

A Multilayer Perceptron-Based Electric Load Forecasting Scheme via Effective Recovering Missing Data

Jihoon Moon[†] · Sungwoo Park^{**} · Eenjun Hwang^{***}

ABSTRACT

Accurate electric load forecasting is very important in the efficient operation of the smart grid. Recently, due to the development of IT technology, many works for constructing accurate forecasting models have been developed based on big data processing using artificial intelligence techniques. These forecasting models usually utilize external factors such as temperature, humidity and historical electric load as independent variables. However, due to diverse internal and external factors, historical electrical load contains many missing data, which makes it very difficult to construct an accurate forecasting model. To solve this problem, in this paper, we propose a random forest-based missing data recovery scheme and construct an electric load forecasting model based on multilayer perceptron using the estimated values of missing data and external factors. We demonstrate the performance of our proposed scheme via various experiments.

Keywords : Smart Grid, Electric Load Forecasting, Missing Data Handling, Deep Learning

효과적인 결측치 보안을 통한 다층 퍼셉트론 기반의 전력수요 예측 기법

문지훈[†] · 박성우^{**} · 황인준^{***}

요약

정확한 전력수요 예측은 스마트 그리드의 효율적인 운영에 있어 매우 중요하다. 최근 IT 기술이 획기적으로 발전되면서, 인공지능 기법을 이용한 빅 데이터 처리를 기반으로 정확한 전력수요를 예측하는 많은 연구가 진행되고 있다. 이러한 예측 모델은 주로 외부 요인과 과거 전력수요를 독립 변수로 사용한다. 하지만, 다양한 내부적 또는 외부적 원인으로 전력수요 데이터의 결측치가 발생하게 되면 정확한 예측 모델을 구성하기가 어렵다. 이에 본 논문에서는 랜덤 포레스트 기반의 결측치 데이터 보완 기법을 제안하고, 보완된 데이터를 기반으로 한 다층 퍼셉트론 기반의 전력수요 예측 모델을 구성한다. 다양한 실험을 통해 제안된 기법의 예측 성능을 입증한다.

키워드 : 스마트 그리드, 전력수요 예측, 결측치 보완, 심층 학습

1. 서론

최근 스마트 그리드(Smart Grid)는 전 세계적으로 문제가 되는 온실가스 배출 및 에너지 부족 문제 등에 대한 해결책으로 많은 주목을 받고 있다[1-5]. 스마트 그리드는 ICT(Information and Communication Technology)를 기존 전력 그리드에 결합하여 전력 공급 업체와 소비자 간에 실시간 정보를 교환함으로써, 에너지 효율성을 최적화하는 차세대 지능형 전력 그

리드이다. 구체적으로, 스마트 그리드의 핵심 구성요소 중 하나인 에너지관리시스템(Energy Management System, EMS)은 냉방기기 사용량, 조명 사용량 등과 같은 스마트 그리드 시스템 내에서 에너지 소비와 관련된 데이터를 분석하여 수요 측면에서 에너지 절감을 위한 방안을 찾아 에너지를 절약한다[5]. 또한, 공급 측면에서는 향후 에너지 소비량에 대한 예측을 수행하여, 예측치를 바탕으로 전력 생성(Power Generator) 및 에너지저장장치(Energy Storage System) 등에 대한 효율적인 스케줄링을 수행하여 에너지 생산 및 저장 비용을 절감한다[4-7]. 이로 인해, 최근 ESS와 신재생에너지의 시장이 빠른 속도로 성장하고 있다[6]. 더욱이 ESS를 기반으로 스마트 그리드의 효율적인 운영을 위해서는 정확한 전력수요 예측이 필수적으로 요구되며[4], 이를 바탕으로 전기요금에 저렴한 밤에 전기를 충전하고, 전기요금이 비싼 낮에 방출하는 식으로

* 이 연구는 한국전력공사의 2018년 착수 에너지 거점대학 클러스터 사업에 의해 지원되었음(R18XA05).

† 준회원: 고려대학교 전기전자공학과 석·박사통합과정

** 준회원: 고려대학교 전기전자공학과 석사과정

*** 종신회원: 고려대학교 전기전자공학과 교수

Manuscript Received: October 5, 2018

Accepted: October 29, 2018

* Corresponding Author: Eenjun Hwang(ehwang04@korea.ac.kr)

경제적 이득을 기대할 수 있다. 더 나아가 스마트 그리드 시장이 확산된다면 전기를 생산하는 발전사에서도 많은 양의 온실가스를 발생시키는 화석연료 기반 발전소와 비교적 생산 원가가 높은 오래된 발전소의 전기 생산을 후 순위로 적용하여 효율적인 전기 생산이 가능하게 된다[8, 9].

전력수요 예측은 예측 범위에 따라 크게 네 가지로 분류할 수 있다[10]. 먼저, 예측이 하루 이하의 범위에서 수행될 때, 이를 초단기 전력수요 예측(Very Short-Term Load Forecasting)이라 한다. 또한, 하루에서 2주의 범위에 대한 전력수요 예측을 단기 전력수요 예측(Short-Term Load Forecasting)이라 한다. 일반적으로 초단기 전력수요 예측과 단기 전력수요 예측을 통틀어 단기 전력수요 예측이라 한다. 예측 범위가 2주에서 3년 내의 전력수요 예측은 중기 전력수요 예측(Mid-Term Load Forecasting)이라고 하며, 예측 범위가 3년 이상이면 이를 장기 전력수요 예측(Long-Term Load Forecasting)이라고 한다. 여기서, 하루 1시간 이하의 예측을 수행하는 VSTLF 또는 STLF은 일일 스마트 그리드 스케줄링을 위해 필수적인 절차이며, 최대수요전력 대응을 포함한 경제적 이득 발생을 기대할 수 있다. 그러나 단기 전력수요 예측은 건물의 복잡한 전력수요 패턴과 불확실한 외부 요인들로 인한 수요의 잦은 변동으로 인해 정확하게 예측하는 것은 쉽지 않다[2]. 이러한 변동에 영향을 미치는 요인으로 건물의 용도, 시간대, 전기요금 체계 및 특별 행사, 거주자의 일정, 기후 조건(기온, 습도 등), 조명 또는 공조(Heating, Ventilation and Air Conditioning) 시스템과 같은 하위 수준의 시스템 구성 요소들이 있다[11]. 또한, 전력수요 예측은 현재 시각과 이전 시각의 전력수요량 간의 밀접한 상관관계를 보이므로 이를 효과적으로 고려해야 한다. 그리하여 정확한 전력수요 예측을 위해, 기존의 통계적 기법, ANN(Artificial Neural Network)나 SVR(Support Vector Regression) 등을 이용한 기계 학습, 그리고 최근 많은 분야에서 우수한 성능을 보인 심층 학습(Deep Learning)으로 총 3가지의 방법론을 이용하여 많은 연구가 진행되었다[12].

하지만 블랙아웃, 데이터의 송수신 과정에서 데이터 손실, 전기 계량기의 고장 등으로 인해, 전력 데이터의 결측치가 발생할 수 있다. 결측치 구간이 짧으면 선형 보간법(Linear Interpolation)을 통해 어느 정도 보완이 가능하지만, 결측치 구간이 길면 선형 보간법으로는 외부 요인과 패턴을 충분히 반영하지 못하기 때문에, 이러한 값을 반영한 예측 모델은 정확한 전력수요 예측을 기대하기 어렵다[13]. 앞서 기술한 문제점을 해결하기 위해, 본 논문은 예측 모델을 학습하기 위한 입력 변수로 먼저 전력수요 시점이 설명 가능한 외부 요인으로 구성한다. 그리하여, 비선형 회귀 분석에 효과적인 알고리즘으로 ANN, SVR, RF(Random Forest)를 이용하여 전력수요 예측 모델을 학습시키고, 10겹 교차검증(10-Fold Cross-Validation)을 통하여 가장 작은 오차율을 갖는 알고리즘을 검출한다. 검출된 예측 모델은 결측치 구간에 적용하여 전력수요를 추정한다. 그리하여 외부 요인과 결측치의 추정값을 입력 변수로 구성하여, 1층부터 4층의 은닉층으로 구성된 다층 퍼셉트론 기반의 예측 모델을 설계하고 최적의 단기 전력수요 예측 모델을 확인한다. 본 논문에서 제안한 예측 모델은 다양한 결측치 보

완 방법을 이용하여 구성된 예측 모델뿐만 아니라 다른 기계 학습 기반의 예측 모델과 선행 연구의 전력수요 예측 모델[14, 15]와의 비교를 통해 제안한 목적의 당위성을 입증하고자 한다.

