

마스크 조건 TimeGAN 기반 비정규 센싱 데이터의 정규 시계열 복원

김현정, 이연서, 김파란하늘, 현장훈*

국립한밭대학교

20231052@edu.hanbat.ac.kr, *jhhyeon@hanbat.ac.kr

Masked TimeGAN for Regular Time-Series Reconstruction from Irregular Sensing Data

Hyeonjeong Kim, Yeonsoo Lee, Paranhaneul Kim, Janghun Hyeon*

Hanbat National University

요약

이상 탐지 시스템의 일반화를 위해서는 일관된 센싱 주기로 정렬된 입력이 요구된다. 그러나 실제 현장 데이터는 제조사나 설계의 차이와 통신 오류로 인해 배터리마다 주기가 다르며 센싱 주기가 길게 지속되는 경우, 기존 보간 기법만으로는 신호의 고정 주기적 특징을 안정적으로 복원하는 데 한계가 있다. 본 논문은 이런 한계를 보완하기 위해 생성형 모델을 마스크 기반 생성형 모델로 확장하여 관측값은 보존하고 결측만 복원함으로써 불규칙한 센싱 데이터를 단일 고정 격자로 일관되게 변환하고 이상 탐지 시스템의 이식성을 고려한 고정 주기 재구성 방법을 제안한다. 실험 결과, 제안한 모델은 유효한 조건 내에서 MAE 최대 19.3%, MSE 최대 20.5% 감소를 보여 선형 보간 대비 시점 또는 구조적 패턴을 더 안정적으로 복원함을 확인하였다.

I. 서론

ESS(에너지 저장 시스템) 배터리에서 화재와 같은 이상 탐지 시에 일관된 센싱 주기의 데이터를 이용하면 이상 탐지 성능의 안정성을 기대할 수 있다[1]. 센싱 주기가 서로 다를 때, 센싱 주기가 길수록 정보량이 줄어들어 짧은 이상 신호에 대한 탐지가 어려워지기 때문에 이상 탐지 성능의 불안정성이 커진다. 그러나 실제 환경에서는 운영 조건과 장비 설정 차이로 인해 ESS마다 센싱 주기가 달라 일관된 센싱 주기 데이터를 확보하기 어렵다. 따라서 혼재 센싱 주기 데이터를 공통의 고정 주기 격자로 재구성하는 전처리가 필요하다.

고정 주기 격자 재구성은 선형 보간(Linear)[2]이나 3차 회선 보간(Cubic)[2]과 같은 수학적 기법으로 가능하다. 하지만 이러한 방법들은 [그림 1-④]같이 전체적인 데이터 패턴을 고려하지 않고 인접한 데이터에만 의존하기 때문에 샘플링 주기가 길어질수록 세밀한 변동을 정밀하게 복원하는 데 한계가 존재한다.

본 연구는 관측된 센싱 주기보다 짧은 센싱 주기를 갖는 고정 주기 시계열을 생성하도록 관측값을 기반으로 비관측 값만 보간하는 생성형 모델을 제안한다. 이를 통해 전체적인 패턴을 고려하면서 혼재 센싱 주기 데이터를 고정 주기의 격자로 일관되게 재구성하는 방법을 제안한다.

II. 본론

본 연구는 [그림 1-⑤]와 같이 입력으로 사용하기 위해 다른 주기로 리샘플링된 데이터와 해당 센싱 주기 시점의 위치를 나타내는 마스크를 입력으로 사용한다. 또한 기존 시계열 생성형 모델의 손실을 수정하여 비관측 구간만을 복원하도록 한다. 이를 기반으로 학습된 모델은 테스트 데이터에 적용되어 임의의 혼재 주기 데이터에 대해 고정 주기 데이터로 재구성하는 성능을 평가한다.

1. 데이터 전처리

데이터는 배터리 rack을 기준으로 여러 module로 나누어지며, 각 module은 다시 여러 개의 cell로 구성된 계층적 구조를 가진다. 이 구조에서 설정된 주기를 가지도록 개별 cell 단위마다의 전압값을 저장한다.

