

기계학습을 활용한 기상예측자료 기반 태양광 발전량 예측 향상기법

정진화 · 채영태

청주대학교 건축공학과

Improvement for Forecasting of Photovoltaic Power Output using Real Time Weather Data based on Machine Learning

Jin-Hwa Jeong and Young-Tae Chae

Department of Architecture, Cheongju University, Cheongju, Korea

Abstract : This study investigated a selection of machine learning model to forecast electric power output from photovoltaic arrays based on forecasted weather data and historic solar radiation data. It tested two approaches to improve forecasting accuracy of power output with three typical machine learning algorithms such as Random Forest(RF), Artificial Neural Network(ANN), and Support Vector Machine(SVM). A forecasting power output was conducted with conventional weather forecasting data from national weather service which does not include solar radiation. The other approach has two steps, forecasting solar radiation with weather forecasting data and historic solar radiation data then it forecasts the electric power output of photovoltaic arrays. It has been studied the importance variables incorporated with the power output forecasting. The results show that the forecasting accuracy of the power output improves by using forecasted solar radiation data and Random Forest outperforms on this power output forecasting problem among other machine learning algorithms.

Key words : Photovoltaic power output forecasting, Machine learning model selection, Solar radiation, Weather data

1. 서 론

1.1. 연구 배경 및 목적

전 세계적으로 기후변화에 대응하기 위해 신재생 에너지 관련 산업이 활성화되고 있다. 신재생에너지는 당면한 기후변화에 대한 문제해결 및 온실가스 배출 저감에 매우 효과적이며 특히 다양한 신재생에너지 중 태양광 에너지는 고갈되지 않는 에너지원과 유지보수가 용이함 등으로 인해 가장 많이 활용되고 있다(Lee & Kim, 2016). 실제 태양광 에너지 시장의 규모는 2011-2016년 전 세계 기준 1순위인 풍력 에너지(285GW)의 뒤를 이어 태양광 에너지가 260GW의 시장규모를 나타냈으며 2017-2022년 기준 태양광 에너지는 438GW의 규모로 풍력에너지(321GW)를 앞설

것으로 예상된다(IEA, 2017).

최근 태양광 발전설비의 보급확대로 인하여 태양광 발전량을 보다 정확하게 예측하기 위한 연구가 국내외적으로 진행되고 있다. 하지만 급격히 변화하는 외부 기상 조건으로 인해 태양광 발전량 예측 성능 저하 문제가 발생되고 있으며 주로 단일 기계학습모델에 의존한 태양광 발전량 예측 연구가 진행되고 있다. 또한 태양광 발전량에 직접적인 영향을 미치는 시간별 태양일사량은 현재 국내 기상예보 데이터에 포함되고 있지 않기 때문에 기상예보에 의존한 태양광 발전량 예측 방법의 개선이 시급한 실정이다.

이에 본 연구에서는 실시간 기상정보를 활용한 3가지 기계학습모델 기반의 일사량 및 태양광 발전량 예측모델을 선정, 태양광 발전량 예측 성능 향상을 위해 각 입력변수 중요도 판별 및 일사량 유무에 따른 선정된 기계학습 모델 기반의 태양광 발전량 예측 성능 평가가 수행된다.

교신저자: 채영태 (우 28503) 충청북도 청주시 청원구 대성로 298, 청주대학교 건축공학과
전화번호 : +82-43-229-8479
E-mail : ychae@cju.ac.kr

