Prediction of Manufacturing Plant's Electric Power Using Machine Learning

*Kyoe-Rae Yeom, Hyo-Sub Choi
Embedded SW, KETI(Korea Electronics Technology Institute)
22, Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do,
Republic of Korea
e-mail : {*duarufp06, hschoi}@keti.re.kr

Abstract

In this paper, we apply the data accumulated through E-IOT platform to machine learning method to find significant variables first and predict the electric power generated in manufacturing process by using these variables. Pre-processing such as resampling of data was carried out before the prediction. In order to select the significant variables, 25 variables were derived using Lasso (least absolute shrinkage and selection operator), one of the machine learning techniques. We used Deep Learning 's LSTM technique, one of the field of machine learning for the prediction.

I. Introduction

According to the Korean Energy Economics Institute in 2014, the manufacturing industry accounted for 59.4% of total domestic energy consumption.[1]. In addition, the rate of increase in industrial electricity rates has risen sharply to 40.8% since 2010, raising manufacturing costs[2]. Therefore, there is a great demand for energy saving in the manufacturing industry, and various measures have been implemented for energy saving.

The development of IoT (Internet of Things) technology has improved the conditions for collecting various data in the process, leading to an increase in data size. This creates the opportunity to apply large-scale data to machine learning techniques to find energy savings points.

In fact, Google DeepMind has consistently accumulated a variety of data through thousands of sensors installed in the data center. The accumulated data was used to train predictive models through machine learning techniques[3]. The trained model was used to recommend when data center cooling is required. This model reduced energy use by 40%.[4].

As in the case of Google Deep Mind, we have built a platform that continuously accumulates real-time electric power data and PI(plant information) data that occurs on the manufacturing process. We call this platform as E-IoT platform. In this paper, We first analyze which of the PI data accumulated from the E-IoT platform is closely related to electric power. And we use these PI data to create a prediction model for future optimization.

II. The Main Issue

A. Characteristic of the data for analysis

In this paper, the process of the manufacturing factory to be analyzed can be divided into three stages. The data to be analyzed here are the variables corresponding to the first stage process (hereinafter referred to as A process). These variables are 1556 in total, and have the following characteristics. The amount of data collected is 15 GB, and the period is from July 2016 to June 2017. This data is displayed in seconds. Typical features of this data are the flow, concentration, storage level, pressure, temperature and electric power of the individual process areas.

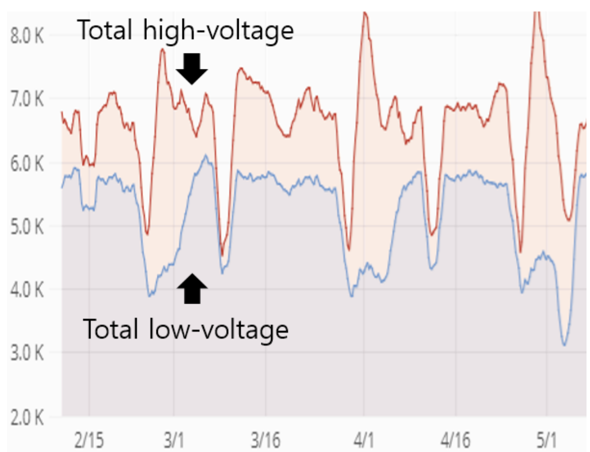

Fig. 1. Total high-voltage is always higher than low total low-voltage.

Among them, the target feature used for feature selection and predictive analysis is the total high voltage electric power corresponding to the A process. As shown in Figure 1, the total high-voltage electric power is always larger than the total low-voltage electric power, and the variation is larger, so that there is a lot of room for electric power saving.

B. Pre-Processing

Although there are 1556 variables, Due to the performance limitation and difference in frequency of variables, we decided to reduce the variables to improve the analysis performance. The figure below shows how the difference in data frequency is severe.

