

순환신경망을 이용한 태양광 발전량 예측 모델 설계

(Design of Photovoltaic Power Generation Prediction Model with Recurrent Neural Network)

김 한 호 [†] 탁 해 성 ^{††} 조 환 규 ^{†††}
(Hanho Kim) (Haesung Tak) (Hwan-gue Cho)

요 약 스마트그리드는 태양광 발전을 포함한 신·재생에너지의 발전량을 예측하고 이를 기반으로 효율적인 전력 생산과 소비를 가능하게 한다. 기존 태양광 발전량 예측 연구들은 시계열에 뛰어난 순환신경망 기법들을 적용 및 비교한 연구가 거의 없다. 또한 학습에 사용되는 과거 데이터의 길이에 대한 고려가 없어 모델의 예측 성능이 떨어졌다. 본 연구에서는 임베디드 변수 선택 기법을 이용하여 태양광 발전에 영향을 미치는 요인을 찾아내고, 시계열 순환신경망 기법들(RNN, LSTM, GRU)에 다양한 과거 데이터 길이를 넣는 실험을 진행하였다. 이 과정에서 가장 뛰어난 성능을 보이는 예측 요인들을 찾고 예측 모델을 설계하였다. 설계한 태양광 발전량 예측 모델은 다른 변수 설정을 사용할 때와 비교하여 더욱 뛰어난 예측 성능을 보이는 것을 확인하였다. 또한 기존 연구들과의 비교를 통하여 본 연구에서 개발한 태양광 발전량 예측 결과가 더 뛰어난 성능을 보이는 것을 확인하였다.

키워드: 스마트그리드, 순환신경망, 태양광 발전, 예측 모델

Abstract The Smart Grid predicts the power generation amount of renewable energy and enables efficient power generation and consumption. Existing PV power generation prediction studies have rarely applied and compared recurrent neural network techniques that are superior to time series. Furthermore, in the reported studies, there is no consideration of the length of past data used for learning, leading to lowered prediction performance of the model. In this study, we used the embedded variable selection techniques to find the factors influencing PV power generation. Subsequently, experiments were carried out to insert various past data length into the recurrent neural networks (RNN, LSTM, GRU). We found the optimal prediction factors and designed a prediction model based on the outcomes of the experiments. The designed PV power generation prediction model shows better prediction performance compared to other factor settings. In addition, better performance based on the prediction rate is confirmed in the present study as compared with the existing researches.

Keywords: smart grid, recurrent neural network, photovoltaic power generation, prediction model

- 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 기초연구실육성사업 연구임(NRF-2015R1A4A1041584)
- 이 논문은 2018 한국컴퓨터종합학술대회에서 '딥러닝을 이용한 태양광 발전 영향 요인 선별 및 예측 모델 설계'의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 부산대학교 전기전자컴퓨터공학과
quant1216@pusan.ac.kr

^{††} 비 회 원 : 부산대학교 전기전자컴퓨터공학과
tok33@pusan.ac.kr

^{†††} 종신회원 : 부산대학교 전기컴퓨터공학부 교수(Pusan Nat'l Univ.)
hgcho@pusan.ac.kr
(Corresponding author임)

논문접수 : 2018년 8월 13일
(Received 13 August 2018)

논문수정 : 2019년 3월 19일
(Revised 19 March 2019)

심사완료 : 2019년 3월 29일
(Accepted 29 March 2019)

Copyright©2019 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 제46권 제6호(2019. 6)

1. 서론

최근 통신 기술과 사물인터넷 기술의 발달로 사회적 문제를 해결하고자 하는 많은 시도가 있다. 그 중 전력 수요 증가 문제와 환경오염 및 자원 고갈 문제를 해결하기 위해 차세대 전력망으로 스마트그리드가 연구되고 있다[1]. 스마트그리드는 기존 전력망에 정보통신기술(ICT)을 더해 전력 생산과 소비 정보를 실시간으로 주고받음으로써 에너지 효율을 높이고자 하는 차세대 전력망이다.

기존 전력망은 최대 수요량에 맞춰 예비율을 두고 일반적으로 예상 수요보다 15% 정도 많이 생산하는 중앙 집중형 발전 형태로, 공급자 중심으로 설비를 운영한다. 전기를 생산하기 위해 화석연료로 대규모 발전을 하는데 이렇게 생산한 전기가 동시에 버려지기도 하는 효율성이 떨어지는 구조다. 하지만 스마트그리드는 실시간 정보 교환과 예측으로 에너지 낭비를 막고, 신·재생에너지를 활용하는 분산 발전 형태를 추구한다. 스마트그리드는 신·재생에너지의 효율을 최대화하여 효율적인 전력 생산과 소비를 가능하게 한다[2]. 스마트그리드의 성공적인 운용을 위해서는 수많은 신·재생 에너지원에 대한 예측 모델이 요구된다. 본 연구에서는 수많은 신·재생 에너지원 중 보급률이 높고 보편적으로 사용되는 태양광 발전량에 대해 예측하고자 한다[3].

