

The Prediction and Analysis of the Power Energy Time Series by Using the Elman Recurrent Neural Network

Chang-Yong Lee · Jinho Kim[†]

Dept. of Industrial and Systems Engineering, Kongju National University

엘만 순환 신경망을 사용한 전력 에너지 시계열의 예측 및 분석

이창용 · 김진호[†]

공주대학교 산업시스템공학과

In this paper, we propose an Elman recurrent neural network to predict and analyze a time series of power energy consumption. To this end, we consider the volatility of the time series and apply the sample variance and the detrended fluctuation analyses to the volatilities. We demonstrate that there exists a correlation in the time series of the volatilities, which suggests that the power consumption time series contain a non-negligible amount of the non-linear correlation. Based on this finding, we adopt the Elman recurrent neural network as the model for the prediction of the power consumption. As the simplest form of the recurrent network, the Elman network is designed to learn sequential or time-varying pattern and could predict learned series of values. The Elman network has a layer of “context units” in addition to a standard feedforward network. By adjusting two parameters in the model and performing the cross validation, we demonstrated that the proposed model predicts the power consumption with the relative errors and the average errors in the range of 2%~5% and 3kWh~8kWh, respectively. To further confirm the experimental results, we performed two types of the cross validations designed for the time series data. We also support the validity of the model by analyzing the multi-step forecasting. We found that the prediction errors tend to be saturated although they increase as the prediction time step increases. The results of this study can be used to the energy management system in terms of the effective control of the cross usage of the electric and the gas energies.

Keywords : Time Series Prediction, Elman Recurrent Neural Network, Power Energy Consumption, Forecast, Volatility

1. 서 론

산업사회의 필수적인 요소인 에너지의 효율적 관리는 에너지 생산 못지않게 중요한 문제로 대두되고 있다. 특히 상이한 유형의 에너지를 혼용하는 것은 잉여 에너지의 교차 활용 측면에서 효율적인 방법으로 간주되고 있으며, 에너지 혼용 중에서도 전기 에너지와 가스 에너지

의 교차 사용은 대표적인 사례가 될 수 있다. 에너지 교차 사용에서 시간에 따른 전기 에너지 사용량의 변화는 가스 에너지로 전환하는 시점을 결정하는데 중요한 요인이 된다.

본 연구에서는 A 테크노파크에서 2016년 측정된 전력 사용량 시계열 데이터를 사용하여 전력 사용량 예측을 위한 모형을 설정하고, 모형의 적합 및 교차 검증 등을 수행하여 예측의 정확성 등을 분석하고, 그 결과를 향후 에너지 혼용을 위한 관리 시스템 개발에 적용하고자 한다.

전력 에너지의 예측은 실용적인 측면이 비교적 강한 관계로 모형과 적용에 대해 많은 연구가 수행된 분야이다.

Received 6 February 2018; Finally Revised 13 March 2018;

Accepted 14 March 2018

[†] Corresponding Author : kjh@kongju.ac.kr

예측 모형 측면을 보면 뉴로퍼지 모형(neuro-fuzzy model)을 사용하여 다양한 리드타임에 대한 전력 수요를 예측한 연구[16, 20], 다중 회귀분석을 사용한 지역전력 수요 예측 연구[14], 지지도 벡터 회귀(support vector regression)를 이용한 월별 전력 수요 예측[13] 등이 있다. 적용 분야 측면을 보면, 기온과 전력 수요 간의 관계를 반영하여 최대 전력 수요를 예측한 연구[10], 장기간 최대 전력 수요 예측[11], 지역별 전력 사용량을 예측한 연구[1], 천연 자원의 가격 예측[19], 그리고 건조 공정의 에너지 수요 예측[12] 등을 들 수 있다. 기존의 많은 연구들은 시계열 데이터의 모형을 결정하기 전에 데이터에 내재한 특성에 대한 분석이 부족하다. 이것은 적절한 모형 설정과 예측에 부정적인 영향을 미치게 됨으로 데이터의 심층적 분석이 선행되어야 하고, 이를 통하여 설정된 모형의 정당성을 확보할 수 있다.

일반적으로 시계열 데이터 사이에는 시간에 대한 상관관계가 존재하기 때문에 시간 차이에 대한 상관관계의 크기와 유형은 예측 모형을 설정하는데 중요한 역할을 한다. 대부분의 시계열 데이터에는 선형 상관관계가 존재하기 때문에 이를 위한 모형으로 ARIMA(auto-regressive integrated moving average)[5] 혹은 선형 회귀 등을 사용한다. 만약 추가적으로 비선형(혹은 고차원) 상관관계가 존재하는 경우에는 이에 맞는 모형 설정이 필요하므로 모형을 선정하기 전에 고려하는 시계열 데이터의 특성을 파악하는 것이 중요하다.

일반적으로 비선형 상관관계는 측정된 시계열 데이터의 대상(여기서는 전력 에너지)의 특성이기 보다 시계열 데이터 자체의 특성에 더 가깝기 때문에, 모든 유형의 전력 에너지 시계열 데이터에 비선형 상관관계가 존재한다고 가정하기는 어렵다. 다만 비선형 상관관계의 존재 유무는 예측 모형을 설정하는데 중요한 요소로 작용할 수 있기 때문에 시계열 데이터의 대상에 무관하게 비선형 상관관계를 분석하는 것은 시계열 데이터 예측에 유용하다. 이러한 관점에서 본 논문에서는 A 테크노파크의 전력 에너지 시계열 데이터에 대하여 비선형 상관관계 분석을 수행하고, 그 결과에 부합한 예측 모형의 설정과 적용을 통하여 전력 에너지 사용량을 예측하였다.

이를 위하여 전력 사용량 시계열 데이터에 대하여 비선형 변환인 변동성(volatility)[15] 척도를 사용하여 표본 분산[21] 및 비경향 섭동 분석(DFA, detrended fluctuation analysis)[17]을 수행하였고, 이를 통하여 전력 사용량 시계열 데이터에는 비선형 상관관계가 존재함을 입증하였다. 이러한 결과를 바탕으로 비선형 상관관계를 포함할 수 있는 모형인 Elman 순환 신경망(Elman recurrent neural network)[7]에 시계열 데이터를 적용하여 전력 에너지 예측을 실행하였다. 데이터의 속성이 많은(혹은 차원이

