • 박범도(모델 개발) FFT 기반 전처리 및 GELU CNN-GRU AE 모델 설계 및 구현 • 권우현(모델 개발) CNN-GRU 하이퍼파라미터 튜닝 및 Isolation Forest 연동 구현 • 장현석(데이터 증강) MI-FGSM 알고리즘 구현 및 클래스 불균형 해소 실험 수행 • 정민성(데이터 분석) 적대적 샘플 생성 실험 및 ϵ 임계값에 따른 성능 변화 분석 • 박준성(성능 비교 연구) LSTM-AE 및 USAD 모델 구현, 데이터 해상도별 비교 실험 • 이정섭(데이터 전처리) LEAD1.0, KPI 등 공개 데이터셋 전처리 및 정규화 수행 • 유동건(실험 지원) 모델 성능 지표(F1, Precision, Recall) 측정 및 결과 시각화 • 이지상, 김준연(연구 자문) 최신 이상치 탐지 동향 조사, 실험 설계 검토 및 논문 작성 지도

5. 결과물 사진
또는 동작 화면

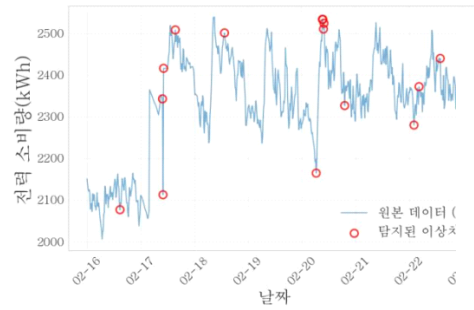


그림 4. (2020년 2월 16일 ~ 2020년 2월 22일)
이상치 검출 결과

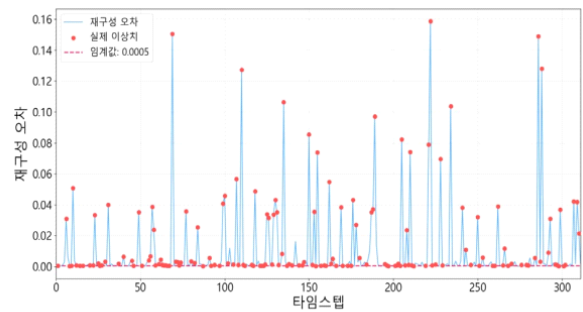


그림 2 이상치 탐지 재구성 오차 분포($\epsilon = 0.01$)

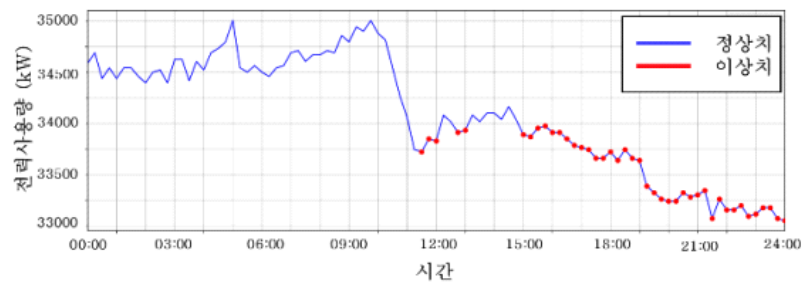


그림 3. 실제 산업 전력 데이터셋의 전력사용량

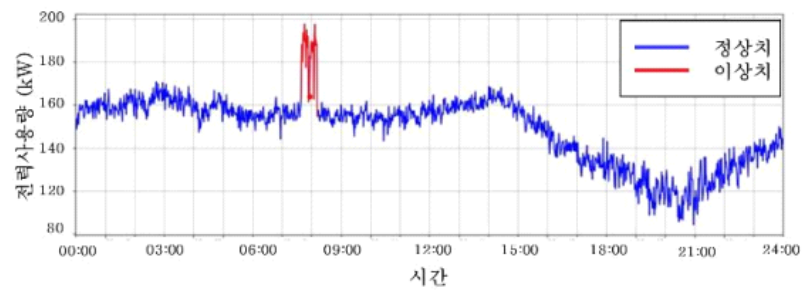



그림 4. KPI 데이터셋의 전력사용량

6. 기타

- 학술대회 발표 성과: 본 과제 수행 결과를 바탕으로 '대한전자공학회 하계종합학술대회' 등 주요 학술대회에서 총 3건의 논문을 발표하였음.
- 후속 연구 계획: 본 연구에서 검증된 모델을 기반으로, 금융(사기 탐지) 및 보안(침입 탐지) 분야로의 도메인 확장 가능성을 확인하였으며, 향후 실시간 스트리밍 데이터 처리를 위한 파이프라인 최적화 연구를 진행할 예정임.


