

Design of Short-Term Load Forecasting based on ANN Using Bigdata

빅데이터를 이용한 인공지능망 기반 시간별 전력수요 예측

Ji-Won Lee · Hyung-Jun Kim · Mun-Kyeom Kim

이지원* · 김형준* · 김문겸†

Abstract

Short-term power load prediction is crucial for scheduling generators, securing transmission capacity, and determining economic market price. This paper proposes a Bigdata-based Artificial Neural Network (B-ANN) model for a short-term load forecasting by utilizing decision-tree method. The bigdata has been composed of 2018 time index, meteorological data, exchange rates, oil price, and past power demand in Seoul. The decision-tree method is applied to classify the data in accordance with the value of decreasing entropy. In addition, the resilient back propagation algorithm (RPROP) is applied for ANN that derives fast results and result in reduced error through variable learning rates. The applicability and effectiveness of the proposed model are verified by conducting simulations for forecasting 1-day hourly power load. The results confirm that the proposed model has decreased the mean absolute percentage error (MAPE) dramatically in comparison with other forecasting methods such as multiple regression analysis, time series analysis, and auto-regressive integrated moving average (ARIMA) model.

Key Words

Neural Network, Backpropagation Algorithm, Market Price, Transmission congestion, Electricity market

1. 서 론

전력수요예측은 안정적인 전력수급과 경제적인 계통운영에 필수적이며 예측기간에 따라서 장기, 중기, 단기, 실시간으로 구분된다. 중장기 수요예측은 전력소 건설계획이나 유지보수를 위한 예측이며 단기 수요예측은 발전계획을 위한 예측이다. 실시간 예측은 주파수 제어 및 경제급전을 반영하여 이후 5분의 전력을 예측한다. 특히 24시간부터 168시간을 예측해야 하는 단기 전력수요예측은 발전원 운영계획, 송전용량 확보와 전력가격 계산에 중요한 역할을 한다 [1]. 따라서 정밀한 단기 전력수요예측을 위해 많은 연구가 진행되고 있다 [1]-[4]. [1]은 시간별 전력수요예측의 오차값을 줄이는 순환형 신경망 구조를 가지는 인공지능망 알고리즘을 제시하였다. [2]는 기온변화에 따른 하계 전력사용의 특징을 해석하고 수요와 기온간의 민감도를 이용하여 월요일을 제외한 평일의 시간별 전력수요예측을 보여준다. [3]은 기온과 전운량을 포함하는 과거 5년의 데이터를 이용하여 전력수요 민감도를 통해 평일/월요일/주말의 전력수요를 예측하였다. [4]는 단기 전력수요의 예측 정확성을 향상시키기 위해 유전 알고리즘을 최적화한 인공지능망 모델을 제안하였다.

단기 전력수요예측에 영향을 미치는 요인에는 대표적으로 기

상요인과 과거 전력소비패턴이 있다. 먼저 기상요인에서, 사계절의 특징을 가진 한국의 특성상 하계에는 냉방기 사용, 동계에는 난방기 사용으로 인해 전력소비가 증가하여 일별 기온이 전력수요예측에 필수적인 요인이 된다. 그리고 시간별 기온에 따라 피크전력이 나타나는 시간도 달라진다. 기상청에서 일별 최고·최저·평균 기온 데이터뿐만 아니라 시간별 기온, 풍속, 강수량, 습도, 전운량에 대한 과거 데이터와 예측 데이터를 제공함으로써 더 다양한 기상요소를 이용할 수 있다. 하지만 최근 지구온난화로 인한 이상기온 등으로 인해 전력수요예측에 오차가 발생한다. 전력예측의 또 다른 요인으로는 전력소비패턴이 있다. 요일에 따라 전력패턴이 다른 양상을 가지는데 근무일과 공휴일은 서로 다른 패턴을 보이며 월요일 새벽의 전력은 전력소비가 적은 주말의 영향이 이어져 낮은 전력을 소비한다 [5]. 최근에는 더 정밀한 단기 전력수요예측을 위해서 기상요소 외에 전력에 영향을 주는 기상 외 요소를 분석하고 데이터화하여 예측에 활용하는 연구가 활발하게 진행되고 있다 [6]-[7].

최근 빅데이터에 대한 관심도가 높아지면서 많은 데이터를 처리하는 기법이 다양한 분야에서 주목받고 있다 [8]. 대표적으로 수치형 데이터에 대한 확률을 기반으로 데이터를 분석하는 통계적 분석과 데이터 마이닝 기법이 있다. 통계적 분석에는 평균,

† Corresponding Author : Dept. of Energy System Engineering, Chung-Ang University, Korea.
E-mail: mkim@cau.ac.kr
<https://orcid.org/0000-0002-0901-7816>

* Dept. of Energy System Engineering, Chung-Ang University, Korea.
<https://orcid.org/0000-0002-5958-9708>

Received : May 8, 2020 Revised : May 18, 2020 Accepted : May 28, 2020

Copyright © The Korean Institute of Electrical Engineers

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

최빈값, 분산, 표준편차 등을 구하는 기술 통계량, 두 변수간의 선형적 관계를 규명하는 상관분석, 독립변수와 종속변수 사이의 수학적식을 구하는 회귀분석 같은 전통적 기법이 사용된다. 데이터 마이닝은 대용량의 데이터를 분석하기 위한 의사결정트리(Decision Tree), 패턴인식, 인공지능 등의 기법이 있다 [9].

