

# 배터리 노화 패턴 필터링을 통한 Transformer-Isolation Forest 모델 제안

김민준, 고동영, 김선혁  
국립공주대학교 전기전자제어공학부

Kmj728mmin5636@smail.kongju.ac.kr, dyko3113@smail.kongju.ac.kr,  
seonh@kongju.ac.kr

## Proposal of a Transformer-Isolation Forest Model Through Battery Degradation Pattern Filtering

Min Jun Kim, Dong Young Ko, Seon Hyeog Kim  
Department of Electrical, Electronic and Control Engineering, Kongju National  
University

### 요약

본 논문은 Transformer 기반 특징 추출과 Isolation Forest 기반 필터링을 결합한 Transformer-Isolation Forest 모델을 제안한다. 제안 모델은 Transformer로 배터리 데이터의 잠재 특징을 추출한 뒤, Isolation Forest를 이용해 정상 상태 패턴에서 벗어나는 데이터를 제거한다. NASA 데이터셋을 이용한 실험을 통해 제안 모델의 필터링 성능을 검증하였으며, 이는 배터리 수명 예측 모델의 데이터 품질을 향상시키는 효과적인 전처리 프레임워크를 제시한다.

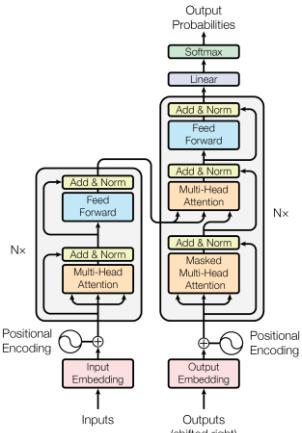
### I. 서론

배터리의 잔존 수명(SOH: State Of Health) 예측 정확도는 데이터 품질에 크게 좌우되지만, 실제 운용 데이터는 노이즈와 이상 신호로 인해 순수한 열화 패턴 분석이 어렵다. 이러한 문제를 해결하기 위해, 본 논문에서는 Transformer와 Isolation Forest를 결합한 Transformer-Isolation Forest 모델을 제안한다. 제안 모델은 Transformer가 정상 상태 데이터의 패턴을 학습해 특징을 추출하면, Isolation Forest가 이 패턴에서 벗어나는 이상 데이터를 필터링하는 2 단계 접근법을 사용한다. 본 논문은 Transformer-Isolation Forest 모델의 노화 패턴 필터링 성능 검증을 통해, 향후 배터리 예측 모델의 정확도를 높일 수 있는 새로운 전처리 프레임워크의 효용성을 입증하고자 한다.

### II. 본론

#### 2.1. Transformer 기반 특징 추출

Transformer 모델은 기존의 순환(Recurrence) 및 컨볼루션(Convolution) 구조를 배제하고 오직 어텐션(Attention) 메커니즘에만 의존하는 모델이다. 구조적으로는 대부분의 경쟁력 있는 신경 시퀀스 변환 모델과 마찬가지로 인코더(Encoder)-디코더(Decoder) 구조를 갖는다. 인코더는 입력 시퀀스  $(x_1, \dots, x_n)$ 를 연속적인 표현인  $z = (z_1, \dots, z_n)$  시퀀스로 매핑하며, 디코더는 이  $z$ 를 기반으로 출력 시퀀스  $(y_1, \dots, y_m)$ 를 한 요소씩 생성하는 자기회귀(autoregressive) 방식으로 작동한다. Transformer는 인코더와 디코더 모두에 스택(Stack)된 셀프 어텐션(Self-attention) 및 점별 완전 연결 계층(Point-wise feed-forward layer)을 사용하여



<그림 1> Transformer 모델 구조[1]

<그림 1>과 같은 구조를 구현한다. 이 아키텍처의 핵심인 셀프 어텐션 메커니즘은 시퀀스 내 모든 요소들 사이의 복잡한 잠재적 특징 및 의존 관계를 계산하여 시퀀스 내의 복잡한 잠재적 특징 및 의존 관계를 효과적으로 학습하는 데 특화되어 있다[1].

