

V2G 환경의 전력 수급 의사결정 지원을 위한 SARIMA기법과 LSTM기법의 전력사용량 1일 예측 연구[†]

이정현¹ · 김재성² · 안영호³ · 조완섭⁴

¹²충북대학교 빅데이터협동과정 · ³레티그리드 · ³충북대학교 경영정보학과 ·

접수 2019년 6월 24일, 수정 2019년 7월 10일, 게재확정 2019년 7월 11일

요 약

최근 제기되는 원전이슈와 함께 현재의 국내 전력수급은 위기에 맞닿아 있고 향후에도 국내 전기 차 충전소 보급 정책과 한반도 정치적 이슈에 따라 추가적인 전력 수요가 급증할 것이라 예상된다. 본 논문은 상시전력과 대체전력의 전환, 에너지 거래 등에 있어서 Peak 전력사용량과 그 시간대를 예측하는 것에 연구 목적을 둔다. 따라서 24시간 기반의 단기 예측을 분석 목표로 설정하였고, 시계열 분석기법인 SARIMA (seasonal ARIMA) 모형과 인공신경망 기법의 하나인 LSTM (long short term memory)을 통해 분석 및 예측을 진행하였다. 향후에도 24시간 기반의 전력수요 예측모형이 V2G (Vehicle to Grid) 또는 ESS (energy storage system)와 관련된 현업에 활용 및 확장될 수 있도록 기대한다.

주요용어: 인공신경망, 전력 Peak 부하, ESS, LSTM, RNN, SARIMA, V2G.

1. 머리말

최근 에너지 절감과 효율을 높이기 위한 다양한 연구가 진행되었고, 이러한 연구에서는 다양한 인자와 요인의 식별, 연관관계 그리고 수요 예측이 주를 이루고 있다. 전력 수요 예측에 관한 연구는 다양한 데이터마이닝 분석 기법과 증장기/단기/초단기 등 다양한 예측기간으로 결과를 도출하고 있다. IoT장비들이 폭발적으로 증가하는 시점에서 다양한 데이터들을 대상으로 예측분석을 하기 위해서는 예측의 과정에 실시간적인 데이터가 지속적으로 반영되어야 정확도를 높일 수 있다. 또한 특정 정보서비스 내에서 예측분석이 모듈 형태로 수행될 때, 예측 알고리즘의 실행이 매시간 또는 적어도 일 단위 주기에 미치지 못한다면 예측 결과를 에너지 운영관리 의사결정에 적극적으로 활용하기에는 부족함이 있다.

본 논문에서 다루고자 하는 내용의 범위는 에너지 저장장치 (energy storage system; ESS)를 운영하는 대형건물이나 산업용 건물에서 ESS의 가용량과 가용시간을 더욱 확대하기 위한 고려사항을 포함하고 있다. ESS는 초기 구축비용이 크기 때문에 해당 건물의 전력사용 패턴 중에서 Peak 사용량과 Peak 지속시간에 기반하여 ESS 용량을 산정한다. 이렇게 구축된 ESS를 효과적으로 활용 및 운영하기 위한

[†] 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터지원사업의 연구결과로 수행되었음(IITP-2018-0-01396).

¹ (28644) 충북 청주시 서원구 충대로 1, 충북대학교 빅데이터협동과정, 박사과정.

² (28644) 충북 청주시 서원구 충대로 1, 충북대학교 빅데이터협동과정, 박사.

³ (05621) 서울특별시 송파구 백제고분로41길 8 송파빌딩, 레티그리드 대표.

⁴ 교신저자: (28644) 충북 청주시 서원구 충대로 1, 충북대학교 경영정보학과 교수.

E-mail: wscho@chungbuk.ac.kr

방법 중 전기자동차 등에 저장된 전력 에너지를 건물의 ESS (grid)로 충전시켜주는 V2G (vehicle to grid) 기술의 접목은 기존 ESS의 운용범위를 확대시킬 수 있다.

본 논문에서는 수도권에 위치한 산업용 건물A에서 1시간 단위의 전력사용 데이터를 수집, SARIMA (seasonal ARIMA)기법과 LSTM (long short term memory)기법을 사용하여 향후 24시간 동안의 Peak 전력량과 그 시간대를 예측하였다. 예측 성능의 확인은 예측 결과로 도출된 24시간 데이터와 기존 실제 데이터를 비교하여 도출하였다. 1시간 주기의 전력사용량 데이터를 통해 24시간을 예측기간으로 결과를 도출하는 것은 예측의 신뢰성을 높이는데 목적을 두고 기존 연구와 차별을 두는 점이다.

본 논문의 구성은 2장에서 ESS (energy storage system)와 V2G (vehicle to grid)에 대한 기술, 시계열 분석기법 중의 하나인 SARIMA와 인공신경망 및 LSTM 기법을 소개하고 전력데이터를 활용한 몇몇의 연구 사례를 소개한다. 3장에서는 실험 데이터의 소개와 실험환경을 설명하면서 시계열모형의 수립 및 예측결과와 LSTM 기법의 파라미터 설정을 통한 실험결과를 도출하였고, 4장에서 결론과 향후 기대효과를 제시했다.

2. 선행 연구

본 논문의 선행연구에서는 국내 ESS 투자확대에 따른 경제성 분석과 효율을 높일 수 있는 알고리즘의 개발사례를 들었다. 또한 V2G기술을 소개하면서 이 기술이 ESS와 접목될 때의 시너지 효과와 효율을 높일 수 있는 기회를 몇몇의 연구에서 확인했다. 또한 본 논문에서의 예측모형 수립과 예측 결과 도출에 사용한 SARIMA기법과 RNN 그리고 LSTM 기법을 간략하게 소개하였다. 이어서 전력데이터를 활용한 예측분석의 연구 사례에서는 전력 수요 패턴에 대해 시간적 요소를 위주로 구분하여 다양한 예측 결과를 제시하였다.

2.1. ESS (energy storage system)

ESS는 전력에너지를 미리 저장했다가 필요한 시간대에 사용할 수 있는 에너지 저장 시스템이다. 국내에서는 2011년 지식경제부에서 발표한 ‘에너지 저장 기술 개발 및 산업화 전략 K-ESS 2020’에서 6.4조원 규모의 연구 개발 및 설비 투자를 추진한 바 있다.

한편 공공기관을 비롯한 여러 시설에 ESS의 구축 및 활용 관련된 최근 연구가 있었다. Kim 등 (2014)의 연구에서는 고속열차 운행에 필요한 Peak 전력사용량 분석 및 예측을 통해 고속철 변전소에 설치되는 ESS 최적 용량을 산정하고 그 경제성을 분석하였다. 고속열차의 전력 사용량 15분 데이터를 분석하여 정격 15MW의 대용량 부하 환경에서 Peak 전력에 대응하기 위한 ESS의 용량 (6MW의 정격 출력과 3.5MWh의 전지용량, 2011~2012년 기준, 특정 변전소 기준)을 산정하였고, 구축비용 (약53억원)에 대한 회수기간 (8.7년)을 평가하였다. Byeon 등 (2016)의 연구에서 ESS의 효율적인 가동을 위해서 과거 데이터를 기반으로 Peak 전력을 저감시킬 수 있는 최적운전 기법이 적용된 프로그램을 개발하였다.