1.2. 관련 연구 동향

최근 기상자료를 활용한 다양한 블랙박스 모델(Back box model) 및 물리적 모델(physical model) 기반의 태양광 발전량 예측에 관한 많은 연구가 국내외적으로 이루어지고 있다. 하지만 물리적 모델은 대상이 되는 시스템의 특성 등 이전 상황 및 조건을 고려하여 모델을 구성해야 하는 어려움이 있다. 반면 블랙박스 모델의 경우 어떠한 입력변수와 목표변수의 물리적 특성과 관계없이 데이터만으로 모델 구성이 가능하다는 장점으로 많이 활용되고 있다(Lennart, 2001). Lee와 Kim (2016)은 실제 기상데이터를 활용한 SVR(Support Vector Regression)모델 기반의 태양광 발전량에 가장 큰 영향을 미치는 일사량 예측을 수행하여 예측된 일사량으로부터 운량에 따른 태양광 발전량 예측을 수행하였다. 일사량 예측 결과, 305일 평균 MAPE=24.80%의 성능을 나타냈다. 앞선 예측된 일사량을 활용하여 운량에 따른 태양광 발전량 예측을 수행하였으며 운량 ≤13의 태양광 발전량 예측 결과, 평균 MAPE=14.90%, 운량 > 13의 경우 평균 MAPE=45.90%의 예측 성능을 보였다. Lee, Lee, Lee와 Lim, Ji (2015)는 실제 기상데이터를 활용한 ANFIS(Adaptive Neuro-Fuzzy Inference System)모델 기반의 각 입력인자에 따른 태양광 발전량 예측 비교분석을 수행하였다. 예측 결과, 일사량(과거 실측 데이터), 일조시간, 운량, 최고기온 데이터가 모두 활용되었을 경우 평균 MAPE=7.71%의 예측 결과가 도출되었다. Shi, Lee, Liu와 Yang, Wang (2012)은 실제 기상데이터를 활용한 SVM(Support Vector Machine)모델 기반의 천공상태에 따른 태양광 발전량

예측 성능 비교 분석을 수행하였다. 예측 결과, 맑은 날의 경우 평균 MRE=4.90%, 안개 낀 날의 평균 MRE=8.20%, 비오는 날의 평균 MRE=9.1%, 흐린 날의 평균 MRE=12.40%의 성능이 도출되었다. 위와 같이 국내의 경우 각 입력변수에 따른 단일 기계학습모델의 태양광 발전량 예측 성능 비교 분석에 관한 연구가 주를 이루고 있고, 국외의 경우 각 입력변수 및 기계학습모델간의 예측 성능 비교 분석 등 태양광 발전량 예측 성능 향상에 관한 다양한 연구가 진행되고 있다. 하지만 다양한 기계학습모델을 활용한 일사량 및 태양광 발전량의 기계학습모델 선정과 각 입력변수별 중요도에 따른 선정된 기계학습모델 기반의 태양광 발전량 예측 성능 향상에 관한 종합적인 연구는 아직 이루어지지 않고 있다.

2. 연구 방법

2.1. 기계학습모델

본 연구에서는 다양한 지도학습의 기계학습모델 중 활용분야별(의사결정, 분류, 예측) 성능이 우수하여 활용도가 높은 Random Forest (이하 RF) 모델, Support Vector Machine (이하 SVM) 모델, Artificial Neural Network (이하 ANN) 모델의 3가지 기계학습모델이 활용되었다. 앙상블(Ensemble) 학습 기반의 RF모델의 경우 전체 데이터 내에서 배깅(Begging) 및 부스팅(Boosting) 기법으로 여러 샘플 데이터를 추출 및 분류한다. 이후 여러 샘플 데이터에서 이전 학습 결과를 바탕으로 다음 학습 데이터의 샘플 가중치 조정을 통해

Table 1. Set Condition of Three Machine Learning Model

Machine learning model	Element	Configuration
Random Forest	Number of nodes	10000
	Tree depth	10
	Child node size	5
Artificial Neural Network	Architecture type	Feedforward neural network
	Type of training	Batch
	Number of hidden neurons	4
	Activation function	Hyperbolic tangent
	Partition dataset	70% / 30%
	(Training / Testing)	
Support Vector Machine	Kernel type	RBF(Radial Basis Function)
	RBF Gamma	0.1
	Gamma	1.0
	Regression precision(epsilon)	0.1

