



# 기상 예보를 이용한 머신러닝 알고리즘 기반 태양광 발전량 예측 기법

김 정 원\*

## A Solar Power Prediction Scheme Based on Machine Learning Algorithm from Weather Forecasts

Kim Jeongwon\*

### 요 약

태양광 에너지는 스마트그리드 시스템에서 신재생에너지로서 중요한 위치를 차지하고 있는데 기상 상황에 따라 발전량의 차이가 큰 편이다. 따라서 기상 상황에 따른 정확한 발전량의 예측은 스마트 그리드의 신뢰성과 운영의 효율성을 향상시킬 수 있다. 이 정확한 발전량 예측 문제를 해결하기 위해 본 논문은 특정 지점의 발전량 예측을 자동화하는 기법을 제안한다. 특정 지점에서의 발전량을 제공함에 있어서 해당 지점의 과거 발전량과 기상 데이터에 기반하여 머신 러닝 모델을 훈련하여 최적의 알고리즘을 찾아내고 기상 예보에 따라 미래의 발전량을 예측한다. 또한 발전량 예측에 있어서 어떠한 기상 변수가 중요한지 분석하여 기상 변수 적용에 따른 모델의 성능을 향상시킨다. 실험결과 최대 90% 이상의 정확도를 나타내었으며 가조시간, 일사량, 일조시간 등이 발전량에 중요 변수임을 확인하였다.

### Abstract

Solar energy occupies an important position as a renewable energy in the smart grid system and varies greatly depending on weather conditions. Therefore, accurate forecasting of power generation according to meteorological conditions can improve the reliability and operational efficiency of smart grid. To address this exact prediction problem, this paper proposes automatically creating site-dependent prediction scheme. In providing power generation at a specific point, the machine learning model is trained based on past power generation as well as meteorological data of the site to find the optimal algorithm and predict future power generation according to the weather forecast. In addition, it analyzes the importance of the meteorological variables in the forecasting of the power generation and improves the performance of the model. Experimental results showed maximum 95% accuracy, and confirmed that tuning time, solar radiation, and daylight hours are important parameters for power generation.

### Keywords

solar power, weather, machine learning, smart grid, renewable energy

\* 신라대학교 컴퓨터소프트웨어공학부  
- ORCID: <http://orcid.org/0000-0001-7069-8203>

• Received: Jul. 18, 2019, Revised: Sep. 22, 2019, Accepted: Sep. 25, 2019  
• Corresponding Author: Kim Jeongwon  
Department of Computer Software Engineering, Silla University, 140 Baegyang-daero(Blvd), 700beon-gil(Rd.), Sasang-Gu, Busan, Korea  
Tel.: +82-51-999-5749, Email: [jwkim@silla.ac.kr](mailto:jwkim@silla.ac.kr)

## I. 서론

스마트그리드는 전력을 생산, 분배, 그리고 소비를 관리하는 시스템으로 태양광이나 풍력 같은 친환경 재생 에너지의 비중을 높이는 것이 주 목적중 하나이다. 그러나 이 친환경에너지는 예측이 불가능하고 관리하기가 쉽지 않은 반면 스마트 그리드는 에너지 소비자에게 언제라도 필요한 전력을 공급해야 한다. 이를 위해 스마트 그리드는 전력 수요를 항상 모니터링하여 수요의 증감에 따라 필요한 발전소의 발전량을 조절한다. 다행히 가정, 공장 등의 전력 수요는 누적 자료에 의해 쉽게 예측 가능한데 전력을 수급하기 위한 발전소의 선택과 공급 시기를 거의 정확히 만족시킬 수 있다.

