

다변량 시계열 데이터를 활용한 Transformer 기반의 LMP 예측 모델 연구

조운호, 조재영, 김선혁
국립공주대학교

202102082@smail.kongju.ac.kr, whwodud02@smail.kongju.ac.kr, seonh@kongju.ac.kr

Transformer based on multivariate time series data A Study on the LMP Prediction Model

Yun Ho Jo, Jae Yeong Jo, Seon Hyeog Kim
Kongju National University

요약

본 논문에서는 변동성이 큰 지역별 한계가격(LMP)를 정확하게 예측하기 위해 순환적 시간 인코딩이 적용된 Transformer 기반의 다변량 시계열 예측 모델을 제안한다. 제안 모델은 데이터의 주기성을 명시적으로 학습하기 위한 전처리 기법과 상태 추정 개념을 응용하여 인코더의 마지막 시점 특징을 추출하는 구조를 채택하였다. 제안 모델의 우수성을 입증하기 위해 LSTM, GRU 등 기존 딥러닝 모델과 성능을 비교 실험하였다. 실험 결과 제안 모델은 $R^2=0.6730$, $MAE=1.82\$/MWh$ 를 기록하며 비교군 대비 가장 높은 예측 성능을 보였다. 이는 불확실성이 높은 전력 시장 데이터에서 시계열의 장기 의존성을 효과적으로 학습하고 최신 상태 정보를 보존하는 모델 설계가 필수적임을 시사한다.

I. 서론

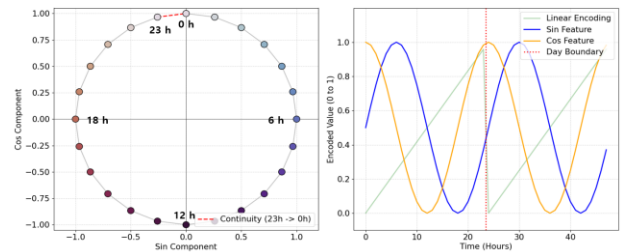
전 세계적으로 탄소 중립 달성을 위한 재생에너지원의 보급이 확대되면서 전력 시스템의 운영 환경이 급변하고 있다. 한국 전력 시장 또한 분산형 전원 중심으로 전환되고 있으며 이에 따라 전력 도매 가격 결정 방식 또한 기존의 계통한계가격(SMP) 체제에서 지역별 송전 혼잡과 손실을 반영하는 지역별 한계가격(LMP) 체제로의 전환이 예고되고 있다. 그러나 LMP는 송전망의 물리적 제약과 수급 상황을 실시간으로 반영하기 때문에 변동성이 매우 크고 예측이 난해하다[1,3].

기존의 시계열 예측 연구에서는 주로 LSTM이나 GRU와 같은 순환 신경망(RNN) 모델이 활용되었으나 이는 입력 시퀀스가 길어질수록 초기 정보가 손실되는 기울기 소실 문제나 급격한 가격 변동을 포착하는데 한계가 있었다[2]. 이를 극복하기 위해 본 연구에서는 Transformer 모델을 전력 가격 예측에 도입한다. 특히 전력 데이터 고유의 주기성을 반영한 순환적 시간 인코딩과 마지막 시점 특징 추출 기법을 제안하여 예측 정확도를 극대화하고자 한다.

II. 본론

2.1. 데이터 구성 및 선정 근거

본 연구에서는 재생에너지 비중이 높아 변동성이 큰 전력 시장 환경을 모의하기 위해 북미 캘리포니아 전력 시장(CAISO)의 실제 운영 데이터를 활용하였다. 데이터 수집 기간은 2023년 1월 1일부터 12월 31일까지이며 1시간 단위로 샘플링된 총 8,760개의 시계열 데이터를 구축하였다. LMP 예측의 정확도를 높이기 위해 단일 변수가 아닌 가격 결정 메커니즘에 직접적인 영향을 미치는 5가지 핵심 변수(과거 48시간의 LMP 값, 전력수요, 태양광, 풍력, 천연가스 가격)를 선정하여 다변량(Multivariate) 데이터셋을 구성하였다.



