

# Development of a System for Predicting Photovoltaic Power Generation and Detecting Defects Using Machine Learning

Seungmin Lee<sup>†</sup> · Woo Jin Lee<sup>\*\*</sup>

## ABSTRACT

Recently, solar photovoltaic(PV) power generation which generates electrical power from solar panels composed of multiple solar cells, showed the most prominent growth in the renewable energy sector worldwide. However, in spite of increased demand and need for a photovoltaic power generation, it is difficult to early detect defects of solar panels and equipments due to wide and irregular distribution of power generation. In this paper, we choose an optimal machine learning algorithm for estimating the generation amount of solar power by considering several panel information and climate information and develop a defect detection system by using the chosen algorithm generation. Also we apply the algorithm to a domestic solar photovoltaic power plant as a case study.

**Keywords :** Solar Photovoltaic Power Generation, Detecting of Solar Panel Defects, Prediction of Power Generation, Machine Learning

## 기계학습을 이용한 태양광 발전량 예측 및 결함 검출 시스템 개발

이 승 민<sup>†</sup> · 이 우 진<sup>\*\*</sup>

## 요 약

여러 개의 태양전지들이 붙어 있는 태양광 패널을 이용하여 전력을 생산하는 태양광 발전은 최근 신재생 에너지 기술로 빠르게 성장하고 있는 분야이다. 하지만 태양광발전의 단점 중 하나인 불규칙한 전력 생산문제로 인해, 장비 및 패널 결함에 빠르게 대응하지 못하는 문제가 발생한다. 이 연구에서는 다양한 기후데이터와 패널 정보를 이용하여 태양광발전량 예측 방법들을 비교하여 최적의 예측 알고리즘을 평가하고 이를 기반으로 태양광발전소 결함 검출 시스템을 개발하여 국내 태양광 발전소에 적용한 사례를 기술한다.

**키워드 :** 태양광발전, 패널결함 검출, 발전량 예측, 기계학습

## 1. 서 론

2000년부터 2014년 사이 태양광 발전의 연간 성장률은 전력발전 기술 중 가장 빠르게 증가하고 있는 분야이다. 태양광 발전량은 전 세계적으로 2004년 4GW에서 2014년말 177GW로 증가하였다. 특히 최근 3년 동안 평균 35GW가 증가하면서 신재생에너지 분야에서 가장 두드러진 성장세를 나타내었다. 태양광발전의 수요 증가와 필요성에도 불구하고 기술 및 발전설비 구축의 어려움, 관련인력 부족 등의 문제를

갖고 있다. 특히 기후변화 및 먼지, 기타 이물질에 의한 감소, 그 외의 기타 패널 결함 문제는 현재 오퍼레이터가 실사를 통한 확인 작업으로 결함 판별을 진행하고 있으며, 관련 인력부족으로 전력의 손실 및 장비 방치가 이루어지고 있는 실정이다. 태양광발전은 특성상 기후조건에 따른 발전량 변동이 심하며 기온, 전운량, 일사량 등 단일 데이터를 이용한 단순한 비교만으로 발전 상태를 예측할 수 없다. 태양광발전의 그 필요성과 수요증대에 따라 패널결함 검출 자동화 시스템의 필요성이 요구되며, 시스템 구축에는 태양광발전량의 정확한 예측에 관한연구가 선행되어야 한다.

예측 알고리즘으로 신경망[1], SVM[2], 딥러닝[3]을 사용하고 예측의 오차를 파악할 때 가장 많이 사용되는 평균 제곱근 오차(RMSE)를 이용하여 최적의 알고리즘을 선정한다. 알고리즘 구조 변경 및 상수 수정을 통해 예측율을 확대시킨 예측 모델을 제안한다. 그리고 예측된 결과를 국내 제주

※ 본 논문은 교육부 및 한국연구재단의 BK21 플러스 사업(경북대학교 컴퓨터학부 Smart Life 실현을 위한 SW인력양성사업단)으로 지원된 연구임(21A20131600005).

† 준 회 원 : 경북대학교 컴퓨터학부 공학석사

\*\* 정 회 원 : 경북대학교 컴퓨터학부/소프트웨어기술연구소 교수  
Manuscript Received : August 9, 2016  
Accepted : August 29, 2016

\* Corresponding Author : Woo Jin Lee(woojin@knu.ac.kr)

지역 데이터에 적용하여 결함 검출 시스템을 개발한다.

본 논문은 다음과 같이 구성된다. 제 2장의 연구배경에서 태양광 발전 현황을 소개하고 기존의 태양광발전량 예측에 관련된 방법을 소개 및 비교한다. 제 3장에서는 기후데이터 수집 및 발전량 데이터 수집하고 SVM, 신경망, 딥러닝을 이용해 실험하여 RMSE 값으로 알고리즘을 평가하여 최적의 알고리즘을 선정한다. 제 4장에서는 태양광발전 결함검출 방법에 대해 논한다. 제 5장에서는 최적의 결과를 보인 알고리즘을 이용해 국내 태양광발전소 결함검출 시스템을 구현하고 제 6장에서 결론을 기술한다.

