

# 현재 기상 정보의 이동 평균을 사용한 태양광 발전량 예측

이 현 진<sup>†</sup>

## Use of the Moving Average of the Current Weather Data for the Solar Power Generation Amount Prediction

Hyunjin Lee<sup>†</sup>

### ABSTRACT

Recently, solar power generation shows the significant growth in the renewable energy field. Using the short-term prediction, it is possible to control the electric power demand and the power generation plan of the auxiliary device. However, a short-term prediction can be used when you know the weather forecast. If it is not possible to use the weather forecast information because of disconnection of network at the island and the mountains or for security reasons, the accuracy of prediction is not good. Therefore, in this paper, we proposed a system capable of short-term prediction of solar power generation amount by using only the weather information that has been collected by oneself. We used temperature, humidity and insolation as weather information. We have applied a moving average to each information because they had a characteristic of time series. It was composed of min, max and average of each information, differences of mutual information and gradient of it. An artificial neural network, SVM and RBF Network model was used for the prediction algorithm and they were combined by Ensemble method. The results of this suggest that using a moving average during pre-processing and ensemble prediction models will maximize prediction accuracy.

**Key words:** Prediction of Solar Power Generation, Neural Networks, Moving Average, Ensemble

### 1. 서 론

신재생 에너지 분야에서 태양광 발전은 가장 빠르게 수요와 공급이 증가하고 있는 분야이다. 2015년도에 전 세계적으로 신규로 태양광 발전량이 50GW 도입되었으며, 총 태양광 발전량은 231GW로 전년대비 28.1% 증가하여 가장 두드러진 성장세를 보이고 있다[1]. 태양광 발전의 필요성 및 공급이 증가하고 있지만, 관련 기술 및 설비 구축이 어렵고, 관련 인력의 부족의 문제를 가지고 있다. 도서 및 산간 지형과 같은 지역은 기존 발전소의 전력을 쉽게

사용할 수 없기 때문에 필요성이 증가하고 있다. 하지만, 접근성 때문에 설비 구축과 유지 보수에 대한 문제를 가지고 있다. 이러한 지역에서는 지형의 특수성과 설비 용량의 한계 때문에 태양광 발전만으로 발전 수요를 충족시키지 못하기 때문에 태양광 발전과 내연 기관 발전을 병행하고 있다.

태양광 발전은 특성상 기상 조건에 따른 발전량 변동이 심하고, 단일 데이터를 이용한 단순 비교만으로는 발전 상태를 정확하게 예측하기 어렵다. 태양광 발전과 내연 기관 발전을 병행하는 환경에서는 태양광 발전량에 대한 예측이 정확해야 내연 기관

※ Corresponding Author : Hyunjin Lee, Address: (03132) #321, Jongno Biz-well, 23 Samil-daero 30-gil, Jongno-gu, Seoul, Korea, TEL : +82-2-708-7863, FAX : +82-2-708-7749, E-mail : hjlee@mail.kcu.ac

Receipt date : Jul. 25, 2016, Approval date : Aug. 2, 2016

<sup>†</sup> Dept. of Computer Science & Software, Korea Soongsil Cyber University

발전의 운용 계획 및 연료에 대한 공급 계획을 정확하게 세울 수 있다.

태양광 발전량은 설비의 노후화, 기상 환경 등 다양한 요소에 영향을 받기 때문에 예측이 어려운 분야이다. 이승민 등은 입력은 기상 데이터로 사용하고, 출력은 단위 발전량 값으로 하여 오류 역전파 알고리즘 (Error Backpropagation)으로 태양광 패널의 결함을 검출하는 연구를 수행하였다[2]. 송재주 등은 태양광발전소의 입지 선정에 위해 신경망을 이용하여 월별 일사량을 예측한 후 예측된 일사량으로 발전량을 예측하는 연구를 수행하였다[3]. Sharma 등은 SVM을 사용하여 기상 데이터를 이용한 발전량 분류기를 만들어 태양광 발전량 예측에 사용하였으며, SVM에서 사용하는 벡터를 고차원으로 만들어 단일 차원 벡터에 비해 약 27%의 예측 성능 향상을 하였다[4]. 이창성 등은 데이터 전체를 사용하여 퍼지 클러스터링 기법으로 서브그룹으로 분할하는 ELM(Extreme Learning Machine) 기반의 모듈라 네트워크를 이용한 태양광 발전 시스템 모델링 기법을 제안하여 기존의 ELM기법에 비해 향상된 결과를 보였다[5]. 박찬곤은 마이크로 인버터를 적용한 태양광 발전시스템의 노후예측판단 시스템을 제안하였다[6].