#### 2.2. Isolation Forest 기반 데이터 필터링

Isolation Forest는 기존의 정상 데이터 모델링 (Profiling) 방식에서 벗어나, 이상치(Anomaly)가 갖는 '소수이며 동떨어진(few and different)' 특성을 이용하여 데이터셋으로부터 직접 이상치를 분리(Isolation)하는 비지도 학습 알고리즘이다. 이 기법의 핵심 아이디어는, 이상치가 정상 데이터에 비해 더 적은 수의 분리 (partitioning)로 고립될 수 있다는 점에 기반한다. 다수의 격리 트리(iTree)로 구성된 앙상블(ensemble)을 통해

각 데이터 포인트의 평균 경로 길이를 측정하고, 이를 정규화하여 이상치 점수(Anomaly Score)를 산출한다. 데이터 인스턴스가  $n$ 개일 때, 평균 경로 길이의 기준 값  $c(n)$ 은 식 (1)과 같이 계산된다.

$$c(n) = 2H(n-1) - (2(n-1)/n), \quad (1)$$

여기서  $H(i)$ 는 조화수(harmonic number)이다. 이 기준 값과 특정 데이터  $x$ 의 평균 경로 길이  $E(h(x))$ 를 이용하여 최종 이상치 점수  $s(x, n)$ 를 식 (2)와 같이 계산한다.

$$s(x, n) = 2 \frac{E(h(x))}{c(n)}, \quad (2)$$

이 점수는 1에 가까울수록 이상치로 판단된다. Isolation Forest는 선형 시간 복잡도를 가지며 고차원 데이터에서도 빠르고 효율적으로 이상치를 탐지할 수 있다[2].

### 2.3. Transformer–Isolation Forest 제안 모델

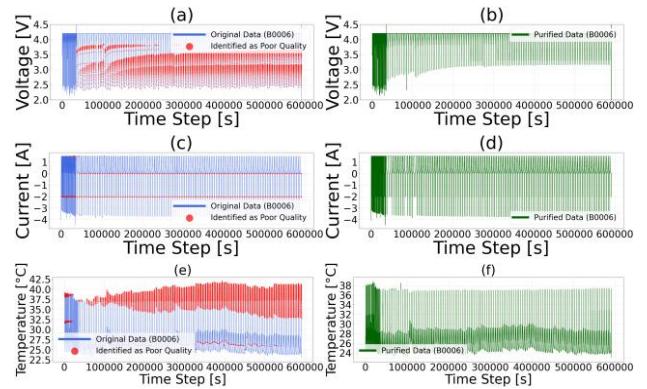
본 논문에서 제안하는 Transformer–Isolation Forest 프레임워크는 배터리 시계열 데이터로부터 고품질의 SOH 특징 벡터를 추출하기 위한 2 단계 파이프라인으로 구성된다. 1 단계: 특징 추출(Feature Extraction)에서는 Transformer의 인코더가 배터리 운용 데이터(전압  $V$ , 전류  $I$ , 온도  $T$ )의 장기적인 패턴 정보를 학습하여, 고차원의 잠재 특징 벡터  $Z$ 를 추출한다. 2 단계: 특징 필터링(Feature Filtering)에서는 추출된 잠재 특징 벡터  $Z$ 를 Isolation Forest에 입력하여 정상 상태 패턴에서 벗어나는 이상 특징을 탐지하고 필터링한다. 이를 통해 정제된 특징 벡터  $Z$ 를 최종적으로 획득한다.

