

확산 확률모형 기반 송전선로 고장데이터 생성기법

이인화, 이경영, 윤성국

숭실대학교

neverforget1114@naver.com, dlru7755@naver.com, sgyoon@ssu.ac.kr

Data Generation Method of Transmission Line Fault Data Using Denoising Diffusion Probabilistic Models

In-Hwa Lee, Kyeong-Yeong Lee, Sung-Guk Yoon

Soongsil Univ.

요약

본 논문은 송전선로 고장데이터 부족 문제를 해결하기 위해 확산 확률모형(Denoising Diffusion Probabilistic Model, DDPM) 기반 송전선로 고장데이터 생성 기법을 제안한다. 시뮬레이션으로 생성된 고장데이터를 학습한 DDPM을 통해 다양한 고장 유형의 데이터를 생성하고, 생성 데이터의 품질을 실효값(Root Mean Square, RMS) 기반 확률 분포를 이용한 KL-divergence(Kullback-Leibler divergence)로 평가하였다. 사례연구 결과, 제안기법으로 생성된 데이터는 기존 시뮬레이션 데이터 및 비교기법으로 생성한 데이터 대비 실제 고장데이터 분포와 가장 높은 유사도를 보였으며, 이를 통해 제안기법이 실제 고장데이터와의 분포 간극을 효과적으로 완화할 수 있음을 확인하였다.

1. 서론

송전선로 고장데이터는 전력계통의 고장 진단, 고장 유형 분류 및 계통 운용 의사결정 지원 등 다양한 분석과 모델 개발에 활용된다[1]. 이러한 데이터 기반 분석의 신뢰도를 확보하기 위해서는 양질의 고장데이터가 요구된다.

그러나 실제 송전선로에서 발생하는 고장은 빈도가 낮고, 운용 환경 및 제약으로 인해 고장데이터 수집이 제한되는 경우가 많아 충분한 고장데이터를 확보하는 데 한계가 있다. 이에 따라 선행 연구에서는 실제 고장 상황을 모의할 수 있는 물리 기반 시뮬레이션을 활용하여 고장데이터를 생성하고, 이를 바탕으로 다양한 고장 시나리오 분석을 수행해 왔다[2]. 다만 시뮬레이션은 실제 계통의 복잡한 운전 조건과 고장 특성을 완벽히 반영하기 어렵기 때문에, 시뮬레이션을 통해 생성된 고장데이터는 실제 고장데이터의 분포 및 특성과 차이를 가질 수 있다. 이로 인해 시뮬레이션 데이터에 기반하여 학습된 모델은 실제 계통 환경에 적용 시 일반화 성능 확보에 한계를 보이며, 고장 진단 및 고장 유형 분류 정확도가 저하될 가능성이 있다[3].

이러한 한계점을 고려하여 본 연구에서는 확산 확률모형(Denoising Diffusion Probabilistic Model, DDPM)을 활용한 송전선로 고장데이터 생성기법을 제안한다. 사례연구로 DDPM으로 생성한 데이터와 실제 고장데이터 간의 분포 유사도를 KL-divergence(Kullback-Leibler divergence)를 통해 정량적으로 분석한다. 이를 통해 제안한 DDPM 기반 데이터 생성기법이 시뮬레이션 데이터 분포와 실제 고장데이터 분포와의 간극을 줄일 수 있음을 보였다.

II. 본론

2.1 DDPM

DDPM[4]은 점진적인 확률적 확산과 그 역과정을 통해 데이터 생성 과정을 모델링하는 생성 모델이다. DDPM은 그림1과 같이 원본 데이터 x_0 에 단계적으로 노이즈를 더해 완전한 노이즈 x_T 상태로 만드는 정방향 확산 과정을 거친다. 이후 모델은 해당 과정을 역으로 추정하여 단계적으로 노이즈를 제거함으로써 데이터를 복원하는 역방향 생성 과정을 학습한다. 이는

원본 데이터 분포 $p(x)$ 를 직접 추정하는 대신, 가우시안 노이즈가 추가된 데이터를 다시 복원하는 일련의 과정을 학습함으로써 실제 데이터 분포 $p(x)$ 를 모델이 생성하는 분포 $p_\theta(x_0)$ 로 근사하는 방식이다.

