Development of Solar Power Generation Prediction System using Artificial Intelligence

Lee Ji Yun, Shin Dong Ho
Student and Professor, MY PAUL SCHOOL
12-11, Dowontongmi-gil, Cheongcheon-myeon, Goesan-gun,
Chungcheongbuk-do, Republic of Korea
eavatar@hanmail.net

Abstract

In this paper, although photovoltaic power generation has recently shown the most remarkable growth in the field of renewable energy worldwide, defects may occur if power outages or manufacturing facilities remain despite the increase in demand and demand for photovoltaic power generation. A machine learning algorithm that predicts the optimal development to solve these problems was obtained through experiments. By implementing the algorithm in the system, it will be able to contribute to reducing operating costs and popularizing it. Neural network, SVM, and deep learning are used as prediction algorithms, and the optimal algorithm is selected by using the root mean square error (RMSE), which is the most used when identifying prediction errors. We propose a predictive model whose prediction rate is expanded by changing the algorithm structure and modifying constants. Then, a defect detection system is developed by applying the predicted results to the domestic regional data.

Keywords
Artificial Intelligence, Solar Power, Generation Prediction System, Solar Power Generation, AI

1. Introduction
The annual growth rate of photovoltaic power generation from 2000 to 2014 is the fastest growing field among power generation technologies. Solar power is increasing worldwide. In particular, it showed the most remarkable growth in the renewable energy sector. Despite the increasing demand and need for photovoltaic power generation, there are problems such as difficulty in constructing technology and power generation facilities, and a lack of related manpower. In particular, workers are currently identifying defects caused by due diligence due to panel defects such as climate change, dust, and other foreign substances, and power outages and equipment due to lack of manpower are also being neglected. Due to its characteristics, solar power generation varies greatly depending on climate conditions, and the need for an automatic panel defect detection system is emerging because it is impossible to predict power generation with single data such as temperature, charge, and solar radiation. In order to build a system, accurate predictive research on solar power must be preceded. In this paper, six variables were selected for solar power generation prediction: solar radiation, solar radiation, precipitation, cloud, fog, and fog (yellow dust). By collecting climate data, a power prediction algorithm was selected using solar power information sharing and regional data comparison. Neural networks, SVMs, and deep learning are used as prediction algorithms. When selecting an algorithm, we selected the optimal algorithm using the mean square root error (RMSE), which is most commonly used to identify prediction errors.(Takefuji, 2012; Harrington, 2012; REN21, 2011)

2. Body
Internationally, the annual growth rate of photovoltaic power generation between 2000 and 2014 is increasing the fastest among power generation technologies. Solar power generation worldwide increased from 4 GW in 2004 to 177 GW in 2014. It has increased by 30GW in 2011, 38GW in 2012, and 39GW in 2013, showing more remarkable growth in recent years. When analyzing the top 10 countries by country in Table 1, the United States, China, and Japan show the largest growth. The spread of supply due to new facilities for solar power generation is interpreted as a result of a decrease in the cost of solar power generation. Solar power generation costs have continued to decline over the past 30 years, with a 19.3% learning experience or learning ratio (Figure 1): Every time the capacity doubles, the price falls by 19.3%.
(Figure 1; REN21, 2011)

