

시계열 결측치 보간을 위한 전력 사용량의 확률적 예측

신한솔, 이충호, 허태욱
한국전자통신연구원

hansol.shin@etri.re.kr, leech@etri.re.kr, htw398@etri.re.kr

Stochastic prediction of electric power use for time-series missing data imputation

Hansol Shin, Chungcho Lee, Tae-Wook Heo
Electronics and Telecommunications Research Institute (ETRI)

요약

국내 총 에너지 소비량의 60%를 차지하는 산업시설에 관한 에너지 절감 및 탄소중립 방안이 요구되며, 이를 위해서는 에너지 소비현황 파악 및 소비패턴 분석이 필요하다. 본 논문에서는 에너지 소비현황 분석을 위해, 산업시설의 전력 사용량 데이터 수집과정에서 발생하는 결측치를 인공지능 모델인 TFT(Temporal Fusion Transformer)를 활용하여 보간하는 방법을 기술한다. TFT 모델은 다섯 종류의 대상시설에 대해, 예측오차 9.5% 이하로 전력 사용량을 확률적으로 예측했으며, 모델의 평균 신뢰도는 88.7%로 도출되었다.

I. 서론

2019 년도 대한민국의 산업부문 에너지 소비량은 전체 최종 에너지 소비에서 60%를 차지한다[1]. 산업부문 에너지 소비 절감안 도출을 위해, 에너지 소비 현황 측정 및 데이터 분석이 선행되어야 한다. 그러나 에너지 소비 데이터 수집 과정에서 에너지 소비와 관련 없는 문제(e.g. 원격 검침기의 이상, 통신망 품질의 문제)로 인해 완전 무작위 결측이 발생할 수 있다. 따라서, 주어진 데이터만으로 원본 데이터 추론이 필요하며, 추론 결과의 신뢰성 확보를 위해 불확실성 가시화가 가능한 확률적 예측 모델이 필요하다.

본 연구에서는 시계열 데이터의 확률적 예측이 가능한 TFT(Temporal Fusion Transformer)를 이용하여 실제 산업시설에서 측정된 전력 사용량의 예측 모델을 제작했으며, 해당 모델의 성능과 신뢰성을 평가했다. 이러한 AI 기반 결측치 보간 방법을 적용하여, 산업시설의 에너지 소비 현황 및 소비패턴 분석을 수행하고, 에너지 절감 방안을 도출하여 탄소중립에 기여하고자 한다.

II. TFT를 이용한 전력 사용량 결측치 보간

i. Temporal Fusion Transformer

TFT는 장기 시계열 데이터의 예측 성능 향상을 위해, 미래 시점에 미리 알 수 있는 정보들(known inputs)(e.g. 요일, 휴일), 정적 공변량(static covariates)(e.g. 용도, 면적) 및 관측값(observed inputs)(e.g. 에너지 사용량)을 이용하여 모델을 구성한다. 또한, 각 예측시점(point forecasts)마다 분위수(quantile) 기반의 확률적 예측범위(prediction interval)를 식 1과 같이 추론한다. 이 때, $\hat{y}_i(q, t, r)$ 은 시점 t 에서 예측시간간격 τ 만큼 앞선 예측된 q^{th} 분위 예측결과이며, $f_q(\cdot)$ 는 예측 모델, z 는 현재 및 과거에 측정가능한 정보, x 는 미리 알 수 있는 정보, s 는 정적 공변량이다.

$$\hat{y}_i(q, t, r) = f_q(\tau, y_{i,t-k:t}, z_{i,t-k:t}, x_{i,t-k:t+\tau}, s_i) \quad (1)$$