7. 작품사진




이상치 탐지 성능 향상을 위한 MI-FGSM 데이터 증강 기법

장현석¹, 정민성¹, 이충호², 허태욱² *이상금¹

¹국립한밭대학교¹, 한국전자통신연구원²



논문 바로가기



Abstract

· (기후문제 및 이상치 탐지 필요성) 화공 산업은 국내 온실가스 배출의 상당 부분을 차지하며, 효과적인 온실가스 배출량 감축과 기후변화 대응을 위해 신뢰성 높은 전력 데이터 분석과 이상치 탐지(클래스 불균형) 데이터 분석 결과, 정상 데이터 이상치가 적은 클래스 불균형이 존재하였고 이는 탐지 결과의 신뢰성을 저해함.

· (데이터 증강 적용) 학습 데이터를 MI-FGSM에 적용해 생성된 적대적 샘플을 통해 이상치를 증강하고 두 클래스 간 균형을 조절함.

· (성능 분석) MI-FGSM 파라미터인 교란 크기 ϵ 와 모델의 탐지 성능 지표인 F1 점수 간 상관관계를 분석함.

· (ϵ 임계값 도출) ϵ 과 F1 점수의 상관관계 그래프를 통해 F1 점수의 성능이 낮아지는 ϵ 의 임계값을 도출하고 데이터 증강을 위한 MI-FGSM의 ϵ 허용 범위에 대한 가이드라인을 제시함.

실험 및 결과

그림5. Isolation Forest 이상치 분류 결과

· (데이터셋)(그림5) 2020년 1월부터; 15분 단위로 수집된 전력 데이터 중 일의 데이터를 사용함. Isolation Forest 1096개의 데이터 중 55개의 이상

그림6. 이상치 탐지 재구성 오차 분포($\epsilon = 0.5$)

그림7. 이상치 탐지 재구성 오차

· (그림6) $\epsilon = 0.5$ 일 때 MI-FGSM으로 생성된 적대적 샘플은 대부분 재구성 오차의 임계값보다 높음. 기존 이상치 대비 정상 데이터에 근사한 분포를 보이며, F1 점수는 0.9778로 높은 성능이 나옴.

· (그림7) $\epsilon = 0.01$ 일 때의 재구성 오차, 적대적 샘플의 분포가 임계값과 사성을 보임. 또한, F1 점수는 0.6907로 대역 0.2871에 감소함.

그림8. ϵ 와 F1 점수의 상관관계 그래프

· (그림8) 발간적 점에 해당하는 $\epsilon = 0.25$ 지점부터 ϵ 값이 작아짐에 따라 F1 점수가 낮아지며, 0에 수렴. 임계값, ϵ 이 크면 노이즈의 교환 범위가 넓어져 이상치 탐지가 용이하지만, ϵ 이 작아질수록 교란 데이터의 구분 경계가 모호해짐. 따라서, F1 점수가 낮아지는 시작점인 $\epsilon = 0.25$ 를 실험 데이터 정함.

결론

· (연구 목표) 화공 산업 전력 데이터의 클래스 불균형을 해소하고, LSTM-AE의 이상치 탐지 성능에 대한 신뢰성을 향상하고자 MI-FGSM 기반 데이터 증강 방법론을 제안함.

· (실험 과정) 화공 산업 내 점유율이 높은 기업 데이터를 전처리 후, $\epsilon = [0.5, 0.25, 0.1, 0.05, 0.01]$ 범위에서 F1점수를 비교함.

· (실험 결과) $\epsilon = 0.5$ 에서 0.9778의 F1 점수를 달성한 반면, ϵ 이 감소함에 따라 F1 점수가 낮아지는 경향을 보였고, $\epsilon = 0.01$ 에서는 F1 점수는 0.6907임.

· (시사점) ϵ 과 F1 점수의 상관관계를 분석한 결과, ϵ 이 임계값보다 낮아 교환 범위가 과도하게 좁을 경우 적대적 샘플과 정상 데이터 간 구분이 불분명하여 탐지 성능이 저하됨.