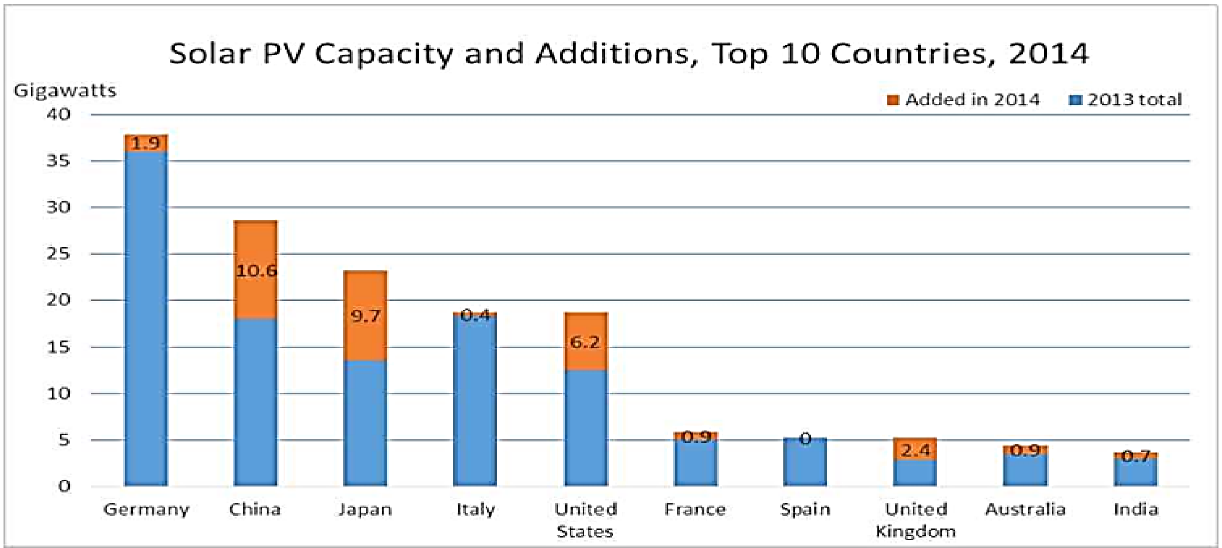

Figure 1. Top 10 countries' increase in solar power generation in 2014

In Korea, as seen in domestic renewable energy production, it has grown rapidly since 2005, and the input of state subsidies is gradually increasing to increase the supply of solar energy. Starting with the 100,000 solar houses distribution project, which began with the government's willingness to supply, the number of solar numbers began to increase, and about 10% of solar power facilities were installed annually after the change in the existing power generation difference system was announced in 2008.(Yoo, 2010).

Despite the increasing demand and necessity for photovoltaic power generation, photovoltaic power generation has problems such as difficulty in constructing technology and power generation facilities, and a lack of related personnel. In general, solar power generation is known to be related to sunlight time, total cloud volume, and solar radiation, but it is not possible to predict power generation status by simple comparison using a single data such as temperature, total cloud volume, and solar radiation.

Support Vector Machine (SVM) is a machine learning algorithm that is used for supervised learning and is mainly used for classification and prediction. If there is data consisting of circles and squares, the vector distance between circles and square data can be measured, and the plane with the maximum distance between the data can be classified as Hyperplane and Linear. This is divided into linear classification possible and linear classification difficult to apply. Data that are difficult to apply linear classification are converted into high-dimensional spaces using kernel functions and classified by returning vector-internal results. Types of kernel functions of SVM include Linear Kernel, Polynomial Kernel, and Radial Basis Function (RBF) kernel, and research on power generation prediction models using RBF.(Harrington, 2012; Sharma, 2011).

Neural Network is an algorithmic system that models the structure and function of the brain in humans or animals, learns the input and output processes accepted from the outside, and predicts the output of the actual input after the learning process. The neural network model may divide the structure into an input layer, a hidden layer, and an output layer. Data are input as values of the input layer and the output layer to learn the weight value of the hidden layer, and the output value according to the input is used for prediction as the learned weight value. A neural network is configured to predict a photovoltaic generator's amount of photovoltaic power. We use climate data as input layer data, construct unit generation values as output layer data, and mainly use Backpropagation algorithms to modify weights by inversely propagating the difference between predicted and actual values to train neural networks. The backpropagation algorithm uses the output of a neural network and the output from that pair and calculates the weight value for each neuron in the hidden layer.(Takefuji, 2012; Harrington, 2012; Sharma, 2011)

Deep learning is derived based on Artificial Neural Networks (ANNs). Although it showed excellent performance in the field of recognition such as image, voice, and cursive writing, it was recognized as unrealistic because it took a lot of time to learn in advance. However, with the advent of powerful GPUs since the 2000s, the time for calculating metrics and vectors has been shortened, and with the advent of big data, a large amount of data can be synthesized and analyzed, and used for learning, increasing interest in deep learning. Deep learning has many kinds of algorithms, so it is necessary to understand and use its characteristics. A deep Neural Network (DNN) is an artificial neural network consisting of multiple hidden layers between the input and output layers and has performed well in the field of language modeling.(Takefuji, 2012; Lee and Lee, 2015; Song, 2014).

Convolutional Neural Network (CNN) is a deep neural network designed to use preprocessing. It shows good performance in video and voice, so it is mainly applied to ARS, an automatic voice recognition service. Recurrent Neural Network (RNN) refers to a neural network with Direct Cycle applied between units constituting artificial neural networks, showing the highest recognition rate in cursive recognition.(Mikolov, 2010).