## 2. 연구 배경

### 2.1 태양광 발전현황

2000년부터 2014년 사이 태양광 발전의 연간 성장률은 전력발전 기술 중 가장 빠르게 증가하고 있는 분야이다. 태양광 발전량은 전 세계적으로 2004년 4GW에서 2014년 177GW까지 증가하였다. 특히 Fig. 1[4]에서 보여지듯, 2011년에 30GW, 2012년에 38GW, 2013년에 39GW가 증가하며 최근 들어 더욱 두드러진 성장세를 보이고 있다.

Fig. 2의 국가별 상위 10개국을 분석해보면 중국, 일본, 미국이 가장 큰 성장세를 보이고 있다[4]. 또한 아시아 지역의 신규 설비용량이 대폭 증가하며 유럽국가의 설비용량을 뛰어넘고 있다.

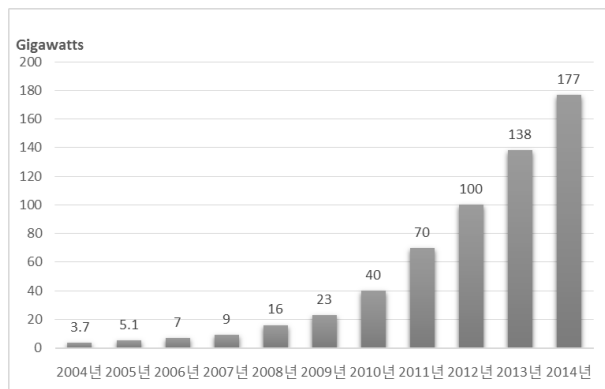


Fig. 1. Cumulative Solar PV Capacity

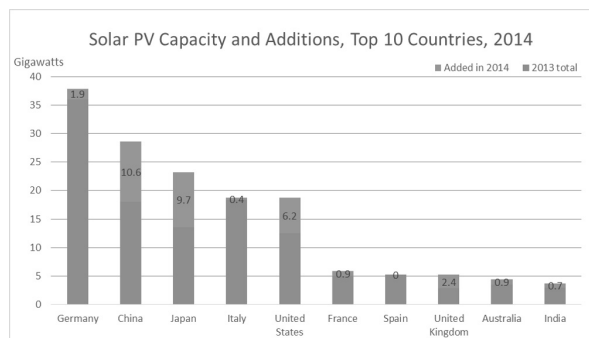


Fig. 2. Solar PV Capacity and Additions, 2014

이러한 보급 확산은 태양광 발전의 비용하락으로 인한 결과물로 해석된다. 태양광 발전비용은 지난 30년간 지속적으로 하락했으며, 이는 19.3%의 학습비율(learning experience or learning ratio: 용량이 2배가 될 때마다 가격은 19.3% 하락)로 나타났다. 제조공정의 확대 및 개발노력에 의한 실적과 가격 개선전망을 고려할 때, 앞으로 상당기간 지속될 것으로 예측된다.

### 2.2 태양광 발전의 문제점

태양광발전의 수요 증가와 필요성에도 불구하고 기술 및 발전설비 구축의 어려움, 관련인력 부족 등의 문제를 갖고 있다. 특히 Fig. 3(a)는 먼지 및 기타 이물질에 의한 발전량 감소, Fig. 3(b)는 패널 결함 문제로 인한 발전량 감소 문제를 보여준다. 이러한 문제로 인하여 오퍼레이터는 실사를 통해 결함 판별을 진행해야 하므로 인건비로 인한 발전단가 상승을 피할 수 없다. 그리고 이 또한 인력부족으로 인한 전력의 손실 및 장비 방치[1]가 이루어지고 있다[5].

일반적으로 태양광발전은 일조시간, 전운량, 일사량과 관계가 있다고 알려져 있으나 기온, 전운량, 일사량 등 단일 데이터를 이용한 단순한 비교만으로 발전 상태를 예측할 수 없다. 그러므로 태양광발전의 그 필요성과 수요증대에 의하여 패널결함 검출 자동화 시스템의 필요성이 요구되며, 그 이전에 태양광발전량의 예측연구가 선행될 필요성이 있다.

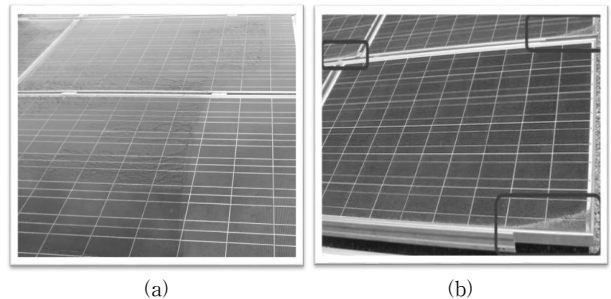


Fig. 3. Power Output Reduction Factor

### 2.3 SVM 활용 발전량 예측 연구

SVM(Support Vector Machine)[2]은 기계학습 알고리즘으로 지도학습(Supervised Learning)을 위해 사용하는 학습 모델로서, 분류와 예측에 주로 사용된다. Fig. 4(a)와 같이 원과 사각형으로 이루어진 데이터가 존재하는 경우, 원과 사각형 데이터들간의 벡터 거리를 측정하여 데이터간의 거리가 최대가 되는 평면을 초평면(Hyperplane)이라 하고 선형(Linear)으로 분류할 수 있다. Fig. 4(a)는 원과 사각형이 선형 분류가 가능한 데이터이다. Fig. 4(b)는 선형 분류를 적용하기 어려운 데이터이며, 커널함수를 이용하여 고차원 공간으로 변환하고 벡터 내적 결과를 리턴하여 분류한다. SVM의 커널함수의 종류로는 Linear Kernel, Polynomial Kernel, RBF(Radial Basis Function) kernel이 있으며 RBF을 이용한 발전량 예측모델 연구가 좋은 성과를 나타내었다[6].