향후 연구



LSTM Autoencoder와 Unsupervised Anomaly Detection 모델의 시간 해상도별 이상치 탐지 성능 분석

박준성¹, 이정섭¹, 유동건¹, 이충호², 허태욱², *이상금¹

¹*국립한밭대학교¹, 한국전자통신연구원²

EcoAI Lab

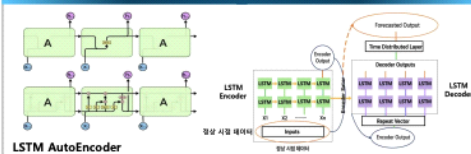
Abstract

- 본 연구는 산업 전력 데이터 내 이상치를 탐지하여 에너지 낭비를 줄이고 전력 효율을 향상시키는 것을 목표로 하며, 이를 위해 시계열 기반 비지도 학습 모델의 성능을 분석.
- LSTM(Long Short-Term Memory) Autoencoder와 USAD(Unsupervised Anomaly Detection) 모델을 적용하여, 시간 해상도에 따른 이상 탐지 성능 차이를 비교.
- 실험에는 전력해상도의 LEAD1.0, 고해상도의 KPI, 중해상도의 실제 산업 전력 데이터셋을 활용하고, F1-score, Precision, Recall 지표를 통해 정량적 성능 평가를 수행.
- USAD는 중저해상도 데이터에서 LSTM Autoencoder보다 우수한 성능을 보이며, 이는 실제 산업 환경에서의 생시 제거나 저장 한계를 고려했을 때 실용적임을 시사.
- 시계열 데이터의 해상도에 따라 모델의 이상 탐지 성능이 달라질 수 있으며, 데이터 특성에 맞는 모델 선택 기준의 필요성을 강조.

개요

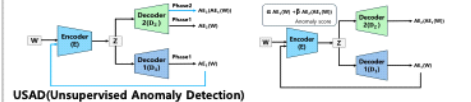


이상치 탐지 모델



LSTM AutoEncoder

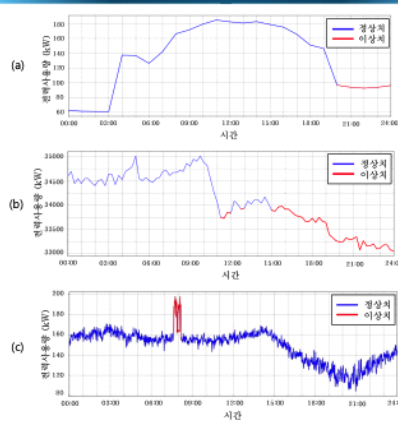
- LSTM 기반의 Autoencoder는 시계열 데이터의 시간적 패턴을 효과적으로 학습하며, 입력과 복원 값 간의 차이를 바탕으로 이상 여부를 판단.
- 시계열 데이터의 장기 의존성 반영에 강점이 있어, 정상 패턴이 반복되는 산업 설비와 같은 환경에서 효과적으로 활용 가능.



USAD(Unsupervised Anomaly Detection)

- 하나의 Encoder와 두 개의 Decoder로 구성되어, 두 단계의 학습을 통해 정상 데이터의 패턴을 효과적으로 학습하고 입력과 복원 값의 차이를 이용해 이상치를 탐지.
- 정상과 비정상 사이의 경계에 위치한 애매한 이상치 구간까지 효과적으로 구분할 수 있음.

데이터 구성 및 전처리



데이터의 시간별 전력사용량 그래프 (a) LEAD1.0 데이터셋 (b) 실제 산업 전력 데이터셋 (c) KPI 데이터셋

- 이상치 탐지 성능 평가를 위해, 두 공개 데이터셋 LEAD1.0(저해상도, 1시간 간격)과 KPI-Anomaly-Detection(고해상도, 1분 간격)과, 실제 산업 환경에서 15분 간격으로 수집된 중해상도 산업 전력 데이터를 활용.
- LEAD와 KPI 데이터셋에는 수작업 또는 전문가에 의해 라벨링 된 이상치 정보가 포함되어 있으며, 실제 산업 데이터는 라벨 없이 모델 성능을 간접 평가하는 기준으로 사용.

데이터 전처리

- LEAD1.0 : 산업 전력과 유사한 건물을 선택
- KPI-Anomaly-Detection : 산업 전력의 정상적인 분포를 보이는 항목을 선택
- 세 데이터셋 모두 : 전력 사용량은 Min-Max 정규화를 적용해 입력 특성 간 스케일 차이 조정.

실험 방법 및 결과 분석

	LEAD1.0	Industrial Power	KPI
Train	8361	10064	69649
Validation	1747	3451	21269
Test	1747	3453	21284

표 1) 세 데이터셋을 훈련/검증/테스트 6:2:2의 비율로 분할한 표.

