

LIME을 활용한 전력부하예측의 특징 중요도 설명

이채윤¹, 이해성², 박민재³, 이용걸¹, 서정욱⁴, 강민구⁴, 안현^{1*}한신대학교 소프트웨어융합학부¹한국전력공사 전력연구원, 스마트배전연구소²대림대학교 컴퓨터정보학부³한신대학교 IT영상콘텐츠학과⁴e-mail : {pos06090, pattern, jwseo, kangmg, hyunahn}@hs.ac.kr, seaster@kepc.co.kr,
mjpark@daelim.ac.kr

Explaining Feature Importance for Electric Power Load Forecasting Using LIME

Chae-Yoon Lee¹, Haesung Lee², Minjae Park³, Yonggeol Lee¹, Jeongwook Seo⁴,Mingoo Kang⁴, Hyun Ahn^{1*}Div. of Software Convergence, Hanshin University¹Smart Power Distribution Laboratory, KEPCO Research Institute²Div. of Computer Information, Daelim University³Dept. of IT Transmedia Contents, Hanshin University⁴

요 약

최근에 많은 산업분야에서 적용되고 있는 AI 모델들의 설명성 문제가 대두되고 있다. 본 논문은 전력 시스템 운영에 중요한 전력부하예측 모델의 특징 중요도를 XAI 기법인 LIME을 적용하여 설명하고자 한다. 실험에는 배전 선로 부하데이터를 이용하여 랜덤포레스트 회귀 모델과 시퀀스 형태의 특징 벡터를 입력으로 받는 LSTM 모델을 각각 학습시키고, LIME 적용하여 생성한 설명 결과를 비교 및 분석한다.

1. 서 론

최근에 많은 산업분야에서 적용되고 있는 AI 모델들의 설명성(explainability) 문제가 대두되고 있다. 개발된 AI 모델이 현장에 배치된 후에 지속적인 유지보수가 필요한데, 모델의 처리 과정과 결과를 해석하기가 어려워 적절한 개선 방안을 도출하기 어려운 실정이다. 이를 해결하기 위해 설명 가능한 인공지능(eXplainable Artificial Intelligence, 이하 XAI)에 대한 연구가 활발하게 진행되고 있다.

본 논문에서는 전력부하 예측 모델의 특징 중요도를 XAI 기법인 LIME을 적용하여 설명하고자 한다. 실험에는 한국전력공사의 배전선로 부하데이터를 이용하여 랜덤포레스트 회귀 모델(random forest regressor)과 LSTM(Long Short-term Memory) 딥러닝 모델을 각각 학습한다. 다음으로, LIME 기법을 적용하여 특징 중요도 분석 결과를 비교한다.

2. 설명 가능한 인공지능

XAI 기법은 기계학습 과정에서 모델의 입력부터 출력까지의 처리과정과 예측 결과에 대한 투명하고 해석 가능한 설명을 제공하는 목적으로 연구되고 있다. 이러한 XAI의 적용으로부터 산출되는 설명을 통해 다음과 같은 측면[1]에서 AI 모델을 검증할 수 있다.

- 책임성(accountability): 모델이 잘못된 결과를 출력할 경우, 이에 대한 근본원인을 분석
- 신뢰성(trust): 모델의 신뢰성이 중요한 고위험 분야의 경우, 모델의 동작 과정에 대한 명확한 이해가 제공되었는지를 검증

- 준수여부(compliance): 모델에 의한 의사결정이 조직정책, 산업표준, 그리고 정부 규제에 부합하는 지를 검증
- 성능(performance): 모델의 성능을 검증하고, 이를 개선 및 최적화하기 위한 방안을 도출
- 강화된 제어(enhanced control): 명확하게 규명되지 않은 모델의 취약점을 모델의 배치(deployment) 이전에 식별하여 배치 이후의 모델 제어 능력을 강화

XAI를 대표하는 분석 기법으로는 PDP(Partial Dependence Plots), LRP(Layer-wise Relevance Propagation), LIME(Local Interpretable Model-agnostic Explanations[2]), 그리고 SHAP(SHapley Additive exPlanations[3]) 등이 있다. 그 중에서 본 논문에서 사용하는 LIME은 비선형의 AI 모델을 근사하는 설명 가능한 선형 모델을 학습시켜 모델의 지역적 설명(local explanation)을 제공하는 XAI 기법이다. LIME의 설명모델 생성은 다음 수식에 따라 동작한다.