본 논문에서는 보안상의 이슈로 네트워크가 제한되는 환경이나 도서 및 산간 지역에 상시 네트워크를 연결하여 기상 예보 정보를 자동으로 수집하는 시스템이 구축되지 않은 기존 설비에 적용하기 위하여 센서에 의해 수집할 수 있는 데이터를 사용하여 태양광 발전 예측을 할 수 있는 시스템을 제안하였다. 기상 데이터 중 일사량, 강수량, 온도, 습도의 네 가지를 태양광 발전량 예측에 대한 변수를 선정하였다. 미래의 발전량 예측을 위해 기상 데이터의 시계열 특성을 반영하여 한 시점이 아닌 일정 구간에 대한 기상 데이터의 이동 평균(Moving Average)을 활용하는 방법을 제안하였다. 신경회로망, SVM, RBF 네트워크로 48시간의 단기 예측을 수행하기 위하여 각 알고리즘별로 1시간 이후를 예측하는 모델부터 48시간 이후의 발전량을 예측하는 모델까지 48개의 모델을 생성하여 앙상블 기법 (Ensemble Method)과 구간별 계수를 적용하여 최적의 예측 모델을 생성하였다.

본 논문의 구성은 다음과 같이 구성된다. 2장에서

신경회로망, SVM, RBF 네트워크와 같은 예측 방법에 대해 살펴본다. 3장에는 제안하는 방법의 구성과 입력특징과 데이터의 전처리방법, 앙상블기법에 대해 설명한다. 4장에서는 실 데이터에 대한 실험을 통해 알고리즘을 평가하고, 5장에서는 결론을 맺는다.

## 2. 예측 방법

### 2.1 신경회로망(Neural Networks)

신경회로망은 인간의 뇌의 구조와 기능을 모델화하여 학습하는 알고리즘으로 패턴인식, 데이터마이닝, 제어 등 다양한 분야에 응용되고 있다[7]. 태양광 발전 예측을 위하여 다층 퍼셉트론 (Multi-Layer Perceptron) 신경회로망을 입력층, 은닉층, 출력층의 3개의 층으로 구성하였다. 가중치의 학습은 오류 역전파 (Error Backpropagation) 알고리즘을 사용하였다. 다층 퍼셉트론의 출력  $f(x)$ 는 식 (1)과 같다.

$$f(x) = K(\sum_i w_i x_i) \quad (1)$$

여기서,  $K()$ 는 활성화 함수 (activation function)이다. 활성화함수는 식 (2)와 같이 Logistic 함수를 사용하였다.

$$K = \frac{1}{1 + e^{-x}} \quad (2)$$

$x_i$ 는 입력 데이터이고,  $w_i$ 는 신경회로망 노드의 가중치 (weight) 값이다.

본 논문에서는 입력층의 노드는 34개, 은닉층의 노드는 입력노드의 1.5배인 50개, 출력층의 노드는 발전량을 예측하기 위해 1개를 사용했다. 48시간의 발전량을 예측하기 위하여 이와 같은 신경망을 48개를 만들었고, 신경망의 출력은 각각 1시간 이후, 2시간 이후에서 48시간 이후의 발전량을 예측하게 구성하였다. 학습된 신경망은 (Fig. 1)과 같다.

### 2.2 Support Vector Machine (SVM)

SVM은 기계 학습 분야의 알고리즘으로 신경정보학, 문자 인식, 이미지 인식, 로봇틱스 등 다양한 분야에서 사용되고 있다[8]. SVM은 커널 (kernel) 함수를 통하여 비선형 문제를 선형 방법론으로 해결할 수 있지만, 사용자가 커널을 선택한 결과에 따라 성능차이가 많이 발생된다. SVM은 일반적으로 이

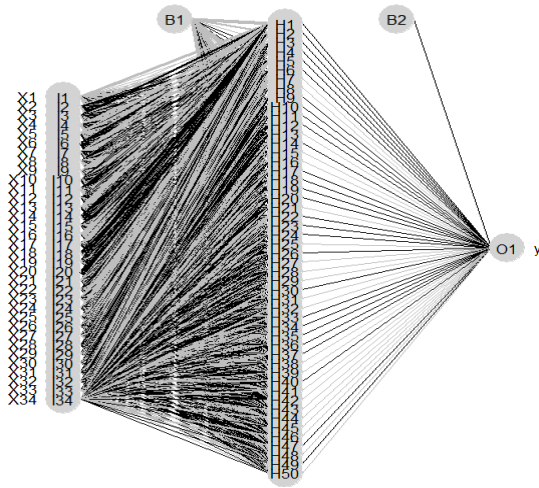


Fig. 1. Artificial neural networks model.