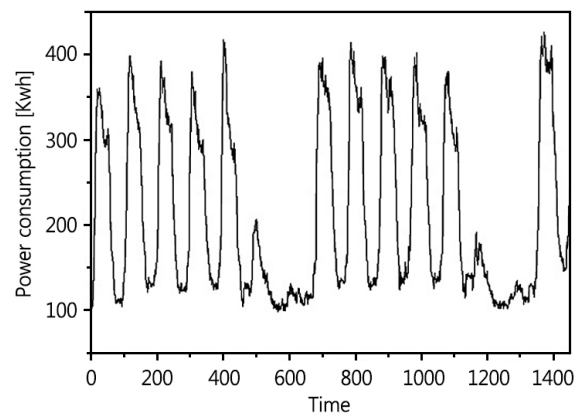
큰 경우에는 LSTM(long-short term memory)[8]과 같이 Elman 순환 신경망보다 더 복잡한 순환 신경망을 사용할 수 있으나, 본 연구에서 사용한 전력 에너지 시계열 데이터는 시간을 유일한 속성으로 가진 데이터임으로 이 경우에 적합한 Elman 순환 신경망을 사용하였다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 시계열 데이터의 비선형 변환인 변동성 시계열에 대하여 표본 분산 및 비경향 섭동 분석 등을 적용하여 전력 시계열 데이터 사이에는 비선형 상관관계가 존재함을 입증하였다. 이를 바탕으로 제 3장에서는 전력 사용량 예측을 위하여 비선형 예측 모형 중 하나인 Elman 순환 신경망을 전력 사용량 예측에 적용하였고, 모형 적합, 교차 검증, 다중 단계 예측 등을 수행하고 그 결과를 논하였다. 그리고 마지막 장에서는 본 논문의 결론과 향후 연구에 대하여 논하였다.

2. 전력 시계열의 비선형 상관관계

2.1 변동성 척도

본 연구에서 사용한 데이터는 2016년 1월 1일 0시에 2016년 12월 31일 24시까지 A 테크노파크에서 사용한 전력을 매 15분 간격으로 측정된 시계열 데이터이며, <Figure 1>에 그 일부를 나타내었다. 분석에 사용된 데이터는 전력 소모량(kWh 단위) 시계열 데이터로 시간 속성만 가지고 있으며 $x = \{x(1), x(2), \dots, x(N)\}$ 로 표현한다.



<Figure 1> A Part of the Power Consumption Time Series, The Horizontal Axis is in the Unit of 15 Minutes

시계열 데이터에 내재한 비선형 상관관계를 규명하기 위한 방법 중 하나는 시계열 데이터로부터 비선형 변환을 통해 생성한 시계열 데이터의 상관관계를 분석하는

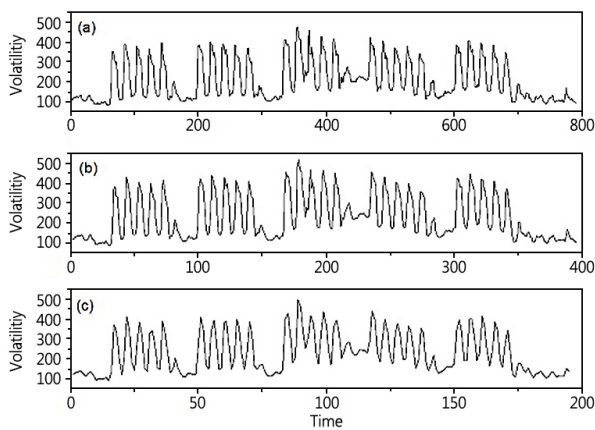
것이다. 비선형 변환을 통하여 생성한 시계열 데이터 사이에 상관관계가 존재한다는 것은 원래 시계열 데이터에 비선형 상관관계가 있다는 것을 간접적으로 입증하는 것임으로 전력 사용량 시계열 데이터에 내재한 비선형 상관관계를 분석하기 위하여 비선형 변환을 고려하였다.

일반적으로 많이 사용되는 비선형 변환은 금융 시장의 섭동(fluctuation)을 설명하기 위해 사용되는 척도인 변동성[15]이다. 변동성은 일정 기간 동안 인접한 데이터 사이의 변화량에 대한 비선형 변환으로 정의되는데 금융 시장으로 유입되는 정보의 양과 밀접한 관계가 있다. 즉, 일정 기간 동안 시장에 유입되는 정보가 크면 거래량도 이에 비례하고 따라서 변동성도 크게 된다. 특히 변동성은 Black-Scholes 모형[3]이나 Cox-Ross-Rubinstein 모형[6] 등 옵션 가격 모형에서 중요한 역할을 한다. 일반적으로 윈도우 크기(window size)가 w 인 경우 t 번째 변동성 $v_w(t)$ 는 시계열 데이터의 비선형 변화량인 절대값을 사용하여

$$v_w(t) \equiv \frac{1}{w} \sum_{i=wt-w+1}^{wt} |x(i+1) - x(i)|, \quad (1)$$

$$t = 1, 2, \dots, n = N/w$$

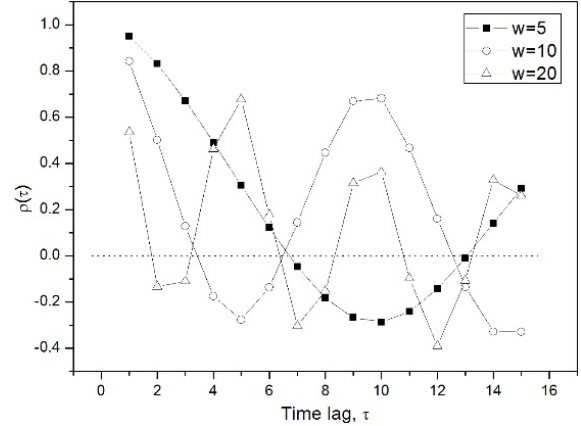
로 정의할 수 있다. 변동성은 주어진 윈도우 크기에서 비선형 변화량의 평균에 해당하며, $v_w(t)$ 역시 $x(t)$ 와 마찬가지로 시간 t 에 대한 시계열 데이터가 된다. 부수적으로, 2장 전체에서 N 은 전력 에너지 시계열 데이터의 개수를 나타내고, n 은 변동성 시계열 데이터의 개수를 나타낸다.



<Figure 2> Plots of Volatilities of Different Window Sizes : (a) $w = 5$, (b) $w = 10$, and (c) $w = 20$. The Horizontal Axis is in the Unit of 15 Minutes

서로 다른 윈도우 크기($w = 5, 10, 20$)에 대한 변동성 시계열 데이터를 <Figure 2>에 나타내었다. <Figure 2>를 통해 볼 때, 변동성의 전반적인 형태는 윈도우 크기에 대하

여 불변(invariant)임을 알 수 있는데, 이것은 변동성의 섭동은 윈도우 크기와 무관한 성질을 의미한다.



