Real-time Monitoring Solution to Detect Symptoms of System Anomalies

● Toshiya Hanamori ● Toshihiro Nishimura

Conventionally, system anomaly detection has been based on setting thresholds and rules. Today, as systems have diversified and the targets of analysis have become increasingly numerous and complicated, the large number of parameters to monitor has made it impossible for analysts to grasp the relationships between them. In addition, resolving these issues with data analysis requires advanced expertise, which hinders data utilization. Fujitsu's symptom monitoring solution provides a model for detecting symptoms of anomalies. It uses the results of applying machine learning to operational data in normal times to automatically detect any "state different from usual" and then narrows down the target data to identify the cause of the anomaly. This enables high-accuracy and real-time detection of symptoms of anomalies at the site level without requiring advanced analysis know-how. This paper presents a solution that enables real-time monitoring to detect problem symptoms such as information and communications technology (ICT) system failures and suspension of a manufacturing line due to failure of operating equipment.

1. Introduction

The New Industrial Revolution is putting the Internet of Things (IoT) and big data to practical use as typified by overseas initiatives such as Industrie 4.0 and Industrial Internet. In Japan, this transformation is becoming increasingly applicable to advancing machine-to-machine (M2M) systems and equipment in the manufacturing industry, enhancing industrial plants, visualizing human behavior, etc.

To achieve such sophisticated systems and equipment, the method used for monitoring of anomalies must likewise change: it must evolve from the setting of threshold values on the basis of an individual's skills to the application of advanced analytical techniques such as machine learning that combine previously collected data with data from newly installed sensors. In an information and communications technology (ICT) system, for example, this new form of monitoring could be used to learn common transitions in server operation, which would enable symptoms of a silent fault to be detected, such as when a message that is normally output under certain conditions fails to be output. Detection of fault symptoms in this way can prevent a system crash or other problems from occurring. This monitoring method could also be used to analyze correlation in fluctuations of load, temperature, and current on a manufacturing line to achieve early detection of abnormal conditions. This would enable maintenance to be performed before the line stops, thereby maintaining stable operation in a supply chain that includes one's own company.

An effective approach to analyzing large volumes of data in real time to detect anomalies is to combine in-memory big data processing such as by the increasingly popular Apache Spark engine with analysis by advanced machine learning. In addition, the key to successful anomaly detection is to construct a cycle that immediately applies the results of analysis to actual business operations and continuously feeds back the effects of doing so to the system.

2. Issues

A variety of analysis tools have recently become available as free and open source software (FOSS) in a form that can be used at low cost in the cloud. However, developing an analysis model using machine

learning requires a specialist such as a data scientist or data analyst having considerable know-how in the field of analysis. This has been a major factor hindering the effective use of data in frontline departments.

At the same time, maintaining high throughput and good detection response while ensuring data consistency in real-time big data processing can be a complicated process as compared to storage-type data processing. It would require, for example, a flow control method for handling a tremendous amount of incoming data, and a method for restarting a data-processing application after an interruption. Such a process cannot easily be achieved simply by combining individual products. For this reason, there are cases in which data processing is performed in batch instead of in real time despite a drop in detection response.

Furthermore, as the return on investment in analysis systems is unclear, many customers are hesitant to invest large sums of money in the development of data analysis schemes and construction of big data processing systems. There is thus a need for a data usage approach that begins with a low-cost "small start" requiring no advanced analysis know-how and that expands the scope of application in a step-by-step manner on the basis of the results.

3. Predictive Supervision solution

Fujitsu has developed FUJITSU Business Application Operational Data Management & Analytics (ODMA)[1] as a comprehensive set of solutions that enables frontline departments to use big data effectively. This includes an "ODMA Predictive Supervision" solution that provides technology for automatically creating a high-accuracy anomaly detection model by applying machine learning to data obtained during normal system operation. This solution also provides tools for data extraction, visualization, and other tasks to facilitate the use of this technology. In short, the ODMA Predictive Supervision solution systematizes a series of tasks needed for analysis. The anomaly detection model created in this way is provided in a state that can be executed on a real-time processing platform based on OSS such as Apache Spark or Elasticsearch and enhanced with Fujitsu operational know-how. This model can therefore be put to work immediately. The above features enable the user to perform the tasks from solution deployment to checking of results in a relatively short period of time.

