Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 6 Issue 6, June - 2019

The Imbalanced Data Processing for the Data Stream Processing in the Smart Distribution Panel

Donghyok Suh[1]
Department of Display Engineering, Dankook University, Cheonan-si, Chungnam, 31116, Korea
hanhwaco@naver.com

Moonsu Lee[2]
Samduk Electrical CO., Ltd.
Kimpo-si,10117, Korea

***Abstract*—In this paper, we propose a method to manage the imbalance data in the smart switchboard system. Instead of applying uniform over-sampling or down-sampling, we propose a method to adjust the low - weighted data in the contextual information, while retaining the weighted data in the acquisition of contextual information. Considering that the goal of the smart system is to recognize the situation in the real world, it is reasonable to process the imbalance data considering the weight in the context information acquisition, and measure the depth of event in the data stream to support it.**

Keywords— IOT, Smart Panel, Context Inference, Imbalanced data, Data Processing

I. INTRODUCTION

In recent years, researches on data processing in the Internet environment of objects have been actively conducted. This network system has the advantage of being able to continuously monitor in real-time so that it is widely utilized in the field of detecting abnormal situations change quickly and responding to countermeasures corresponding to it.

Whether a sensor network system or an object internet system, a sensor is attached to the terminal of the network. It is based on the sensing activity of the sensor and the data generated as a result.

There is a problem in this case. In case that different sensors are used together in any system, the occurrence frequency of event data of the sensor may be different from each other. If a sensor reports an event once every 10 seconds, some sensors can report an event once every 20 seconds and once every 30 seconds. In this case, when the data stream is divided and analyzed at a constant time interval in the environment where the data stream continuously flows, the number of event data among the sensors is different. This case was called an unbalance between classes by previous data processing researchers.

To realize the Smart Switchboard System, various sensors can be installed inside and outside the Smart Switchboard. In this case, the frequency of events on the data stream reported by the sensor is different and eventually, the data set of the sensors becomes unquantified data. Even in the case of an event that detects a situation involving a very serious problem, the incidence is very low. But the importance of the event can be very high.

This study suggests how to deal with unbalanced data generated in the data stream environment in the Smart Switchboard System. Through this, we intend to contribute to more effective data processing by securing the data processing technology that is the basis for developing Smart Switchboard.

This paper is organized as follows. Section 2 summarizes the related research and section 3 describes what we suggest in this study. In section 4, the proposed method has experimented and the results are evaluated. Conclusions are given in section 5.

II. RELEVANT STUDIES

A. Real time data stream processing

Sliding windows are used for real-time processing and analysis of streaming data. In the sliding window technique, some or all windows overlap. Performance is degraded because redundant data is processed redundantly. Hadoop is a batch processing system that is effective for batch processing large amounts of data but does not support real-time streaming processing [1]. Spark of an in-memory-based distributed data processing system provides large-scale processing of real-time data to solve streaming processing not supported by Hadoop [2]. There is a problem in that the load and performance of a real-time node are not considered in a distributed memory environment. The Slider technique supports streaming data processing but does not take into account the load and performance of real-time nodes [3].

The other four users save the result of the previously computed data in memory and reuse the data to process the query, so that duplicate operation of the data is not performed. Sets input data of a certain time interval to a window and processes it in units of windows. When the streaming data is input, the SHA-1 algorithm. And stores the hash key value in the form of key value in a temporary queue. Data is allocated to each node by a job scheduler. The assigned data is processed by each node and stored in memory. If there is a previously processed result, duplicate operations can be avoided because data is taken from memory and reused [4]. Stream Reasoning technology is very useful for analyzing big data transmitted from smart devices on the Internet and providing intelligent services through accurate situation

recognition. Stream-rearing is logical reasoning that processes large-data streams mixed with many heterogeneous and bad data in real time to support decision-making systems used by multiple users at the same time [5]. Stream-rezzing is performed using real-time distributed processing based on cloud computing.