진 분류에 사용된다. 데이터를 구성하는 두 클래스를 가장 잘 구별할 수 있는 초평면 (hyper plane)을 찾는 것을 목표로 한다. 커널을 사용해서 선형 공간을 비선형 공간으로 변화시켜서 선형 분류가 불가능한 데이터를 선형 분류가 가능하게 한다. 일반적인 커널 함수로는 선형 커널, 가우시안 커널, 다항식 커널 등이 있다. SVM을 예측에 사용할 때는 커널 함수를 이용하여 데이터의 공간을 비선형 공간으로 변화시키고, 이 공간에서 회귀 분석을 사용하여 예측을 수행한다[9]. 본 논문에서는 다항식(Polynomial) 커널을 사용하였고, 신경회로망과 동일한 입력과 출력을 사용하였다.

### 2.3 RBF 네트워크 (Radial Basis Function Network)

RBF 네트워크는 다층 퍼셉트론과 유사하게 분류나 예측 문제를 해결하기 위하여 사용된다. RBF 네트워크는 다층 퍼셉트론과 달리 입력값에 대한 목표값을 대응시켜 학습하는 지도학습과 목표값 없이 입력값만으로 경쟁을 통해 학습하는 자율학습의 두 가지를 모두 할 수 있다[10]. 기본적인 구조는 다층 퍼셉트론과 유사하지만, 은닉층은 단 한 층이고, 입력값을 그대로 받아들인다. RBF 네트워크는 은닉층의 각 노드에 중심점이라는 벡터를 가지고 있으며, 이 중심점을 비교사 학습으로 학습한다.

RBF는 그 함수의 중심에서 가장 큰 값을 갖고, 중심에서 멀어질수록 0에 가까워지는 함수이다. 일반적으로  $N$ 차원의 실수 입력 벡터  $x$ 에 대해  $K$ 개의

RBF로 구성된 중간 은닉층을 가지는 RBF는 식 (3)과 같이 정의된다.

$$f(x) = \sum_{k=1}^K w_k \Phi(\|x - C_k\|) \quad (3)$$

여기서,  $C_k$ 는 RBF의 중심이고,  $w_k$ 는  $k$ 번째 RBF 뉴런의 출력과 출력 뉴런과의 연결 강도,  $\Phi$ 는 RBF로 보통 가우시안 함수를 사용한다. RBF 네트워크는 계산이 단순하고, 수행시간이 빠른 장점이 있지만, 중심이 잘못되거나 은닉 노드의 수가 잘못되면 수렴하지 않을 가능성이 높아지고, 입력 자료에 민감하다.

## 3. 제안하는 예측 모델

### 3.1 제안하는 시스템의 구성

제안하는 예측모델의 구성은 (Fig. 2)와 같다. 기상 데이터 중 일사량, 강수량, 온도, 습도의 네 가지를 태양광 발전량 예측에 대한 변수로 사용하였다. 미래의 발전량을 예측을 위해 기상 데이터의 시계열 특성을 반영하여 한 시점의 데이터가 아닌 일정 구간의 데이터에 대한 이동 평균을 활용하는 방법을 제안하였다. 예측은 신경회로망, SVM, RBF 네트워크로 48시간의 단기 예측을 수행하기 위한 각각의 48개 모델을 생성하였다. 예측 오차는 평균 제곱근 오차 (Root Mean Squared Error; RMSE)를 사용하

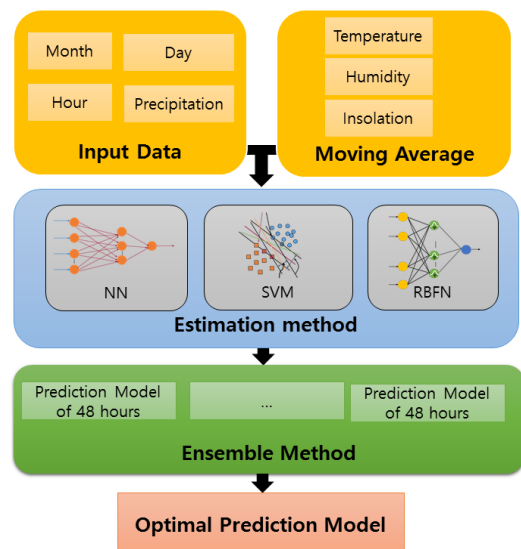


Fig. 2. Proposed Method of Solar Power Generation.