단기 전력예측 기법에는 통계적 기법과 인공지능형 기법이 있는데, 통계적 기법에는 지수평활법과 시계열 분석법 등이 있다. 지수평활법[10]은 실제값과 예측값 사이의 가중 평균을 이용하여 미래의 값을 예측하는 기법으로, 국내에서는 평일 하루 전 전력수요예측을 위해 지수평활법을 응용한 기법을 사용하고 있다. [11]에서는 지수평활법을 이용하여 시간별 전력예측을 진행하였다. 시계열 기법은 일간, 주간, 계절 등과 같은 주기성을 갖는 대상을 예측하는 기법으로, 최근에는 주기가 일정하지 않은 비정상상의 데이터를 가지고 예측을 진행하는 ARIMA(Autoregressive integrated moving average model)에 대한 연구가 진행되고 있다. [12]에서는 ARIMA 기반의 예측 모델을 통해 대학 캠퍼스의 시간별 전력 사용량 예측을 진행하였다. 하지만 위 두 가지 통계적 기법 모두 비선형적 모델 구현이 어렵고 이상기후 같은 돌발상황을 반영하여 예측을 진행하지 못한다는 단점을 가진다 [13]. 단기 전력예측을 위한 인공지능형 기법에서는 서포트 벡터 머신(Support vector machine, SVM)[14]과 인공신경망(Artificial neural networks, ANN) 기법이 사용된다. [15]는 신호나 시스템의 데이터 신호의 잡음을 제거하는 웨이블릿 변환과 인공신경망을 결합한 모델을 통해 동계와 춘계의 시간별 전력을 예측한다.

본 논문은 기상요소와 기상 외 요소를 기반으로 빅데이터를 처리하는 데이터 마이닝 기법 중 하나인 의사결정트리를 이용하여 전력과의 상관성을 분석하고자 한다. 기존 연구가 고려하던 기온과 전운량 요인뿐만 아니라 풍속, 강수량, 습도 등의 기상 요소를 통해 데이터를 구성한다. 그리고 시간별 전력수요와 상관관계가 약해보이는 기상 외 요소도 고려하여 빅데이터를 구축하고 전력수요와의 연결성 분석을 위해 의사결정트리를 사용한다. 2018년 서울의 평일(근무일) 기상데이터와 시간별 전력데이터, 그리고 기상 외 데이터를 수집한다. 본 논문은 빅데이터에 기반한 ANN 모델을 적용한 수요예측모델(B-ANN)을 제안하여 단기 시간별 전력수요예측 결과 값을 다른 예측 알고리즘의 예측결과와 비교하여 모델의 정확성을 증명한다.

본 논문의 2장에서는 논문에서 사용될 모델에 관해 서술하였으며 3장에서는 사례 연구에 관하여 서술하였다. 4장의 결론으로 마무리한다.

2. 모델 설명

2.1 의사결정트리 기법

의사결정트리는 데이터의 분류나 예측을 목적으로 통계학에서 개발 및 발전되었다[16].

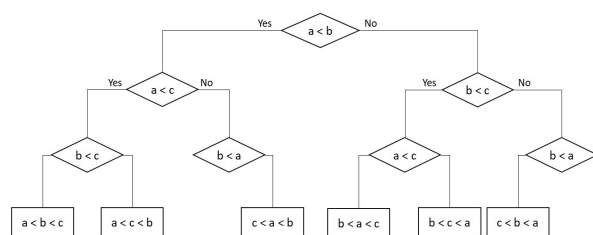


그림 1 의사결정트리 모형

Fig. 1 Decision tree model

의사결정트리는 순환적 분할 방식을 이용하여 트리를 구축하는 방식이다. 순환적 분할이란 하향식 귀납법으로 상위노드에서 분리된 데이터를 분리규칙에 의해 서로 유사한 데이터로 세분화하고 최종 분류가 이루어질 때까지 반복하는 분할 기법이다. 의사결정트리는 분류나 예측을 목적으로 쓰이지만 분석의 과정과 결과를 직관적으로 보여주기 때문에 분석과정의 설명이 필요한 경우에도 적용된다.

의사결정트리에서 서로 유사한 데이터를 세분화 하는 방식은 순도의 증가, 또는 불순도나 불확실성의 감소를 기준으로 형성된다. 순도를 계산하는 방법에는 두 가지가 있다. 첫 번째, 엔트로피(entropy)는 확률변수의 불확실성을 수치로 나타낸 것으로, 값이 높을수록 불확실성이 높아진다. 분할 전 A 영역에 포함되는 전체 m 개인 경우 중에 k 인 경우의 엔트로피와 분할 후 d 개의 영역에 각각 속하는 경우의 엔트로피를 나타낸 식은 다음과 같다.

$$Entropy(A) = - \sum_{k=1}^m p_k \log_2(p_k) \quad (1)$$

여기서, p_k 는 A 영역에 포함되는 m 개의 경우 중 k 범주에 속하는 경우의 비율을 의미한다.

$$Entropy(A) = \sum_{i=1}^d R_i \left(- \sum_{k=1}^m p_k \log_2(p_k) \right) \quad (2)$$

여기서, R_i 은 분할 전 경우 가운데, 분할 후 i 영역에 속하는 경우의 비율을 말한다.

두 번째로 지니계수(Gini index)는 엔트로피와 비슷하게 불순도와 다양성을 계산하는 기법이다. 분할 전 불순도의 지니계수와 d 개의 부분집합으로 나뉜 후의 지니계수에 대한 식은 다음과 같다.

$$G \cdot I(A) = 1 - \sum_{k=1}^m p_k^2 \quad (3)$$

$$G \cdot I(A) = \sum_{i=1}^d R_i \left(1 - \sum_{k=1}^m p_k^2 \right) \quad (4)$$

그림 2에서 두 개의 부분집합으로 나뉘었을 때의 비율(p)에 따라, 의사결정트리의 가지치기를 위한 오분류오차(misclassification error)와 위 두 가지 기법을 표현하였다. 엔트로피와 지니계수

모두 값을 최소로 하는 분류가 최적이라고 해석된다.

