Advanced Science and Technology Letters
Vol.145 (NGCIT 2017), pp.1-5
http://dx.doi.org/10.14257/astl.2017.145.01

Abnormal Data Detection with CEP Engine for Smart Factory

Won-chang Lee[1], Jae-Han Cho[2] and LeeSub Lee[3]

[1,2,3] Kumoh National Institute of Technology
61, YangHo-Ro, Gumi, South Korea
{ [1] lwczzang, [2]jaehanfs, [3] eesub }@kumoh.ac.kr

Abstract. The smart factory is not suitable for real-time big data processing using a system such as Hadoop. Therefore, it is mandatory to develop a CEP based real-time analysis system for increasing data processing requirements. In the smart factory system, existing error processing is concentrated only on out of range errors. However, there are various kinds of errors. Accurate detection and handling of errors is an important part of the productivity of manufacturing.

This research proposes a real–time analysis system based on CEP (Complex Event Processing) which detects abnormal data in terms of time-series using Least Square Method(LSM). The proposed method will provide high performance real time detection of error data and the way of figure out the window size which generates optimum error the detection rates.

Keywords: Big data; Least Square Method; Error detection; Apache Storm; Mutation Testing.

1 Introduction

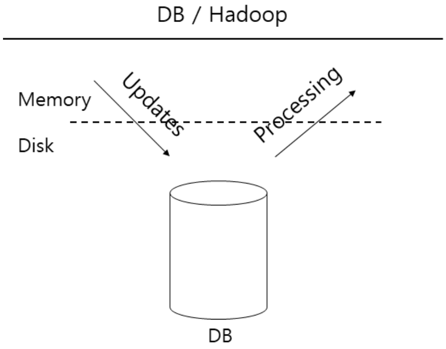
Nowadays, the smart factory is a hot issue. Since Hadoop is a framework of using total collection analysis, Big data processing using a system such as Hadoop is not suitable to the smart factory that requires real time processing [1]. New and existing facilities are also increasingly being converted to smart factories. Therefore, it is necessary to develop a CEP (Complex Event Processing) [2] based real-time analysis system for increasing data processing requirements. CEP is a recent emerging technology, mainly developed by Apache and Microsoft. This is a technique for real-time big data processing with a structure that is stored after analyzing unlike the existing big data framework.

Data is generated from multiple sensors in smart factory environment. Therefore, it is necessary to process a large amount of data in real time. As shown in the Figure 1, since Hadoop stores the received data in the database and then processes the data, it is not suitable for real-time processing because it is stored on disks. As shown in Figure 2, on the other hand, the CEP handles sensor data directly from memory without storing it on disk.

In this smart factory system, existing error processing is concentrated only on out-of-range error detection [3]. However, there are various kinds of errors. Accurate

ISSN: 2287-1233 ASTL

detection and handling of various errors is also an important part of the efficiency for manufacturing. This study suggests the utilizing LSM (Least Square Method) to detect abnormal data through time series.



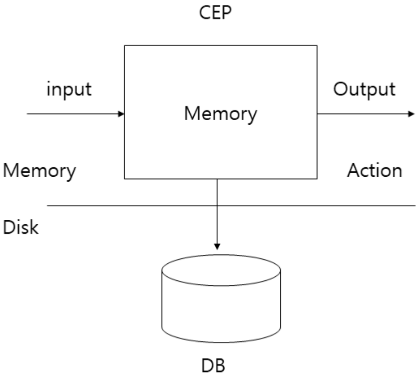


Fig. 1. Using Hadoop in Data Analysis

Fig. 2. Using CEP in Data Analysis

2 Abnormal Data Detection with CEP Engine for Smart Factory

The most popular CEP engines are Microsoft StreamInsight, Apache Spark, and Apache Storm. Microsoft StreamInsight has a problem that is dependent on Microsoft's environment [4]. Apache Spark uses a batch-oriented approach with Hadoop [5]. Apache Storm, an open-source software produced by Twitter, is a technology that allows large-scale data to be analyzed in real time. If Hadoop is a large-scale distributed processing system specialized for batch analysis, Storm is a distributed processing system specialized for real-time analysis. That is why we applied the Apache Storm [6].

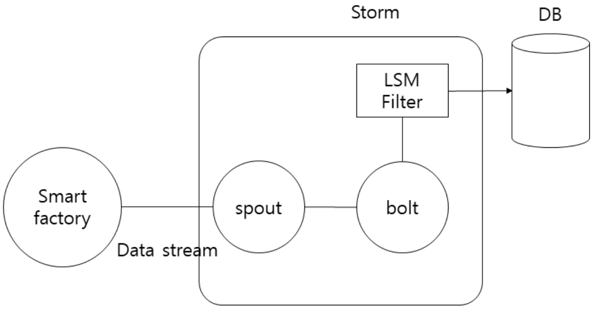


Fig. 3. System Architecture of the CEP based Analysis

Figure 3 shows the structure of the CEP based analysis system. The data generated by the sensors of the smart factories are transferred to the spout module where they are converted into tag value format data [7]. These data are consumed by the bolt and filtered by the filter attached to the bolt. This filter is made by applying LSM [8][9]. Consequently, the filtered sensor data is stored in the database and the subsequent processing proceeds.

