

다수의 병렬 가중 모델 기반 스마트팩토리 에너지 사용량 예측 알고리즘에 관한 연구

여현영⁺, 심서영⁺, 홍수민⁺, 차채연^{*}, 박형곤^{**}

이화여자대학교 전자전기공학전공⁺, 이화여자대학교 스마트팩토리융합전공^{*}

{eyhy1016, yeni2000, hsum916, chaeyeon.cha}@ewhain.net, hyunggon.park@ewha.ac.kr

Energy Consumption Prediction Algorithm for Smart Factory Based on Multiple Parallel Weighted Models

Hyeon-yeong Yeo⁺, Seoyoung Sim⁺, Sumin Hong⁺, Chaeyeon Cha^{*} and Hyunggon Park^{**}

Department of Electronic and Electrical Engineering, Ewha Womans University^{*}

Graduate Program in Smart Factory, Ewha Womans University^{*}

요 약

본 논문은 머신러닝을 통한 스마트팩토리의 에너지 소비량 예측의 정확성을 높이고자 다수의 병렬 가중 모델을 활용하는 방법을 제안한다. 제안하는 방법은 여러 일 동안 수집된 하루 에너지 사용량 데이터를 하루씩 나누어 다수의 모델에 병렬로 학습시킨 후 각각의 모델의 성능을 평가한 가중치를 예측값에 곱하여 합하는 방식으로 최종 예측치를 얻는다. 실제 스마트팩토리 데이터를 활용한 실험을 통하여 제안한 방법이 하나의 머신러닝 모델에 직렬로 학습시키는 방법보다 평균 제공된 편차가 작고 예측 성능이 우수함을 확인하였다.

1. 서 론

스마트팩토리는 4차 산업 혁명의 핵심 요소 중 하나로 다양한 센서 등을 활용하여 상황을 인지하고 실시간으로 반응할 수 있는 지능형 공장을 의미한다[1]. 이러한 스마트팩토리의 빅데이터 수집, 생산 라인 모델링, 예측 및 생산 지능화에 대한 연구가 전 세계적으로 활발히 이루어지고 있다 [2]. 그 중 스마트팩토리의 핵심 기술인 IoT(Internet of Things)와 관련하여 주목받고 있는 스마트미터는 실시간으로 에너지 사용량 데이터를 수집할 수 있고 더 나아가 본 논문의 목적인 에너지 사용량 예측을 통한 에너지 효율성 향상 및 생산 관리 결정에 중요한 역할을 한다[3]. 그러나 생산 활동 전반이 자동화, 지능화 및 자율화된 스마트팩토리의 특성상 작동 시간 및 에너지 사용량에 대한 변동 폭이 크고 데이터 수집 및 측정 과정에서 잡음이 포함될 가능성이 있다. 이러한 변동 폭은 머신러닝을 통한 예측 성능을 저하시켜 예측 성능 향상에 근본적인 한계치를 만든다[4], [5]. 본 논문에서는 이러한 머신러닝의 성능을 향상시키기 위하여 활용 데이터를 병렬로 학습시킨 다수의 개별적인 머신러닝 모델을 만든 후 각 모델의 예측 결과에 가중치를 주어 최종 예측값을 얻는 방법을 제안한다.

II. 본론

A. 문제 정의

본 논문에서는 n 일 동안 수집된 하루 에너지 사용량 데이터를 통해 $n+1$ 일째 하루 에너지 사용량을 예측하고자 한다. $i(i=1, \dots, n)$ 일 하루를 균등한 시간 간격으로 나누어 총 m 번 에너지 사용량을 측정하였을 때, $t(t=1, \dots, m)$ 번째 시간 간격에서 수집된 에너지 사용량 데이터를 $y_i(t)$ 라 하자. i 일 하루 동안 수집된 에너지 사용량 벡터를 $\mathbf{x}_i$ 라 하면, 벡터 $\mathbf{x}_i$ 는 다음과 같이 표현된다.