태양광 발전량 예측은 스마트그리드라는 개념이 보편화되기 전에도 수많은 연구가 수행되었다. 기존 태양광 발전량 예측 연구들은 시계열에 뛰어난 신경망 기법들을[4] 적용 및 비교한 연구가 거의 없다. 또한 과거 데이터의 길이에 대한 실험 없이 고정된 길이의 데이터를 사용하여 모델의 예측 성능이 떨어졌다.

최근 개발된 순환신경망 기법들은 다양한 길이의 과거 데이터를 입력 데이터로 사용할 수 있어, 이전보다 변수에 대한 다양한 실험을 진행할 수 있게 되었다. 이에 따라 본 연구에서는 신경망 기법, 기상 변수 그리고 과거 데이터 길이에 대한 실험을 진행하여 태양광 발전에 영향을 미치는 최적의 예측 요인을 찾고 이를 기반으로 예측 모델을 설계하고자 한다.

2. 관련 연구

태양광 발전량을 예측하고자 하는 연구는 국내·외 모두 활발하게 이루어지고 있으며 스마트그리드가 대두되기 이전부터 태양광 발전량을 예측하는 연구는 많이 이루어져 왔다. 관련 연구를 설명하기 위해 수많은 예측 연구들이 결과 지표로 사용하는 MAPE(Mean Absolute Percentage Error)와 MSE(Mean Squared Error)를

설명한다. MAPE와 MSE는 예측 결과와 실제 결과의 차이를 절댓값 또는 제곱을 이용해서 구한 후 평균을 구한 값이다. MAPE는 식 (1) MSE는 식 (2)로 표현된다. n 은 평가에 사용한 데이터의 개수, Y_i 는 실제 값, \hat{Y}_i 은 예측된 값을 나타낸다.

$$MAPE = \frac{100\%}{n} \cdot \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (1)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2)$$

MAPE는 퍼센티지(percentage)로 표현되어 이해가 쉽다는 장점이 있지만, Y_i 와 \hat{Y}_i 의 차이가 클 경우 100%를 넘는 오차율이 나오는 단점이 있다. 또한 $Y_i = 0$ 을 포함하는 데이터는 비교 대상에 포함할 수 없다는 단점이 존재한다[5]. 반면에, MSE는 $Y_i = 0$ 을 포함하는 데이터를 다룰 수 있다는 장점이 있지만, 데이터의 크기에 의존적이기 때문에 동일한 스케일링이 진행되지 않은 실험 결과와는 비교가 어렵다는 단점이 있다[6].

Yasuhiko Hosoda[7]는 기온과 태양광 발전량의 데이터를 이용하여 태양광 발전량 예측을 수행하였다. 날씨에 따라 맑음, 흐림, 비 세 가지로 나누고 클러스터링을 이용하여 유사한 데이터들을 그룹화하였다. 그룹화한 데이터 각각에 대하여 회귀식을 정의하고 회귀식의 매개변수를 $H\infty$ 필터링 방식을 적용하여 찾아내었다. 연구 결과 MAPE 13.87%의 예측을 수행하였다. 정진화[8]는 일사량에 영향을 미치는 기온들을 선별하고 랜덤포레스트, 인공신경망, SVM(Support Vector Machine)을 이용하여 태양광 발전량을 예측하였다. 연구 결과 MAPE 11.71%의 예측을 수행하였다.

Atsushi Yona[9]는 인공신경망 기술과 RNN(Recurrent Neural Network)을 적용하여 예측을 수행하였으며 과거 24시간의 일사량, 기온, 대기 일사량, 습도 등을 이용하여 미래 24시간의 일사량과 기온을 예측하고 이에 산출식을 이용하여 태양광 발전량을 예측하였다. 연구 결과 MAPE 14.97%의 예측을 수행하였다. André Gensler[10]는 순환신경망 기술인 LSTM(Long Short Term Memory)을 접목하여 태양광 발전량을 예측하였다. 해당 연구에서는 과거 6시간의 기온과 일사량 데이터를 이용하여 예측을 수행하였고 해당 연구를 통하여 MSE 0.0049의 예측을 수행하였다.

본 연구에서는 기존에 연구되었던 RNN, LSTM과 같은 시계열 순환신경망 기법들에 최근 개발된 GRU(Gated Recurrent Unit) 순환신경망 기법을 추가하여 모델을 생성한다. 또한 과거 데이터 산정 시간에 대한 실험을 진행하여 성능을 향상시키고자 한다.

A)

<https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36>

3. 데이터 전처리 및 변수 선택

태양광 발전과 같은 친환경 에너지 발전은 기상에 많은 영향을 받기 때문에 태양광 발전량 데이터와 기상 데이터를 이용하여 태양광 발전량을 예측하였다. 태양광 발전량 데이터는 남부발전의 부산발전본부 내의 태양광 발전량 데이터를 사용하였고 기상 데이터는 부산발전본부 해당 지역의 기상 데이터가 존재하지 않아 주변 지역의 기상 데이터를 거리에 따른 가중치를 주어 평균값을 구하여 사용하였다. 데이터는 2015년 1월 1일부터 2017년 9월 19일까지의 자료를 사용하였다.