여 알고리즘을 평가하였다. 최적의 예측 모델을 만들기 위해 이 알고리즘에 의한 모델들을 결합한 앙상블 기법 (Ensemble Method)으로 최적의 모델을 생성 하였다. 전 구간(48시간)에 걸쳐서 동일한 계수를 설정하는 것이 아니라, 구간을 나누어 다른 계수를 적용하여 예측 성능을 높였다.

### 3.2 입력특징과 데이터 전처리

태양광 발전에 영향을 미치는 요소는 다양하지만, 그 중 일사량, 강수량, 기온, 전운량, 습도, 안개가 가장 큰 영향을 미친다[11]. 이러한 요소 중에서 기상청의 정보로부터가 아닌 시스템 상에서 직접 얻을 수 있는 데이터인 일사량, 강수량, 기온, 습도를 태양광 발전량 예측을 위한 입력 변수로 사용하였다. 기상 데이터와 태양광 발전량 데이터는 단위 및 분포가 모두 달라 학습을 위해서는 정규화 변환이 필요하다. 정규화는 식 (4)를 이용하여 데이터의 값이 0~1의 분포를 갖도록 변형하여 학습 알고리즘에 적용하였다.

$$\bar{d}_i = \frac{d_i - \min(d)}{\max(d) - \min(d)} \quad (4)$$

원시 자료의 특징은 현재 시점의 월(month), 일(day), 시(hour), 온도(temperature), 습도(humidity), 강우량(precipitation), 일사량(insolation)의 7개이다. 이들 중에서 월, 일, 시, 강우량 4개는 원본 그대로 사용하고, 온도, 습도, 일사량 3개는 시간의 흐름에 영향을 받기 때문에, 일정 구간의 과거 데이터를 같이 사용하였다. 과거 데이터를 그대로 사용할 수도 있지만, 특징의 개수가 많아지기 때문에 다음과 같이 전처리를 수행하였다. 먼저, 각 데이터들의 이동 평균을 계산한다. 1시간 단위의 입력 데이터에 대해서 최적의 구간의 크기를 찾아내기 위하여, 6, 12, 18, 24로 구간의 크기를 변경하면서 이동 평균을 계산하였다.

$x$ 는 각 입력 데이터를 의미하고,  $n$ 은 전체 데이터 크기이고,  $k$ 는 이동 평균을 계산하기 위한 창(창의 크기 (6, 12, 18, 24)이다. 데이터 내의 임의의 위치  $i$ 에 대하여, 창은 이동 평균을 계산하기 위하여  $x_j$ 개의 데이터를 보유하고 있다. 창은 기준 입력  $x_i$ 를 제일 오른쪽에 보유하고 있다. 기준 입력  $x_i$ 가 데이터 집합의 시작 지점과 끝 지점에 있을 때는 전체 창의 크기를 가질 수 없기 때문에 그에 따른 보정을 수행

하였다. 식 (5)는  $k$ 개의 데이터에 대한 이동 평균을 계산한 것이다. 여기서,  $d$ 는 온도(temperature), 습도(humidity), 일사량(insolation)이다.

$$avg_{d,i} = \begin{cases} \frac{1}{k} \sum_{j=i-k_i}^i x_j & \text{if } i < k \\ \frac{1}{k} \sum_{j=i-k}^i x_j & \text{if } k \leq i \leq n-k \\ \frac{1}{k_i} \sum_{j=i-k_i}^i x_j & \text{if } i > n-k \end{cases} \quad (5)$$

이동 평균을 계산하면서 사용된 데이터에 대한 최대값( $\max_d$ )과 최소값( $\min_d$ )을 계산한다. 그리고, 이동 평균을 계산한 구간의 기울기를 계산한다. 기울기는 전체 기울기와 전반부 기울기, 후반부 기울기의 세 가지를 계산한다.

$$gradT_d = \begin{cases} x_i - x_1 & \text{if } i < k \\ x_i - x_{i-k} & \text{if } k \leq i \leq n-k \\ x_i - x_{i-k_i} & \text{if } i > n-k \end{cases} \quad (6)$$

$$gradF_d = \begin{cases} x_{i-k/2} - x_1 & \text{if } i < k \\ x_{i-k/2} - x_{i-k} & \text{if } k \leq i \leq n-k \\ x_{i-k/2} - x_{i-k_i} & \text{if } i > n-k \end{cases} \quad (7)$$

$$gradL_d = \begin{cases} x_i - x_{i-k/2} & \text{if } i < k \\ x_i - x_{i-k/2} & \text{if } k \leq i \leq n-k \\ x_i - x_{i-k/2} & \text{if } i > n-k \end{cases} \quad (8)$$

