Data Management System for Photovoltaic Solar Environment

H Tak, D Lee, H Kim and H G Cho

1Department of Electrical Computer Engineering, Pusan National University, Busan, Korea

Email: hgcho@pusan.ac.kr

Abstract. As communication environments and Internet-of-Things technologies evolve, smart grid is being studied for collecting and predicting facility data in real time. In this paper, we propose a management system that collects the data of a photovoltaic system and predicts the amount of power generation. For this purpose, we designed an environment that can collect solar data, adjacent environment data, and facility information in real time. In addition, to expand the limits of previous power generation prediction studies, a power generation prediction experiment was conducted using recurrent neural network-based methods. The experimental results are compared with those of previous studies.

1. Introduction

Smart grid [1] is attracting attention as next-generation power grid for solving the problems of power demands, environmental pollution, and resource depletion. Studies on optimizing the smart grid network and predicting energy generation are actively being carried out. In this study, our aim is to predict solar power generation, which is the most widely used source of renewable energy [2].

Existing solar power generation prediction studies employ regression analysis or machine learning techniques. However, these studies have problems in that the length of the training data is fixed and the prediction performance is poor. Currently, deep learning techniques are widely used for prediction, but there are few cases of using neural network techniques to predict solar power generation. The recently developed neural network techniques can use historical data of various lengths as input data. As a result, they can flexibly cope with the length of training data while using complicated variables.

In this paper, we propose a management system that can measure the power data produced by a power facility and predict the solar power generation using information about the solar panels, related equipment, and environment. To do this, we describe the database design, data accumulation method, and solar power generation forecasting model that are necessary for constructing the management system. In addition, an optimized prediction model using three recurrent neural network (RNN) methods and environmental data sets was evaluated experimentally.

2. Related Work

Previous studies on smart grid and solar power generation have been conducted. Hosoda and Namerikawa [3] clustered solar power data using the type of weather, and predicted power generation using a regression equation for the grouped data. Yona et al. [4] modeled solar radiation, temperature, atmospheric solar radiation, and humidity over 24h period using artificial neural networks, RNNs, and...
predicted solar power generation accordingly. Gensler, Henze, and Sick [5] predicted solar power generation using a long short-term memory (LSTM) network, which is a type of RNN.

In the system described in this paper, data from the solar module and environmental information are collected using sensors. Hence, we also reference a study using environmental sensors [6].

3. Management System Architecture
The proposed management system consists of a data extraction module, data analysis module, and solar power generation prediction module. The data extraction module measures the data of the solar module, environmental sensor, and DAB converter. The measured data are sent to a database over Wi-Fi. The data analysis module calculates data statistics and determines abnormalities using past data. The solar power generation prediction module predicts power generation using past data and RNNs. Figure 1 shows the architecture of the proposed system. As the data are accumulated, the prediction module predicts the amount of power using a model and data of various length. The prediction results are stored in the database. In the proposed system, it is possible to check the data measured at each facility, perform statistical analysis, and evaluate the results of power generation prediction.

4. Database Construction and Data Gathering
In the data management system proposed in this paper, solar module, DAB converter, and environmental sensor data are collected. The procedure is shown in Figure 2. Each measurement module consists of an Arduino and Wi-Fi shield. The measured data are stored in the database remotely using PHP. The data generated from the solar panel are stored in the Solar Database, the data measured in the environmental sensor is stored in the Weather Database, and the data generated in the DAB converter is stored in the DAB Database. Each data is stored in the server's database after time and place information is added.

5. Prediction System for Power Generation System
For solar power generation prediction, we used the mean absolute percentage error (MAPE) and mean squared error (MSE) to evaluate the forecast results. MAPE and MSE are obtained by calculating the
difference between the predicted result and the actual result using the absolute value or square and then calculating the average value as follows:

\[
\text{MAPE} = \frac{100\%}{n} \cdot \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|
\]

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)
\]

Here, \( n \) is the number of data used in the evaluation, \( Y_i \) is the actual value, and \( \hat{Y}_i \) is result of prediction.