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} L(f, g; \pi_x) + \Omega(g) \quad (1)$$

위의 수식은 지역성 기반 손실(locality-aware loss)과 모델 복잡도와 같이 두 개의 항으로 구성된다. 먼저, 지역성 기반 손실 함수 $L(f, g; \pi_x)$ 는 원본 모델 f , 설명 모델 g , 그리고 설명하고자 하는 인스턴스 x 와 g 의 학습을 위해 생성한 인스턴스 z 사이의 근접성을 나타내는 π_x 로 구성된다. 다음으로 모델 복잡도 함수 $\Omega(g)$ 는 LIME에 의해 생성된 설명 모델 g 의 복잡도를 나타낸다. 예를 들어, 선형회귀(linear regression) 기반의 설명 모델을 생성할 경우 가중치 개수가 모델의 복잡도로 측정된다. 결론적으로 LIME 기법은 두 항의 합을 최소화하는 방향

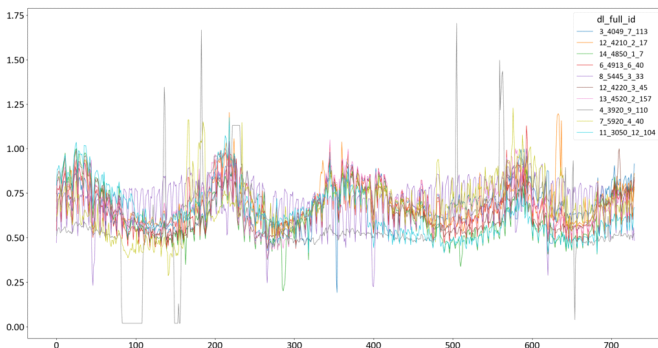
으로 설명 모델을 탐색하여, 지역 충실도(local fidelity)와 해석성(interpretability) 모두 갖춘 설명을 생성하도록 동작한다.

3. 실험 결과

3.1 전력부하예측 모델 학습

전력 분야에서 전력부하예측 문제는 전력 시스템의 운영 관점에서 매우 중요한 분석 이슈이다. 본 논문에서는 실험을 위해 한국전력공사의 총 6,800 개의 배전선로 부하 데이터셋을 사용하였다. 데이터셋은 각 배전선로의 시간별(hourly) 부하값으로 구성된 단일변량 시계열이므로, 특정 중요도를 설명하기 위해서는 다변량 시계열이 확보되어야 한다. 이를 위해, 부하값을 기반으로 시계열 분해(timeseries decomposition)와 시계열 특징 생성 라이브러리인 VEST[4]를 사용하여 총 16개의 특징들로 구성된 일별(daily) 시계열 데이터셋을 생성하였다.

배전선로 별로 상이한 패턴들을 고려하여 계층적 군집화(hierarchical clustering) 알고리즘을 적용하여 시계열들을 분류하였으며, 그림 1과 같이 전력부하예측 모델 학습을 위한 특정 시계열 군집을 선택하였다.