이런 고차원 벡터를 이용하여 기후 데이터를 이용한 초평면을 구하고 발전량 분류기(Classifier)를 만들어 태양광발전량 예측에 이용한 연구는 단일 벡터를 이용한 경우에 비해 약 27%의 예측 성능향상이 이루어졌다[7].

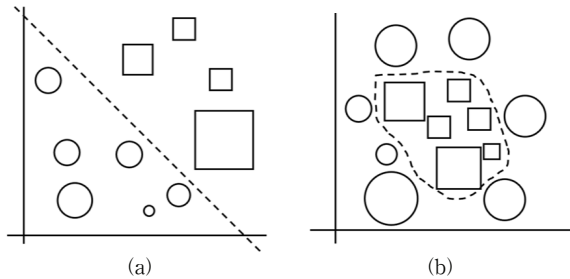


Fig. 4. Data Consisting of Circles and Squares

2.4 신경망 활용 발전량 예측 연구

신경망(Neural Network)[1]은 인간이나 동물에 뇌의 구조와 기능을 모델화하여, 외부로부터 받아들이는 입력과정과 출력과정을 학습시키고 학습과정이 끝난 후 실제 입력에 대한 출력을 예측하는 알고리즘 시스템이다. 패턴인식, 데이터 마이닝, 제어 등 다양한 분야에 응용되고 있고, 목적에 맞도록 출력형태, 신경망 구조, 계층수를 정의하여 사용한다. 신경망 모델은 입력층, 은닉층, 출력층으로 구조를 나눌 수 있다. 학습용 데이터를 입력층과 출력층의 값으로 입력하여 은닉층의 가중치값을 학습하고, 학습된 가중치값으로 입력에 따른 출력값을 예측에 사용한다.

일반적으로 태양광발전량 예측을 위한 신경망 구성에서는 기후 데이터를 입력층의 데이터로 사용하고, 단위발전량 값을 출력층 데이터로 구성한다. 그리고 신경망을 훈련시키기 위해 예측값과 실제값의 차이를 역으로 전파하여 가중치를 수정하는 역전파(Backpropagation) 알고리즘을 주로 사용한다[8]. 역전파 알고리즘은 신경망의 출력과 그 쌍에서 오는 출력을 이용하며, 은닉층의 뉴런마다 가중치값을 계산을 통해 알아낸다. 가중치를 가지고 지역 에러값을 연산하며, 끊임없이 가중치 연산을 진행한다[9]. 정해진 Error범위 내에 오차가 발생하는 경우 연산이 끝나게 되고 연산된 가중치를 저장하여 Training 과정이 끝난 후 실제 예측분야에 사용한다. 이때 학습용 데이터가 충분치 않고 잡음이 있는 경우 Overfitting이 발생할 수 있으며, 학습률 과다일 경우 학습속도는 빠르지만 오차계수에 따른 불안정성문제를 내재하고 있다. 그 외로 태양광발전소의 입지선정을 위해 신경망을 이용하여 월별 일사량을 예측하고 보정하여 예측된 일사량으로 발전량을 예측하는 연구도 진행중이다[10].

각 연구에서는 MAPE, RMSE 값을 이용하여 예측의 정확성을 나타내고 있다. 위 두 값은 예측이 얼마나 잘 이루어져 있는지 정확도와 정밀도를 수치화한 값이다.

3. 기계학습 알고리즘 실험

3.1 실험 데이터 추출

1) 태양광발전량 데이터 선정

태양광발전량 데이터를 얻기 위해 PVOutput.org의 데이터를 이용하였다. PVOutput.org에서는 무료로 데이터를 공유하고 실시간으로 태양광 발전량 비교, 모니터링, 데이터 소비량을 공유하는 웹사이트이다[11].