2.2. V2G (vehicle To grid)

V2G는 그리드 환경과 자동차 간의 에너지 거래를 위한 기술과 개념이 포함되어 있다. 이는 전기자동차에 기반한 서비스 중 하나로서 전기요금에 싸고 공급량이 충분한 시간대에 전기자동차를 충전하고 피크시간대에 다시 전력망으로 되팔게 함으로써 전체적인 발전요구량을 낮춘다 (Ansari 등, 2015). 이러한 배경에서 Kim과 Hwang (2016)의 연구에서는 피크시간대에 전기자동차 배터리에 충전했던 전력을 다시 계통에 공급하는 V2G 기술은 전력 최대수요를 낮추어 높은 단가의 발전소 건설비용과 연료비용을

회피할 수 있는 효과를 고려하였다. 이러한 효과를 통해 V2G는 전기차 (EV), 플러그인 하이브리드 자동차 (PHEV) 등을 충전했다가 주행 후 남은 전기를 다시 전력망으로 송전하게 해주는 기술이며, 전기가 움직이는 에너지저장장치 (ESS)의 역할을 할 것이라고 전망하기도 했다 (Kim 등, 2017).

이와 같이 V2G 기술은 풍력/태양열 발전소처럼 생산능력이 일정치 않거나, 전력 수요가 큰 변전소 등에 우선 적용될 수 있다. Hwang 등 (2014)의 연구에서는 기상상황에 따라 발전량이 유동적인 풍력/태양열 발전소 체계에 V2G를 적용하여 전력의 안정성을 높이고자 했다.

현재 관련 연구기술의 방향으로는 전기자동차가 전력계통에 직접 송전하는 것은 안전관리 측면의 이슈가 존재하기 때문에, ESS에 여러 대의 전기자동차가 연결된 형태의 V2G가 고려될 수 있다.

2.3. SARIMA (seasonal ARIMA)

Lee (2017)의 연구에 따라, ARIMA 모형은 과거의 관찰된 값들로 설명되는 자기회귀 모형 (autoregressive model)과 시계열을 오차항으로 설명하는 모형으로, 이동평균 모형 (moving average model)을 혼합하고 차분 과정을 거친 모형이다. 이러한 ARIMA 모형에 주기적인 특성이나 계절성을 반영한 모형이 SARIMA (seasonal autoregressive integrated moving average)모형이다. SARIMA 모형은 ARIMA 모형처럼 예측하고자 하는 시점으로부터 가까운 과거 자료들을 이용할 뿐만 아니라, 자료의 주기적인 특성, 계절성을 반영했기 때문에 이전 주기의 자료를 추가적으로 활용한다. $SARIMA(p, d, q) \times (P, D, Q)_s$ 모형의 일반식은 식(2.1)과 같다.

$$\begin{aligned}
 \phi_p(B)\Phi_P(B)W_t &= \theta_q(B)\Theta_Q(B)e_t, \\
 \phi_p(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\
 \Phi_P(B) &= 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}, \\
 \theta_q(B) &= 1 - \theta_1 B - \dots - \theta_q B^q, \\
 \Theta_Q(B) &= 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}, \\
 W_t &= \nabla^d \nabla_s^D Z_t,
 \end{aligned} \tag{2.1}$$

여기서 Z_t 는 원시계열 자료, W_t 는 차분된 시계열 자료, t 는 시간을 나타내는 연산자, e_t 는 $N(0, \sigma^2)$ 을 따르는 오차항 (백색잡음), B 는 후향연산자, p 는 AR항의 차수, q 는 MA항의 차수, d 는 차분의 차수, P 는 계절 AR항의 차수, Q 는 계절 MA항의 차수, D 는 계절차분의 차수, s 는 계절 시차를 나타낸다.

본 연구에서 사용된 데이터는 시계열 특성을 가지고 있고, 계절성 (seasonality)이나 주기적 특성 (periodicity)을 확인하기 위해서 시계열 분석 (time series analysis) 기법중 하나인 SARIMA기법을 분석 모형의 일부로 활용하였다.

2.4. RNN (recurrent neural networks: 순환신경망)

RNN은 인공신경망의 한 종류로 시계열 개념이 추가되어 있어서 데이터의 시간적 순서 관계가 중요한 시계열 데이터에 적합한 모형이다. 은닉계층에 이전 정보가 지속되도록 순환구조를 가지고 학습을 진행한다. RNN의 처리 과정은 먼저 입력이 시간에 따라 입력층 (input layer)을 통해 은닉층 (hidden layer)으로 입력되면, 이전 은닉층의 출력값이 현재의 은닉층의 입력으로 들어가게 되고, 이 두 값을 토대로 출력값을 도출한다. 이러한 반복적인 과정으로 과거의 데이터가 다음의 데이터에 어떤 영향을 주는지 학습할 수 있고, 이런 장점으로 시계열 예측에 사용되는 추세다.

Figure 2.1은 RNN의 구조를 나타내는데, x_t 는 시간 스텝에서의 입력값으로 메모리 부분의 h_t 를 거쳐 다음 스텝에 반영되어 계산되는 은닉층에 해당된다. o_t 는 시간 스텝 t 에서의 출력값으로 매 시간 스텝마다 값이 나오는 출력층이다.

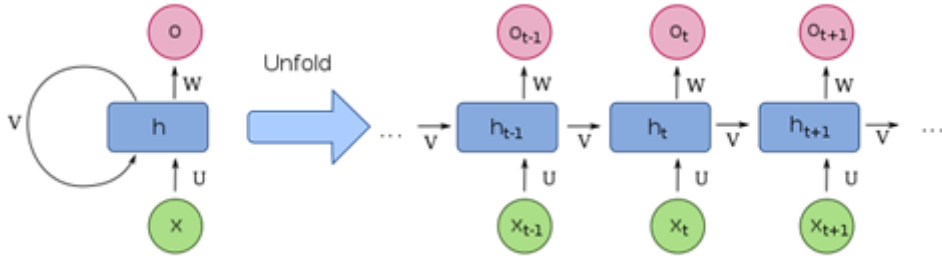


Figure 2.1 Structure of RNN (Wikipedia, 2018)

그러나 RNN의 학습과정에서 기울기 소실 문제 (vanishing gradient)가 발생하여 데이터의 장기 의존성 (long-term dependency)을 고려하는데 한계가 있다. 이러한 문제를 고려한 RNN의 변형모형이 LSTM이다.

2.5. LSTM (long short term memory)

LSTM은 Hochreiter와 Schmidhuber (1997)에 의해 제안된 방법으로 RNN의 기울기 손실문제를 해결할 수 있다. 셀 스테이트 (cell state)를 통해 과거 학습결과를 전달하는 구조로 장기 의존성 문제를 해결하는 부분은 LSTM 기법의 핵심이라 할 수 있다. 셀 스테이트는 3개 요소인 망각게이트 (forget gate), 입력게이트 (input gate), 출력게이트 (output gate)로 구성된다. 이 요소들을 활용하여 정보를 더하거나 제거하는 기능을 수행하고, 선택적으로 정보를 반영할 수 있다.