<Figure 3> Results of the Auto-Correlation Coefficients in Terms of the Time Lag with Different Window Sizes

변동성 시계열 데이터에 내재한 선형 상관관계를 분석하기 위하여 시간 지연 τ 에 따른 자기 상관 계수(auto-correlation coefficient)[5] $\rho(\tau) = \frac{\text{cov}\{v(t), v(t+\tau)\}}{\text{var}\{v(t)\}}$ 를 고려하였다. 자기 상관 계수는 시계열 데이터 사이의 선형 상관(linear correlation) 정도를 살펴보는 양으로 $-1 \leq \rho(\tau) \leq 1$ 로 주어진다. 특히 $\rho(\tau) = 1$ 은 완전 상관, $\rho(\tau) = -1$ 은 완전 반상관(anti-correlation), 그리고 $\rho(\tau) = 0$ 은 상관관계가 없음을 나타낸다. 서로 다른 윈도우 크기에 대한 변동성 시계열 데이터에 대하여 자기 상관 분석을 수행하였으며, 그 결과를 <Figure 3>에 나타내었다. <Figure 3>을 통해 볼 때, 윈도우 크기가 클수록 큰 시간 지연에 대한 자기 상관관계가 양수임으로 이 시간 지연에 대하여 선형 상관관계가 존재함을 알 수 있다.

2.2 표본 분산 분석

변동성 시계열 $v_w(t)$ 를 간편하게 표현하기 위하여 윈도우 크기 w 을 생략하여 $v(t)$ 라 표현하고, $v(t)$ 사이의 상관관계를 분석하기 위해 표본 분산(sample variance)을 사용하였다. 표본 분산은 평균의 분산에서 출발한다. 변동성 시계열 데이터 $v = \{v(t) : t = 1, 2, \dots, n\}$ 가 분산이 σ^2 인 모집단에서 추출된 값이라 가정하면 [즉, $\text{var}(v(t)) = \sigma^2$, $t = 1, 2, \dots, n$], 변동성의 평균인 \bar{v}_n 의 분산은

$$\begin{aligned} \text{var}(\bar{v}_n) &= n^{-1}\sigma^2 + n^{-2}\sum_{i \neq j}^n \text{cov}\{v(i), v(j)\} \\ &= n^{-1}\sigma^2 + n^{-2}\sigma^2\sum_{\tau=1}^n \rho(\tau) \end{aligned} \quad (2)$$

으로 주어진다. 여기서 $\rho(\tau)$ 는 시간 지연(time delay) τ 에 대한 자기 상관 함수(auto-correlation function)이다. 만약 시계열 데이터가 서로 독립이면 $\rho(\tau) = 0$ 이 됨으로 $var(\bar{v}_n) \propto n^{-1}$ 이 된다. 이를 멱급수(power-law) 형태인 $var(\bar{v}_n) \propto n^{-\beta}$ 로 표현하면, $\beta = 1$ 에 해당한다. 만약 $\tau \gg 1$ 인 경우에 $\rho(\tau) \neq 0$ 이면 $var(\bar{v}_n) \propto n^{-\beta}$ 에서 $\beta < 1$ 임이 알려졌다[21]. 이것은 만약 데이터 사이에 상관관계가 존재하면 표본 평균의 분산은 독립인 경우보다 더 천천히 감소함을 의미한다. 즉, 시계열 데이터 사이에 상관관계가 존재하면 표본 분산은 표본 개수에 반비례하는 것보다 더 느리게 변할 것임으로 $\beta < 1$ 이 된다. 따라서 데이터 사이의 상관관계 여부를 분석하기 위해서는 지수 β 를 추정해야 하는데, 본 연구에서는 아래의 방법[21]을 사용하였다.

Step 1 : 주어진 n 개의 시계열 데이터를 k 개의 데이터로 구성된 m_k 개의 부분집합으로 나눈다. 즉, $n = km_k$ 이 되도록 k 와 m_k 를 정한다.

Step 2 : j 번째 부분집합에 속한 k 개 표본에 대한 평균을 $\bar{v}_j(k) = \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} v(i)$ 라 하면, 주어진 시계열 데이터는 m_k 개의 부분집합으로 구성됨으로 m_k 개의 표본 평균 $\bar{v}_j(k) (j=1, 2, \dots, m_k)$ 을 계산할 수 있다. 표본 분산 $s^2(k)$ 는 m_k 개의 표본 평균 $\bar{v}_j(k) (j=1, 2, \dots, m_k)$ 에 대한 분산이며 다음과 같이 계산한다. $s^2(k)$ 는 k 에 대한 함수로

$$s^2(k) \equiv \frac{1}{m_k - 1} \sum_{j=1}^{m_k} \{ \bar{v}_j(k) - \bar{v}(k) \}^2 \quad (3)$$

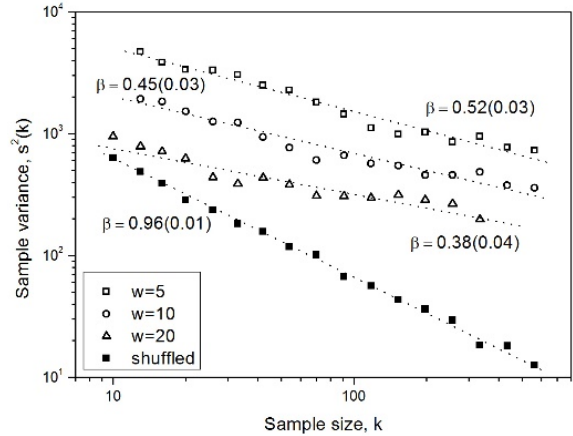
으로 주어지며, 여기서 $\bar{v}_j(k) \equiv \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} v(i)$,

$\bar{v}(k) \equiv \frac{1}{m_k} \sum_{j=1}^{m_k} \bar{v}_j(k)$ 이다.

Step 3 : $s^2(k) \propto k^{-\beta}$ 에서 양변에 로그(logarithm)를 취하여 $\log k$ 에 대하여 $\log s^2(k)$ 의 변화인 기울기 β 를 추정한다.

윈도 크기 $w = 5, 10, 20$ 에 대한 변동성 시계열 데이터를 사용하여 표본 분산을 구하였으며, 그 결과를 <Figure 4>에 나타내었다. <Figure 4>를 통해 볼 때 모든 윈도 크기에 대하여 $\hat{\beta} < 1$ 임으로 주어진 윈도 크기에서 변동성 데이터 사이에는 시간적으로 상관관계가 있음을 알 수 있다. 특히 윈도 크기가 커짐에 따라 표본 분산은 더 느리게 감소함을 알 수 있다(즉, w 이 커짐에 따라 추정된 β 는 작아짐). 이것은 윈도 크기가 큰 변동성은 작은 변

동성에 비하여 상대적으로 장기간에 걸친 변화량의 절대 값에 대한 평균임으로, 만약 데이터 사이에 상관관계가 있다면 장기간에 걸친 변화량에 대한 상관관계는 단기간 경우보다 클 것으로 예상할 수 있기 때문이다.