개인 사업자는 태양광발전소의 일별 데이터를 PVOutput에 제공하며 지역별 분류를 통해 평균 데이터와 비교, 발전소 문제 발생여부를 파악하기 쉽도록 제공한다. PVOutput.org에서 제공하는 미주 지역의 태양광발전소 분포는 서부의 캘리포니아주에 411개, 동부의 뉴욕과 메사추세츠주에 394개로 가장 많이 분포하였다. 캘리포니아주의 샌프란시스코의 기후는 최고기온 평균 18.45℃, 최저기온 평균 9.42℃, 강수량 41.7mm, 평균 강수일수 4일로 여름에는 서늘하고 겨울에는 따뜻한 지중해성 기후를 갖고 있어, 태양광발전량의 출력이 대부분 일정하게 나타난다. 이에 비해 메사추세츠주의 보스턴 기후를 보면 최고기온 평균 15℃, 최저기온 평균 6.44℃로 최고-최저 기온 차는 샌프란시스코와 비슷하지만, 강수량 87.85mm, 평균 강수일수 8.22일로 습윤 대륙성 기후와 온난 다습한 기후의 중간을 위치하고 있어 태양광발전량의 출력 변화가 심하여, 발전량예측에 메사추세츠주의 데이터가 가장 적합하였다. 메사추세츠주의 태양광발전소 중 일정량 이상의 데이터와 대량의 규모의 발전량 데이터를 갖춘 Concord Cape 발전소의 발전량 데이터를 추출하여 본 실험의 예측알고리즘에 이용한다.

2) 기후데이터 추출

태양에너지 최적 활용을 위한 기상자원 분석 보고서[12]와 Chen, C.[13]의 연구에 따르면 6가지 요소인 일사량, 강수량, 기온, 전운량, 습도, 안개가 태양광발전에 가장 큰 영향을 미친다. 이 6가지 변수를 태양광발전량 예측에 관한 입력변수로 선정하여 예측 실험에 이용한다. NCEI(National Centers for Environmental Information)의 산하기관인 NOAA(National Oceanic and Atmospheric Administration)의 NSRDB(National Solar Radiation Database)[14]에서는 기온, 습도, 강수량, 가시거리(안개), 전운량을 제공하고 있어 콩코드시의 메타 데이터를 추출하고, 태양광발전량 데이터와 기간이 일치하도록 일별 데이터로 변환해 실험에 사용한다. 일사량 데이터는 Utah대학의 날씨 관측 데이터베이스인 Meso West[15]에서 동일 기간의 메타 데이터를 추출하였다. Meso West에서는 한 도시 내에도 관측지가 여러 곳이 있어 보다 정확한 데이터를 얻을 수 있다.

3.2. 알고리즘 성능 평가 방법

실험결과와 예상값이 평균적으로 얼마만큼 떨어져 있는가를 판단하는 방법으로 평균 제곱근 오차인 RMSE(Root Mean Square Error) 값을 사용한다. 실제값과 예측한 값에서 나타나는 오차를 제공하여 평균한 값의 제곱근을 의미한다[16].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [y(x_i) - y'(x_i)]^2} \quad (1)$$

$y(x_i)$ :  $x_i$  에서 실제 측정값  
 $y'(x_i)$ :  $x_i$  에서 예측값  
 $n$ : 데이터의 수

RMSE는 위의 Equation (1)을 따른다. 이는 실제 발전량과 예측값이 평균적으로 얼마만큼 차가 있는지 확인하는 척도이며 RMSE가 적을수록 보다 정확한 예측이 되었다고 말할 수 있다.

### 3.3 기계학습 알고리즘 적용

#### (1) 신경망을 이용한 예측실험

학습에서는 입력층의 노드수와 은닉층의 노드수, 출력층의 노드수를 설정하여 신경망 구조를 결정한다. 0.0에서 1.0 사이의 값으로 정규화된 실험용 데이터를 입력값과 출력값으로 이용하여 학습률을 높인다. 학습량, 은닉층의 단위수, 최대 에러율의 선정에는 분명한 규칙이 없고, 잘못된 초기값은 과대적합(Overfitting) 문제를 보이므로 적절한 값을 선정하여 반복적인 학습 실험을 통해 최적의 초기값을 찾는다[17].

학습방법으로 역전파(Backpropagation) 알고리즘을 사용하는데 Fig. 5와 같은 형태의 학습과정을 반복한다. 반복적인 학습을 통해 가중치 값은 최종적으로 수렴되어 학습과정을 멈추게 된다.

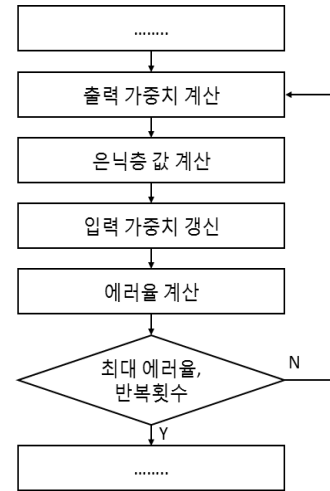


Fig. 5. Neural Network Learning Process

실험데이터로 선정된 데이터의 기후데이터를 입력층, 발전량데이터를 출력층으로 선정하여 Fig. 6과 같이 입력층 6개, 출력층 1개의 신경망을 구성한다. 은닉층의 개수는 3, 5, 7, 9, 11개로 실험하고, 1000회의 epoch를 갖는다. 또한 최대 에러율은 0.0001, 0.001, 0.01, 학습률 0.0001, 0.001, 0.01, 0.1을 선정하여 60회 진행하였으며, 예측된 결과는 RMSE값을 산출하여 평가한다. 실험결과는 Table 1과 같이 은닉층 개수 5, 최대 에러율 0.01, 학습률 0.001일 때 RMSE값은 0.140689으로 최적의 예측 결과를 도출하였다.