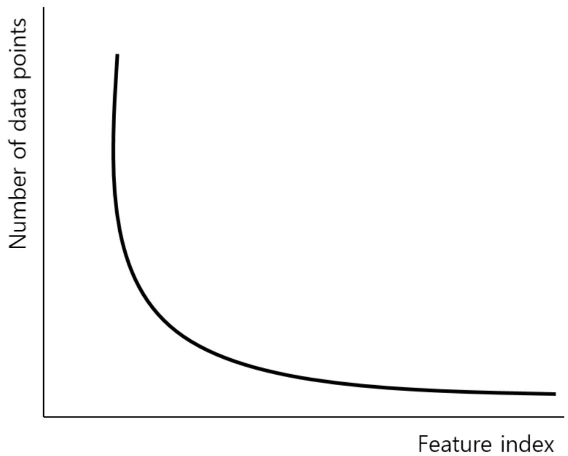

Fig. 2. Number of data points when sorting variables in descending order

When variables are sorted in descending order by the number of data points, the variables representing the steep slope of the graph are mostly those for electric power. Since only a large number of data can't be used as a material for predictive analysis, we decided to resampling the data in 1-minute increments in order to obtain data for analysis.

For example, if one feature has one data point per minute, there are 525,600 data points per year. The smaller the number, the more likely the data will be filled with the average value through resampling. The more likely it is that there is no data value in a minute. To prevent data from being deformed by resampling, we decided to exclude variables with less than 525,600 data points from analysis. Therefore, 1323 out of a total of 1556 were excluded and 232 were selected for analysis.

C. Feature selection by Lasso

We used Lasso (Least Absolute Shrinkage and Selection Operator) technique in the machine learning method to derive meaningful variables from the 232 variables and use them for prediction analysis.

The Lasso technique is a kind of regression model, in which the coefficient of a variable determined to be unnecessary in a model is converged to 0 and excluded from the model.

This method uses L1, one of the regularization methods. As the alpha value corresponding to the penalty of the model increases, the number of variables selected in the model decreases. As the alpha value decreases, the number of variables selected in the model increases. Lasso is a useful method for feature selection because it can automatically select a meaningful variable[5].

We set the alpha value to 0.03 in the Lasso model and derive 25 out of a total of 232 variables. As a result, the 25 variables are as follows.

- 11 variables for flow rate
- 9 variables for electric power
- 3 variables for tank level
- 2 variables for concentration

D. Prediction by LSTM

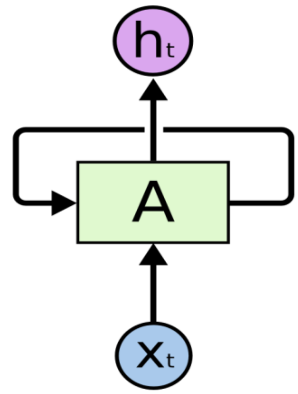

Fig. 3. Basic structure of RNN

LSTM(Long Short-Term Memory) is a type of RNN, which is a sub-sector of machine learning. The RNN, which is useful in the field of deep learning, reflects the structure that past data can affect in the future as shown above Fig. 3(It is formed in a loop so that past data can affect the future)[6]. In other words, it is an excellent tool when events occur consecutively.

However, the basic RNN has a problem of long-term dependency. The problem is that the model's performance drops sharply when the distance between input information that the model needs to learn is large. This problem was solved by introducing Cell State and three kinds of gates into LSTM[7].

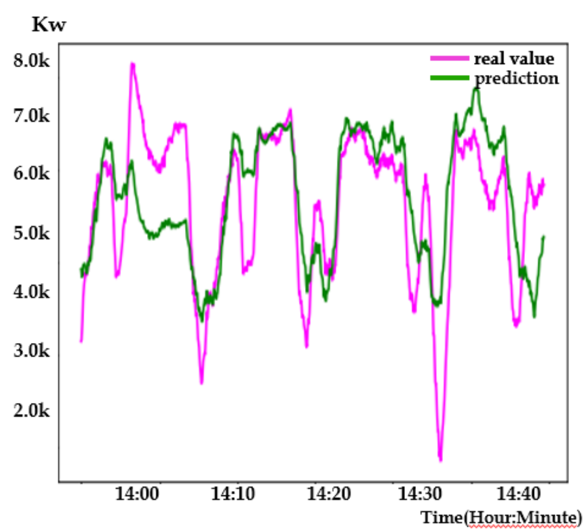

Fig. 4. Expanded Graph of Linear Regression
(The less fluctuating line is the predicted)

Figure 4 above shows the result of linear regression before predicting with LSTM model. The predictions of linear regression appear to predict electric power fluctuations in general. However, when the power fluctuates rapidly, the predicted value is almost wrong. Unlike RNN, linear regression does not reflect past data volatility. In linear regression, only input values are considered.