B. Smart Switchboard Safety Inspection System

The switchboard is a very important facility because it is responsible for the power of all facilities. ARC may occur in a high voltage switchboard, and there is a high risk of explosion or fire due to ARC. The shipboard switchboard shall be able to withstand the electrical and mechanical stresses caused by shock and vibration. And power switches and other power switches and similar devices [6] [7]. Park and others have installed one main board and several sensor boards in MV (Medium Voltage) to monitor the ARC occurrence in real time for the corresponding sector in the switchboard. The sensor board receives the command of the main board and senses the arc, temperature, and humidity at a specific time, transfers it to the main board, and sends an interrupt signal to the main board in the form of an interrupt when the set ARC value is exceeded [8].

There are various causes of the arc flash that occur inside the switchboard such as bad contact of the bus, tracking, short circuit, animal invasion, and operator error. Arc flash accidents are the biggest accidents in the middle of the switchboard accident. If the duration is long, it causes a breakdown of the power panel and the power panel. It also causes a danger to the adjacent user as well as power failure and fire due to equipment failure. The temperature of the arc flash caused by an electric accident is more than 35,000 degrees, and it is impossible to measure the temperature by using the sensor because it is close to the solar surface temperature [9]. Lee, Sangjun et al. Designed a miniature metal enclosure switchboard with improved arc characteristics. The arc generated in the inside was quickly discharged to the arc discharge port designed at the upper part, thereby preventing spill damage and personal injury. By limiting the hot gas and heat emission angle, the safety of the user is considered and the damage inside the switchgear is minimized [10].

III. THE IMBALANCED DATA PROCESSING FOR THE DATA STREAM PROCESSING IN THE SMART DISTRIBUTION PANEL

The smart system recognizes the situation by relying on the data obtained from the sensing result of the sensor and the event information included therein. The sensed data that the sensor detects and reports to the host is continuously transmitted in the connected state and thus has a stream form. Among the incoming data, events that are useful for context recognition are included. The problem may be that the frequency of events varies depending on the sensor. In data processing, there are many cases in which the size of a data set to be analyzed is the same. Therefore, there is a need for a data processing method in the case where the sizes of the data sets are different from each other.

Unbalance data per sensor

Various sensors can be installed in the switchboard and electrical equipment. A temperature sensor can be installed, and an infrared sensor is installed where it is difficult to directly measure the temperature or where it is difficult to install the temperature sensor. It is possible to detect the state of electricity flowing through the electric line through the current sensor. A vibration sensor can be installed to monitor the mechanical stability of the switchboard and electrical equipment. In this case, depending on the situation, the sensors may have different frequencies of events. Events from temperature sensors are frequently reported when the heat is generated inside or outside the switchboard. However, the vibration sensor may have a low frequency of events. Conversely, in other situations, the event of the vibration sensor is frequently reported, but the event of the temperature sensor may not occur the same as the vibration sensor.

Different frequencies may give clues to infer the situation. However, this study presupposes conditions that have the same frequency. If there are differences in dataset size between two datasets of the same size, there are three ways. The first is to reduce many datasets to the same size as the smaller datasets. Downsampling. The second is to increase the size of the smaller dataset. oversampling. The third is bi-directional sampling.

In this study, we propose that if two datasets are different in size, we should adjust the sampling to the other dataset with the more meaningful dataset as the original. In other words, to recognize the situation, the side that detects the change of the weight is changed to the original data, and the data set to be compared is downsampling if the data amount is large, and oversampling if the data amount is small.

Such a method is valid. It is better to analyze the original data as it is important to recognize the situation and contribute to it and allow a little adjustment for low weight data. This is based on the assumption that the size of the dataset should be adjusted and if the dataset size differs under these conditions.

The proposed algorithm is as follows.

1. *ds1* is the dataset transmitted by sensor *S1*.

and *ds2* is the data set transmitted by the sensor *S2*.

ev1n (n = 1,2,3, ⋯ n) is an event in *ds1*.

ev2n (n = 1,2,3, ⋯ n) is an event in *ds2*.

2. | *ds1* | = | *ds2* | , Normal data analysis and processing are executed.

3. | *ds1* | | *ds2* | *Ev1n - ev1n* = + 1 | And | *ev2n - ev2n* = + 1 |.