To optimize the solar power generation prediction module, four weather factors that affect power generation were selected: temperature, humidity, solar irradiance, and duration of sunshine. For the prediction experiment, we designed five data sets shown in Table 1 using different combinations of factors. A comparison experiment was conducted using three RNN models: a basic RNN, a LSTM network and a gated recurrent unit (GRU), which is a network with excellent learning ability for time series data.

| Data Set | Input Data Parameter |
|----------|----------------------|
| 1        | Photovoltaic power generation |
| 2        | Photovoltaic power generation, solar irradiance, and temperatures |
| 3        | Humidity and duration of sunshine |
| 4        | Solar irradiance and temperatures |
| 5        | Solar irradiance, temperatures, humidity, and duration of sunshine |

6. Experiment
In this study, we trained the three RNN models using historical data that varied in length from one hour to three days. If the data exceeded three days in length, there was no fluctuation in prediction rate, so longer periods were not evaluated. Figures 3–7 show the variation in MSE and solar power generation prediction results for each RNN and various lengths of historical training data for the data sets listed in Table 1.

There are differences in the results for each data set, but the error rate generally decreases when data for the past 24 h are used. However, the model overfit the data if historical data of a certain length are used. The experimental results show that the LSTM networks trained on 24 h of historical data obtain the best results. Figure 8 shows that the error rate is lowest when predictions are made by inputting temperature, humidity, solar irradiance, and duration of sunshine.

In existing studies, the prediction results are evaluated using MAPE or MSE. However, because the previous models and proposed model cannot be tested on the same data, the results reported by the authors on their own data are shown. The results are compared in Table 2, which shows that the error rate of the proposed prediction model is low.

![Figure 3](image-url)  
*Figure 3. Prediction results for data set 1. (a) MSE with respect to the length of the historical training data. The error rate for LSTM with 24h of training data is the lowest (0.0024). (b) Forecasted results of LSTM (trained on 24h).*
Figure 4. Prediction results for data set 2. (a) MSE with respect to the length of the historical training data. The error rate at LSTM with 24h of data and GRU with 21h of data are the lowest at 0.0022. (b) Forecasted results of LSTM.

Figure 5. Prediction results for data set 3. (a) MSE with respect to the length of the historical training data. The lowest error rate (0.0022) was obtained by LSTM with 24h of data and GRU network with more than 24h of data. (b) Forecasted results of LSTM.

Figure 6. Prediction results for data set 4. (a) MSE with respect to the length of the historical training data. The error rate for LSTM with 24h of data is the lowest (0.002). (b) Forecasted results of LSTM.

Figure 7. Prediction results for data set 5. (a) MSE with respect to the length of the historical training data. The error rate of LSTM with 24h of data is the lowest (0.0019). (b) Forecasted results of LSTM.
Figure 8. Comparison of the prediction error rate of LSTM with 24 h of past data on each data set. In data sets 2 and 3, the duration of sunshine is considered, but overfitting occurs and the error rate increases. Using the temperature and solar irradiance data together lowers the error rate with respect to the use of past photovoltaic power generation data alone. The lowest error rate is obtained when all four weather factors are used.

Table 2. Performance Comparison with existing Photovoltaic Power Research

| Author                          | MAPE | MSE   |
|---------------------------------|------|-------|
| Hosoda and Namerikawa [3]       | 13.87| N/A   |
| Yona et al. [4]                 | 14.97| N/A   |
| Gensler, Henze, and Sick [5]    | N/A  | 0.0049|
| Proposed Model                  | 7.29 | 0.0019|

7. Conclusion and Future work
In this paper, we proposed a data management system for smart grids. To do this, we introduced the modules needed to implement such a system. In addition, a prediction model based on an RNN was proposed and evaluated by experiments on historical solar power generation data of different lengths. LSTM was found to have excellent prediction performance, and it was confirmed that all four evaluated weather variables (temperature, humidity, solar irradiance, and duration of sunshine) were influential. In the future, we will implement an algorithm for optimal power allocation using the prediction model proposed in this study.

8. Reference
[1] Sung D, Song N, Go G, Cha J, Bae G and Jang H, 2013, Communications of the Korean Institute of Information Scientists and Engineers, 31, 10
[2] Lee M and Kim S, 2013, Journal of Korean Institute of Information Technology, 11, 181
[3] Hosoda Y and Namerikawa T, 2012, SICE Annual Conference, 119
[4] Yona A, Senju T, Saber A T, Funabashi H, Sekine H and Kim C, 2007, Intelligent Systems Applications to Power Systems, 1.
[5] Gensler A, Henze J and Sick B, 2016, IEEE international conference on systems, man, and cybernetics (SMC), 2858.
[6] Ausilio A, 2012, Behavior research methods, 44, 305.

Acknowledgments
This research was supported by Basic Research Laboratory through the National Research Foundations of Korea funded by the Ministry of Science, ICT and Future Planning (NRF-2015R1A4A1041584).