

## 서포트 벡터 회귀를 이용한 24시간 앞의 태양광 발전량 예측

Forecasting of 24\_hours Ahead Photovoltaic Power Output Using Support Vector Regression

---

저자 (Authors)	이강혁, 김우제 Kanghuyk Lee, Woo-Je Kim
출처 (Source)	<a href="#">한국정보기술학회논문지 14(3)</a> , 2016.3, 175-183(9 pages) <a href="#">The Journal of Korean Institute of Information Technology 14(3)</a> , 2016.3, 175-183(9 pages)
발행처 (Publisher)	<a href="#">한국정보기술학회</a> Korean Institute of Information Technology
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06644885">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06644885</a>
APA Style	이강혁, 김우제 (2016). 서포트 벡터 회귀를 이용한 24시간 앞의 태양광 발전량 예측. 한국정보기술학회논문지, 14(3), 175-183
이용정보 (Accessed)	건국대학교 202.30.38.*** 2021/04/15 16:50 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# 서포트 벡터 회귀를 이용한 24시간 앞의 태양광 발전량 예측

이강혁\*, 김우제\*\*

## Forecasting of 24\_hours Ahead Photovoltaic Power Output Using Support Vector Regression

Kanghuyk Lee\*, Woo-Je Kim\*\*

본 연구는 서울과학기술대학교 교내 연구비지원에 의하여 연구되었음.

### 요 약

태양광 발전은 기상에 많은 영향을 받기 때문에 발전량이 일정하지 않아 예측에 어려움이 있다. 24시간 앞의 태양광 발전량을 예측하는 것은 한국 전력거래소의 전력시스템과 스마트그리드 시스템에 안정적으로 연계되어 효율적인 에너지관리를 가능하게 할 수 있다. 본 연구에서 발전량 예측 모델을 수립하기 위하여, 2013년부터 2014년까지의 발전량, 기상실측, 기상예보 데이터를 수집하였다. 그리고 기계학습 알고리즘인 서포트 벡터 회귀를 이용하여 일사량 예측모델을 수립하였고, 예측된 일사량으로부터 최종적으로 발전량을 예측하는 모델을 수립하였다. 또한, 발전량 예측을 저해하는 요인들의 영향을 분석하여 본 24시간 후의 태양광 발전량을 예측하는 모델의 일반적인 성능을 도출하였다.

### Abstract

Since photovoltaic(PV) power system is subject to a lot of impacts on weather conditions especially solar insolation, it is difficult to predict photovoltaic power generation. Accurate forecast of 24-hour ahead PV power output makes stable link to Korea Power Exchange and Smart Grid power system and enable efficient energy management. In this study, actual historical power output and meteorological information data from 2013 to 2014 year were collected. First we establish the solar insolation forecasting model using Support Vector Regression(SVR) of machine learning algorithm, and second establish a SVR based model to predict PV power output from the predicted solar insolation. Also, by analyzing the effects of factors that inhibit the PV power prediction, the general performance of the model for predicting 24-hours ahead PV power output was derived.

### Keywords

photovoltaic power predicting, solar insolation predicting, support vector regression, machine learning,

\* 서울과학기술대학교 데이터사이언스과

\*\* 서울과학기술대학교 산업정보시스템공학과 교수 (교신저자)

· 접수 일: 2015년 11월 18일

· 수정완료일: 2016년 01월 08일

· 게재확정일: 2016년 01월 11일

· Received: Nov. 18, 2015, Revised: Jan. 08, 2016, Accepted: Jan. 11, 2016

· Corresponding Author: Woo-Je Kim

Dept. of DataScience, Seoultech, 232 Gongneung-ro, Nowon-gu, Seoul, 01811, Korea

Tel.: +82-2-970-6465, Email: jihjim@naver.com

## I. 서 론

현재 인류가 의존하고 있는 대부분의 에너지는 화석에너지이다. 하지만 증가하는 인구와 빠른 산업의 발전으로 인하여 많은 에너지 자원을 필요로 하게 되었고, 이는 화석에너지의 고갈로 이어지고 있다. 또한, 화석연료는 에너지로 전환되는 과정에서 이산화탄소 및 온실가스를 배출한다. 이는 지구온난화의 원인이 되고 극심한 환경 문제를 초래하게 된다. 이러한 이유로 최근 화석연료의 의존도를 줄이며 고갈될 염려 없이 풍부하고 자연환경에 영향을 미치지 않는 신재생에너지에 대한 중요성이 증가하고 있다[1].