또한, 각 데이터의 최대값과 최소값의 차이를 식 (9)와 같이 계산하고, 그들 간의 차이를 식 (10)과 같이 계산한다. 또한 식 (11)과 같이 평균 간의 차이도 계산한다.

$$\begin{cases} D_{temp} = \max_{temp} - \min_{temp} \\ D_{hum} = \max_{hum} - \min_{hum} \\ D_{insol} = \max_{insol} - \min_{insol} \end{cases} \quad (9)$$

$$\begin{cases} D_1 = D_{temp} - D_{hum} - D_{insol} \\ D_2 = D_{hum} - D_{insol} - D_{temp} \\ D_3 = D_{insol} - D_{temp} - D_{hum} \end{cases} \quad (10)$$

$$\begin{cases} D_{a1} = avg_{temp} - avg_{hum} - avg_{insol} \\ D_{a2} = avg_{hum} - avg_{insol} - avg_{temp} \\ D_{a3} = avg_{insol} - avg_{temp} - avg_{hum} \end{cases} \quad (11)$$

따라서 입력 데이터는 원시 자료 7개, 이동 평균 관련 자료 8개, 기울기 자료 9개, 속성간의 차이 자료 9개로 모두 34개의 데이터가 만들어진다.

### 3.3 앙상블 학습

태양광 예측 모델의 성능을 높이기 위하여 신경

회로망, SVM, RBF네트워크의 결과를 결합하는 앙상블 학습(Ensemble Learning)을 하였다. 각각의 신경회로망, SVM, RBF네트워크의 구성과 학습은 2장에서 살펴보았다. 일반적인 앙상블은 학습 데이터를 변화시키면서 학습시켜 여러 모델을 만들고, 이를 결합하여 하나의 분류기를 생성하는 것이다. 하지만, 본 논문에서는 학습 데이터는 고정된 상태에서 여러 가지 학습 알고리즘을 사용해서 각 모델의 결과에 대한 투표 (Voting) 기법을 사용했다. 본 논문에서는 신경회로망, SVM, RBF 네트워크 모델이 각각 48개씩 만들어진다. 이 모델들을 12시간씩의 4그룹으로 나누어, 다른 투표를 수행하였고, 그 방법은 식 (12)와 같다.

$$F(x) = \begin{cases} a_{1,1}f_1(x) + a_{1,2}f_2(x) + a_{1,3}f_3(x) & \text{if } t \leq 24h \\ a_{2,1}f_1(x) + a_{2,2}f_2(x) + a_{2,3}f_3(x) & \text{if } t > 24h \end{cases} \quad (12)$$

세 모델의 가중치  $a_{1,1}, a_{1,2}, a_{1,3}$ 는 예측하는 값의 구간에 따라 상이한 값을 할당하였다. 가중치의 집합은 1개의 조합 (1, 0, 0), (0, 1, 0), (0, 0, 1), 2개의 조합 (1/2, 1/2, 0), (1/2, 0, 1/2), (0, 1/2, 1/2), 3개의 조합 (1/3, 1/3, 1/3)의 7개의 집합을 만들고, 각각의 예측 결과를 비교하였다. 실험을 통하여 알고리즘들을 조합한 후 최적의 결과를 가중치로 설정하였다.

## 4. 실험환경 및 결과

### 4.1 실험 데이터

캘리포니아 태양 에너지 통계(California Solar Statistics)에서 제공하는 미국 캘리포니아에 있는 태양광 발전기에 대한 공개된 태양광 발전량 데이터 중에서 2010년 10월부터 2015년 8월까지 15분 간격으로 수집된 데이터를 사용하였다[12]. (Fig. 3)은 캘리포니아 태양 에너지 통계에서 제공하는 옐로(Yolo)주 데이비스(Davis)시에 있는 발전소의 2014년도의 발전량 데이터의 분포이다. 데이비스시는 전형적인 지중해 기후로 여름에는 건조하며 덥고, 겨울에는 시원하고 비가 자주 온다. 발전량이 여름과 겨울이 일정하지 않고, 차이가 있다.