3 Time Series Analysis for Abnormal Data Detection

There are various types of errors in the data generated in the manufacturing process. Previous studies have focused only on detection of data that is outside of the error tolerance range, but data on a pattern of time series error cannot be detected. As shown in Figure 4, the datum indicated by the arrow cannot be physically generated, so it can only be seen as the noise of the sensor. If this data is not filtered, the accuracy of the data may be a problem.

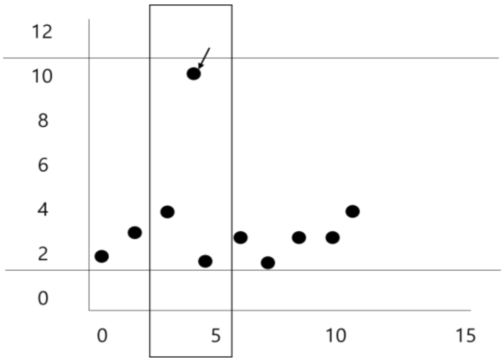


Fig. 4. An example of undetected Sensor data

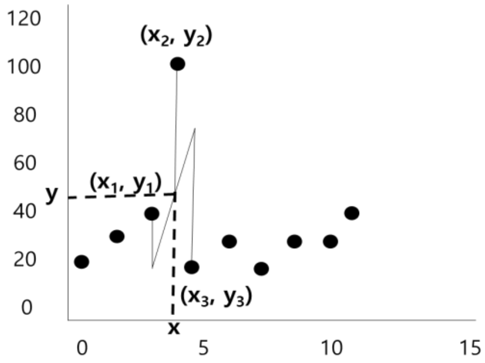


Fig. 5. An example of sensor data detection using LSM

Figure 5 shows how to filter these abnormal data using LSM. Assuming that the size of the window is 3, the LSM can calculate the coordinates (x, y) of the expected value by letting the average value of the X axis be x and the average value of the Y axis be y. After calculating the difference value between the expected value and the actual value of the fourth data, if the value exceeds the predetermined threshold, it is found that there is a problem with the data.

4 Conclusion

Recently, new facilities are being built as smart factories and the transition from existing facilities to smart factories are increasing. Therefore, demands for processing data generated from a large number of sensors are rapidly increasing. In this paper, we propose real-time big data processing of smart factories and detection of abnormal data using LSM with CEP engine suitable for smart factory The paper applies CEP based on Apache Storm which processes input data in memory.

There are various kinds of errors in sensor data generated in smart factories. However, existing researches focus only on detection of data outside the error range of data, so that abnormal data that can be detected in a time series cannot be filtered. The detection and processing of this error are directly related to the accuracy of the data. For this purpose, this study proposes a method to detect abnormal data using LSM.

Future research will include implementing the proposed system and verifying the accuracy of the model through simulation using mutation techniques. It also tests the various sizes of windows to determine the optimal window size. Although this study focuses on detecting abnormal data or noise from sensors, future studies will include methods to detect various types of abnormal data.

References

1. Xizoming Zhang, Guang Wang, “Hadoop-Based System Design for Website Intrusion Detection and Analysis, pp.1171-1174, 2015 IEEE international Conference on Smart city (2015)
2. Wenlu Yang, “Computing data quality indicators on Big Data streams using a CEP”, computational intelligence for Multimedia Understanding, pp.29-30, 2015 Internation workshop (Oct. 2015)
3. Marcus Kurth, Carsten Schleyer, Daniel Feuser, “Smart factory and education: An integrated automation concept”, pp.1057-1061, 2016 IEEE 11^{th} conference of Industrial Electronics and Applications (2016)
4. Mohamed Ali, Badrish Chandramouli, Jonathan Goldstein, Roman Schindlauer, “The Extensibility Framework in Microsoft StreamInsight “, IEEE, ICDE Conference (2011).
5. Seyoon Ko, Joong-Ho Won. "Processing large-scale data with Apache Spark", The Korean Jaurnal of Applied Statistice, pp.1077-1094. (2016).
6. SoonHyun Kwon, Dongwan Park, Hyochan Bang, Youngtack Park, “Real-time and Parallel Semantic Translation Technique for Large-Scale Streaming Sensor Data in an IoT Environment “, korea information science society, 1th, pp.1-10. (2007)

7. Qin Guo, jiwei Huang, “A complex event processing based approach of multi-sensor data fusion in IoT sensing systems”, pp.548-551, 2015 4th International Conference on computer science and network Technology (2015)

8. Bilawal Rehman, Chongru Liu and Lili Wang, “Least Square Method: A Novel Approach to Determine Symmetrical Components of Power System”, pp.39-44, J Electr Eng Technol(2017)

9. Haotian Chi, “A Discussions on the Least-Square Method in the Course of Error theory and Data Processing, pp.486-489, 2015 International Conference on computational intelligence and Communication Networks (2015)