$$\mathbf{x}_i = (y_i(1), y_i(2), \dots, y_i(m)). \quad (1)$$

n 일 동안 수집된 하루 에너지 사용량 벡터의 집합을 $\mathbf{X}_n$ 이라 하자. 집합 $\mathbf{X}_n$ 은 다음과 같이 표현된다.

$$\mathbf{X}_n = \{\mathbf{x}_i | i=1, \dots, n\} \quad (2)$$

n 일 동안 수집된 에너지 사용량 벡터의 집합 $\mathbf{X}_n$ 을 머신러닝의 훈련 집

합으로 사용하여 예측한 $n+1$ 일의 에너지 사용량 벡터를 $\hat{\mathbf{x}}_{n+1}$ 라 하면, 벡터 $\hat{\mathbf{x}}_{n+1}$ 는 다음과 같이 표현된다.

$$\hat{\mathbf{x}}_{n+1} = (\hat{y}_{n+1}(1), \hat{y}_{n+1}(2), \dots, \hat{y}_{n+1}(m)) \quad (3)$$

본 논문에서는 예측의 정확도를 평가하기 위해 평균 제곱근 편차(RMSE)를 사용하였으며 각 시간 간격 t 에서 예측한 에너지 사용량과 실제 에너지 사용량을 통해 구한다.

$$RMSE(\mathbf{x}_{n+1}, \hat{\mathbf{x}}_{n+1}) = \frac{1}{m} \sqrt{\sum_{t=1}^m |y_{n+1}(t) - \hat{y}_{n+1}(t)|^2} \quad (4)$$

변동 폭이 큰 데이터를 기반으로 일반적인 머신러닝을 통해 $\hat{\mathbf{x}}_{n+1}$ 을 예측할 때, 예측의 정확도가 떨어지는 경향이 있다[4], [5]. 본 논문에서는 이를 해결하기 위해 데이터 전처리와 제안 방법을 통해 파동이 큰 데이터에서도 예측의 성능을 높이고자 한다.

B. 제안 방법

본 논문에서는 지수 이동 평균으로 잡음을 제거한, 하루 동안 1분 간격으로 총 1,440번($m=1,440$) 측정된 하루 동안의 에너지 사용량 데이터 $\mathbf{x}_i$ 를 사용하여 다수의 머신러닝 모델을 병렬로 학습시킨 뒤, 학습된 모델의 예측 성능을 기반으로 한 가중 모델을 제안한다. 본 논문에서 제안하는 모델은 그림 1과 같다. 하나의 머신러닝 모델을 M^i 라 하면, M^i 는 $\{\mathbf{x}_i, \mathbf{x}_{i+1}\}$ 를 기반으로 i 일을 입력하였을 때 $i+1$ 일의 하루 에너지 사용량을 예측하도록 학습한다. $n+1$ 일을 예측하기 위해 활용되는 다수의 병렬 머신러닝 모델들은 $n-1$ 일 동안 수집된 데이터 집합 $\mathbf{X}_{n-1}$ 을 $\{\mathbf{x}_i, \mathbf{x}_{i+1}\}$ ($i=1, \dots, n-2$)로 나누어 병렬로 학습하는 과정을 통해 총 $n-2$ 개 생성된다.

제안하는 모델은 학습된 M^i 에 $\mathbf{x}_{n-1}$ 을 입력하였을 때 n 일에 대해 예측한 성능을 기반으로 가중치를 정한다. 머신러닝 모델 M^i 가 n 일에 대하여 예측한 결과를 에너지 사용량 벡터 $\hat{\mathbf{x}}_n^i$ 라 하면, $\hat{\mathbf{x}}_n^i$ 와 실제 n 일에 대한 데이터 $\mathbf{x}_n$ 을 통해 각각의 모델 M^i 에 대한 $RMSE(\mathbf{x}_n, \hat{\mathbf{x}}_n^i)$ 인 e^i 를 계산한다. 편차와 예측 성능은 반비례하므로 이 편차의 역수로 가중치 w^i 를 얻게