본 논문의 나머지 구성은 다음과 같다. 2장에서는 단기 전력수요 예측에 대한 관련 연구를 기술한다. 3장은 예측 모델을 구성하기 위한 전처리 단계에 관해 기술한다. 4장은 결측치 추정 모델과 이를 통한 다층 퍼셉트론 기반의 전력수요 예측 모델 구성에 관해 설명한다. 5장은 실험 설계 및 예측 모델에 대한 실험 결과에 대해 구체적으로 기술한다. 마지막으로 6장에서는 본 논문의 실험 결과에 대한 평가 및 향후 연구 방향의 제시와 함께 결론을 맺는다.

2. 관련 연구

이 장에서는 스마트 그리드 에너지관리시스템의 효율적인 운영을 위해 수행되었던 단기 전력수요 예측에 관한 연구 사례를 소개한다. 전력수요 예측을 위하여 과거부터 통계적 접근과 기계 학습[16-18]을 이용한 많은 연구가 수행되어 왔으며, 최근에는 심층 학습 기반의 연구[19]가 예측 성능을 한층 발전시켰다.

2.1 통계적 접근

통계적 접근은 독립 변수와 종속 변수 간의 상관관계나 인과관계를 파악하기 쉽다는 장점을 가진다. [20]은 온도와 연간 인구수 및 과거 전력 데이터를 이용한 퍼지선형회귀 분석 기반의 전력수요 예측 모델을 구성하였다. 이는 통상적인 통계 예측 방법(이동평균법, 지수평활법, ARIMA 등)에서 과거 전력수요량만 적용할 경우 외부 요인을 반영하지 못하여 다소 예측 정확도가 낮은 한계를 개선하였다. [21]은 하루의 전력수요 주기 패턴과 선형회귀 분석을 이용한 예상 출력 패턴 등을 반영하여 단계적(Stepwise) 회귀와 라소(Lasso) 회귀 분석 기반의 예측 모델을 구성하였다. [11]은 대학 캠퍼스의 전력수요 특징을 분석하기 위해, 시간, 온도, 습도 등과 같은 전력수요와 밀접한 관련을 갖는 환경 요인들을 고려하여 Curve Fitting, Similar Day, MR(Multiple Regression)을 이용하여 15분 단위의 초단기 전력수요 예측 모델을 구성하였다. [22]는 과거의 전력수요와 온도를 입력 변수로 구성하여 양자 평활 스플라인 회귀 분석(Quantile Smoothing Spline Regression)에 기반을 둔 전력수요 예측 모델을 구성하였다. 이러한 통계적 접근 방식들은 단순한 수요 패턴에는 우수한 성능을 보이지만, 복잡한 수요 패턴에는 다소 부정확한 예측 성능을 보인다[5, 17]. 또한, 선형회귀 분석의 입력 변수들이 명목 척도나 비선형 상관관계를 갖는다면, 각 변수의 적절한 가중치를 부여하기 어렵다[17]. 이러한 통계적 접근 방식의 단점을 극복하기 위하여, SVR, ANN 등과 같은 기계 학습을 이용한 전력수요 예측 모델이 우수한 예측 성능을 산출하였다[17, 23].

2.2 기계 학습

기계 학습에서 지도 학습은 훈련 데이터(Training Data)로부

터 하나의 함수를 유추해내는 방법이다. STLF를 위해 전력수요와 밀접한 관계가 있는 다양한 입력 변수들을 구성하여 기계 학습 기반의 예측 모델들이 제안되었다. [24]는 SVR을 이용하여 건물 에너지 예측 모델을 개발하였으며, 뉴욕시 컬럼비아 대학 캠퍼스에 위치한 다가구 주택의 전력수요를 예측하였다. [25]는 이탈리아에 위치한 사무용 건물 에너지와 슬로바키아 동부 지역의 에너지를 예측하기 위해, EMD(Empirical Mode Decomposition)와 SVR을 이용하여 예측 모델을 개발하였다. [26]은 기존의 외부 요인에 이벤트 정보를 추가한 ANN과 SVR을 이용하여 캐나다 대형 엔터테인먼트 건물의 단기 전력수요 예측 및 알고리즘의 예측 정확성을 비교하였다. [27]은 RF, ANN, FIR(Fuzzy Inductive Reasoning), ARIMA(Autoregressive Integrated Moving Average) 등의 다양한 기계 학습 알고리즘을 통해 바르셀로나에 있는 3개의 건물의 시간대별 전력수요 예측 및 알고리즘의 예측 정확성을 비교하였다.

컴퓨터의 성능이 크게 향상됨에 따라, ANN 기반의 전력수요 예측 기법들이 제안되었다. [28]은 ANN을 기반으로 토리노의 첼리니 병원 건물의 전력수요를 예측하였으며, [29]는 정확한 STLF를 위해 ELM(Extreme Learning Machine)과 SDPSO(Switching Delayed Particle Swarm Optimization) 알고리즘을 결합한 하이브리드 학습 방법을 제안하였다. [3]은 매개 변수의 최적화 프로세스가 간편한 SRWNN(Self-Recurrent Wavelet Neural Network)을 이용하여, 용도가 다른 두 건물의 시간대별 전력수요와 과거 24시간 기준 최대전력수요 예측 정확도를 MLP(Multilayer Perceptron), WNN(Wavelet Neural Network)과 비교하였다. [30]은 온도 및 습도와 같은 기상 요소를 고려한 STLF를 위해 Bat 알고리즘 기반의 역전파 접근법을 제안하였으며, ANN 기반의 예측 모델보다 학습 시간이 현저히 줄어들어 효율적인 예측 방법인 것을 입증하였다.

2.3 심층 학습

최근 IoT(Internet of Things)와 하드웨어의 발전으로 인해 다양한 연구 분야에서 심층 학습을 사용하고 있으며[7, 31, 32], 전력수요 예측에서도 심층 학습을 적용하여 정교한 예측 모델을 구성한 연구들이 제안되었다. [7]은 RBM(Restricted Boltzmann Machine)과 사전 학습 없이 ReLU(Rectified Linear Unit)으로 구성된 2가지의 DNN(Deep Neural Network) 기반 전력수요 예측 모델을 구성 및 비교하였다. [33]은 1D-CNN(Convolutional Neural Network) 및 Pooling Layer를 통한 전력수요 예측 모델을 제안하여 전통적인 기계 학습 알고리즘과 MLP의 비교를 통해 전력수요 예측을 수행하였다. [31]은 다섯 가지의 LSTM(Long Short-Term Memory) 기반의 예측 모델을 구성하여 PV(Photovoltaics) 출력에 대한 예측 정확도 성능을 비교 및 평가하였다. 또한, [32]는 집적 부하(Load on System Level)와 개별 부하(Loads in Individual Households) 간의 불일치 평가 및 비교를 위한 밀도 기반 클러스터링을 이용하였다. 다음으로, 과거 전력수요, 시간 정보를 반영하여 거주자의 생활양식 패턴을 인식하여 LSTM 기반의 가구별 단기 전력수요 예측을 수행하였다.

하지만, 앞서 기술한 예측 모델들은 결측치를 보완하는 방법에 대해서는 주로 다루지 않았다. 몇 가지 연구에서는 결측치의 해결 방안으로 선형 보간법을 사용하였으며, 이는 장기 결측치 보완으로 수요 패턴을 반영할 수 없기 때문에 적합하지 않다. 또한, 본 논문은 일일 스마트 그리드 운영 스케줄링을 위한 목적이기 때문에, 예측 시점의 하루 전(Day-Ahead)에 전력수요를 예측해야 한다. 그리고 한국의 전기요금은 RTP(Real-Time Pricing)이 아닌 TOU(Time Of Use)와 CPP(Critical Peak Pricing)로 이루어져 있다. 그로 인해 학습된 정보를 이전의 정보와 함께 반영하여 이후의 정보를 활용하는 LSTM을 이용하는 것은 적합하지 않다. 이에 본 논문은 기계 학습 알고리즘을 적용하여 결측치를 효과적으로 보완하고, 심층 학습인 MLP 모델을 기반으로 스마트 그리드 환경에서 실현 가능한 단기 전력수요 예측 모델을 구성하는 것을 목표로 한다.