3.1 기상 데이터의 결측치 처리

발전량 데이터는 결측치가 없는 반면 기상 데이터는 48217개의 결측치가 존재하였다. 시간간수량은 비가 온 날에만, 일사량과 일조시간은 맑은 날에만 값을 기록하여 결측치는 0으로 처리하였다. 풍향의 경우 1시간 전의 값과 동일한 경우가 많아 1시간 전의 값으로 결측치를 처리하였다. 풍속의 경우 풍향의 변화가 없을 시 값의 변화가 선형적으로 변할 것으로 간주하여 선형 보간법을 이용하여 결측치를 처리하였다.

해면기압의 경우 연속된 4개의 값이 결측치로 되어있는데 결측치 이전 이후의 측정값 변화가 없어 결측치 주변 데이터를 그대로 사용하였다. 전운량의 경우 3시에서 18시까지 1시간마다 기록을 하지만 18시에서 3시까지는 3시간 단위로 기록을 하였다. 전운량은 선형적인 값의 변화가 예상되므로 선형 보간법을 이용하여 결측치를 처리하였다. 각 속성에 따른 결측치 개수와 처리 방법을 정리한 것은 표 1과 같다.

표 1 기상데이터 결측치 개수 및 처리 방법

Table 1 The number of missing value and processing method of meteorological data

Attribute	Number of missing value	Processing Method
Temperatures	0	×
Precipitation	21579	Set to 0
Wind direction	2	1 hour ago datum
Wind speed	2	Linear interpolation
Atmospheric pressure	4	Linear interpolation
Humidity	0	×
Cloud amount	5227	Linear interpolation
Duration of sunshine	10637	Set to 0
Solar irradiance	10766	Set to 0
Total	48217	

3.2 데이터 분리 및 스케일링

스케일링을 진행하는 데 있어 훈련 데이터와 검증 데이터를 분배하는 것이 매우 중요하다[11]. 태양광 전력 데이터와 기상 데이터 모두 계절의 영향을 받으므로 검증 데이터가 너무 짧은 경우 계절의 영향을 인식하지 못한다. 이를 방지하기 위해 2015-2016년의 데이터를 훈련 데이터로, 2017년의 데이터를 검증 데이터로 사용하였다. 발전량 데이터와 기상 데이터는 모두 정규화(normalization)를 이용하여 스케일링을 진행하였다.

3.3 임베디드 변수 선택 방법을 이용한 변수 선택

일반적으로 태양광 발전량에 영향을 주는 요인으로는 일사량을 먼저 고려하고 다음 요인으로 기온을 고려한다. 표 2는 추후 사용할 예측 모델 중 하나를 이용하여 일사량만 사용할 때, 일사량과 기온을 사용할 때, 모든 기상 데이터를 사용하였을 때의 예측 오차를 나타낸다. 일사량만 사용할 때보다는 일사량과 기온을 사용할 때 더 낮은 오차를 가지고 일사량과 기온을 사용할 때보다는 모든 기상 데이터를 사용할 때 더 낮은 오차를 가진다는 것을 알 수 있다. 일사량과 기온은 발전량 예측에 큰 영향을 미치지만, 이 외에도 태양광 발전량 예측에 영향을 미치는 요소가 있음을 알 수 있다.

모든 기상 데이터를 사용한 경우에는 유의미한 변수도 있지만 불필요한 변수도 추가됨으로써 오차를 더 증가시킬 수도 있다. 이러한 문제를 해결하기 위하여 현재 기상 데이터에서 유의미한 변수와 불필요한 변수를 찾아내는 변수 선택 과정을 수행하였다. 변수 선택 과정은 임베디드 변수 선택 방법을 이용하였다.

임베디드 변수 선택 방법은 모델의 학습 및 생성 과정에서 최적의 변수를 선택하는 기법으로 대표적으로 Lasso[12], Ridge[13], Elastic Net[14] 회귀 기법이 있다(표 3). 이 세 가지 기법은 각 독립 변수(기상 데이터)들이 종속 변수(태양광 발전량 데이터)에 어느 정도의 영향을 미치는지 분석하는 방법이다.

임베디드 변수 선택 방법에는 변수의 영향력에 대한 페널티인 α 값이 존재한다. α 값이 클수록 변수 하나에

표 2 태양광 발전에 영향을 미치는 다양한 변수 조합에 따르는 예측 오차

Table 2 Predictive error according to various combinations of variables affecting photovoltaic power generation

Variable combinations	Train data error	Validation data error
Solar irradiance	105.66	62.05
Solar irradiance & Temperatures	62.77	34.80
All Variables	55.57	31.59

표 3 임베디드 변수 선택 기법의 종류와 특징

Table 3 Types and characteristics of Embedded Variable Selection Techniques

Name	Regularization	Variable selection ability	Information loss risk
Lasso	L1(Absolute)	High	High
Ridge	L2(Square)	Low	Low
Elastic Net	L1 + L2	Medium	Medium