<Figure 4> Results of the Sample Variance with Different Window Size Together with the Result of the Shuffled Data

이러한 결과를 통해 볼 때 고려하는 시계열 데이터 사이에는 비선형(혹은 고차원) 상관관계가 존재함을 알 수 있다. 이 결과를 뒷받침하기 위하여 변동성 시계열 데이터를 무작위로 섞어 시간에 따른 순서를 고려하지 않은 데이터(shuffled data)에 대하여 β 를 추정한 결과를 <Figure 3>에 추가로 나타내었다. <Figure 3>에서 볼 수 있듯이 이 경우의 추정치 $\hat{\beta} = 0.96 \approx 1$ 으로 변동성 데이터를 무작위로 섞은 시계열 데이터는 시간에 대해 서로 독립임을 알 수 있다.

2.3 비경향 섭동 분석

표본 분석을 통해 규명한 변동성 시계열 데이터의 상관관계에 대한 유형을 분석하기 위하여 비경향 섭동 분석(DFA, Detrended Fluctuation Analysis)[17]을 변동성 시계열 데이터에 적용하였다. DFA는 일반적인 평균자승평방근(root-mean-square) 분석을 수정한 것으로 시계열 데이터에 내재할 가능성이 있는 지역적 경향(local trend)을 제거하여 순수한 장거리 상관관계(long-range correlation)를 규명하기 위한 방법이다. DFA는 시계열 데이터에 존재할 수 있는 지역적 경향을 제거하기 때문에 인위적인 요인을 배제할 수 있는 장점이 있다.

n 개 변동성 데이터 $v = \{v(t) | t = 1, 2, \dots, n\}$ 를 크기가 T 인 상자(box)로 구분하여 M 개 상자로 나누면, $n = TM$ 을 만족한다. 비경향 섭동 함수는 각 상자에 대하여 데이터의

평균 섭동 정도를 나타내는데, 이 때 데이터의 지역적 경향을 나타내는 회귀식을 사용하여 데이터의 경향성을 제거한다. 즉, 상자 크기가 T 일 때 k 번째 상자에 대한 비경향 섭동 함수(detrended fluctuation function) $F_k(T)$ 는

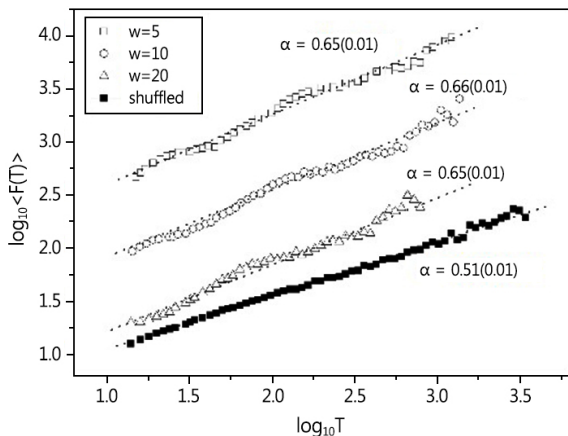
$$F_k(T) = \sqrt{\frac{1}{T} \sum_{i=kT-T+1}^{kT} \{v(i) - z_{i,T}\}^2}, \quad (4)$$

$$k = 1, 2, \dots, M$$

으로 주어진다. 식 (4)에서 $z_{k,T}$ 는 크기가 T 인 k 번째 상자에 속한 데이터의 지역적 경향을 선형 회귀 모형으로 적합시킨 회귀식을 나타내며, $z_{i,T} = \hat{a}_{k,T}i + \hat{b}_{k,T}$ 로 표현할 수 있다. 여기서 $\hat{a}_{k,T}$ 와 $\hat{b}_{k,T}$ 는 각각 선형 회귀식의 기울기와 절편을 나타내며 상자 크기 T 와 상자 순서 k 에 의존한다.

M 개의 $F_k(T)$ 에 대한 평균을 $\langle F(T) \rangle = \frac{1}{M} \sum_{i=1}^M F_k(T)$ 라 두면, 데이터의 섭동을 상자 크기 T 에 대한 함수로 표현할 수 있다. 특히 $\langle F(T) \rangle$ 가 상자 크기 T 에 대하여 축척(scaling)에 무관한 성질을 가지고 있는 경우에는 $\langle F(T) \rangle \propto T^\alpha$ 로 표현할 수 있다. 여기서 지수 α 는 축척 지수로 양의 값을 가지며, 축척 지수 값을 통하여 고려하는 시계열 데이터의 특성을 구별할 수 있다. 즉, $\alpha = 0.5$ 인 경우에는 시계열 데이터는 서로 독립으로 상관관계가 없으며, $0.5 < \alpha < 1$ 인 경우에는 데이터들 사이에 장거리 상관관계가 존재한다[17].

서로 다른 윈도우 크기 $w=5, 10, 20$ 에 대한 변동성 시계열 데이터에 대하여 DFA를 실행하였으며, 그 결과를 <Figure 5>에 나타내었다. <Figure 5>에서 볼 수 있듯이 윈도우 크기 w 에 무관하게 DFA의 축척 지수는 $0.5 < \alpha < 1$ 임으로 변동성 시계열 데이터 사이에는 장거리 상관관계가 존재함을 알 수 있다. 이에 반하여 변동성 시계열 데이



<Figure 5> The Long-Range Correlation of the Volatilities and the Shuffled Data by Using DFA

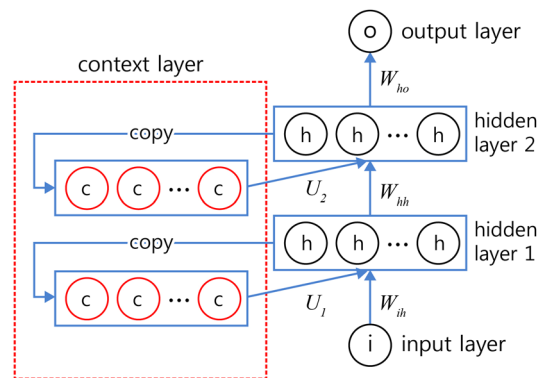
터를 무작위로 섞어 만든 시계열 데이터에 대한 DFA 축척 지수는 $\alpha \approx 0.5$ 임으로 데이터 사이에는 서로 독립임을 알 수 있다. <Figure 5>를 통해 볼 때, 또한 섭동성의 상관관계 크기는 윈도우 크기 w 에 대체적으로 무관하고, 윈도우 크기는 변동성을 정의하는 시간 범위에 해당하므로 섭동성은 시간 범위에 의존하지 않는 특성을 가지고 있음을 알 수 있다. 따라서 전력 시계열 데이터의 비선형 상관관계는 정상적(stationary)이지 않는 성질을 가지고 있음을 알 수 있다.