신재생에너지의 핵심인 태양광 에너지는 기상 조건과 지역적 특성에 따라 예측이 쉽지 않은데 대규모 그리드 시스템은 분산되어 있는 태양광 발전소의 전력 예측은 시간이 많이 소요되고 소규모 그리드의 경우는 수동으로 계산하는 불편함이 있다. 따라서 그리드에서 태양광 에너지의 비중을 높이고 운영의 효율성 제고를 위해서 정확한 예측이 필요한데 주로 기상 데이터를 이용한 모델이 다수 제안되고 있다. 이 모델들은 과거의 기상 데이터만 이용하는 것이 주류를 이루고 있고 기상 예보를 이용하는 기법도 일부 제안된 바 있다. 또한 태양의 위치에 기반한 기법도 있는데 이는 태양의 위치에 따라 태양에너지의 강도가 변하는 점에 기반하여 발전량을 예측하는 것이다.

기상데이터와 태양광의 세기에 기반하여 태양광 패널의 발전량을 예측하는 통계적 기법에는 주로 신경망 기법이나 최근에는 텐서플로우 같은 딥러닝이 사용되는데 본 연구도 이러한 통계적 기법에 따라 특정 지점의 태양광 발전량을 자동으로 예측하는 기법을 제안하고자 한다. 제안하는 기법의 차별점은 크게 특정 지점, 과거와 미래의 기상 데이터 사용, 그리고 과거의 발전량을 사용하는 것이다. 태양광 세기의 범위가 광범위하여 특정지점에서의 태양광 세기를 반영 하고 과거 기상 데이터하의 발전량은 미래의 해당 기상 조건하의 발전량과 높은 상관관계가 있으므로 적용한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 태양광 발전량 예측 기법에 대한 기존 연구들을 분석하고 3장에서는 기상 데이터 분석 방법, 발전량 예측 모델을 제안하고 4장에서는 제안한 기법의 성능 측정 결과를 제시하고 효율성을 분석한다. 그리고 5장에서 본 연구의 결론과 향후 과제를 제시한다.

## II. 관련 연구

태양광 발전량의 예측에는 주로 기상 데이터를 사용하는 방법이 주류를 이루고 있는데 먼저 과거의 데이터에 기반한 방법들이다. 이 기법은 현재 또는 과거의 기상 조건과 미래의 태양광 발전량은 높은 상관관계를 가진다는 사실에 기반하는데 신경망 기법[1], 회귀분석법[2] 등이 활용되었다. 또한 이전 기상 데이터의 시간적 구간을 확장하여 시계열 분석하는 방법인 ARMA(Autoregressive Moving Average) [3], RNN(Recurrent Neural Network)[4] 등을 이용하는 기법이 제안된 바 있다. 이런 기법들은 상당한 예측 정확도를 보여주고 있으나 미래의 기상 데이터를 사용하지 않는 점에서 최적의 기법이라 보기는 힘들다.

본 논문과 같이 기상 예보를 사용하는 다수의 기법들이 제안되었는데 미래 특정 시점의 기상예보를 기반으로 발전량을 예측하는 것이다[5]-[8]. 하지만 기상 예보의 정확도에 따라 알고리즘의 성능이 좌우되고 기상 예보 데이터가 현재의 기상 데이터보다 구체적이지 않으며 데이터의 양도 적은 편이다. 한편 과거의 데이터와 예보를 동시에 이용하는 기법들도 제안되고 있는데 적응적 선형 시계열 모델에 기반하는 방법[9]과 퍼지 결정 트리 모델에 과거 데이터와 예보를 동시에 적용하는 기법 등이 제안된 바 있다[10].

본 연구는 과거와 미래의 기상 데이터에 기반하여 발전량을 예측하는 점에서 기존의 기법들과 유사하다. 즉 과거 기상 데이터에서 발전량은 미래의 동일한 기상 조건에서 발전량과 높은 상관관계가 있음은 의심의 여지가 없다. 본 연구에서는 다양한 기상 변수에서 어떤 변수가 모델의 성능을 좌우하는지 분석하고 기상 변수들 사이에 어떠한 상관관

계가 있는지 분석하고자한다. 또한 다양한 머신 러닝 알고리즘을 발전량 예측에 적용하여 최선의 모델을 찾아내는데 기존의 연구는 태양광의 세기나 일조량에 기반하여 모델을 훈련하는데 본 연구에서는 해당 지점의 발전량에 기반하여 모델의 정확도를 높이고자 한다.