Table 1. Neural Network Learning Results

은닉층수	최대 에러율	학습률	RMSE	은닉층수	최대 에러율	학습률	RMSE	은닉층수	최대 에러율	학습률	RMSE
3	0.0001	0.0001	0.177036	5	0.01	0.0001	0.149455	9	0.001	0.0001	0.153235
3	0.0001	0.001	0.178475	5	0.01	0.001	0.140689	9	0.001	0.001	0.151185
3	0.0001	0.01	0.149903	5	0.01	0.01	0.146697	9	0.001	0.01	0.146441
3	0.0001	0.1	0.172496	5	0.01	0.1	0.17356	9	0.001	0.1	0.169478
3	0.001	0.0001	0.172837	7	0.0001	0.0001	0.17246	9	0.01	0.0001	0.151867
3	0.001	0.001	0.158102	7	0.0001	0.001	0.153586	9	0.01	0.001	0.150185
3	0.001	0.01	0.14886	7	0.0001	0.01	0.152503	9	0.01	0.01	0.146288
3	0.001	0.1	0.173232	7	0.0001	0.1	0.171725	9	0.01	0.1	0.181676
3	0.01	0.0001	0.155978	7	0.001	0.0001	0.166436	11	0.0001	0.0001	0.174536
3	0.01	0.001	0.149035	7	0.001	0.001	0.155004	11	0.0001	0.001	0.159644
3	0.01	0.01	0.149539	7	0.001	0.01	0.147378	11	0.0001	0.01	0.149554
3	0.01	0.1	0.1756	7	0.001	0.1	0.18838	11	0.0001	0.1	0.176752
5	0.0001	0.0001	0.172024	7	0.01	0.0001	0.144953	11	0.001	0.0001	0.152108
5	0.0001	0.001	0.157754	7	0.01	0.001	0.145667	11	0.001	0.001	0.154946
5	0.0001	0.01	0.150951	7	0.01	0.01	0.144852	11	0.001	0.01	0.144551
5	0.0001	0.1	0.171763	7	0.01	0.1	0.176343	11	0.001	0.1	0.172675
5	0.001	0.0001	0.157696	9	0.0001	0.0001	0.176835	11	0.01	0.0001	0.148997
5	0.001	0.001	0.157235	9	0.0001	0.001	0.15458	11	0.01	0.001	0.146892
5	0.001	0.01	0.152108	9	0.0001	0.01	0.146446	11	0.01	0.01	0.150466
5	0.001	0.1	0.186937	9	0.0001	0.1	0.172471	11	0.01	0.1	0.197629

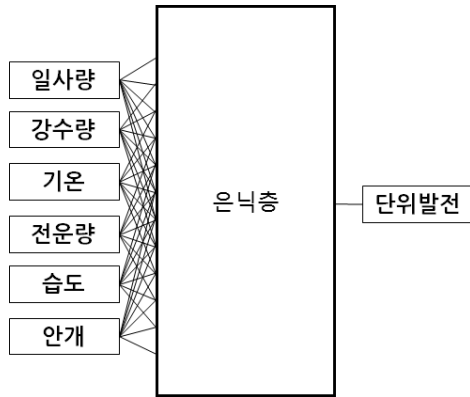


Fig. 6. Neural Network Architecture

(2) SVM을 이용한 예측 실험

SVM(Support Vector Machine)은 기계학습 분야의 알고리즘으로의 분류, 회귀에서 뛰어난 성능을 보였다. 특히 2000년에 National Taiwan University에서 개발된 LIBSVM 라이브러리[18]는 SVM 소프트웨어에서 가장 널리 사용된다. 본 실험에서는 LibSVM Version 3.2을 사용하고 nu-SVM 모델과 RBF-Kernel(Radial Basis Function Kernel)을 사용하였으며, 신경망과 동일한 입출력 데이터를 사용하였다. nu-SVM[2]은 [0, 1] 사이의 숫자 값에 최적화된 SVM 모델이다. 실험결과 RMSE 값은 0.15067784의 값을 보였다.

(3) 딥러닝을 이용한 예측실험

딥러닝 알고리즘 중 심층 신경망을 이용하여 그림 7과 같이 Depth2인 구조를 만들어 예측 실험을 진행하였다. 심층 신경망의 학습과정 신경망과 동일하게 은닉층의 개수는 3, 5, 7, 9, 11개로 실험하고, epoch 1000회, 최대 에러율은 0.0001, 0.001, 0.01, 학습률 0.0001, 0.001, 0.01, 0.1을 선정하여 60회 진행한다.

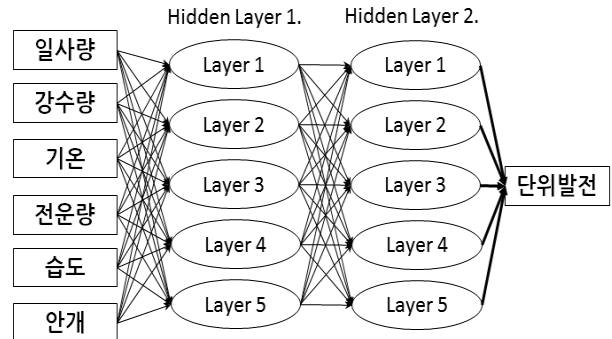


Fig. 7. Deep Learning Architecture

Table 2의 딥러닝 학습결과를 보면 은닉층의 수가 5개, 최대 에러율 0.0001, 학습률 0.01일 때 RMSE값이 0.127648으로 가장 좋은 효율을 보였다.