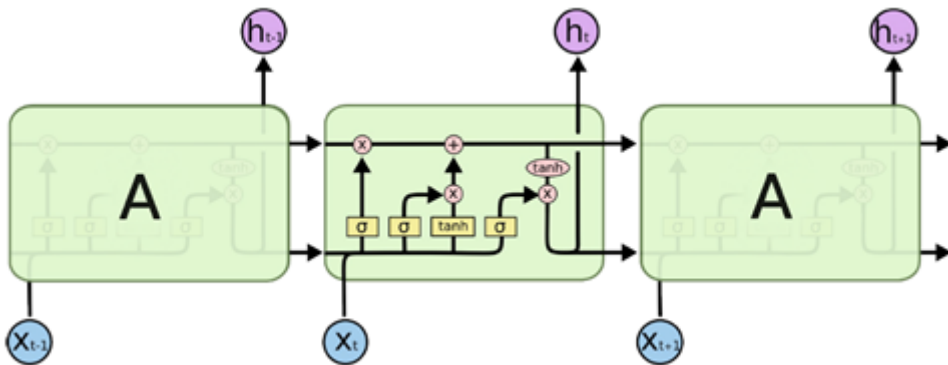


Figure 2.2 Structure of LSTM(Christopher, 2015)

Figure 2.2에서와 같이, 먼저 망각 게이트는 현재의 입력과 이전 입력의 은닉층 출력이 σ 로 표현된 레이어 (sigmoid; 활성화함수)를 통과하게 되는데 입력 정보를 얼마나 삭제시킬 것인지 선택하게 된다. 두 번째는 입력 게이트이다. 망각 게이트를 통과한 값에 어떠한 값을 저장해 줄 것인지 선택하는 과정으로 또 다른 sigmoid 함수와 tanh 함수를 이용해 새로운 입력을 얼마나 셀 스테이트에 반영해 줄 것인지를 결정한다. 마지막으로 출력게이트에서는 마지막 sigmoid 함수와 셀 스테이트에서 나온 출력을 통과시킨 마지막 tanh 함수로 새로운 출력 후보로 만든다. 이 두 값의 곱이 다음 은닉층과 현재 은닉층의 출력으로 빠져나가게 된다 (Sibi 등, 2013; Salehinejad 등, 2018; Krizhevsky 등, 2012).

본 연구에서 사용된 데이터는 시계열 특성을 가지고 있고, 전력사용량 1일 예측을 위한 최적 모형 수립을 위해서 RNN 기법 중 LSTM 기법을 활용하였다. LSTM 알고리즘은 몇가지 파라미터를 설정하여 모형을 수립할 수 있고, 각 파라미터를 다양한 값으로 조정하면서 최적 모형을 수립하였다.

2.6. 예측 분석 사례

다양한 비즈니스와 연구에서는 IoT 센서 및 인프라와 데이터 통신방식이 접목된 형태로 네트워크 환경이 확대, 다양한 전력데이터가 여러 기법을 통해 분석에 활용되고 있다. Choi와 Lee (2015)의 연구에 따르면 에너지 기술과 ICT기술의 융합은 기존에 정채되어 있던 에너지 분야에서의 기술 혁신을 이루어 내기 위한 원동력이 된다. 에너지 분야에 있어서 정보통신기술의 사용은 서로 다른 건물 및 장비간의 에너지 거래 및 공유를 가능하게 하며, 에너지 생산 및 소비 단계에서의 공급 및 수요예측을 통한 효과적인 에너지 사용을 가능하게 한다. 또한 통신기술을 통해 수집한 정보를 분석하여 기존에 예측하지 못했던 다양한 부분에서의 에너지 절감이 가능하다.

Park과 Yoon (2017)의 연구에서는 전력 수요 패턴인 요일, 휴일, 날씨, 사회적 이벤트 등에 대해 선행연구를 진행하면서 국내 전력거래소에서 수집된 5년간의 시간단위 데이터를 활용하여 4년간 데이터를 training data로 사용하고 나머지 1년간 데이터를 validate data로 사용하여 분석 결과를 비교 및 평가하였다. 데이터 셋에서의 각 시간대를 독립변수로 간주, 24개의 독립변수를 가지고 군집분석을 수행하였다. 군집분석에 사용된 기법은 의사결정나무와 랜덤포레스트, 나이브베이즈, SVM이고, 각 분석 기법에 대해 예측 정확도를 산출했다.

Moon 등 (2017)의 연구에서는 대학 캠퍼스 중 일부 건물을 대상으로 총 4년간의 일간 전력 사용량 데이터를 통해서 의사결정나무를 이용하여 데이터를 분류하였고, 테스트 셋에서 유사 시계열 형태의 데이터 셋을 가지고 ARIMA 기반의 예측모형을 구성하였다. 학사 기간의 특성에 따라 몇 가지 전력 사용 패턴이 확인되었고, 해당 기간이나 주말에 따라 9개의 ARIMA 모형을 제시하였다.

Ryu 등 (2017)의 연구에서는 미국 연구소 내 지원시설을 대상으로 분석을 수행하였으며 총 91일간의 일간 전력 데이터 중 86일 간의 데이터를 통해 시계열 모형을 수립하고 선정된 모형의 적절성을 검토한 뒤, 나머지 5일간의 실제 데이터와 예측된 5일간의 데이터를 비교하여 성능을 평가하였다.

Lee 등 (2018)의 연구는 몇 가지 가전제품을 사용한 테스트 공간에서 수집한 전력 사용량 데이터를 가지고 LSTM 기법을 활용한 전력사용량 예측을 하였다. LSTM의 파라미터를 최적화하기 위해 LSTM layer 개수와 셀의 개수, look_back, Epoch, batch_size 파라미터들을 다양한 값으로 설정하면서 예측모형을 수립, 예측 결과를 도출하였다. 또한 Shin과 Kim (2018)의 연구에서는 태양광 발전량 예측을 위해서 LSTM 기법을 활용하였다. 초단기 예측 중에서도 다음날 24시간의 발전량을 예측하기 위해 하루 단위 (24개의 타임스텝)로 학습을 진행하였고, 모형 수립과 예측결과를 도출하였다.

Lee 등 (2019)의 연구는 본 논문에서 사용된 동일한 데이터를 활용하여 시계열 분석 기법인 계절지수 평활법과 SARIMA 기법을 통해서 예측모형을 수립하고 전력사용량의 1일 예측을 결과로 도출하였다.

3. 모형 수립

Lee 등 (2019)의 연구에서는 본 논문에서 사용된 동일한 데이터를 활용하여 시계열 분석모형인 가법적 계절지수평활법 (additive seasonal exponential smoothing method)과 SARIMA (seasonal ARIMA)모형을 가지고 예측분석을 진행하였다. 여기에서 결과로 도출된 SARIMA 모형의 예측결과와 본 논문에서의 LSTM 모형의 예측결과를 비교하여 성능을 도출하고자 한다.

3.1. 데이터 set

본 연구에서 사용된 데이터는 도시 외곽에 위치한 A사 건물에서 수집된 전력사용량 데이터이다. 해당 건물의 수배전반 전력설비에 설치된 다수의 IoT센서에서는 전력량 및 전력 품질에 관한 다양한 데이터가 수집된다. 이 데이터들은 기존 전력량계가 제공하는 전력량 이외에 다양한 전력 품질 데이터가 포함되어 있고, 매초 측정되고 저장된다.

본 논문에서는 다양한 전력 품질 데이터 중 전력사용량 데이터를 학습과 검증에 사용하였고, 데이터 수집주기는 1시간 간격으로 1시간 동안의 전력사용 데이터 중 MAX 값 (최대 전력사용량)으로 측정된 데이터를 사용하였다. 데이터의 수집기간은 2015년 5월 2일 ~ 2015년 7월 27일까지의 총 87일간의 데이터이고 수집된 데이터의 전체 사용량 그래프는 Figure 3.1과 같다. 학습은 87일중 86일간의 시간별 데이터를 가지고 수행하고, 나머지 87일째 되는 1일간의 24시간에 대해 예측을 하되, 그 예측결과를 실제 데이터와 비교하여 성능을 측정하였다.