Data from PVOutput.org was used to obtain solar power generation data. PVOutput.org is a website that shares data for free and compares, monitors, and shares data consumption in real-time. Private businesses provide daily data on solar power plants to PVOutput, and PVOutput provides a comparison with average data through regional classification to make it easier to identify plant problems. According to the distribution of solar power plants in the Americas provided by PVOutput.org, there are 411 in western California and 394 in eastern New York and Massachusetts. According to this distribution, the San Francisco climate has a cool and warm Mediterranean climate in summer and has a constant solar power output, with an average temperature of 18.45℃, a minimum temperature of 9.42℃, average precipitation of 41.7mm, an average precipitation of 4 days. However, in Boston, Massachusetts, the highest-low temperature difference is similar to San Francisco with an average of 15℃ and a minimum temperature of 6.44℃, but the rainfall is 87.85mm and an average of 8.22 days, which is located in the middle of a wet continental climate and a warm humid climate, making Massachusetts's data the most suitable for power generation forecast.(Meteorological Agency, 2008; Chen, 2011).

According to the weather resource analysis report for optimal utilization of solar energy and research by Chen and C, six factors such as solar radiation, precipitation, temperature, total cloud volume, humidity, and fog have the greatest influence on solar power generation. These six variables are selected as input variables for solar power generation prediction and used for prediction experiments. The National Solar Radiation Database (NSRDB) of the NOAA (National Oceanic and Atmospheric Administration), an affiliate of the NCEI (National Centers for Environmental Information), provides temperature, humidity, precipitation, visibility, and total cloud, extracting metadata from Concorde and using it for daily experiments to match solar power data. For solar radiation data, metadata for the same period was extracted from MesoWest, a weather observation database at the University of Utah. In Meso West, there are several observation sites within a city, so more accurate data can be obtained. (Meteorological Agency, 2008; Chen, 2011).

The Concord cape can import data from February 12, 2012, to October 27, 2015, but the actual data from November 3, 2014, to December 2014 are missing. In addition, in Meso West, data is collected semi-annually and uploaded through a verification process, so data from 2015 is not available. 977 data for learning were selected from March 1, 2012, to November 2, 2014. Climate data and solar power data have different units and distributions, so a process of converting them to suit the experiment is needed.

$$\tilde{d}_i \quad := \quad \frac{d_i - \min\{d\}}{\max\{d\} - \min\{d\}}$$

Using the above normalization equation, the data value is changed to have a distribution from 0 to 1 and applied to the machine learning algorithm. Normalized data helps the learning of machine learning algorithms perform well.(Anthony, et al. 2009; Chai, 2014).
The root mean square error (RMSE) value, which is the root square error, is used to determine how far the experimental results and expected values are on average. It means the square root of the average value by squaring the error that appears from the actual value and the predicted value.(Chai, 2014).

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

$y(x_i)$: Actual measurements at x1
$y'(x_i)$: Predictions at x1
n : Number of data

The RMSE follows the formula above. This is a measure of how much difference the actual power generation amount and the predicted value differ on average, and it can be said that the less RMSE, the more accurate the prediction.

In learning, the neural network structure is determined by setting the number of nodes in the input layer, the number of nodes in the hidden layer, and the number of nodes in the output layer. The learning rate is increased by using experimental data normalized to a value between 0.0 and 1.0 as input and output values. There is no clear rule in selecting the learning amount, the number of units of the hidden layer, and the maximum error rate, and the incorrect initial value shows an Overfitting problem, so select the appropriate value and find the optimal initial value through repeated learning experiments. As a learning method, an algorithm is used, and Backpropagation, a method of gradually modifying and transmitting weights from the highest output layer to the lower layer, is used. Climate data of data selected as experimental data are selected as input layers and power generation data are selected as output layers to form a neural network of six input layers and one output layer. The number of hidden layers is 3, 5, 7, 9, and 11, and it has 1000 epochs. In addition, the maximum error rate was 0.0001, 0.001, 0.01, and the learning rates 0.0001, 0.001, 0.01, and 0.1 were selected and proceeded 60 times, and the predicted result was evaluated by calculating the RMSE value(Srivastava, 2014).

Support Vector Machine (SVM) was classified as an algorithm in the field of machine learning and showed excellent performance in regression. The LIBSVM library, developed by National Taiwan University in 2000, is the most widely used SVM software. In this experiment, LibSVM Version 3.2 was used, the nu-SVM model and the Radial Base Function Kernel (RBF-Kernel) were used, and the same input/output data as the neural network was used.

