

기상 데이터를 활용한 산업시설 전력 소비량의 확률적 예측

신한솔, 정운철, 이충호, 허태욱
한국전자통신연구원

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

Stochastic prediction of electric use in industrial facilities using weather data

Hansol Shin, Wun-Cheol Jeong, Chunggho Lee, Tae-Wook Heo
Electronics and Telecommunications Research Institute (ETRI)

요약

본 연구는 기상청에서 제공하는 관측 및 예보 데이터를 활용하여 산업시설의 전기 사용량을 예측하고, Temporal Fusion Transformer(TFT) 모델의 확률적 시계열 예측 성능을 평가하였다. 에너지 다소비 산업시설 4 곳의 데이터를 수집하여 예보 데이터를 입력한 모델과 입력하지 않은 모델의 예측 성능을 비교한 결과, 예보 데이터가 예측 성능을 저해할 수 있음을 발견하였다. 이 연구는 기상 데이터를 활용한 예측 모델이 산업시설의 전력 소비 예측에 유용할 수 있음을 보여주며, 각 시설의 특성에 맞는 데이터 입력과 변수 중요도 분석의 필요성을 강조한다.

I. 서론

산업 부문의 에너지 소비량은 국내 최종 에너지 소비량의 60%에 달하며, RE100 이행 및 탄소중립을 위한 에너지 전환의 중요성이 확대되고 있다. 따라서, 산업 부문의 에너지 절감 운영 방안 도출 및 전력 수요량 추산을 위한, 에너지 소비 예측이 요구된다. 그러나 생산활동의 변동성이 크고, 다양한 공정의 혼재되어 있어 소비 패턴이 복잡하며, 시설 및 공정별로 기후에 대한 민감도가 상이하므로 불확실성이 내재된 산업시설의 에너지 소비의 예측은 어려운 과제이다.

본 연구에서는 기상청에서 제공하는 초단기실황 기상 데이터와 초단기예보 데이터를 이용하여, 산업시설 전력 소비량 예측 성능을 개선했다. 전력 소비량 예측을 위해, TFT(Temporal Fusion Transformer) 기반의 확률적 시계열 모델을 제작했으며, 해당 모델을 이용하여 각 시설의 에너지 소비에 대한 기상 영향을 분석했다. 이를 통해 에너지 소비 패턴 및 소비 현황 분석을 수행하여, 산업시설의 탄소중립에 기여하고자 한다.

II. 기상 예보 데이터 기반 전력 사용량 예측 TFT 모델

i. 기상 예보 데이터 및 산업시설 전력 소비 데이터

기상청에서는 5km 해상도 격자로 구성된 동네예보 구역에 대한 대표 방재기상 관측값을 1 시간 간격의 '초단기실황' 자료로 제공한다. 또한, 동일한 지역 해상도에 대해 매시 30 분마다 6 시간 이내에 대해 발표하는 '초단기예보' 데이터도 공개하고 있다. 본 연구에서는 실황 및 예보 데이터 중, 건구온도, 상대습도, 강수량, 강수형태, 하늘상태, 풍속 데이터를 수집했다.

에너지 다소비 산업시설의 전력 소비 데이터를 이용하여 TFT 모델의 예측 성능 및 기상 영향을 분석했다. 연구에 포함된 4 개소의 산업시설은 위치, 업종, 연면적 등에서 다양성을 보이며, 이들 시설에서의 전력 사용은 제

품 생산 과정에만 할당되었다. 모델의 학습과 성능 검증에 사용된 산업시설의 특성은 표 1 과 같다.