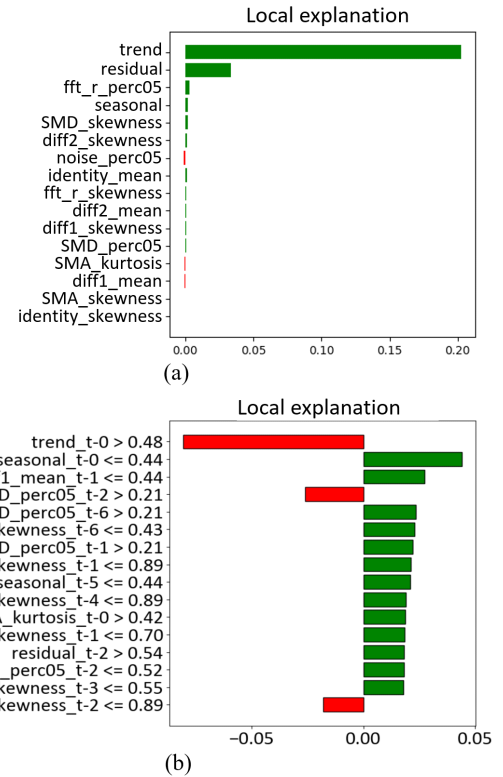
(그림 1) 2,890 개의 배전선로 데이터로 구성된 시계열 군집 샘플

모델 학습에는 시계열 예측에 널리 사용되고 있는 랜덤포레스트 회귀와 LSTM 딥러닝 기법을 사용하여 각각 모델을 학습하였다. LSTM의 경우 시퀀스 길이 7의 many-to-one 예측 모델로 학습함에 따라, 단일 예측에 과거 7일 동안의 특징 벡터가 입력으로 사용된다.

3.2 특징 중요도 설명

LIME은 사용된 학습 데이터의 유형(예: 이미지, 텍스트, 시계열 등)에 따라 다른 시각화 형태의 설명을 제공한다. 본 실험에서는 시계열 예측 모델의 특징 중요도를 시각화하기 위해 LimeTabularExplainer 클래스를 이용하여 설명 모델들을 생성하였다.

그림 2는 LIME을 적용하여 생성된 특징 중요도 분석 결과를 나타낸다. 랜덤포레스트 회귀 모델(그림 2(a))의 경우, 시계열 분해를 통해 생성된 추세(trend)와 잔차(residual) 값이 모델 예측에 가장 큰 영향을 미치는 두 특징으로 분석된다. 반면에, LSTM 모델(그림 2(b))의 경우, 시퀀스 형태의 특징 벡터가 입력으로 사용되므로, 특징 중요도 결과를 설명하기가 용이하지 않다. 예를 들어, 설명 모델의 특징 조건인 $0.32 < \text{SMA_skewness_t-6} \leq 0.43$ 의 경우, 6일 전의 단순이동평균(Simple Moving Average)의 편포도(skewness)를 나타내며 특징 중요도가 0.025인데, 이러한 정보가 시계열 예측 모델을 설명하는데 한계점이 있다고 판단된다.



(그림 2) 랜덤포레스트 회귀 모델(a)과 LSTM 모델(b)의 특징 중요도 분석 결과

4. 결 론

본 논문에서는 XAI 기법인 LIME을 활용하여 전력부하예측 모델의 특징 중요도를 설명하였다. 그러나, LSTM과 같이 many-to-one 예측 모델의 경우 설명성 측면에서 한계점이 있는 것으로 분석되었다. 향후에는 시계열 예측에 유용한 다른 XAI 기법을 적용하여 특징 중요도를 분석하고자 한다.

Acknowledgement

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2021R1F1A1063160).

참 고 문 헌

- [1] Sharma, A. Decrypting Your Machine Learning Model Using LIME, Towards Data Science, 2018. <https://bit.ly/2ZACyLk>
- [2] Ribeiro, M. T., Singh, S. and Guestrin, C., "Why Shoud I Trust You?: Explaining the Predictions of Any Classifier," In Proc. of the 22nd International Conference on Knowledge Discovery and Data Mining, pp. 1135-1144, 2016.
- [3] Lundberg, S. M. and Lee, S. -I., "A Unified Approach to Interpreting Model Predictions," In Proc. of the 31st International Conference on Neural Information Processing Systems, 2017.
- [4] Vitor, C., Moniz, N. and Soares, C., "VEST: Automatic Feature Engineering for Forecasting," Machine Learning, pp. 1-23, 2021.