신재생에너지 중에서도 특히, 태양광 발전이 각광 받고 있다. 태양광 발전은 태양의 빛 에너지를 변환시켜 전기를 생산하는 발전 기술로, 햇빛을 받으면 광전효과에 의해 전기를 발생하는 태양전지를 이용한 발전방식이다. 이러한 태양광 발전시스템은 태양전지(Solar Cell)로 구성된 모듈(Module)과 축전지 및 전력변환장치(Inverter)로 구성되어 있다. 태양광 발전의 장점은 에너지원이 청정·무한하고 필요한 장소에서 발전이 가능하고 유지보수가 용이한 점과 수명(20년 이상)이 길다는 것이다[2]. 그렇기 때문에 현재 다른 신재생에너지 보다도 태양광사업에 대한 투자가 전체 80%이상을 차지한다. 2015년 10월 기준 국내 신재생에너지 발전소의 개수는 태양광 발전소가 13,503개로 그 외에 모든 발전소를 합한 393개보다 30배 이상 많이 존재하는 것으로 나타난다. 반면, 태양광 발전의 단점으로는 초기투자비와 발전단가가 높고, 설치장소가 한정적이고, 에너지 밀도가 낮아 큰 설치면적이 필요하다는 점과 전력생산량이 지역별 기상상태(일사량)에 의존한다는 점에서 발전량 예측에 어려움이 존재한다.

본 연구에서는 태양광 발전량이 기상상태(일사량)에 의존하여 발전량 예측이 어렵다는 점에 초점을 두고 24시간 앞의 태양광 발전량을 시간별로 예측할 수 있는 모델을 수립하기 위한 연구를 진행하였다. 24시간 앞의 태양광 발전량을 시간별로 예측하는 것에 대한 필요성으로는 첫째, 한국 전력거래소의 전력계통에 안정적으로 연계되어, 효율적인 에

너지 관리를 가능하게 할 수 있다는 것이다. 한국 전력거래소에서는 발전 하루 전에 발전사들로부터 공급가능용량을 입찰 받아 발전계획을 수립하게 된다. 발전 당일에도 실시간 급전 지시를 통해 발전소들을 통제하며 발전을 운영하게 된다. 하지만 기상상태에 의존하는 태양광 발전량은 예측이 어렵기 때문에 전력거래소의 통제에서 벗어나 있다. 그러므로 24시간 앞의 태양광 발전량을 시간 별로 정확히 예측할 수 있다면, 전력거래소의 전력계통에 안정적으로 연계되어 효율적인 에너지 관리가 가능하게 될 것이다. 둘째로, 차세대 지능형 전력망인 스마트 그리드 시스템에서도 온실가스 등의 환경문제에 대응하기 위해 신재생에너지인 태양광 발전시스템을 결합하고 있다[3]. 이러한 스마트그리드 시스템에서도 안정적으로 연계되어 효율적인 에너지 관리를 가능하게 할 것이다. 따라서 본 연구에서 24시간 앞의 태양광 발전량을 시간별로 예측하는 발전량 예측모델을 수립하였다.

본 논문의 구성으로 2장에서는 태양광 발전량 예측에 영향을 미치는 요인 및 방법에 관련된 연구들을 소개하고 3장에서는 예측 절차 및 알고리즘에 대해 소개한다. 그리고 4장에서는 예측 모델을 수립하기 위해 수집된 데이터를 소개하고 실험 설계 및 자세한 실험 과정을 설명한다. 그리고 마지막 5장에서 결론을 맺는다.

## II. 관련 연구

### 2.1 태양광 발전량에 영향을 미치는 요인

태양광 발전에 영향을 미치는 요인을 크게 정리하면 기상요인과 지리요인 그리고 설비요인으로 구분할 수 있다. 첫 번째, 기상요인으로 태양광 발전은 근본적으로 태양의 빛 에너지를 변환시켜 전기를 생산하기 때문에 태양에서 오는 빛의 복사(輻射)인 일사량이 발전량에 가장 큰 영향을 미치는 요인이다. 또한 태양발전 시스템의 모듈을 구성하는 태양전지는 반도체이기 때문에 고온에서 효율이 떨어져 태양광 발전량에 영향을 미치게 된다. 그 외 자외선, 일조시간, 운량, 상대습도, 풍향, 풍속, 강우량,

강설량, 황사, 미세먼지 등이 태양광 발전량에 영향을 미치는 기상요인이다. 두 번째, 지리요인은 기상상태가 유사한 지역일 지라도 서로 다른 발전소에서 동일한 발전량이 생산되지 않는다. 이는 기상요인 외에도 지리적인 요인이 영향을 미치는 것을 의미하고 지리적 요인으로서는 위도, 경도 및 해발높이 등이 있다. 마지막 설비요인으로 고정형, 고정가변형, 추적형으로 구분되는 발전시스템의 태양전지 모듈 경사각의 차이에 따라 발전량이 다르게 나타난다. 또한 축전지 및 전력변환장치의 효율이 발전량에 영향을 미치게 된다[4].

하지만 본 연구는 24시간 앞의 태양광 발전량을 시간별로 예측하는 것이다. 24시간 앞의 태양광 발전량을 예측하기 위해서는 발전량에 영향을 미치는 요인들의 24시간 앞의 정보를 알 수 있어야 한다. 하지만 다음날의 정확한 기상정보를 아는 것은 사실상 불가능하다. 그러므로 본 연구에서는 기상청에서 제공되는 기상예보데이터를 활용하였다. 기상예보 정보로는 하늘상태, 강수확률, 습도, 온도, 풍향, 풍속이 있다. 그리고 기상실측 정보는 일사량과 운량을 활용하였고, 추가적으로 기존 연구에서 활용되지 않았던 자외선B 정보를 활용하였다. 지리적요인인 위도, 경도, 해발높이는, 본 연구에서 단일 발전소의 발전량을 예측하기 때문에 고정적인 요인으로 생략하였다. 하지만 기존 연구에서 활용되지 않은 태양고도 정보를 활용하였다. 태양고도는 월별로 차이를 보이기 때문에 월마다 차이를 보이는 발전량에 영향을 미칠 것으로 판단하였다. 마지막 설비요인도 지리적요인과 같이 고정적인 요인으로 생략하였다.