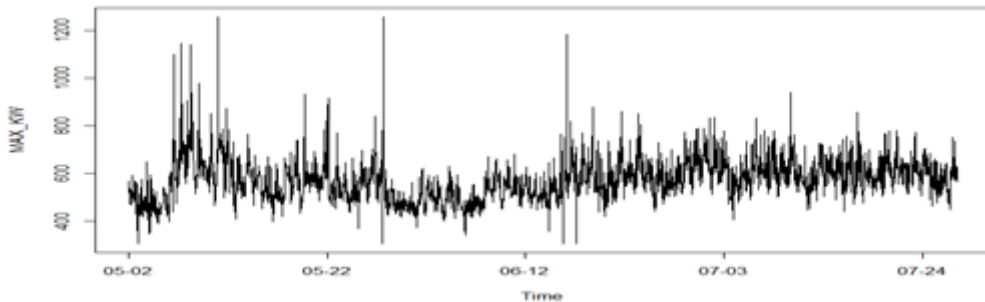


Figure 3.1 Observation data

3.2. SARIMA 모형의 식별

Lee (2017)의 연구에 따라 Box-Jenkins 모형 중 seasonal ARIMA모형은 시계열자료의 예측분석에 널리 사용되어 왔고, 이 모형을 통해 시계열 자료를 다루고자 할 때는 모형의 식별과 추정 그리고 검정 단계를 거친다.

먼저 시계열모형에서의 식별 (model identification) 과정에서는 자료의 산점도를 통해서 시계열의 경향이나 계절성, 이상치, 일정하지 않은 분산, 비정규성 등 비정상성 (non-stationary)을 갖고 있는지 판단한다.

Figure 3.2에서는 SARIMA 모형 수립에 사용된 데이터에 대해 분해법을 적용하였다. 여기에서는 추세성과 계절성, 랜덤 경향을 파악할 수 있다. 추세는 증가한 후 감소하고 다시 증가한 후 유지되는 경향

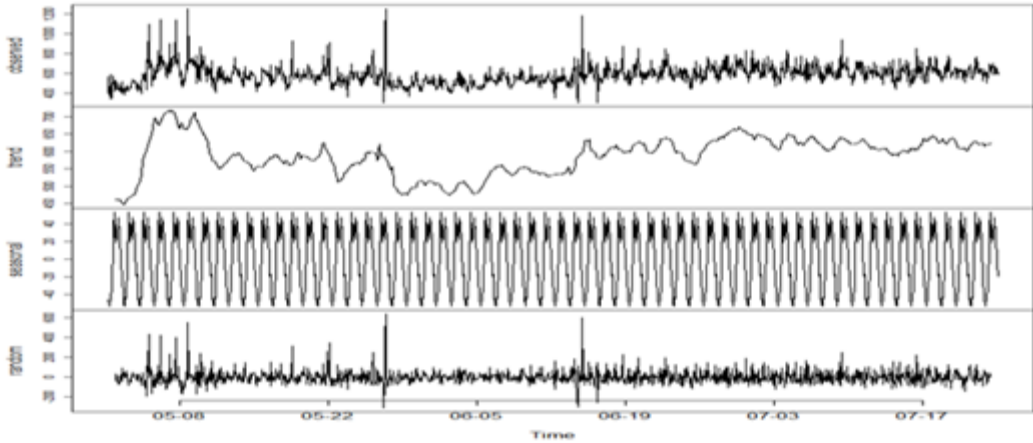


Figure 3.2 Decomposition of time series data

이 있었고, 시간대 별로 전력소모량이 크고 적음을 나타내는 계절성의 확인도 필요함을 판단하였다. 이러한 특성들을 통해서 본 연구에서는 ARIMA 모형에 계절성을 반영한 SARIMA 모형을 사용하였다.

데이터의 자기상관의 존재 여부를 확인하기 위해서 Figure 3.3에서와 같이 ACF와 PACF를 도출하였고, 이는 비정상 시계열 자료이기 때문에 정상화 과정이 필요하다. 정상성을 만족시키기 위해 1차 차분, 계절 차분 (24차 차분), 로그 변환을 해주었는데, 1차 차분의 필요성이 명확치 않아서 R에서 지원하는 `ndiffs()` 함수를 사용하여 1차 차분이 필요함을 확인하였다. 이 자료는 24시간 주기로 계절성을 가지기 때문에 계절차분을 적용, 분산의 안정화를 위해 로그 변환을 실시하였다.

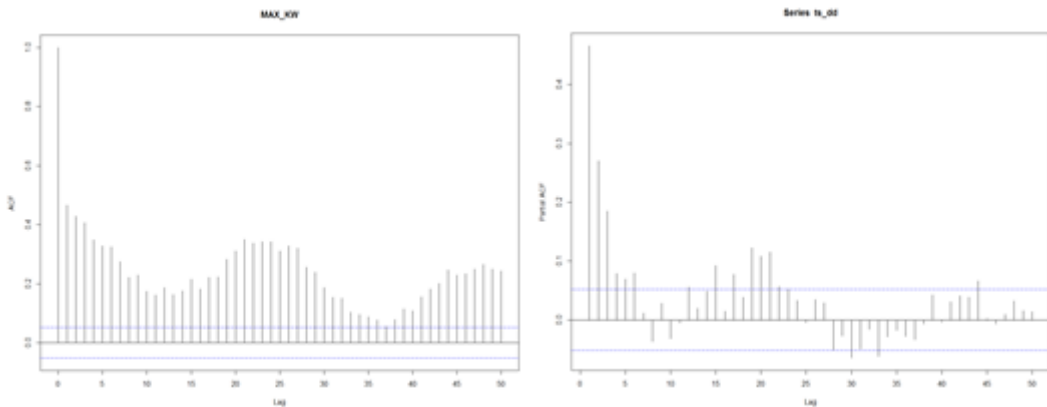


Figure 3.3 ACF and PACF on observation data

로그변환과 1차 차분, 계절차분을 처리한 결과, Figure 3.4와 같이 0을 중심으로 데이터가 산포되어 있고, 분산이 비교적 안정화 되었다. 이를 통해 정상성 만족을 알 수 있다. 또한 여기에서 변동성이 변하는 경향도 있는데, 향후 연구에서는 ARCH 또는 GARCH 모형을 적용한 분석도 고려해 볼 수 있다.