3. 데이터 수집 및 전처리

본 논문은 한진 i-Smart를 통해 2015년 1월 1일부터 2016년 12월 31일까지 인천에 위치한 “K” 공장의 30분 단위 전력 사용량 데이터를 수집하였다. 수집한 데이터를 통해 예측 모델을 구성하기 위해서는 전력수요와 밀접한 관계를 갖는 외부 요인들을 고려해야 한다. 또한, 외부 요인을 기계 학습 알고리즘의 학습에 맞게 전처리 과정이 필요하다. 본 논문의 단기 전력수요 예측 모델링의 구조는 Fig. 1과 같다. 먼저, 기상청과 한국전력공사를 통해 전력수요와 높은 상관관계를 갖는 외부 요인들을 수집하고, 이를 통해 전력수요 추정 모델을 구성하기 위한 전처리 과정을 수행한다. 이는 다음 절에 더욱 자세하게 기술한다. 다음으로, 기계 학습 및 심층 학습 기반의 단기 전력수요 예측 모델링은 다음 장에 더욱 자세하게 기술한다.

3.1 날짜 및 시간 정보

전력 사용량 데이터는 시계열 데이터이므로 해당 시점에 속한 날짜 및 시간 정보를 가지고 있다. 또한, 날짜 및 시간 정보는 전력수요의 트렌드를 나타내기 때문에, 이를 표현할 수 있는 모든 변수들을 고려해야 한다. 본 논문에서 고려한 날짜 및 시간 정보는 월, 일, 시, 분, 요일, 공휴일이다.

월, 일, 시, 분은 연속성을 가지지만, 수열의 형태를 보이므로 주기성을 반영하기 어렵다. 예를 들어, 어제의 23시와 오늘의 0시는 실제로 인접한 정보이지만, 수열로 표현할 때 23의 차이를 갖게 된다. 따라서 본 논문에서는 Equations (1)~(7)까지의 식을 통해 수열 형태의 1차원 데이터를 2차원 공간의 연속적인 데이터로 변환하여 주기성을 반영한다[14]. 분의 경우, 0과 30이라는 두 가지 경우에만 존재하므로, Equation (1)과 같이, 해당 시간에 시와 분 모두 반영할 수 있도록 구성한 뒤, Equations (2), (3)에 적용한다. EoM_{month} 는 End of Month로 해당 달의 마지막 날을 나타낸다. 예를 들어, 1월은 31, 2월은 28/29, 3월은 31 등이다. 그리하여 Equations (2)~(7)까지 총 6개의 입력 변수를 구성하여 예측 시점의 날짜 및 시간을 표현한다.

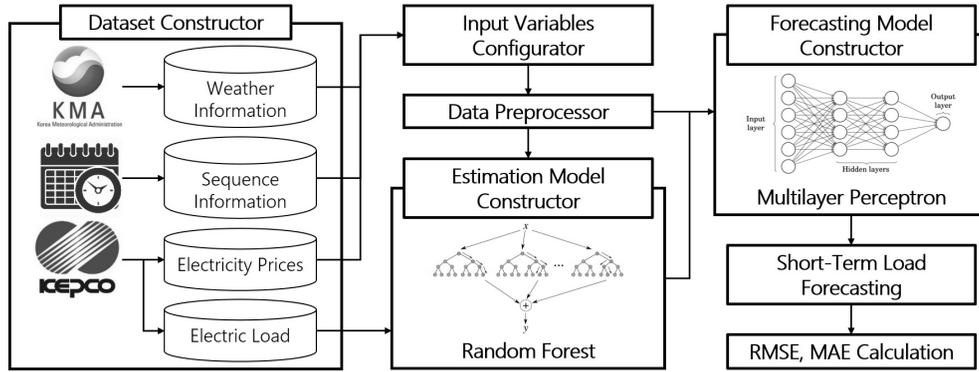


Fig. 1. Our Framework for Short-Term Load Forecasting

$$hour_{min} = hour + (minute/60) \quad (1)$$

$$hour_x = \sin((360/24) \times hour_{min}) \quad (2)$$

$$hour_y = \cos((360/24) \times hour_{min}) \quad (3)$$

$$day_x = \sin((360/EOM_{month}) \times day) \quad (4)$$

$$day_y = \cos((360/EOM_{month}) \times day) \quad (5)$$

$$month_x = \sin((360/12) \times month) \quad (6)$$

$$month_y = \cos((360/12) \times month) \quad (7)$$

대부분의 공장 가동은 평일에 운영이 이루어지며, 토요일, 일요일, 국경 공휴일 등과 같은 공휴일에는 운영하지 않으므로, 일반적으로 평일과 공휴일은 서로 다른 수요 패턴을 보인다. 이러한 요소를 반영하기 위해, 명목 척도인 요일과 공휴일 정보를 예측 모델에 반영한다. 이를 위해, 7가지의 요일은 해당 요일만 1이 할당되도록, 공휴일 정보 반영에서 평일은 0, 공휴일은 1을 갖도록 명목 척도로 구성한다. 그리하여 본 논문에서는 예측 시점을 나타낼 수 있는 날짜 및 시간 정보를 총 14개를 사용한다.

3.2 기상 정보

에어컨, 난방기 등과 같이 많은 전기가 필요로 하는 제품들의 사용은 기온이나 습도와 같은 기상 요인에 큰 영향을 갖는다. 그리하여 기상 정보는 전력수요 예측의 많은 연구에서 필수적으로 고려되었다[10]. 기상청의 동네예보는 국내 지역별 날짜, 기상, 기온, 풍속, 습도 등을 제공한다. 본 논문에서는 기상청에서 제공하는 종관기상관측자료를 사용한다. EMS의 실현 가능성을 높이기 위해, 동네예보를 고려하여 온도, 풍속, 습도로 총 3가지의 기상 변수를 사용한다. 또한, 전력수요와 연관성 높은 변수를 추가하기 위해, Equations (8), (9)를 적용하여 체감온도와 불쾌지수를 산출한다. 그리하여 총 5가지의 기상 정보를 입력 변수로 반영한다.

$$WC = 13.12 + 0.6215 \times T - 11.37 \times WS^{0.16} + 0.3965 \times T \times WS^{0.16} \quad (8)$$

$$DI = 1.8 \times T - 0.55(1.8 \times T - 26) \times (1 - 0.01 \times H) + 32 \quad (9)$$

여기서, T 는 온도(Temperature), H 는 습도(Humidity), WS 는 풍속(Wind Speed), WC 는 체감온도(Wind Chill), DI 는 불쾌지수(Discomfort Index)를 의미한다.

3.3 전기요금 정보

스마트 그리드 운영의 목적 중 하나는 전기요금의 감소로 인한 경제적 이득 발생이므로[5, 6], 다양한 전력수요 예측 연구에서 전기요금을 입력 변수로 고려하였다[17, 18]. 이에 본 논문은 전기요금에 대한 정보를 입력 변수로 고려한다. 한편에서는 건물의 용도에 따라 계절(월) 및 시간대에 따라 경부하, 중간부하, 최대부하로 총 3가지 단계의 각기 다른 전기요금을 책정한다. 공장이 속한 산업용의 시간대에 따른 전기요금 정보는 Table 1과 같다. 앞서 기술한 요일, 공휴일 정보와 같이 해당 시점에 대해 해당하는 시간은 1, 그렇지 않으면 0을 갖도록 총 3가지의 입력 변수를 구성하였다.