## 2.2 태양광 발전량 예측 연구

기존 태양광 발전량 예측 연구를 살펴보면, 먼저 월 단위의 발전량을 예측하는 장기예측 연구와 지역별 발전효율을 예측하는 연구가 존재한다[5][6]. 월 단위의 발전량 장기예측과 지역별 발전효율을 예측하는 것은, 태양광 발전소를 건설하는데 초기 투자비용이 많이 발생하기 때문에, 장기적으로 미루어 봤을 때 안정적인 발전 수익을 올릴 수 있는지를 판단하여 투자를 계획하기 위해서이다.

장기예측과는 반대로 단기예측은 태양광 발전소를 효율적으로 운영하거나 다른 전력계통과 안정적인 연계를 가능하게 하기 위해 필요하다. 단기예측에도 실시간으로 발전량을 예측하는 초단기 예측[7][8]과 다음날인 24시간 앞의 발전량을 예측하는 연구가 존재한다.

24시간 앞의 발전량을 예측하는 대부분의 연구는 기계학습 알고리즘인 인공신경망 모델과 SVR(Support Vector Regression) 모델을 이용하여 발전량을 예측하였다. 인공신경망 모델에서도 FFNN(Feed-Forward Neural Network), RBFNN(Radial Basis Function Neural Network), RNN(Recurrent Neural Network)를 이용하여 일사량을 예측하고 발전량 계산식을 통해서 발전량을 유도한 방법과 위 3가지 인공신경망을 앙상블 모델을 만들어 예측한 방법이 있다[9][10]. 그리고 날씨를 맑은 날, 흐린 날, 비오는 날로 분류하여 인공신경망 모델을 각각 적용한 방법이 있다[11]. SVR 모델을 이용한 24시간 앞의 발전량을 예측하는 방법으로는 기상상태를 구름, 안개, 맑음, 비 4가지로 나눠 각 기상상태별로 SVR 모델을 적용하여 발전량을 예측한 방법이 있다[12]. 하지만 위 논문들에서 예측모델을 수립하고 모델의 예측 성능을 평가하는데 사용되는 날이 5일, 12일, 60일 등 단기간의 샘플로 평가를 진행하였다. 이는 예측성능이 일반적으로 모든 날에 적용된다고 말하기 어려워 보인다. SVR 모델을 이용한 다른 연구에서는 1년 동안의 발전량을 예측하였다. 하지만 예측성능은 기존 논문들에 비해 떨어졌고, 그 원인에 대한 추가 분석은 하지 않았다[13].

본 연구도 24시간 앞의 태양광 발전량을 시간별 예측 모델을 수립하는 것으로 SVR 모델을 사용하였다. 기상예보 및 기상실측 데이터 그리고 파생변수를 생성하여 일사량 예측모델을 수립하고 예측된 일사량으로부터 최종적으로 발전량 예측 모델을 수립하였다. 그리고 모델평가는 2014년 동안의 매달 6일에서 말일까지 총 305일을 사용하여 일반적인 예측성능을 나타내었다. 또한, 예측성능을 저해하는 요인의 영향을 분석하였고 구름두께의 불확실성을 도입하여 하늘상태 예보에 따른 예측성능을 두 가지로 분류하였다.

### III. 예측 알고리즘

#### 3.1 예측 절차

본 연구에서 24시간 앞의 태양광 발전량 예측모델 수립은 크게 2단계를 통해 이루어진다. 첫 번째로 태양광 발전량에 가장 직접적인 영향을 미치는 일사량을 예측하는 모델을 수립한다. 현재 기상청에서 다음날의 일사량 정보를 제공하지 않기 때문에, 다음날의 태양광 발전량을 예측하기 위해서는 먼저 일사량 예측모델이 수립되어야 한다. 일사량 예측모델은 기상예보와 기상실측 데이터 그리고 태양 고도정보를 사용하여 수립하게 된다. 그리고 최종적으로 예측된 일사량과 기상예보, 기상실측, 태양고도 데이터로부터 발전량 예측모델을 수립하게 된다. 각 예측 모델을 수립하는 과정에서 기계학습 모델인 SVR을 이용하였고 격자 탐색(Grid Search)방법으로 모델에 적합한 파라미터를 찾아 모델을 최적화하였다. 동시에 후진제거(Backward Elimination)방법으로 모델성능에 영향을 미치지 않는 변수를 제거함으로써 모델을 간소화하여 최종 예측모델을 수립하였다.