사용한 데이터는 센서 오류 및 방전에 의해 의미 없는 0값이 많아 스케일

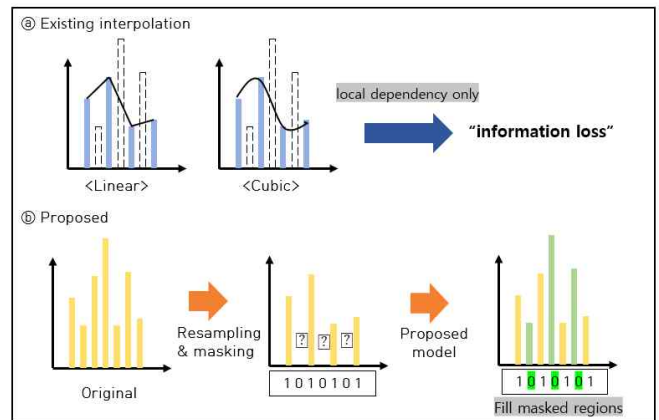


그림 1. 기존 보간의 한계 및 제안 방법

링 과정에서 왜곡되지 않도록 하기 위해 Min-Max 스케일링을 적용한다. 스케일링 과정에서 데이터 누수를 방지하기 위해 학습 데이터에서의 Min, Max 값을 이용한다. 이후 [그림 1-⑥]의 리샘플링 과정을 수행한다. 리샘플링 과정은 혼재 주기 환경을 모사하기 위해 고정된 주기로 센싱된 데이터를 다양한 센싱 주기로 리샘플링하여 센싱 주기 별로 리샘플링 데이터를 생성한다. 리샘플링 시, 해당 주기 시점의 값은 관측 영역으로 판단하여 그대로 보존하고 그 외 시점은 0으로 채운다. 또한 별도의 채널을 만들어 리샘플링 데이터에서 실제 값이 존재하는 시점은 1로, 0으로 채워진 시점은 0으로 표시하는 마스크 벡터를 생성한다.

2. 마스크 조건부 생성형 모델

본 논문은 잠재 공간에서 시간적 동역학을 보존하며 실제와 유사한 시퀀스를 생성하는 시계열 생성 모델 TimeGAN[3]을 기반으로 비관측 구간 보간을 통해 고정 주기로 재구성하는 네트워크를 제안한다. 기존 TimeGAN은 전체 시퀀스 생성을 목적으로 설계되어 관측 값을 보존하고 결측 구간만 선택적으로 생성할 수 없다. 따라서 TimeGAN의 입력을 확장하고 손실을 재설계함으로써 결측 구간을 보간하여 고정 주기 센싱 데이터를 생성하도록 한다.

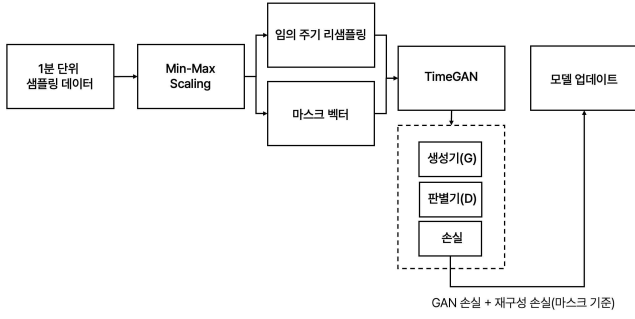


그림 2. 마스크 기반 확장 모델 학습 구조

2.1 모델 구조

제안된 모델은 혼재 주기 데이터와 해당 위치의 관측 여부를 나타내는 마스크 벡터를 입력으로 비관측 구간을 모델이 직접 복원하도록 한다. 생성기는 리샘플링된 데이터와 마스크 벡터를 입력으로 받아 관측 구간에서는 원래 값을 유지하고 비관측 구간에서는 새로운 값을 생성하여 고정 주기로 재구성한다. 판별기는 이렇게 복원된 시계열과 실제 원본 시계열을 구분한다. 학습 과정에서는 입력 데이터를 무빙 윈도우 방식으로 일정 길이의 시퀀스로 분할하여 배치 단위로 학습시키며 이를 통해 장기적 패턴과 국소적 패턴을 동시에 학습한다[그림 2].

2.2 손실 함수 설계

본 연구는 기존 TimeGAN 손실 구조를 기반으로 손실 함수를 수정 및 확장한다. 기존 TimeGAN은 재구성 손실(Reconstruction Loss), 감독 손실(Supervised Loss), 적대적 손실(Adversarial Loss), 임베딩 손실(Embedding Loss)로 구성된다. 본 연구에서는 재구성 손실을 비관측 구간에 초점을 맞추어 학습하도록 하는 인페인팅 손실(Inpainting Loss)로 수정한다. 또한 관측 구간에서 입력값과 복원된 값이 동일하게 유지되도록 하는 아이덴티티 보존 손실(Identity Preservation Loss)을 추가한다.

2.2.1 인페인팅 손실

$$L_{inp} = E[\| (1 - M) \odot (X - \hat{X}) \|] \quad (1)$$

기존 재구성 손실은 임베더-리커버리 네트워크의 출력 $\hat{X}$ 과 원본 X 의 차이를 윈도우 전체 시계열 구간에서 계산한다. 그러나 본 연구는 비관측 영역 손실에 대해 최적화해야하기 때문에 해당 주기 시점의 위치는 1, 그 외 시점을 0으로 하는 마스크 벡터 M 에 따라서 마스크 벡터 M 이 비관측(0)일 때만 손실에 반영되도록 하는 인페인팅 손실(1)로 수정한다.