주어지는 페널티가 커져 더 적은 변수 선택을 수행한다. 본 연구에서는 다양한 α 값에 대하여 실험을 진행하였고 결과는 표 4와 같다. 기상 변수는 변수들 사이에 다중공선성 문제가 존재하기 때문에 일정하진 않지만 모든 경우에 일사량이 선택되었고 α 를 낮출수록 기온, 습도, 일조시간이 영향력 있는 변수로 선택되었다. $\alpha=0.001$ 일 때의 Lasso와 다른 변수 선택 결과를 비교해보면 기온, 습도, 일사량, 일조시간 변수가 약 88%의 설명력(R^2)을 가진 것을 알 수 있다. 그 외 나머지 5개 변수는 약 0.5% 이내의 영향력을 가지기 때문에 기온, 습도, 일사량, 일조시간을 최종 변수로 선택하였다.

4. 순환신경망을 이용한 태양광 발전량 예측 모델

연구를 통하여 태양광 발전량에 영향을 미치는 4가지 기상 변수를 선별하였다. 선별한 변수의 성능을 평가하기 위하여 일반적으로 사용되는 일사량과 기온만을 이용할 때, 본 연구에서 선별한 기상 변수를 이용할 때, 그리고 과거 태양광 발전량 데이터를 추가적으로 입력 받는 경우를 고려하여 5가지 입력 데이터 셋을 설계하였다. 태양광 발전량을 예측하기 위해 사용된 입력 데이터 셋은 표 5와 같으며 모든 데이터 셋의 출력 데이터는 태양광 발전량이다. 각 데이터 셋에 대하여 3가지 순환신경망 기법과 과거 데이터 산정 시간에 대한 실험을 진행하여 태양광 발전량 예측에 뛰어난 순환신경망 기법과 최적의 과거 데이터 산정 시간을 찾도록 하였다.

표 4 α 값에 따른 변수 선택 결과Table 4 The result of variable selection by α value

Name	α	R^2	Selected Variables
Lasso	0.01	0.8361	Solar irradiance, Duration of sunshine
Ridge	0.01	0.8871	All variables
Elastic	0.01	0.8406	Humidity, Solar irradiance, Duration of sunshine
Lasso	0.005	0.8608	Humidity, Solar irradiance
Ridge	0.005	0.8871	All variables
Elastic	0.005	0.8673	Temperatures, Humidity, Solar irradiance, Duration of sunshine
Lasso	0.001	0.8829	Temperatures, Humidity, Solar irradiance, Duration of sunshine
Ridge	0.001	0.8871	All variables
Elastic	0.001	0.8849	Temperatures, Atmospheric pressure, Humidity, Cloud amount, Solar irradiance, Duration of sunshine

표 5 예측 모델의 입력 데이터 종류

Table 5 Input data of the prediction model

No	Input Data
No.1	Photovoltaic power generation
No.2	Photovoltaic power generation, Solar irradiance, Temperatures
No.3	Photovoltaic power generation, Solar irradiance, Temperatures, Humidity, Duration of sunshine
No.4	Solar irradiance, Temperatures
No.5	Solar irradiance, Temperatures, Humidity, Duration of sunshine

4.1 순환신경망 모델 및 과거 데이터 산정 시간

순환신경망 모델은 시계열 데이터에 뛰어난 학습 능력을 보이는 3가지 순환신경망 모델인 RNN(Recurrent Neural Network)[15], LSTM(Long Short Term Memory)[16], GRU(Gated Recurrent Unit)[17]를 사용하였다. 최소 1시간 전부터 최대 과거 10일 전까지의 과거 데이터 산정 시간을 사용하여 실험하였으며, 은닉층과 은닉 노드의 수는 Jeff Heaton[18]의 책을 참고하여 조정하며 진행하였다. Tensorflow와 Keras를 사용하였으며 최종적으로 선택된 모델의 설계 사항은 표 6과 같다.

표 6 최종 모델 설계 세부 사항

Table 6 Final model specification

Item	Name
Model	RNN, LSTM, GRU
Hidden Layer	1 layers
Hidden Node	10 nodes
Optimizer	Adam Optimizer
Others	Default

4.2 태양광 발전량 예측 모델 실험 결과

과거 데이터 산정 시간이 3일을 넘어가는 경우에는 모든 데이터에 대해서 3일째의 데이터와 비슷한 결과가

도출되었기 때문에 과거 1시간에서 과거 3일까지의 실험 결과를 수록하였다. 그림 1-11은 표 5의 각 데이터셋에 대하여 순환신경망 모델과 과거 데이터 산정 시간에 따른 MSE 변화량, 그리고 MSE가 가장 낮을 때의 태양광 발전량 예측 결과이다. 실험 결과를 보면 과거 데이터 산정 시간에 따라 오차가 상승하는 경우가 있는데, 이는 과적합이 발생한 것으로, 훈련 데이터에서는 오차가 감소하지만 검증 데이터에서는 오차가 상승하는 것을 확인하였다.

