The Golden Academy Conference Series

2017 International Symposium for Advanced Computing and Information Technology (ISACIT 2017)

Aug 18 – Aug 21 2017

Conference Proceedings

Proudly presented to you by

APSCIT

Asia Pacific Society for
Computing and Information Technology

3. Introduction to the Lossless Source Coding and the Context Tree Weighting Method (T. KAWABATA)
4. Hybrid uplink traffic scheduling algorithm in Fixed Mobile Convergence (FMC) networks: A Comparative Study of Performance (I.S. HWANG)
5. Data analytics and simulation tools for urban mobility of the future (J. DAUWELS)
6. Efficient Tools to Implement Web-Based Systems (L.H. CHANG)
7. Robust texture image representation by scale selective local binary patterns (Z.H. GUO)
11. Error Data Analysis of the Photovoltaic Energy Monitoring System with the confidence interval of Multivariate Linear Regression (E.Y. BYUN)
12. Metamodel based Photovoltaic Monitoring System for Heterogeneous Renewable Energies (W.S. JANG)
13. Analysis of Integrated Renewable Energy Monitoring System Data using KNN for Pre-Processing (J.H. LEE)
14. Improvement for Effective Commination Cost of Solar Energy Integrated Monitoring System based on Long Range, Low Power Wireless Platform (LoRa Technology) (B.K. PARK)
15. A study on Failure Judgement Method of Photovoltaic System based on Logistic Classification (J. PARK)
16. An Experimental Practice with the lightweight test maturity model to improve test process of Korean Small & Medium Sized Companies (K. KIM)

Sat/8/19

No. 2 Meeting Room

Abstract ID: 11

Error Data Analysis of the Photovoltaic Energy Monitoring System Using the prediction interval for Multivariate Linear Regression

Eun Young BYUN, So Young MOON, R. Young Chul KIM, Hyun Seung SON, Hongik University, Korea

Abstract

Today, it is difficult to apply classification and clustering to the Photovoltaic monitoring system because of the low frequency of the error data. To solve this problem, we use the multivariate linear regression of normal data. It measures the average and standard error for normal data and predicts response variable(y) for a new data. In this paper, we analyze the error data using the difference between an actual and predicated response variable for a new data. To clarify this difference, we use prediction interval of predicated response variable. It defines a range of normal data excluding outliner. If new data is contained in this area, it is most likely normal data. On the other hand, if a new data is not contained in this area, it is most likely an error data. For the evaluation of this approach, we confirm that actual error data are contained in this prediction interval. The result is that error data are not contained in this area. Therefore, we will detect an anomaly of this system, rapidly. In the future, we will predict the error by adding information that can be collected from this system.

Keywords: Photovoltaic monitoring system, Multivariate linear regression, Error data analysis, Prediction interval

1. Introduction

Recently, various data have been generated by the development of ICT (Information and Communication Technologies). Therefore, there is increasing the influence of big data [1]. Companies can gain three main effects by using big dataas follows: 1) detecting an anomaly quickly through the analysis of the data pattern, 2) predicting the future, and 3) identifing exactly the problem on the present situation and then suggest a solution [2]. It allows effective management of the system. Especially monitoring system based on big data is being introduced to increase the generating efficiency of Renewable Energy. Among various energy sources, the number of photovoltaic power generation plants have increased sharply. Therefore, a stable operation and reliability of this plant is important [3,4]. In this paper, we propose a strategy to detect an anomaly of the system by analizing the real-time data of the photovoltaic monitoring system developed during the project. This strategy aims to increase generating efficiency by optimizing this system. However, a problem is the imbalance between normal data and error data in this system. To solve this problem, we describe the strategy using multivariate linear regression and developed them using Tensorflow [5-6].

Section 2 describes some problem of imbalanced data in related studies, and section 3 defines the multivariate regression strategy for error data analysis. Section 4 evaluates out strategy. Finally, provides the concluesions and direction of future work.

2 Related Work

The imbalanced data causes an overfitting of high frequency data. The result of the learning has more than 90% accuracy, but not properly trained(because of overfitting). There are ways to solve this problem :

- Try Collect More Data
- Try Changing Your Performance Metrics
- Try Resampling Your Dataset
- Try Different Algorithm
- Try Penalized Models

The simplest way is collecting more data. However, there is no difference in proportion of data. Besides, there are many other ways. In this paper, we perform different algorithm called the regression analysis, not classfication/clustering.