Table 1. Season & Time-Period Classification

Classification	March ~ October	November ~ February
Off-peak	23:00 ~ 09:00	23:00 ~ 09:00
Mid-peak	09:00 ~ 10:00	09:00 ~ 10:00
	12:00 ~ 13:00	12:00 ~ 17:00
	17:00 ~ 23:00	20:00 ~ 22:00
On-peak	10:00 ~ 12:00	10:00 ~ 12:00
	13:00 ~ 17:00	17:00 ~ 20:00
		22:00 ~ 23:00

4. 단기 전력수요 예측 모델링

이 장에서는 기계 학습 및 심층 학습을 이용한 단기 전력수요 예측 모델에 대해 설명한다. 먼저, 앞서 구성한 22개의 외부 요인들을 Equation (10)과 같이 최소-최대 정규화(Min-Max Normalization) 과정을 통해 추정 모델 및 예측 모델을

위한 입력 변수로 반영한다.

$$z = \frac{x - \min(x)}{[\max(x) - \min(x)]} \quad (10)$$

다음으로, 결측치 보안에 강인한 기계 학습 기법인 랜덤 포레스트[34]를 기반으로 비선형 회귀 분석 모델을 구성한 다음 해당 결측치 부분에 대한 전력수요를 추정한다. 이는 4.1절에 더욱 자세히 기술한다. 다음으로, 전력수요 패턴을 반영하기 위해, 앞서 수행한 결측치의 추정값을 포함하여 예측 시점에 근접한 과거 전력수요량과 공휴일 여부에 대한 새로운 입력 변수를 반영한다. 이는 4.2절에 자세하게 기술한다. 마지막으로, 새로운 입력 변수와 예측 시점을 나타낼 수 있는 외부 요인을 이용하여 심층 학습 기법 중 하나인 다층 퍼셉트론을 기반으로 단기 전력수요 예측 모델을 구성한다. 이는 4.3절에 자세히 기술한다.

4.1 랜덤 포레스트 기반의 전력수요 추정 모델링

결측 데이터는 사회과학, 자연과학 등의 다양한 분야에서 관측 또는 실험하여 얻어지는 데이터에 종종 발생하는 실제 문제이다. 전기, 전자, 컴퓨터 및 통신 분야에서도 결측치는 다양한 요인들로 인해 발생할 수 있으며, 이는 전력 사용량 데이터를 관리하는 한전에서도 여러 가지 이유로 인해 전력 데이터의 결측치가 발생할 수 있다. 결측 데이터는 다양한 통계 분석에서 완전한 데이터 집합을 요구하기 때문에 이를 사전에 효율적으로 처리해야 한다. 특히, 전력수요와 같은 시계열 데이터는 인접 데이터 간의 상관관계가 매우 높으며, 중간에 결측 데이터가 생기면 해당 시기의 변화를 탐지하지 못하는 문제가 발생한다[34].

랜덤 포레스트는 의사결정나무(Decision Tree)의 앙상블 학습 형태이다. 랜덤 포레스트의 장점은 대규모 데이터에서 실행 가능하며, 변수를 제거하지 않고 많은 변수를 다룰 수 있기 때문에 높은 정확도를 제공한다[5, 35]. 또한, ANN 및 SVR과 비교하여 초매개변수(Hyperparameter)의 세부 조정(Tuning)이 간편하다. 이러한 이유로 랜덤 포레스트를 이용하여 결측치 보안을 해결하고자 하는 연구가 수행되었다. [34]는 랜덤 포레스트가 통계적 기법보다 다변량 및 대규모 데이터에서 발생하는 변수의 복잡한 상호작용(Interaction)과 비선형 회귀 분석을 통한 결측치 해결에 더욱 효과적임을 나타내었다. 또한, [13]은 날짜/시간과 기상 정보만을 고려한 랜덤 포레스트 기반의 단기 전력수요 추정 모델을 제안하였으며, ANN, SVR, DT, GBM(Gradient Boosting Machine) 등의 기계 학습 기법과 비교하여 결측치 보안에 더욱 적합하다는 것을 입증하였다. 이에, 본 논문은 랜덤 포레스트를 기반으로 전력수요의 결측치 데이터 추정 모델을 구성한다. 랜덤 포레스트의 초매개변수는 무작위 입력 변수 선택(Number of Observations)과 하위 나무의 개수(Number of Trees)가 있다. 전력수요 추정 모델링을 위해, 앞서 기술한 22개의 입력 변수를

이용하여 결측치가 제외된 데이터 집합을 구성한다. 다음으로, 10겹 교차검증을 통하여 최적의 초매개변수의 값을 검출하고, 이를 적용한 전력수요 추정 모델을 구축한다. 그리하여, 전력수요 추정 모델에서 결측 데이터의 입력 변수를 적용하여 도출되는 값을 전력수요 추정값으로 사용하며, 이를 결측치에 보완한다.

4.2 과거 데이터를 이용한 전력수요 패턴 반영

전력수요는 시계열 데이터이기 때문에, 주/월/년의 변화보다 하루 단위로 변하는 특성이 더 크다. 그리하여 과거의 인접 데이터 간의 상관관계가 매우 높으며 이러한 패턴을 효율적으로 반영할 수 있어야 한다. 최근 LSTM이 시계열 데이터 처리에서 우수한 성능을 입증하였으나, 본 논문은 스마트 그리드의 일일 스케줄링 운영을 목표로므로, 예측 시점의 하루 전에 예측 시점의 전력수요를 예측해야 한다. 그렇기 때문에 학습된 정보를 이전의 정보와 함께 반영하여 이후의 정보를 활용하는 LSTM은 적합하지 않다.

본 논문은 전력수요의 패턴을 학습하기 위해, 예측 시점에서 근접한 과거 전력수요량을 반영한다. 단기 전력수요를 예측할 때, 예측일과 인접한 과거 전력수요를 예측 모델의 입력 변수로 이용하게 될 경우에 예측 정확도가 향상된다. 하지만 다수의 과거 전력수요량을 입력 변수로 반영하게 될 경우에는 학습의 과적합(Overfitting)과 예측 모델 구성을 위한 훈련 시간이 지나치게 길어질 수 있다는 단점이 있다. 그리하여, 본 논문에서는 예측 시점부터 3일 전까지의 과거 정보만을 고려하여 입력 변수로 이용한다[36]. 여기서 예측하고자 하는 시점부터 예측 시점까지는 미래 시점으로 전력수요량을 반영할 수 없다. 스마트 그리드의 실현 가능성을 위해, 예측 시점의 하루 전(T-48)부터 2일 전(T-143)까지의 모든 전력수요와 공휴일 정보를 반영한다. 예를 들어, 6월 10일 오전 2시 30분이 지난 직후라면, 예측 시점은 6월 11일 오전 2시 30분의 전력수요이며, 이를 위해 6월 8일 오전 3시부터 6월 10일 오전 2시 30분까지 총 96개의 전력수요량을 모두 반영한다. 또한, 예측 시점의 하루 전(D-1)과 2일 전(D-2)의 공휴일 정보로 2개의 입력 변수를 포함하여 총 98개의 변수를 최소-최대 정규화 과정을 통해 단기 전력수요 예측 모델을 구성한다.

4.3 다층 퍼셉트론 기반의 전력수요 예측 모델링

MLP는 FFNN(Feed-Forward Neural Network) 구조 중 하나이며, 입력층, 은닉층, 출력층으로 총 3가지의 계층으로 구성되어 있다. 최근 2개 이상의 Hidden Layer로 구성된 MLP가 다른 기계 학습 기법보다 정확한 전력수요를 예측하였다[1, 2, 7].

본 논문에서 사용하는 MLP의 구조는 Fig. 2와 같다. X는 전력수요 예측 시점에 해당하는 외부 요인을 의미하며, Y는 예측 시점의 전력수요 예측값이다. MLP 기반의 전력수요 예측 모델을 구성하기 위해서는 여러 가지 요소를 고려해야 한다. 대표적으로 은닉층의 수, 은닉층에 포함된 노드의 수, 활성화 함수(Activation Function) 등이 있다.