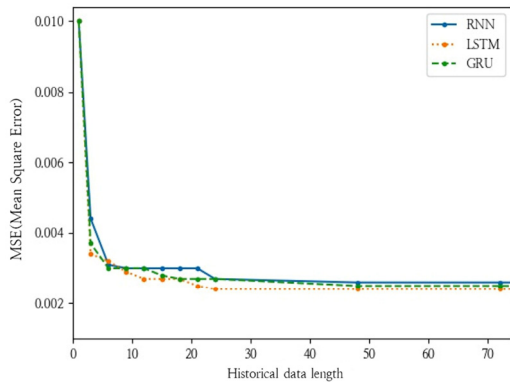


그림 1 데이터 1번의 과거 데이터 산정 시간에 따른 MSE 곡선. LSTM은 24시부터 수렴하는 반면 RNN과 GRU는 그 이후에 수렴하였다. LSTM 24시간에서의 오차율이 0.0024로 가장 적음을 확인하였다.

Fig. 1 The MSE curve according to the past data length of data #1. LSTM converges from 24 hours, while RNN and GRU converge thereafter. It was confirmed that the error rate at LSTM with 24 hours past data is the lowest at 0.0024

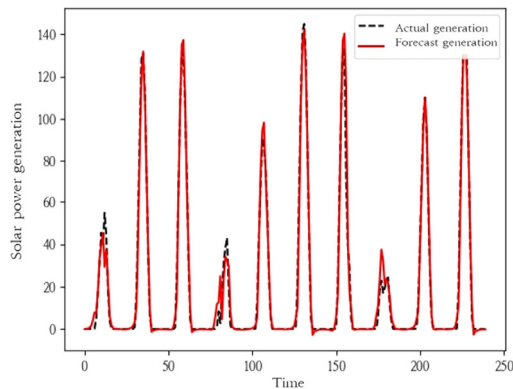


그림 2 데이터 1번의 LSTM + 24시간 과거 데이터의 예측 결과

Fig. 2 Forecast of LSTM with 24 hour past data of data #1

실험 결과, 데이터마다 약간의 차이는 있지만 약 24 시간까지는 오차율이 감소하는 경향을 보인다. 하지만 특정 길이의 과거 데이터 산정 시간에서는 훈련 데이터에 과적합되어 오차율이 증가하는 모습도 관찰할 수 있었다.

모든 실험에서 LSTM 기법과 과거 24시간의 데이터를 함께 사용할 때 뛰어난 성능을 보였으며, 데이터 간의

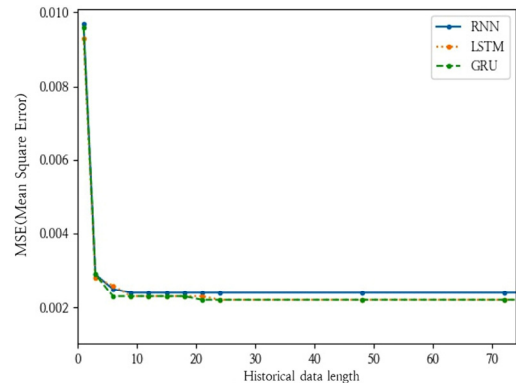


그림 3 데이터 2번의 과거 데이터 산정 시간에 따른 MSE 곡선. RNN은 9시간부터 수렴하는 반면 LSTM은 24시간, GRU는 21시간에 수렴하였다. LSTM 24시간과 GRU 21시간에서의 오차율이 0.0022로 가장 적음을 확인하였다.

Fig. 3 The MSE curve according to the past data length of data #2. RNN converges from 9 hours, while LSTM converges from 24 hours and GRU converges from 21 hours. It was confirmed that the error rate at LSTM with 24 hours past data and GRU with 21 hour past data are the lowest at 0.0022

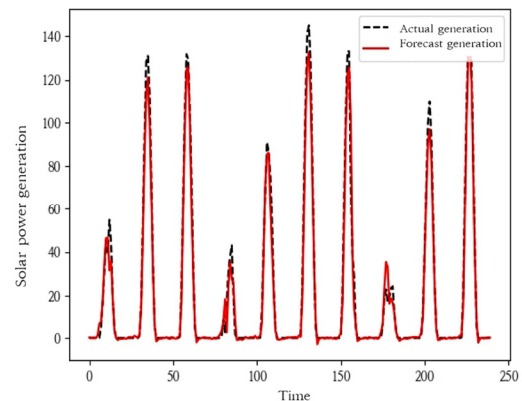


그림 4 데이터 2번의 LSTM + 24시간 과거 데이터의 예측 결과

Fig. 4 Forecast of LSTM with 24 hour past data of data #2

비교를 통하여 본 연구에서 선별한 4가지 기상변수(기온, 습도, 일사량, 일조시간)들이 다른 기상 변수 조건보다 오차율이 낮은 것을 확인하였다.

4.3 기존 태양광 발전량 예측 연구들과의 비교

기존 연구의 결과는 표 7에 MAPE(Mean Absolute

Percentage Error)와 MSE(Mean Squared Error)로 비교해 나타내었다. 기존 연구와 동일한 상황에서의 비교는 어려워, 기존 연구에서 사용한 상황과 데이터를 비슷하게 적용하여 실험을 진행하였다. 실험 결과 본 연구에서 제안한 예측 모델이 더 낮은 오차율을 가지는 것을 확인하였다.