NSRDB (National Solar Radiation Database) [13]에서 기온, 습도, 강수량을 제공하고 있고, 일사량 데이터는 Utah 대학의 Meso West[14]에서 제공하고 있다. 실험에 사용된 데이터는 2013년 1월부터

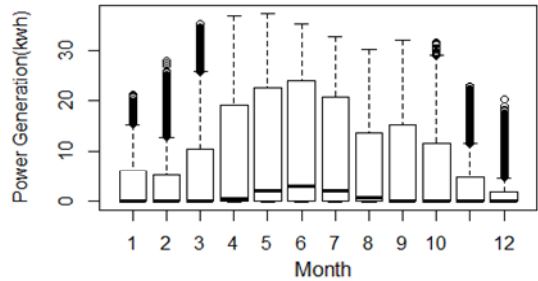


Fig. 3. 2014 monthly power generation.

2014년 12월까지 24개월간의 데이터를 1시간 간격으로 정렬한 17,520의 데이터를 학습용 데이터로 하였고, 2015년 1월부터 2015년 8월까지의 데이터를 테스트용 데이터로 하였다.

### 4.2 알고리즘 성능 평가 방법

학습 결과와 목표 값이 평균적으로 얼마나 떨어져 있는가를 판단하는 방법으로 식 (13)과 같이 평균 제곱근 오차를 사용한다. 실제값과 예측값의 오차를 제공하여 평균한 값의 제곱근이다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2} \quad (13)$$

여기서,  $F_i$ 는 예측 모델에 의한 예측값이고,  $O_i$ 는 실제 값이며,  $N$ 은 데이터의 수이다. RMSE가 작을수록 예측이 정확한 것이다.

### 4.3 개별 알고리즘의 학습 결과

이동 평균의 구간을 변경시키면서 1시간 이후의 태양광 발전량을 예측한 결과는 (Fig. 4)와 같다. 가장 예측 결과가 좋은 것은 이동 평균의 구간이 24일 때이다. Table 1은 세 개의 알고리즘에 대하여 1시

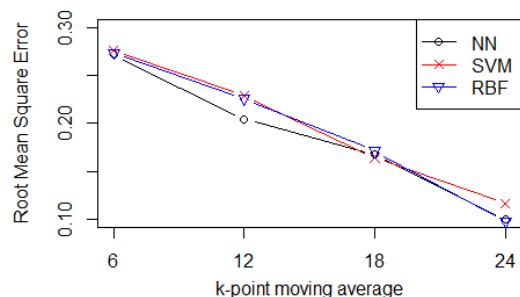


Fig. 4. Power generation error for one hour later by windows.

Table 1. Overall results at 24 point moving average

Algorithm	training (RMSE)	test (RMSE)
Neural Networks	0.0999352	0.138536
SVM	0.116426	0.162621
RBFN	0.098069	0.173894

간 단위로 48시간을 예측했을 때의 평균 제공근 오차값이다.

Fig. 5는 48시간의 구간을 예측했을 때 실제값과의 차이이며 신경회로망이 우수한 성능을 보였다.

#### 4.4 모델 앙상블(Ensemble)

세 알고리즘 중 신경회로망(Neural Networks)과 SVM은 Fig. 4와 같이 전구간 (1시간~48시간)에 걸쳐서 고른 성능을 보였다. RBFN은 전반부 (1시간~24시간)의 성능과 후반부 (25시간~48시간)의 성능에 차이를 보였다. 식 (12)의 앙상블 알고리즘의 계수는  $a_{1,1} = a_{1,2} = a_{1,3} = 1/3$ 으로 하고,

$a_{2,1} = a_{2,2} = 1/2, a_{2,3} = 0$ 으로 계수를 설정하였을 때, 학습 데이터에 대하여 앙상블 기법은 0.076523의 RMSE 값을 보였고, 테스트 데이터에 대해서는 0.129584의 RMSE 값을 보였다. Table 1의 결과와 비교하면, 학습데이터에 대해서는 22.0%~34.3%의 성능 향상을 테스트 데이터에 대해서는 6.5%~25.5%의 성능 향상을 보였다. (Fig. 6)은 (Fig. 5)에서 살펴본 구간과 동일한 48시간의 데이터에 대한 앙상블 기법의 예측 결과이다.

제안하는 앙상블 태양광 발전량 예측 모델을 재사용할 수 있는지를 확인하기 위하여 24일간의 데이터를 48시간 간격으로 적용한 결과는 (Fig. 7)과 같다. 분석 모델의 수행 결과 테스트 데이터의 처음 수행 결과뿐만 아니라 마지막 수행 결과도 유사한 결과를 보이며, 분석 모델이 재사용 가능함을 확인할 수 있었다.