It implements a web-based solar power defect detection system. Deep learning is learned using historical data from Jeju's climate data and Jeju's Bottari energy plant power generation data from the National Climate Data Center of the Korea Meteorological Administration, and the results learned are applied to the detection system. In addition, the defect detection rate is increased by using the panel information inside the plant to determine whether there is a defect. Web Server uses Apache2 and Tomcat7.0, MySql5 as DBMS, Hibernate3.0, Slf4J as Web App Libraries, and Neuroph2.9 to construct deep neural networks.(Meteorological Agency, 2008).

The solar power data of the Botari Energy Power Plant in Jeju City from October 5, 2014, to July 11, 2015, were used as learning data for algorithm learning. Climate data from the same period is extracted from the domestic national climate data center and used for learning. Of the six climate information used, only five data excluding fog data could be extracted, and hourly data is provided, so it is converted into daily data and used for learning by obtaining the highest and lowest data of the year (Figure 2).

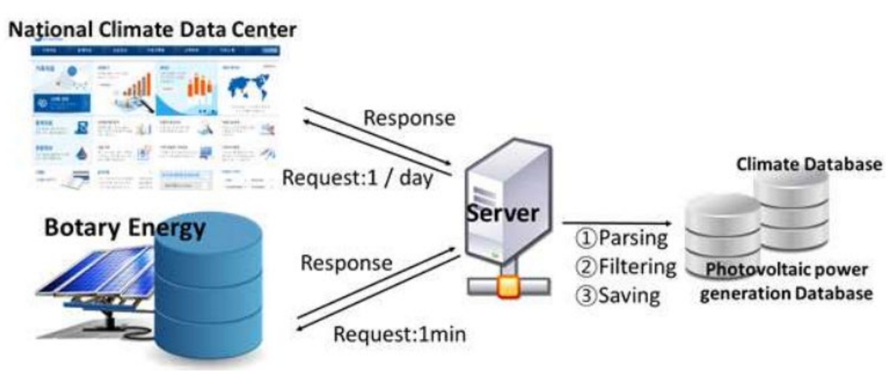

Figure 2. Server for data storage

Using the learned results, a data storage server for a web-based photovoltaic defect detection system is configured as shown in Figure 2. Climate data is requested daily to form a database and plant data is changed in real-time, so data is parsed and stored at 1-minute intervals.

Figure 3 shows the implementation completion screen of the solar power defect detection system. Power generation information for each inverter and module is configured to be monitored on the screen in real-time, and the power generation forecast is displayed as a value using climate data and power generation data from the Korea Meteorological Administration in the prediction algorithm.