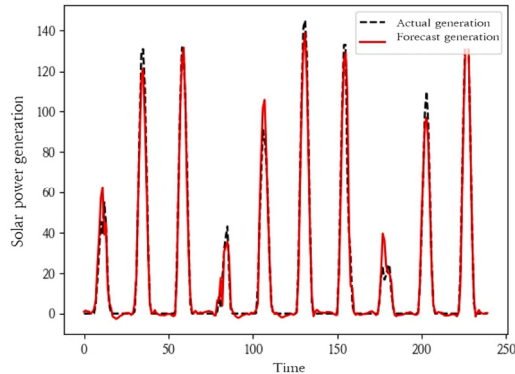


그림 5 데이터 3번의 과거 데이터 산정 시간에 따른 MSE 곡선. RNN은 24시간부터 수렴하였고 LSTM은 15시간에 오차가 잠시 상승하나 과적합이 일어났다가 해소되며 24시간부터 수렴하였다. GRU는 24시간에 수렴하였다. LSTM 24시간과 GRU 24시간 이후에서의 오차율이 0.0022로 가장 적음을 확인하였다.

Fig. 5 The MSE curve according to the past data length of data #3. RNN converges from 24 hours. LSTM get overfitting at 15 hours and then converges from 24 hours. GRU converges from 24 hours. It was confirmed that the error rate at LSTM with 24 hour past data and GRU with over 24 hours past data are the lowest at 0.0022

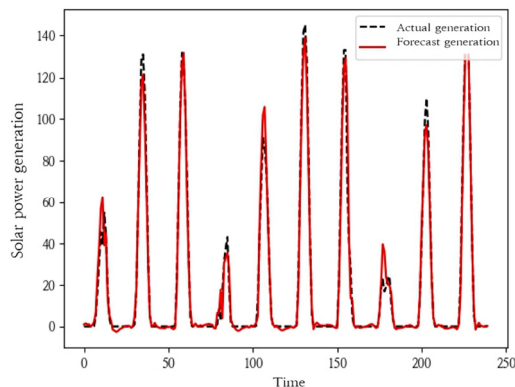


그림 6 데이터 3번의 LSTM + 24시간 과거 데이터의 예측 결과

Fig. 6 Forecast of LSTM with 24 hour past data of data #3

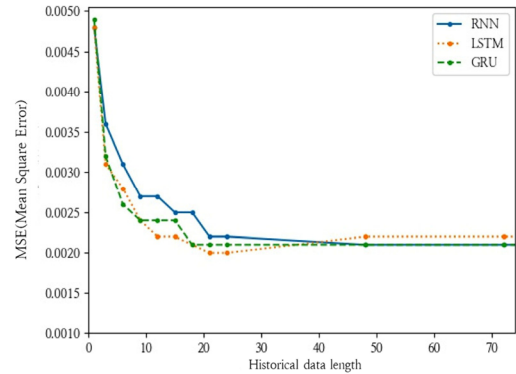


그림 7 데이터 4번의 과거 데이터 산정 시간에 따른 MSE 곡선. RNN은 24시간 이후에 수렴하였고 LSTM은 21-24시에 최저점을 찍고 이후 과적합으로 오차율이 증가하였고 GRU는 18시부터 수렴하였다. LSTM 24시간에서의 오차율이 0.002로 가장 적음을 확인하였다.

Fig. 7 The MSE curve according to the past data length of data #4. RNN converged after 24 hours. LSTM recorded the lowest error at 21-24 hours and an increase in the error occurred due to overfitting. GRU converges from 18 hours. It was confirmed that the error rate at LSTM with 24 hour past data is the lowest at 0.002

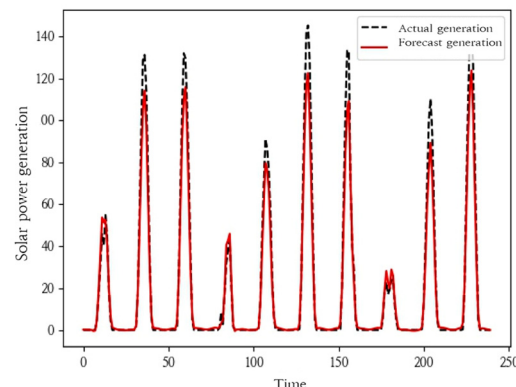


그림 8 데이터 4번의 LSTM + 24시간 과거 데이터의 예측 결과

Fig. 8 Forecast of LSTM with 24 hour past data of data #4

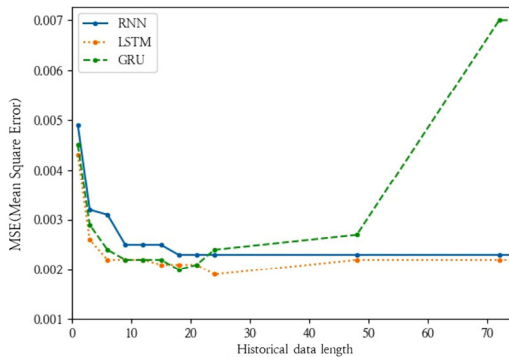


그림 9 데이터 5번의 과거 데이터 산정 시간에 따른 MSE 곡선. RNN은 18시간부터 수렴하였고 LSTM과 GRU는 각각 24시, 18시에 최저점을 찍지만 이후 과적합이 되어 오히려 오차율이 증가하였다. LSTM 24시간에서의 오차율이 0.0019로 가장 적음을 확인하였다.