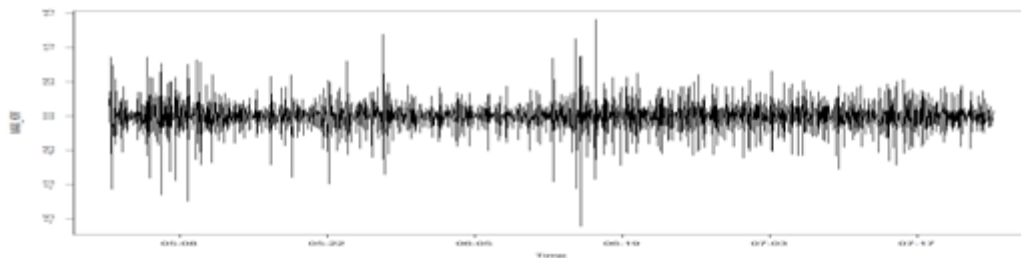


Figure 3.4 Apply log, diff and seasonal diff on observation data

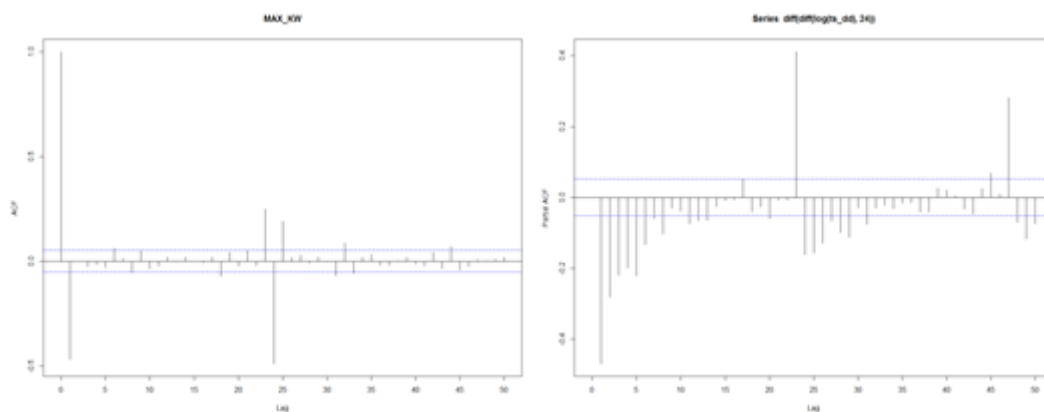


Figure 3.5 ACF and PACF after applying log, diff and seasonal diff

로그 변환, 1차 차분, 계절차분을 수행한 자료에 대해 ACF와 PACF를 확인해보면 Figure 3.5에서와 같이 ACF에서 시차 1, 24에서 최하를 나타내고 절단되는 형태를 보이고, PACF에서 시차 24, 48에서 점차 감소하는 형태를 보이고 있다. 이를 통해서 계절성 MA 차수를 1 또는 2로 식별할 수 있다. 비계절성 파트를 식별하기 위해서 ACF를 보면 1, 24 근처에서 하나 최하를 찍고 절단되는 형태를 보이고 PACF는 1, 24, 48 근처에서 지수적으로 감소하는 형태를 보이므로 비계절성 MA 차수로 1 또는 2를 식별하였다. 따라서 $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ 을 잠정적인 모형으로 선정할 수 있음을 고려하였다.

3.3. SARIMA 모형의 추정

$SARIMA(p, d, q) \times (P, D, Q)_s$ 모형을 추정하는데 있어서 p, d, q, P, D, Q 의 값을 찾는 것이 이 부분에서의 핵심이다. 우선 $d = 1, D = 1$ 이라는 것을 정상성 검정을 통해 알 수 있다. 다음으로 p, q, P, Q 의 값을 추정하여야 하는데 p, q, P, Q 를 추정하기 위해서 AIC (Akaike's information criterion) 통계량과 BIC (Bayesian information criterion) 통계량 등을 이용한다. AIC와 BIC는 모형의 적합도와 간명성을 동시에 고려하는 통계량이다. 모형의 AIC, BIC 값이 작을수록 더 적합한 모형이고 식 (3.1)으로부터 도출된다.

$$AIC = -2\ln(\hat{L}) + 2k, \quad BIC = -2\ln(\hat{L}) + \ln(n)k, \quad (3.1)$$

여기서 \hat{L} 은 maximum likelihood, k 는 모수의 개수, n 은 관측수이다.

모형이 아무리 높은 예측력을 가지고 있다 하더라도 모형이 간결하지 않으면 실용성이 떨어진다. 모형을 추정함에 있어서 모수 축약의 원칙에 따라 각 계수들을 2까지만 설정 했다.

Table 3.1 AIC, BIC, RMSE on SARIMA model

	AIC	BIC	RMSE
$SARIMA(0, 1, 1) \times (0, 1, 2)_{24}$	-1890.03	-1874.268	0.1197656
$SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$	-1890.67	-1869.648	0.1197733
$SARIMA(0, 1, 2) \times (0, 1, 1)_{24}$	-1888.56	-1867.538	0.1197525
$SARIMA(0, 1, 2) \times (0, 1, 2)_{24}$	-1889.03	-1862.757	0.1197604
$SARIMA(0, 1, 1) \times (0, 0, 1)_{24}$	-1799.5	-1783.683	0.1291253
$SARIMA(0, 1, 1) \times (0, 0, 2)_{24}$	-1811.86	-1790.777	0.1284579
$SARIMA(0, 1, 2) \times (0, 0, 1)_{24}$	-1803.4	-1782.315	0.1288492
$SARIMA(0, 1, 2) \times (0, 0, 2)_{24}$	-1817.95	-1791.587	0.1280833

최적의 모형을 찾기 위해 AIC와 BIC 그리고 RMSE를 Table 3.1와 같이 도출하였다. 또한 계절차분을 적용한 모형과 적용하지 않은 모형의 AIC, BIC, RMSE를 확인했을 때, AIC는 $SARIMA(0, 1, 1) \times (0, 1, 2)_{24}$ 모형이 가장 낮은 값을 보였지만 다른 모형들과의 AIC 차이가 크지 않다. 반면에 BIC를 확인하면 $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ 모형이 -1874.268로 가장 작은 값을 보였다. 다른 모형들과의 BIC 값 차이가 AIC보다 다소 큰 경향을 보였다. 이러한 근거로 본 연구에서는 BIC를 기준으로 최적의 모형을 선정하였다. 최적의 모형으로 추정된 $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ 모형의 계수와 식은 Table 3.2와 식 (3.2)와 같다.

Table 3.2 Coefficients on $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ model

	Estimate	std. Error	z value	Pr(> z)
ma1	-0.873326	0.013930	-62.674	< 2.2e-16***
sma1	-0.971069	0.015197	-63.900	< 2.2e-16***

$$\nabla^1 \nabla_{24}^1 Z_t = (1 - \theta_1 B)(1 - \Theta_1 B^{24})e_t, \quad e_t \sim WN(0, \sigma^2). \quad (3.2)$$

3.4. SARIMA 모형의 검정

추정된 모형이 통계적으로 유의한지 판단하기 위해서 모형 검정이 필요하고, 이는 잔차검정을 통해서 확인한다. 잔차의 자기상관이 존재한다면 추정의 단계로 돌아가 모형을 다시 추정하게 되고, 자기상관이 나타나지 않는다면 해당 모형이 적합하다고 할 수 있다. Figure 3.6에서 확인되는 바와 같이 잔차는 0을 기준으로 분포하고 있고 특정한 패턴은 보이지 않으므로 i.i.d. 가정을 충족하는 것으로 보인다. 잔차의 ACF도 1 이상의 시차에서 모두 유의하게 작은 값을 가지므로 독립성을 만족한다고 할 수 있다. 또한 백색잡음 검정을 위한 포트맨토 검정의 유의확률이 모든 시차에서 0.05보다 크므로 귀무가설 $H_0 : \rho(k) = 0, k \geq 1$ 을 기각할 수 없기 때문에 백색잡음이 자기상관성이 없다고 할 수 있다. 따라서 $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ 모형은 통계적으로 유의하다고 판단되어 SARIMA의 최적 모형으로 선정하였다.