3. 순환 신경망을 사용한 시계열 데이터 예측

3.1 Elman 순환 신경망

제 2장의 분석 결과를 통해 볼 때, 전력 사용량 시계열 데이터는 선형 상관관계에 더하여 비선형 상관관계가 존재함을 알 수 있었다. 특히 시간에 대한 비선형 상관관계가 존재함을 변동성 시계열 데이터의 분석을 통하여 입증하였으므로 과거를 “기억”하는 기능이 포함된 비선형 모형이 적합함을 알 수 있다. 이를 위하여 본 연구에서는 최근 심층 학습(deep learning)[18]과 연관되어 주목을 받고 있는 비선형 모형 중에서 과거의 기억을 효율적으로 포함시킬 수 있는 Elman 순환 신경망[7]을 전력 사용량 시계열 데이터의 예측 모형으로 사용하고자 한다. Elman 순환 신경망은 일반적인 전향 전달 신경망(feed-forward neural network)에 기억을 위한 계층을 추가로 포함하고 있다(<Figure 6> 참조).

Elman 순환 신경망은 시간 t 에서 입력값 $x(t)$ 를 사용하여 다음 시간 $t+1$ 의 출력값 $x(t+1)$ 을 예측하기 위한 것으로 일반 신경망과 마찬가지로 입력(input layer), 은닉(hidden layer), 그리고 출력층(output layer) 등 세 종류의 층으로 구성되어 있고, 전향 전달과 역 전파(feed-for-



<Figure 6> The Diagram of an Elman Recurrent Neural Network Used in this Study

ward back-propagation)의 특징을 가지고 있는 신경망이다. <Figure 6>에서 볼 수 있듯이 Elman 순환 신경망은 일반 신경망 구조에 더하여 기억 능력을 보유하기 위하여 맥락층(context layer)을 추가로 가지고 있다.

시간 $t+1$ 에서 맥락층 노드는 시간 t 의 은닉층 노드를 복사한 것으로 은닉층의 과거 시간에서 값을 저장하는 역할을 한다. 따라서 Elman 순환 신경망에서는 입력층 데이터만 은닉층으로 전달되는 것이 아니라 추가로 과거를 기억하고 있는 맥락층 데이터 역시 은닉층으로 전달되는 방식을 취한다. 즉 과거의 출력이 다시 입력이 되는 피드백(feedback) 방식으로 기억 능력을 보유하도록 함으로, Elman 순환 신경망은 두 유형의 입력 데이터(현재 입력층과 과거 맥락층)를 가지고 있다. 또한 맥락층 노드 개수는 은닉층 노드 개수와 동일하고, 초기 시간 $t=1$ 에서 맥락층은 아무런 역할을 하지 않는다.

본 연구에서 사용한 Elman 순환 신경망을 수식으로 표현하면 다음과 같다.

$$\begin{aligned}
 t=1 : h_1(t) &= \phi_{h_1}(W_{ih} \cdot x(t)), & (5) \\
 h_2(t) &= \phi_{h_2}(W_{hh} \cdot h_1(t)), \\
 \hat{x}(t) &= \phi_o(W_{ho} \cdot h_2(t)).
 \end{aligned}$$

$$\begin{aligned}
 t > 1 : h_1(t) &= \phi_{h_1}(W_{ih} \cdot x(t) + U_1 \cdot h_1(t-1)), & (6) \\
 h_2(t) &= \phi_{h_2}(W_{hh} \cdot h_1(t) + U_2 \cdot h_2(t-1)), \\
 \hat{x}(t) &= \phi_o(W_{ho} \cdot h_2(t)).
 \end{aligned}$$

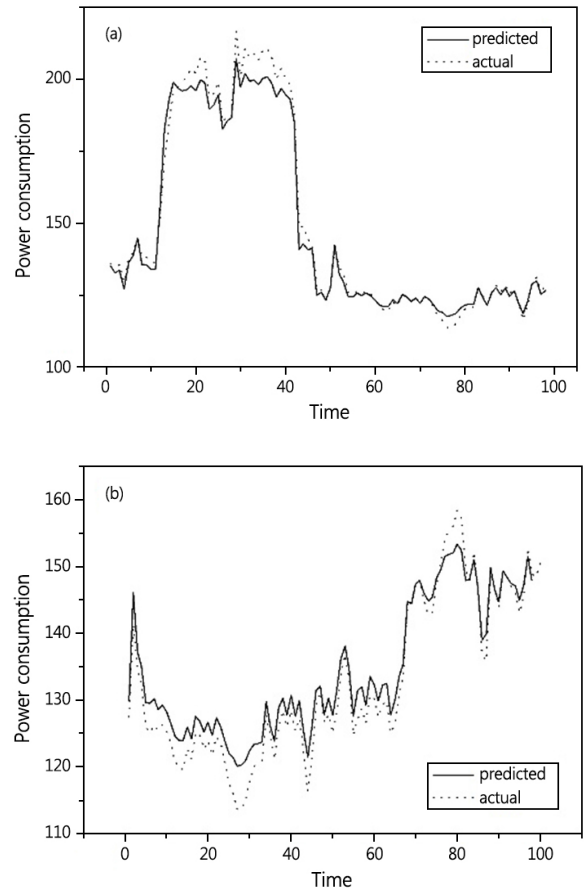
여기서 $x(t)$, $h_1(t)$ 와 $h_2(t)$, 그리고 $\hat{x}(t)$ 는 노드값들로 각각 입력층, 은닉층 1과 은닉층 2, 그리고 출력층 값을 나타낸다. 또한 W_{ih} , W_{hh} , W_{ho} 는 각각 입력층과 은닉층, 은닉층 1과 은닉층 2, 그리고 은닉층 2와 출력층 사이의 노드들을 연결하는 가중치를 나타내며, U_1 과 U_2 는 각각 맥락층과 은닉층 1 그리고 맥락층과 은닉층 2 사이의 노드들을 연결하는 가중치를 나타낸다. 마지막으로 $\phi_{h_1}(\cdot)$, $\phi_{h_2}(\cdot)$, $\phi_o(\cdot)$ 는 활성화함수(activation function)로 보통 시그모이드 함수(sigmoid function) $\phi(x) = \frac{1}{1+e^{-x}}$ 를 사용한다.