Table 2. Neural Network Learning Results

은닉층수	최대 에러율	학습률	RMSE	은닉층수	최대 에러율	학습률	RMSE	은닉층수	최대 에러율	학습률	RMSE
3	0.0001	0.0001	0.163138	5	0.01	0.0001	0.137902	9	0.001	0.0001	0.143659
3	0.0001	0.001	0.145025	5	0.01	0.001	0.133265	9	0.001	0.001	0.147541
3	0.0001	0.01	0.139374	5	0.01	0.01	0.134547	9	0.001	0.01	0.142741
3	0.0001	0.1	0.14875	5	0.01	0.1	0.174984	9	0.001	0.1	0.169641
3	0.001	0.0001	0.144658	7	0.0001	0.0001	0.162572	9	0.01	0.0001	0.144377
3	0.001	0.001	0.149495	7	0.0001	0.001	0.146573	9	0.01	0.001	0.138987
3	0.001	0.01	0.145994	7	0.0001	0.01	0.137595	9	0.01	0.01	0.130851
3	0.001	0.1	0.164097	7	0.0001	0.1	0.177374	9	0.01	0.1	0.16198
3	0.01	0.0001	0.138138	7	0.001	0.0001	0.158689	11	0.0001	0.0001	0.157717
3	0.01	0.001	0.141993	7	0.001	0.001	0.136535	11	0.0001	0.001	0.136828
3	0.01	0.01	0.136326	7	0.001	0.01	0.131478	11	0.0001	0.01	0.141306
3	0.01	0.1	0.160191	7	0.001	0.1	0.168021	11	0.0001	0.1	0.165541
5	0.0001	0.0001	0.153929	7	0.01	0.0001	0.139992	11	0.001	0.0001	0.140954
5	0.0001	0.001	0.143032	7	0.01	0.001	0.128042	11	0.001	0.001	0.144574
5	0.0001	0.01	0.127648	7	0.01	0.01	0.147708	11	0.001	0.01	0.129728
5	0.0001	0.1	0.156846	7	0.01	0.1	0.163612	11	0.001	0.1	0.155295
5	0.001	0.0001	0.141019	9	0.0001	0.0001	0.153245	11	0.01	0.0001	0.141123
5	0.001	0.001	0.137832	9	0.0001	0.001	0.160148	11	0.01	0.001	0.135452
5	0.001	0.01	0.130712	9	0.0001	0.01	0.135697	11	0.01	0.01	0.134367
5	0.001	0.1	0.157825	9	0.0001	0.1	0.171108	11	0.01	0.1	0.158749

#### 4. 태양광발전 결함 검출 시스템 설계 및 구현

이 장에서는 웹기반 태양광발전 결함 검출 시스템을 설계 및 구현한다. 기상청 국가기후데이터센터[19]의 제주지역 기후 데이터와 제주도 보타리 에너지 발전소 발전량 데이터의 과거 데이터를 이용해 딥러닝을 학습하고, 검출 시스템에 학습된 결과를 적용한다. 또한 발전소 내부 패널정보를 이용해 결함여부를 판별하여 결함 검출율을 높인다.

##### 4.1 태양광발전 결함 검출 방법

태양광발전 결함 검출 방법은 실시간으로 주변 정보와의 비교를 통해 실시하는 방법과 발전량 예측 결과를 이용하여 판단하는 방법으로 나눌 수 있다. 만약에 주변 태양광 발전소가 있는 경우, 같은 기후의 인접 발전소 발전량 데이터를 이용하여 실시간 비교해 발전효율을 알아낸다. PVOutput.org[11] 웹사이트처럼 인근 발전소가 많은 경우 유용하게 사용할 수 있으나, 국내 발전소의 개수가 부족하여 이용하기 용이치 않다. 주변 비교 대상 발전소가 없어도 발전소내의 패널 데이터를 실시간으로 비교하여 결함을 추출할 수 있다. 패널 데이터는 패널온도, 전압, 에러코드 등을 갖고 있어 실시간 확인해 정확한 패널 결함 상태를 점검할 수 있다.

발전량 예측치를 이용하면 패널 불량을 포함한 발전량 감소 현상을 찾아 낼 수 있다. 발전량을 0에서 1사이로 정규화시켜 RMSE값을 얻었으므로 예측값 대비 실제발전량이 0.127648이하의 값인 경우 결함으로 볼 수 있다. 발전량 예측치를 활용하는 경우 실시간으로 패널 불량을 확인할 수는 없다. 발전량 예측을 위해 사용되는 일부 기상청 데이터는 하루가 지나야 공개됨으로 이러한 방법은 일일 경과후에 불량 패널에 대한 정보를 리포트한다.