#### 3.2 서포트 벡터 회귀(SVR)

SVR은 본래 패턴인식 분야에서 두 개의 범주를 분류하기 위한 학습 알고리즘인 SVM(Support Vector Machine)의 회귀문제로 확장된 방법이다. 본 연구에서 사용한 SVR은 Vapnik에 의해 제안된  $\epsilon$ -SVR(Vapnik, 1995)로, 기본아이디어는 학습데이터가 주어졌을 때, 모든 학습데이터의 목표변수가 주어진 편차  $\epsilon$ 안에 존재하도록 하는 가장 평평한 회귀함수인  $f(x)$ 를 찾는 것이다. 회귀함수  $f(x)$ 는 다음 식 (1)과 같고 에러를 최소화 하며 가장 평평한 함수를 찾는 최적화식은 다음 식 (2)와 제약조건 식 (3)으로 나타낼 수 있다.

$$f(x) = \langle w, x \rangle + b \quad (1)$$

$$\text{minimize} \quad \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (2)$$

$$\text{subject to} \quad y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \quad (3)$$

$$\langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0 \quad \text{for all } i = 1, \dots, l$$

식 (2)의  $\frac{1}{2} \|\omega\|^2$ 은 함수의 평평한 정도를 의미한다. 그리고  $C$ 는 0보다 큰 상수로 패널티(Penalty)를 의미하며, 함수  $f$ 의 평평한 정도와 편차  $\epsilon$ 보다 큰  $y_i$ 를 어느 정도 포함할 것인지를 균형(Trade-off)을 결정하는 모수이다.  $C$ 값을 크게 설정한다면, 편차  $\epsilon$ 을 벗어나는  $y_i$ 에 대한 패널티를 높게 주겠다는 의미가 된다. 즉, 편차  $\epsilon$ 를 벗어나는  $y_i$ 를 적게 하겠다는 의미이다. 결과적으로 SVR모형은 구불구불한 함수로 학습데이터에 과대적합(Overfitting)될 것이다. 그러므로 좋은 성능을 보이는 일반화된 SVR 모형으로 적합하기 위해서 적절한  $C$ 값을 선택하는 것은 중요한 문제이다. SVR 함수를 찾기 위한 최적화 문제는 식 (2)와 식 (3)으로부터 라그랑지 함수를 통해 해결할 수 있다.

SVR의 강점은 비선형 문제에서도 적용됨에 있다. 회귀함수  $f(x)$ 를 구하기 위한 라그랑지안 문제를 풀기위해 오직 두 벡터의 내적 값만이 필요하다. 이는 커널함수(Kernel Function)에 의해 선형으로 표현할 수 있는 고차원에서의 두 벡터의 내적 값을 구할 수 있게 된다. 커널함수에는 RBF(Radial Basis Function), Polynomial, Sigmoidal 커널이 존재한다 [14][15].

본 연구에서는 이론적으로 무한대의 차원으로 확장이 가능한 가우시안 함수인 RBF 커널을 사용하였다. RBF 커널 함수는 식 (6)과 같다.

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (6)$$

그러므로 본 연구에서 SVR모델을 최적화하기 위해 필요한 파라미터는  $\epsilon$ ,  $C$ ,  $\sigma$  3가지를 필요하게 된다.

### IV. 발전량 예측 실험

#### 4.1 데이터 설명

본 연구에서 태양광 발전량을 예측하기 위해 수집한 데이터로는 첫 번째로 태양광 발전량 데이터이다. 3MW 규모의 영암태양광 발전소로부터 2013년부터 2014년까지의 시간당 발전량(kWh) 데이터를 수집하였다. 두 번째 데이터는 기상실측 데이터로 운량, 일사량, 자외선B 정보가 있다. 세 번째 데이터는 동네예보 데이터로 하늘상태, 강수확률, 온도, 상대습도, 풍향, 풍속 예보정보를 수집하였다. 두 기상데이터는 기상청으로부터 수집하였다. 수집 위치는 실측기상이 측정되는 영암태양광발전소로부터 12km 떨어진 목포기상대이다. 운량과 기상예보데이터는 3시간 단위로 구성되어있고 자외선과 일사량은 1시간 단위로 구성되어있다. 마지막 데이터는 Keisan Online Calculator로부터 목포기상대의 시간당 태양의 고도 정보를 수집하였다.

#### 4.2 실험 설계

먼저 예측 모델을 수립하기 위해 2013년부터 2014년도의 데이터를 학습데이터와 검증데이터 그리고 평가데이터로 설정하였다. 학습데이터는 2013년 1월 8일부터 2013년 12월 31일로 설정하였고, 모델에 적합한 파라미터를 찾기 위한 검증데이터는 2014년 매달 1일부터 5일로 설정하였다. 마지막으로 모델 평가데이터는 검증데이터를 제외한 2014년 매달 6일부터 말일까지로 설정하였다.