### III. 머신러닝 알고리즘과 기상 데이터에 기반한 태양광 발전량 예측 시스템

#### 3.1 기상 데이터 수집 및 분석

태양으로부터 나오는 에너지의 활용방법은 태양광과 태양열로 구분되며 태양광은 주로 태양전지를 이용한 태양광 발전에 활용되는데 태양에너지의 강도는 주로 일사량으로 나타내며 단위는 kWh/m<sup>2</sup>/day (하루에 도달하는 단위면적당 에너지량)를 사용한다. 기상 데이터와 태양 에너지와의 관계를 분석하기 위하여 한국에너지기술연구원의 홈페이지(<https://kredc.kier.re.kr>)에서 실험 대상 지역의 월단위 태양광 에너지 데이터를 수집하였다.

그림 1은 실험지역의 연중 일사량 자료인데 1, 2월과 11월, 12월의 동절기는 일사량이 봄, 여름보다 낮은 것을 알 수 있는데 이것은 기온이나 일조율, 그리고 일사율과 밀접한 관련이 있음을 짐작할 수 있다.

태양에너지와 관련이 있는 기상 변수로는 기온, 풍속, 습도, 운량, 일사율, 일조율, 그리고 청명일수

등이 있는데 이들의 과거데이터로부터 일사량과의 상관관계를 분석하여 발전량을 예측할 수 있다.

표 1은 기상 변수간의 상관관계와 기상 변수 각각과 일사량과의 상관관계를 분석한 결과로서 각 값은 피어슨 분석에 의한 상관계수이다.

분석에 사용된 기상 데이터의 단위는 기온은 0.1℃, 풍속은 0.1m/s, 습도는 %, 운량은 1/10, 일사율은 청명일기준 %, 일조율 0.1%, 패널의 설치경사각은 실험지역에 기반하여 33°, 그리고 일사량은 kWh/m<sup>2</sup> 이다.

피어슨 분석에 의하면 상관계수의 절대값이 커질수록 두 기상 데이터간의 상관 강도는 높아지는데 양의 값은 두 기상 데이터가 선형적으로 커지고, 음의 값은 선형적으로 감소한다. 상관계수 값 자체가 높다고 해서 일사량을 결정하는 중요 변수가 될 수도 있지만 반드시 그렇다고 할 수는 없다.

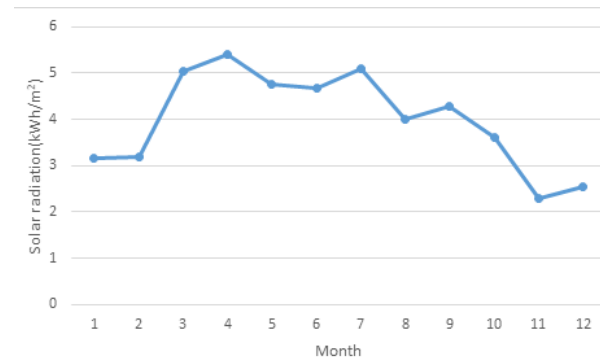


그림 1. 실험 지역의 일사량  
Fig. 1. Solar radiation at experiment area

표 1. 기상 데이터간의 상관관계

Table 1. Correlation between weather data

	Temperature	Wind speed	Humidity	Cloud	Rate of sunshine	Sunshine hours	Clear days
Temperature	1.00	-0.22	0.97	0.85	0.51	-0.80	-0.85
Wind speed	-0.22	1.00	-0.12	0.10	0.17	-0.12	-0.08
Humidity	0.97	-0.12	1.00	0.93	0.47	-0.90	-0.92
Cloud	0.86	0.10	0.93	1.00	0.53	-0.99	-0.99
Rate of sunshine	0.51	0.17	0.47	0.53	1.00	-0.47	-0.58
Sunshine hours	-0.81	-0.12	-0.90	-0.99	-0.47	1.00	0.98
Clear days	-0.86	-0.08	-0.92	-0.99	-0.58	0.98	1.00