##### 4.2 데이터 추출

2014년 10월 05일에서 2015년 07월 11일까지의 제주도 보타리 에너지발전소의 태양광발전량 데이터를 알고리즘 학습

을 위한 학습용 데이터로 이용하였다. 동일 기간의 기후 데이터를 국내 국가기후데이터센터에서 추출하여 학습에 이용한다. 국가기후데이터센터는 본 논문에서 사용한 6가지 기후 정보 중 안개 데이터를 제외한 5가지의 데이터만 추출 가능하였으며 시간별 데이터를 제공하고 있어, 일별데이터로 변환 및 연중 최고 및 최저치의 데이터를 얻어 정규화 시켜 0.0~1.0 값으로 변환하여 학습에 사용한다. BackPropagation과 MultiLayerPerceptron 클래스를 이용하여 학습시키고 학습된 결과를 결함검출 시스템에 이용한다.

학습된 결과를 이용하여, 웹기반 태양광발전 결함 검출 시스템을 위한 데이터 저장 서버를 Fig. 8과 같이 구성한다. 기후 데이터는 익일 오전 10시에 공지되므로 일별 요청하여 데이터베이스를 구성하고 발전소 데이터는 실시간으로 변경되므로 1분 간격으로 데이터를 파싱하여 저장한다.

##### 4.3 태양광발전 결함검출 시스템 구현

시스템의 개발환경으로는 Web Server는 Apache2와 Tomcat7.0을 이용하고, DBMS로 MySQL5, Web App Libraries로 Hibernate3.0, Struts2를 사용한다. 또한 Neuroph2.9을 사용하여 심층신경망을 구성한다. Neuroph[20]는 Apache 제단의 SourceForge팀에서 개발한 Java언어의 신경망 프레임워크이다. 심층신경망의 Neuron을 구성하기 쉽고 학습과정이 빠르다는 장점이 있다.

Fig. 9는 태양광발전 결함 검출 시스템의 구현완료 화면이다. 실시간으로 인버터와 모듈별 발전정보를 화면에 모니터링 가능하게 구성하고, 기상청에서 가져온 기후데이터와 발전데이터를 예측 알고리즘에 이용한 값으로 발전 예측량을 표시하였다.

Fig. 10은 시간별 태양광발전 패널의 평균온도를 표시한 화면으로 패널 온도가 음수값인 경우, 데이터가 미검신된 경우, 모듈정보에 에러 코드가 있는 경우를 따로 표시하여 운영자가 웹상에서 쉽게 확인할 수 있다.

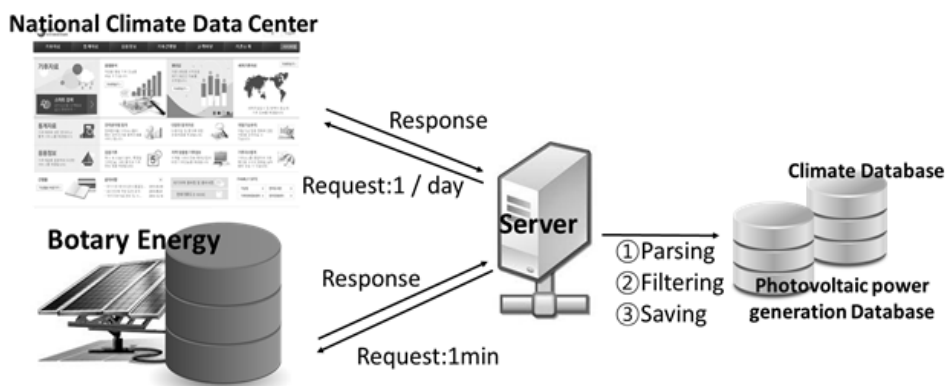


Fig. 8. Server for Data Storage

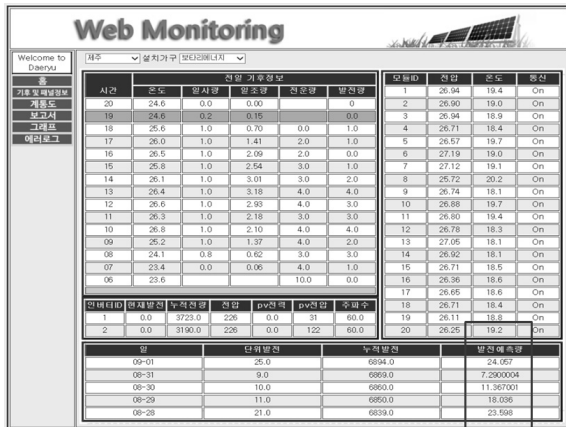


Fig. 9. Solar Power Failure Detection System Implementation Completed Screen

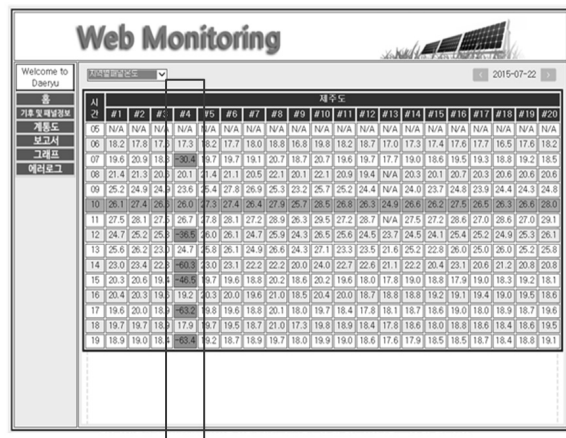


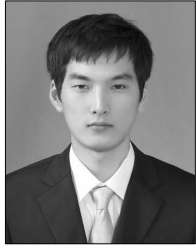
Fig. 10. Solar Power System Fault Detection - Regional Panel Temperature