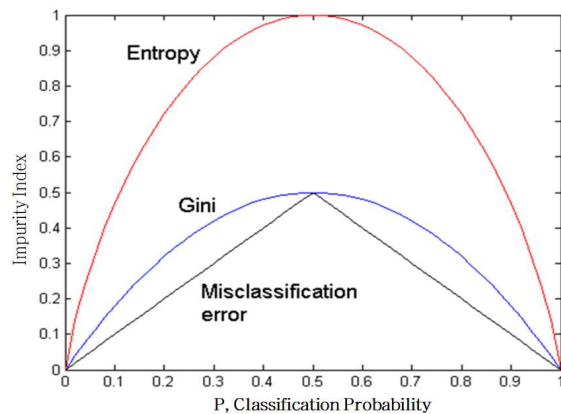


그림 2 엔트로피와 지니계수의 비교곡선
Fig. 2 Comparative curve of Entropy and Gini index

지니계수 계산이 엔트로피 계산보다 빠르지만 엔트로피는 로그를 취하여 정규화 과정을 거쳤기 때문에 정확한 데이터의 비율을 분석하는 데에 용이하다. 따라서 본 논문은 엔트로피를 사용하는 의사결정트리 시뮬레이션을 사용한다.

2.2 인공신경망 모델

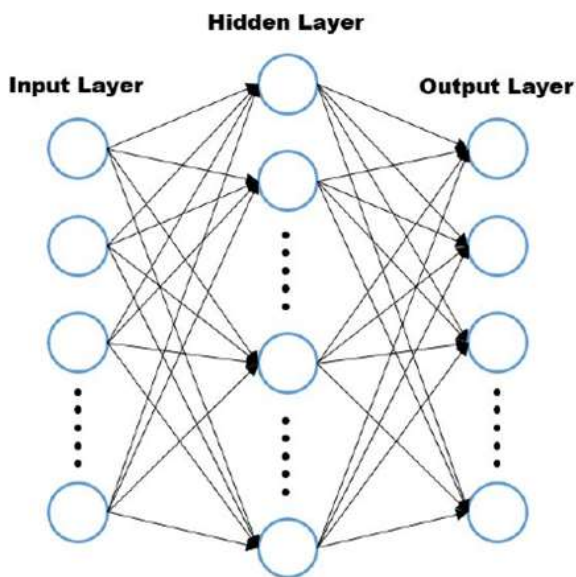


그림 3 인공신경망 모형 [17]
Fig. 3 Artificial Neural Network

인공신경망은 사람의 뉴런세포에서 비롯된 통계적 학습 알고리즘으로, 입력층, 은닉층 그리고 출력층으로 구분된다(그림 3 참고). 입력층과 은닉층, 은닉층과 은닉층 사이를 연결하는 가중치에 따라 결과값이 실제값에 가까워 질 수도, 결과값이 수렴하지 않을 수도 있다. 인공신경망에는 출력값의 오차를 줄이기 위한 가중치 조절에 따른 두 가지 대표적인 알고리즘이 존재한다.

2.2.1 순방향 신경망(Feed forward neural network) 모델

순방향 신경망은 인공신경망의 기본적인 구조를 가진 모델로, 하나 이상의 은닉층을 가지고 있어서 입력데이터의 비선형적 특징들을 추출할 수 있다 [18-19]. 통신 시스템, 동적 시스템 제어와 데이터 분류 같은 비선형 데이터 처리에 적용되었다 [20]. 은닉층이 2개 이상인 다층퍼셉트론이며 식은 다음과 같다.

$$net_{hidden} = \omega_0 + \sum_{i \in put} \omega_{i \in put} x_{i \in put} \quad (5)$$

net_{hidden} 은 은닉층으로 전달할 데이터 값이고 $\omega_0, \omega_{i \in put}$ 는 i 번째 입력데이터의 가중치이며, $x_{i \in put}$ 은 입력데이터를 의미한다. 위 식을 통해 은닉층으로 전달된 데이터는 시그모이드 활성화 함수를 적용하여 은닉층의 입력데이터로 사용된다.

$$o_{hidden} = f(\neq t_{hidden}) \quad (6)$$

$$f = \frac{1}{1 + e^{-x}} \quad (7)$$

$$net_{hidden} = \omega_0 + \sum_{hidden=1}^j \omega_{hidden} o_{hidden} \quad (8)$$

$$o_{output} = f(net_{output}) \quad (9)$$

o_{hidden} 은 은닉층의 입력데이터이고 net_{output} 은 출력층으로 전달할 데이터 값을 의미하며, ω_{hidden} 는 i 번째 입력데이터의 가중치를 의미한다. 초기에 입력한 출력값과 결과값인 o_{output} 간의 오차들의 합을 구하여 총 오차를 줄이기 위한 가중치 조절을 시행한다.

$$e = \frac{1}{2} \sum_{output=1}^k (t_{output} - o_{output})^2 \quad (10)$$

$$\Delta \omega_{hidden, output} = -\eta \frac{\partial e}{\partial \omega_{hidden, output}} \quad (11)$$

ϵ 은 오차를 의미하며 t_{output} 은 초기 출력값이다. η 은 학습률로, 학습속도를 조절한다. 속도가 너무 빠르면 결과값이 수렴하지 못할 가능성이 있다. 이 과정을 오차값이 일정 수준 이하가 될 때까지 반복한다.