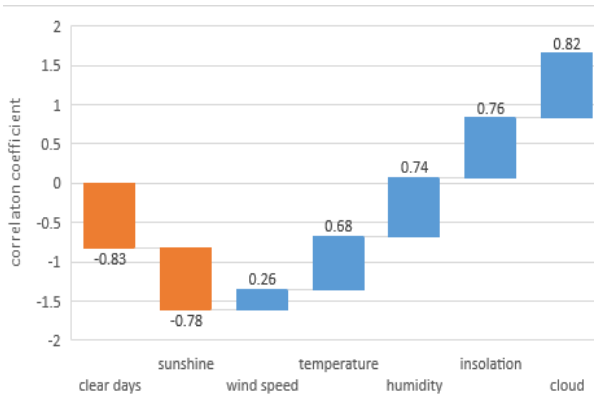


그림 2. 기상 데이터와 일사량의 상관관계  
Fig. 2. Correlation between weather data and solar radiation

표 1과 그림 2에서 보면 상관계수는 운량, 일사율, 습도, 기온, 풍속, 일조율, 그리고 청명일수 순인데 운량과 일사율이 일사량에 밀접한 관련이 있음을 알 수 있고 반대로 날씨가 청명하다고 해서 태양에너지가 높지는 않다는 것을 나타내고 있다. 따라서 이들 기상 변수가 머신러닝 모델에서 주요한 특징이 될 것임을 기대할 수 있다. 한편 표 1의 상관계수가 일사량을 결정하는 주요 요인의 중요도를 나타내고 있다고 보기는 어렵지만 다른 기상 변수에 대한 상대적 중요도는 결정할 수는 있다.

### 3.2 머신러닝 모델

이 절에서는 본 논문과 관련된 다양한 발전량 예측에 사용된 머신러닝 알고리즘에 대해 논한다. 먼저 Linear Regression은 간단하면서 독립변수와 종속변수 사이의 관계를 예측할 수 있으며 SVR (Support Vector Regression)은 극단값의 효과를 최소화하기 위해서 임계치보다 작은 예측 오류값은 무시하는 일종의 Linear Regression이다. SVR은 다양한 커널을 제공하는데 본 연구 환경에서 가장 좋은 성능을 보인 poly를 사용하였다. k-NN은 가장 간단한 알고리즘으로 훈련 데이터셋을 단순 저장하는 것이 모델을 만드는 과정의 전부이고 테스트 포인트에 대해 가장 가까운 데이터 포인트를 찾는 것이다. MLP Regression은 다층 퍼셉트론 모델로서 과적합 방지를 위해서 부분적인 비용 함수를 적용하고 반복적으로 모델을 훈련시킨다.

마지막으로 트리 모델로서 RFR(Random Forest Regression), 에이다부스트(AdaBoost), 그레이디언트 부스팅(Gradient boosting)을 적용하였는데 데이터에 존재하는 복잡한 상호 관계에 따른 숨겨진 패턴들을 발견하는 이 모델은 예측변수들 사이의 관계를 단순 트리 모델로 표시할 수 있고 해석이 비교적 쉽다는 점이 있다. 그레이디언트 부스팅은 residual fitting method로서 가장 인기 많은 베이스모델은 결정트리로서 다른 통계모형과는 다르게 가정이 적고, 범주형이든 연속형이든 제약 없이 쉽게 만들 수 있다. 본 연구에서는 상기의 7개 모델을 적용하여 성능을 평가한다.

본 논문에서 적용한 알고리즘들의 예측 정확도를 평가하기 위해서 2년간의 데이터를 대상으로 훈련 집합과 테스트 집합으로 분리하여 MSE(Mean Squared Error), R-squared 값, 즉 결정 계수를 계산하는데 MSE의 경우 데이터의 스케일에 따라 값의 변화가 심하므로 이 값만 보고 판단하기 어려움이 있으므로 결정 계수인  $R^2$ 를 적용하였다. 이를 통해 모델의 적합도가 충실하게 반영되므로 표본이 모델에 의해 얼마나 정확하게 예측되는지를 알 수 있는데 1은 100% 정확하게 예측함을 의미한다. 아래의 식 (1), (2)는 MSE,  $R^2$  계산에 사용한 공식으로  $t_i$ 는 실제값,  $\bar{t}_i$ 는 실제값의 평균이고  $y_i$ 는 예측값,  $\bar{y}_i$ 는 예측값의 평균이다.