### 2.4. 실험 설정 및 모델 학습

본 논문에서의 실험 데이터는 미국 항공 우주국(NASA) PCoE (Prognostics Center of Excellence)의 배터리 데이터셋을 사용하였다. SOH 가 100%에 해당하는 초기 20 사이클의 데이터를 배터리 정상 상태 기준 데이터로 정의하고 모델 학습에 사용하였다. Transformer 인코더는 자기지도학습(Self-Supervised Learning) 방식인 SimSiam(Simple Siamese) 아키텍처를 기반으로 학습되었다[3]. 이는 별도의 레이블 없이 데이터 본연의 특성을 학습함으로써, 열화가 진행되지 않은 이상적인 운용 패턴의 특징을 모델이 스스로 학습하도록 유도하기 위함이다.

### 2.5. 제안 모델의 필터링 성능 검증

제안된 Transformer–Isolation Forest 모델의 성능을 검증하기 위해 NASA 데이터셋의 B0006 배터리 데이터에 훈련된 모델을 적용하였다. 그 결과, 제안 모델은 정상 상태 패턴을 벗어난 노이즈 및 열화가 진행된 데이터들을 성공적으로 탐지 및 필터링하는 것을 확인하였다. <그림 2>는 모델이 학습한 정상 상태 패턴을 원본 데이터에서 식별한 결과와 필터링을 통해 각 파라미터가 정제된 결과를 시각화한 것이다. 이 필터링 결과는 제안한 모델이 데이터의 불순물을 효과적으로 제거하여 고품질의 특징 지표를 생성할 수 있음을 입증한다.



<그림 2> 원시 데이터 및 정제 데이터

(a) 전압의 원시 데이터, (b) 전압의 정제 데이터, (c) 전류의 원시 데이터, (d) 전류의 정제 데이터, (e) 온도의 원시 데이터, (f) 온도의 정제 데이터

## III. 결론

본 논문에서는 Transformer 와 Isolation Forest 를 결합한 새로운 Transformer–Isolation Forest 모델을 제안하고, 배터리 노화 데이터의 필터링 실험을 통해 그 성능을 성공적으로 입증하였다. 제안 모델은 Transformer의 특징 추출 능력과 Isolation Forest의 효율적인 이상치 탐지 능력을 결합하여, 복잡한 시계열 데이터에 포함된 노이즈와 열화 패턴을 효과적으로 분리해냈다. 실험 결과는 제안한 모델이 데이터의 불순물을 제거하고 순수한 정상 상태 특징을 추출하는데 매우 효과적임을 명확히 보여주었다. 이를 통해 본 연구는 배터리 데이터 전처리를 위한 새로운 프레임워크를 제시했다는 점에서 중요한 의의를 갖는다. Transformer–Isolation Forest 모델은 향후 배터리 수명 예측 모델의 정확성을 높이는 핵심 전처리 모듈로 활용될 수 있으며, 나아가 다양한 산업 분야의 시계열 데이터 이상치 탐지 문제에도 확장 적용될 수 있을 것으로 기대한다.

## ACKNOWLEDGMENT

본 연구는 산업통상자원부 및 한국산업기술진흥원의 사업지원(KIAT) (NO. RS-2025-04752996), 그리고 국립공주대학교 학술연구지원사업의 의하여 연구되었음.

## 참고 문헌

- [1] A. Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NIPS)*, 2017.
- [2] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," *2008 Eighth IEEE International Conference on Data Mining*, 2008.
- [3] X. Chen and K. He, "Exploring Simple Siamese Representation Learning," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [4] 권상욱, 한동호, 박성윤, 김종훈, "LSTM 모델기반 리튬이온 배터리 SOH 예측알고리즘 구현", *대한전기학회논문지*, 제 68 권, 제 10 호, 1214-1221 쪽, 2019년 10월.
- [5] 강민지, 김서림, 김선혁, 허태우, 권구영, "LSTM-VAE 기반 공장 이상 모니터링 기술 개발", 2025년도 *한국통신학회 동계종합학술발표회* 논문집, 2025년 2월, 156-157 쪽.