Table 1. Properties of target facility

#	Location	Industrial sector	Floor area [m ²]	Purpose of electricity use
1	Seosan	Chemical	26,953	Production
2	Gimcheon	Chemical	14,741	Production
3	Hwaseong	Metal	12,801	Production
4	Cheongju	Food	11,219	Production

ii. Temporal Fusion Transformer 모델

TFT 는 시계열 예측 정확도 개선을 위해 다양한 입력을 활용한다. 미래 예측에 필요한 사전에 알려진 정보(예: 요일, 휴일, 날씨 예보), 변경되지 않는 특성(예: 시설 용도 및 면적), 그리고 관측된 데이터(예: 에너지 사용량, 날씨)로 모델이 구성된다. 또한, 해당 모델은 각 예측 시점에 대해 분위수 기반의 확률적 예측 범위를 식 1 과 같이 추정한다. 이 때, 각 예측시점 t 에서 예측시간간격 τ 를 고려하여 q^{th} 분위의 예측결과 $\hat{y}_i(q, t, \tau)$ 을 계산하며, 이는 예측 모델 $f_q(\cdot)$ 를 통해, 현재 및 과거의 측정가능한 정보 z , 미리 알 수 있는 정보 x , 그리고 정적 공변량 s 를 활용하여 계산된다.

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

TFT 모델은 Python 패키지[1]를 활용하여 구축했고, 입력 데이터 구성은 표 2 와 같다. 입력 데이터로는 과거 하루 동안의 15 분 간격 데이터(=4*24)를 사용했으며, 출력 데이터로는 다음 6 시간의 15 분 간격 전력 소비 데이터(=4*6)를 설정했다. 모델에 사용된 학습, 검증, 테스트 데이터는 각각 2020 년 1 월부터 10 월까지, 11 월, 그리고 12 월에 대한 데이터를 사용했다. 모델의 정확도는 결정계수(R^2)를 사용하여 평가했다.

Table 2. TFT model inputs

Types	Variables
Observed	Electric use[kW/15min], temperature[°C], humidity[%], precipitation[mm/h], precipitation type, sky condition, wind speed[m/s]
Known	Holidays, day, hour, minute, day of week, week of year, temperature[°C], humidity[%], precipitation[mm/h], precipitation type, sky condition, wind speed[m/s]
Static	Location, industrial sector, floor area[m ²], heating and cooling area[m ²]

III. 예측 모델의 정확도 및 기상영향 분석 결과

기상 데이터가 정확도에 미치는 영향을 확인하기 위해, 관측 및 예보 기상 데이터 미입력 모델(A)과 입력 모델(B)로 구분했다(표 3). 시설#2에서는 기상 데이터에 의해 정확도가 증가했다(그림 1, 2). 이때, 노란색 음영은 예측 불확실성 범위인 예측 결과의 10 번째 및 90 번째 백분위 사이이다. 시설#2에서는 기상 데이터 입력 모델의 불확실성이 낮으며, 전력 소비 증감 거동을 정확하게 모사한다. 반면, 시설#3에서는 기상 데이터를 입력한 모델의 예측 성능이 감소했다(그림 3, 4). 특히, 토요일(12월 5일)의 에너지 소비가 기존 대비 감소하여, 두 경우 모두 해당 기간의 불확실성 범위가 타 기간보다 넓었다.

Table 3. TFT models' coefficient of determination (R²)

#	Predicted without weather (A)	Predicted with weather (B)	Difference (C=A-B)
1	0.97	0.98	0.01
2	0.79	0.92	0.13
3	0.09	0.89	<u>-0.01</u>
4	0.87	0.92	0.05