$$MSE = \frac{1}{N} \sum_{i=1}^N (t_i - y_i)^2 \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (t_i - y_i)^2}{\sum_{i=1}^N (t_i - \bar{t}_i)^2} \quad (2)$$

## IV. 실험 및 성능 평가

기상 예보에 따른 태양광 발전량의 예측을 위해 2017년 7월부터 2019년 6월까지 실험지역 (위도 35도6분10초, 경도 129도2분25초)의 2년간 태양광 발전량을 기록하고 기상청 기상예보와 실제 기상데이터를 저장하였다. 실험에 사용된 태양광 패널은 120W 용량의 패널 2장을 병렬로 연결하여 배터리

에 충전하고 실제 발전량데이터는 충전 전류값이 0보다 큰 시각에서 0보다 작은 시각까지 배터리의 충전 전류량의 차이를 하루 발전량으로 정하고 데이터베이스에 기록하였다. 머신 러닝 알고리즘의 학습을 위해 2년간의 데이터를 훈련데이터와 테스트 데이터로 구분하여 모델을 훈련하였으며 예측 모델은 scikit-learn 머신러닝 패키지에 있는 알고리즘을 파이썬으로 구축하였다.

표 2. 모델의 성능  
Table 2. Performance of model

Algorithm	MSE	$R^2$
k-NN	52927.6744	0.67
Linear regression	44673.1629	0.83
SVR(poly)	43269.3592	0.86
AdaBoost	47800.4472	0.89
RFR	38675.2506	0.91
MLP	36440.3623	0.93
Gradient boosting	36630.9446	0.94

표 2는 각 알고리즘의 MSE와  $R^2$  를 나타내고 있는데 결정계수 값이 0.67에서 0.94로 다양하게 나타나고 있으나 MSE의 경우 값만 가지고 모델의 정확도를 예측하는 것이 어렵다는 것을 알 수 있다. 이 결과에서는 그레이디언트 부스팅이 가장 좋은 결과를 보이고 있고 k-NN의 성능이 최하인데 단순 거리 값으로는 예측이 힘들다는 것을 보여준다.

태양광 발전량을 예측하는 머신러닝 알고리즘들의 성능을 평가하는데 있어서 각 모델의 비용함수

를 사용하지만 어떤 기상 변수가 각 모델의 성능을 결정하는 핵심 요소인지 확인하는 것과 각 모델에서 변수들의 중요도 패턴을 분석하는 것도 모델의 신뢰성을 높이는 방법이다. 본 실험에서는 여러 모델들의 집합을 통해 예측 결과를 만들어내는 앙상블 기법들이 우수한 성능을 보이므로 이들 모델들의 Gini 중요도, 즉 MDI(Mean Decrease in Impurity)를 통해 각 변수의 중요도를 분석하고자 한다.

그림 3은 레이디언트 부스팅, RFR, 에이다부스트, 그리고 MLP 모델의 Gini 중요도를 나타내고 있다. 이들 모델 모두 비슷한 패턴을 보이고 있는데 가조시간, 합계 일조시간, 합계일사 등이 높은 중요도를 보이고 있다. 가장 높은 성능을 보이는 그레이디언트 부스팅의 경우 가조시간은 41%, 합계일사는 19%, 합계일조시간은 16.8%의 중요도를 보이고 있는데 이들 변수의 합이 76.8%로서 모델의 성능을 결정하는 중요 변수임을 알 수 있다. 이는 표 1에서 운량과 일사율이 일사량과 상관관계가 높은 것이 유효함을 나타내고 있다.