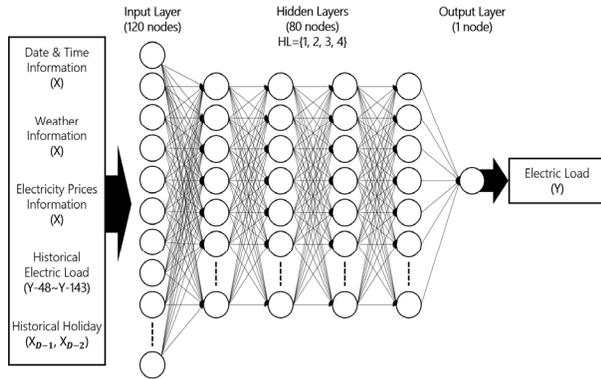


Fig. 2. Architecture of Multilayer Perceptron for Our Forecasting Model

일반적으로 MLP의 은닉층 수가 증가하면 우수한 성능을 기대할 수 있다. 하지만 학습 과정에서 과적합 문제가 자칫 발생할 수 있으므로, 지나치게 은닉층의 수를 늘리는 것은 적절하지 않다. 은닉층의 수는 ANN 또는 SNN(Shallow Neural Network)의 구성인 1층부터 DNN의 구성인 2층에서 4층까지의 모델을 각기 구성하여 각 예측 모델에 대한 예측 정확성을 통해 최적의 MLP 구조를 검증한다. MLP 모델의 노드 구성에서 입력층은 120개의 노드로 구성되어 있으며, 은닉층은 입력층의 노드 수와 출력층의 노드 수를 더하여 2/3를 적용해 층마다 80개의 노드로 구성하였다[37]. 마지막으로, 출력층은 1개의 노드(해당 시점의 전력수요량)로 구성하였다. 은닉층에서 사용할 활성화 함수로는 ReLU(Rectified Linear Unit)를 사용하였다. ReLU는 일반적으로 ANN에서 쓰이는 sigmoid, tanh 함수 등과 비교하였을 때, 선형성을 가지고 있으며, SGD(Stochastic Gradient Descent)의 수렴 속도를 빠르게 만들어준다는 장점이 있어, 많은 연구에서 주로 쓰이는 활성화 함수이다. MLP 모델의 구조를 구성한 다음 학습에 고려해야 하는 초매개변수로 Learning Rate와 Epoch 등이 있다. Learning Rate가 큰 경우에는 Training Epoch을 작게 설정하기 때문에 학습 시간은 짧지만 최적의 결과가 나오지 않을 가능성이 있다. 또한, Learning Rate가 작은 경우에는 최적의 해를 찾을 수 있다는 장점이 있지만 Training Epoch 당 Weight의 변화량이 적기 때문에 학습에 걸리는 시간이 길다. 본 논문은 결측치 보완에 대한 예측 성능을 더욱 정밀하게 평가하기 위해, Learning Rate는 0.000001, Epoch은 10,000으로 충분한 학습이 가능한 모델로 구성하였다. 그 외에 MLP 모델 구성으로 Batch Size는 144, Optimizer는 RMSProp[38]을 사용하였으며, 손실 함수는 MSE(Mean Square Error)보다 이상치에 덜 민감한 Huber Loss[39]를 사용하였으며, $\delta=1$ 로 설정하였다. Huber Loss의 식은 Equation (11)과 같다.

$$L_{\delta}(y, f(x)) \begin{cases} \frac{1}{2}(y-f(x))^2 & \text{for } |y-f(x)| \leq \delta, \\ \delta(|y-f(x)| - \frac{1}{2}\delta^2) & \text{otherwise.} \end{cases} \quad (11)$$

5. 실험 및 결과

본 연구에서는 구성된 예측 모델의 성능을 평가하기 위해 RMSE(Root Mean Square Error)와 MAE(Mean Absolute Error)를 이용하며, RMSE와 MAE의 수식은 Equations (12), (13)과 같다. 수식에서 A_t 는 t 시점의 실제 전력수요량, F_t 는 t 시점의 전력수요 예측값이며, n 은 관측치의 수이다.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (F_t - A_t)^2}{n}} \quad (12)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |F_t - A_t| \quad (13)$$

5.1 랜덤 포레스트를 이용한 결측치 보완

본 논문은 앞서 기술한 해당 시점을 나타낼 수 있는 외부 요인을 입력 변수로 구성하여 랜덤 포레스트 기반의 단기 전력수요 추정 모델을 구성한다. 랜덤 포레스트 모델링을 위해 주로 고려하는 초매개변수는 무작위 입력 변수 선택과 하위 나무의 개수이다. [40]은 최적의 하위 나무의 개수로 64부터 128까지 구성하는 것을 제안하였다. 이는 128개를 초과하였을 때, 랜덤 포레스트의 성능은 과적합으로 대부분 좋아지지 않았으며, 때에 따라 더 악화시킬 수 있다고 기술하였다. 따라서 결측치 추정 모델의 구성에서 하위 나무의 개수를 128개로 선정하였다. 무작위 입력 변수 선택에서 Python의 scikit-learn[41]은 아래와 같은 옵션을 제공한다.

- Auto: max_features=n_features
- Sqrt: max_features=sqrt(n_features)
- Log2: max_features=log2(n_features)

본 논문은 무작위 입력 변수 선택의 3가지 옵션을 모두 고려하여 10겹 교차검증을 수행하였다. 또한, 제한한 기법의 당위성을 입증하기 위해 비교 대상으로 SVR, ANN을 함께 구성하여 실험을 진행하였으며, Grid Search를 통해 구성된 SVR, ANN의 초매개변수는 Table 2와 같이 선정하였다. SVR은 scikit-learn, ANN은 TensorFlow[42] 라이브러리를 사용하여 각각 실험을 진행하였다.

Table 2. SVR/ANN Model Configuration

Algorithm	Hyperparameters
Support Vector Regression	Kernel: RBF kernel Cost: 1 Epsilon: 0.1 Gamma: 0.125
Artificial Neural Network	Number of hidden nodes: 15 Number of hidden layers: 1 Activation function: ReLU Learning rate: 0.000001 Batch size: 144 Epochs: 1000 Optimizer: RMSProp Loss Function: Huber loss

Table 3은 랜덤 포레스트의 각 무작위 입력 변수 선택을 적용한 모델과 SVR, ANN으로 구성된 결측치 추정 모델의 10겹 교차검증 결과이다. Auto를 이용하여 구성된 모델이 RMSE/MAE 부분에서 가장 좋은 성능을 확인할 수 있었다. 본 논문에서 수집한 전력 사용량 데이터에서는 총 632개의 결측 데이터가 존재하였으며, Auto를 적용한 랜덤 포레스트 기반의 추정 모델을 기반으로 결측 데이터의 추정값을 도출하여 결측 데이터를 모두 보완하였다. Fig. 3을 통해 랜덤 포레스트의 변수 중요도를 나타내었다. 전력수요 추정 모델에서 전력수요의 중요한 외부 요인으로 크게 공휴일의 여부, 전기요금에서 경부하의 여부, 시간을 나타내는 Equation (3)의 $hour_y$ 등을 확인할 수 있었다.

Table 3. 10-Fold Cross Validation

Algorithm	RMSE (kWh)	MAE (kWh)
Random Forest (auto)	5.398	2.767
Random Forest (sqrt)	6.975	3.911
Random Forest (log2)	6.832	3.807
Support Vector Regression	10.686	6.115
Artificial Neural Network	9.980	6.534

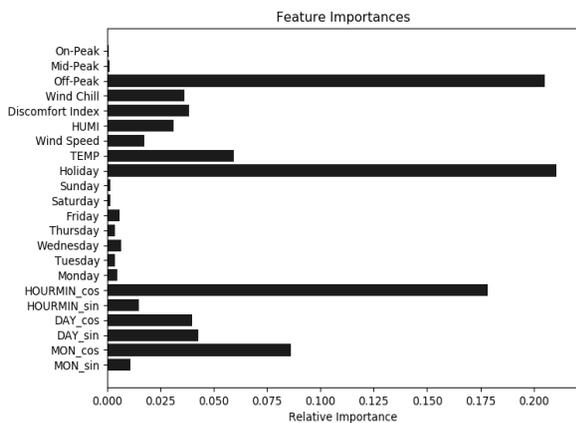


Fig. 3. Feature Importances in Random Forest

5.2 예측 모델 구성을 위한 데이터 분할

본 논문은 앞서 3장에서 기술한 것과 같이, 한전 i-Smart를 통해 2015년 1월 1일부터 2016년 12월 31일까지 2년의 기간으로 인천에 위치한 “K” 공장의 30분 단위 전력 사용량 데이터를 수집하였다. Table 4와 같이, 전체 데이터 집합에서 훈련 집합과 평가 집합을 5:5의 비율로 나누어, 훈련 집합은 2015년 1월 4일부터 2016년 1월 3일까지 총 365일의 기간으로, 평가 집합의 기간은 2016년 1월 4일부터 2016년 12월 31일까지 총 363일로 분할하였다. 2015년 1월 1일부터 2015년 1

월 3일까지의 데이터 집합은 과거 전력수요에 관한 정보가 존재하지 않는다. 이는 모델을 학습할 때 과적합의 가능성이 있어 훈련 집합에서 입력 변수를 위한 정보로만 사용하였다. 훈련 집합 튜플(Tuple)의 개수는 30분 단위를 기준으로 365일의 시점을 모두 고려하였을 때, 17,520개에서 487개의 결측치가 존재하였으며, 결측치를 모두 보완하여 모델을 학습하였다. 평가 집합 튜플의 개수는 예측 시점을 모두 고려하였을 때, 총 17,424개가 존재해야 한다. 하지만, 145개의 결측치가 존재하였기에 실제 예측 정확성을 평가할 때에는 17,279개를 적용하였다.