In addition, the technology used for automatically creating the symptoms detection model in the ODMA Predictive Supervision solution implements a machine learning function that is a key component of Human Centric AI Zinrai,[2] Fujitsu's systematic approach to providing its artificial intelligence (AI) technologies developed and enhanced over time.

4. Solution configuration

Fujitsu's ODMA Predictive Supervision solution has a three-stage structure aimed at providing flexible services in line with customer objectives (**Figure 1**).

1) Streaming data processing platform

This platform enables basic functions such as data collection and processing, streaming data processing, and detection and assessment to be used on the basis of common operations and security measures. It includes an execution platform for in-memory processing and parallel distributed processing to accommodate huge volumes of data.

2) Analysis library

This library includes data models for processing, diverse analysis models for anomaly detection (unsupervised learning) and predictive analytics (supervised learning), and user interface (UI) components for screen displays.

3) Templates

These templates can be used for diverse purposes, such as to visualize real-time data, provide a dashboard display for presenting detected anomalies, and provide a comprehensive view of plant operating conditions. They can be customized to fit the target application and put to use quickly.

Analysis models included in the analysis library provide a way of using effective analysis techniques as needed from among a wide variety of techniques. Similarly, templates can be added as needed to support new types of customer business operations. The following section describes a streaming data processing platform as a common infrastructure for the Predictive Supervision solution and an anomaly detection function as the first analysis model to be provided.

5. Streaming data processing platform

The streaming data processing platform provides "analysis flow" and "operational know-how" based on

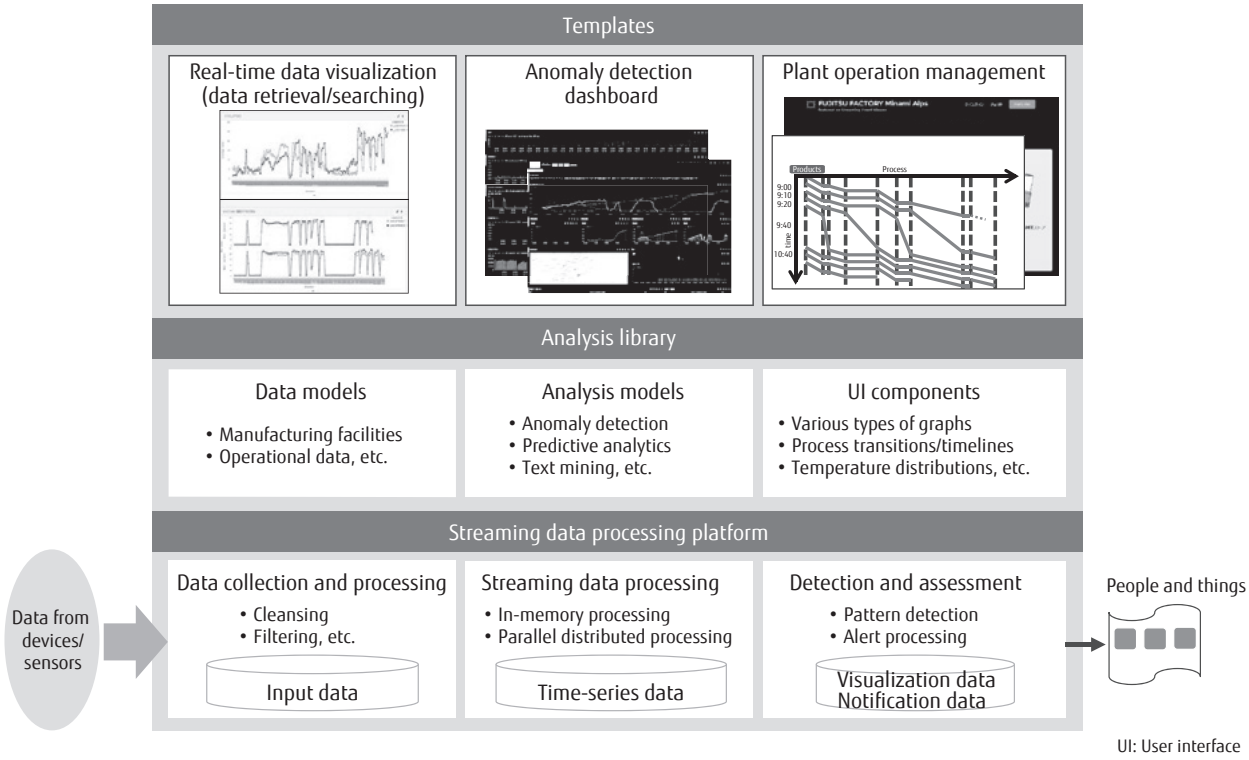

Figure 1
Configuration of ODMA Predictive Supervision solution.

open technologies as an easy-to-use integrated and general framework (**Figure 2**).