2.2.2 탄력성 역전파 신경망(Resilient back propagation algorithm, RPROP) 모델

탄력성 역전파 신경망은 인공신경망 중 결과값의 오차를 줄이기 위해 은닉층 간의 학습률을 변동하게 하는 알고리즘으로 이루어져있다. 오차의 크기에 따라 변화하는 학습률은 수렴속도와 결과의 정확성을 향상시킨다[21].

$$b_i(k+1) = b_i(k) + \Delta b_i(k) \quad (12)$$

$$\Delta b_i(k) = -\text{sign}\left(\frac{\partial e}{\partial b_i}\right) \Delta_i(k) \quad (13)$$

$$\text{if } \frac{\partial \epsilon(k-1)}{\partial b_i} \frac{\partial \epsilon(k)}{\partial b_i} \geq 0 \quad (14)$$

$$\text{then } \Delta_i(k) = \min(\eta^+ \cdot \Delta_i(k-1), \Delta_{\max})$$

$$\text{if } \frac{\partial \epsilon(k-1)}{\partial b_i} \frac{\partial \epsilon(k)}{\partial b_i} < 0 \quad (15)$$

$$\text{then } \Delta_i(k) = \max(\eta^- \cdot \Delta_i(k-1), \Delta_{\min})$$

여기서, b_i 는 i 번째 은닉노드의 가중치를 말하며 k 는 internal 횟수, ϵ 는 오차를 나타낸다. Δ_i 는 b_i 의 스텝사이즈를 의미하며 Δ_{\max} 와 Δ_{\min} 는 스텝사이즈의 최대, 최소값을 나타낸다. 학습률은 $0 < \eta^- < 1 < \eta^+$ 를 의미한다.

그림 4는 순방향 신경망과 탄력성 역전파 신경망의 학습률 특성을 간단하게 보여주는 그림이다. 순방향 신경망은 고정된 학습률을 가지기 때문에 변화율이 0이 되는 최적 가중치 ω 값을 찾지 못할 수 있다. 그에 반해, 탄력적 학습률은 학습결과와 실제값의 오차에 따라 스텝사이즈를 조절하기 때문에 최적 가중치를 정확하게 구할 수 있다. 따라서 본 논문은 오차에 따라 학습률을 변화시켜 최적 가중치를 찾는 RPROP 알고리즘을 채택하여 전력수요예측을 진행한다.

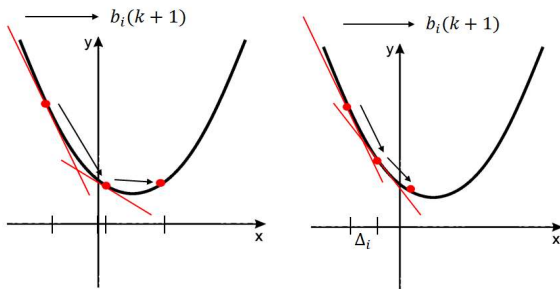


그림 4 고정된 학습률과 탄력적 학습률에 따른 알고리즘 비교
Fig. 4 Comparison of algorithms based on fixed and resilient learning rates

2.3 제안된 전력수요예측 모델

본 논문에서는 빅데이터 기반 RPROP 인공지능망 모델을 제안하여, 향상된 단기 전력수요예측 결과값을 증명하고자 한다. B-ANN 모델은 다음 알고리즘 흐름도를 따라 전력수요예측을 진행한다.

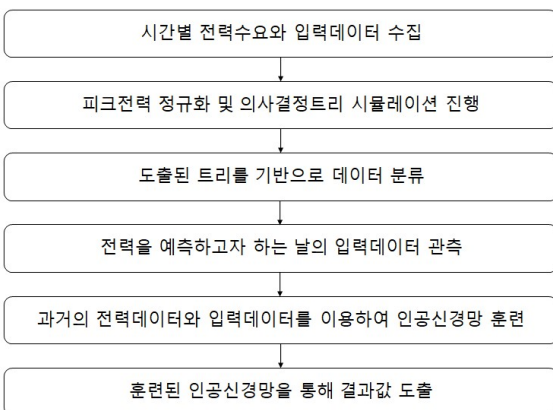


그림 5 제안된 B-ANN 모델 흐름도
Fig. 5 Proposed B-ANN model flow diagram

그림 5는 본 논문에서 제안된 모델의 흐름도이다. 첫번째로 시간별 전력수요와 기상 및 기상 외 데이터를 수집한다. 본 논문은 기존에 고려되지 않았던 시간별 기상 데이터와 환율, 유가 등의 기상 외 데이터를 수집하여 전력수요와의 연관성을 분석하고자 한다. 두번째로 앞서 수집된 빅데이터 정보 분석을 위해 피크전력 값을 정규화하고, 정규화된 데이터를 가지고 의사결정트리 시뮬레이션을 진행한다. 의사결정트리는 데이터 분류에 특화된 통계적 기법으로, 유사한 데이터를 세분화하여 결과값이 영향을 크게 받는 입력값과 영향을 주지 않는 입력값을 분류한다. 따라서 본 논문은 의사결정트리 시뮬레이션을 통해 전력수요에 영향을 주는 입력데이터를 분석하고, 군집화된 결과노드를 통해 유사한 특징을 가진 데이터를 그룹화한다. 이는 비슷한 데이터를 가지고 인공신경망을 학습시켜, 더 향상된 전력수요예측을 도출할 것이다. 네번째로 예측하고자 하는 날을 결정하여 해당 날의 기상데이터와 기상 외 데이터가 유효한지 판단한다. 입력데이터가 시간별로 존재하지 않는다면 보간법을 이용하여 시간별 입력값을 설정한다. 데이터 유효성을 판단한 후, 미래의 입력데이터가 포함되는 그룹의 데이터를 가지고 인공신경망을 학습시킨다. 앞에서 설명한 RPROP 알고리즘을 바탕으로, 기상요소 그리고 기상 외 요소의 입력데이터와 시간별 전력수요데이터를 이용하여 인공신경망을 학습시킨다. 인공신경망의 학습이란, 입력데이터와 전력수요데이터의 선형적, 비선형적 상관성을 분석하여 최적의 가중치를 도출하는 것이다. 마지막으로, 학습된 모델에 미래의 입력데이터를 입력하여 결과값을 도출한다. 본 논문은 모델의 정확성 판단을 위해 Mean Absolute Percentage Error(MAPE)를 오차값으로 사용하였으며 식은 다음과 같다.