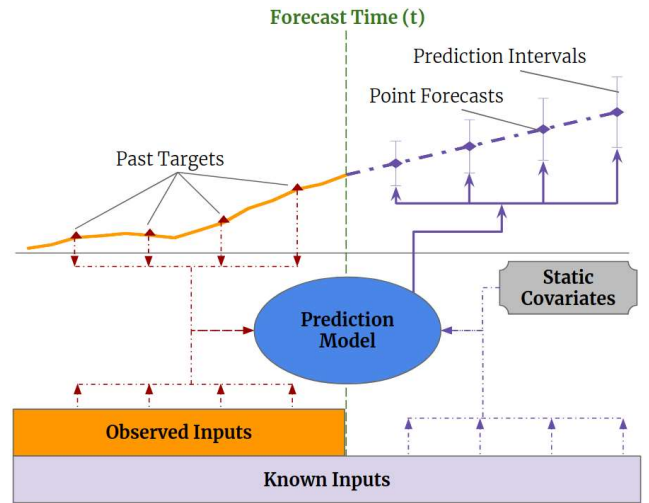


Figure 1. Illustration of multi-horizon forecasting [2]

ii. 에너지다소비 산업시설 전력 예측 모델

본 연구에서는 에너지다소비 산업시설의 15 분간격 전력 데이터를 사용하여 TFT 모델의 예측 성능을 검토했다. 모델 학습 및 성능 검증에 사용된 산업시설 5 개소의 속성은 표 1과 같다. 대상 산업시설들은 위치, 업종, 연면적이 다양하게 분포되어 있으며, 대상 산업시설에서 전력은 제품 생산에만 사용되었다.

Table 1. Properties of target industrial facility

#	Location	Industrial sector	Floor area [m ²]	Purpose of electricity use
1	Incheon	Food	56,912	Production
2	Yeosu	Chemical	16,362	Production
3	Pyeongtaek	Energy	9,758	Production
4	Seongnam	Water treatment	2,221	Production
5	Cheonan	Steel	5,685	Production

본 연구에서는 Darts[3] python 패키지를 이용하여 TFT 모델을 제작했으며, 입력 및 출력은 표 2 와 같다. 이 때, 입력 시계열은 과거 1 주일에 해당하는 672 개의 15 분간격 데이터(=4*24*7)이며, 출력 시계열은 1 일에 해당하는 96 개의 15 분간격 데이터(=4*24)이다. 대상 산업시설의 전력 사용량 측정기간이 상이하여, TFT 모델 학습 및 검증을 위해, 시설별로 다른 기간에서 데이터를 추출했다(표 3). 모델 오차율은 CVRMSE(Coefficient of the Variation of the Root Mean Square Error)를 이용하여 평가했으며, 15% 이하를 신뢰도 기준으로 적용했다.

Table 2. Inputs and outputs of TFT model

Known inputs	Observed inputs	Static covariate inputs	Outputs
Holidays, day, hour, minute, day of week, week of year	Electric use [kW/15min]	Location, industrial sector, floor area[m ²], heating and cooling area[m ²]	Electric use [kW/15min]

Table 3. Sequence extraction periods

#	Metered	Train	Validation	Test
1	'22.10.14.- '22.12.31.	'22.10.14.- '22.12.15.	'22.12.16.- '22.12.23.	'22.12.24.- '22.12.31.
2	'22.07.16.- '22.12.31.	'22.07.16.- '22.11.27.	'22.11.28.- '22.12.14.	'22.12.15.- '22.12.31.
3	'20.12.16.- '21.06.25.	'20.12.16.- '21.05.16.	'21.05.17.- '21.06.04.	'21.06.05.- '21.06.24.
4	'22.08.25.- '22.12.31.	'22.08.25.- '22.12.05.	'22.12.06.- '22.12.18.	'22.12.19.- '22.12.31.
5	'22.07.31.- '22.12.31.	'22.07.31.- '22.11.31.	'22.12.01.- '22.12.15.	'22.12.16.- '22.12.31.

III. TFT 모델 검증 결과

각 시설별 테스트 기간에 대한 모델 정확도 검증 결과는 그림 2-6 과 같다. TFT 모델의 예측오차(CVRMSE)는 산업시설#1 부터 #5 까지 순서대로 9.5%, 1.5%, 5.6%, 6.4%, 7.2% 이다. 그림 2-6 에서 노란색 음영은 예측의 10 번째 및 90 번째 백분위 사이의 범위로, 예측결과의 불확실성이다. 특히, 전력 사용량이 증가하거나, 휴일에 전력을 사용한 경우, 불확실성이 증가한다(그림 2). 이는 TFT 모델의 학습에 사용된 산업시설#1 의 전력 소비 데이터가 타 대상시설 대비 적었기 때문으로 판단된다(표 3). 또한, 이상치가 발생한 경우, TFT 모델을 이용하여 필터링 및 보정을 수행할 수 있다(그림 5).