2.2.2 아이덴티티 보존 손실

$$L_{id} = E[\| M \odot (X - \hat{X}) \|] \quad (2)$$

아이덴티티 보존 손실(2)은 관측 값을 훼손하지 않도록 강제하는 역할을 한다. 마스크 벡터 M 을 통해 관측된 위치(1)에서만 출력 $\hat{X}$ 과 원본 X 가 같아지도록 오차를 계산한다.

3. 실험 설계

실험은 제안한 모델과 전통적인 보간 기법인 선형보간(Linear), 3차 회선보간(Cubic)을 비교한다. 비교 과정에서 3.3V 이하의 값을 방전 값 또는 센서 오류로 인한 값으로 판단하여 3.3V 이하의 값은 평가에서 제외한다. 또한 정답이 NaN인 구간은 단순 결측이 아닌 측정 종료로 해석하여 이 구간에서 모델이 3.3V 이상을 예측하면 패널티를 부여하여 평균 절대 오차(Mean Absolute Error)와 평균 제곱 오차(Mean Squared Error)로 성능을 평가한다.

4. 결과

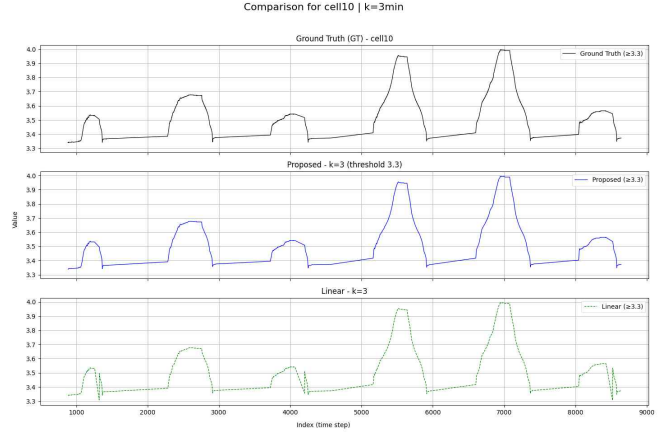


그림 3. 제안 방법 및 선형 보간 결과 예시 ($k=3$)
[그림 3]은 선형 보간과 제안 모델의 재구성 예를 비교한 것이다. 선형 보간은 고정 주기화 자체는 가능하지만 왜곡이 나타나는 반면, 제안 모델은 상승, 하강 시점 지속 등 구조적 패턴을 더 안정적으로 복원함을 볼 수 있다.

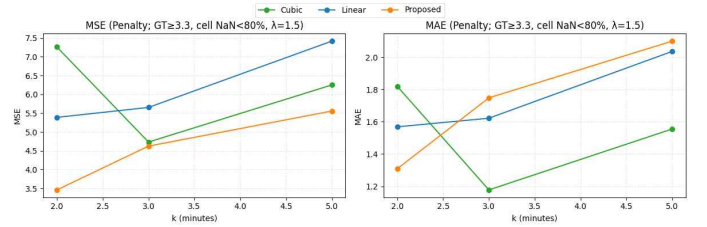


그림 4. 기존 보간기법과 제안 모델 보간의 성능 비교

[그림 4]를 볼 때, 제안된 모델은 전통적인 보간 기법보다 짧은 샘플링 간격에서 최저 오차를 보이며 긴 간격에서도 경쟁력을 유지하는 균형적인 특성을 보인다. 하지만 평균 절대 오차 관점에서는 장주기에서 선형, 3차 회선 보간이 일부 구간에서 더 낮은 값을 보이는 한계가 존재한다.

III. 결론

본 연구는 관측값을 그대로 보존하고 결측만 복원하도록 학습하는 마스크 조건 구조를 통해 서로 다른 샘플링 주기를 일관된 주기로 정규화하는 방법을 제안한다. 제안한 모델은 유효한 조건에서 실용 범위($k < 5$ 분)에서는 평균 절대 오차 최대 19.3%, 평균 제곱 오차 최대 20.5% 감소를 확인하여 샘플링 주기가 다양한 환경에서도 일정한 주기로 보정이 가능함을 보인다. 한편 조건이 없는 환경에서는 단순 보간(Linear, Cubic)이 점 추정 정확도에서 더 유리한 경우가 존재해 제안 모델의 우월성을 모든 상황으로 일반화하기 어렵다. 그럼에도 제안한 방식은 보간법의 구조적 한계를 답러닝으로 보완하는 대안으로서 다양한 주기 환경에서 신뢰 가능한 시계열 복원을 제공한다.

참고 문헌

- [1] Liu W., Zhang Z., Zhao Z., and Zhang W., "A Framework for Anomaly Cell Detection in Energy Storage Systems Based on Daily Operating Voltage and Capacity Increment Curves," Batteries, vol. 11, no. 8, p. 316, 2025
- [2] Burden R. L. and Faires J. D., Numerical Analysis, 9th ed., Brooks/Cole, 2010.
- [3] Yoon J., Jarrett D., and van der Schaar M., "Time-series Generative Adversarial Networks," Advances in Neural Information Processing Systems (NeurIPS), vol. 32, 2019,