	LEAD1.0	Industrial Power	KPI
Model	LEAD1.0	USAD	USAD
Result	0.6210	0.8179	0.7777
F1 Score	0.7019	0.8696	0.8195
Precision	0.6039	0.8147	0.8036
Recall	0.6999	0.8639	0.8362

표 2) LEAD1.0, KPI와 실제 산업 전력 데이터셋에서 LSTM Autoencoder와 USAD의 이상 탐지 성능 비교

- (표 1) 이상치 탐지 성능을 평가하기 위해 산업 전력 데이터 구조를 반영한 세 가지 시계열 데이터셋을 훈련/검증/테스트 6:2:2 비율로 활용.
- (표 2) USAD 모델이 중저해상도 데이터에서는 더 높은 탐지 성능을 보이며, KPI와 같은 고해상도 데이터에서는 상대적으로 낮은 성능을 보임.
- 변화 폭이 작고 이상치가 드물게 발생하는 환경에서 USAD의 최대 학습 구조가 미세한 이상치에도 민감하게 반응.
- 실제 산업 환경에서 자주 발생하는 중저해상도 데이터 수집 환경을 고려할 때, USAD의 Decoder 구조를 통해 미세한 이상치 변화에도 민감한 이상치 탐지에 강점을 지닌 모델로 평가.

결론 및 향후 연구

- LSTM Autoencoder와 USAD를 활용하여 서로 다른 시간 해상도를 가진 산업 전력 데이터셋에서의 이상치 탐지 성능을 비교.
- 실험 결과, 데이터셋의 해상도에 따라 두 모델의 성능 차이가 나타나며, 중저해상도 환경에서는 USAD가 높은 탐지 성능을 나타내고, LSTM Autoencoder는 상대적으로 낮은 탐지 성능을 보임.
- USAD는 복잡한 LSTM 구조 없이 이상치에 민감하게 반응할 수 있어 중저해상도 데이터가 주를 이루는 산업 환경에 적합한 대안.
- 향후 두 모델의 강점을 결합한 하이브리드 이상치 탐지 모델을 설계하여 해상도 변화나 실시간 환경에서도 강건한 이상치 탐지 성능을 유지하는 시스템을 구축할 예정.
- 다양한 해상도의 산업 전력 데이터의 신뢰성을 향상시키고, 에너지 효율 최적화 및 탄소 배출 저감 목표 달성에 기여하고자 함.



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

박범도¹, 권우현¹, 이충호, 허태욱², *이상금¹

*국립한밭대학교¹, 한국전자통신연구원²

EcoAI Lab

Abstract

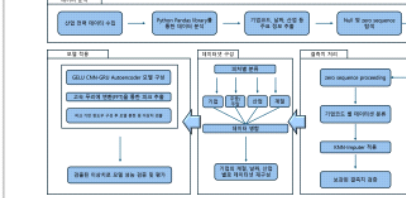
• (연구 배경 및 목표) 산업 현장의 복잡한 전력 데이터로부터 설비 고장 예방 및 에너지 효율 최적화를 목표로, FFT 기반의 동적 윈도우 구성과 GELU CNN-GRU 오토인코더를 결합한 이상치 탐지 방법론을 제안
• (데이터 전처리 및 주기 추출) 1차 차분으로 데이터의 변화 패턴을 강조하고, FFT 분석으로 주요 주파수 성분을 식별하여 상위 주기를 객관적으로 추출, 이를 동적 분석 윈도우로 설정
• (다중 윈도우 기반 이상 탐지 방법) 추출된 주기들로 다중 스케일 윈도우를 구성, CNN으로 지역적 패턴을, GRU로 시간적 의존성을 학습하는 오토인코더 모델을 설계, GELU 활성화 함수를 학습 안정성을 확보
• (DTW 및 Isolation Forest 기반 이상치 탐지) 시간 축 왜곡에 강인한 오차 측정을 위해 DTW를 활용하고, 계산된 오차 값을 분포에 비지도 학습인 Isolation Forest를 적용하여 이상치 윈도우를 최종 판별
• (연구 의의) 제안된 다단계 방식은 실제 산업 현장의 다양한 이상 상황에 정교하게 대응 가능하며, 스마트 팩토리의 데이터 기반 예지 보전 시스템 구축을 위한 핵심 기술로 활용될 수 있음

Introduction & Research Background



산업 설비의 예지보전과 에너지 관리 효율화, 시스템 안정성 확보를 위해서는 전력 데이터의 이상 탐지가 필수적이다. 기존의 통계적 방법은 복잡한 이상 패턴을 포착하는 데 한계가 있으며, 딥러닝 기반 Autoencoder 구조가 이상 탐지의 효과적인 대안으로 주목받고 있다. 본 연구는 FFT 기반 다중 주기 분석, CNN-GRU Autoencoder, DTW, Isolation Forest 등을 통합하여, 실제 산업 현장에서 다양한 시간 스케일의 이상 상황을 신뢰성 있게 탐지하는 방안을 제시한다.