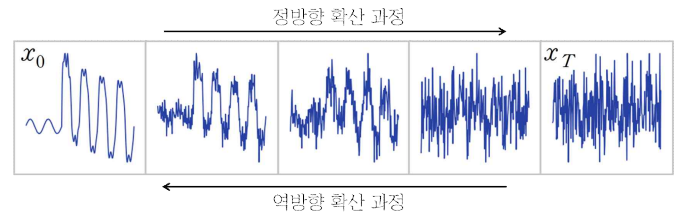


그림 1 DDPM을 이용한 확산 기반 데이터 생성 과정의 개념도

실제 학습 과정에서는 각 확산 단계 t 에 따라 생성된 x_t 로부터 추가된 가우시안 노이즈 ϵ 를 예측하도록 신경망을 학습하며, 이에 따른 학습 목적함수는 수식 (1) 과 같다.

$$L_{DDPM} = \mathbb{E}_{x_0, \epsilon, t} [\|\epsilon - \epsilon_\theta(x_t, t)\|^2] \quad (1)$$

여기서 ϵ 은 가우시안 분포에서 샘플링된 실제 노이즈이며, $\epsilon_\theta(x_t, t)$ 는 신경망이 예측한 노이즈이다. 이와 같은 학습 과정을 통해 DDPM은 생성 분포 $p_\theta(x_0)$ 가 학습 데이터 분포 $p(x)$ 를 근사하도록 학습된다.

2.2 제안기법

본 연구에서는 다양한 송전선로 고장데이터 생성을 위해 DDPM 기반 고장유형 데이터 생성기법을 제안한다. 그림 2는 제안하는 DDPM 기반 데이터 생성기법의 전체 개념도이다. 제안기법은 1) DDPM 모델 훈련 단계와 2) KL-divergence 기반 최적 데이터셋 선정의 총 두 단계로 구성된다.

1) DDPM 모델 훈련: DDPM 모델 훈련을 위해 송전선로 고장 상황을 모의한 시뮬레이션 데이터를 사용한다. 시뮬레이션으로 생성된 고장데이터는 DDPM이 근사하고자 하는 목표 데이터 분포 $p(x)$ 를 정의하며, 모델은 확산 기반 생성 과정을 통해 생성 분포 $p_\theta(x_0)$ 가 해당 분포를 근사하도록 파라미터 θ 를 학습한다. 이를 위해 고장 유형, 선로 길이, 위상각 등의 파라미

터를 변화시켜 다양한 고장 시나리오를 구성하고, 이에 따른 고장데이터를 생성한다. 본 연구에서 사용한 DDPM은 조건부 생성 모델이 아니므로 고장 유형별로 모델을 개별 학습한다.

2) KL-divergence 기반 최적 데이터셋 선정: DDPM 모델을 통해 생성된 데이터는 확률적으로 샘플링되므로, 동일한 모델이라 하더라도 샘플링 과정에 따라 생성 결과의 데이터 분포가 달라질 수 있다. 따라서 생성된 데이터셋이 학습에 사용된 시뮬레이션 데이터셋의 분포를 얼마나 충실히 반영하는지를 정량적으로 평가할 필요가 있다. 본 연구에서는 고장 유형의 특성을 고려하여 각 데이터셋에 대해 실효값(Root Mean Square, RMS)에 기반한 확률 분포를 구성하고, 시뮬레이션 데이터셋과 DDPM으로 생성된 데이터셋 간의 분포 차이를 KL-divergence를 통해 비교·분석한다. KL-divergence는 두 확률 분포 간의 차이를 정량적으로 측정하는 지표로, 그 값이 작을수록 두 분포 간의 유사도가 높음을 의미한다. 이에 따라 본 연구에서는 훈련된 DDPM 모델을 이용하여 복수의 생성 데이터셋을 생성한 후, KL-divergence 값이 가장 작은 데이터셋을 최적의 생성 데이터셋으로 선정한다. 이러한 KL-divergence 기반 데이터셋 선정 방법의 구체적인 절차와 타당성은 선행연구[5]를 따른다.

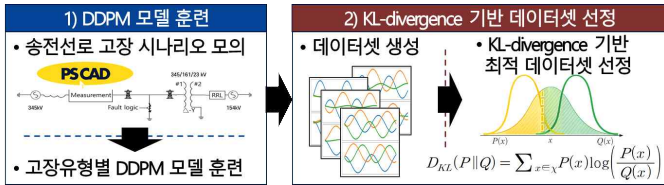


그림 2 데이터 생성과정 전체 개념도

2.3 사례연구

본 사례연구의 목적은 제안한 DDPM 기반 송전선로 고장데이터 생성 기법이 실제통 고장데이터의 분포를 얼마나 효과적으로 모사할 수 있는지를 정량적으로 검증하는 것이다. 이를 위해 시뮬레이션 데이터, 비교기법, 그리고 제안기법으로 생성된 데이터셋을 대상으로 실제통 고장데이터와의 분포 유사도를 비교·분석한다.

DDPM 모델의 학습을 위한 시뮬레이션 고장데이터는 국내 345 - 154 kV 계통을 단순화한 모델을 기반으로 PSCAD를 사용하여 생성하였다. 고장 유형은 AG, BG, CG, ABG, ACG, BCG, AB, AC, BC, ABC의 총 10가지로 구성하였으며, 송전선로 길이는 10 km에서 100 km까지 10 km 간격으로 변화시켰다. 또한 고장 위상각은 0°에서 180°까지 30° 간격으로 설정하였다. 고장 유형별로 총 10개의 DDPM 모델을 개별 학습한 후, 각 고장 유형에 대해 1,000개의 고장데이터셋을 생성하였다. 각 고장데이터는 A, B, C의 3개 상으로 구성되며, 고장 발생 이후 2주기(128 샘플)를 대상으로 하고 샘플링 주파수는 3,840 (60×64) Hz이다.