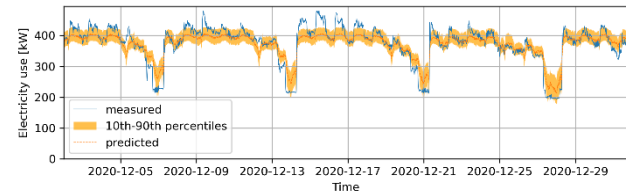


Figure 1. Prediction of facility#2 without weather data

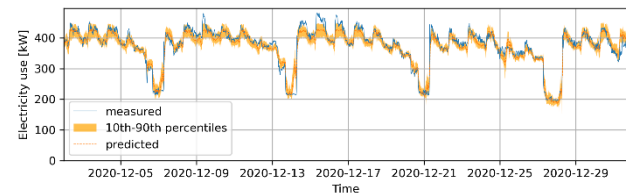


Figure 2. Prediction of facility#2 with weather data

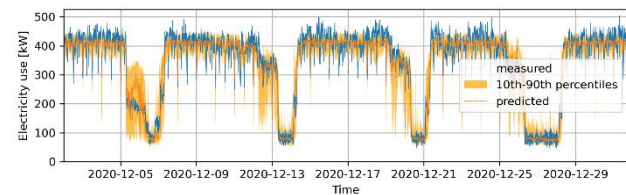


Figure 3. Prediction of facility#3 without weather data

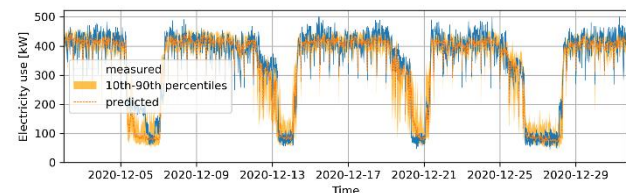


Figure 4. Prediction of facility#3 with weather data

TFT 모델은 과거 관측 데이터를 인코더 레이어에서 입력받고, 미래에 예정된 정보를 디코더 레이어에서 입력받으며, 각 레이어에서 변수 중요도가 계산된다. 기상 데이터 입력 모델에 대한 각 레이어별 변수 중요도 비율은 표 4, 5 와 같다. 기상 데이터 미입력 모델에서는 인코더 레이어에서 과거 전기 소비량의 중요도 비율이 모든 시설에서 가장 높았다(41.6%~74.4%). 표 4 에 따르면, 기상 데이터 입력에 따라 정확도가 향상된 시설#1, #2, #4의 경우 전기 소비량, 시간 변수, 관측된 날씨 순서로 예측에 중요한 변수로 나타났으며, 기상 예보 데이터는 예측에 거의 사용되지 않았다. 반면, 정확도가 감소한 시설 #3에서는 인코더 레이어와 디코더 레이어에서 기상 예보 데이터가 가장 중요한 변수로 선정되었다. 이는 전력 소비량 예측 시 관측된 날씨 데이터가 산업시설 전력 소비량 예측의 정확성을 향상시킬 수 있지만, 기상 예보 데이터는 오히려 예측의 정확성을 저해할 가능성이 있음을 시사한다.

Table 4. Importance ratio of variable types in encoder

#	Observed electric use	Observed weather	Forecasted weather	Time-based variables
1	72.6%	10.1%	0.0%	17.3%
2	46.3%	16.1%	0.0%	37.6%
3	1.2%	15.9%	64.2%	18.7%
4	64.3%	9.1%	0.0%	26.6%

Table 5. Importance ratio of variable types in decoder

#	Forecasted weather	Time-based variables
1	0.0%	100.0%
2	0.0%	100.0%
3	60.8%	39.2%
4	0.0%	100.0%

IV. 결론

본 연구는 기상 데이터를 활용하여 산업시설의 전력 소비량을 예측하는 TFT 모델의 성능을 평가하였다. 기상청에서 제공하는 초단기실황과 초단기예보 데이터를 이용하여 예측 성능을 향상시키고, 각 시설의 에너지 소비에 대한 기상 영향을 분석하였다. 분석 결과, 기상 데이터는 일부 시설에서 예측 정확성을 향상시켰으나, 시설 #3에서는 기상 예보 데이터가 예측 성능을 저해하였다. 이는 관측된 날씨 데이터가 예측에 더 유효하며, 기상 예보 데이터는 불확실성을 증가시킬 수 있음을 시사한다. 따라서, 각 시설의 특성에 맞는 데이터 입력과 변수 중요도 분석이 필요하다. 본 연구에서 제안한 방법론은 산업시설의 에너지 소비 패턴 분석과 탄소중립 목표 달성에 기여할 수 있는 유용한 도구로 활용될 수 있을 것이다.

ACKNOWLEDGMENT

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

참 고 문 헌

[1] 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.