## 5. 결 론

본 논문에서는 도서 및 산간 지역에 상시 네트워크를 연결하여 기상 예보 정보를 자동으로 수집하는 시스템이 구축되지 않은 기존 설비를 이용하거나 보안상의 이유로 네트워크 사용이 어려운 환경에서 적용할 수 있는 태양광 예측 모델을 제안하였다. 기상

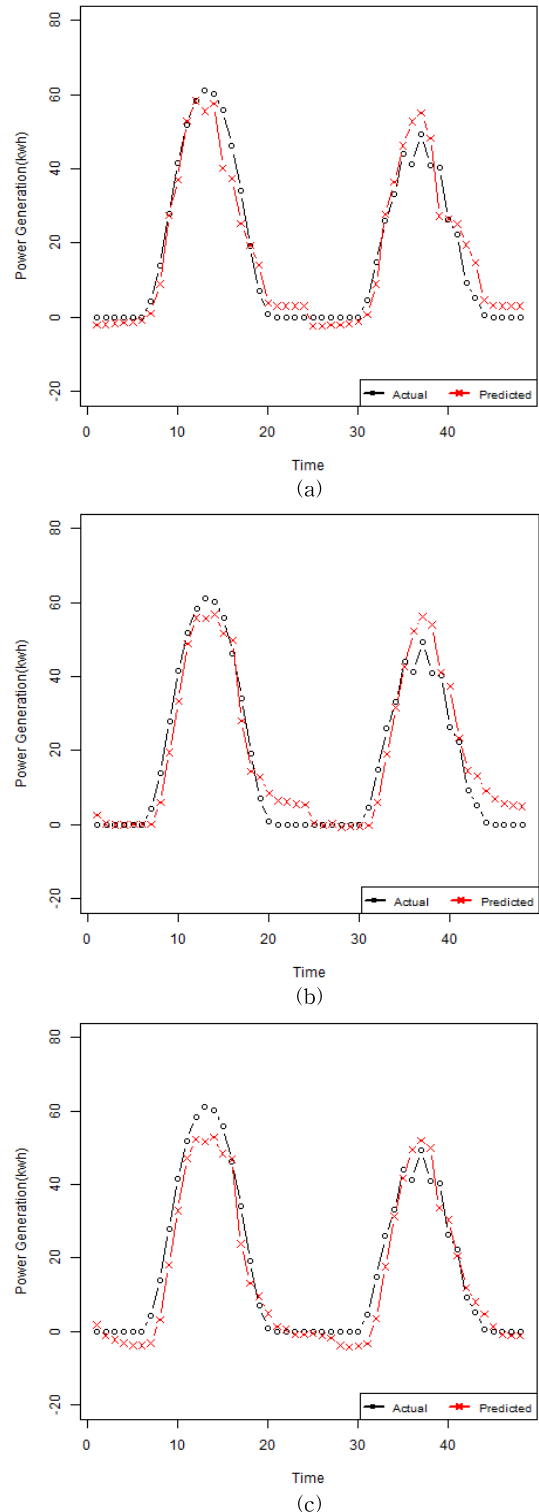


Fig. 5. 48 hours prediction results of each algorithms (a) Neural Networks, (b) SVM and (c) RBFN

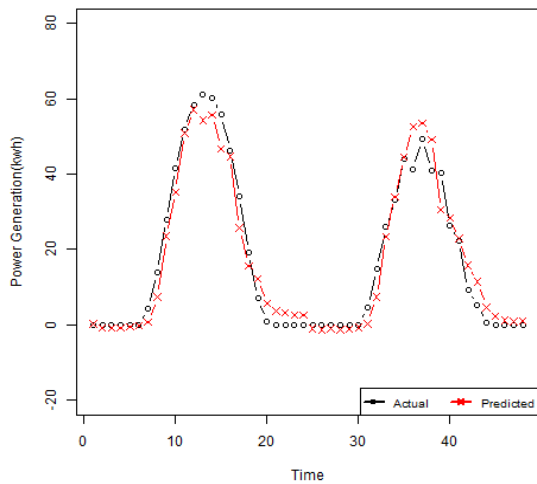


Fig. 6. 48 hours prediction result of Ensemble method.