학습이 진행된다(Andy & Matthew, 2002). 위와 같은 학습방법을 통해 기존 의사결정나무(Decision tree)알고리즘에서의 일반화 성능저하 및 과적합(Overfitting) 문제가 개선되었다고 보고된 바 있다(Ramon & Sara, 2006). SVM모델의 경우 선형 데이터에서의 분류뿐만 아니라 비선형 데이터에서의 성능이 뛰어나다(Fung & Managasarian, 2005). 기본적으로 SVM모델은 분류 경계(hyperplane)를 정해 마진(margin)을 최대화하여 데이터를 분류하고, 특히 비선형 데이터일 경우 커널 트릭(Kernel trick)을 활용하여 고차원 특징 공간에서의 분류가 수행되어 분류 성능 향상에 기여한다고 보고된 바 있다(Simon & Daphne, 2001). 다층 퍼셉트론(Multi-Layer Perceptron)기반의 FFNN(Feed Forward Neural Network)모델의 경우 3개의 층(입력층, 은닉층, 출력층)으로 구성되어 각 층간의 이어진 노드를 통해 가중치를 조정함으로써 예측이 진행된다(Samanta & Balushi, 2003). 특히 은닉층에서는 입력변수에 따라 학습된 가중치의 결합 및 조정을 통해 오류를 최소화하여 출력층으로 전달하는 과정이 진행된다. 위 설명된 각 기계학습모델 기본설정조건은 Table 1과 같다.

2.2. 연구방법

국내 기상청에서는 Table 2와 같이 3시간 간격의 외기온, 상대습도, 풍향, 풍속, 천공상태, 강수확률 등의 11가지 항목의 기상예보 데이터가 제공되고 있다(KMA). 이에 해당지역인 경기도 용인시(우리동네 기상예보 좌표, X=62, Y=121)의 기상예보 데이터를 활용하여 사전적으로 변수 중요도 분석이 실시되었으며 그 수치가 0.02 미만인 최고온도(tmx), 최저온도(tmn), 강수코드(pty), 강수량(rd), 적설량(sd)을 제외한 외기온(temp), 상대습도(reh), 풍향(wd), 풍속(ws), 천공상

Table 3. Set Condition of Photovoltaic Generation System

Element	Variable
Type	SM-250PG8
Direction	South east
Capacity	50 kW
Module	250w (10 series × 20 parallel)
Rated voltage	30.8 V
Efficiency	15.08 %

태(sky), 강수확률(pop)의 6가지 항목의 기상예보 데이터가 활용되었다. 일사량의 경우 기상예보 데이터가 아닌 기상실측 데이터 형식으로 제공되고 있어 이를 태양광 발전량 예측의 입력변수로 활용하기 위해서는 우선적으로 일사량 예측모델이 수립되어야 한다. 이에 기상예보 데이터를 활용한 3가지 기계학습모델(RF, SVM, ANN) 기반의 일사량 및 태양광 예측모델이 선정되었다. 본 연구의 목표 변수인 태양광 발전량 시스템의 경우 경기도에 위치한 소형 규모의 업무용 건물에 설치되었으며 시스템의 개요는 Table 3과 같다.

수집된 1개월(16/12/01~16/12/31)의 기상예보 데이터 및 일사량 데이터를 활용한 3가지 기계학습모델(RF, SVM, ANN) 기반의 7월(17/01/01~17/01/07)간의 일사량 및 태양광 발전량 예측 모델이 선정되었다. 이후 태양광 발전량에 주요한 입력변수의 중요도 판별을 위해 6가지 기상예보 데이터를 활용한 7월(17/01/01~17/01/07)간의 변수 중요도 분석이 실시되었다. 중요도 판별을 통해 가장 높은 변수 중요도를 갖는 입력변수와 목표변수(일사량, 태양광 발전량)간의 상관관계 분석을 통한 변수 중요도 검증이 수행되었으며 두 변수의 선형관계 파악을 위한 피어슨(Pearson) 상관계수가 활용되었다. 식(1)은 상관계수 공식이며 여기서 r 은