예측 모델 평가에 사용할 지표는 MAE(Mean Absolute Error)값으로 절대 평균 오차를 의미한다. 즉 실측값으로 부터의 ±오차의 평균을 나타낸다고 볼 수 있다. 그리고 MAPE(Mean Absolute Percentage Error)지표를 사용하였다. MAPE는 평균 절대 오차율로 실측값으로부터 벗어난 예측값에 대한 비율의 평균을 의미한다. MAE와 MAPE를 산출하는 공식은 다음과 같다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (7)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - f_i}{y_i} \right| \quad (8)$$

식 (10), (11)에서  $y_i$ 는 실측값을 의미하고  $f_i$ 는 예측값을 의미한다. 그리고  $n$ 은 데이터의 수를 의미한다. 본 연구에서 MAPE값은 하루단위 총 발전량을 기준으로 하였다.

태양광 발전량은 24시간을 주기로 태양이 뜨는 시각부터 지는 시간까지 시계열 분포를 나타낸다. 그래서 본 연구에서는 실측데이터를 24시간 주기를 갖는 7일 동안의 자기회귀(AR(7) - Autoregressive Order by 7)를 파생변수로 추가하여 태양광 발전량을 예측하는 설명변수로 사용하였다.

마지막으로 24시간 앞의 예측 모델을 수립하기 위한 설명변수들의 단위가 다르기 때문에 0-1정규화(0-1 Normalization)를 시켜, 모든 변수들이 0에서 1사이의 값을 갖도록 하였다. 0-1정규화는 다음과 같이 산출할 수 있다.

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (9)$$

#### 4.3 일사량 예측

먼저 발전량을 예측하기 전에 발전량에 가장 큰 영향을 미치는 일사량 예측 모델을 수립하였다. 일사량을 예측하기 위해 사용되는 설명변수인 기상예보정보는 3시간 단위로 예보되어지고 있다. 하지만 일사량 값은 1시간 단위로 구성되어있다. 그렇기 때문에 3시간 단위의 설명변수를 1시간 단위의 목표 변수에 매핑할 필요가 있다. 본 연구에서는 선형 보간법을 사용하여 3시간 단위 변수를 1시간 단위 변수에 매핑하였다. 보간법은 두 변수 사이에 있는 값을 평균하여 추정하는 방법으로 다음 식 (10)에 의해 산출할 수 있다.

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_2)}{x_1 - x_2} (x - x_0) \quad (10)$$

일반적으로 구름이 많은 날에는 일사량이 적고 맑은 날에는 일사량이 높듯이, 운량은 일사량에 영

향을 미치는 변수이다. 하지만 0에서 10사이로 측정되는 실측 운량과 달리 하늘상태 예보는 1에서 4로 예보가 된다. 그러므로 본 연구에서는 더 정확한 운량 값을 사용하기 위해 하늘상태를 설명변수로 운량을 종속변수로 하여 2013년을 학습하고 2014년의 운량을 예측하였다. 즉 일사량 예측모델에서 학습은 실측운량으로 학습하지만 예측에서는 운량예측 정보를 사용하는 것이다.

일사량과 가장 연관 깊은 변수는 자외선 값이다. 하지만 운량과 같이 자외선 값은 실측값이기 때문에 24시간 앞의 일사량을 예측하는 모델을 수립하는데 학습데이터에 그대로 사용할 수 없다. 그래서 과거 하늘상태 예보에 따른 계절 및 시간 별 자외선 값의 평균과 1~3사분위수로 다음날의 자외선 값을 예측하여 변수로 활용하였다. 예를 들어 2014년 1월 1일 12시에 하늘상태가 1로 예보되었다면, 이때의 자외선 값은 과거 2013년 계절이 겨울(12~2월)인 12시 중 하늘상태가 1로 예보된 날들의 자외선 값의 평균과 1~3사분위수를 사용하는 것이다. 자외선과 마찬가지로 일사량도 하늘상태 예보에 따른 과거 계절 및 시간 별 평균과 1~3사분위수를 파생변수로 추가하여 사용하였다.

표 1. 일사량 예측모델을 수립하기 위한 설명변수  
Table 1. Independent variables for predicting solar insolation

분류	설명변수 명
기상예보	하늘상태, 강수확률, 상대습도, 온도, 풍향, 풍속
파생변수	일사량 AR(7), 운량 실측(학습) 및 예측(예측), 2013년 자외선 평균 및 1~3사분위, 2013년 일사량 평균 및 1~3사분위
기타변수	태양고도

표 2. 일사량 예측모델 파라미터 및 변수 선택 결과  
Table 2. The result of selection of parameters and variables in solar insolation predicting model

제거변수(누적)	C	$\epsilon$	$\sigma$	MAE
일사량2사분위	20	0.001	0.01	0.198
일사량1사분위	20	0.005	0.01	0.198
일사량3사분위	20	0.005	0.01	0.198
풍속	30	0.005	0.01	0.199

표 1은 일사량예측에 사용되는 설명변수들을 나타내고, 2013년 동안의 학습데이터로 SVR 모델을 적합하였다. 그리고 검증데이터를 활용하여 모델평가 지표인 MAE를 기준으로 파라미터  $C$ ,  $\epsilon$ ,  $\sigma$ 를 격자탐색 방법으로 모델에 적합한 값을 찾았다. 동시에 모델 성능에 영향을 미치지 않는 변수들을 후진제거 방법으로 제거하여 모델을 간소화함으로써, 최종적인 일사량 예측 SVR모델을 수립하였다. 표 2는 모델의 파라미터 및 변수 선택 결과를 나타낸다. 최종 선택된 변수는 일사량의 1~3사분위수를 제외한 나머지 모든 변수들로 나타났다.