3.2 Elman 순환 신경망의 모수 조정 및 적합

Elman 순환 신경망의 학습은 예측값(혹은 출력값) $\hat{x}(t+1)$ 와 목표값 $x(t+1)$ 의 차이를 사용한 손실 함수(혹은 비용함수)가 최소가 되도록 모형의 모수인 가중치 W 와 U 를 최적화하는 것을 의미한다. 다양한 손실 함수가 가능하나 일반적으로 평균제곱오차(mean square error) $L(W, U) = \sum_t \{\hat{x}(t+1) - x(t+1)\}^2$ 을 가장 많이 사용한다. 일반적인

신경망과 유사하게 Elman 신경망은 두 가지 모수, 은닉층의 노드 개수와 학습율(learning rate)을 가지고 있다. Elman 신경망의 입력값과 출력값은 각각 $x(t)$ 와 $\hat{x}(t+1)$ 임으로 입력층과 출력층의 노드는 1개인 반면, 은닉층 노드 개수는 정해진 값이 없고 매개변수에 해당한다. 또한 신경망에서 가중치의 학습은 오차 역전파법(error backpropagation) [18]을 적용한 경사 하강법(gradient descent)을 주로 사용하여 손실 함수가 최소가 되도록 가중치를 갱신한다. 가중치의 갱신은 $W \leftarrow W - \alpha \frac{\partial L(W; U)}{\partial W}$ 와 $U \leftarrow U - \alpha \frac{\partial L(W; U)}{\partial U}$ 로 표현할 수 있고, 이때 α 를 학습율이라 한다.

Elman 신경망의 실행을 위하여 RSNN R-package[9]를 사용하였으며, 최적의 모수 값을 선택하기 위하여 두 모수의 값들을 변화시키면서 모형의 예측 성능을 평균제곱근오차(식 (9))를 사용하여 비교하였다. 그 결과 $\alpha = 0.1$ 그리고 은닉층 노드 개수는 12개일 때 예측 성능이 가장



<Figure 7> The Comparison of the Fitted Values to the Elman Neural Network [solid line] and the Actual Time Series [dotted line] for Two Different Region in time. The Values of the Parameters are $\alpha = 0.1$ and the Number of Hidden Node is 12. The Horizontal Axis is in the Unit of 15 Minutes

우수하여 이 값들을 모수로 조정하였다. 조정된 모수와 전력 사용량 시계열 데이터를 사용하여 Elman 신경망을 학습시켰으며, 그 결과의 일부를 <Figure 7>에 나타내었다. <Figure 7>을 통해서 볼 수 있듯이 Elman 신경망 모형은 전력 사용량 시계열 데이터를 비교적 충실히 적합하고 있음을 알 수 있다.

3.3 교차 검증

Elman 신경망을 사용한 예측 모형의 검증을 위하여 교차 검증(cross validation)을 실행하였다. 시계열 데이터의 교차 검증은 인접한 데이터들 사이에 상관관계로 인하여 일반 데이터를 위한 교차 검증 방법을 그대로 적용하기 힘들다.

본 논문에서 사용한 방법은 시계열 데이터의 교차 검증에서 가장 널리 사용하는 방법으로, “evaluation on a rolling forecasting origin”[2]이라 불리는데, 이것은 “예측을 검증하는 시점(origin)이 시간이 증가하는 방향으로 두루마리처럼 말린다”는 의미이다. N 개로 구성된 시계열 데이터에 대한 교차 검증에서 훈련 데이터(train data)와 시험 데이터(test data)는 교차 검증을 실행하는 시간인 $\tau(\tau \leq N-1)$ 를 기준으로

$$x_{train} = \{x(1), x(2), \dots, x(\tau)\}, \quad x_{test} = x(\tau+1) \quad (7)$$

로 나눈다. 훈련 데이터는 Elman 신경망 모형의 학습(혹은 적합)을 위해 사용하고, 시험 데이터는 학습된 모형의 성능을 평가하는데 사용한다. 초기값 $\tau = \tau_0$ 부터 $\tau = N-1$ 까지 τ 값을 1씩 증가시키면서 x_{train} 을 사용하여 모형의 학습과 예측을 수행하기 때문에 훈련 데이터가 점차적으로 증가하는 특징을 가지고 있다(<Figure 8(a)> 참조).

위의 방법과 유사하나 훈련 데이터의 크기를 고정시키는 방법도 시계열 데이터의 교차 검증 방법으로 널리 사용된다. 이 방법은 훈련 데이터의 고정된 크기를 t_0 ($\tau > t_0$)라 하면 아래 식으로 표현할 수 있다(<Figure 8(b)> 참조).

$$x_{train} = \{x(\tau - t_0 + 1), x(2), \dots, x(\tau)\}, \quad (8)$$

$$x_{test} = x(\tau+1)$$

본 논문에서 사용한 두 가지 교차 검증 절차를 의사 코드로 표현하면 <Figure 8>과 같다.

위에서 언급한 두 가지 교차 검증 방법을 훈련 데이터의 개수를 조절하면서 실행하였다. 즉, 전체 데이터에서 훈련 데이터가 차지하는 비율 p 를 0.5에서 0.9까지 0.1

(a) Time series: $\{x(1), x(2), \dots, x(N)\}$

```
For  $\tau = \tau_0$  to  $N - 1$ 
  set  $x_{train} = \{x(1), x(2), \dots, x(\tau)\}$ 
  model  $\leftarrow$  Elman( $x_{train}$ )
   $\hat{x}(\tau + 1) \leftarrow$  prediction(model)
  compare  $\hat{x}(\tau + 1)$  with  $x(\tau + 1)$ 
End For
```

(b) Time series: $\{x(1), x(2), \dots, x(N)\}$ and t_0

```
For  $\tau = \tau_0$  to  $N - 1$ 
  set  $x_{train} = \{x(\tau - t_0 + 1), x(2), \dots, x(\tau)\}$ 
  model  $\leftarrow$  Elman( $x_{train}$ )
   $\hat{x}(\tau + 1) \leftarrow$  prediction(model)
  compare  $\hat{x}(\tau + 1)$  with  $x(\tau + 1)$ 
End For
```

<Figure 8> Pseudo Code for the Cross Validation : Method (a) and method (b)

단위로 변화시키면 두 가지 교차 검증 방법을 실행하였다. 따라서 식 (7)과 식 (8)에서 τ 의 초기값은 $\tau_0 = Np$ 이고 예측값 개수는 $N - \tau_0$ 개가 된다. 또한 교차 검증 방법 (b)의 경우에는 훈련 데이터 개수를 $t_0 = 0.5N$ 으로 고정시켰다.