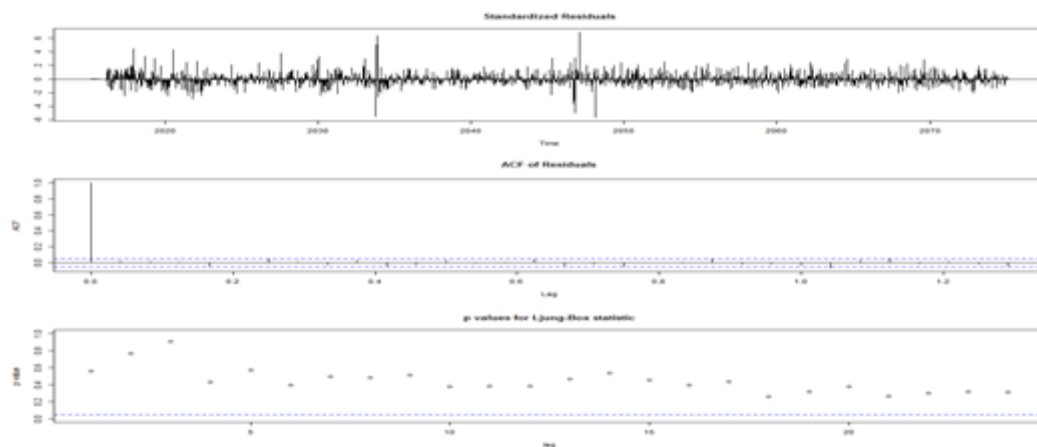


Figure 3.6 Check of Residuals on $SARIMA(1, 1, 2) \times (0, 1, 1)_{24}$ model

3.5. LSTM 모형수립

LSTM 분석기법을 활용하기 위한 실험 환경은 리눅스 기반에서 Python 3.6.7, Keras 2.2.4 그리고 Tensorflow 1.12.0을 사용하였다. 또한 사용된 LSTM 알고리즘에서의 네트워크 구성은 Figure 3.7과 같다.

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 100)	40800
dense_1 (Dense)	(None, 1)	101
Total params: 40,901		
Trainable params: 40,901		
Non-trainable params: 0		

Figure 3.7 Configuration of LSTM networks

LSTM은 100개의 노드로 구성되어 있고, 이를 처리하는 내부 파라미터가 40,800개 이다. 최종 출력 계층의 노드는 1개이며, LSTM과 출력계층 간에 적용되는 가중치 수는 101개 이다. 이 중 100개는 입력값에 대한 가중치이고, 1개는 평균값을 조절하는 가중치이다.

실험에 사용한 파라미터는 Batch_size, Windows_size, Epochs를 활용하였다. Batch_size는 몇 개의 데이터 셋을 학습하여 연결 계수들의 값을 수정시킬 것인지를 결정하는 파라미터이다. Window_size는 과거 몇 시간의 데이터를 참조하여 다음 한 시간의 데이터를 예측할 것인가를 결정하고, Epoch는 주어진 데이터들의 반복 학습 횟수를 결정하는 파라미터이다. 예를 들어 데이터 샘플이 1시간 주기의 데이터인 환경에서 Window_size=10이면 과거 10시간 데이터를 참조하여 그 다음 한 시간을 예측하게 된다.

최적의 LSTM 모형을 선정하기 위한 정확도는 예측 오차율을 통해 확인하였다. 오차율을 구하기 위하여 파라미터를 조정하면서 실제값과 예측값 간의 평균 제곱근 오차 RMSE (root mean square er-

ror)를 각각 도출하였고 도출식은 식 (3.3)과 같다.

$$RMSE = \sqrt{\frac{1}{n} \left(\sum_{t=1}^n (real - predict_t)^2 \right)}. \quad (3.3)$$

3개의 파라미터를 통해 최적의 RMSE 값을 측정하기 위해서 3단계의 파라미터 조정 과정을 거쳤다. 먼저 몇몇 연구에서 사용된 값 (Window_size=16, Epochs=100)을 고정값으로 놓고 Batch_size 파라미터를 1부터 512까지 증가시키면서 RMSE를 측정하였다. 개별적인 실험과정에서 Batch_size는 과적합 (overfitting) 이슈로 인해 수치를 512 초과하여 증가시키지 않았고, Batch_size=1 또는 Window_size=1의 조건에서는 데이터 학습의 유효성 측면의 이슈가 있었으므로, 모델 선정에서 제외하였다.

이를 통해 최소 RMSE를 갖는 3개의 후보를 Table 3.3와 같이 도출하였고, 1차적인 후보 모델은 Batch_size가 8, 32, 64를 갖는 모델이다.

Table 3.3 Selection of candidate model from 1st adjusting LSTM parameters

	Batch_size	Window_size	Epochs	RMSE	Rank
1	1	16	100	62.125	-
2	2	16	100	76.288	
3	4	16	100	67.997	
4	8	16	100	62.292	1
5	16	16	100	64.356	
6	32	16	100	64.295	3
7	64	16	100	64.141	2
8	128	16	100	66.042	
9	256	16	100	65.708	
10	512	16	100	66.422	

여기에서 도출된 후보 모델들에 대해서 이번에는 Epochs를 고정값, Batch_size는 8, 32, 64로 각각 고정, Window_size를 1부터 512까지 증가시키면서 RMSE를 도출하였다. 여기에서 Window_size는 과적합 (overfitting) 이슈로 인해 수치를 512 초과하여 증가시키지 않았다. 2차적인 후보로 선정된 모델들은 Table 3.4에서와 같이 Batch_size와 Window_size가 각각 (8, 64), (32, 128), (64, 128)인 모델들이 낮은 RMSE 값으로 도출되었다.

Table 3.4 Selection of candidate model from 2nd adjusting LSTM parameters

	Batch_size	Window_size	Epochs	RMSE	Rank
1	8	1	100	75.394	-
2	8	2	100	65.001	
3	8	4	100	87.707	
4	8	8	100	66.284	
5	8	16	100	62.292	
6	8	32	100	65.527	
7	8	64	100	59.524	1
8	8	128	100	65.304	
9	8	256	100	62.280	
10	8	512	100	61.753	
11	32	1	100	63.523	-
12	32	2	100	66.988	
13	32	4	100	72.689	
14	32	8	100	67.266	
15	32	16	100	63.093	
16	32	32	100	63.670	
17	32	64	100	62.245	
18	32	128	100	60.860	3
19	32	256	100	65.276	
20	32	512	100	62.834	
21	64	1	100	63.801	-
22	64	2	100	65.585	
23	64	4	100	65.862	
24	64	8	100	68.348	
25	64	16	100	62.524	
26	64	32	100	65.788	
27	64	64	100	68.980	
28	64	128	100	60.707	2
29	64	256	100	65.490	
30	64	512	100	68.538	

2차 후보로 선정된 3개 모델의 Batch_size와 Window_size인 (8, 64), (32, 128), (64, 128)의 각 모델을 고정값으로 두고, Epochs만 10부터 300까지 증가시키면서 가장 낮은 RMSE 값이 도출되는 모델을 확인하였다.

결론적으로 Table 3.5에서와 같이 최종 모델의 수립을 위해 설정된 파라미터를 조정하면서 최적의 오차를 얻었다. 최적 모형은 Batch_size=64, Windows_size=128, Epoch=150 조건에서, RMSE의 최소값이 58.220로 도출되어 최종 LSTM 모형으로 선정했다.