Table 2. Used Weather Forecast Data for Korea Meteorological Administration

Variable	Input condition	Range	Unit
Outdoor air temperature(temp)	O	-	°C
Relative humidity(reh)	O	0 / 100	%
Wind direction(wd)	O	0 / 7	-
Wind speed(ws)	O	-	m/s
Sky code(sky)	O	1 / 4	-
Rainfall probability(pop)	O	0 / 100	%
Max outdoor air temperature(tmx)	X	-	°C
Min Max outdoor air temperature(tmn)	X	-	°C
Rainfall code(pty)	X	0 / 3	-
Rainfall depth(rd)	X	-	mm
Snow depth(sd)	X	-	mm

상관계수, I_i 는 입력변수, O_i 는 목표변수, n 은 데이터의 수를 의미한다. 마지막으로 태양광 발전량 예측 성능 향상을 위해 입력변수 및 예측된 일사량 데이터를 활용한 선정된 기계학습모델 기반의 Case별 7일(17/01/01~17/01/07)간의 태양광 발전량 예측 성능 비교 분석이 실시되었다. 일사량 및 태양광 발전량 예측 성능 평가를 위해 MAPE(Mean Absolute Percentage Error) 지표가 활용되었으며 식(2)와 같이 계산된다. 여기서 A_i 는 실제 발전량, F_i 는 예측 발전량, n 은 데이터의 수를 의미한다.

$$r = \frac{n(\sum I_i O_i) - (\sum I_i) \cdot (\sum O_i)}{\sqrt{(n \sum I_i^2 - (\sum I_i)^2) \cdot (n \sum O_i^2 - (\sum O_i)^2)}} \quad (1)$$

$$MAPE = \frac{\sum_{i=1}^n |A_i - F_i / A_i|}{n} \times 100(\%) \quad (2)$$

3. 연구 결과

3.1. 기계학습모델 선정

3.1.1. 일사량 예측모델 선정

태양광 발전량 예측의 입력변수로 일사량 예측 데이터를 활용하기 위해 6가지 항목의 기상예보 데이터를 활용한 3가지 기계학습모델(RF, ANN, SVM) 기반의 일사량 예측모델이 선정되었다. 동절기 기간에 해당하는 31일간(16/12/01~16/12/31)의 학습기간 데이터를 활용하여 7일간(17/01/01~17/01/07)의 일사량 예측이 수행되었다. 시간별 일사량 예측에 대한 오차율(MAPE)을 활용한 각 모델 성능 검토 결과, RF(MAPE=12.94%) > SVM(MAPE=27.62%) > ANN(MAPE=30.69%) 순

Table 4. Forecasting Results of Solar Radiation Based on 3 Machine Learning Model

Date	Model		
	MAPE (%)		
	RF	SVM	ANN
17/01/01	10.56 %	35.75 %	37.88 %
17/01/02	17.00 %	18.68 %	19.16 %
17/01/03	12.20 %	22.46 %	32.27 %
17/01/04	11.73 %	39.66 %	35.22 %
17/01/05	10.01 %	32.41 %	41.07 %
17/01/06	15.56 %	15.66 %	19.74 %
17/01/07	13.51 %	28.70 %	29.50 %
Mean	12.94 %	27.62 %	30.69 %

의 결과가 도출되었으며 RF모델의 경우 SVM모델, ANN모델에 비해 Figure 1과 같이 최대 일사량 구간에서의 예측 성능이 우수하였다. Table 4는 일별 일사량 예측 성능에 대해 일별로 정리한 것이다.