Elman 신경망 모형의 성능 평가를 위해 실제 값과 예측 값을 사용한 예측의 정확도는 다양한 척도를 통해 나타낼 수 있는데, 본 연구에서는 평균상대오차백분율(MAPE, Mean Relative Percentage Error)과 평균제곱근오차(RMSE, Root Mean Square Error)를 사용하였다.

$$MAPE = \frac{1}{N - \tau_0} \sum_{\tau = \tau_0}^{N-1} \left| \frac{x(\tau+1) - \hat{x}(\tau+1)}{x(\tau+1)} \right| \times 100, \quad (9)$$

$$RMSE = \sqrt{\frac{1}{N - \tau_0} \sum_{\tau = \tau_0}^{N-1} \{x(\tau+1) - \hat{x}(\tau)\}^2}$$

<Table 1> The Results of MAPE and RMSE with Different Ratio p . The Variances are in the Parentheses. The Used Parameters are $\alpha = 0.1$ and the Number of Nodes in the Hidden Layer is 12

	p	0.5	0.6	0.7	0.8	0.9
MAPE	method (a)	2.69	3.05	1.94	3.79	1.55
	method (b)	3.05	2.95	2.51	4.19	1.64
RMSE	method (a)	2.53	2.78	6.73	4.82	4.85
	method (b)	2.93	2.73	7.60	5.01	5.19

위의 방법을 적용하여 예측의 성능을 비교하였고, 그 결과를 <Table 1>에 내었다. <Table 1>을 통해 볼 때, 상대오차는 약 2%~5%이고 평균오차는 약 3kWh~8kWh이고, 성능은 훈련 데이터의 크기인 비율 p 에 대체적으로 무관함을 알 수 있다. 또한 MAPE와 RMSE의 결과를 통해 볼 때 두 가지 검증 방법은 큰 차이가 없음을 알 수 있다

3.4 다중 단계 시계열 예측

시계열 예측은 일반적으로 다음 시간 단계에 대한 예측 즉, 단일 단계 예측(one-step forecast)을 의미하는데, 가끔 다중 단계 예측(multi-step forecast)이 필요한 경우가 있다. 다중 단계 예측을 위한 방법은 직접법(direct multi-step forecast), 재귀법(recursive multi-step forecast), 그리고 혼합법(direct-recursive hybrid forecast) 등 세 가지가 있다[4]. 직접법은 훈련 데이터를 사용하여 각 단계마다 독립된 모델을 적합 시키는 방법으로 h 단계 예측치를 위한 모형은 $\hat{x}(\tau+h) = f_h(x(t), x(\tau-1), \dots, x(\tau-t_0+1))$ 로 주어진다. 이 방법은 예측 모형 $f_h(\cdot)$ 이 서로 다른 시간 단계 h 에 대해 독립적이기 때문에 모형들 사이의 의존성을 파악하기 어려운 단점이 있다. 재귀법은 동일한 모형을 여러 시간 단계에 공통으로 적용하는데, 이 때 이전 단계에서 예측한 값을 다음 단계의 예측을 위한 훈련 데이터의 일부로 사용하기 때문에 h 단계 예측치를 위한 모형은 $\hat{x}(\tau+h) = f_h(\hat{x}(\tau+h-1), x(\tau+h-2), \dots, x(\tau-t_0+1))$ 로 표현된다. 따라서 재귀법은 예측 오차가 계속 누적되어 예측 단계가 길어짐에 따라 성능이 급속히 떨어지는 단점이 있다.

혼합법은 직접법과 재귀법을 결합한 것으로 이전 단계의 예측치를 훈련 데이터의 일부로 사용하여 다음 단계의 예측을 실행하되 독립된 모형을 적합시키는 방법으로 $\hat{x}(\tau+h) = f_h(\hat{x}(\tau+h-1), \hat{x}(\tau+h-2), \dots, x(\tau-t_0+1))$ 로 표현할 수 있다. 혼합법은 직접법처럼 독립된 모형을 사용하되 혼합법처럼 예측치를 훈련 데이터의 일부로 사용함으로써 모형들 간의 의존성을 고려하고, 또한 이전 단계의 예측치를 사용하여 독립된 모형을 다시 적합시킴으로써 오차의 누적을 방지할 수 있다. 이러한 장점으로 인해 본 논문에서는 혼합법을 사용하여 다중 단계 시계열 예측을

실행하였다. 다중 단계 $m=1, 2, \dots, 10$ 에 대하여 제 3.3절의 두 교차 검증 방법을 적용하여 예측을 실행하였으며 그 결과를 <Table 2>에 나타내었다. <Table 2>를 통해 볼 수 있듯이 다중 단계가 증가할수록 예측 오차는 커지는 경향이 있으나 급격하게 증가하지 않고 대체적으로 수렴함을 알 수 있다. 또한 두 가지 교차 검증 방법을 사용한 결과 역시 큰 차이가 없으며 시간 단계에 비교적 무관함을 알 수 있다.

4. 결론 및 요약

본 연구에서는 전기와 가스 에너지의 혼용을 위한 선행적 연구로 전력 에너지 사용량 예측을 위한 모형과 그 방법에 대한 연구를 수행하였다. 이를 위하여 A 테크노파크에서 2016년에 15분 간격으로 측정된 전력 사용량 시계열 데이터를 Elman 순환 신경망 모형에 적합하고 그 결과를 예측에 적용하였다.

전력 사용량 시계열 데이터를 위한 모형을 설정하기 위하여 시계열 데이터에서 비선형 변환된 변동성 시계열 데이터에 대하여 표본 분산과 비경향 섭동 분석 등을 사용하여 상관관계를 분석하였다. 변동성 시계열 데이터 사이에는 무시할 수 없을 정도의 상관관계가 내재하였고, 이것은 전력 사용량 시계열 데이터에는 비선형 상관관계가 존재함을 의미하였다. 이러한 결과를 바탕으로 시계열 데이터에 내재하는 비선형 상관관계를 포함할 수 있는 비선형 예측 모형인 Elman 순환 신경망을 설정하여 전력 에너지 예측에 적용하였다.

Elman 순환 신경망을 적용하기 위하여 검정 데이터를 사용하여 모수인 학습율과 은닉층 노드 개수 최적화를 수행하였으며, 최적의 매개변수 값과 시계열 데이터를 사용하여 모형을 적합시켰다. 모형의 적합은 실제 시계열 데이터를 충분히 설명할 수 있었고, 이를 바탕으로 시계열 데이터의 교차 검증을 위해 널리 사용되는 두 가지 방법을 적용하여 모형의 정확도를 검증하였다. 또한 예측의 정확도를 평가하기 위하여 평균상대오차백분율과 평균 제공근 오차 등 두 가지 척도를 사용하였다. 평가 결과를 통해 볼 때, 적합된 모형의 예측치는 훈련 데이터의 크기에 대체적으로 무관하였으며 상대오차는 약 2%~5%이고

<Table 2> The Accuracy of Prediction for the Multi-Step Forecast by $p = 0.7$

	m	1	2	3	4	5	6	7	8	9	10
MAPE	method(a)	0.72	1.54	3.42	4.87	3.91	4.08	2.61	4.11	4.90	5.03
	method(b)	0.69	1.33	3.81	5.20	4.34	4.55	2.98	4.55	5.38	5.28
RMSE	method(a)	1.31	3.36	5.40	7.49	6.54	6.77	4.84	6.68	7.74	8.92
	method(b)	1.18	2.04	8.83	8.61	7.69	7.98	6.24	7.94	8.80	8.71

평균오차는 약 3kWh~8kWh임을 알 수 있었다. 본 연구에서 제안한 방법의 유용성을 검증하기 위하여 기존 연구 [1, 10, 11, 13, 14, 16, 20, 19]들의 결과와 비교하였다. 7개의 기존 연구의 상대오차에 대한 평균과 표준편차는 각각 6.54%와 4.86%이었으며, 이것은 본 연구에서 제시한 방법의 유용성을 입증하는 결과라 할 수 있다.