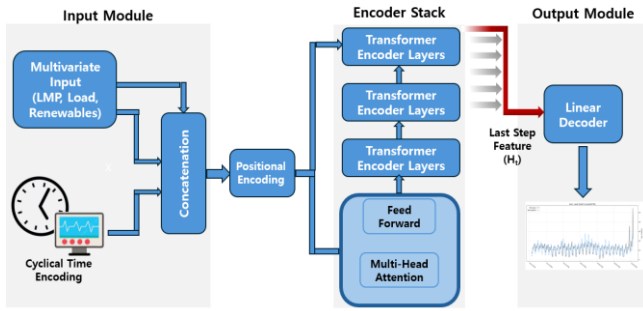
[그림 1] 순환적 시간 인코딩 시각화

2.2. 데이터 전처리 및 순환적 시간 인코딩

수집된 이중 데이터는 물리적 단위와 스케일의 차이가 크므로 경사 하강법 기반 학습의 안정성을 위해 모든 변수에 대해 Min-Max 정규화를 수행하여 데이터의 범위를 [0,1]로 통일하였다. 또한, 시계열 데이터의 전처리의 핵심은 시간 정보의 연속성을 보존하는 것이다. 기존의 선형 인코딩 대신 삼각함수를 활용한 순환적 시간 인코딩 기법을 적용하였다. [그림 1]은 해당 기법의 기하학적 구조와 데이터의 연속성을 시각화한 결과이다. [그림 1]의 좌측은 시간 데이터를 Sin과 Cos 성분으로 변환하여 2차원 평면에 투영한 결과이다. 0시와 23시의 데이터 포인트가 기하학적으로 인접해 있어 시간의 순환적 흐름이 위상 공간상에서 완벽한 원형 구조로 보존됨을 알 수 있다. [그림 1]의 우측은 48시간 동안의 시계열 데이터 변화를 나타낸다. 기존 선형 인코딩 방식은 하루가 지나면 시점에서 값이 급격히 하락하는 불연속성을 보인다. 이는 딥러닝 모델이 23시와 0시를 전혀 다른 특성으로 오인하게 만드는 원인이 된다. 반면, 제안하는 순환 인코딩 방식은 날씨가 변하는 경계선에서도 데이터가 끊김 없이 부드럽게 이어지는 연속성을 보장한다.

$$t_{enc} = [\sin(\frac{2\pi t}{T}), \cos(\frac{2\pi t}{T})] \quad (1)$$

식(1)에서 t 는 현재 시점의 시간 값을 나타내며, T 는 해당 시간 정보의 전체 주기를 의미한다. 이를 통해 모델은 시간의 흐름에 따른 전력 소비 패턴의 주기성을 효과적으로 학습할 수 있다.



[그림 2] Time-Aware Transformer 모델 구조

2.3. Time-Aware Transformer 모델 구조

[그림 2]은 제안하는 Time-Aware Transformer 모델의 전체적인 구조를 나타낸 그림이다. 제안하는 모델은 시계열 데이터의 장기 의존성을 학습하기 위해 Transformer의 인코더 구조를 기반으로 설계되었다. 입력된 과거 48 시간의 데이터는 위치 인코딩과 결합된 후 4개의 층으로 적층된 인코더 블록을 통과하며 셀프 어텐션 메커니즘을 통해 현재 시점의 가격 결정에 중요한 과거 시점들에 가중치를 부여하게 된다. 본 연구의 핵심은 인코더의 최종 출력 처리 방식에 있다. 일반적인 방식과 달리 본 연구에서는 인코더의 출력 중 가장 마지막 시점의 특징 벡터만을 추출하여 디코더의 입력으로 사용하였다. 이는 과거의 정보들을 종합한 후 예측 시점과 가장 가까운 최신 상태 정보를 집중적으로 반영하여 단기 예측의 정확도를 높이기 위함이다. 손실 함수로는 가격 스파이크와 같은 이상치 예측에 강건한 성능을 보일 수 있도록 예측값과 실제값 차이의 제곱을 최소화하는 평균 제곱 오차(MSE)를 채택하였다.