Table 4. Split Data into Training and Test Set

Data set	Period (Day)	Number of tuples	Missing value
Training set	2015. 01. 04 ~ 2016. 01. 03 (365)	17,033	487
Test set	2016. 01. 04 ~ 2016. 12. 31 (363)	17,279	145

5.3 다층 퍼셉트론 기반의 전력수요 예측

본 논문에서 제안한 단기 전력수요 예측 모델의 성능을 평가하기 위해, 3가지의 경우를 고려하여 다양한 은닉층의 구성으로 전력수요를 예측하였다.

- Case 1: 전체 평가 집합
- Case 2: 훈련 집합의 결측치 시점
- Case 3: 입력 변수에 결측치 추정값이 포함된 시점

Case 1의 전체 평가 집합은 평가 집합에서 결측치를 제외한 모든 시점을 예측하며 총 17,269개의 시점을 예측하였다. Case 2의 훈련 집합의 결측치 시점은 2015년에 결측치가 존재하였던 시점의 동일한 시점을 2016년도에 예측하는 것이다. 예를 들어, 2015년 4월 26일 18시 30분 전력수요 데이터가 결측 데이터라면 이를 랜덤 포레스트 기반의 추정 모델로 결측치를 보완한 다음, MLP 모델을 구성하여 2016년 4월 26일 18시 30분의 전력수요를 예측하는 것이다. Case 2의 예측 시점은 총 487개이다. 또한, Case 3은 평가 집합에서 입력 변수에 결측치 추정값이 포함하는 시점을 뜻한다. 예를 들어, 2016년 7월 8일 0시 30분 전력수요 데이터가 결측 데이터라면 먼저 랜덤 포레스트 추정 모델의 추정값으로 보완한다. 다음으로, MLP 예측 모델을 이용하여 2016년 7월 9일에서 10일을 예측할 때, 모델의 입력 변수로 2016년 7월 8일 0시 30분의 추정값이 속하므로 이 시점에 관한 예측 정확도를 평가한다. Case 3의 예측 시점은 2016년도 결측치는 145개이었으나, 입력 변수는 과거 2일의 전력수요량 모두를 반영하므로 실제 예측 시점은 총 239개이었다.

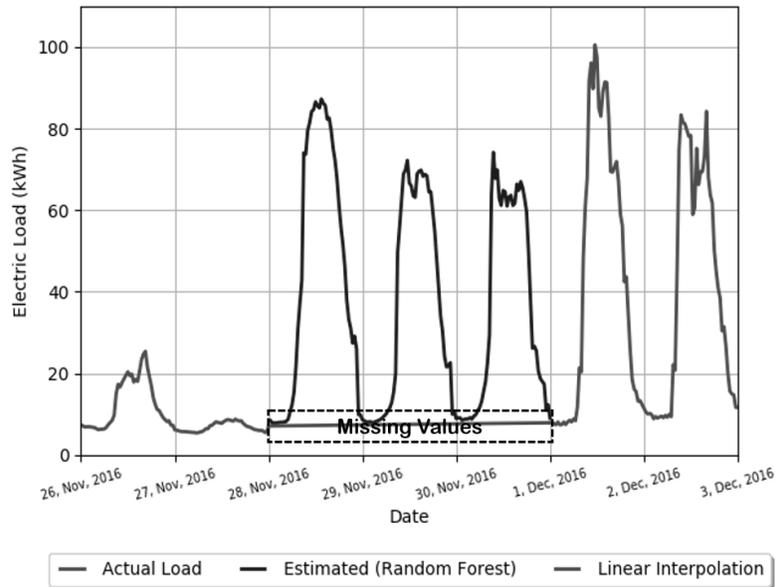


Fig. 4. Recovering Missing Values Using the Random Forest

또한, 본 논문은 3가지의 결측치 보완 방법과의 비교를 통해 제안한 방법의 당위성을 입증한다. 먼저 Original은 과거 전력 수요량과 공휴일 정보를 반영하지 않고, 앞서 구성한 22개의 외부 요인으로 평가 집합을 예측한다. Zero는 결측 데이터의 값을 0으로 보완하여 예측 모델을 구성하고 예측 성능을 평가한다. Linear Interpolation은 결측 데이터의 이전과 이후 전력수요량에 선형 보간법을 이용하여, 결측 데이터를 보완한다. 그리하여 본 논문에서 제안한 랜덤 포레스트 기반의 전력수요 추정 모델을 통해 결측치를 보완한 방법으로 예측 정확성을 비교한다. 실험 결과를 Table 5에 나타내었으며, 대체로 제안한 기법이 다른 결측치 보완 방법보다 더욱 향상된 예측 정확성을 확인하였다. 또한, 은닉층의 구성은 전반적으로

로 확인한 성능 차이를 보이진 않았으나, 3층과 4층으로 구성된 MLP 기반의 예측 모델이 다양한 MLP 모델들 중 가장 우수한 전력수요 예측 기법임을 확인하였다. 특히, Case 3에서 제안한 기법은 다른 보완 방법보다 월등한 예측 성능을 보였다. 이는 Fig. 4와 같이, 장기 결측치가 발생하였을 때, 제안한 기법은 전력수요 패턴을 효과적으로 반영하여 추정된 값들이 예측 모델의 입력 변수로 구성되기 때문에 0이나 선형 보간법으로 보완한 값보다 더욱 정교한 예측 모델을 구축할 수 있었다.

5.4 기계 학습 및 선행 연구와의 예측 성능 비교

본 논문은 제안한 모델을 앞서 기술했듯이 비선형 회귀 분

Table 5. RMSE(MAE) Results with Multilayer Perceptron

Case	Number of hidden layers	Original	Zero	Linear interpolation	Random forest
Case 1	1	18.211 (10.619)	12.414 (7.503)	12.186 (7.077)	11.639 (6.818)
	2	19.031 (11.088)	12.288 (7.180)	11.965 (6.884)	11.626 (6.793)
	3	18.325 (10.501)	12.139 (7.126)	12.134 (6.625)	11.510 (6.936)
	4	18.272 (10.187)	12.255 (7.201)	11.934 (6.757)	11.651 (6.599)
Case 2	1	11.356 (7.234)	7.009 (4.519)	7.371 (4.628)	7.094 (4.507)
	2	11.870 (7.715)	7.737 (4.993)	7.207 (4.469)	6.920 (4.530)
	3	11.683 (7.239)	7.603 (4.881)	7.235 (4.534)	6.695 (4.246)
	4	11.096 (6.648)	7.750 (4.849)	7.333 (4.589)	7.141 (4.376)
Case 3	1	18.149 (11.397)	22.324 (12.585)	22.674 (12.443)	11.827 (6.680)
	2	18.921 (12.169)	21.291 (11.543)	21.982 (11.954)	10.926 (6.664)
	3	18.684 (11.475)	20.138 (10.573)	23.104 (12.025)	11.261 (6.604)
	4	17.897 (11.164)	21.877 (11.315)	22.184 (11.562)	8.903 (5.772)

Table 6. RMSE(MAE) Distribution for Each Model

Model	Original	Zero	Linear interpolation	Random forest
SVR	21.874 (13.536)	15.338 (9.470)	14.140 (8.372)	13.446 (7.915)
RF	19.535 (11.439)	12.938 (7.664)	12.780 (7.889)	12.462 (7.422)
SVR [14]	21.593 (13.435)	-	-	-
ANN [14]	19.325 (11.556)	-	-	-
ANN [15]	19.423 (11.982)	-	-	-
Proposed model	18.272 (10.187)	12.255 (7.201)	11.934 (6.757)	11.651 (6.599)