3.1.2. 태양광 발전량 예측모델 선정

태양광 발전량 예측모델 선정을 위해 6가지 항목의 기상예보 데이터를 활용한 3가지 기계학습모델(RF, ANN, SVM) 기반의 태양광 발전량 예측모델이 선정되었다. 일사량 예측모델 선정과 동일한 동절기 기간에 해당하는 31일간(16/12/01~16/12/31)의 학습기간 데이터를 활용하여 7일간(17/01/01~17/01/07)의 태양광 발전량 예측이 진행되었다. 태양광 발전량 예측 결과, RF(MAPE=17.81%) > ANN(MAPE=30.26%) > SVM(MAPE=31.92%) 순의 결과가 도출되었으며 앞서 설명한 일사량 예측문제와 동일한 형태로 RF모델이 ANN모델, SVM모델에 비해 Figure 2와 같이 기저 발전량 및 최대 발전량 구간에서의 예측 성능이 우수

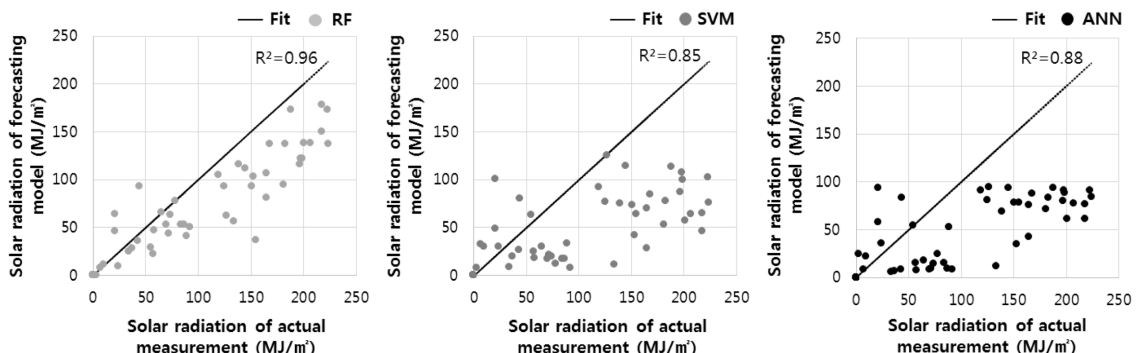


Figure 1. Comparison of solar radiation between actual measurement and forecasting model using R-squared.

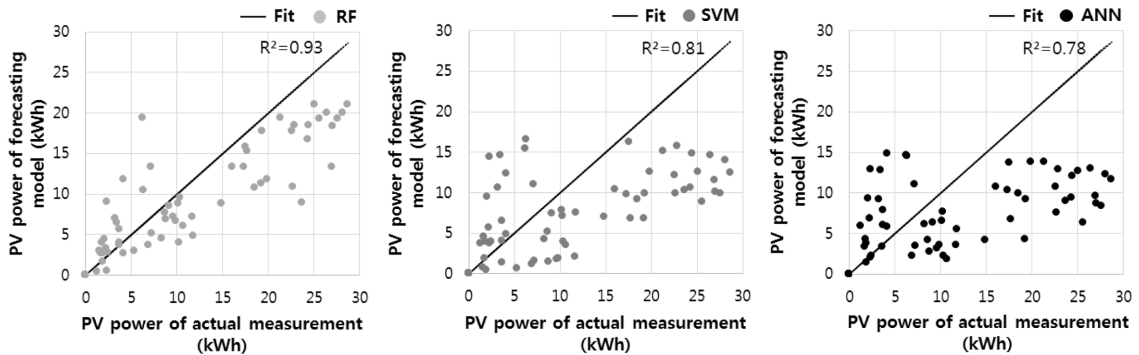


Figure 2. Comparison of PV power output between actual measurement and forecasting model using R-squared.

Table 5. Forecasting Results of PV Power Output Based on 3 Machine Learning Model

Date	Model	MAPE (%)		
		RF	ANN	SVM
17/01/01		14.70%	23.14%	37.59%
17/01/02		17.13%	31.06%	32.59%
17/01/03		20.41%	27.67%	28.59%
17/01/04		19.48%	37.81%	36.29%
17/01/05		14.42%	26.52%	23.60%
17/01/06		22.03%	33.04%	35.58%
17/01/07		16.47%	32.57%	29.18%
Mean		17.81%	30.26%	31.92%

하였다. Table 5는 일별 태양광 발전량 예측 성능에 대해 일별로 정리한 것이다.