The characteristics of the LSTM model in this paper are as follows. Each of the 25 feature variables resampled in 1 minute has 525,600 data points (1 year). And 25 total variables were scaled between 0 and 1. This LSTM model has a learning rate of 0.01 and has 5 hidden layers, 10 sequence lengths, and 100 batch sizes. 70% of the total data was used as training data and the remaining 30% as test data.

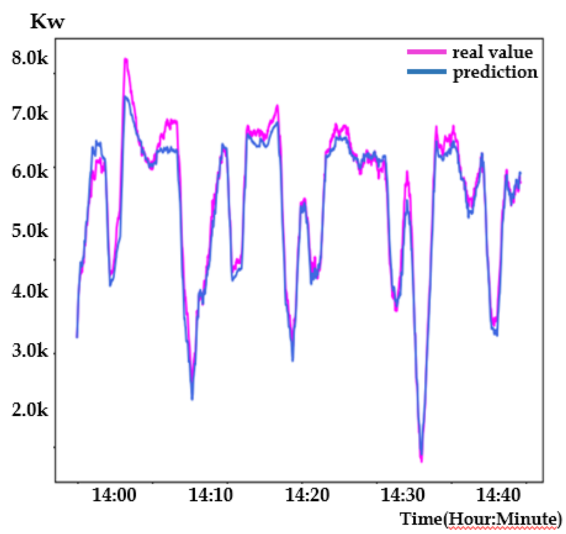

Fig. 5. The predicted and actual values of LSTM model

The prediction results are shown in Fig. 2. The LSTM high voltage electric power prediction model showed a mean absolute error of 0.07 and an accuracy of 79%. The prediction model will be used for future electric power optimization.

III. Conclusions

In this paper, we use the Lasso method to derive the variables closely related to the total high - voltage electric power using the PI variables generated in the process part, and predict the future electric power usage by using the derived parameters in the LSTM model. However, using the target value as the total high-voltage electric power, it was not possible to grasp the individual trends of the more detailed electric power of some processes. PI data is expressed in the details of the process, while the target value, total high-voltage electric power, is derived by integrating the A process.

The analysis direction to go forward is as follows. In the future, before proceeding with the process optimization, we need to get a closer look at the detailed process of the A process and divide the total electric power into the detailed electric power to conduct a deeper analysis. In addition, we need to perform sensitivity analysis on the electric power derived from the above feature selection for each detail electric power to see which variable responds most sensitively. Finally, based on field experience, we will collect and accumulate more important variables to create a more sophisticated model.

Acknowledgment

This paper was conducted as part of the energy technology development project of the Ministry of Trade, Industry and Energy and the Korea Energy Technology Evaluation and Planning. (20172020108780: Development and Demonstration of Energy-IoT Platform Technology for Energy Saving in Base Industry)

References

[1] Korean Energy Economics Institute, “2014 Energy Consumption Survey”, 121p, 2014.
[2] Kim Dae-Wook, Kim Kwang-In, Choi Woo-Jin, “A study on the Adequacy of the Industrial Electricity Price in the Korean Electricity Market", Journal of Regulation Studies, 24(2), pp. 31, 2015.
[3] Jim Gao, “Machine Learning Applications for Data Center Optimization”, Google, pp. 1-2, 2014
[4] Richard Evans, Jim Gao, “DeepMind AI Reduces Google Data Centre Cooling Bill by 40%”, Available at https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/
[5] V. Fonti, E. Belitser, “Feature Selection using LASSO”, VU AMSTERDAM, pp. 4-5, 2017.
[6] “Understanding LSTM Networks”, Available at http://colah.github.io/posts/2015-08-Understanding-LSTMs/
[7] “Long Short-Term memory”, Available at https://en.wikipedia.org/wiki/Long_short-term_memory