III. 실험 결과

3.1. 실험 환경 및 비교 설정

제안 모델의 성능 검증을 위해 과거 48 시간의 데이터를 입력 받아 향후 24 시간의 LMP를 예측하는 실험을 수행하였다. 데이터 분할은 시계열의 순서를 유지하며 8:2 비율을 적용하였다. 제안 모델의 구조적 우수성을 입증하기 위해 시계열 예측에 사용되는 LSTM 및 GRU 모델과 성능을 비교하였다. 모든 모델은 동일한 전처리 데이터와 손실 함수를 사용하여 공정한 비교를 수행하였다.

3.2. 결과 분석

각 모델의 예측 성능 비교 결과는 <표 1>과 같다. 실험 결과 제안하는 Transformer 모델이 모든 지표에서 가장 우수한 성능을 기록하였다.

<표 1> 예측 모델 성능 비교 결과

모델	R^2	MAE [\$/MWh]	RMSE [\$/MWh]	SMAPE [%]
GRU	0.4236	2.28	3.07	22.07
LSTM	0.5390	2.04	2.87	20.72
Transformer	0.6730	1.82	2.43	19.57

비교 모델인 GRU와 LSTM은 각각 R^2 점수 0.4236과 0.5390을 기록하며 일정한 예측 성능을 보였으나 제안하는 Transformer 모델에는 미치지 못했다. 제안 모델은 R^2 점수 0.6730, MAE 1.82\$/MWh를 기록하여 비교 모델 대비 가장 낮은 오차율을 보였다. 이러한 성능 향상의

주된 원인은 Transformer의 Self-Attention 메커니즘에 있다. RNN 계열 모델은 시퀀스가 길어질수록 정보가 희석되는 경향이 있는 반면 Transformer는 과거 48 시간의 모든 시점을 동시에 참조하여 가격 변동에 영향을 미치는 중요 시점을 직접적으로 학습할 수 있다. 또한 인코더의 마지막 시점 특성을 추출하는 방식이 최신 시장 상황을 즉각적으로 반영하여 예측의 민감도를 높이는 데 기여한 것으로 분석된다. [그림 3]에서도 제안 모델이 실제 가격의 변동 추세를 가장 유사하게 추종함을 확인할 수 있다.



[그림 3] 제안 모델의 15일간 예측 결과

IV. 결론

본 연구에서는 한국형 LMP 시장의 도입에 대비하여 순환적 시간 인코딩과 Transformer를 결합한 LMP 예측 모델을 제안하였다. CAISO 데이터를 활용한 비교 실험을 통해 제안 모델이 기존의 GRU 및 LSTM 모델보다 우수한 예측 정확도($R^2=0.6730$)를 보임을 확인하였다. 특히 순환적 시간 인코딩 기법은 시간적 불연속성을 제거하여 전력 데이터의 주기성을 효과적으로 학습하는 데 기여했으며 마지막 시점 특징 추출은 급변하는 시장 상황을 예측에 즉각 반영하는데 중요한 역할을 하였다. 다만, 본 연구에서 수행한 단일 노드 LMP 예측은 인접 지역과의 송전 혼잡이나 전력 조류와 같은 공간적 상관성을 직접적으로 반영하지 못한다는 점에서 SMP 예측과 유사한 한계를 가질 수 있다. 따라서 향후 연구에서는 한국 전력 시장의 실제 계통 토폴로지 정보를 반영하여 발전소 간의 거리 및 송전 제약에 따른 공간적 상관성을 고려한 GNN 기반의 시공간 LMP 예측 모델로 확장할 계획이다.

ACKNOWLEDGMENT

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

참고 문헌

- [1] 송재도, "LMP와 송전이용요금 개념의 중복 적용을 통한 지역별 차등요금", 자원·환경경제연구, 제 34 권, 제 2 호, pp.231-256, 2025.
- [2] 임상택, 김준영, A.M.Z.Khaki, 최아영, "딥러닝 기반 지역별 한계 가격제 예측 기술 개발", 한국차세대컴퓨팅학회 논문지, 제 21 권, 제 2 호, pp.49-61, 2025.
- [3] X.Wang, Y.Zhang, and C.Kang. (2021). "Locational marginal price forecasting using Transformer-based deep learning network," *IEEE Power & Energy Society General Meeting(PESGM)*.