Objective

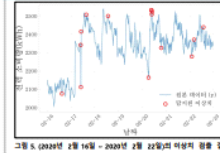


Materials & Methods

데이터 전처리

- 데이터셋: (2020~2022년) 3년간 15분 단위로 수집된 특정 산업체의 전력 사용량 데이터
- 결측치 처리: sklearn.KNN Imputer를 이용, 시계열 데이터 패턴 고려 및 연속성 확보
- 1차 차분 (Differencing): 비정상성을 완화하고 급격한 변화를 감지하여 동적 윈도우 구성에 활용

Results & Discussion



- 다중 스케일 이상치 탐지: FFT 기반 데이터 특성에 맞는 다중 윈도우 설정으로, 순간적인 피크나 강하 같은 단기 이상과 미묘한 패턴 변화를 보이는 장기 이상을 모두 효과적으로 탐지
- 강인한 탐지 성능: DTW 오차와 Isolation Forest의 결합을 통해 단순한 값의 변동뿐 아니라, 시간적 왜곡이 포함된 복잡한 형태의 이상 패턴까지 정밀하게 식별
- 실제 산업 이상 상황 매핑: 탐지된 이상치들은 실제 현장의 설비 가동 중단, 예기치 않은 부하 발생, 공정 라인 비정상 가동 등 다양한 잠재적 위험 상황과 연결 가능

Future Works

- (탐지 방법 향상) 본 연구는 1차 차분, FFT 기반 주기 분석, GELU-CNN-GRU Autoencoder, DTW, Isolation Forest의 다단계 결합으로 산업 전력 데이터의 다양한 이상 패턴을 효과적으로 탐지할 수 있음을 입증함
- (확장성) 향후 모델의 지속적인 발전을 통해 보다 다양한 산업 데이터셋으로 확장 및 이상 원인 분석, 실시간 예측 모델과의 연동을 통하여 자동화·고도화 추진 예정
- (스마트 팩토리 적용) 제안된 방법론은 스마트 팩토리의 실시간 모니터링, 에너지 효율 관리 및 설비 예지보전을 통해 산업 현장에서의 실질적 기여를 기대함

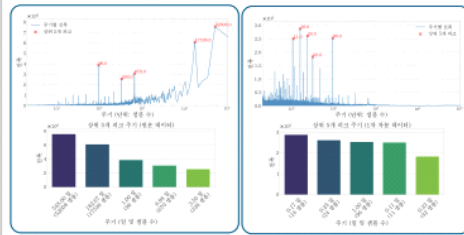
(향후 모델 개선 방안)

CNN 레이어를 GCN으로 구성하여 모델 훈련

- 더 강력한 시계열 데이터 반영
- 데이터 간 관계에 더 집중
- 지역 예측력이 더 강함
- MLSTM 네트워크에 적용 가능

그림 6 CNN 구조도

FFT 기반 주기 추출 & 동적 윈도우 구성

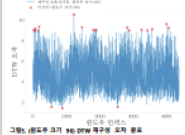


GELU CNN-GRU Autoencoder 모델

- CNN 계층으로 윈도우 내 지역적 패턴을 추출
 - GRU 계층이 시계열의 시간적 의존성을 학습하여 잠재적 패턴에 압축 후 다시 원본 자료로 복원하는 구조
- GELU 활성화 함수를 적용하여 학습 안정성 및 성능을 향상

구분	모델명	평균 정확도 (%)
CNN 계층	Conv1D	85.2
	Conv2D	88.5
	Conv3D	90.1
	Conv4D	91.5
GRU 계층	GRU1	82.3
	GRU2	85.7
	GRU3	87.9
	GRU4	89.4

DTW 재구성 오차 기반 IsolationForest 이상치 탐지



- 모델이 재구성한 윈도우와 원본 윈도우 간의 형태적 유사성을 시간 축 왜곡에 강인한 DTW(Dynamic Time Warping)로 계산하여 재구성 오차를 산출
- 이 오차 분포에 비지도 학습인 Isolation Forest를 적용하여 이상치 윈도우를 최종 판별