그림 4는 가장 좋은 성능을 보인 그레이디언트 부스팅 알고리즘의 학습 세트에서 예측 발전량과 실제 발전량을 비교한 그래프이다. 전반적으로 예측 발전량의 추이가 실제 발전량을 따라가는 것을 알 수 있다. 다만 피크 발전량에 있어서는 예측량이 실제 발전량을 따라가지 못하는 경우가 다수 보이는데 이것은 기상 예보 값이 실제 기상 상황과 일부 차이가 발생된 것이 영향을 미친 것으로 분석된다.

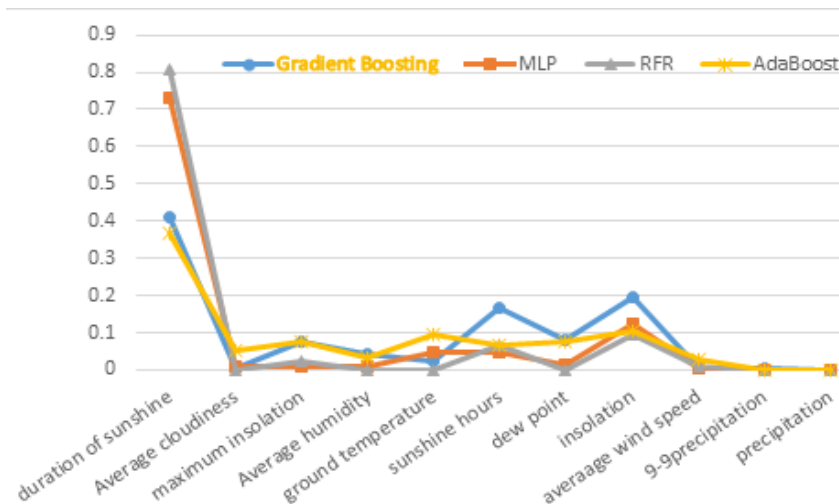


그림 3. Gini 중요도  
Fig. 3. Gini importances

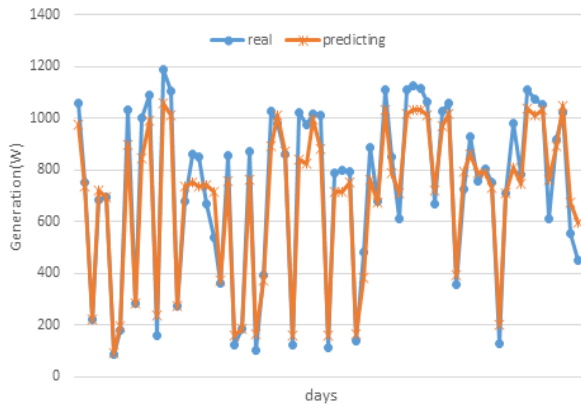


그림 4. 태양광 발전 예측  
Fig. 4. Power generation prediction

상기의 분석에 의하면 태양광 발전에 있어서 태양광이 하루에 존재하는 시간인 가조시간이 태양광이 지표면에 직접 도달하는 시간인 일조시간보다 모델의 Gini 분석에서 높은 중요도를 가지는 것은 패널에 태양광이 직달하는 것도 중요하지만 반사에 의해 도달하는 태양광도 모델의 성능에 영향을 미친다는 것을 보여주고 있다. 그리고 일사량도 순간적인 태양에너지를 나타내므로 가조시간에 비해 중요도가 낮은 것을 나타내고 있다.

한편 본 연구에 적용된 머신러닝 알고리즘들에서 앙상블 기법들이 타 기법에 비해 나은 성능을 보이는데 가조시간, 일조시간, 일사량, 온도, 습도 등 다양한 변수가 모델의 성능에 관련되어 있는 발전량 예측 알고리즘에서 각 라운드마다 레코드와 열을 재표본 추출하여 라운드마다 잔차가 큰 레코드들이 가중치를 높여 모델을 생산하기 때문으로 분석된다.