4. | *ds1* | > | *ds2* | *Ev1n - ev1n* = + 1 & *lt*; */ RTI & gt;* > | *ev2n - ev2n* = + 1 | , We oversample *ds2*

| *ds1* | > | *ds2* | *Ev1n - ev1n* = + 1 *& lt; / RTI & gt;* <| *ev2n - ev2n* = + 1 | , Downsample *ds1.*

| *ds1* | <| *ds2* | *Ev1n - ev1n* = + 1 *& lt; / RTI & gt;* > | *ev2n - ev2n* = + 1 |, If *ds2* is downsampled,

| *ds1* | <| *ds2* | *Ev1n - ev1n* = + 1 *& lt; / RTI & gt;* <| *ev2n - ev2n* = + 1 | If so, oversample *ds1.*

5. Execute data analysis and processing for ds1 and *ds2.*

Since the incoming data set in this study is stream type, there is no processing end. In other words, the sensor installed inside and outside the smart switchboard continuously detects the phenomenon accompanying the change of the surrounding situation and reports it to the host. The following figure shows the algorithm proposed in this study.

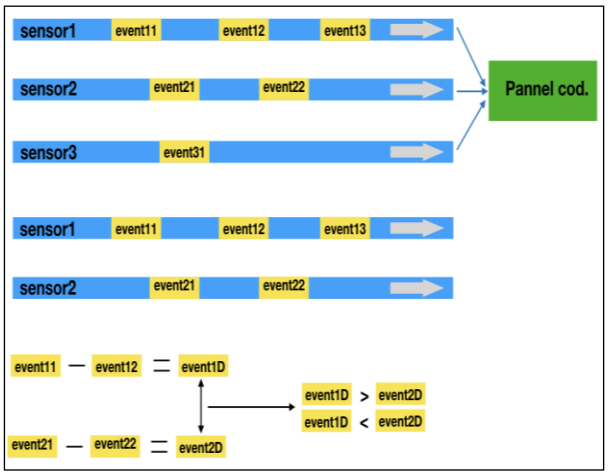

Fig. 1. Processing of unbalanced data for situational awareness in Smart Switchboards

It is necessary to determine which sensor data set is to be weighted in the data stream. In this study, the depth of the event is measured. The gap between the previous event and the current event is measured, and the increase in the depth of the event is determined to be larger. This suggests that the changing tendency of the event provides clear data that can be used to more clearly determine the situation.

IV. The Experiment and Evaluation

To test the proposed method, each sensor of temperature and humidity sensors continuously report the sensed value to the host. The data reported from the sensor to the host are summarized in the following figure.

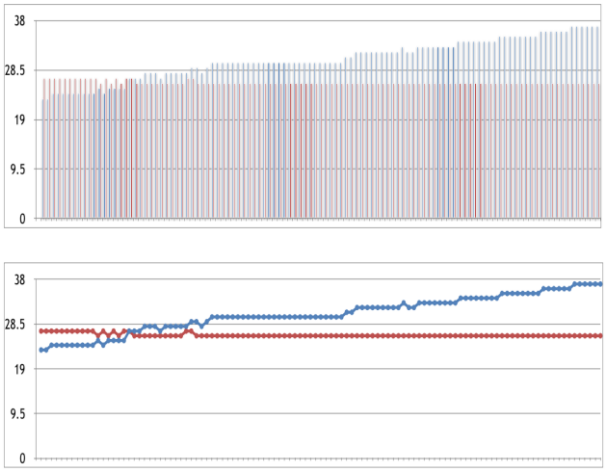

Fig. 2. Data reported from sensor

Temperature sensors and humidity sensors detect and report unbalanced data. Temperature sensor events are frequent. At this time, it is necessary to decide how to deal with the imbalance of event data. Applying the proposed algorithm, we can see that the data set sent by the temperature sensor is even bigger at the event depth.