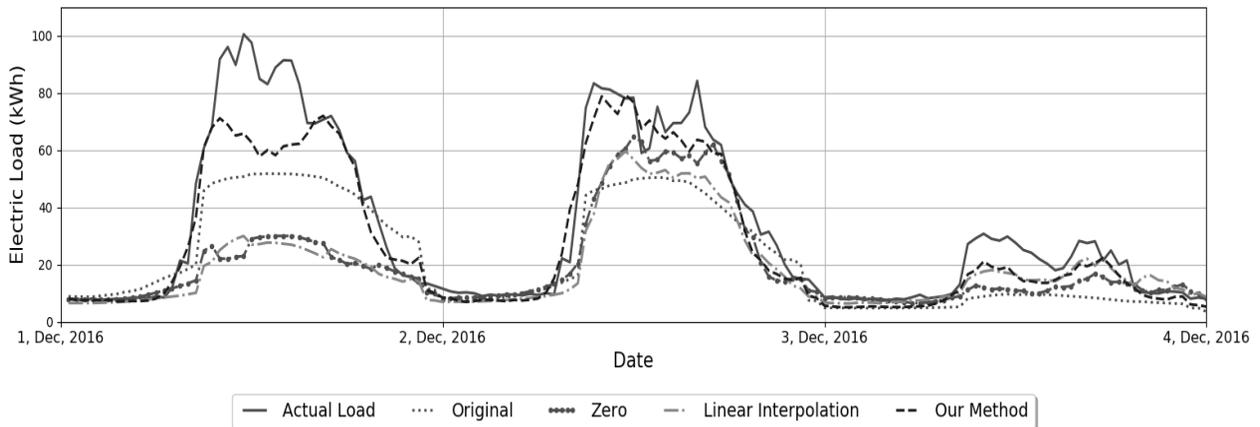


Fig. 5. Short-Term Electric Load Forecasting

석에 효과적인 알고리즘으로 SVR 및 랜덤 포레스트뿐만 아니라 선행 연구[14, 15]에서 제안된 모델과의 예측 성능을 비교하였다. 제안한 모델은 랜덤 포레스트로 결측치를 보완한 다음 4층의 은닉층으로 구성된 MLP 기반의 전력수요 예측 모델이다. SVR과 랜덤 포레스트의 내부 구조는 Table 2와 5.1절에 각각 기술한 것과 동일한 조건으로 실험하였다. [14, 15]는 단기 전력수요 예측을 위한 선행 연구로 SVR와 ANN 기반의 예측 모델이다. 본 논문에서 제안한 방법론이 더욱 발전된 예측 기법임을 입증하기 위해, 선행 연구에서 제시한 입력 변수와 초매개변수의 설정을 동일한 환경으로 구성하여 평가 집합의 전력수요를 예측하였다. Table 6에서 나타냈듯이, 제안한 예측 모델이 다른 기계 학습 기법이나 선행 연구보다 우수한 예측 정확성을 확인할 수 있었다.

앞서 Fig. 5에 나타낸 것과 같이, 전력 사용량 데이터에서 2016년 11월 28일 0시 30분부터 12월 1일 0시까지 144개의 시점에 결측치가 발생하였다. 그리하여, 4층의 은닉층으로 구성된 MLP 모델을 이용해 Case 3을 토대로 30분 단위의 전력수요를 예측한 결과를 Fig. 5에 나타내었다. Fig. 5의 예측 기간은 결측치의 추정값이 입력 변수로 반영된 기간인 2016년 12월 1일 0시 30분부터 3일 11시 30시까지의 30분 단위로 143개의 예측 시점에 관한 전력수요량이다. Original은 이전 전력수요를 반영하지 않고, 훈련 집합에서 외부 요인으로만 학습하여 평가 집합에 적용한 예측 모델이다. 그리하여, 2015

년의 낮은 전력수요량을 기반으로 학습되었기에 실제 전력수요보다 낮은 전력수요를 예측한다는 것을 확인하였다. Zero와 Linear Interpolation은 결측치 이후의 예측 시점에 전력수요가 0이나 선형 보간법의 값으로 반영되어 실제 전력수요보다 낮은 전력수요를 예측하였다. 다만 예측 시점이 결측 데이터와 멀어질수록 실제 전력수요량을 반영한 예측 결과를 도출하여, 2016년 12월 2일부터는 어느 정도 추세가 반영된 전력수요의 예측 결과를 확인하였다. 본 논문에서 제안한 기법이 이전 전력수요의 패턴을 반영하여 모든 결측치 보완 방법보다 정확한 단기 전력수요를 예측하였다. 하지만 결측치 이후 시점인 2016년 12월 1일의 피크전력에 관한 예측이 정확하지 않았다. 블랙아웃과 같은 상태를 방지하기 위해서는 전력수요 예측 모델이 실제 수요보다 항상 높게 예측하면서도 실제 수요와 가깝게 예측해야 안전한 모델이라 할 수 있다. 이는 피크전력 대응에는 다소 적합하지 않아 향후 연구로 이러한 문제를 개선시켜야 할 여지가 존재한다.

6. 결 론

본 논문은 스마트 그리드의 에너지관리시스템에서 정교한 단기 전력수요를 예측하기 위해, 다수의 결측치가 포함된 데이터 집합을 대상으로 결측치의 보완 방법과 이를 반영한 심층 학습 기반의 전력수요 예측 기법을 제안하였다. 먼저, 예

측 모델을 학습하기 위한 입력 변수로 전력수요 시점이 설명 가능한 외부 요인으로 구성한다. 다음으로, 랜덤 포레스트를 이용하여 효과적으로 결측치를 추정하고, 이를 반영하여 MLP 기반의 전력수요 예측 모델을 구축한다. 다양한 실험을 바탕으로 제안한 기법이 다른 결측치 보완 방법이나 기계 학습 알고리즘보다 더욱 정교한 예측 성능이 보임을 확인하였다. 하지만, 결측치 이후의 전력수요 예측에서 피크전력 시점에 대해 다소 부정확한 예측 정확성을 도출하여 피크전력 대응에는 적합하지 않다는 것을 확인하였다.

향후 Many-To-Many 방식으로 결측치의 보완을 MLP나 LSTM을 통해 더욱 정교한 예측 모델을 구성할 예정이다. 또한, 앞서 결측치 이후의 예측 시점에서 피크전력과 연관성이 높은 새로운 입력 변수를 추가하여, 다양한 통계 기법이나 기계 학습을 통해 정교한 피크전력 예측 모델을 구성할 계획이다. 이와 함께, ESS나 CCHP(Combined Cooling, Heat and Power)의 스케줄링을 위한 연구도 병행하여 실제 스마트 그리드의 운영 방안을 모색할 예정이다.