Table 3.5 Selection of candidate model from final adjusting LSTM parameters

	Batch_size	Window_size	Epochs	RMSE	Rank
1	8	64	10	82.222	
2	8	64	30	87.966	
3	8	64	60	83.088	
4	8	64	100	59.524	
5	8	64	150	64.426	
6	8	64	200	64.299	
7	8	64	300	70.087	
8	32	128	10	62.437	
9	32	128	30	70.582	
10	32	128	60	62.411	
11	32	128	100	59.949	
12	32	128	150	58.648	2
13	32	128	200	66.618	
14	32	128	300	83.851	
15	64	128	10	72.298	
16	64	128	30	68.390	
17	64	128	60	76.633	
18	64	128	100	60.707	
19	64	128	150	58.220	1
20	64	128	200	59.198	3
21	64	128	300	65.203	

4. 예측 결과

4.1. SARIMA와 LSTM모형의 예측

추정된 $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ 모형의 24시간 예측결과와 Batch_size=64, Windows_size=128, Epoch=150의 파라미터를 가진 LSTM 모형의 24시간 예측결과 데이터를 Table 4.1에서와 같이 정리하였다. LSTM의 예측은 학습된 데이터를 통해 이후 1시간씩을 예측하여 총 24시간의 예측결과가 도출되었다.

4.2. $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ 모형과 LSTM 모형의 성능 비교

Prediction sum-of-squares (PRESS)는 적합치와 관측치 사이 편차의 측도이다. 이를 사용하여 모형의 예측 능력을 평가할 수 있고, 일반적으로 PRESS 값이 작을수록 모형의 예측 능력이 더 좋다. PRESS는 식 (4.1)과 같이 나타낸다.

$$PRESS = \sum_{i=1}^n (y_i - f_i)^2. \quad (4.1)$$

이 경우 n 은 예측한 데이터의 개수, y_i 는 실제 값, f_i 는 예측 값으로 표현하고, $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ 모형과 LSTM 모형에 대한 PRESS 결과는 Table 4.2과 같다.

Table 4.1 Following 24 points forecasting on SARIMA and LSTM model

	Real point value	ARIMA point forecast	LSTM point forecast
07-27 00:00	502.90	576.26	493.13
07-27 01:00	534.88	572.72	497.69
07-27 02:00	490.16	570.91	521.38
07-27 03:00	448.17	571.08	539.40
07-27 04:00	601.05	572.02	544.37
07-27 05:00	600.53	577.84	571.59
07-27 06:00	610.33	584.51	586.15
07-27 07:00	626.62	601.01	593.24
07-27 08:00	572.33	632.67	595.84
07-27 09:00	593.77	652.54	583.63
07-27 10:00	750.44	675.63	589.61
07-27 11:00	744.31	656.36	628.44
07-27 12:00	591.71	643.50	621.85
07-27 13:00	606.80	649.35	595.63
07-27 14:00	582.02	657.90	616.57
07-27 15:00	566.64	643.01	616.77
07-27 16:00	734.75	668.30	622.81
07-27 17:00	611.60	659.24	649.05
07-27 18:00	632.71	648.25	606.15
07-27 19:00	574.57	648.14	598.56
07-27 20:00	600.20	622.15	586.16
07-27 21:00	622.30	616.21	582.83
07-27 22:00	623.30	596.69	580.01
07-27 23:00	565.25	603.76	579.06

Table 4.2 Result of PRESS between SARIMA and LSTM model

	$SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$	LSTM
PRESS	82684.91	80246.55

궁극적으로 Table 4.2에서 확인되는 바와 같이 두 모형의 최종적인 성능 비교는, LSTM 모형에서 도출된 PRESS 값이 80246.55로 $SARIMA(0, 1, 1) \times (0, 1, 1)_{24}$ 모형의 값보다 더 적기 때문에 LSTM 모형이 더 좋은 예측 성능을 가지고 있다. 실제값과 SARIMA의 예측값 그리고 LSTM의 예측값 각각의 24개 시점에 대한 예측 그래프는 Figure 4.1과 같이 표현하였다.

예측결과로 표현된 LSTM과 SARIMA 모형의 두 그래프를 보면, 시각적으로는 큰 차이를 보이지 않는다. 그럼에도 LSTM을 최종모형으로 선택한 것은 그 PRESS 값이 SARIMA 모형의 값보다 더 작고 값의 차이가 적지 않다. 또한 SARIMA는 이론적으로 충실한 모형이고 잘 정립된 모형이며, 식별과 모수추정, 모형검진을 거친 통계적 모형으로 이론적인 강점이 있지만, RNN기반의 딥러닝 네트워크인 LSTM 기법이 시계열 데이터를 처리하는데 더욱 효과적이라 할 수 있다. 또한 새로운 데이터 입력에 대해 추가적인 분석을 수행할 때 LSTM 기법이 사용된다면, 파라미터 값만 변경하여 해당 자료의 특성을 반영할 수 있는 일반화의 장점도 있다.

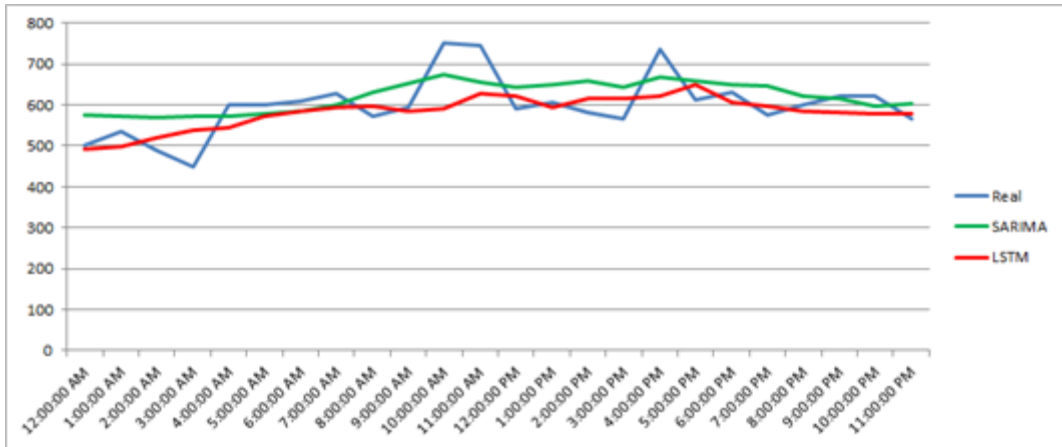


Figure 4.1 Comparison 24 points of forecasting result using SARIMA and LSTM method

4.3. 실험결과

본 연구에서의 목표는 87일중 86일간의 시간별 데이터를 가지고 학습을 수행하고 이후 1일간 24시간에 대한 예측값과 실제값을 비교하여 성능을 도출하는 것이다. SARIMA와 LSTM 여러 모델에서 최적 예측 모형을 도출하였고, 두 모형에 대한 성능 비교를 위한 PRESS를 통해 LSTM의 예측결과가 더 우수한 성능으로 도출되었다.

전반적으로는 안정적인 결과가 나왔지만 등락의 폭이 클 경우에는 예측오차가 커지는 경향이 있었다. 이러한 시계열 데이터를 활용한 LSTM 예측 연구에서는 과도한 학습량과 학습횟수가 오히려 과적합 이슈가 발생할 가능성이 있지만, 다양한 실험을 통해 세부 파라미터 값들이 적절히 조절되고, 기타 환경적 예측요소들이 선택되어질 수 있다면 향상된 정확도와 오차율을 기대할 수 있다. 따라서 현업 등의 환경에서 향후 계속적으로 축적되는 데이터에 대해서 예측분석이 필요할 경우, LSTM 모형을 통해 향후 24시간에 대한 전력사용 및 Peak전력 예측에 활용할 수 있다.