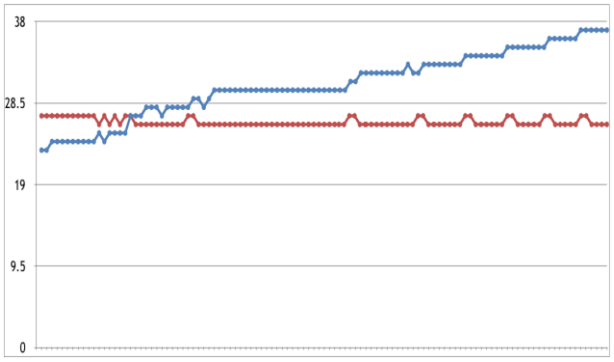

Fig. 3. Processing result for unbalanced data

The above table shows the result of oversampling the data of the humidity sensor. This can be beneficial if you need to process data for the same class. The results of this experiment can provide useful data for inferring the fatigue of electrical equipment in smart switchboards.

Electrical equipment fatigue is gradually accumulated and increases. When estimating the fatigue of electrical equipment, it is necessary to estimate the deterioration of the insulator. Temperature events and humidity events are accompanied by no humidity events in this experiment. This may be a sensitivity problem of the humidity sensor, or it may be another environmental issue. The sampling of imbalance data was applied to the deficit of the humidity sensor event which occurs if the event of the humidity sensor must be followed in the event of the temperature sensor according to the suggestion in this study. At this time, temperature sensor data with a

larger depth of event was not sampled, and oversampling of the humidity sensor data was performed. Through this, we could compensate for the gap between classes and process data.

V. CONCLUSION

It is good practice to use a wide variety of sensors when using sensors and networks to recognize real-world situations without human intervention. It is expected that the situation information can be improved when various kinds of sensors are used rather than using a single sensor. However, data that are detected and reported by different sensors often have unbalanced data attributes.

In this paper, we propose a method to manage the imbalance data in the smart switchboard system. Instead of applying uniform oversampling or downsampling, we propose a method to adjust the low - weighted data in the contextual information, while retaining the weighted data in the acquisition of contextual information. Considering that the goal of the smart system is to recognize the situation in the real world, it is reasonable to process the imbalance data considering the weight in the context information acquisition, and measure the depth of event in the data stream to support it.

REFERENCES

[1] J., and Ghemawat,S., “MapReduce: simplified data processing on large clusters”, Proc. conference on Symposium on Operating Systems Design & Implementation, pp.137-150, 2004

[2] Zaharia,M., Chowdhury,M., Das,T., Dave,A., Justin M., McCauly,M., Michael J. F., Shenker,S., and Stoica,I.,”Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing”, NSDI, pp.15-28, 2012

[3] Bhatotia,P., Dischinge,M., Rodrigues,R., and Acar,U.A., “Slider: Incremental Sliding-Window Computations for Large-Scale Data Analysis”, Middleware, pp.61-72, 2014

[4] Yook,M.S., Han,J.E., Lim,J.T., Bok,K.S.,and Yoo,J.S., “Incremental Query Processing Scheme Considering Load of Nodes in Distributed In-memory Environments”, Korea Information Science Society,pp.175-177,2015

[5] Valle, E. D. , Ceri, S. , Harmelen, F. v., and Fensel,D.,” It’ s a Streaming World! Reasoning upon Rapidly Changing Information”, IEEE Intelligent Systems, Vol. 24, pp83–89, 2009

[6] Kim,I.W., “A Study on Remote Monitoring Method of Power Facilities Using Smart Phones”, Master Thesis,pp.6-39,2011

[7] Lee,Y.D., and Jeng, S. H., “Design and Implementation of Busbar Contact and Temperature Sensing System", The Journal of the Korea Information and Communications Society, Vol.21,No.2, pp.380-384 ,2017

[8] Park,H.S., Yoon,J.S., Kim,T.O., and Kang,G.H., “ARC Monitoring System for Ship and Marine Switchboard”, The Korean Institute of Electrical Engineers, pp.1125-1126,2018

[9] IEEE Std 1584-2002, “IEEE Guide for Performing Arc-Flash Hazard Calculations”,2004

[10] Lee,S.J., Kim,Y.J., and Kim, Y.H., “structure design of the compact metal clad switchgear for improving the internal arc characteristics”, The Korean Institute of Electrical Engineers, pp.321-322,2016

[11] Koo,B.Y., and Chang,C.K., ”The Mitigation Method of High Energy Arc Fault(HEAF) damage in Nuclear Power Plant Onsite Low Voltage Switchgear”, The Korean Institute of Electrical Engineers, pp.295-296,2018