최종 수립된 일사량 예측 모델로부터 테스트 데이터의 일사량을 예측 평가한 결과 MAE 값은 0.205, MAPE값은 0.248로 나타났다.

#### 4.4 발전량 예측

4.3절로부터 예측된 일사량으로부터 최종적인 발전량 예측모델을 수립하였다. 먼저 일사량과 마찬가지로 3시간 단위의 예보 및 운량 변수를 보간법으로 1시간 단위의 발전량 변수에 매핑하였다. 또한, 자외선B, 발전량에 대한 과거 하늘상태 예보에 따른 계절 및 시간 별 평균 및 1~3사분위수를 파생변수로 생성하여 설명변수에 사용하였다. 최종적으로 발전량 예측에 사용된 설명변수들은 표 3과 같다. 일사량 예측에서 사용된 설명변수들에 더해 일사량 변수를 추가하였다. 단, 운량과 같이 예측 모델을 학습할 때는 일사량 실측값을 사용하고 예측할 때는 예측값을 사용하였다.

표 3. 태양광 발전량 예측모델을 수립하기 위한 설명변수  
Table 3. Independent variables for predicting PV power output

분류	설명변수 명
기상예보	하늘상태, 강수확률, 상대습도, 온도, 풍향, 풍속
파생변수	발전량 AR(7), 일사량 실측(학습) 및 예측(예측) 운량 실측(학습) 및 예측(예측), 2013년 자외선 평균 및 1~3사분위, 2013년 발전량 평균 및 1~3사분위
기타변수	태양고도

그리고 2013년 동안의 학습데이터로 모형을 적합하였고, 검증데이터를 활용하여 SVR 모형의 적합한 파라미터를 찾고 변수 후진제거 방법으로 모형을 간소화함으로써, 최종적으로 발전량 예측모형을 수립하였다. 표 4는 발전량 예측모형의 파라미터 및 변수 선택 결과를 나타낸다. 운량, 풍속, 자외선2사분위, 하늘상태, 발전량평균 변수가 제거되었고 최종적으로 선택된 변수는 일사량, 강수확률, 온도, 습도, 풍향, AR(7), 자외선 평균 및 1,3분위, 발전량 1~3사분위수로 나타났다.

표 4. 태양광 발전량 예측모형 파라미터 및 변수 선택결과  
Table 4. The result of selection of parameters and variables in PV power output predicting model

제거변수(누적)	C	$\epsilon$	$\sigma$	MAE
운량	20	0.010	0.005	188.1
풍속	25	0.010	0.005	184.6
자외선2사분위	15	0.005	0.005	184.3
하늘상태	25	0.010	0.005	184.1
발전량평균	25	0.010	0.005	184.1
발전량1사분위	25	0.005	0.005	184.3

최종 수립된 발전량 예측 모델로부터 테스트 데이터의 발전량을 예측 평가한 결과 MAE값은 193.8, MAPE값은 0.288로 나타났다.

#### 4.5 예측성능을 저해하는 요인 분석

24시간 앞의 발전량을 예측하는 것은 다음날의 기상상태에 많은 영향을 받는다. 그러므로 본 연구에서는 기상예보의 정확성이 예측모델의 성능을 높일 수 있다. 또한, 본 연구에서는 발전시스템의 오류를 고려하지 않았다. 이번 4.5절에서는 발전시스템의 오류가 발생했을 가능성이 높은 날과 하늘상태예보를 기준으로 예보가 틀린 날을 추정하여, 발전량 예측모델의 성능에 미치는 영향을 분석하였다.

먼저 태양광 발전시스템의 오류는 실측일사량 및 실측기상정보(자외선, 운량, 온도, 습도, 풍향, 풍속, 태양고도)를 설명변수로 활용하여 발전량을 예측하였다. 그리고 실측 일사량으로 발전량을 예측했음에도 불구하고 MAPE값이 40%이상으로 나타난 일들은 발전시스템의 오류가 발생했을 것이라고 추정하였다. 추정결과 총 7일(04/28, 06/06, 06/17, 07/06, 08/07, 11/24, 12/17)이 추정되었다.

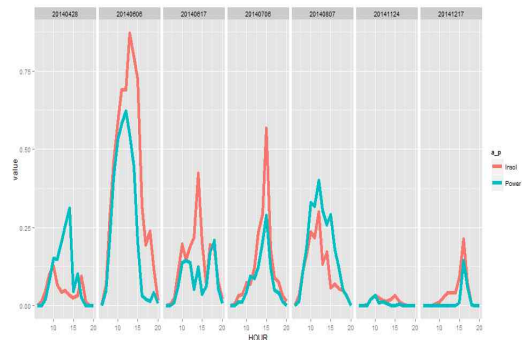


그림 1. 시스템에러 추정일의 발전량과 일사량 비교  
Fig. 1. Comparing the day of estimated internal error between power output and insolation

그림 1은 발전시스템에러 추정일에 대한 일사량과 발전량의 비교 그래프이다. 일사량과 발전량의 값을 0-1사이의 값으로 정규화 하여 선그래프로 나타내었고 실제 발전량과 일사량이 다른 것을 관찰할 수 있다.