3.2. 입력변수 중요도 및 예측 향상 기법

3.2.1. 입력변수 중요도

변수 중요도는 기계학습모델을 활용하여 여러 입력 변수와 목표 변수간의 상대적인 중요도를 나타낸 것으

Table 6. Results of PV Power Output of Feature Extraction for Weather Forecast Data

Indicators	Variable	reh	ws	temp	pop	wd	sky
Variable importance	RF	0.63	0.17	0.06	0.13	-	-
	ANN	0.30	0.12	0.22	0.12	0.13	0.12
	SVM	0.59	0.18	0.17	0.05	0.01	-
Mean		0.51	0.16	0.15	0.10	0.05	0.04

로 0-1 범위 내 수치 결과로 도출된다(Carolin, 2007). 변수 중요도 결과를 통해 예측에 유효한 입력변수 도출이 가능하며 중요도에 따라 예측에 활용될 변수 사용유무를 판단할 수 있게 된다.

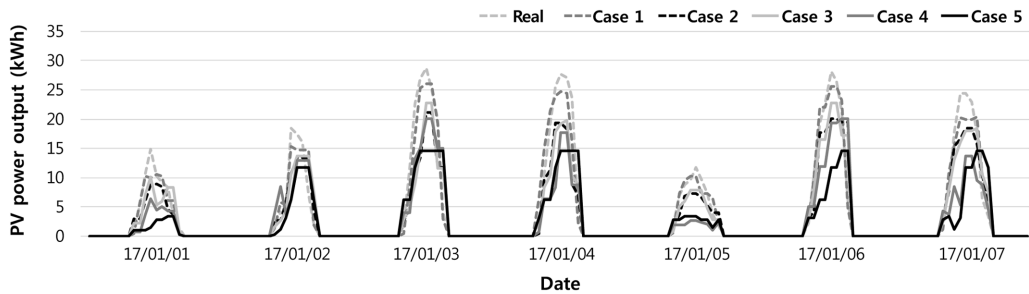
태양광 발전량의 입력변수의 중요도 분석을 위해 3가지 기계학습모델(RF, ANN, SVM) 기반의 동절기 기간에 해당하는 7일(17/01/01~17/01/07)간의 6가지 항목의 기상예보 데이터가 활용되었다. 기상예보 데이터의 변수 중요도 해석결과, Table 6과 같이 상대습도(reh) > 풍속(ws) > 외기온(temp) > 강수량(pop) > 풍향(wd) > 천공상태(sky) 순의 결과가 도출되었다.

Table 7. Correlation of Relative Humidity, Solar Radiation and PV Power Output

Date	Variable	Value			Correlation	
		reh (mean)	Solar radiation (sum)	PV power output (sum)	Solar radiation	PV power output
17/01/01		75.08	485.00	65.03		
17/01/02		61.46	533.00	71.06		
17/01/03		46.38	1033.00	142.74		
17/01/04		51.38	1044.00	140.79	-0.82	-0.81
17/01/05		72.17	375.00	61.56		
17/01/06		54.92	948.00	139.07		
17/01/07		66.98	919.00	127.98		

Table 8. Used Input Variables of PV Power Output Forecasting by Case

Element \ Case	Case 1 (all)	Case 2 (solar radiation x)	Case 3 (10% >)	Case 4 (20% >)	Case 5 (50% >)
reh	O	O	O	O	O
ws	O	O	O	O	X
temp	O	O	O	X	X
pop	O	O	O	X	X
wd	O	O	X	X	X
sky	O	O	X	X	X
radiation	O	X	X	X	X