$$MAPE(\%) = \frac{1}{n} \sum_{n=1}^n \left| \frac{\text{실제전력수요} - \text{예측전력수요}}{\text{실제전력수요}} \right| \times 100 \quad (16)$$

3. 사례 연구

3.1 입력데이터 분석

본 논문에서는 입력데이터로 최대기온(high.degree), 최저기온(low.degree), 평균기온(avg.degree)과 평균습도(avg.wet), 강우량(rain), 풍속(wind.speed), 전운량(cloud)와 같은 기상데이터[22] 뿐만 아니라 환율(dollar, yen, euro), 유가(Brent, WTI, Dubai, Oman) 및 천연가스 가격(natural_gas) [23]을 사용하여 의사결정트리를 구축하였다.

그림 6은 2018년 서울의 평일 피크전력수요를 나타낸 그림이다. 하계와 동계에 냉난방 기기의 사용이 늘어나서 전력수요가 증가함을 볼 수 있다. 의사결정트리 시뮬레이션을 위해 표 1과 같이 피크전력을 정규화하였다.



그림 6 2018년 서울의 평일 피크전력수요
Fig. 6 2018 Weekday peak power demand in Seoul

표 1 정규화된 전력수요

Table 1 Regularized power demand

Power(MWh)	Regularized
~ 65,000	-3
65, ~ 70,	-2
70, ~ 75,	-1
75, ~ 80,	0
80, ~ 85,	1
85, ~ 90,	2
90,000 ~	3

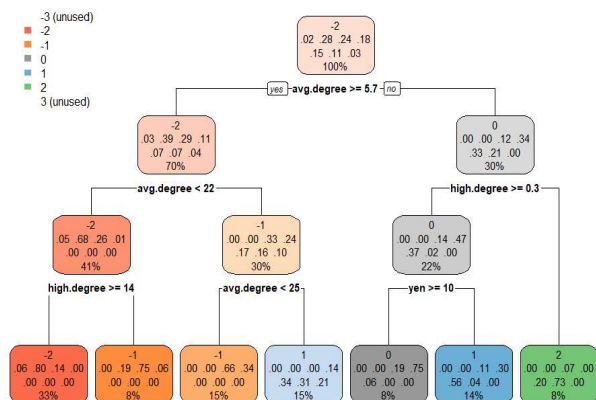


그림 7 기상요소와 기상 외 요소를 고려한 2018년 서울의 피크전력의 의사결정트리

Fig. 7 Decision tree of 2018 peak power in Seoul considering weather factors and out-of-climate factors

그림 7은 15개의 입력데이터와 2018년 평일 서울의 피크전력을 사용하여 도출된 의사결정나무이다. 마지막 7개의 노드가 분류된 데이터들을 의미하며, -3부터 3의 레벨로 구분된 피크전력들이 낮은 순서부터 분류되었음을 보여준다. 좌측부터 group 1, 2, ...라고 명명한다. .00으로 표시된 부분은 노드에 포함되는 전력의 레벨 경우를 표현하였으며 00%은 전체의 경우 중 그 노드에 포함되는 경우의 비율을 나타내었다. 가장 낮은 전력 레벨인 -3과 가장 높은 전력 레벨인 3은 노드에 나타나지 않았으나 가운데 노드에서 .21이라는 확률로 레벨 3의 전력의 경우가 포함되어 있는 것을 보아 평균기온이 25도 이상일 때 높은 전력수요에 주의해야한다. group1의 데이터양이 전체 중 33%로 가장 많고, 그 중 -3단계의 피크

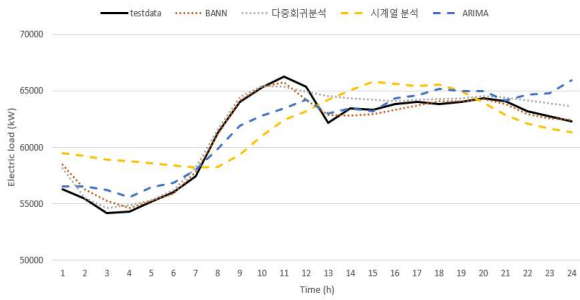
전력을 가지는 케이스가 6%, -2단계의 피크전력을 가지는 케이스가 80%이다.

본 논문에서는 시간별 전력수요예측의 정확성을 위해 마지막 노드를 기준으로 1년 데이터를 7가지로 그룹화하여 입력데이터에 따라 민감하게 변화하는 전력수요를 분석한다. 사계절 또는 월별로 나누어 전력수요예측을 진행하는 것은 일별 전력패턴을 평균치로 고려하기 때문에 시간별 전력수요예측에 오류를 유발할 수 있으며, 건물에너지 소비가 많은 도시형 네트워크 내에서 피크 값 예측에 불리할 수 있다.