4.4절의 발전량 예측모델 평가에서 시스템에러 추정일을 제거하였을 때 MAPE값은 0.267로 나타났다. 그러므로 시스템에러에 대한 요인이 예측성능(MAPE)에 미치는 영향은 약 2%로 보인다.

다음은 기상예보가 틀린 날은 하늘상태예보를 기준으로 태양광이 강해지는 시점인 10시부터 16시까지의 실제운량과의 오차가 2이상인 날을 기상예보가 틀린 날이라고 가정하였다. 이때, 하늘상태예보와 운량은 0-1사이의 정규화된 값을 사용하였다. 테스트 데이터의 기간인 305일 중 54일이 예보가 틀린 날로 나타났다. 그리고 예보가 틀린 날과 시스템에러 추정일을 제거하였을 때 MAPE값은 0.225로 MAE값은 167.1로 나타났다. 그러므로 예측할 수 없는 틀린 기상예보에 대한 요인이 예측성능에 미치는 영향은 약 4%로 보인다.

#### 4.6 구름두께의 불확실성 고려

운량은 일사량에 영향을 미치는 변수이다. 하지만 구름의 두께에 따라 일사량에 차이가 발생한다. 구름의 두께가 얇을 경우 태양광이 투과되어 일사량이 많이 발생한다. 하지만 구름의 두꺼울 경우는 투과되는 태양광이 약해 일사량이 적게 발생하게 된다.

하지만 본 연구에서 사용한 운량은 단지 하늘을



덮고 있는 구름의 양을 의미하는 값이다. 그러므로 구름의 두께가 고려되지 않아 운량이 높은 날에, 일사량이 많이 발생할 수 있고 적게 발생할 수도 있어, 일사량의 편차가 크다. 하지만 구름의 두께정보는 확인할 수 없고 운형으로 간접적으로 유추할 수 있지만 다음날의 운형까지 예측하는 것은 사실상 어렵기 때문에 본 연구에서는 구름의 두께요인은 고려하지 않았다. 그러므로 다음날의 하늘상태 예보에 따라 예측모델을 두 개로 분리하였다.

6시부터 20시까지의 운량이 합이 13이상인 날과 이하인 날을 구분하였다. 즉, 하루 총 15의 운량 중 13이상인 날에 구름 구개의 불확실성을 적용하였다. 그리고 운량 합이 13이상인 날과 13이하인 날을 구분하여 MAPE값과 MAE값은 나타냈을 때, 다음 표 5와 같다.

표 5. 구름의 두께를 고려한 모델 평가  
Table 5. The model evaluation considering of cloud thickness

	운량 <= 13	운량 > 13
MAE	161.1	185.5
MAPE	14.9%	45.9%

## V. 결 론

본 연구는 태양광 발전시스템이 전력거래소와 스마트그리드 전력시스템에 안정적인 연계 및 효율적인 에너지관리를 가능하게하기 위해 24시간 앞의 태양광 발전량을 시간별로 예측하는 모델을 수립하였다.

발전량을 예측하기 위해 기상청으로부터 수집한 기상예보 및 실측정보와 태양고도 그리고 파생변수를 생성하여 먼저 일사량을 예측하였고, 예측된 일사량으로부터 최종적으로 발전량을 예측하였다. 예측에 사용된 알고리즘은 기계학습 방법인 SVR를 사용하였다. 그리고 예측 모델을 최적화하기 위해 격자 탐색 방법으로 모델에 적합한 파라미터를 선택하였고 후진제거 방법으로 변수를 제거하여 모델을 간소화함으로써, 최종 발전량 예측 모델을 수립하였다.

2013년 데이터를 학습하여 예측 모델을 수립하였고 2014년 매월 6일부터 말일까지를 예측 모델을 평가하는데 사용하였다. 예측 결과 MAE값은 193.8,

MAPE값은 0.288로 나타났다. 그리고 예측성능을 저해하는 요인의 영향을 분석하였고, 구름 두께의 불확실성을 고려하여 하늘상태예보 값을 기준으로 모델을 분리하였다. 결과로 하늘상태 예보 운량이 13이하인 날에 MAPE값은 14.9%로 나타났고 운량이 13초과인 날에는 MAPE값이 45.9%로 나타났다. 다음날의 하늘상태가 맑은 날에는 태양광 발전량 예측이 잘 되었지만, 구름이 많은 날에는 구름의 두께라는 불확실 요소로 인해 예측이 잘 안 되는 것으로 나타났다.