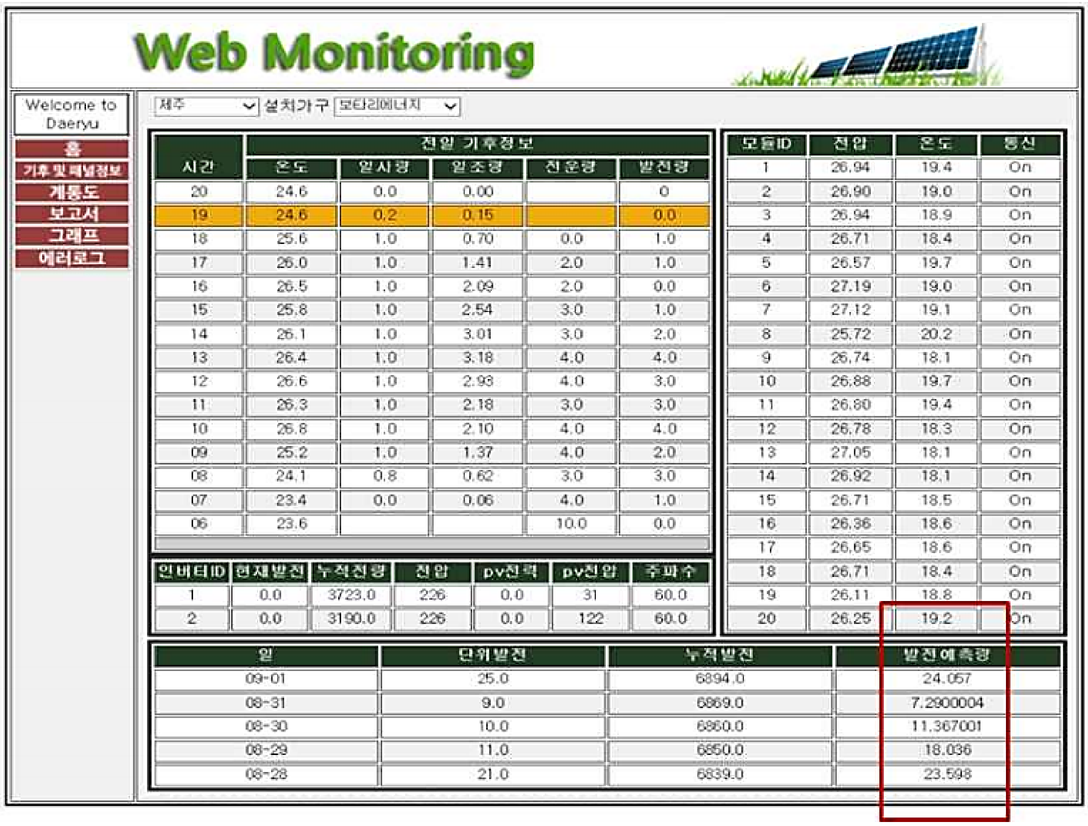

Figure 3. Photovoltaic power generation defect detection system implementation completion screen

3. Conclusion

Due to their nature, photovoltaic power plants have many difficulties in early identification in the event of defects. To detect defects, the amount of power generation was predicted and the defect was determined through comparison with the actual measured amount of power generation. To select the optimal prediction algorithm, climate data and power generation data were normalized and applied to SVM, neural networks, and deep learning. As a result of using the RMSE value as an evaluation method of the algorithm, deep learning showed the best prediction rate. To apply this to domestic data, 280 days of data were obtained from the National Climate Data Center and the Jeju Botari Power Plant and used as learning data. The solar power defect detection system can contribute to the decline and popularization of solar power plants in Korea by predicting power generation using learned deep learning algorithms and determining defects using panel data.

References

Takefuji, Y, Neural network parallel computing, *Springer Science & Business Media,* Vol. 164, 2012.
Harrington P., Machine learning in action, *Manning,* 2012.

NGIAM, J., "Multimodal deep learning," *Proceeding of ICML-11 conference on machine learning*, pp. 689-696, 2011.

REN21, *Renewables 2011 global status report*, REN21 Secretariat, 2011.

Yoo k. J., ″Solar power development status″, *Electrical Journal*, pp. 39-44, 2010.

Sharma, N., Sharma, P., Irwin, D., Shenoy, P., "*Predicting Solar Generation from Weather Forecasts Using Machine Learnin*g," 2011.

Lee S., Lee W., ″Panel defect detection method based on back-propagation algorithm using climate data and photovoltaic data″, *Spring Conference of the Korea Information Processing Society meeting papers,* Vol. 22, No. 1, 2015.

Song J. J., Lee S. H., Jung Y. S., ″Analysis of power generation prediction model for solar power generation″, *Digital Policy Research*, Vol. 12, No. 3, pp. 243-248, 2014.

Mikolov, T., "Recurrent neural network based language model," *Proceeding of INTERSPEECH Conference*, pp.26-30, 2010.

Meteorological Agency, *Meteorological Resource Analysis Report for Optimal Utilization of Solar Energy*, 2008.

Chen, C. S., "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.

Anthony, M., L. Peter. B., Neural network learning: Theoretical foundations, *cambridge university press,* 2009.

Lee. D. W., Seasonal Thermal Energy Stroage System*, The Magazine of the Society of Air-conditioning and Refrigerating Engineers of Korea*, Vol. 45, No. 9, pp. 14-21, 2016,

Chai, T., Roland, R. D., "Root mean square error (RMSE) or mean absolute error (MAE)–Arguments against avoiding RMSE in the literature," the literature," *Geoscientific Model Development*, Vol. 7, No. 3, pp. 1247-1250, 2014.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp.1929-1958, 2014.

Biographies

Lee Ji Yun is student in MY PAUL SCHOOL. She is interested in artificial intelligence, deep learning, cryptography, robots, mechanical engineering, block chains, drones, autonomous vehicles, etc., and is conducting related research.

Shin Dong Ho is Professor and Teacher in MY PAUL SCHOOL. He obtained his Ph.D in semiconductor physics in 2000. He is interested in artificial intelligence, deep learning, cryptography, robots, block chains, drones, autonomous vehicles, the Internet of Things, metaverse, virtual reality, and space science, and is conducting related research.