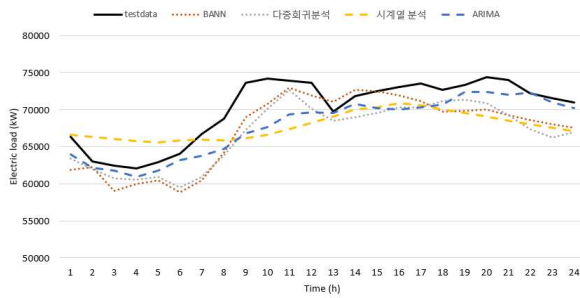
3.2 전력수요예측 결과 평가

전력수요예측을 위한 프로그램으로는 RStudio를 사용하였다. 본 논문에서는 활용한 B-ANN 모델의 적합성을 평가하기 위해 다중회귀분석, 시계열분석법과 ARIMA 모델을 사용한 각각의 전력수요 예측값의 오차를 비교하였다. 다중회귀분석은 두 개의 독립변수와 하나의 종속변수 간의 선형적 수식을 도출하여, 독립변수의 대입으로 결과를 도출 예측하는 기법이다. 독립변수는 기상요소와 기상 외 요소가 해당되며 종속변수는 시간별 전력수요가 해당된다. 시계열분석은 시간적 주기성을 갖는 데이터를 분석하여 예측하는 기법이다. 그래서 시간적 패턴을 가지는 시간별 전력수요예측에 적용될 수 있다. ARIMA 모형은 과거의 시간적 데이터를 기반으로 과거 데이터의 추세까지 반영하여 예측을 진행한다. 과거의 데이터를 학습하고 예측에 반영하는 기법으로 시간별 전력수요예측에 사용될 수 있다.

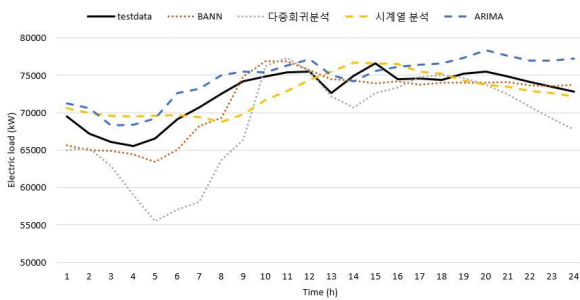
그림 8은 실제값인 testdata와 B-ANN, 다중회귀분석, 시계열분석 그리고 ARIMA 모델을 통해 동일데이터를 가지고 시간별 전력수요를 예측한 결과값을 보여준다. 그림 8의 (a)에서 B-ANN과 다중회귀분석이 좋은 결과값을 보여주었다. 그림 8의 (b)에서는 4개의 모형이 모두 실제값보다 낮은 값의 예측을 보였는데, 이는 group2에 낮은 피크값을 가지는 케이스가 19%정도 분포되어 있기 때문에 전체적인 예측값이 낮았던 것으로 해석된다. (c)에서는 10시부터 21시에 높은 전력이 분포되었다. B-ANN 모형은 15시의 전력이 늘어나는 부분을 예측하지 못하였다. 그래서 오차값이 2.9%로 약간 높게 측정되었다. (d)에서 ARIMA 모형은 전체적으로 높은 예측값을 도출하였고 B-ANN 모형은 16시의 과예측 외에는 좋은 결과값을 도출하였다. 그림 8의 (e)에서 B-ANN 모형은 10시와 11시를 피크전력으로 예측하였다. 전체적으로 시계열 분석은 평평한 전력값을 도출하였는데, 이는 주간 또는 계절별 단위의 주기성을 고려하지 않은 그룹 분류법 때문이라고 해석된다. 그림 8의 (f)에서 전력이 높아졌다가 낮아지는 21시와 22시에 대부분의 모델이 과한 예측값을 보였다. 이는 본 논문에서 고려한 입력데이터 외에 전력수요에 영향을 준 요인이 존재한다는 해석을 가능하게 한다. (g)에서 B-ANN과 다중회귀분석이 가장 좋은 예측을 보였다.