3. Multivariate Linear Regression Analysis Strategy

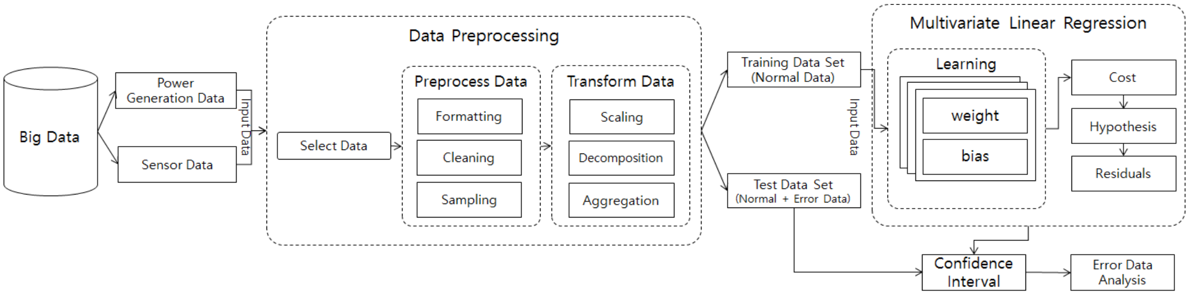


Figure 1 The Structure of Error Data Analysis

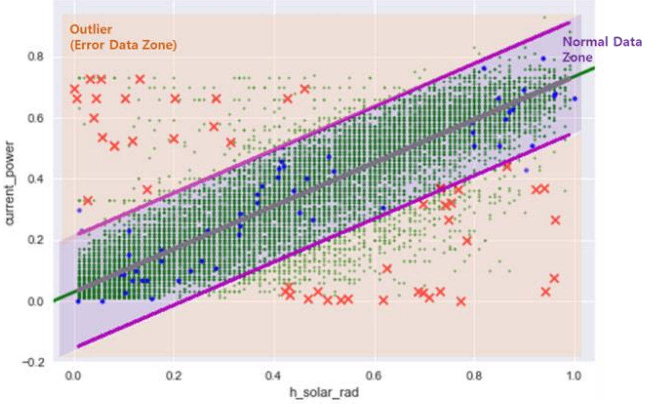
Fig. 1 shows the structure of the error data analysis using the multivariate linear regression. First, we preprocess the data according to the algorithm. The following is data preprocessing steps :

- Step 1(Select Data) : We specify the dependent variable as the power, and specify the independent variable as the month, times, horizontal solar radiation, slop solar radiation which have a effect on the power.
- Step 2(Preprocess Data) : We remove data with zero power in Cleaning work.
- Step 3(Transform Data) : We normalize data in scaling work and decompose data of Date type into month and time.

After, we create training/test data set, and define a regression line and prediction interval using the multivariate linear regression. We define the normal data and error data zone based on the prediction interval. It is hard to graph regression line based on many independent variable. We graph a regression line that is the relationship between power and horizontal solar radiation which has the most effect on power at Chater 4. The prediction interval is the normal data zone, and the other interval is the error data zone. We will analyze the error data by confirming whether a new data is including in error data zone or not.

4 Evaluation

We evaluate the strategy using the test data sets. Fig. 2 shows the evaluation result. We can confirm the distribution of normal data(o) and error data(x) and measure the accuracy. There is a possibility to exist 5 percent of error data in prediction interval, because it is the range that 95 percent interval based on overall data.



n = 110	Predicted: Normal Data	Predicted: Error Data	
Actual: Normal Data	57	3	= 60
Actual: Error Data	1	49	= 50
	= 58	= 52	

Accuracy = (57+49) / (57+3+1+49) = 0.96

Figure 2 Evaluation Result

5 Conclusion

In this paper, we analyze the error data in big data of photovoltaic monitoring system, and detect an anomaly rapidly. Therefore, we will increase the generating efficiency. However, it is difficult to apply classification and clustering because of the imbalance between normal and error data in this system. To solve this problem, we use multivariate linear regression on normal data and the prediction interval for analyzing error data. In future work, we will detailed analyze big data using the inverter output current, inverter output voltage, inverter input current, inverter input voltage, etc. that can be collected from this system. Based on this, we will expand not only post-processing but also preprocessing for the failure of this system by predicting errors. Error prediction will be study to support not only post-processing but also preprocessing for the failure of this system.

Acknowledgements:

This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea 2015-2018 (NRF-2015H1C1A1035548).

References

1. Lee Sang Un, Lee Jung Gyu, ”A System of Smart Integrated Monitoring and Analysis Based on Big Data”, The Korean Institute of Broadcast and Media Engineers, pp. 106-109, July, 2015
2. Cheon Hong Mal, ”The Development Success Factors for Integrated Big Data Monitoring System”, Journal of Culture Industry, Vol. 15, No. 2, pp. 133-142, June, 2015.
3. Korea Energy Agency, ”2016 Renewable Distribution Statistics”, 2016.
4. Korea Energy Agency, ”2016 New&Renesable Energy White paper”, 2016
5. M Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, “Tensorflow: A system for large-scale machine learning”, USENIX Association, 2016.
6. M. Abadi, A. Agarwal, P. Barham, E. Breydo, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”, 2016.

Notes