Fig. 9 The MSE curve according to the past data length of data #5. RNN converges from 18 hours. LSTM and GRU take the lowest points at 24 and 18 hours, respectively. However, due to overfitting, an increase in the error rates occurred. It was confirmed that the error rate at LSTM with 24 hours past data is the lowest at 0.0019

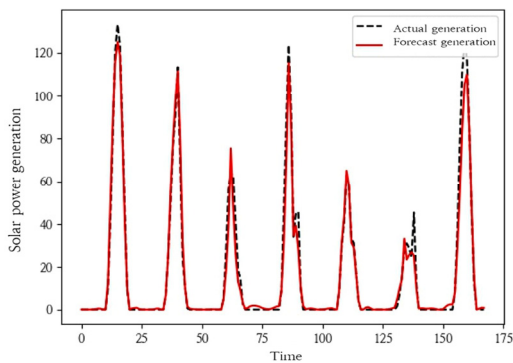


그림 10 데이터 5번의 LSTM + 24시간 과거 데이터의 예측 결과
Fig. 10 Forecast of LSTM with 24 hour past data of data #5

표 7 기존 태양광 발전 연구와의 성능 비교

Table 7 Performance comparison with existing photovoltaic power research

Author	MAPE	MSE
Yasuhiko Hosoda [7]	13.87	N/A
Jin-Hwa Jeong [8]	11.71	N/A
Atsushi Yona [9]	14.97	N/A
André Gensler [10]	N/A	0.0049
This Study	7.29	0.0019

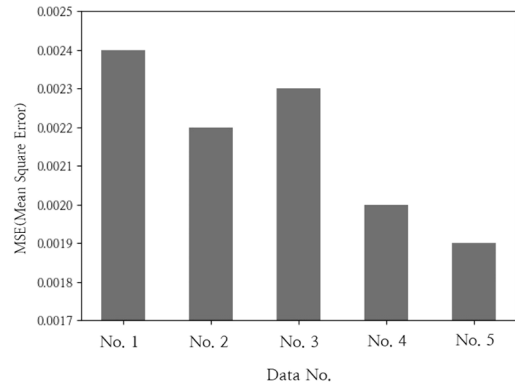


그림 11 모든 데이터 셋의 LSTM + 24시간 과거 데이터의 태양광 발전량 오차를 비교. 데이터 2번과 3번의 경우 태양광을 추가로 사용하였지만, 훈련 데이터에 과적합되어 오차율이 커진 것을 확인하였다. 태양광 발전량 예측을 위해 과거 태양광 발전량 데이터만을 사용할 때보다 기온과 일사량 데이터를 함께 사용할 때 더 오차율이 낮고 본 연구에서 찾은 4가지 기상 변수(기온, 습도, 일사량, 일조시간)를 사용하였을 때, 가장 오차율이 낮은 것을 알 수 있다.

Fig. 11 Comparison of PV power generation prediction error rate of LSTM with 24 hour past data at each data set. In the case of data 2 and 3, we used sunlight additionally, but the occurrence of overfitting was confirmed along with an increase in the error rate. Use of temperature and solar irradiance data together has a lower error rate compared to the use of past PV power generation data alone. Also, it was confirmed that the lowest error rate occurred when the four weather factors (temperature, humidity, solar irradiance, duration of sunshine) were used.

5. 결론 및 향후 연구

본 연구에서는 기존의 태양광 발전량 예측 연구에서 수행하지 않았던 순환신경망 기법과 과거 데이터 길이에 대한 실험을 진행하여 최적의 순환신경망 기법, 기상 변수 그리고 과거 데이터 산정 시간을 찾았다. 해당 과정에서 LSTM 순환신경망 모델이 태양광 발전에 뛰어난 성능을 보이는 것을 확인하였고, 4가지 기상 변수(기온, 습도, 일사량, 일조시간)가 큰 영향을 미치는 것을 확인하였다. 또한 과거 24시간의 데이터를 이용할 때, 태양광 발전량의 예측률이 가장 높음을 확인하였다. 이러한 최적 요인들을 이용하여 태양광 발전량 예측 모델을 설계하였고, 설계한 예측 모델은 기존 연구들보다 더 뛰어난 성능을 보이는 것을 확인하였다. 이를 정리하면

다음과 같다.