향후 연구로, 더 좋은 24시간 후의 태양광 발전량 예측모델을 수립하기 위해서 첫 번째로, 본 연구에서는 학습데이터의 기간이 약 1년으로 다양한 기상 정보를 반영하기는 부족한 양이었다. 충분한 학습데이터를 확보할 필요가 있다고 생각한다. 두 번째로, 같은 운량임에도 일사량의 편차를 발생시키는 다음날의 구름의 두께에 대해 고려할 수 있는 새로운 요인에 대한 연구가 필요하다고 생각한다.

## References

- [1] Kim Geungi, "Prospects of Alternative Energy Development to Solve the Fossil Fuel Exhaustion and Global Warming", Hankuk University of Foreign Studies, international business, Feb. 2008.
- [2] Myeong-Hwan Seo, Gi-seok Kim, and Sun-Hyung Kim, "A Development of the Solar Position Algorithm for Improving the Efficiency of Photovoltaic Power Generation", Proceedings of KIIT Summer Conference, pp. 46-51, June 2009.
- [3] SungMin Rue, "Study on Smart Grid Trend and Policy Enforcement Way", Journal of KIIT, Vol. 12, No. 7, pp. 163-177, July 2014.
- [4] Cha Wangcheol, "A Study on the Prediction of the Annual Power Generation through the Analysis on Factors Affecting Photovoltaic Power Generation", Dept. of Electrical Engineering Graduate School of Soongsil University, June 2015.
- [5] Wang-Cheol Cha, Joung-ho Park, Uk-rae Cho, and Jae-Chul Kim, "Design of Generation Efficiency Fuzzy Prediction Model using Solar Power Element Data", The Transactions of the Korean

- Institute of Electrical Engineers, Vol. 63, No. 10, pp. 1423-1427, Oct. 2014.
- [6] Yu-seong Noh, Yong-taek Lim, Woo-sik Yoo, and Jae-gon Kim, "A study on the estimating power generation for a PV solar plant using weather forecast", Korea Industrial And Systems Engineering, Vol. 2013 No. 1, pp. 161-166, 2013.
- [7] Jae-Ju Song, Yoon-Su Jeong, and Sang-Ho Lee, "Analysis of prediction model for solar power generation", Journal of Digital Convergence, Vol. 12 No. 3, pp. 243-248, Mar. 2014.
- [8] Kim Kwang-Deuk, "The Development of the Short-Term Predict Model for Solar Power Generation", The Korea Solar Energy Society, Vol. 33, No. 6, pp. 62-69, Dec. 2013.
- [9] Yona A, Senjyu T, Saber A. Y, Funabashi T, Sekine H, and Chul-Hwan Kim, "Application of Neural Network to One-Day-Ahead 24 hours Generating Power Forecasting for Photovoltaic System", Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on, pp. 1-6, Nov. 2007.
- [10] A. Chaouachi, R. M. Kamel, R. Ichikawa, H. Hayashi, and K. Nagasaka, "Neural Network Ensemble based Solar Power Generation Short-Term Forecasting", World Academy of Science, Engineering & Technology, No. 30, No. 6, pp. 1258-1263, June 2009.
- [11] Changsong Chen, Shanxu Duan, Tao Cai, and Bangyin Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network", Solar Energy, Vol. 85, No. 11, pp. 2856-2870, Nov. 2011.
- [12] Jie Shi, Wei-Jen Lee, Yongqian Liu, Yongping Yang, and Wang Peng, "Forecasting power output of photovoltaic system based on weather classification and support vector machine", Industry Applications Society Annual Meeting (IAS), Vol. 48, No. 3, pp. 1064-1069, May-June 2012.
- [13] Joao Gari da Silva Fonseca Jr., Takashi Oozeki, Takumi Takashima, Gentarou Koshimizu, Yoshihisa Uchida, and Kazuhiko Ogimoto, "Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan", Progress in Photovoltaics, Vol. 20, No. 7, pp. 874-882, Nov. 2012.
- [14] Alex J. Smola, Bernhard Schölkopf, "A Tutorial on Support Vector Regression", Statistics and Computing, Vol. 14, No. 3, pp. 199-222, Aug. 2004.
- [15] Klaus-Robert Müller, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, and Bernhard Schölkopf, "An Introduction to Kernel-Based Learning Algorithms", *IEEE transactions on neural networks*, Vol. 12, No. 2, pp. 181-201, Mar. 2001.

## 저자소개

### 이 강 혁 (Kanghyuk Lee)



2013년 2월 : 명지대학교 수학과 (이학사)  
 2014년 3월 ~ 현재 : 서울과학기술대학교 데이터사이언스학과 (석사과정)  
 관심분야 : Data- Mining, Machine-Learning, Recommender System

### 김 우 제 (Woo-Je Kim)



1986년 : 서울대학교 산업공학과 (학사)  
 1988년 : 서울대학교 산업공학과 (석사)  
 1994년 : 서울대학교 산업공학과 (박사)  
 1988년 ~ 1991년 : 동양경제연구소 연구원

1999년 ~ 2001년 : University of Michigan Visiting Scholar

2003년 ~ 현재 : 서울과학기술대학교 글로벌융합산업공학과 교수

관심분야 : IT서비스, 소프트웨어 공학, 최적화, 스마트그리드