References

- [1] C. Fan, F. Xiao, S. Y. Hea, and Y. Zhao, "A short-term building cooling load prediction method using deep learning algorithms," *Applied Energy*, Vol.195, pp.222-233, 2017.
- [2] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, "Deep belief network based electricity load forecasting: An analysis of Macedonian case," *Energy*, Vol.115, pp.1688-1700, 2016.
- [3] H. Chitsaz, H. Shaker, H. Zareipour, D. Wood, and N. Amjadi, "Short-term electricity load forecasting of building in microgrids," *Energy and Buildings*, Vol.99, pp.50-60, 2015.
- [4] W. Lee, J. Jung, and M. Lee, "Development of 24-hour Optimal Scheduling Algorithm for Energy Storage System using Load Forecasting and Renewable Energy Forecasting," in *Proceedings of the IEEE Power & Energy Society General Meeting*, Chicago, pp.1-5, 2017.
- [5] J. Moon, K.-H. Kim, and E. Hwang, "A Short-Term Electric Load Forecasting Scheme Using 2-Stage Predictive Analytics," in *Proceedings of the IEEE International Conference on Big Data and Smart Computing*, Shanghai, pp.219-226, 2018.
- [6] W.-W. Kim, J.-S. Shin, and J.-O. Kim, "Operation Strategy of Multi-Energy Storage System for Ancillary Services," *IEEE Transactions on Power Systems*, Vol.32, No.6, pp.4409-4417, 2017.
- [7] S. Ryu, J. Noh, and H. Kim, "Deep Neural Network Based Demand Side Short Term Load Forecasting," *Energies*, Vol.10, No.1, pp.1-20, 2017.
- [8] T. Niet et al., "Hedging the risk of increased emissions in long term energy planning," *Energy Strategy Reviews*, Vol.16, pp.1-12, 2017.
- [9] Y.-M. Wi, S. Kong, J. Lee, and S.-K. Joo, "Demand-Side Management Program Planning Using Stochastic Load Forecasting with Extreme Value Theory," *Journal of Electrical Engineering & Technology*, Vol.11, No.5, pp.1093-1099, 2016.
- [10] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, Vol.32, No.3, pp.914-938, 2016.
- [11] X. Ke, A. Jiang, and N. Lu, "Load Profile Analysis and Short-term Building Load Forecast for a University Campus," in *Proceedings of the IEEE Power & Energy Society General Meeting*, Boston, pp.1-5, 2016.
- [12] B. Zhang, J.-L. Wu, and P.-C. Chang, "A multiple time series-based recurrent neural network for short-term load forecasting," *Soft Computing*, Vol.22, No.12, pp.4099-4112, 2018.
- [13] S. Park, J. Moon, and E. Hwang, "A Comparison of Various Machine Learning Techniques for Missing Values Recovering in Load Prediction," in *Proceedings of the KSIC Spring Conference*, Busan, pp.1-2, 2018.
- [14] J. Moon, S. Jun, J. Park, Y.-H. Choi, and E. Hwang, "An Electric Load Forecasting Scheme for University Campus Buildings Using Artificial Neural Network and Support Vector Regression," *KIPS Transactions on Computer and Communication Systems*, Vol.5, No.10, pp.293-302, 2016.
- [15] J. Park, J. Moon, and E. Hwang, "An Electric Load Forecasting Scheme with High Time Resolution Based on Artificial Neural Network," *KIPS Transactions on Software and Data Engineering*, Vol.6, No.11, pp.527-536, 2017.
- [16] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renewable and Sustainable Energy Reviews*, Vol.73, pp.1104-1122, 2017.
- [17] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, Vol.50, pp.1352-1372, 2015.
- [18] L. Hernandez et al., "A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings," *IEEE Communications Surveys & Tutorials*, Vol.16, No.3, pp.1460-1495, 2014.
- [19] A. Almalaq and G. Edwards, "A Review of Deep Learning Methods Applied on Load Forecasting," in *Proceedings of the IEEE International Conference on Machine Learning and Applications*, Cancun, pp.511-516, 2017.
- [20] M. R. Sarkar, M. G. Rabbani, A. R. Khan, and M. M. Hossain, "Electricity Demand Forecasting of Rajshahi City in Bangladesh Using Fuzzy Linear Regression Model," in *Proceedings of the International Conference on Electrical Engineering and Information Communication Technology*, Dhaka, pp.1-3, 2015.
- [21] G. Dudek, "Pattern-based local linear regression models for short-term load forecasting," *Electric Power System Research*, Vol.130, pp.139-147, 2016.
- [22] A. Gerossier, R. Girard, G. Kariniotakis, and A. Michiorri,

- “Probabilistic Day-Ahead Forecasting of Household Electricity Demand,” *CIREC - Open Access Proceedings Journal*, Vol.2017, No.1, pp.2500-2504, 2017.
- [23] A. S. Ahmad et al., “A review on applications of ANN and SVM for building electrical energy consumption forecasting,” *Renewable and Sustainable Energy Reviews*, Vol.33, pp.102-109, 2014.
- [24] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, “Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy,” *Applied Energy*, Vol.123, pp.168-178, 2014.
- [25] L. Ghelardoni, A. Ghio, and D. Anguita, “Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression,” *IEEE Transactions on Smart Grid*, Vol.4, No.1, pp.549-556, 2013.
- [26] K. Grolinger, A. L’Heureux, M. A. M. Capretz, and L. Seewald, “Energy Forecasting for Event Venues: Big Data and Prediction Accuracy,” *Energy and Buildings*, Vol.112, pp.222-233, 2016.
- [27] S. Jurado, A. Nebot, F. Mugica, and N. Avellana, “Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques,” *Energy*, Vol.86, pp.276-291, 2015.
- [28] A. Bagnasco, F. Fresi, M. Saviozzi, F. Silvestro, and A. Vinci, “Electrical consumption forecasting in hospital facilities: An application case,” *Energy and Buildings*, Vol.103, pp.261-270, 2015.
- [29] N. Zeng, H. Zhang, W. Liu, J. Liang, and F. E. Alsaadi, “A switching delayed PSO optimized extreme learning machine for short-term load forecasting,” *Neurocomputing*, Vol.240, pp.175-182, 2017.
- [30] S. S. Reddy, “Bat algorithm-based back propagation approach for short-term load forecasting considering weather factors,” *Electrical Engineering*, Vol.100, No.3, pp.1297-1303, 2018.
- [31] M. Abdel-Nasser and K. Mahmoud, “Accurate photovoltaic power forecasting models using deep LSTM-RNN,” *Neural Computing and Applications*, pp.1-14, 2017.
- [32] W. Kong et al., “Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network,” *IEEE Transactions on Smart Grid*, pp.1-11, 2017.
- [33] P.-H. Kuo and C.-J. Huang, “A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting,” *Energies*, Vol.11, No.1, pp.1-13, 2018.
- [34] F. Tang and H. Ishwaran, “Random forest missing data algorithms,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, Vol.10, No.6, pp.363-377, 2017.
- [35] M. W. Ahmad, M. Mourshed, and Y. Rezgui, “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption,” *Energy and Buildings*, Vol.147, pp.77-89, 2017.
- [36] K.-H. Kim, R.-J. Park, S.-W. Jo, and K.-B. Song, “24-Hour Load Forecasting Algorithm Using Artificial Neural Network in Summer Weekdays,” *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, Vol.31, No.12, pp.113-119, 2017.
- [37] G. Panchal, A. Ganatra, Y. P. Kosta, and D. Panchal, “Behaviour Analysis of Multilayer Perceptrons with Multiple Hidden Neurons and Hidden Layers,” *International Journal of Computer Theory and Engineering*, Vol.3, No.2, pp.332-337, 2011.
- [38] G. Hinton, N. Srivastava, and K. Swersky, Lecture 6a-overview of mini-batch gradient descent [Internet], <https://class.coursera.org/neuralnets-2012-001/lecture>.
- [39] J. Niu, J. Chen, and Y. Xu, “Twin support vector regression with Huber loss,” *Journal of Intelligent and Fuzzy Systems*, Vol.32, No.6, pp.4247-4258, 2017.
- [40] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How Many Trees in a Random Forest?” in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, Springer, Berlin, Heidelberg, pp.154-168, 2012.
- [41] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, Vol.12, pp.2825-2830, 2011.
- [42] M. Abadi et al., “Tensorflow: a system for large-scale machine learning,” *OSDI*, Vol.16, pp.265-283, 2016.



문 지 훈

<https://orcid.org/0000-0001-9524-5729>

e-mail : johnny89@korea.ac.kr

2015년 한성대학교 정보통신공학과(학사)

2015년~현 재 고려대학교

전기전자공학과 석·박사통합과정

관심분야 : 스마트 그리드, 시계열 분석, 수요 예측, 기계 학습, 데이터 마이닝 등



박 성 우

<https://orcid.org/0000-0001-5469-0332>

e-mail : psw5574@korea.ac.kr

2018년 홍익대학교 컴퓨터공학과(학사)

2018년~현 재 고려대학교 전기전자공학과

석사과정

관심분야 : 인공 지능, 데이터 마이닝,

빅 데이터 분석, 수요 예측 등



황 인 준

<http://orcid.org/0000-0002-0418-4092>

e-mail : ehwang04@korea.ac.kr

1988년 서울대학교 컴퓨터공학과(학사)

1990년 서울대학교 컴퓨터공학과(석사)

1998년 Univ. of Maryland at College Park
전산학과(박사)

1998년~1999년 Bowie State Univ. 조교수

1999년~1999년 Hughes Research Lab. 연구교수

1999년~2004년 아주대학교 정보통신전문대학원 조/부교수

2004년~현 재 고려대학교 전기전자공학과 교수

관심분야: 데이터베이스, 빅데이터 처리, 영상 처리, 음악 검색,
멀티미디어 검색, 소셜 네트워크 데이터 분석