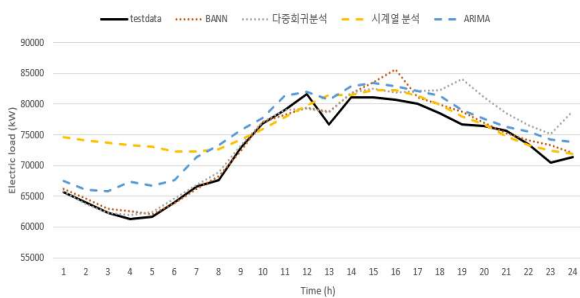
(a) group1



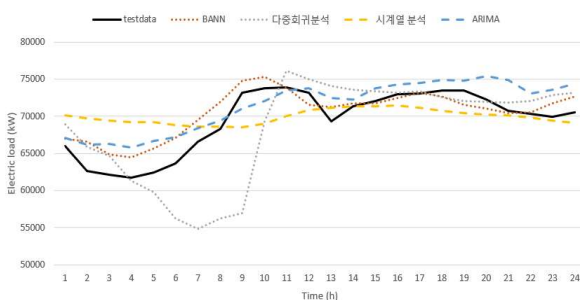
(b) group2



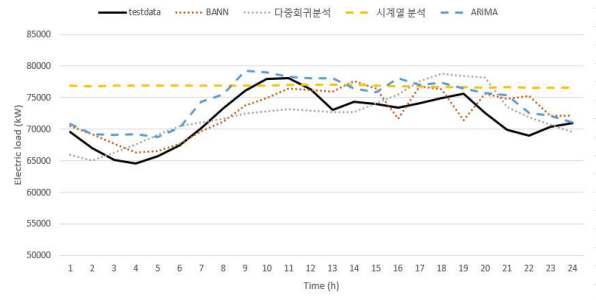
(c) group3



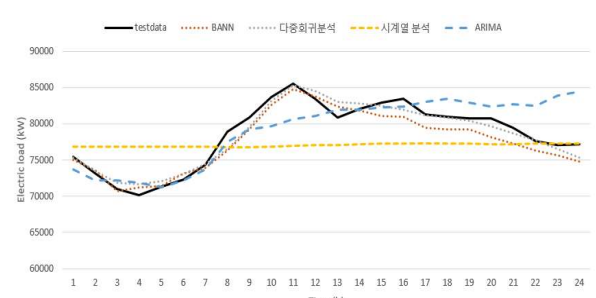
(d) group4



(e) group5



(f) group6



(g) group7

그림 8 각 그룹의 전력수요 실제값과 예측값

Fig. 8 Actual and predicted power demand for each group

표 2 각 케이스별 전력수요예측 오차값

Table 2 MAPE value for each case

	MAPE(%)							
	g1	g2	g3	g4	f5	g6	g7	avg
다중 회귀분석	0.986	4.628	5.016	2.703	4.948	4.727	1.391	3.486
시계열분 석	1.284	6.015	0.636	12.62	0.073	9.198	0.841	4.239
ARIMA 모델	0.701	3.154	2.945	4.002	3.027	3.795	0.825	2.635
B-ANN 모델	0.789	2.792	2.950	2.003	3.046	3.620	1.444	2.378

표 2는 전력예측을 수행하는 모형에 따른 각 그룹별 오차값을 나타낸다. 그룹별로 오차값의 차이가 존재하는 이유는 그룹별로 일(日) 수가 달라 데이터의 양의 차이가 존재하기 때문이다. 데이터 양의 차이에 의해 인공지능망이 학습하는 정도의 차이가 발생할 수 있다. 다만 데이터의 양이 많음이 항상 좋은 예측을 보이지는 않는다. group1에서는 네 가지 기법 모두 1.3%이하의 오차값을 가지기 때문에 7개의 그룹 중 가장 좋은 예측을 보인다. 특히 앞선 그림에서 볼 수 있듯이 피크전력 값이 존재하는 오후타임에 제안된 모델이 가장 정확하게 전력수요를 예측하고 있다. group2에서는 B-ANN 모델이 2.8%로 가장 낮은 MAPE 값을 가진다. 다음 그룹에서는 시계열 분석이 가장 낮은 오차값을 가졌지만, MAPE의 특성 상 24 시간동안의 예측 전력의 오차값을 나타냈기 때문에 부분적으

로 과학계 예측하고 적게 예측한 모델은 서로 보완되어 오차 값이 적게나왔다. 따라서 오차값 뿐만 아니라 시간별 예측값을 비교하며 모형의 적합성을 판단해야 한다. 앞선 그림에서, B-ANN 모델이 실제 전력데이터와 근사한 예측을 진행하였음을 보여준다. group4과 group6에서 가장 오차율이 낮은 모델은 본 논문이 제시한 B-ANN 모델이다. 시계열 분석은 데이터의 시간적 주기성을 기반으로 예측을 진행하기 때문에 입력데이터의 양에 따른 영향을 많이 받음을 보여준다. ARIMA 모델은 과거의 추세에 따른 전력예측을 진행하기 때문에 앞선 데이터들의 전력이 증가하는 추세라면 group6 처럼 과도한 전력을 예측할 수 있다. group5는 앞선 group3에서와 같은 이유로 시계열분석이 아닌, 10시~12시의 전력예측을 가장 정확하게 진행한 B-ANN 모델이 예측에 적합하다. 마지막으로 group7에서는 다중회귀분석과 B-ANN 모델이 좋은 예측값을 보인다. 평균적으로 본 논문에서 제시한 모형을 이용한 전력예측오차가 2.38%로 가장 낮은 MAPE 값을 가진다.

4. 결론

안정적인 전력수급과 경제적인 계통운영을 위해 정확한 전력수요예측은 필수적이다. 특히 단기수요예측은 발전원 운영 계획, 시장 가격 형성 등에 중요한 역할을 한다. 단기 전력수요예측에 가장 큰 영향을 주는 기상요인은 피크전력수요와 시간별 전력소비에 변동을 준다. 그리고 과거 전력패턴, 환율 또는 유가 등 기상 외 요인도 전력수요의 변동을 야기할 수 있다. 본 논문에서는 빅데이터를 효율적이고 직관적으로 분석할 수 있는 의사결정트리 기법을 기반으로 인공지능경망을 적용한 모델(B-ANN)을 제안하여 시간별 수요예측을 진행하였다. 의사결정트리 기법은 입력데이터와 전력소비의 상관성을 분석하여 빅데이터를 분류하였고 인공지능경망은 기존 통계적 기법보다 향상된 예측값을 도출하면서, 예측정확도를 향상하였다. 학습데이터로는 2018년 서울의 시간별 기상데이터, 환율 및 유가, 그리고 시간별 전력수요를 사용하였다. 정규화된 피크전력수요를 가지고 데이터를 그룹화하였으며, 그룹 데이터를 학습시켜 하루 24시간의 전력수요예측을 진행하였다. 제안된 B-ANN 모델의 적합성을 입증하기 위해 다중회귀분석과 시계열분석, ARIMA 모형으로 수요예측을 진행하여 결과를 비교하였다. 그 결과 다중회귀분석 모델의 오차는 3.49%, 시계열분석은 4.24%, ARIMA 모형의 오차는 2.64%, 인공지능경망 모델의 오차는 2.38%로, 본 논문에서 제시한 B-ANN 모델이 가장 낮은 오차값을 가진다.