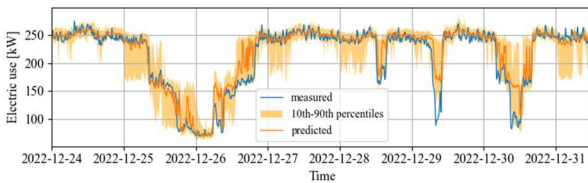


Figure 2. Measured data vs. predicted data (facility #1)

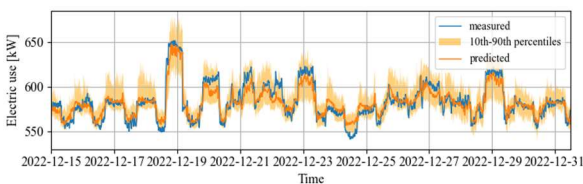


Figure 3. Measured data vs. predicted data (facility #2)

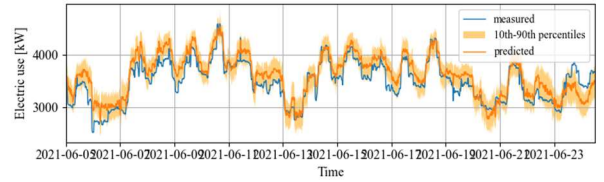


Figure 4. Measured data vs. predicted data (facility #3)

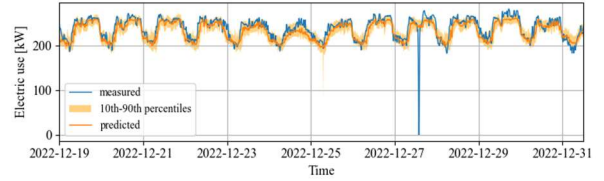


Figure 5. Measured data vs. predicted data (facility #4)

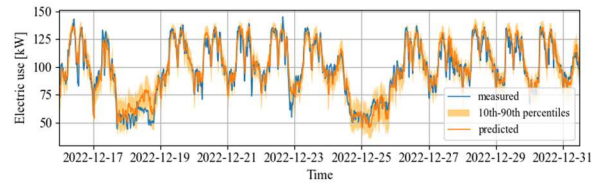


Figure 6. Measured data vs. predicted data (facility #5)

TFT 모델의 0 번째부터 100 번째 분위 예측결과에 대한 오차의 분포를 이용하여, 예측오차가 15% 이하일 확률을 신뢰도로 판단했다. 이 때, 산업시설#1 부터 #5 까지 모델의 신뢰도는 68.3%, 100%, 97.0%, 95.1%, 83.2% 로서, 평균 88.7%로 계산되었다.

IV. 결론

본 연구에서는 산업시설 전력 사용량 결측치 보간을 위해, TFT 를 이용하여 전력 사용량을 예측했다. TFT 모델의 예측 오차는 9.5% 이하였으며, 신뢰도는 평균 88.7% 로 신뢰가능한 예측 모델을 제작할 수 있었다. 추후 연구에서는 모델 성능 및 신뢰성 향상을 위한 데이터 불균형 문제의 해결과, 산업시설 에너지 절감방안 도출을 위한 TFT 모델의 시계열 attention weight 를 이용한 전력 사용패턴의 분석이 필요하다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. RS-2023-00237018)

참 고 문 헌

- [1] Ministry of Trade, Industry and Energy (MOTIE). (2022). *Energy Consumption Survey 2020* (No. 115005). Korean government.
- [2] Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748-1764.
- [3] Herzen, J., Lässig, F., Piazzetta, S. G., Neuer, T., Tafti, L., Raille, G., ... & Grosch, G. (2022). Darts: User-friendly modern machine learning for time series. *The Journal of Machine Learning Research*, 23(1), 5442-5447.