### 5. 결론

태양광발전소는 그 특성상 결함 발생 시 조기 파악에 많은 어려움이 따른다. 본 논문에서는 결함검출을 위해서 발전량을 예측하여 실제 측정된 발전량과 비교를 통한 결함 판별을 진행하였다. 최적의 예측 알고리즘을 선정하기 위해서 기후데이터와 발전량 데이터를 정규화 시킨 뒤 SVM, 신경망, 딥러닝에 적용하였다. 알고리즘의 평가방법으로 RMSE 값을 이용한 결과 딥러닝이 가장 우수한 예측율을 보였다. 선정된 알고리즘을 국내 데이터에 적용하기 위해 국가기후데이터센터와 제주도 보타리발전소에서 280일간의 데이터를 얻어 학습데이터로 이용하였다. 태양광발전 결함검출 시스템에서는 학습된 딥러닝 알고리즘을 이용한 발전량 예측 및 패널 데이터를 이용한 결함 판별을 진행하여 국내 태양광발전소 발전량 하락 및 대충화에 기여할 수 있을 것으로 판단된다. 향후연구로는 기상 데이터를 실시간으로 전송받아 발전량 예측 모델을 구동하여 실시간 결함 검출 시스템을 구축하고자 한다.

### References

- [1] Takefuji, Yoshiyasu, "Neural network parallel computing," *Springer Science & Business Media*, Vol.164, 2012.
- [2] Harrington Peter, *Machine learning in action*, Manning, 2012.
- [3] Jiquan Ngiam, "Multimodal deep learning," *ICML-11 Conference on Machine Learning*, pp.689-696, 2011.
- [4] REN21, *Renewables 2011 Global Status Report*, REN21 Secretariat, 2011.
- [5] Justine Sanchez, "Potential PV Problems & New Tools for Troubleshooting" [Internet], <http://www.homepower.com/articles/solar-electricity/design-installation/potential-pv-problems/>.
- [6] Radial\_basis\_function\_kernel [Internet], <https://en.wikipedia.org/wiki/>.
- [7] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting Solar Generation from Weather Forecasts Using Machine Learning," *IEEE International Conference on Smart Grid Communications*, 2011.
- [8] Seung Min Lee and Woo Jin Lee, "Backpropagation Algorithm based Fault Detection Model of Solar Power Generation using Weather Data and Solar Power Generation Data," *The 2015 Spring Conference of the KIPS*, Vol.22, No.1, 2015(in Korean).
- [9] Backpropagation [Internet], <https://en.wikipedia.org/wiki/>.
- [10] Jae-Ju Song, Sang-Ho Lee, and Yoon-Su Jeong, "Analysis of prediction model for solar power generation," *The Society of Digital Policy & Management*, Vol.12, No.3, pp.243-248, 2014(in Korean).
- [11] PVoutput [Internet], <http://pvoutput.org/about.html/>.
- [12] Korea Meteorological Administration, *Weather Resource Analysis Report for the Optimal Utilization of Solar Energy*, 2008(in Korean).
- [13] Changsong Chen, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Solar Energy*, Vol.85, No.11, pp.2856-2870, 2011.
- [14] NOAA, Unedited Local Climatological Data Samples [Internet], <http://www.ncdc.noaa.gov/ulcd/ULCD/>.
- [15] MesoWest [Internet], <http://mesowest.utah.edu/>.
- [16] Tianfeng Chai and R. Roland Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, Vol.7, No.3, pp.1247-1250, 2014.
- [17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, Vol.15, No.1, pp.1929-1958, 2014.
- [18] A Library for Support Vector Machines [Internet], <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [19] Korea Meteorological Administration [Internet], <http://www.kma.go.kr/>.
- [20] Java Neural Network Framework [Internet], <http://neuroph.sourceforge.net/>.



이 승 민

e-mail : seung.min.lee.k@gmail.com  
2009년 영진전문대학(전문학사)  
2009년~2012년 BNSoft 주임연구원  
2016년 경북대학교 컴퓨터학부(공학석사)  
관심분야: 임베디드 소프트웨어 테스트,  
기계학습



이 우 진

e-mail : woojin@knu.ac.kr  
1992년 경북대학교 컴퓨터학과(학사)  
1994년 KAIST 전산학과(공학석사)  
1999년 KAIST 전산학과(공학박사)  
1999년~2002년 ETRI 선임연구원  
2002년~현 재 경북대학교 컴퓨터학부/  
소프트웨어기술연구소 교수  
관심분야: 임베디드 소프트웨어 테스트, 임베디드 소프트웨어  
개발환경