- 태양광 발전량 예측에 가장 중요한 변수는 기온, 습도, 일사량, 일조시간이다. 해당 변수들을 함께 사용할 때 가장 뛰어난 예측성적을 보인다.
- 과거 태양광 발전량 데이터를 입력으로 사용할 경우 과적합이 발생하여 오히려 성능이 저하되는 문제가 발생한다.
- 과거 24시간의 기상 데이터가 미래 태양광 발전량을 예측하기에 적합하다. 그 이상의 데이터를 이용할 경우 성능 향상이 거의 없고 과적합의 위험성이 존재한다.
- LSTM 모델이 다른 순환신경망 모델보다 태양광 발전량을 예측하는데 뛰어난 성능을 보인다.

본 연구의 다음 목표는 마이크로그리드에서의 효율적인 전력 사용을 가능하게 하는 것이다. 마이크로그리드는 소규모 지역에서 전력의 자급자족을 목표로 하는 독립형 스마트그리드를 의미한다. 마이크로그리드는 스마트그리드보다 신·재생에너지원에 대한 의존도가 높고 전력의 효율적인 소비가 매우 중요시된다. 본 연구에서 개발한 태양광 발전량 예측 모델을 이용하여 태양광 발전량을 예측하고 이 정보를 기반으로 최적 전력 배분을 수행하는 알고리즘을 연구하고자 한다.

References

- [1] D. Sung, N. Song, G. Go, J. Cha, G. Bae and H. Jang, "Convergence of Power System Technology and Information Communication Technology in Smart Grid," *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 31, No. 3, pp. 10-21, Mar. 2013. (in Korean)
- [2] M. Lee and S. Kim, "Analysis of Smart Grid Technologies and Domestic and Foreign Policy Trends," *Journal of Korean Institute of Information Technology*, Vol. 11, No. 8, pp. 181-187, May, 2013. (in Korean)
- [3] J. Su, *Energy supply and demand brief May issue*, pp. 1-15, Korea Energy Economics Institute, 2017.
- [4] J. C. B. Gamboa, "Deep Learning for Time-Series Analysis," *arXiv preprint arXiv:1701.01887*, Jan. 2017.
- [5] C. Tofallis, "A better measure of relative prediction accuracy for model selection and model estimation," *Journal of the Operational Research Society*, Vol. 66, No. 8, pp. 1352-1362, Aug. 2015.
- [6] E. L. Lehmann and G. Casella, *Theory of point estimation, 2nd Ed.*, pp. 99. Springer Science & Business Media, 2006.
- [7] Y. Hosoda and T. Namerikawa, "Short-term photovoltaic prediction by using H ∞ filtering and clustering," *SICE Annual Conference*, pp. 119-124, Aug. 2012.
- [8] J. Jeong and Y. Chae, "Improvement for Forecasting of Photovoltaic Power Output using Real Time Weather Data based on Machine Learning," *Journal of The Society of Living Environment System*, Vol. 25, No. 1, pp. 119-125, Feb. 2018. (in Korean)
- [9] A. Yona, T. Senjyu, A. Y. Saber, T. Funabashi, H. Sekine and C. Kim, "Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system," *Intelligent Systems Applications to Power Systems*, pp. 1-6, Nov. 2007.
- [10] A. Gensler, J. Henze and B. Sick, "Deep Learning for solar power forecasting - An approach using AutoEncoder and LSTM Neural Networks," *Systems, Man, and Cybernetics (SMC)*, pp. 2858-2865, Oct. 2016.
- [11] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, Mar. 2015.
- [12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B(Methodological)*, Vol. 58, No. 1, pp. 267-288, 1996.
- [13] S. L. Cessie and J. C. V. Houwelingen, "Ridge estimators in logistic regression," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 41, No. 1, pp. 191-201, 1992.
- [14] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 67, No. 2, pp. 301-320, Mar. 2005.
- [15] T. Mikolov, S. Kombrink, L. Burget, J. Černocký and S. Khudanpur, "Extensions of recurrent neural network language model," *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5528-5531, May, 2011.
- [16] H. Sak, A. Senior and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," *Fifteenth annual conference of the international speech communication association*, pp. 338-342, Sep. 2014.
- [17] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, Dec. 2014.
- [18] J. Heaton. *Introduction to Neural Networks in Java*, pp. 157-158. Heaton Research Inc., 2008.



김 한 호

2016년 부산대학교 통계학과 졸업(학사)
2018년 부산대학교 전기전자컴퓨터공학과 졸업(석사). 관심분야는 데이터마이닝, 딥러닝

**탁 해 성**

2013년 부산대학교 정보컴퓨터공학부 졸업(학사). 2015년 부산대학교 전기전자컴퓨터공학과 졸업(석사). 2002년~현재 부산대학교 전기전자컴퓨터공학과 박사과정. 관심분야는 데이터마이닝, 기계학습

**조 환 규**

1984년 서울대학교 계산통계학과(이학사). 1986년 KAIST 대학원 전산학과(공학석사). 1990년 KAIST 대학원 전산학과(공학박사). 1990년~현재 부산대학교 전기컴퓨터공학부 교수. 관심분야는 알고리즘, 응용그래프이론, 생물정보학