**Figure 3.** Forecasting results of PV power output by input variable condition based on RF model.

RF모델과 SVM모델의 경우 풍향, 천공상태를 제외한 변수 중요도 순위가 동일하나 ANN모델의 경우 상대 습도, 외기온을 제외한 중요도 차이가 크지 않은 것을 확인할 수 있었다. 실제 가장 높은 변수 중요도를 갖는 상대습도를 중심으로 식(2)를 활용하여 상관관계를 분석한 결과 Table 7과 같이 일사량 및 태양광 발전량과 매우 강한 음의 상관관계가 나타났다. 이는 실제 흐린 날은 상대 습도가 높아짐에 따라 일사량이 적어지는 것으로 판단된다. 이후 변수 중요도 결과를 바탕으로 입력 변수의 추가 및 제거조건(Case)에 따른 태양광 발전량 예측 성능이 검증되었다.

3.2.1. 입력변수에 따른 예측 향상 기법

입력변수의 중요도 결과를 바탕으로 각 기상예보 데이터 및 일사량 예측 데이터의 유무에 따라 앞서 일사량 예측 및 태양광 발전량 예측에 가장 우수한 성능을 보인 RF모델 기반의 태양광 발전량 예측 성능 비교 분석이 실시되었다.

태양광 발전량 예측의 경우 5가지 Case에 대해 수행되었으며 각 Case별 활용된 입력변수는 Table 8과 같이 정리된다. Case 1은 6가지 항목의 기상예보 데이터 및 예측된 일사량 데이터가 모두 활용되었을 경우,

Case 2는 6가지 항목의 기상예보 데이터가 활용되었을 경우, 그리고 Case 3은 변수 중요도 0.10 미만의 입력변수가 제외된 경우이다. Case 4는 변수 중요도 0.20미만인 입력변수가 제외된 경우, Case 5는 변수 중요도 0.50미만인 입력변수가 제외된 경우를 나타낸 것이다. Figure 3은 각 Case별 태양광 발전량 예측 결과를 나타낸 것이며 Table 9는 일별 태양광 발전량 예측 결과가 정리된 것이다. Case 1인 6가지 항목의 기상예보 데이터 및 일사량 예측 데이터가 활용되었을 경우 해당

Table 9. Forecasting Results of PV Power Output by Case Based on RF Model

Date \ Case	MAPE (%)				
	Case 1	Case 2	Case 3	Case 4	Case 5
17/01/01	12.78%	14.70%	17.05%	21.26%	24.70%
17/01/02	11.33%	17.13%	21.16%	25.24%	23.73%
17/01/03	7.40%	20.41%	20.05%	26.24%	32.76%
17/01/04	16.64%	19.48%	22.13%	25.95%	31.55%
17/01/05	9.69%	14.42%	12.81%	23.12%	23.28%
17/01/06	8.79%	22.03%	18.02%	27.78%	27.01%
17/01/07	15.33%	16.47%	17.83%	24.19%	35.93%
Mean	11.71%	17.81%	18.44%	24.83%	28.42%

기간 내 일 최대 발전량, 일출, 일몰 구간 태양광 발전량 예측 성능이 가장 우수하였으며 일사량 예측 데이터가 없는 조건 내에서 Case 2와 Case 3에서의 평균 MAPE가 20% 미만에서 안정된 예측 성능이 확보되었다.

4. 결 론

본 연구에서는 국내 기상청에서 제공하는 실시간 기상정보를 활용한 3가지 기계학습모델 기반의 일사량 및 태양광 발전량 예측모델을 선정, 태양광 발전량 예측 성능 향상을 위해 각 입력변수의 중요도 판별 및 일사량 유무에 따른 선정된 기계학습 모델 기반의 태양광 발전량 예측 성능 평가가 진행되었으며 연구 결과는 다음과 같다.

(1) 기존 단일 기계학습모델이 아닌 3가지 예측모델 간의 성능 비교 분석이 가능하였으며 일사량 및 태양광 발전량 예측모델 선정 결과, RF모델의 예측 성능이 가장 우수하였다.