제안기법의 유효성을 검증하기 위해, 제안기법을 통해 생성된 데이터셋을 실제 송전선로에서 발생한 고장데이터와 비교하였다. 실제통 고장데이터는 약 9년간 수집된 총 51건의 고장데이터로 구성된다. 비교 대상 데이터셋은 표 1에 정리하였다. 각 데이터셋에 대해 실제통 고장데이터를 기준으로 KL-divergence를 계산하고, 그 결과를 그림 3에 도시하였다.

표 1 실제통 고장데이터 기준 분포와 비교 대상 데이터셋 구성

데이터셋 유형	생성 방법
시뮬레이션	PSCAD 기반 고장 시뮬레이션
비교기법 [5]	GAN(Generative Adversarial Networks) 기반 생성
제안기법	DDPM 기반 데이터 생성

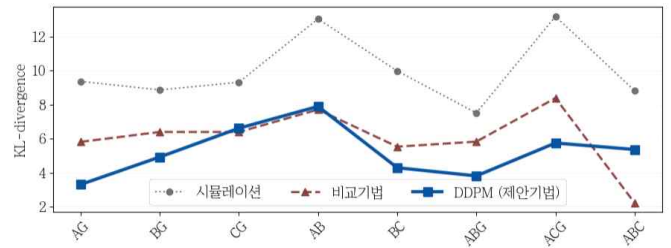


그림 3 실제통 고장데이터셋과의 KL-divergence 값 비교

분석 결과 DDPM 기반 생성 데이터셋은 시뮬레이션 데이터셋보다 낮은 KL-divergence 값을 나타냈다. 이는 시뮬레이션 데이터와 실제통 고장데이터 간의 분포 간극이 DDPM을 통해 완화될 수 있음을 의미한다. 또한 DDPM 기반 생성 데이터셋은 전반적으로 GAN 기반 생성 데이터셋보다 낮은 KL-divergence 값을 보였다. 이는 판별자 기반의 적대적 학습을 통해 분포를 암묵적으로 근사하는 GAN과 달리 확산 확률모형이 점진적인 확률적 생성 과정을 통해 고장데이터의 분포 특성과 다양성을 보다 안정적으로 반영할 수 있기 때문으로 해석된다. 이러한 결과는 확산 기반 생성 모델이 송전선로 고장데이터 생성에 있어 시뮬레이션 데이터의 한계를 보완할 수 있음을 시사한다.

III. 결론

본 연구에서는 DDPM 기반 송전선로 고장데이터 생성기법을 제안하였다. RMS 기반 KL-divergence 분석 결과 시뮬레이션 고장데이터를 학습한 DDPM을 통해 생성된 데이터는 시뮬레이션 및 비교기법 대비 실제통 고장데이터 분포와 가장 높은 유사도를 보였다. 이는 DDPM 기반 데이터 생성기법이 시뮬레이션 데이터와 실제통 고장데이터 간의 분포 차이를 완화할 수 있음을 시사한다. 후속 연구에서는 제안기법의 적용이 송전선로 고장 진단 및 분류 모델의 일반화 성능 향상으로 이어지는지를 분석할 예정이다.

ACKNOWLEDGMENT

본 연구는 한국전력공사의 2025년 착수 사외공모 기초연구 사업의 지원을 바탕으로 수행되었습니다. (과제번호:R25X003-10)

참고 문헌

- [1] A. Firos, N. Prakash, R. Gorthi, M. Soni, S. Kumar, and V. Balaraju, "Fault detection in power transmission lines using AI model," IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), pp. 1-6, Feb. 2023.
- [2] Y. Xi, W. Zhang, F. Zhou, X. Tang, Z. Li, X. Zeng, and P. Zhang, "Transmission line fault detection and classification based on SA-MobileNetV3," Energy Reports, vol. 9, pp. 955-968, Nov. 2023.
- [3] T. G. Kim, S. H. Lim, K. M. Song, and S. Yoon, "LSTM-based Fault Classification Model in Transmission Lines for Real Fault Data," Trans. Korean Inst. Electr. Engineers, vol. 73, no. 3, pp. 585-592, Mar. 2024.
- [4] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Advances in Neural Information Processing Systems (NeurIPS) 33, pp. 6840-6851, Dec. 2020.
- [5] K. Y. Lee, S. H. Lim, T. G. Kim, K. M. Song, and S. G. Yoon, "GAN-Based Data Augmentation Technique for Various Transmission Line Fault Data," Trans. Korean Inst. Electr. Engineers, vol. 73, no. 8, pp. 1318-1326, Aug. 2024.