5. 결론

현재의 국내 전력수급 상황에서 ESS를 활용한 에너지 저장을 위한 연구는 중요한 부분이다. 심야의 전력을 저렴하게 충전하고 낮 시간대의 Peak 전력시에 효과적으로 대응하기 위해서는 24시간 기반의 전력사용량과 Peak 시간대의 기간을 예측하는 것이 가장 중요하다. 또한 비용 효율적으로 구축된 기존 ESS의 용량을 더욱 효과적으로 활용하기 위해서도 LSTM 기법을 통한 예측분석이 핵심적인 부분으로 간주될 수 있다. 기존 연구사례에서의 중장기적인 예측결과도 물론 쓰임이 있겠지만, 보다 높은 신뢰성으로 뒷받침할 수 있는 24시간 기반의 단기예측이 현업에서의 관련 시스템에서 기본적으로 중요하고도 중요한 역할로 반영될 것이다. 이러한 핵심적인 예측 기능은 전력 수급이나 대체전력 전환 등에 필요한 관리자의 사결정에 주요한 도움을 줄 것이다.

많은 연구에서와 같이 ESS의 활용도를 높이기 위해서 ESS와 V2G의 조합에 관한 연구도 전망이 밝다. 이에 따라 에너지 저장과 거래를 위한 에너지 Pool 형태의 기술 연구도 진행되고 있다. 이러한 영역에서도 Peak 전력의 예측 뿐만 아니라 전력 수요와 공급을 위한 예측분석에 LSTM 기법이 대응될 수

있고, 이러한 예측 결과를 기반으로 하는 의사결정을 수립할 수 있을 것이다.

국내 거의 대부분의 현업에서는 예측에 필요한 데이터의 양이 충분치 않음에도 불구하고 최적의 예측 결과가 요구된다. 따라서 현재 주어진 데이터를 가지고도 최적의 결과를 도출할 수 있도록 LSTM 파라미터의 추가적인 조정과 다양한 인공신경망 분석기법을 통한 추가적인 실험이 필요하다. 또한 더욱 정확한 예측을 위해서 기온과 강수량 등의 환경적 요소들도 모형 수립에 반영되어야 할 것이다. 향후 연구에서는 전력 사용량에 대한 대표적인 개형을 구분하는 등의 프로파일 시퀀스 추출을 통한 단기 예측으로 현업에서 필요로 하는 실질적인 연구를 수행하고자 한다.

References

- Ansari, M. et al. (2015). Coordinated bidding of ancillary services for vehicle-to-grid using fuzzy optimization. *IEEE Transaction on Smart Grid*, **6**, 261-270.
- Byeon, G. S., et al. (2016). Development of daily operation program of battery energy storage system for peak shaving of high-speed railway substations. *The Transactions of the Korean Institute of Electrical Engineers*, **65**, 404-410.
- Choi, J. W. and Lee, I. D. (2015). Energy ICT convergence with big data services. *Journal of the Korean Data & Information Science Society*, **26**, 1141-1154.
- Christopher, P. (2015). *Understanding LSTM Networks*, <https://www.cse.iitk.ac.in/users/sigml/lec/Slides/LSTM.pdf>
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, **9**, 1735-1780.
- Hwang, J. H., et al. (2014). The V2G operating method considering micro-grid. *The Transactions of the Korean Institute of Electrical Engineers*, 137-137.
- Kim, H. C. (2017). Trends and implications of V2G business in Korea and other countries. *Jeju Research Institute, Brief of policy and Issue*, **281**.
- Kim, J. H. and Hwang, S. W. (2016). A study on the decision of appropriate subsidy levels regarding electric vehicles for V2G as load management resources. *The Transactions of the Korean Institute of Electrical Engineers*, **65**, 264-268.
- Kim, S. K., et al. (2014). Sizing and economic analysis of battery energy storage system for peak shaving of high-speed railway substations. *The Transactions of the Korean Institute of Electrical Engineers*, **63**, 27-33.
- Krizhevsky, A. et al. (2012). Imagenet classification with deep convolutional neural networks. *In Proceeding of the 25th International Conference on Neural Information Processing Systems*, Nevada, U.S.A., 1097-1105.
- Lee, D. G., et al. (2018). Power consumption prediction scheme based on deep learning for powerline communication Systems. *Journal of IKEEE*, **22**, 822-828.
- Lee, J. H., et al. (2019). Analysis of time series to support decision making on V2G using energy consumption data. *Journal of the Korean Data & Information Science Society*, **30**, 401-414.
- Lee, S. D. (2017). *Time series analysis using SAS and R*, Free Academy.
- Moon, J. H., et al. (2017). Power consumption forecasting scheme for educational institutions based on analysis of similar time series data. *Journal of KIISE*, **44**, 954-965.
- Park, D. I. and Yoon, S. H. (2017). Clustering and classification to characterize daily electricity demand. *Journal of the Korean Data & Information Science Society*, **28**, 395-406.
- Ryu, J. A., et al. (2017). Forecasting daily building heating energy demand using ARIMA model. *Architectural Institute of Korea*, **37**, 608-611.
- Salehinejad, H. et al. (2018). *Recent advances in recurrent neural networks*.
- Shin, D. H. and Kim, C. B. (2018). Short term forecast model for solar power generation using RNN-LSTM. *Journal of Advanced Navigation Technology*, **22**, 233-239.
- Sibi, P. et al. (2013). Analysis of different a activation function using back propagation neural networks. *Journal of Theoretical and Applied Information Technology (JATIT)*, **47**, 1264-1268.
- Wikipedia (2018). *Recurrent neural network*, https://en.wikipedia.org/wiki/Recurrent_neural_network

Daily forecasting of energy demand using SARIMA and LSTM method to support decision making on V2G[†]

Jeong Hyun Lee¹ · Jae Sung Kim² · Young Ho Ahn³ · Wan Sup Cho⁴

¹²Department of Bigdata, Chungbuk National University

³RETIGRID

⁴Department of MIS, Chungbuk National University

Received 24 June 2019, revised 10 July 2019, accepted 11 July 2019

Abstract

Domestic energy demand is currently at risk and is related to the gradual phase-out of nuclear power plants and the expansion of EV (electric vehicle) charging stations. And also, there will be additional demand due to political problems on the Korean Peninsula. The purpose of this study is to predict the energy peak load and their time zone of day when switching from main power to ESS (energy storage system) or energy trading. So, our goal is to get more reliable results through a short time period analysis based on 24-hours. And SARIMA (seasonal ARIMA) and long short term memory (LSTM) were used to perform this study. We hope this model of analysis can be used positively on real world related with V2G (vehicle to grid) or ESS in the near future.

Keywords: ESS, energy peak load, LSTM, SARIMA, V2G.

[†] This research is supported by a research fund of university ICT center on Ministry of Science and ICT/IITP (IITP-2018-0-01396).

¹ Ph.D., Chungbuk National University, 1 ChungDae-ro, Seowon-gu, Chungbuk 28644, Korea.

² Ph.D. candidata, Chungbuk National University, 1 ChungDae-ro, Seowon-gu, Chungbuk 28644, Korea.

³ CEO, RETIGRID, Songpa bldg, 8 Baekjegobun-ro 41-gil, Songpa-gu, Seoul 05621, Korea.

⁴ Corresponding author: Professor, Department of MIS, Chungbuk National University, 1 ChungDae-ro, Seowon-gu, Chungbuk 28644, Korea. E-mail: wscho@chungbuk.ac.kr