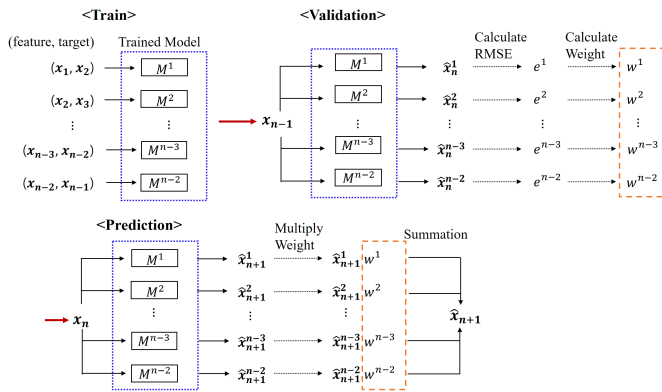


그림 1. 병렬 생성 모델을 활용한 제안 방법.

되며 가중치 w^i ($i=1, \dots, n-2$)는 다음과 같이 표현된다.

$$w^i = \frac{1/e^i}{\sum_{i=1}^{n-2} 1/e^i} \quad (5)$$

제안 모델은 병렬로 학습된 각각의 머신러닝 모델 M^i 에 x_n 데이터를 입력해 $n+1$ 일을 예측한 $\hat{x}_{n+1}^i$ 를 추출하고 M^i 마다 계산된 가중치 w^i 를 예측값 $\hat{x}_{n+1}^i$ 에 각각 곱한 후 더하여 $n+1$ 일에 대한 예측값 $\hat{x}_{n+1}$ 을 얻는다.

$$\hat{x}_{n+1} = \sum_{i=1}^{n-2} w^i \hat{x}_{n+1}^i \quad (6)$$

III. 구현 및 실험 결과

본 논문에서 제안한 모델의 성능을 평가하기 위하여 독일 쾰른-아헨 지역의 FINESCE 시범 스마트팩토리에서 1분 간격으로 수집한 4가지 기계 레이저 절단기(laser cutting), 벤딩기(bending), 3D 로봇 웰터(3D robot welter), 레이저 형상기(laser shaping)의 하루 동안 에너지 사용량 데이터를 사용하였다[6]. 결측값이 없는 데이터를 활용하기 위하여 2015년 5월 18일부터 6월 15일까지 4주 동안 수집된 데이터를 사용하여 에너지 사용량을 분 단위로 모두 더한 총 에너지 사용량 데이터를 활용하였다. 실험에서 사용한 머신러닝 모델은 Decision Tree, Adaboost 및 kNN(k-Nearest Neighbors)이며 본 논문에서는 각각을 DT, AB, kNN으로 나타낸다.

제안 방법의 병렬 모델 개수에 따른 예측 성능을 평가를 위해 n ($n=4.5 \dots, 17$)을 지정하여 데이터 길이에 맞춰 머신러닝 모델 M^i 를 각각 2, 3, ..., 15 개 생성하여 $n+1$ 일의 에너지 사용량을 예측하였다. 그림 2에서 실험에 사용한 3개의 머신러닝 모델에 대해 병렬 모델을 13개 사용하였을 때 예측 성능이 가장 우수하고 모델을 2개 사용한 경우보다 성능이 평균 14.41% 향상된 것을 확인할 수 있다. 이는 경향성이 없는 데이터를 학습할 때, 데이터를 많이 사용할수록 모델이 정확하게 학습하기 때문이다. 또한, 본 논문에서는 일반적인 머신러닝 방법과 제안 방법의 성능을 비교하기 위해 활용 데이터 길이 $n=15$ 로 하였으며, 제안 방법에서는 병렬 모델 개수를 13개로 지정하여 15일 동안 수집한 데이터로 16번째 날의 하루 에너지 사용량을 예측하였다. 표 1에서 임의의 4일에 대해 제안 방법이 일반적인 머신러닝 방법에 비해 AB 최대 2.2%, DT 최대 17.2%, kNN 최대 8.2%로 예측 성능이 향상된 것을 확인할 수 있다.