Fujitsu applies Apache Spark and Spark Streaming at the center of this platform. The former can perform in-memory parallel distributed processing while the latter can perform streaming processing above Apache Spark. In addition to being scalable in accordance with the amount of data being handled, this platform can perform a series of tasks from real-time processing and treatment of incoming data to data analysis. These are achieved through know-how accumulated in the development of Java or SQL software.

Fujitsu also applies Elasticsearch and Kibana to enable the flexible creation of dashboard displays to visualize incoming data and analysis results. The ODMA Predictive Supervision solution equips this OSS with operational know-how that includes non-functional requirements such as data guarantee and security.

6. Anomaly detection function

The word "anomaly" simply means a "state different from usual." Of course, an unusual state does not necessarily mean a problem, but it can be thought of as a state that may lead to a problem in the system.

The anomaly detection function can handle numerical data in time-series form with a fixed interval. However, not all targeted data will necessarily be obtained at fixed intervals or without some loss. In such cases, there is a need for a data interpolation operation prior to inputting the data into the anomaly detection function. This operation can be accomplished using the streaming data processing platform described above.

Anomaly detection can be divided into two phases: preparation and operation (**Figure 3**). The preparation phase is used to generate a model of the system's "usual state" by applying machine learning (pre-training) to stored data. This process uses various types of information including correlation between different types of stored sensor data to determine the overall system state at each time point of data storage and to calculate the occurrence probability and transition probability of each state.

The operation phase compares real-time incoming data with the model of the system's "usual state" to

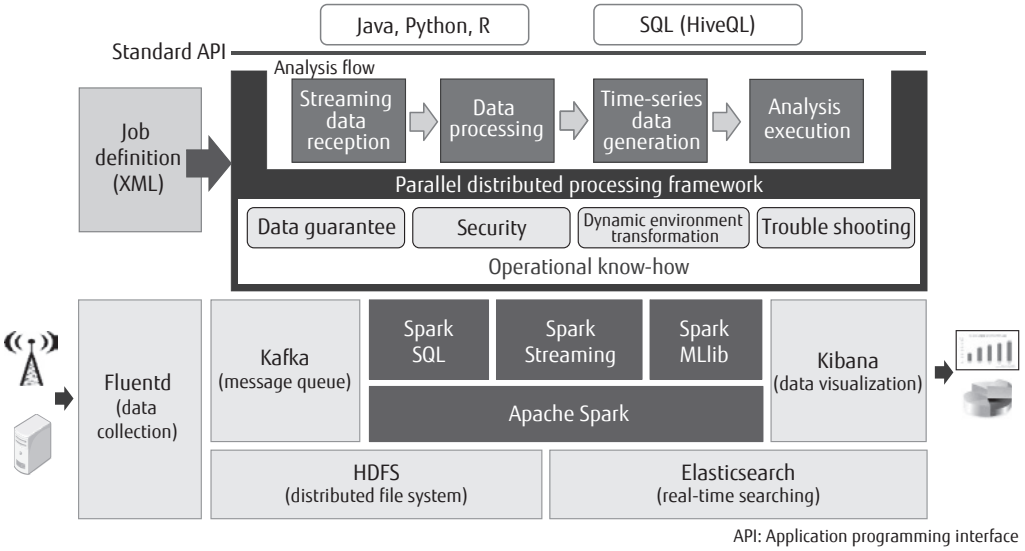

Figure 2
Streaming data processing platform.