향후 다양한 데이터를 이용하여 산업용, 가정용 등 전력사용 용도에 영향을 주는 데이터를 분석 및 분류하여 향상된 정확도를 갖는 모델을 구축하려 한다. 그리고 GDP와 경제성장을 등 입력데이터의 확장을 통해 B-ANN 모델이 중장기 전력수요예측에도 활용될 것으로 기대한다.

Acknowledgements

본 연구는 2020년도 한국연구재단(2020R1A2C1004743)의 지원과 2018년도 한국전력공사(R18XA06-75)의 지원에 의하여 이루어진 연구로서, 관계부처에 감사드립니다.

References

- [1] J. Buitrago and S. Asfour, "Short-Term Forecasting of Electric Loads Using Nonlinear Autoregressive Artificial Neural Networks with Exogenous Vector Inputs," *Energies*, vol. 10, pp. 1-24, Jan. 2017.
- [2] B. H. Koo, H. S. Kim, H. S. Lee, and J. H. Park, "Short-term electric load forecasting for summer season using temperature data," *Trans. KIEE*, vol. 64, pp. 1137-1144, 2015.
- [3] S. W. Jo, B. S. Kwon, and K. B. Song, "Day ahead 24-hours load forecasting algorithm using latest weather forecasting," *Trans. KIEE*, vol. 68, pp. 416-422, 2019.
- [4] D. H. Kim, H. J. Jo, M. S. Kim, J. H. Roh, and J. B. Park, "Short-term load forecasting based on deep learning model," *Trans. KIEE*, vol. 68, pp. 1094-1099, 2019.
- [5] C. H. Kim, "Analysis of the daily and hourly power load patterns," *Korea Energy Economics Institute Research Paper*, Mar. 2013.
- [6] C. H. Kim, "A study of mixed data sampling(MIDAS) model for electricity demand forecasting," *Korea Energy Economics Institute Research Paper*, Dec. 2014.
- [7] K. Kandanand, "Forecasting electricity demand in Thailand with an artificial neural network approach," *Energies*, vol. 4, pp. 1246-1257, Aug. 2011.
- [8] Y. C. Jung, "Utilization strategy of big data for official statistics," *Korea Information Society Development Institute (KISDI)*, 2016.
- [9] S. M. Gorade, A. Deo, and P. Purohit, "A study of some data mining classification techniques," *International Research Journal of Engineering and Technology (IRJET)*, vol. 04, pp. 3112-3115, Apr. 2017.
- [10] C. H. Kim, "Estimating short-term load forecasting in Korea using multiple exponential smoothing," *The Korea Electrotechnology Research Institute (KERI)*, 2013.
- [11] C. H. Kim, "Study on within-week and within-day seasonal patterns of electricity demand in Korea," *Korea Energy Economics Institute Policy Issue Paper*, Jun. 2013.
- [12] J. H. Moon, J. W. Park, S. H. Han, and E. J. Hwang, "Power consumption forecasting scheme for educational institutions based on analysis of similar time series data," *Journal of KIIE*, vol. 44, pp. 954-965, Sept. 2017.
- [13] Korea Electric Power Corporation, "Development of integrated demand management portal based on demand forecasting," Jan. 2009.
- [14] S. B. Roh and S. K. Oh, "Design of SVM-based polynomial neural networks classifier using particle swarm," *Trans. KIEE*,

- vol. 67, pp. 1071-1079, Aug. 2018.
- [15] S. R. Salkuti, "Short-term electrical load forecasting using hybrid ANN-DE and wavelet transforms approach," *Electrical Engineering*, vol.100, pp. 2755-2763, Dec. 2018.
- [16] J. Han and M. Kamber, "Data mining: concepts and techniques," Morgan Kaufmann Publishers, 2001.
- [17] S. H. Kwon, J. W. Lee, and G. H. Chung, "Snow damages estimation using artificial neural network and multiple regression analysis," *Journal of the Korean Society of Hazard Mitigation*, vol. 17, pp. 315-325, 2017.
- [18] Simon Haykin, "Feedforward neural networks : an introduction," 2004.
- [19] M. K. Kim, "A new approach to short-term price forecast strategy with an artificial neural network approach: application to the Nord pool," *Journal of Electrical Engineering& Technology*, vol. 10, July 2015.
- [20] J. N. Hwang, S. Y. Kung, M. Niranjana, and J. C. Principe, "The past, present, and future of neural networks for signal processing," *IEEE Signal Processing Magazine*, vol. 14, pp. 28-48, 1997.
- [21] P. Priyadarshi, "Rprop and improved Rprop+ based constant modulus type (RCMT) blind channel equalization algorithm for QAM signal," *Journal of Information and Optimization Sciences*, vol. 40, pp. 351-366, Mar. 2019.
- [22] Korea Meteorological Agency, "ASOS data set."
- [23] Korean Statistical Information Service, "Exchange rates, oil price, and power load data," OpenAPI.

저자소개



이지원 (Ji-Won Lee)

2020년 중앙대학교 에너지시스템공학부 졸업.
2020년~현재 동 대학원 에너지시스템공학과 석사 과정.



김형준 (Hyung-Jun Kim)

2019년 중앙대학교 에너지시스템공학부 졸업.
2019년~현재 동 대학원 에너지시스템공학과 석사 과정.



김문겸 (Mun-Kyeom Kim)

2004년 고려대학교 전기공학부 졸업.
2006년 서울대학교 대학원 전기·정보공학부 졸업 (석사).
2010년 동대학원 전기·정보공학부 졸업(공박).
2011년~2014년 동아대학교 전기공학과 조교수.
2014년~현재 중앙대학교 에너지시스템공학부 교수.