IV. 결론

본 논문에서는 변동폭이 큰 스마트팩토리 데이터의 예측 성능을 향상하기 위해 병렬로 생성된 머신러닝 모델의 예측 성능에 따른 가중치를 부여하여

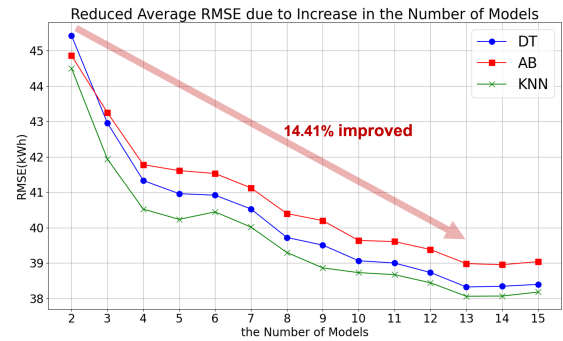


그림 2. 병렬 모델 개수에 따른 제안 방식의 예측 성능.

표 1. 직렬 학습 모델 1개, 병렬 모델 13개 예측 성능 비교

RMSE (kWh)	Day 1			Day 2			Day 3			Day 4		
Series	DT	AB	kNN	DT	AB	kNN	DT	AB	kNN	DT	AB	kNN
Proposed	40.10	40.84	39.90	42.81	43.26	42.50	40.37	40.92	40.27	30.03	30.93	29.58

최종 예측을 하는 다수의 병렬 가중 모델을 제시하였다. 실험을 통해 기존 방법보다 우수한 예측 성능을 확인하였으며 병렬 모델의 수가 증가할수록 학습에 사용된 데이터가 많아 성능이 더욱 향상되는 것을 확인할 수 있었다. 잡음이 있고 경향성이 없는 데이터에 대하여 제안 방법을 통한 예측을 통해 높은 예측 정확성을 보일 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2021-0-00739, 분산/협력AI 기반 5G+ 네트워크 데이터 분석 기능 및 제어 기술 개발, No. 2019-0-00024, 네트워크 자동화를 위한 개방형 네트워크 데이터 분석 기반 지도형 애자일 머신러닝 기술 개발)과 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1A2B5B01002528).

참고 문헌

- [1] 정병도, "스마트 팩토리 연구동향," TTA(한국정보통신기술협회) journal, 181호, pp. 46-51, Jan. 2019.
- [2] X. Xu and Q. Hua, "Industrial Big Data Analysis in Smart Factory: Current Status and Research Strategies," *IEEE Access*, Vol. 5, pp. 17543-17551, 2017.
- [3] F. Shrouf, J. Ordieres and G. Miragliotta, "Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm," *IEEE International Conf. on Industrial Engineering and Engineering Management*, pp. 697-701, 2014.
- [4] S. Gupta and A. Gupta, "Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review," *Procedia Computer Science*, Vol. 161, pp. 466-474, 2019.
- [5] G. Li, J. Zrimec, B. Ji, J. Geng, J. Larsbrink, A. Zelezniak, J. Nielsen, and M. K. Engqvist, "Performance of Regression Models as a Function of Experiment Noise," *Bioinformatics and Biology Insights*, Vol. 15, pp. 1-10, Jan. 2021.
- [6] *Smart Energy Data: Aachen / Cologne Smart Factory*, Aachen / Cologne, Germany, Aug. 25, 2015. [Online] Available: https://data.lab.fiware.org/dataset/smart_energy_data-aachen-cologne-smart-factory