(2) 태양광 발전량에 대한 3가지 기계학습모델 기반의 각 입력변수의 중요도 판별결과, 상대습도>풍속>외기온 순의 중요도 결과가 도출되었으며 보다 정밀한 해석을 위해 일 평균 상대습도와 일사량 및 태양광 발전량의 상관관계 분석 결과, 매우 높은 음의 관계를 갖는 것을 확인할 수 있었다.

(3) 각 입력변수의 중요도 및 일사량 유무에 따른 태양광 발전량 예측 성능 비교 분석 결과, case 1인 6가지 항목의 기상예보 데이터 및 일사량이 활용되었을 경우 가장 우수한 예측 성능 결과가 도출되었으며 일사량을 제외한 case 2와 변수 중요도 0.10 미만의 변수를 제외한 case 3에서 평균 MAPE 기준 20% 미만의 안정된 예측 성능 확보가 가능하였다.

(4) 기상예보 데이터를 활용한 일사량 예측 데이터 유무에 따른 태양광 발전량 예측 결과, 일사량 데이터가 있을 경우 일사량 데이터가 없을 경우에 비해 MAPE 기준 6.10%의 향상된 예측 성능이 도출되었다.

향후 기계학습모델의 튜닝을 통해 최적화된 기계학습모델을 활용한 태양광 발전량 예측 향상 성능평가, 활용된 데이터의 프로세싱을 통한 예측 성능 비교분석에 관한 연구가 수행될 예정이다.

감사의 말

이 논문은 2018년도 청주대학교 연구장학 지원에 의한 것임

REFERENCES

- Lee, K. H. & Kim, W. J. (2016). Forecasting of 24_hours ahead photovoltaic power output using support vector regression. *Korean Journal of KITT*, 14(3), 175-183. <http://dx.doi.org/10.14801/jkiit.2016.14.3.175>
- IEA (2017). RENEWABLES 2017, IEA.
- Lennart, L. (2001). Black-box models from input-output measurements. *18th IEEE Instrumentation and Measurement Technology*, 10.1109/IMTC.2001.928802
- Lee, D. J., Lee, J. P., Lee, C. S., Lim, J. Y., & Ji, P. S. (2015). Development of PV power prediction algorithm using adaptive neuro-fuzzy model. *The Transactions of the Korean Institute of Electrical Engineers*, 64(4), 246-250. <http://dx.doi.org/10.5370/KIEEP.2015.64.4.246>
- Shi, J., Lee, W. J., Liu, Y., Yang, Y., & Wang, P. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications*, 48(3), 1064-1069. <http://dx.doi.org/10.1109/IAS.2011.6074294>
- Ramon, D. U. & Sara, D. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Biomaterials*, <https://doi.org/10.1186/1471-2105-7-3>
- Andy, L. & Matthew, W. (2002). Classification and regression by random forest. *R News*, 2(3), 18-22.
- Glenn, M. F. & Mangasarian, O. L. (2005). Multicategory proximal support vector machine classifiers. *Machine Learning*, 59(1-2), 77-97.
- Simon, T. & Daphne, K., (2001). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2, 45-66. <http://dx.doi.org/10.1162/153244302760185243>
- Samanta, B. & Al-balushi, K. R. (2003). Artificialneural-network based fault diagnosticsofrolling element bearings usingtime-domain features. *Mechanical Systems and Signal Processing*, 17(2), 317-328. <http://dx.doi.org/10.1006/mssp.2001.1462>
- KMA, http://www.weather.go.kr/weather/lifenindustry/sevice_rss.jsp, KMA.
- Carolyn, S., Anne, L.B., Achim, Z., & Torsten, H. (2007). Bias in random forest variable importance measures : Illustrations, sources and a solution. *Methodology Article*, <https://doi.org/10.1186/1471-2105-8-25>

투 고 일: 2018. 1. 22
수정접수일: 2018. 2. 23
게재승인일: 2018. 2. 23