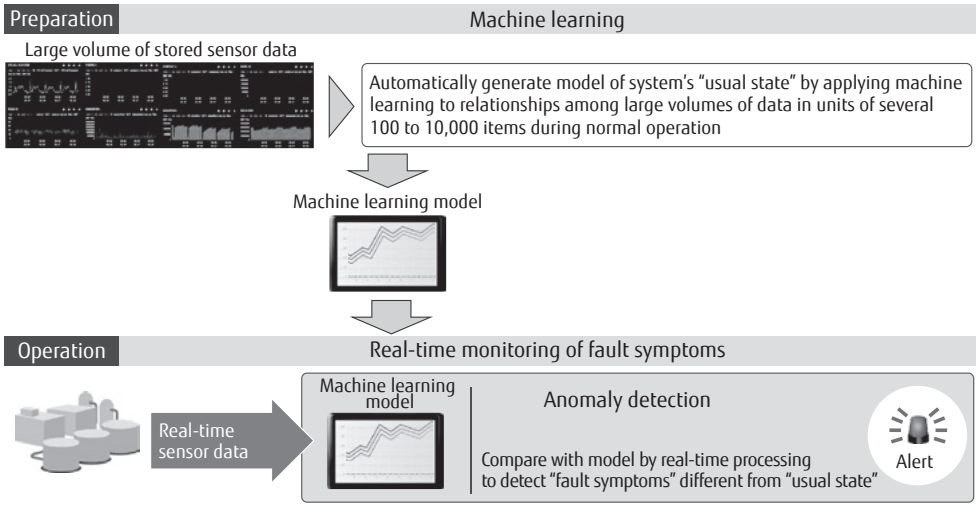

Figure 3
Anomaly detection function.

output the extent to which they differ in the form of an index called an "anomaly score." This index ranges in value from 0 to 1: a value close to 0 indicates a system close to its "usual state," and a value close to 1 indicates a system close to a "state different from usual." When the occurrence probability of the current state calculated from real-time incoming data is low and the transition probability from the prior state to the current state is low, the value is close to 1.

The anomaly score depends on the data used in pre-training. If the amount of this pre-training data is small, the possibility exists that a state that is essentially normal may nevertheless not be included in that data, resulting in the output of a high anomaly score when that state appears in real time. With this in mind, we have equipped the anomaly detection function with an online learning function so that training can be applied immediately to data input in real time. Online learning can be turned on and off. When turned on, a state that initially has a high anomaly score can

gradually take on a lower score if it continues to appear over time.

7. Proof of concept demonstrations

The streaming data processing platform has been implemented on the basis of knowledge gained from the proof of concept (PoC) demonstrations summarized below.

- Real-time processing and rule-based assessment of data collected by the FUJITSU Cloud Service IoT Platform[3] and real-time data visualization.
- Display of ship-related operating conditions such as course and fuel consumption based on data generated during the running of a ship.
- Collection of data generated by facilities and equipment on a factory's manufacturing line; real-time visualization of progress and stagnation in the manufacturing process in timeline form.

In addition, Fujitsu originally applied anomaly detection technology to server operating conditions in the monitoring of ICT equipment and achieved success in detecting states different from usual. At the time of this writing, PoC demonstrations using sensing data from plant and factory lines had begun in collaboration with several customers.

8. Conclusion

This paper introduced the functions and technologies of the Predictive Supervision solution as part of Fujitsu's Operational Data Management & Analytics (ODMA) comprehensive set of solutions. Fujitsu will first provide the anomaly detection technology portion of its ODMA Predictive Supervision solution as a function, but note that there are other real-time data analysis techniques that include symptoms detection. Indeed, there are both commonly used analysis techniques and proprietary analysis techniques developed by Fujitsu Laboratories. Fujitsu plans to put these technologies into a form that can be easily provided to customers while performing PoC demonstrations with the aim of providing an extensive and useful analysis library.

Although the ODMA Predictive Supervision solution constitutes a system that can process big data, it can be initially implemented as a small start. We ask any enterprise interested in the detection of symptoms of anomalies using big data to examine the features that the ODMA Predictive Supervision solution has to offer.

References

1) Fujitsu: Fujitsu Provides Big-Data Solutions to Transform the Front Lines of Business.
http://www.fujitsu.com/global/about/resources/news/press-releases/2014/0513-01.html
2) Fujitsu: Fujitsu Takes Systematic Approach to Artificial Intelligence with "Human Centric AI Zinrai."
http://www.fujitsu.com/global/about/resources/news/press-releases/2015/1102-01.html
3) Fujitsu: Fujitsu Releases FUJITSU Cloud IoT Platform, a Platform Service for IoT Data.
http://www.fujitsu.com/global/about/resources/news/press-releases/2015/0610-01.html

Toshiya Hanamori
Fujitsu Ltd.
Mr. Hanamori is currently engaged in the development of analysis solutions using IoT.

Toshihiro Nishimura
Fujitsu Ltd.
Mr. Nishimura is currently engaged in the development of analysis solutions using IoT.