단일 단계 예측에 더하여 다중 단계 예측도 실행하였으며, $m = 1, 2, \dots, 10$ 에 대하여 두 가지 교차 검증을 실행하였고, 또한 두 정확도 척도를 사용하여 예측 오차를 추정하였다. 오차는 단계가 커짐에 따라서 증가하는 현상을 보이고 있으나 증가하는 정도는 완만하여 수렴하였다. 이러한 분석을 통해 볼 때 전력 에너지 시계열 데이터의 예측을 위해 사용한 Elman 순환 신경망 적용은 비교적 우수한 결과를 도출하는 모형임을 알 수 있었다.

본 연구의 결과는 향후 전력 에너지 시계열 데이터 예측 모형과 전력 소모량 상황을 비교 검토하고, 날씨 등 다양한 정보를 포함한 예측 모형의 신뢰성 검증을 통해 보다 효율적인 에너지 관리를 위한 전력 에너지의 부하 조절과 전력과 가스의 혼용을 위한 의사결정 시스템에 적용할 수 있을 것으로 판단된다.

Acknowledgement

This work was supported by the Power Generation and Electricity Delivery Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry and Energy, Republic of Korea (No. 2016-2020108170).

References

- [1] Ahn, B., Choi, H., and Lee, H., Regional Long-term/Mid-term Load Forecasting using SARIMA in South Korea, *Journal of the Korea Academia-Industrial cooperation Society*, 2015, Vol. 16, No. 12, pp. 8576-8584.
- [2] Bergmeir, C., Hyndman, R., and Koo, B., A note on the validity of cross-validation for evaluating autoregressive time series prediction, *Computational Statistics and Data Analysis*, 2018, Vol. 120, No. C, pp. 70-83.
- [3] Black, F. and Scholes, M., The Pricing of Options and Corporate Liabilities, *J. Polit. Econ.*, 1973, Vol. 81, No. 3, pp. 637-654.
- [4] Bontempi, G., Ben Taieb, S., and Le Borgne, Y., Machine Learning Strategies for Time Series Forecasting. In : Aaufaure MA., Zimányi E. (eds) Business Intelligence, eBISS 2012. Lecture Notes in Business Information Processing, Vol, 138, Springer, Berlin, Heidelberg.
- [5] Box, G., Jenkins, G., and Reinsel, G., Time Series Analysis : Forecasting and Control, 4th Edition, Wiley, Hoboken, New Jersey, 2008.
- [6] Cox, J., Ross, S., and Rubinstein, M., Option pricing : A simplified approach, *J. Financial Econ.*, 1979, Vol. 7, pp. 229-263.
- [7] Elman, J., Finding Structure in Time, *Cognitive Science*, 1990, Vol. 14, pp. 179-211.
- [8] Hochreiter, S. and Schmidhuber, J., Long short-term memory, *Neural Computation*, 1997, Vol. 9, No. 8, pp. 1735-1780.
- [9] <https://cran.r-project.org/web/packages/RSNNS/RSNNS.pdf>.
- [10] Jung, H. and Song, K., Daily Maximum Electric Load Forecasting for the Next 4 Weeks for Power System Maintenance and Operation, *The Transactions of The Korean Institute of Electrical Engineers*, 2014, Vol. 63, No. 11, pp. 1497-1502.
- [11] Kim et al., Weekly Maximum Electric Load Forecasting for 104 Weeks by Seasonal ARIMA Model, *Journal of the Korean Institute of Illuminating and Electrical Installation Engineer*, 2014, Vol. 28, No. 1, pp. 50-56.
- [12] Lee, C., Song, G., and Kim, J., Correlation Analyses of the Temperature Time Series Data from the Heat Box for Energy Modeling in the Automobile Drying Process, *Journal of Society of Korea Industrial and Systems Engineering*, 2014, Vol. 37, No. 2, pp. 27-34.
- [13] Lee, H. and Shin, H., Electricity Demand Forecasting based on Support Vector Regression, *IE Interfaces*, 2011, Vol. 24, No. 4, pp. 351-361.
- [14] Nam, B., Song, K., Kim, K., and Cha, J., The Spatial Electric Load Forecasting Algorithm using the Multiple Regression Analysis Method, *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, 2008, Vol. 22, No. 2, pp. 63-70.
- [15] Pagan, A., The econometrics of financial markets, *J. Empirical Finance*, 1996, Vol. 3, No. 1, pp. 15-102.
- [16] Park, Y. and Wang, B., Neuro-Fuzzy Model based Electrical Load Forecasting System : Hourly, Daily, and Weekly Forecasting, *Journal of Korean Institute of Intelligent Systems*, 2004, Vol. 14, No. 5, pp. 533-538.
- [17] Peng, C., Havlin, S., Stanley, H., and Goldberger A. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, *Chaos*, 1994, Vol. 49, No. 1, pp. 82-87.

- [18] Schmidhuber, J., Deep Learning in Neural Networks : An Overview, *Neural Networks*, 2015, Vol. 61, pp. 85-117.
- [19] Shim et al., A Model on Price Forecasting of Natural Resources with Restricted Market, *Journal of Society of Korea Industrial and Systems Engineering*, 2014, Vol. 37, No. 4, pp. 82-89.
- [20] Sohn, K., Kim, S., and Shon, E., Fuzzy Time Series Models with Triangular Fuzzy Numbers as Parameters, *Journal of Korean Data Analysis Society*, 2001, Vol. 3, No. 2, pp. 149-162.
- [21] Various methods for detecting long-range dependence can be found, for example, in J. Beran, *Statistics for Long-Memory Processes* (Chapman & Hall/CRC, 1994) Ch. 4.4.

ORCID

Chang-Yong Lee | <http://orcid.org/0000-0003-1778-6532>

Jinho Kim | <http://orcid.org/0000-0003-4243-7318>