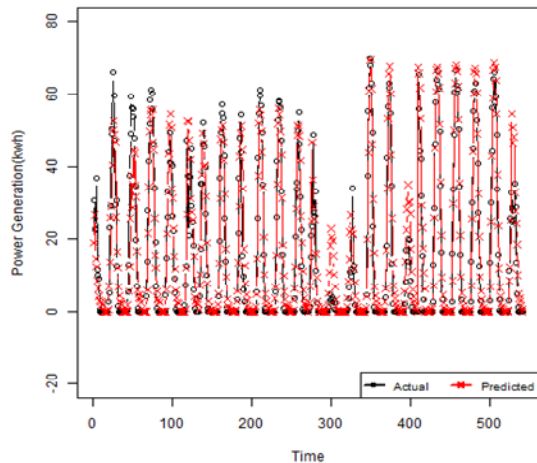


Fig. 7. Long term predict result by Ensemble model.

정보의 시계열 특성을 반영하기 위하여 현 시점을 기준으로 이전 24시간의 온도, 습도, 일사량에 대하여 이동 평균과 기울기를 계산하고, 이들간의 차이를 계산하여 차원을 축소한 새로운 특징을 반영하고, 신경회로망, SVM, RBF 네트워크에 의해 생성된 모델들을 앙상블 기법으로 결합하여 예측 구간별로 다른 계수를 적용하여 태양광 발전량에 대한 예측의 정확도를 높였다.

이 모델을 기상예보 정보와 결합한다면, 더 높은 예측 정확도를 보일 것이다. 또한, 다양한 기상 데이터로 관측할 수 있는 운량, 센서를 통해 수집할 수 있는 기압, 태양광 전지판에 대한 정보 들을 데이터로 사용할 수 있다면 더 좋은 예측 성능을 보일 수

있을 것이다. 향후 연구과제는 다음과 같다. 입력특징의 다양화에 관한 연구와 예측 모델을 자동으로 갱신할 수 있도록 점진 학습(Incremental Learning)을 통해서 한 번 생성된 예측 모델의 성능을 최적으로 유지시키기 위한 연구가 필요하다.

## REFERENCE

- [1] BP Statistical Review of World Energy June 2016, bp.com/statisticalreview, (accessed June, 1, 2016).
- [2] S.M. Lee and W.J. Lee, "Backpropagation Algorithm Based Fault Detection Model of Solar Power Generation Using Weather Data and Solar Power Generation Data," *Proceedings of the Korea Information Processing Society Conference*, Vol. 22, No. 1, 2015.
- [3] J.J. Song, Y.S. Jeong, and S.H. Lee, "Analysis of Prediction Model for Solar Power Generation," *Journal of Digital Convergence*, Vol. 12, No. 3, pp. 243-248, 2014.
- [4] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting Solar Generation from Weather Forecasts Using Machine Learning," *Proceeding of IEEE International Conference on Smart Grid Communications*, pp. 528-533, 2011.
- [5] C. Lee and P. Ji, "Modeling of Photovoltaic Power Systems Using Clustering Algorithm and Modular Networks," *Transactions of the Korean Institute of Electrical Engineers*, Vol. 65P, No. 2, pp. 108-113, 2016.
- [6] C. Park, "Study on the Obsolescence Forecasting Judgment of PV Systems adapted Micro-inverters," *Journal of Korea Multimedia Society*, Vol. 18, No. 7, pp. 864-872, 2015.
- [7] B.D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, 1996.
- [8] J.S. Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge, 2004.

- [9] R. Collobert and S. Bengio, "SVM-Torch: Support Vector Machines for Large-Scale Regression Problems," *Journal of Machine Learning Research*, Vol. 1, No. 2, pp. 143–160, 2001.
- [10] M.D. Buhmann, *Radial Basis Functions: Theory and Implementations*, Cambridge University Press, Cambridge, 2003.
- [11] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h Solar Power Forecasting Based on Weather Type Classification Using Artificial Neural Network," *Journal of Solar Energy*, Vol. 85, No. 11, pp. 2856–2870, 2011.
- [12] Go solar CALIFORNIA, [https://www.californiasolarstatistics.ca.gov/data\\_downloads/](https://www.californiasolarstatistics.ca.gov/data_downloads/), (accessed Jan., 12, 2016).
- [13] National Oceanic and Atmospheric Administration, <http://www.ncdc.noaa.gov/>, (accessed May, 3, 2016).
- [14] Meso West, <http://mesowest.utah.edu>, (accessed Feb., 13, 2016).



이 현 진

1996년 순천향대학교 전산학과  
공학사  
1998년 연세대학교 대학원 컴퓨터  
과학과 공학석사  
2002년 연세대학교 대학원 컴퓨터  
과학과 공학박사

현재 숭실사이버대학교 컴퓨터소프트웨어학과 부교수  
관심분야: 이터닝, 머신러닝, 빅데이터 처리