## V. 결 론

본 논문에서는 특정 지역에서의 태양광 발전량을 예측하는 방법을 제시하고 있는데 과거 기상데이터와 특정 지역의 발전량 데이터를 기반으로 하여 다양한 머신러닝 알고리즘을 통해 예측 모델을 훈련시키고 기상 예보에 따라 발전량을 예측한다. 모델을 훈련하기 전 기상 변수들 간의 상관관계, 그리고 발전량에 중요한 요소인 일사량과의 상관관계를 분석하여 어떠한 기상 변수를 모델의 입력으로 사용할지 결정하였다. 모델의 정확도 평가를 위해 MSE,  $R^2$ 를 통해 성능을 평가하였는데 앙상블 기법들이

나은 성능을 보였고 특히 확률적 그래디언트 부스팅 기법이 가장 좋은 성능을 보였다.

본 연구에서 제시한 발전량의 예측은 자동화가 가능하므로 그리드에서 신재생 에너지의 발전량을 예측하여 소비 전력대비 생산 전력의 예측을 용이하게 할 수 있으며 이는 스마트 홈뿐만 아니라 분산 그리드에도 적용 가능할 것으로 기대되는 한편 독자적인 예측 모델과 이에 대한 성능 분석 방법 등 추가 연구가 필요할 것으로 판단된다.

## References

- [1] A. Chaouachi, R. M. Kamel, and K. Nagasaka, "Neural network ensemble-based solar power generation short-term forecasting", *Journal of Advanced Computational and Intelligence Informatics*, Vol. 14, No. 1, pp. 69-75, Jan. 2010.
- [2] M. R. Hossain, A. M. Oo, and A. S. Ali, "Hybrid prediction method of solar power using different computational intelligence algorithms", In *Proceedings of the Power Engineering Conference*, Vol. 4, No. 1, pp. 76-87, Sep. 2012.
- [3] M. David, F. Ramahatana, P. J. Trombe, and P. Lauret, "Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models", *Solar Energy*, Vol. 133, pp. 55-72, Aug. 2016.
- [4] N. Sharma, J. Gummesson, D. Irwin, and P. Shenoy, "Cloudy computing: Leveraging weather forecasts in energy harvesting sensor systems", In *Proceedings of the 7th Annual IEEE Communications Society Conference, Sensor Mesh and Ad Hoc Communications and Networks*, Boston, MA, USA, pp. 1-9, Jun. 2010.
- [5] N. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning", In *Proceedings of the 2nd IEEE International Conference, Smart Grid Communications*, Brussels, Belgium, pp. 528-533, Oct. 2011.
- [6] B. Amrouche and X. Le Pivert, "Artificial neural

network based daily local forecasting for global solar radiation", *Applied Energy*, Vol. 130, pp. 333-341, Oct. 2014.

- [7] M. Zamo, O. Mestre, P. Arbogast, and O. Pannekoucke, "A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production", *Solar Energy*, Vol. 105, pp. 792-803, Jul. 2014.
- [8] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep Learning for solar power forecasting-An approach using AutoEncoder and LSTM Neural Networks", In *Proceedings of the 2016 IEEE International Conference, Systems, Man, and Cybernetics*, Budapest, Hungary, pp. 2858-2865, Oct. 2016.
- [9] M. Detyniecki, C. Marsala, A. Krishnan, and M. Siegel, "Weather-based solar energy prediction", In *Proceedings of the 2012 IEEE International Conference, Fuzzy Systems*, Brisbane, QLD, Australia, pp. 1-7, Jun. 2012.
- [10] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting", *Solar Energy*, Vol. 83, No. 10, pp. 1772-1783, Oct. 2009.

## 저자소개

김 정 원 (Kim Jeongwon)



1995년 2월 : 부산대학교

전자계산학과(이학사)

1997년 2월 : 부산대학교

전자계산학과(이학석사)

2000년 8월 : 부산대학교

전자계산학과(이학박사)

2000년 9월 ~ 2002년 2월 :

기술보증기금 전문직 차장

2002년 3월 ~ 현재 : 신라대학교 컴퓨터소프트웨어공학부  
교수

관심분야 : 임베디드시스템, 빅데이터, 운영체제, 모바일  
컴퓨팅